

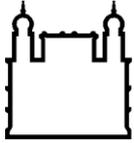
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Doutorado no Programa de Pós-Graduação em Biologia Computacional e Sistemas

MINERAÇÃO *IN SILICO* DE ENZIMAS ANÁLOGAS
ESSENCIAIS NOS GENOMAS DE PATÓGENOS BACTERIANOS E
FÚNGICOS DE *Glycine max*, *Zea mays* E *Solanum lycopersicum*

RANGELINE AZEVEDO DA SILVA

RIO DE JANEIRO
Setembro de 2018



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

RANGELINE AZEVEDO DA SILVA

MINERAÇÃO *IN SILICO* DE ENZIMAS ANÁLOGAS ESSENCIAIS NOS GENOMAS DE PATÓGENOS BACTERIANOS E FÚNGICOS DE *Glycine max*, *Zea mays* E *Solanum lycopersicum*

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutora em Ciências, área de concentração em Bioinformática.

Orientador: Prof. Dr. Antonio Basilio de Miranda

RIO DE JANEIRO

Setembro de 2018

Azevedo da Silva , Rangeline .

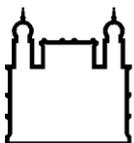
Mmineração in silico de enzimas análogas essenciais nos genomas de patógenos bacterianos e fúngicos de Glycine max, Zea mays e Solanum lycopersicum / Rangeline Azevedo da Silva . - , 2018.
135 f.

Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2018.

Orientador: Antonio Basílio de Miranda.

Bibliografia: Inclui Bibliografias.

1. Agrotóxicos. 2. Enzimas análogas. 3. Genômica comparativa. 4. NISEs.
I. Título.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: RANGELINE AZEVEDO DA SILVA

MINERAÇÃO *IN SILICO* DE ENZIMAS ANÁLOGAS ESSENCIAIS NOS
GENOMAS DE PATÓGENOS BACTERIANOS E FÚNGICOS DE *Glycine max*, *Zea
mays* E *Solanum lycopersicum*

ORIENTADOR: Prof. Dr. Antonio Basílio de Miranda

Aprovada em: 24/09/2018

EXAMINADORES:

Prof. Dr. Nicolas Carels - Presidente (CDTS, FIOCRUZ)

Prof. Dr. Ernesto Raúl Caffarena (PROCC, FIOCRUZ)

Prof. Dr. Diogo Antônio Tschoeke (UFRJ)

Prof. Dr. Floriano Paes Silva Junior (IOC, FIOCRUZ)

Prof. Dra. Adriana Machado Fróes (INPI)

Rio de Janeiro, 24 de Setembro de 2018

A Deus, toda honra e toda glória.

Com amor, à minha querida família, Rubens e Ana Maria
e meus irmãos Rangelson, Laiane,
Laielson e Ana Laura.

AGRADECIMENTOS

Bendiga o Senhor a minha alma! Não esqueça nenhuma de suas bênçãos (Salmos, 103:2). A Deus, obrigada por ter me amparado com seu grande Amor, Graça e Misericórdia. Jamais alguma palavra seria suficiente para expressar minha gratidão e meu amor.

Aos meus amados pais, Rubens e Ana Maria, meus maiores incentivadores, através dessa vitória quero honrar vocês. Obrigada por todo o amor, confiança, dedicação, renúncias, conselhos, cuidado, carinho, orações, por cada lágrima de saudade e por cada sorriso nos reencontros. Obrigada por terem me apoiado incondicionalmente, eu amo vocês.

Aos meus irmãos, Rangelson, Laiane, Laielson e Ana Laura, meus amores. Obrigada por todo o amor, carinho, partilha, cuidado, torcida, orações e por sempre contarem os dias para cada reencontro. Especialmente a Laiane, por ter acompanhado cada passo, cada lágrima, cada batalha vencida. Vocês são parte fundamental da minha vida e fazem parte dessa vitória. Amo vocês.

Obrigada a todos os meus familiares que estiveram torcendo, orando e vibrando por cada conquista, especialmente minha amada tia Solangen. Amo vocês.

À Sanlai e Nara, a distância fortaleceram nossos laços de amizade e reforçaram nosso amor. Obrigada por todas as vezes que me receberam em meio a sorrisos e saudades. Obrigada por terem suportado/compartilhado cada passo dessa minha caminhada, a verdadeira amizade dura pra sempre. Amo vocês.

À Lillian, amiga, irmã, obrigada por seu carinho, amor, companhia, seu apoio. Obrigada por sonhar junto comigo, sei que a minha felicidade também é a sua felicidade. Amo você.

Ao meu querido casal, Jonatas e Lucy, minha família. Agradeço por todo o carinho, amor, acolhimento, partilha, sorrisos, compreensão, orações e conselhos. Obrigada por terem segurado a minha mão quando mais precisei. Vocês fizeram meus dias mais alegres e leves e ajudaram a tornar esse sonho realidade. Amo vocês!

À minha querida amiga Gisele, obrigada por todas as orações, carinho, cuidado e sensibilidade. Obrigada por todas as conversas e cafezinhos que me ajudaram a sorrir, acreditar e olhar firmemente para nosso querido Pai. Amo você!

À Victoria, por sua sensibilidade, amor, carinho e amizade. Obrigada por todas as orações, apoio e torcida. Amo você, Vic!

À Juliana César, minha amiga e companheira de casa. Obrigada por sua companhia, carinho e confiança. Por ouvir meus desabafos e por sempre tentar me lembrar que eu era capaz e que eu conseguiria vencer esta etapa. Amo você!

Aos meus queridos amigos e irmãos da Igreja Presbiteriana do Grajaú, por terem me acolhido com tanto cuidado e amor. Especialmente ao Reverendo Alex e sua esposa Maria Claudia, por terem me acolhido como filha e com tanto carinho.

Às minhas queridas Médicas Arícia, Ana Paula, Fabrícia, Bárbara e Vanessa por terem me recebido e me acolhido no Rio de Janeiro, por terem cuidado de mim como uma irmã. Obrigada por me fazer sentir mais pertinho da Bahia, obrigada pelo carinho!

Ao meu orientador, Dr. Antonio Basílio de Miranda, obrigada por ter aceitado a orientação, por toda a paciência (principalmente com as tabelas rsrs), pela compreensão, sensibilidade, humanidade, amizade, sorrisos e conselhos pra vida! Muito obrigada por tudo! Sentirei saudades!

Ao Dr. Fábio Mota por toda a ajuda com os *scripts*, esclarecimentos, conselhos e todo o apoio. Obrigada!

Ao Dr. Rodrigo Jardim, pelos ensinamentos com lógica de programação, por solucionar todos os problemas com os *scripts*, com a rede, com a internet, com os computadores, servidores e pelas dicas e conselhos. Obrigada!

A Melise e Letícia, obrigada pelo apoio, pela companhia da jornada, conversas, desabafos, ajudas e sorrisos compartilhados. E ao André, pelos cafezinhos, conselhos e muitas risadas compartilhadas! Sentirei saudades!

Ao Leandro, pela amizade, por todas as conversas, desabafos e ajuda nas discussões, análises e conselhos. Muito obrigada!

À equipe do Laboratório de Biologia Computacional e Sistemas, Dr. Alberto D'Avila, Dr. Renata Schama, Fábio Bernardo e Ricardo Junqueira, obrigada pelo apoio!

Ao Instituto Oswaldo Cruz - FIOCRUZ pela infraestrutura necessária a realização deste trabalho.

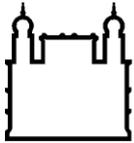
Ao programa de Pós-graduação em Biologia Computacional e Sistemas, a todos professores e funcionários, especialmente a secretária Rose Pani, por ter solucionado os problemas ao longo do curso com paciência, cordialidade e profissionalismo.

À equipe do Lauro Travassos, especialmente Gabriela, Cida, Davi, Leandro, Nadir, Tati, Sr. Rodrigues e Rogério, muito obrigada pelo apoio e os sorrisos todas as manhãs.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Obrigada a todos aqueles que em algum momento fizeram parte dessa etapa, torceram comigo e agora se alegram com mais uma conquista.

"Confia no Deus eterno de todo o seu coração e não se apoie na sua própria inteligência. Lembre-se de Deus em tudo o que fizer, e Ele lhe mostrará o caminho certo." (Provérbios 3:5-6)



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

MINERAÇÃO *IN SILICO* DE ENZIMAS ANÁLOGAS ESSENCIAIS NOS GENOMAS DE PATÓGENOS BACTERIANOS E FÚNGICOS DE *Glycine max*, *Zea mays* E *Solanum lycopersicum*

RESUMO

TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Rangeline Azevedo da Silva

Os últimos vinte anos foram de intenso crescimento na agricultura brasileira, com base na produtividade e expansão das fronteiras agrícolas. Diversas medidas de manejo das lavouras são empregadas, a exemplo da rotatividade de culturas, fertilização química, uso de cultivares resistentes e o uso de agrotóxicos. Os agrotóxicos são uma das principais medidas de controle de pragas e doenças mais amplamente usadas em cultivos de plantas e seu uso indiscriminado representa um risco direto à saúde das populações e ao ambiente em todo o mundo. Os riscos para a saúde humana incluem doenças agudas, como tonturas, enjoos, vômitos, infecções de pele e doenças crônicas a exemplo de malformação fetal, doenças neurológicas e neoplasias malignas. Os danos ao meio ambiente vão desde a contaminação dos corpos d'água até a morte de polinizadores e microbiota do solo. Como resultado, existe uma grande necessidade de desenvolvimento de novas moléculas menos tóxicas a serem empregadas contra patógenos de plantas. Neste trabalho, empregamos uma abordagem *in silico* para estudar os genes que codificam enzimas dos genomas de três plantas comercialmente importantes, soja (*Glycine max*), tomate (*Solanum lycopersicum*) e milho (*Zea mays*), bem como 15 patógenos de plantas (4 bactérias e 11 fungos), com foco em revelar um conjunto de enzimas análogas essenciais que poderiam ser priorizadas como alvos de drogas. Combinando sequência e dados estruturais, obtivemos um conjunto inicial de 568 casos de analogia, dos quais 97 foram validados e refinados, revelando um subconjunto de 29 atividades enzimáticas essenciais com um total de 119 formas estruturais diferentes, a maioria pertencentes a rotas metabólicas centrais, incluindo o metabolismo de carboidratos e de aminoácidos, entre outros. Além disso, outro subconjunto de 26 atividades enzimáticas possui uma estrutura terciária específica para o patógeno, não presente em plantas, homens e polinizadores, o que pode ser importante para o desenvolvimento de inibidores enzimáticos específicos contra doenças de plantas menos nocivas ao homem e ao meio ambiente.

INSTITUTO OSWALDO CRUZ

IN SILICO MINING OF ESSENTIAL ANALOGUE ENZYMES IN BACTERIAL AND FUNGIC PATHOGEN GENES OF *Glycine max*, *Zea mays* E *Solanum lycopersicum*

ABSTRACT

THESIS IN COMPUTATIONAL BIOLOGY AND SYSTEMS

Rangeline Azevedo da Silva

The last twenty years have been intense growth in Brazilian agriculture, based on productivity and expansion of agricultural frontiers. Various measures of crop management are employed, such as crop rotation, chemical fertilization, use of resistant cultivars and the use of pesticides. Pesticides are one of the most widely used pest and disease control measures in plant crops and their indiscriminate use represents a direct health risk to populations and the environment around the world. Risks to human health include acute diseases such as dizziness, nausea, vomiting, skin infections, and chronic diseases such as poor fetal formation, neurological diseases, and cancer. Damage to the environment ranges from the contamination of water bodies to the death of pollinators and soil microbiota. As a result, there is a great need for the development of novel, less toxic molecules to be employed against plant pathogens. In this work, we used an *in silico* approach to study genes that encode for enzymes derived from three commercially important genomes, soybean (*Glycine max*), tomato (*Solanum lycopersicum*) and maize (*Zea mays*), as well as 15 plant pathogens (4 bacteria and 11 fungi) focusing to reveal a set of essential and non-homologous isofunctional (NISEs) enzymes that could be prioritized as drug targets. Combining sequence and structural data, we obtained an initial set of 568 cases of analogy, out of which 97 were validated and refined, revealing a subset of 29 essential enzymatic activities with a total of 119 different structural forms, most belonging to the central metabolic pathways, including carbohydrate and amino acids metabolism, among others. In addition, another subset of 26 enzymatic activities has a specific tertiary structure for the pathogen, not present in plants, men and *Apis mellifera*, which may be important for the development of specific enzyme inhibitors against diseases of plants less harmful to humans and the environment

ÍNDICE

RESUMO	x
ABSTRACT.....	xi
ÍNDICE DE FIGURAS.....	xiv
ÍNDICE DE TABELAS.....	xv
LISTA DE SIGLAS E ABREVIATURAS.....	xvi
1. INTRODUÇÃO	19
1.1 Produção agrícola brasileira	19
1.2 Uso de agrotóxicos no Brasil	20
1.3 Agrotóxicos e a saúde humana.....	22
1.4 Agrotóxicos e meio ambiente.....	27
1.5 <i>Commodities</i> brasileiras.....	31
1.5.1 <i>Glycine max</i> L. Merr. (soja).....	31
1.5.2 <i>Zea mays</i> L. (milho)	32
1.5.3 <i>Solanum lycopersicum</i> L. (tomate).....	33
1.6 Genômica comparativa e a proposição de alvos	34
1.7 Enzimas isofuncionais não homólogas	36
1.8 Bancos de dados Biológicos	38
1.8.1 KEGG	38
1.8.2 SUPERFAMILY e SCOP	39
1.8.3 BRENDA.....	40
1.9 <i>Pipelines</i> de identificação de enzimas	41
1.9.1 Analogous Enzyme Pipeline (AnEnPi).....	41
2. OBJETIVOS.....	42
2.1 Objetivo Geral	42
2.2 Objetivos Específicos.....	42
3. Metodologia	43
3.1 Identificação de enzimas análogas	43
3.1.1 Organismos estudados	43
3.1.2 Identificação de enzimas iso-funcionais não homólogas e essenciais..	45

3.1.3 KEGG e clusterização.....	48
3.2 Inferência da função das proteínas preditas dos organismos	48
3.3 NISEs: identificação, validação estrutural e essencialidade	50
4. RESULTADOS	51
4.1 Preparação dos dados, clusterização e inferência da atividade funcional	51
4.2 Potenciais NISEs: identificação e validação	51
4.3 NISEs essenciais	53
4.4 Enzimas análogas no sistema antioxidante	58
4.5 Formas estruturais específicas das enzimas análogas.....	59
5. DISCUSSÃO	64
5.1 Essencialidade enzimática.....	64
5.1.2 Enzimas do Sistema Antioxidante.....	66
5.1.3 Enzimas do metabolismo central, metabolismo de carboidratos e metabolismo energético	68
5.2 Enzimas não essenciais.....	70
5.3 Enzimas descartadas.....	71
6. PERSPECTIVAS.....	72
7. CONCLUSÕES	73
8. REFERÊNCIAS BIBLIOGRÁFICAS	74
9. Arquivos suplementares.....	95
Tabela Suplementar 1. Validação das sequências de acordo com seus enovelamentos usando o banco de dados SUPERFAMILY e SCOP.	95
10. ANEXOS	105
10.1 Artigo relacionado à tese publicado em periódico Qualis A1.	105
10.2 Artigo publicado em colaboração durante o doutorado.....	124

ÍNDICE DE FIGURAS

Figura 1 Municípios que relataram poluição por agrotóxicos em água, Brasil, 2011. Fonte: Atlas do saneamento IBGE 2011.	30
Figura 2 A. Fluxograma: Identificação de enzimas iso-funcionais não homólogas e essenciais.	46
Figura 2 B. Fluxograma: Identificação de enzimas iso-funcionais não homólogas e essenciais.	47
Figura 3 Classificação funcional das NISEs validados. Os números entre parênteses indicam a quantidade de atividades enzimáticas essenciais.	53

LISTA DE TABELAS

Tabela 1. Classificação toxicológica dos agrotóxicos em função do DL50. Em toxicologia, DL significa Dose Letal Média, e é definido como a dose necessária de uma determinada substância para matar 50% de uma população em estudo. Adaptado; (Copplestone, 1988; World Health Organization, 1991).	24
Tabela 2. Principais culturas brasileiras e seus respectivos polinizadores. Adaptado Cham <i>et al.</i> ,2017).	28
Tabela 3. Descrição das doenças e sintomas causados pelos fitopatógenos avaliados neste estudo.	44
Tabela 4. Descrição dos conjuntos de dados das proteínas preditas dos organismos incluídos neste estudo.	49
Tabela 5 . Número de NISEs potenciais, validadas, específicas e essenciais. Os números entre parênteses indicam o número de atividades enzimáticas identificadas no total.	52
Tabela 6 Enzimas análogas e essenciais.....	55
Tabela 7 Formas enzimáticas alternativas encontradas entre as enzimas do sistema antioxidante.	59
Tabela 8. Formas estruturais enzimáticas específicas dos fitopatógenos.....	61

LISTA DE SIGLAS E ABREVIATURAS

ABRASCO - Associação Brasileira de Saúde Coletiva

Af - *Aspergillus flavus*

AnEpl - *Analogous Enzyme Pipeline*

ANVISA - Agência Nacional de Vigilância Sanitária

Bc - *Botrytis cinerea*

CA - Anidrase carbônica

CAT- Catalase

cDNA - DNA complementar

DEG- *Database Essencial Genes*

DNA - *Deoxyribonucleic acid*

DTC - ditiocarbamatos

EC - *Enzyme Commission*

Et - *Exserohilum turcicum*

EST - *Expresse Sequence Tags*

FAO - *Food and Agriculture Organization*

Fo - *Fsuarium oxysporum*

Gg- *Colletotrichum graminicola.*

GLX-I - Lactoilglutaciona liase

Gm - *Glycine Max*

Gm - *Giberella moniliformis*

GR- Glutaciona redutase

HEPH - Ferroxidase

HMM - *Hidden Models Markov*

IARC - *International Agency for Research on Cancer*

IBAMA - Instituto Brasileiro do Meio Ambiente

IBGE - Instituto Brasileiro de Geografia e Estatística

IUBMB - *International Union Of Biochemistry And Molecular Biology) Enzyme Nomenclature Committee*

KEGG- *Kyoto Encyclopedia of Genes and Genomes*

LMR - Limite Máximo de Resíduo

MAPA - Ministério da Agricultura Pecuária e Abastecimento

Mp - *Moniliophthora perniciosa*

NISEs - *Non-homologous Isofunctional Enzymes*

Pa - *Pantoea ananatis*

PARA - Programa de Análise de Resíduos de Agrotóxicos em Alimentos

PDB - *Protein Databank*

PDTD - *Potential Drug Target Database*

POX - Peroxidase

PR- Pathogenesis related

PRDX - Peroxiredoxina

Ps - *Phytophthora sojae*

Psy - *Pseudomonas syringae*

ROS - *Reactive oxygen species*

Rpe - Fosfato epimerase

Rpi - Ribose-5-fosfato isomerase

Rs - *Raslstonia solanacearum*

SA - Sistema Antioxidante

SCOP- *Structural Classification of Proteins*

SINITOX - Sistema Nacional de Informações Tóxico-Farmacológicas

Sl - *Solanum lycopersicum*

SNPs - *Single Nucleotide Polymorphism*

SOD - Superóxido dismutase

Ss - *Sclerotinia sclerotiorum*

SSR - *Simple Sequence Repeat*

Uniprot - Universal Protein knowledgebase

U.S.EPA - *United States Environmental Protection Agency*

Xa - *Xantomonas axonopodis*

Zm - *Zea mays*

1. INTRODUÇÃO

1.1 Produção agrícola brasileira

No cenário da agricultura brasileira, os últimos vinte anos foram de intenso crescimento com base na produtividade, na expansão e consolidação das fronteiras agrícolas, principalmente nas regiões Centro-Oeste e Norte do país. Apesar do mercado interno absorver a maior parte da produção, esse crescimento foi impulsionado principalmente pela expansão da produção destinada à exportação [1, 2].

Nos anos de 2016/2017 a agricultura brasileira vivenciou um salto, com uma supersafra de 238 milhões de toneladas de grãos, registrando um marco histórico, consolidando o país como um destaque no agronegócio mundial [3]. Segundo dados divulgados pela Secretaria de Relações Internacionais do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), as exportações do agronegócio brasileiro alcançaram a marca de U\$ 9,21 bilhões em junho de 2018, os grãos soja representam um dos principais destaques do setor atingindo a marca recorde para o mesmo período alcançando 10,42 milhões de toneladas [4]. Ainda de acordo com dados MAPA, o Brasil apresenta uma das maiores taxas de produtividade quando comparado a outros países, com potencial de expansão da sua área agrícola em 70 milhões de hectares, onde seria possível obter incrementos de 136% em relação à atual produção nacional de grãos e fibras [5].

Segundo relatório da Divisão de População do Departamento de Assuntos Econômicos e Sociais (DESA) das Nações Unidas, o crescimento das atividades agrícolas está diretamente relacionado com o crescimento da população mundial, as projeções para os próximos anos estimam que a população passará de 7,4 bilhões em 2016 para 8,5 bilhões em 2030, sendo que o maior incremento será proveniente de países em desenvolvimento [6]. Assim, para atender a demanda mundial por alimentos, houve um aumento nas pesquisas no campo das Ciências Agrícolas, onde o principal desafio dos melhoristas de plantas é o desenvolvimento de uma

agricultura sustentável, focada na qualidade dos produtos gerados e diminuição dos danos causados ao meio ambiente.

1.2 Uso de agrotóxicos no Brasil

As medidas atuais de manejo de culturas incluem cultivares resistentes, rotação de culturas, manejo de solo, fertilização química, irrigação e controle de pragas e doenças [7]. Os agrotóxicos são uma das alternativas mais utilizadas para controle de pragas e doenças que causam grandes perdas anuais. Segundo dados divulgados pela Agência Nacional de Vigilância Sanitária (ANVISA) e do Observatório da Indústria dos Agrotóxicos da Universidade Federal do Paraná, realizado em Brasília, DF, em abril de 2012, desde 2008, o Brasil ocupa o primeiro lugar no *ranking* mundial de consumo de agrotóxicos: o crescimento do setor foi de 190%, enquanto o aumento do consumo mundial foi de 93% [8].

De acordo com a Lei nº 7.802, de 11 de julho de 1989 os agrotóxicos ou designações afins são definidos como:

"...produtos e os agentes de processos físicos, químicos ou biológicos, destinados ao uso nos setores de produção, no armazenamento e beneficiamento de produtos agrícolas, nas pastagens, na proteção de florestas, nativas ou implantadas, e de outros ecossistemas, e também de ambientes urbanos, hídricos e industriais, cuja finalidade seja alterar a composição da flora ou da fauna, a fim de preservá-las da ação danosa de seres vivos considerados nocivos" [9].

Os agrotóxicos ou defensivos agrícolas, utilizados para o controle fitossanitário das mais diversas doenças de plantas, são geralmente produtos com elevados custos e riscos ambientais e têm se mantido em evidência devido à resposta rápida que proporcionam, bem como por não exigir conhecimento detalhado dos patógenos envolvidos [10]. O uso indiscriminado de agrotóxicos

provoca o acúmulo de substâncias nocivas no solo e na água, leva ao surgimento de populações de fitopatógenos e pragas resistentes aos compostos químicos, além do desequilíbrio ambiental pela falta de seletividade dos produtos utilizados [11,12]. Assim, a utilização do manejo ecológico de doenças é preconizada com o objetivo de reduzir o uso de bactericidas, fungicidas, herbicidas, dentre outros [13].

Desde 2008, o Brasil é o maior consumidor mundial de agrotóxicos, superando a sua marca a cada ano; Entre 2000 e 2012 o aumento do uso de agrotóxicos foi de 288%. O aumento dos índices de consumo está diretamente relacionado com a expansão das lavouras de culturas transgênicas, como a soja e milho. Em 1994 foram consumidas cerca de 800 toneladas de herbicidas, coincidindo com a liberação e introdução da soja transgênica no país [14]. Em 2009, dos seis pedidos de liberação de sementes com transgenia, quatro sementes foram modificadas geneticamente para tolerar o uso de herbicidas e outros agrotóxicos [15]. Recentemente, em 2014, a Associação Brasileira da Indústria Química (ABIQUIM) contabilizou um incremento de 13% nas vendas de agrotóxicos no Brasil com faturamento de R\$ 25 bilhões [16]. Dados do Censo Agropecuário 2017 mostram que aproximadamente 1.681.000 produtores utilizaram agrotóxicos, ressaltando que não houve necessidade de aplicação no período de referência, representando um crescimento de 20,4% em relação a 2006, quando 1.396.077 produtores declararam ter feito uso de agrotóxicos [17].

A ANVISA possui um Programa de Análise de Resíduos de Agrotóxicos em Alimentos (PARA), onde relatórios apontam irregularidades para o uso de agrotóxicos para os principais alimentos consumidos pela população. Segundo o acompanhamento do PARA 2011/2012, 36% das amostras de 2011 e 29% das amostras alimentares de 2012 apresentaram resultados insatisfatórios; a pesquisa foi realizada com o levantamento de amostras de todas as 26 unidades federativas do Brasil. As irregularidades são divididas em dois grupos, uma quando a amostra contém agrotóxico acima do Limite Máximo de Resíduo (LMR) permitido e outra quando a amostra apresenta resíduos de agrotóxicos não permitidos por lei para aquele alimento. Das amostras insatisfatórias, cerca de 30% se referem a agrotóxicos em que seu uso está sendo reavaliado pela ANVISA, incluindo casos de agrotóxicos em que o uso já foi extinto em diversos países, devido à sua elevada toxicidade [18].

Os resultados do período de 2013 a 2015 apresentam uma melhora nos resultados, mas é importante ressaltar que não entram soja e milho (forma *in natura*) nestes estudos, culturas com os maiores consumo de agrotóxicos do país. Na avaliação mais recente do PARA foram analisadas 12.051 amostras de 25 tipos de alimentos divididos em seis categorias: raiz, tubérculo e bulbo; cereal/leguminosa; hortaliça folhosa; hortaliça não folhosa e fruta. Um total de 9.680 (80,3%) amostras foram consideradas satisfatórias e 4.618 (38,3%) amostras apresentaram resíduos abaixo do LMR. O número de amostras consideradas insatisfatórias foi de 2.371 (19,7%). Do total de 232 agrotóxicos pesquisados, 134 foram detectados nos alimentos [16].

1.3 Agrotóxicos e a saúde humana

Os impactos na saúde pública atingem diretamente diferentes grupos populacionais, desde trabalhadores do campo, moradores do entorno das áreas de agricultura e os consumidores. Segundo o Sistema Nacional de Informações Tóxico-Farmacológicas (Sinitox), órgão ligado à Fundação Oswaldo Cruz (Fiocruz), entre 1999 a 2016 foram registradas 2947 mortes decorrentes de intoxicações causadas por agrotóxicos de uso agrícola no Brasil [20]. De acordo com dados do Sistema de Informação de Agravos de Notificação (Sinan) do Ministério da Saúde, os dados referentes à intoxicação por defensivos agrícolas no período compreendido entre 2007 a 2017 são de aproximadamente 40 mil pessoas, deste total 26 mil pacientes tiveram intoxicação confirmada [21]. Esses dados são subestimados, pois nem todas as ocorrências são registradas; a Organização Mundial da Saúde estima que, para cada caso de intoxicação por agrotóxico notificado, outros 50 casos acabam sendo omitidos ou não registrados [22-24].

A exposição ocupacional aos agrotóxicos é responsável por intoxicações agudas, causando problemas respiratórios e cutâneos, cólicas, vômitos, diarreias, convulsões e morte. Os produtores rurais de pequenas propriedades detêm o maior número de casos de intoxicações registradas, uma vez que não possuem conhecimentos prévios sobre a manipulação adequada dos defensivos agrícolas

[25]. Os demais grupos populacionais sofrem com as intoxicações crônicas, sendo expostos a diversos tipos de agrotóxicos presentes nos alimentos e no meio ambiente [26]. Os sintomas por vezes são tardios, o que dificulta a correlação com os agentes, além de apresentarem efeitos cumulativos que podem causar infertilidade, impotência, abortos, malformações, neurotoxicidade, desregulação hormonal e o surgimento de câncer [27].

A maioria dos pequenos e médios agricultores desconhece a forma adequada de manuseio e destinação de embalagens de agrotóxicos, e o tempo que o princípio ativo permanece agindo na planta, bem como os efeitos nocivos dos mesmos. Assim, a população acaba consumindo níveis elevados de agrotóxicos [28]. Medidas alternativas ao seu uso precisam ser amplamente estudadas e divulgadas visando à diminuição dos danos ao meio ambiente e à saúde humana. Diversos fatores contribuem para o aumento dos níveis de contaminação ambiental e intoxicação humana por agrotóxicos, entre os quais podemos citar o nível de escolaridade dos agricultores, que está diretamente relacionado ao desconhecimento de técnicas de manejo de cultura, descarte de resíduos contaminantes, protocolo de aplicação do produto, uso de equipamentos de proteção individual, dentre outros, além da deficiência de políticas públicas para o controle da venda destes produtos [29].

Os agrotóxicos podem ser classificados de acordo com o grupo de organismos para o qual é direcionado (pragas ou patógenos), segundo a estrutura química das substâncias ativas e de acordo com a toxicidade para o homem e o meio ambiente (Tabela 1).

Tabela 1. Classificação toxicológica dos agrotóxicos em função do DL50. Em toxicologia, DL significa Dose Letal Média, e é definido como a dose necessária de uma determinada substância para matar 50% de uma população em estudo. Adaptado [30,31].

Classe toxicológica	Formulação DL50 Oral (mg/kg)	Faixa indicativa de cor	Dose letal em humanos (quantidade aproximada)
I	Extremamente tóxicos (DL50 < 50 mg/kg de peso vivo)	Vermelho vivo	Algumas gotas
II	Muito tóxicos (DL50 – 50 a 500 mg/kg de peso vivo)	Amarelo intenso	1 - 2 col. Chá
III	Moderadamente tóxicos (DL50 – 500 a 5000 mg/kg de peso vivo)	Azul intenso	2 col. Sopa - 1 copo
IV	Pouco tóxicos (DL50 > 5000 mg/kg de peso)	Verde intenso	1 copo – 1 litro

Os principais tipos de agrotóxicos, de acordo com o grupo de organismos para os quais são direcionados são: i) os inseticidas, controle de insetos; ii) fungicidas, controle de infecções fúngicas; iii) bactericidas, ação contra bactérias (partes aéreas das plantas); iv) herbicidas, controle do crescimento de plantas invasoras; v) fumigantes, ação contra bactérias do solo; vi) raticidas, combatem ratos e outros roedores; vii) moluscocidas, combatem moluscos; viii) nematicidas, ação contra nematóides; ix) acaricidas, controle de proliferação de ácaros (Peres, et al 2003). Quanto aos grupos químicos, são classificados em organofosforados (Malation), organoclorados (DDT), carbamatos (Carbaril), piretróides (Permetrina), dipiridilos, fenoxi-acéticos, dinitrofenóis e clorofenóis, arseniais e inorgânicos, mercuriais orgânicos, dentre outros [32].

Os defensivos agrícolas mencionados a seguir foram encontrados nos alimentos analisados pelo PARA da ANVISA, tanto em níveis acima dos valores do LMR permitidos ou em culturas para as quais não são destinados. O LMR é um índice agrônomo, estabelecido pela ANVISA que delimita a quantidade máxima de resíduo de agrotóxico permitido no alimento, levando em consideração parâmetros

como a aplicação adequada em fase específica desta cultura, desde a cadeia produtiva até o consumo [33]

Os ditiocarbamatos (DTC) são agrotóxicos pertencentes à classe de fungicidas, com classificação toxicológica e ambiental I, usados nas lavouras de soja, milho, tomate, arroz, feijão, batata, café, dentre outros. Diversos estudos apontaram evidências da atuação do DTC no metabolismo celular, que é modificado por interações diretas com diferentes moléculas, como por exemplo proteínas de sinalização, peptídeos e enzimas. O DTC atua no metabolismo oxidoreduzidor das células, inibindo peróxidos e íons superóxidos. Suas propriedades quelantes de metais contribuem para seus efeitos pro-oxidativos, podendo influenciar nas funções de enzimas dependentes de metais tais como zinco e cobre. Em humanos, das três proteínas da classe superóxido dismutase (SOD) presentes no genoma, duas são dependentes de metais. As células em contato com o DTC sofrem estresse oxidativo e desregulações metabólicas que conduzem a danos nos tecidos e apoptose. A família da SOD contribui para a defesa de linha de frente através da desintoxicação de ânions radicais superóxidos reativos. São pertencentes às famílias das metaloenzimas, ocorrendo em três formas moleculares diferentes, e contêm os metais Mn, Fe ou Cu / Zn como grupos prostéticos [34]. Como todas as metaloproteínas, as enzimas SOD enfrentam os desafios de adquirir o cofator apropriado em condições de disponibilidade limitada de metal. Estas formas não possuem o nível de similaridade de sequência de aminoácidos, mas convergiram para catalisar a dismutação de superóxido evolutivamente. O radical superóxido (O_2^-), participa ativamente das respostas de defesa das plantas, inclusive atuando na sinalização celular [35-37]. No entanto, em situações de acumulação de radicais O_2^- , SOD atua sobre a molécula de peróxido de hidrogênio (H_2O_2) e oxigênio molecular (O_2) ($2O_2 + 2H_2O \rightarrow O_2 + H_2O_2 + 2H_2O$), e H_2O_2 é convertido em água (H_2O) para APX (Ascorbato Peroxidase), Catalase (CAT) ou Peroxidase (POX), para eliminar o estresse oxidativo nas células [38]. Estudos usando *Fusarium graminearum* como modelo revelaram a importância da SOD no crescimento das hifas e na virulência de *F. graminearum*. Os resultados dos experimentos apontaram que a germinação dos conídios sob estresse oxidativo extracelular foi significativamente retardada nos mutantes para SOD_1 e que a produção de desoxinivalenol mediada por SOD_1 contribui para a virulência de *F. graminearum* durante a infecção em trigo [39].

O procloraz é um fungicida pertencente ao grupo químico imidazolilcarboxamida com classificação toxicológica I, utilizado nas lavouras de tomate, trigo, cenoura, cevada, manga e mamão. O composto é absorvido via sistema respiratório, via cutânea e mucosas. É um desregulador endócrino, que afeta a produção e síntese de hormônios corticosteróides e sexuais masculinos e femininos, sendo os principais danos relacionados ao aparecimento de malformações fetais (observado em ratos), infertilidade, metabolismo de nutrientes e regulação do sistemas imunológicos [40-46].

O clorotaonil pertence ao grupo químico benzimidazol e Isoftalonitrila, inserido na classe toxicológica II, e é um fungicida de amplo espectro utilizado nas lavouras de tomate, soja, banana, batata, dentre outros. É classificado pela Agência Ambiental Norte-Americana (U.S.EPA) como provável carcinógeno humano (Grupo B2) e pelo *International Agency for Research on Cancer* (IARC) como possível carcinógeno humano (Grupo 2B) [47].

Carbendazim é um fungicida (classe toxicológica III), utilizado nas lavouras de soja, feijão, trigo, algodão e citros. Estudos comprovaram que a contaminação por carbendazim causa aberrações cromossômicas, apresentando forte interação com os microtúbulos celulares e inibindo funções vitais como o transporte intracelular e a divisão celular [48-49]. Esse agrotóxico ainda foi apontado como um dos fatores responsáveis pela desregulação endócrina do sistema reprodutivo masculino de ratos [50-54].

Outro grupo de agrotóxicos são os triazóis, fungicidas de amplo espectro, com modo de ação sistêmico e mesostêmico. Em 2017 o Ministério da agricultura Pecuária e Abastecimento (MAPA), suspendeu o uso de 63 fungicidas utilizados contra a ferrugem asiática (*Phakopsora pachyrhizi*) na cultura da soja. Destes, a metade possui o ingrediente ativo Tebuconazol, pertencente ao grupo químico Triazol. A medida foi aplicada devido à eficiência reduzida por adaptação do fungo [55]. O uso desses agrotóxicos era recomendado para as culturas de soja, milho, tomate, arroz, cenoura, trigo, feijão, dentre outros. De acordo com a bula dos agrotóxicos à base de triazol os mecanismos de toxicidade em humanos não são conhecidos, com classe toxicológica II para o meio ambiente e III para os seres humanos. No entanto, pesquisas apontam que este produto interfere na produção dos

hormônios sexuais feminino e masculino em análises utilizando sistemas *in vitro* de linhagens celulares humanas e *in vivo* [41, 56, 57].

Os riscos para a saúde humana apontados pela exposição aos agrotóxicos são extensos, partindo do princípio da existência de múltiplas formas de contaminações envolvendo diversas classes de defensivos agrícolas. As despesas concernentes com saúde pública poderiam ser melhor remanejadas ou reduzidas caso houvessem estímulos por parte de órgãos responsáveis, no intuito de informar e divulgar sobre uso consciente dos agrotóxicos.

1.4 Agrotóxicos e meio ambiente

Segundo a Associação Brasileira de Saúde Coletiva (ABRASCO) dos 50 produtos mais utilizados nas lavouras brasileiras, 22 são proibidos na União Europeia, Estados Unidos e outros países, onde foram comprovados seus potenciais riscos à saúde humana e ao meio ambiente [58].

O modo de aplicação dos agrotóxicos pode ser determinante para a dispersão de partículas para as regiões circunvizinhas às áreas agricultáveis, atingindo organismos não alvos e influenciando diretamente na desestruturação da biodiversidade. Um dos organismos atingidos diretamente pela aplicação de defensivos agrícolas são os polinizadores, que são peças fundamentais para o correto funcionamento dos ecossistemas [59]. Durante a reunião da Plataforma Intergovernamental sobre Biodiversidade e Serviços de Ecossistêmicos (IPBES) realizada na Turquia em 2014, foram divulgados documentos que apontam que ao menos 75% das culturas do mundo dependem da polinização por abelhas e outros polinizadores para gerar frutos [60, 61]. Em 2017 o IBAMA (Instituto Brasileiro do Meio Ambiente) lançou o Manual de avaliação de risco ambiental de agrotóxicos para abelhas. O protocolo objetivou apresentar as bases conceituais da avaliação de risco de agrotóxicos e as orientações gerais sobre os novos procedimentos de avaliação. O documento também chamou a atenção para as principais culturas de plantas e seus polinizadores (Tabela 2). A ação do IBAMA é uma forma de avaliação

mais abrangente para que os produtos atendam as diretrizes e exigências da área ambiental, utilizando os agrotóxicos de maneira racional e sustentável [62].

Tabela 2. Principais culturas brasileiras e seus respectivos polinizadores. Adaptado [62].

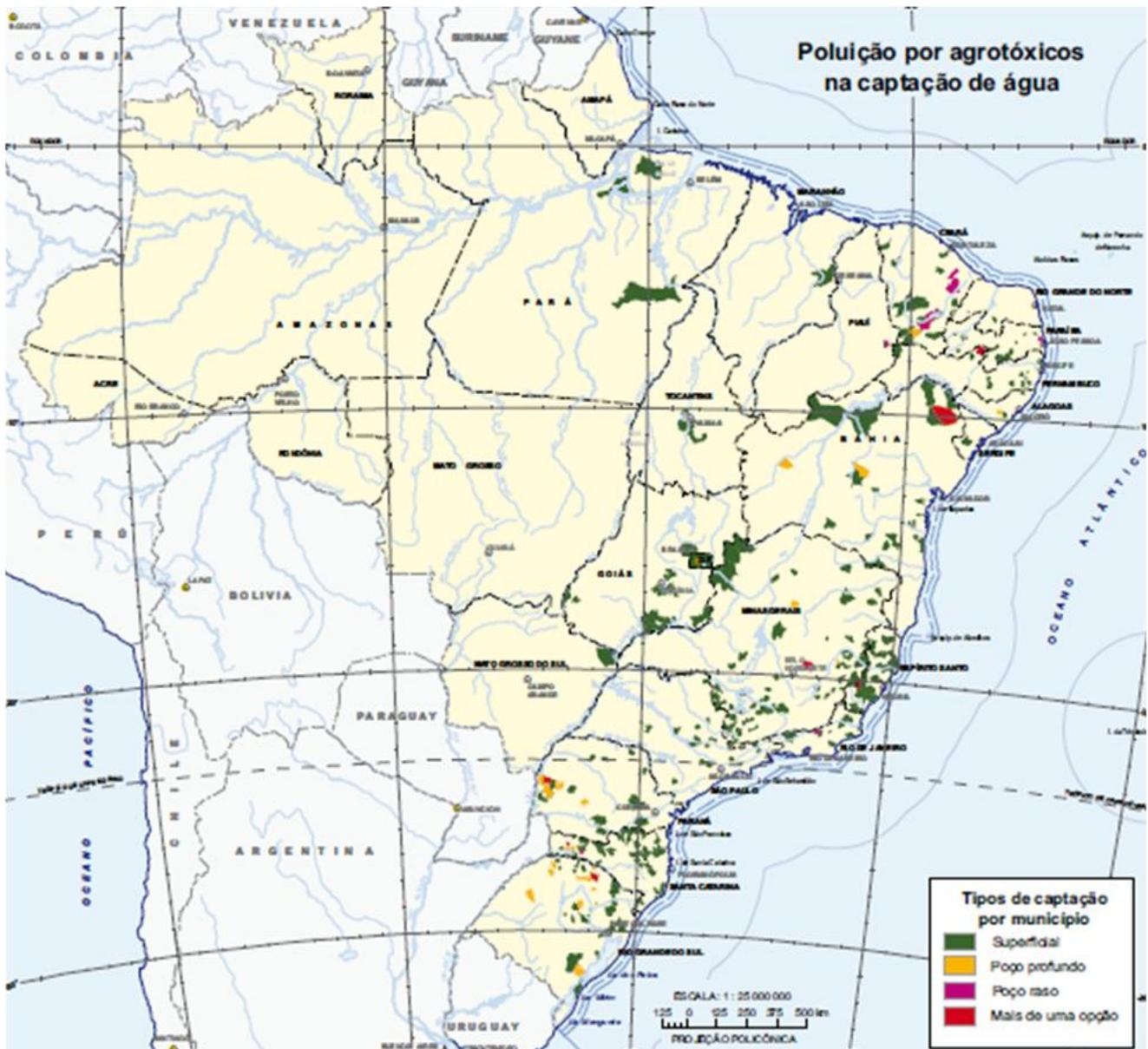
Culturas	Ordens, famílias, gêneros, espécies visitantes das culturas	Referências
Tomate (<i>S. lycopersicum</i>)	<i>Amegilla chlorocyanea/ Apis mellifera/ Augochlora sp./ Augochlorini spp./ Augochloropsis cf. sparsilis/ Augochloropsis electra/ Bombus atratus/ Bombus brevivillus/ Bombus morio/ Centris fuscata/ Centris tarsata/ Centris varia/ Dialictus sp./ Epicharis flava/ Euglossa spp./ Eulaema nigrita/ Exomalopsis analis/ Exomalopsis auropilosa/ Melipona fasciculata/ Melipona quadrifasciata/ Nannotrigona perilampoides/ Oxaea flavescens/ Partamona sp./ Pseudaugochlora erythrogaster/ Pseudaugochlora graminea/ Tetragonisca angustula/ Tetrapedia diversipes/ Trigona fulviventris/ Trigona spinipes/ Xylocopa frontalis/ Xylocopa muscaria/ Xylocopa nigrocincta/ Xylocopa suspecta</i>	[60, 63-65]
Soja (<i>G. max</i>)	<i>Apis mellifera/ Scaptotrigona sp./ Tetragonisca angustula/ Trigonas</i>	[66-68]
Pimentão (<i>Capsicum annuum</i>)	<i>Apis mellifera/ Augochlora cf. morrae/ Augochlora. / Augochlorella acarinata/ Augochloropsis caerulans/ Augochloropsis cupreola/ Exomalopsis auropilosa/ Exomalopsis aureopilosa/ Melipona quadrifasciata/ Melipona scutellaris/ Melipona subnitida/ Nannotrigona testaceicornis/ Paratrigona lineta/ Pereirapis rhizophila/ Trigona spinipes/ Tetragonisca angustula/ Toxomerus sp./ Trigona recurva</i>	[60,69-72]
Mandioca (<i>Manihot esculenta</i>)	<i>Apis mellifera/ Melipona compressipes/ Tetragona clavipes/ Trigona spinipes/ Xylocopa aeneipennis</i>	[73, 74]
Milho (<i>Z. mays</i>)	<i>Apis mellifera/ Tetragonisca angustula/ Trigona spinipes</i>	[75]

Estudos revelam que as contaminações por agrotóxicos atingem também o solo e as águas superficiais e/ou subterrâneas. Os fertilizantes e agrotóxicos são a segunda causa de contaminação das águas no país, seja por erosão e lixiviação do solo, pulverizações aéreas, escoamento superficial e destinação inadequada de embalagens [76]. As contaminações dos corpos d'água são um fator preocupante,

pois os resíduos podem atingir um maior número de pessoas e ecossistemas [77]. As pesquisas realizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) sobre saneamento básico em 2011 divulgaram que diversos municípios afirmaram existir poluição por agrotóxicos na captação de água para o abastecimento urbano, o que abrange os corpos d'água superficiais, poços rasos e poços profundos (Figura 1). Tal fato pode ser justificado pelo grande volume destas substâncias aplicado nas lavouras, e sua proximidade dos pontos de captação. O problema pode ser agravado devido a algumas propriedades dos agrotóxicos, que podem ser persistentes, móveis e tóxicos na água [78].

Figura 1 Municípios que relataram poluição por agrotóxicos em água, Brasil, 2011.

Fonte: Atlas do saneamento IBGE 2011.



1.5 *Commodities* brasileiras

É considerada uma *commodity* uma cultura que é produzida em larga escala e por ter seu preço controlado pelo mercado mundial (comercializado em dólares). Exemplos de *commodities* são milho (*Zea mays* L.), soja (*Glycine max*, L. Merril.) e tomate (*Solanum lycopersicum*), sendo as principais regiões produtoras o Centro-Oeste, Sul e Sudeste do Brasil [79].

1.5.1 *Glycine max* L. Merr. (soja)

A soja (*Glycine max* L. Merr.) é a quarta cultura mais importante do mundo em termos produção e área plantada [80]. De acordo com dados fornecidos pela Companhia Nacional de Abastecimento (CONAB) no segundo semestre desse ano, a estimativa de 35,2 milhões de hectares semeados na atual safra é 3,7% superior ao que foi cultivado na safra 2016/17 e 69,9% maior do que a safra 2006/07, constituindo-se no décimo primeiro aumento consecutivo na área total cultivada com soja no Brasil. A área cultivada com essa proteoleaginosa ultrapassou a área de milho total na safra 1997/98 e, desde então, ocupa o primeiro lugar no *raking* em área semeada no país. O Brasil teve um incremento de 14,5 milhões hectares novos de soja nas últimas 12 safras, sendo a cultura que mais cresce em área cultivada no país. Atualmente corresponde a cerca de 57% da área total semeada com grãos no país [81].

A soja é rica em proteína vegetal, que uma vez ingerida por humanos ou animais, transforma-se em proteína animal. Dezenas de alimentos proteicos possuem esta proteolegínosa na sua composição [82]. Desde o grão *in natura* como alimento funcional até o óleo de soja, o grão é utilizado em diversas preparações alimentícias, a exemplo de chocolates, farelos para alimentação animal, molhos prontos, massas, bebidas, alimentos dietéticos, derivados de carne, dentre muitos outros. Seu uso também é bastante difundido na indústria farmacêutica, indústria de

comésticos, fertilizantes, fabricação de fibras e biocombustível, um combustível sustentável que contribui para reduzir a emissão de poluentes [83].

Para sustentar toda essa cadeia produtiva, diversas medidas de manejo de culturas são empregadas, como já descritas anteriormente, sendo a mais difundida o uso de agrotóxicos. No período compreendido entre 2002 a 2011 o consumo médio de agrotóxicos vem aumentando em relação à área plantada, onde passou de 10,5 para 12 litros por hectare [84-88]. Do montante de 852,8 milhões de litros de agrotóxicos utilizados nas culturas de plantas em 2011, a soja consumiu 40% do volume total entre herbicidas, inseticidas, fungicidas, acaricidas, dentre outros produtos. Logo após estão o milho com 15%, a cana e o algodão com 10%, depois os cítricos com 7%, o café com 3% e o tomate (1%). As demais culturas consumiram 24% do total de agrotóxicos [84,85]. Em 2015 a área plantada de soja foi de 32.206.787 hectares, com uma média de 17,7 litros de agrotóxicos por hectare perfazendo um total de 570.060.129,90 litros [89].

1.5.2 *Zea mays* L. (milho)

O milho (*Zea mays* L.) é pertencente à família Poaceae, uma espécie alógama cultivada em diversas partes do mundo. A sua grande plasticidade genotípica e adaptabilidade permitem o seu cultivo em áreas tropicais, temperadas e subtropicais. O milho é utilizado principalmente na alimentação humana e ração animal devido às suas elevadas qualidades nutricionais, contendo quase todos os aminoácidos conhecidos, com exceção da lisina e do triptofano. A espécie ainda é utilizada para produção de energia a partir do biodiesel e etanol [90].

O milho é um dos principais produtos do agronegócio nacional. A safra 2017 atingiu a marca de 68.481.488 toneladas [91]. Diversos cultivares são disponibilizados para os pequenos e grandes produtores, sendo 50% da produtividade de uma lavoura é atribuída ao cultivar escolhido, de maneira geral. De acordo com a EMBRAPA 315 cultivares foram disponibilizados na safra 2016/2017, dos quais, 214 cultivares (67,93%) apresentavam tecnologias transgênicas: desde o

potencial genético visando produtividade, resistência a doenças, pragas e moléculas de herbicidas como glifosato e glifosinato para o controle de plantas daninhas [92].

A área plantada de milho em 2015 foi de 15.846.517 de hectares, com uma média de 7,4 litros de agrotóxicos por hectare perfazendo um total de 117.264.225,80 de litros [89]. Na cultura do milho as doenças podem ocorrer de forma epidêmica, com sintomas manifestados em até 100% das plantas na lavoura. Os problemas com os patógenos podem ter suas proporções ampliadas em áreas de plantio direto. Nestas situações, é fundamental o manejo da cultura, como a rotatividade, escolha de cultivares tolerantes às principais doenças e uso de agrotóxicos a fim de evitar redução de produtividade. Todos os produtos comerciais registrados no MAPA para o manejo de doenças do milho são pertencentes aos grupos químicos dos triazóis e das estrobilurinas [93].

1.5.3 *Solanum lycopersicum* L. (tomate)

O tomate (*S. lycopersicum* L.) é um fruto pertencente à família solanacea, originário da América do Sul, sendo umas das culturas mais difundidas no mundo [94]. No Brasil, é uma das oleráceas mais consumidas sendo o estado de Goiás o maior produtor com 32% da produção nacional, o que reflete na sua participação expressiva no mercado nacional [95]. A safra nacional do fruto para 2018 foi estimada em 4.432.739 toneladas, um crescimento de 0,6% em relação à safra de 2017 que atingiu a marca de 4.373.047 toneladas [91].

O fruto é comumente consumido em sua forma natural, em preparações industrializadas como extratos, enlatados, purês, molhos, alimentos congelados, papinhas, sopas, dentre outros. O tomate é rico em vitaminas B e C, ferro e fósforo, aminoácidos essenciais, açúcares, fibras e o licopeno, um carotenóide que pode ajudar na prevenção do câncer de próstata e no fortalecimento do sistema imunológico [96-99].

Sua cadeia produtiva apresenta custos elevados por ser uma cultura que demanda tratamentos culturais rigorosos, disponibilidade de altos níveis tecnológicos e utilização de mão-de-obra. É caracterizada também por ser uma cultura suscetível a

estresses abióticos (déficit hídrico) e bióticos (pragas e doenças). Diversas doenças acometem o tomateiro, ocasionando grandes perdas de produtividade e qualidade do fruto. Para controle é recomendado o uso de cultivares resistentes às principais doenças, vistorias frequentes na lavoura, tratos culturais exigentes e uso de defensivos agrícolas [100]. O tomate já foi apontado como um grande consumidor de agrotóxicos. No último relatório ANVISA-PARA as detecções foram menores do que nos anos anteriores. Foram analisadas 730 amostras de tomate de mesa e dessas, 450 amostras apresentaram resíduos em concentrações iguais ou inferiores ao LMR. No total, foram detectados 63 agrotóxicos diferentes dentre os 202 pesquisados: acefato (353 amostras), imidacloprido (333 amostras) e carbendazim (272 amostras). Ainda foram detectadas 126 amostras irregulares de tomate contendo o defensivo agrícola clorpirifós, não autorizado para a cultura em questão [19].

1.6 Genômica comparativa e a proposição de alvos

Com o crescente desenvolvimento das tecnologias de sequenciamento do tipo NGS (*next generation sequencing*), a quantidade de informações genéticas depositadas nos bancos de dados aumentou exponencialmente o que estimulou a demanda por abordagens rápidas, acuradas e automatizadas para analisar genomas de forma comparativa [101]. Assim, a genômica comparativa emergiu como uma poderosa ferramenta para identificar elementos genômicos funcionalmente importantes nos mais diversos organismos [102,103].

A Bioinformática é uma ciência que tem por objetivo a manipulação de informações derivadas dos grandes projetos de genômica, transcriptômica e proteômica, dentre outros. A partir do uso de suas técnicas torna-se possível o armazenamento, distribuição, análise e interpretação de dados e possibilita a análise em larga escala e em curto período, graças aos avanços no campo da informática e engenharia de *softwares* [104-106].

O setor farmacêutico teve uma grande participação no desenvolvimento da bioinformática. O processo de desenvolvimento de fármacos inicia-se com uma

ampla seleção de potenciais alvos, onde a grande parte é descartada por não atender a pré-requisitos necessários. Ao final, somente alguns candidatos são selecionados para testes, e às vezes nenhum alvo atende aos critérios selecionados. No entanto em qualquer situação, o custo é oneroso [107]. Estima-se que 75% dos custos empregados para desenvolver um novo medicamento são usados nas moléculas que serão descartadas. Assim, um dos principais focos da bioinformática tem sido a descoberta de alvos específicos de inibidores enzimáticos a partir do extenso estudo do genoma do hospedeiro e do patógeno, onde alvos com características desejáveis são selecionados, minimizando os custos. Os alvos possuem definição abrangente e podem ser genes, RNAs e mais comumente proteínas; estas desempenham um papel fundamental no processo de patogênese [108]. Avanços na Biologia Estrutural permitiram o desenvolvimento da técnica de desenho racional de fármacos, onde inibidores enzimáticos começaram a ser planejados para alvos específicos [109].

Apesar de toda a importância biológica e econômica das plantas, os dados provenientes do sequenciamento de plantas crescem de uma maneira mais lenta se comparado com o sequenciamento de outros organismos, devido à complexidade na montagem do seu genoma [110]. Assim como nos genomas de outros eucariotos, as plantas possuem sequências de DNA repetitivas, sequências invertidas e transposons. No entanto, o genoma das plantas é mais complexo que os demais genomas eucarióticos e isso se deve ao fato de que plantas apresentam tamanho cromossômico e níveis de ploidia muito variados, o que modifica ainda mais seu conteúdo genético [111].

Um problema muito comum no complexo de interação entre uma planta e fitopatógeno é a dificuldade de conter a infecção por meio de técnicas de manejo da cultura, como podas, rotação de culturas, dentre outros, dada a resistência de indivíduos dentro da população de fitopatógenos [112]. Em decorrência deste fato, técnicas de obtenção de genótipos resistentes à doença têm sido amplamente utilizadas. No entanto, este tipo de prática promove à redução do número de genótipos cultiváveis, favorecendo a seleção de cultivares com uma base genética estreita. Visto que as plantas selvagens possuem um alto grau de heterozigose e ampla adaptabilidade frente às adversidades, no geral elas são pouco produtivas e muito heterogêneas. Com o surgimento das tecnologias que permitiram a seleção

genômica em larga escala, muitos genes já foram perdidos durante os sucessivos eventos de segregação, recombinação e seleção, gerando uma base genética estreita nas culturas modernas [113].

As técnicas avançadas de biologia molecular e genética permitem a identificação de genes diferencialmente expressos no complexo da interação planta vs fitopatógeno, e dessa forma é possível identificar genes que se apresentam como prováveis indutores de resistência em genótipos específicos [114].

No âmbito na interação homem vs patógeno, na maioria das vezes, desde a proposição de fármacos até sua liberação comercial, um longo caminho de testes é percorrido [115-116]. Com o intuito de contornar este problema, novas abordagens vêm sendo propostas. A triagem de alvos por técnicas de bioinformática aumentam as chances de escolher um gene que tenha seu produto validado como alvo, resultando em economia de tempo e custos. Uma abordagem com poucos estudos na área da agricultura tem como base a identificação de enzimas isofuncionais não-homólogas, ou simplesmente análogos funcionais, entre o hospedeiro e seu fitopatógeno. A identificação destas enzimas permite que eventualmente determinadas enzimas chave com funções similares e com conformações diferentes sejam identificadas em vias metabólicas essenciais, onde inibidores específicos podem ser produzidos visando ao bloqueio de um ou mais passos em vias metabólicas importantes do fitopatógeno [117-118].

1.7 Enzimas isofuncionais não homólogas

A reconstrução de vias metabólicas pode expor determinados compostos essenciais à manutenção das funções vitais dos organismos. Dentre estes compostos podemos citar as enzimas, componentes fundamentais para qualquer processo bioquímico, por apresentarem elevada eficiência catalítica, alto grau de especificidade por seus substratos e por funcionarem em soluções aquosas sob determinadas condições de pH e temperatura. Dentre os grupos de moléculas que possuem atividade catalítica, com exceção de alguns RNA específicos, todas as enzimas cumprem essa função e são proteínas [119,120].

Dada a sua importância na maioria dos processos biológicos, as enzimas têm sido alvo de investigação, principalmente no que tange à regulação e catálise de processos vitais nas vias metabólicas onde elas atuam. Nestas, as enzimas catalisam centenas de reações sucessivas envolvidas na degradação de nutrientes, sendo a energia química conservada e transformada, consistindo em processos altamente coordenados e indispensáveis para a manutenção da vida de um organismo [121].

Recentemente, pesquisadores têm empenhado esforços em identificar, caracterizar e reconstruir rotas bioquímicas com as quais estão envolvidas as enzimas isofuncionais não homólogas, (*Non-homologous isofunctional enzymes* - NISEs). Compreende-se por NISEs enzimas que desempenham a mesma função bioquímica, mas são provenientes de origens evolutivas distintas, possuindo diferenças em suas estruturas primárias o que, geralmente, é refletido em diferenças em suas estruturas terciárias [122-123]. Os estudos disponíveis na literatura sobre as NISEs ainda são poucos, principalmente no que diz respeito ao reino Plantae [124,125]. Diversas NISEs foram identificadas em organismos de importância médica, onde foram identificados candidatos que podem ser estudados como potenciais novos alvos terapêuticos para o desenvolvimento de medicamentos contra *Leishmania major*, *Trypanosoma brucei*, e *Trypanosoma cruzi* [118, 126]. Além da importância na área da terapêutica, o estudo das NISEs tem se tornado interessante pelo fato de apresentarem um tema de estudo evolutivo e por auxiliar na reconstrução de vias metabólicas em uma gama de organismos [123].

As ômicas, que constituem em plataformas biotecnológicas de análises de sequências, seus produtos gênicos, proteínas, metabólitos primários e secundário, tomaram força desde então e têm possibilitado muitas descobertas a respeito da identificação, função e comparação de genes, transcritos e proteínas bem como suas interações [127]. Estas informações, quando comparadas às de outros organismos, apresentam-se como uma fonte poderosa na na caracterização de espécies, permitindo a realização de estudos evolutivos e filogenéticos além da exposição de moléculas e rotas bioquímicas exclusivas que podem ser essenciais na descoberta de fármacos e/ou vacinas e, no caso da agricultura, podem auxiliar no combate aos ataques por fitopatógenos [128-130].

Para atender a demanda mundial por alimentos, houve um aumento na pesquisa no campo das Ciências Agrícolas, a fim de expandir a produção e resolver problemas de pragas e doenças das culturas. Assim, os estudos envolvendo o melhoramento genético de plantas tomaram impulso com as investigações por meio do uso extensivo de marcadores moleculares tais como *Single Nucleotide Polimorfism* (SNPs) [131], *Sequence Short Repeat* (SSR) [132], *Expressed Sequence Tags - Sequence Short Repeat* (EST-SSR) [133], estudos de expressão gênica envolvendo a construção de bibliotecas subtrativas de cDNA [134], técnicas de microarranjos [135] e a *Reverse transcription - polymerase chain reaction quantitative real time* (RT-qPCR) [136]. O uso destas metodologias permitiu selecionar caracteres quantitativos e qualitativos de determinadas plantas, onde o objetivo é a obtenção de um cultivar com alta produtividade e resistência às principais linhagens ou cepas de espécies fitopatogênicas. No entanto, o problema persiste, pois com o passar dos anos os fitopatógenos são capazes de suplantar a resistência de um determinado cultivar, sendo necessárias outras medidas de controle [137-144].

Mesmo com os avanços tecnológicos, as vantagens de ser um país essencialmente agrícola com forte potencial de ascensão – com disponibilidade de recursos hídricos e material genético – não atendem as necessidades atuais. São necessárias medidas para minimizar a atual degradação do meio ambiente e os riscos para a saúde humana.

1.8 Bancos de dados Biológicos

1.8.1 KEGG

O *Kyoto Encyclopedia of Genes and Genomes* (KEGG) é uma coleção de banco de dados extensivo que integra informações funcionais genômicas, químicas e sistêmicas [145-147]. O banco de dados KEGG/PATHWAY é uma coleção de mapas de vias metabólicas curadas manualmente representando o conhecimento

sobre as redes de interação molecular para as redes metabólicas já descritas. Cada mapa de referência pode ser visualizado como uma rede de enzimas ou uma rede de números ECs (*Enzyme Commission*) interconectados. O EC é um esquema de classificação numérica atribuída às enzimas; tal nomenclatura foi estabelecida pela *Internacional Union of Biochemistry and Molecular Biology Nomenclature Committee* (NC-IUBMB). O número EC é formado por quatro dígitos, o primeiro corresponde à classe, o segundo dígito corresponde à subclasse, o terceiro à sub-subclasse e o quarto dígito indica a atividade catalítica. As enzimas estão divididas em seis principais classes: oxidoredutases (EC 1), transferases (EC 2), hidrolases (EC 3), liases (EC 4), isomerases (EC 5), ligases (EC 6) [148].

Uma vez que os genes correspondentes à enzima são identificados no genoma, com base na similaridade de sequência e na correlação posicional dos genes, os números de ECs são atribuídos. As vias específicas de um determinado organismo podem ser construídas computacionalmente usando a genômica comparativa, correlacionando genes e enzimas nas vias de referência de acordo com os números EC correspondentes. As vias metabólicas são conservadas entre a maioria dos organismos, especialmente para o metabolismo central de mamíferos e bactérias. Assim, é possível planejar manualmente uma via de referência e, em seguida, gerar vias específicas de organismos. Em contrapartida, as vias regulatórias são muito mais divergentes e difíceis de combinar em diagramas de vias de referência comuns [146].

1.8.2 SUPERFAMILY e SCOP

SUPERFAMILY (<http://supfam.org>) é um banco de dados de acesso livre, onde o principal objetivo da ferramenta é prever a presença de domínios proteicos de estrutura conhecida. O banco de dados possui um repertório de 15.438 modelos ocultos de Markov (*hidden Markov models*- HMM) curados manualmente que representam todos os domínios de estrutura proteicas que se tem conhecimento. A classificação desses domínios proteicos é retirada do banco de dados *Structural Classification of Protein* (SCOP) (<http://scop.mrc-lmb.cam.ac.uk/scop/>). O SCOP é

organizado de maneira hierárquica, classificando os domínios proteicos de acordo com a similaridade estrutural, em níveis crescentes de similaridade e relação evolutiva, incluindo: níveis de classe e padrões de enovelamento (*fold*s) [149].

Uma superfamília pode agrupar domínios de famílias diferentes que possuem um ancestral evolutivo comum. Essa inferência é baseada em análise de dados estruturais, funcionais e de sequência. Dois domínios são estritamente agrupados na mesma Superfamília desde que haja evidência estrutural de um ancestral comum. Em geral, a anotação SCOP ao nível da família também foi considerada funcionalmente consistente, e tem uma relação evolutiva mais próxima, muitas vezes diretamente observável ao nível da sequência de aminoácidos [150-151].

1.8.3 BRENDA

O BRAunschweig ENzyme DAtabase (BRENDA) é um banco de dados abrangente que armazena informações funcionais e moleculares de todas as enzimas que se tem conhecimento com base em uma extensiva revisão de literatura ligadas ao PubMed. O banco de dados BRENDA ainda inclui informações sobre as propriedades de todas as enzimas classificadas, dados sobre a ocorrência, reação catalisada, cinética, substratos/ produtos, inibidores, co-fatores, ativadores, estrutura e estabilidade [152,153].

As informações disponibilizadas constituem em importantes ferramentas para a pesquisa de mecanismos enzimáticos, vias metabólicas, evolução do metabolismo, diagnósticos de doenças e desenvolvimento de fármacos. As enzimas são classificadas de acordo com o sistema EC do IUBMB e as todas as demais informações são cruzadas com outras bases de dados, como o *Protein DataBank* (PDB) [154-156] e o *Universal Protein knowledgebase* (Uniprot) [157], fornecendo dados com a classificação do organismo, sequência e estrutura proteica e referências bibliográficas.

1.9 *Pipelines* de identificação de enzimas

1.9.1 Analogous Enzyme Pipeline (AnEnPi)

O AnEnPi é um *pipeline* desenvolvido em linguagem de programação *perl* que faz uma combinação de ferramentas como o BLAST, HMMer e *scripts* caseiros, para auxiliar na identificação, anotação e comparação de enzimas análogas e homólogas. Para a detecção de casos de analogia o algoritmo baseia-se em duas principais etapas: construção de grupamentos de enzimas homólogas e identificação de casos em que uma dada atividade enzimática é realizada por duas ou mais proteínas sem similaridade significativa entre suas estruturas primárias (enzimas análogas). O AnEnPi é ferramenta eficiente para a detecção e anotação de enzimas análogas e na construção de conjuntos de dados iniciais que podem ser posteriormente curados, principalmente em estudos envolvendo evolução e metabolismo molecular e na identificação de novos alvos potenciais de fármacos. [117].

2. OBJETIVOS

2.1 Objetivo Geral

Identificar enzimas análogas essenciais entre plantas e seus fitopatógenos com o objetivo de propor potenciais novos alvos para o desenvolvimento de inibidores enzimáticos, no intuito de conter infecções causada por fitopatógenos.

2.2 Objetivos Específicos

- Construir grupamentos de atividades enzimáticas similares de acordo com os números EC correspondente;
- Inferir a função das proteínas dos patógenos através da similaridade com os grupos obtidos pelo *pipeline* AnEnPi;
- Identificar e validar enzimas análogas entres as plantas e seus respectivos patógenos;
- Verificar a presença das enzimas análogas validadas para os fitopatógenos em *Homo sapiens*, *Apis mellifera*, *Trichoderma harzianum* e *Bacillus subtilis*;
- Identificar as enzimas essenciais nos patógenos candidatas para o alvos de inibidores a partir da lista de enzimas análogas obtidas;

3. METODOLOGIA

3.1 Identificação de enzimas análogas

3.1.1 Organismos estudados

Três genomas de plantas (*G. max*, *Z. mays* e *S. lycopersicum*) foram extraídos do RefSeq e Uniprot e analisados. Os fitopatógenos foram escolhidos de acordo com a disponibilidade dos genomas nas bases de dados; a maioria de ocorrência cosmopolita. Os patógenos estudados compreendem 11 genomas de fungos e 4 genomas de bactérias, todos patogênicos para uma ou mais espécies de plantas estudadas (Tabela 3). Também foram incluídos os genomas de *Homo sapiens*, *Apis mellifera* (polinizador), *Trichoderma harzianum* (agente de controle biológico encontrado no solo) e *Bacillus subtilis* (bactéria promotora do crescimento em plantas) (Tabela 4).

Tabela 3. Descrição das doenças e sintomas causados pelos fitopatógenos avaliados neste estudo.

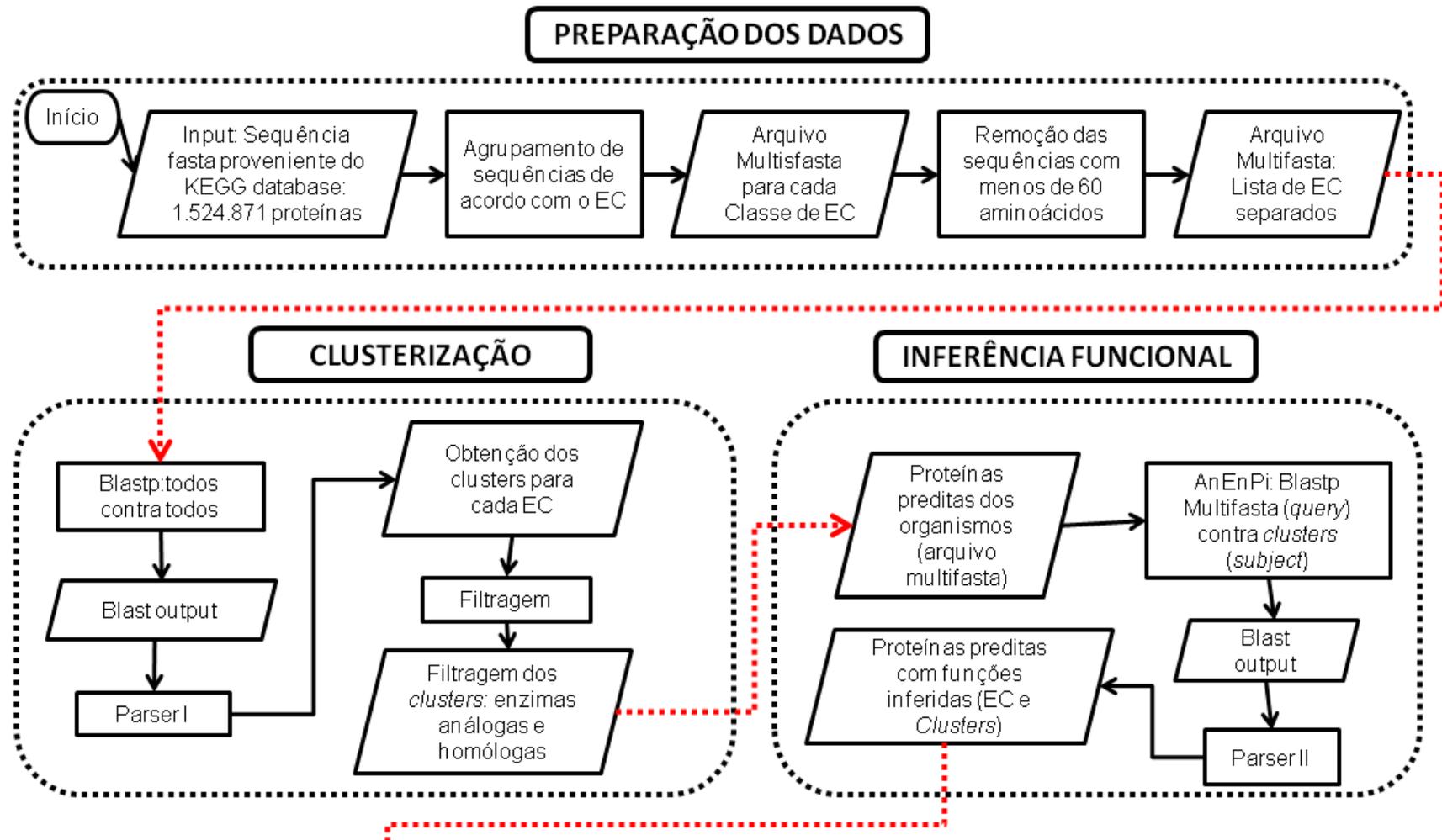
Plantas	Produção anual / ton*	Fitopatógenos	Doença	Sintomas	Distribuição	Referencia
<i>Glycine max</i>	30.651.925 mi	<i>Aspergillus flavus</i>	Podridão de sementes	Podridão de sementes, liberação de aflatoxinas	Cosmopolita	[158]
		<i>Fusarium oxysporum</i>	Podridão basal	Amarelecimento de folhas, nanismo em plantas jovens, murchamento de ramos e morte de plantas	Cosmopolita	[159]
		<i>Phytophthora sojae</i>	Podridão radicular	Podridão de sementes, raízes e caules	Cosmopolita	[160]
		<i>Sclerotinia sclerotiorum</i>	Mofo branco	Ramos laterais e caule mostrando lesões, resultando em pobre preenchimento de vagem ou morte de planta	Cosmopolita	[161]
		<i>Xanthomonas axonopodis</i>	Mancha bacteriana	Pequenas pústulas, rodeadas de halos.	Cosmopolita	[162]
<i>Solanum lycopersicum</i>	170750767mi	<i>Botrytis cinerea</i>	Mofo cinzento	Mofo cinzento em várias partes da planta, como folhas, brotos, cálices jovens, flores e frutos	Cosmopolita	[163,164]
		<i>Fusarium oxysporum</i>	Murcha de Fusarium	Amarelecimento das folhas inferiores, murchamento progressivo de folhas e caule, morte da planta.	Cosmopolita	[165]
		<i>Moniliophthora perniciosa</i>	Vassoura-de-Bruxa	Hipertrofia, rebentamento, proliferação de botões laterais e espessamento dos tecidos infectados.	América do Sul, Ilhas do Caribe e Panamá	[166]
		<i>Pseudomonas syringae</i>	Pinta Bacteriana	Pequenos pontos de aparência necrótica, desfolhação	Cosmopolita	[167]
		<i>Ralstonia solanacearum</i>	Murcha bacteriana	Aparência flácida das folhas mais jovens, degradação dos vasos e tecidos circundantes, morte da planta	Cosmopolita	[168]
<i>Zea mays</i>	1.037.791,518 bi	<i>Aspergillus flavus</i>	Podridão da orelha de Aspergillus	Oídio em grãos, produção de aflatoxinas que são prejudiciais para aves e mamíferos	Cosmopolita	[169,170]
		<i>Colletotrichum graminicola</i>	Podridão de antracnose	As folhas apresentam lesões irregulares, ovais a alongadas, com margens marrons de amarelo a avermelhado características	Cosmopolita	[171]
		<i>Gibberella moniliformis</i>	Podridão de orelha	Os grãos desenvolvem estrias brancas. Produção de micotoxinas conhecidas como fumonisinas	Cosmopolita	[172]
		<i>Exserohilum turcicum</i>	Ferrugem da folha	Lesões necróticas alongadas e fusiformes	Cosmopolita	[173]
		<i>Pantoea ananatis</i>	Mancha branca de milho	Formação de manchas cloróticas verde-escuras, com aspecto de alagamento, causando necrose na região do tecido afetado	Cosmopolita	[174,175]

* [176,177].

3.1.2 Identificação de enzimas iso-funcionais não homólogas e essenciais

As análises foram realizadas em cinco estágios principais: preparação dos dados, clusterização, inferência funcional e validação estrutural e essencialidade de acordo com o fluxograma da metodologia (Figura 2).

Figura 2 A. Fluxograma: Identificação de enzimas iso-funcionais não homólogas e essenciais.



3.1.3 KEGG e clusterização

O conjunto de todas as sequências de enzimas anotadas de acordo com a sua atividade enzimática foi extraído do KEGG (versão 73.0, janeiro de 2015) que continha 1,524,871 sequências de proteínas. A base de dados continha sequências de 298 Eucariotos, 3014 Eubacterias e 175 Archaea. As sequências com menos de 60 aminoácidos foram removidas, dado que estas poderiam representar fragmentos de proteínas.

Para organizar as sequências de proteínas em grupos, realizamos a clusterização das sequências contidas no banco de dados do KEGG. Para este passo, utilizamos o *pipeline* AnEnPi [117]. Foi utilizado um índice de similaridade com um valor de corte de 120 para todas as comparações par a par BLASTp entre todas as proteínas incluídas em um conjunto de dados especificado. Após esta etapa, para cada grupo de sequências inicialmente separadas pelo número EC e, portanto, compartilhando a mesma atividade funcional, obtivemos arquivos contendo um ou mais grupos de estruturas primárias, com base na similaridade de sequência. Se, para uma determinada atividade enzimática (conforme definido pelo seu número EC), apenas um grupo foi produzido no final do passo de agrupamento, então todas as seqüências seriam consideradas homólogas e conseqüentemente a atividade enzimática foi removida da análise. Por outro lado, se mais de um cluster fosse produzido, as sequências agrupadas no mesmo cluster foram consideradas homólogas, com uma pontuação acima de 120. Caso contrário, com uma pontuação menor que 120 as enzimas alocadas em diferentes clusters foram consideradas como potenciais enzimas análogas [122,123].

3.2 Inferência da função das proteínas preditas dos organismos

Os conjuntos de dados das proteínas preditas para cada genoma estudado neste trabalho foram obtidos a partir do UniprotKB (versão 2015_10 and 2016_12, <http://www.uniprot.org/>) e RefSeq (Versão 70, <http://www.ncbi.nlm.nih.gov/>). A inferência funcional das proteínas preditas dos organismos também foi realizada

com AnEnPi usando os grupos de sequências homólogas obtidas após agrupamento (Tabela 4)

Tabela 4. Descrição dos conjuntos de dados das proteínas preditas dos organismos incluídos neste estudo.

Organismos	Database	Acesso NCBI	Referência	#Ptn	Não Carac.	Hip.	Put.	Anot. (%)
<i>Glycine max</i>	RefSeq	NC_016088	[178]	59374	23618	—	1566	61
<i>Aspergillus flavus</i> ^{1*}	RefSeq	GCA_000006275.2	[179]	13287	5380	—	—	59
<i>Fusarium oxysporum</i> ^{2*}	Uniprot	GCA_000222805.1	[180]	17385	16,684	—	1	8
<i>Phytophthora sojae</i> *	RefSeq	AAQY00000000	[181]	26106	—	25279	125	2,8
<i>Sclerotinia sclerotiorum</i> *	RefSeq	AAGT00000000.1	[182]	12902	12,042	—	3	6,6
<i>Xanthomonas axonopodis</i> **	RefSeq	CP004399	[183]	4496	1413	—	35	67
<i>Solanum lycopersicum</i>	Uniprot	AEKE00000000	[184]	31683	28785	—	—	9,1
<i>Botrytis cinerea</i> *	RefSeq	NZ_AAID00000000.1	[185]	14687	—	8,696	—	40
<i>Fusarium oxysporum</i> ^{3*}	Uniprot	GCA_000149955.2	[186]	15811	15,148	—	—	4,3
<i>Moniliophthora perniciosa</i> *	Uniprot	ABRE00000000	[187]	12915	12,741	—	—	1,3
<i>Pseudomonas syringae</i> **	RefSeq	NC_004578.1	[188]	5449	—	1446	—	73
<i>Ralstonia solanacearum</i> **	RefSeq	NC_003295.1	[189]	4400	696	135	1292	56
<i>Zea mays</i>	RefSeq	LPUQ00000000	[190]	59384	—	2363	2300	92
<i>Aspergillus flavus</i> ^{4*}	RefSeq	GCA_000952835.1	[191]	13561	—	5423	5884	16
<i>Colletotrichum graminicola</i> *	RefSeq	ACOD00000000	[192]	11910	—	5,381	—	54
<i>Gibberella moniliformis</i> *	Uniprot	AAIM00000000.2	[193]	17384	13,71	—	—	21
<i>Exserohilum turcicum</i> *	RefSeq	AIHT00000000	[194]	4248	—	11159	1	3,6
<i>Pantoea ananatis</i> **	RefSeq	CP001875	[195]	4302	707	—	14	83
<i>Apis mellifera</i>	Uniprot	AADG00000000	[196]	13514	12511	—	5	7,3
<i>Trichoderma harzianum</i> *	Uniprot	MRYK00000000	[197]	11480	7704	—	3	32
<i>Bacillus subtilis</i> **	Uniprot	NC_000964	[198]	26433	1299	—	301	93
<i>Homo sapiens</i>	Uniprot	CM000663	[199]	63487	1338	—	1071	96

- Sem proteínas nesta categoria

* Fungo

** Bactéria

¹ *A. flavus* NRRL3357, ² *F. oxysporum* Fo5176, ³ *F. oxysporum* 4287, ⁴ *A. flavus* AF70.

#Ptn, número total de proteínas; Não carac., proteínas não caracterizadas; Hip., proteínas hipotéticas; Put., proteínas putativas; Annot.%, porcentagem de proteínas anotadas.

Neste passo, as enzimas foram agrupadas usando BLASTp [200] de acordo com o número EC: as proteínas preditas de plantas e patógenos foram comparadas par a par com cada estrutura de proteica dentro de cada grupo funcional do KEGG. Para a inferência da função bioquímica, utilizou-se um e-valor 10^{-20} , um valor

altamente restritivo que proporciona maior confiabilidade aos resultados [117,126, 201,202]. As sequências com pontuações acima deste limiar foram removidas da análise.

3.3 NISEs: identificação, validação estrutural e essencialidade

A busca de casos de analogia foi realizada através da análise dos grupos produzidos após o passo de clusterização e inferência funcional. Outro módulo do AnEnPi foi usado para rastrear enzimas análogas entre plantas e patógenos. Para validar as NISEs identificadas, classificamos as sequências de acordo com seus enovelamentos usando o banco de dados SUPERFAMILY. A informação neste banco de dados é baseada em uma coleção de modelos HMM [203], que representam os domínios estruturais das proteínas classificadas a partir da base de dados SCOP [204].

Para refinar a predição de casos de analogia funcional, as enzimas heteromultiméricas (codificadas por diferentes genes), as enzimas anotadas com o termo subunidade e as sequências sem dobra associada (predita) foram excluídas da lista final. As enzimas com a mesma classificação enzimática mas apresentando dobras diferentes e, conseqüentemente, pertencentes a diferentes superfamílias, foram consideradas análogas.

A base de dados de genes essenciais (DEG, 14.7, outubro / 2016, <http://www.essentialgene.org/>) foi utilizada como referência para a busca de atividades essenciais nos patógenos estudados. Uma pesquisa no BLASTp foi realizada entre todas as sequências enzimáticas identificadas com análises contra o banco de dados DEG. Um e-valor de 10^{-5} foi usado como limite.

Foi realizado um BLASTp entre todas as sequências enzimáticas identificadas como análogas contra as proteínas preditas dos organismos que não devem ser afetados (*H. sapiens*, *A. mellifera*, *T. harzianum* e *B. subtilis*) por um eventual inibidor para o alvo identificado no fitopatógeno. Um e-valor de 10^{-5} foi usado como limite.

4. RESULTADOS

4.1 Preparação dos dados, clusterização e inferência da atividade funcional

Após a limpeza e preparação, o conjunto de dados inicial obtido do KEGG foi reduzido para 1.225.682 sequências de proteínas distribuídas em 3.893 atividades enzimáticas. Após o agrupamento, este conjunto de dados foi utilizado para a reanotação das proteínas preditas das plantas e fitopatógenos, compreendendo 444.198 sequências individuais em 2.096 atividades enzimáticas das três plantas e seus 15 patógenos. As proteínas preditas de *H. sapiens*, *A. mellifera*, *T. harzianum* e *B. subtilis* também foram reanotadas, compreendendo 114.914 sequências individuais em 2.008 atividades enzimáticas. Houve heterogeneidade na anotação dos conjuntos baixados de proteínas preditas. Antes do procedimento de reanotação, o melhor organismo anotado entre as plantas era *Z. mays*, com aproximadamente 90% de suas proteínas caracterizadas, enquanto *S. lycopersicum* apresentou apenas 9% de suas proteínas anotadas. Entre os agentes patogênicos, *P. ananatis* apresentou 83% de todo o seu proteoma conceitual anotado e *M. perniciososa* apresentou apenas 1,3% de suas proteínas caracterizadas. Após a etapa de inferência funcional, onde apenas as enzimas foram reanalisadas, em média, 15% das proteínas de cada organismo foram associadas a uma atividade enzimática (dados não apresentados).

4.2 Potenciais NISEs: identificação e validação

Inicialmente, foram identificados 568 casos de NISEs potenciais, e deste conjunto 97 casos foram validados (Tabela 2, ver Tabela Suplementar 1 para mais detalhes). As sequências rotuladas com "*subunit*" ou "*chain*" (324 casos), correspondentes a enzimas que exibem a mesma dobra (55 casos) e sequências sem dobramento associado na base de dados SUPERFAMILY (92 casos) foram

excluídas. Casos de analogia foram validados para todos os patógenos estudados: apenas um caso foi encontrado para *P. sojae* e *S. sclerotiorum*, enquanto 14 casos foram encontrados para *A. flavus* AF70. No total, 13 casos de analogia foram encontrados nas comparações entre *G. max* e seus agentes patogênicos, 23 casos entre *S. lycopersicum* e seus patógenos e 61 casos entre *Z. mays* e seus patógenos (Tabela 5).

Tabela 5 . Número de NISEs potenciais, validadas, específicas e essenciais. Os números entre parênteses indicam o número de atividades enzimáticas identificadas no total.

Hospedeiros	Patógenos	Potenciais NISEs	Validadas	Específicas*	Essenciais
<i>G. max</i>	<i>A. flavus</i> ¹	25	4	3	2
	<i>F. oxysporum</i> ²	21	4	4	1
	<i>P. sojae</i>	25	1	1	1
	<i>S. sclerotiorum</i>	21	1	1	0
	<i>X. axonopodis</i>	12	3	2	2
<i>S. lycopersicum</i>	<i>B. cinerea</i>	18	3	2	1
	<i>F. oxysporum</i> ³	30	6	5	2
	<i>M. perniciosa</i>	23	4	2	2
	<i>P. syringae</i>	38	5	4	5
	<i>R. solanacearum</i>	32	5	5	5
<i>Z. mays</i>	<i>A. flavus</i> ⁴	64	14	8	7
	<i>C. graminicola</i>	69	13	7	9
	<i>E. turcicum</i>	62	12	7	9
	<i>G. moniliformis</i>	65	10	6	5
	<i>P. ananatis</i>	63	12	11	7
Total		568	97 (39)	68 (26)	58 (29)

*Número de estruturas terciárias específicas dos patógenos

¹ *A. flavus* NRRL3357

² *F. oxysporum* Fo5176

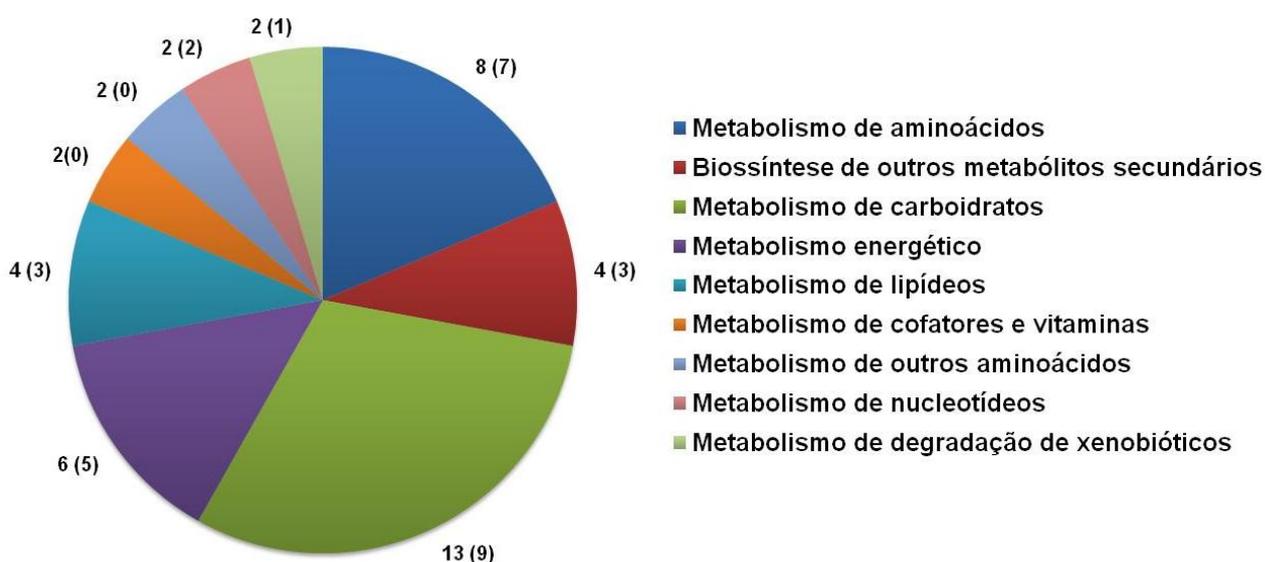
³ *F. oxysporum* 4287

⁴ *A. flavus* AF70.

As NISEs validadas (97 casos), que compreendem 39 atividades enzimáticas diferentes, participam de vias metabólicas centrais, incluindo o metabolismo de carboidratos (13 atividades enzimáticas), metabolismo de aminoácidos (8), metabolismo energético (6), biossíntese de metabólitos secundários (4) e metabolismo lipídico (4). Oito atividades enzimáticas pertencem a outras vias, como a degradação dos xenobióticos, o metabolismo dos cofatores e vitaminas, o metabolismo de nucleotídeos e o metabolismo de outros aminoácidos (Figura 3). É

importante lembrar que uma atividade enzimática pode participar em mais de uma via.

Figura 4 Classificação funcional das NISEs validados. Os números entre parênteses indicam a quantidade de atividades enzimáticas essenciais.



4.3 NISEs essenciais

Após a etapa de validação, realizou-se uma busca por enzimas essenciais, revelando 58 casos de analogia (Tabela 6), envolvendo 29 atividades enzimáticas essenciais diferentes, correspondentes a 119 estruturas diferentes, para todos os organismos analisados neste estudo. No metabolismo de carboidratos, o caso mais frequente foi a enzima catalase, classificado como essencial para três patógenos de *G. max* (*A. flavus*, *F. oxysporum* e *P. sojae*), três patógenos de *S. lycopersicum* (*F. oxysporum*, *P. seryngae* e *R. solanacearum*) e três patógenos de *Z. mays* (*A. flavus*, *E. turcicum* e *C. graminicola*). Os membros da via das pentoses, como 5-fosfato isomerase de ribose, 3-epimerase de ribulose-fosfato e glioxalase I, foram

identificados em três patógenos de *Z. mays* (*A. flavus*, *G. moniliformis* e *C. graminicola*). Outro caso frequente, quinases dependentes de ciclina (CDKs), foi encontrado para quatro dos cinco patógenos de *Z. mays* (*A. flavus*, *E. turcicum*, *C. graminicola* e *G. moniliformis*).

Tabela 6 Enzimas análogas e essenciais.

Hospedeiros	ID Seq. Hospedeiros	Patógenos	NISEs			Dados de Essencialidade	
			ID Seq. Patógenos	Número EC	Enzima	ID DEG**	E-value
<i>G. max</i>	NP_001235974.1	<i>A. flavus</i>	XP_002384918.1	1.11.1.6*	Catalase	DEG10110209	2,00E-68
<i>G. max</i>	XP_003557098.2	<i>A. flavus</i>	XP_002377297.1	1.11.1.7*	Peroxidase	—	—
<i>G. max</i>	XP_006600684.1	<i>A. flavus</i>	XP_002376298.1	1.2.1.3	Aldehyde dehydrogenase (NAD+)	DEG20180006	1,00E-65
<i>G. max</i>	XP_006600243.1	<i>A. flavus</i>	XP_002382374.1	2.6.1.1	Aspartate transaminase	—	—
<i>G. max</i>	NP_001235974.1	<i>F. oxysporum</i>	9FP11 F9FP11_FUSOF	1.11.1.6*	Catalase	DEG10110209	0
<i>G. max</i>	XP_003555725.2	<i>F. oxysporum</i>	F9FYF1_FUSOF	1.15.1.1*	Superoxide dismutase	—	—
<i>G. max</i>	XP_006600243.1	<i>F. oxysporum</i>	F9G466_FUSOF	2.6.1.1	Aspartate transaminase	—	—
<i>G. max</i>	XP_006598804.1	<i>F. oxysporum</i>	F9G2J4_FUSOF	4.4.1.5	Lactoylglutathione lyase	—	—
<i>G. max</i>	NP_001235974.1	<i>P. sojae</i>	XP_009521283.1	1.11.1.6*	Catalase	DEG10110209	8,00E-115
<i>G. max</i>	XP_003557098.2	<i>S. sclerotiorum</i>	XP_001585507.1	1.11.1.7*	Peroxidase	—	—
<i>G. max</i>	NP_001235974.1	<i>X. axonopodis</i>	WP_042823856.1	1.11.1.6*	Catalase	—	—
<i>G. max</i>	XP_006605648.1	<i>X. axonopodis</i>	WP_054320474.1	1.15.1.1*	Superoxide dismutase	DEG20241649	6,00E-18
<i>G. max</i>	XP_006601861.1	<i>X. axonopodis</i>	WP_033483073.1	6.4.1.2	Acetyl-CoA carboxylase	DEG10030125	4,00E-57
<i>S. lycopersicum</i>	K4CN29_SOLLC	<i>B. cinerea</i>	XP_001560519.1	3.1.3.2	Acid phosphatase	—	—
<i>S. lycopersicum</i>	LGUL_SOLLC	<i>B. cinerea</i>	XP_001550649.1	4.4.1.5	Lactoylglutathione lyase	—	—
<i>S. lycopersicum</i>	P21568 CYPH_SOLLC	<i>B. cinerea</i>	XP_001545186.1	5.2.1.8	Peptidylprolyl isomerase	DEG20241291	1,00E-46
<i>S. lycopersicum</i>	K4BVX3_SOLLC	<i>F. oxysporum</i>	A0A0D2YKD1_FUSO4	1.11.1.6*	Catalase	DEG10110209	0
<i>S. lycopersicum</i>	Q7XAV2_SOLLC	<i>F. oxysporum</i>	A0A0D2YE80_FUSO4	1.15.1.1*	Superoxide dismutase	—	—
<i>S. lycopersicum</i>	K4CN29_SOLLC	<i>F. oxysporum</i>	A0A0D2YGA3_FUSO4	3.1.3.2	Acid phosphatase	—	—
<i>S. lycopersicum</i>	Q42875_SOLLC	<i>F. oxysporum</i>	A0A0D2XJE6_FUSO4	3.2.1.4	Cellulase	—	—
<i>S. lycopersicum</i>	Q8GZD8_SOLLC	<i>F. oxysporum</i>	A0A0D2XCV3_FUSO4	3.4.11.5	Prolyl aminopeptidase	DEG20210010	7,00E-14
<i>S. lycopersicum</i>	LGUL_SOLLC	<i>F. oxysporum</i>	A0A0D2XLV4_FUSO4	4.4.1.5	Lactoylglutathione lyase	—	—
<i>S. lycopersicum</i>	P15003 PER1_SOLLC	<i>M. perniciosa</i>	E2LX62_MONPE	1.11.1.7*	Peroxidase	—	—
<i>S. lycopersicum</i>	Q9FVN0 AMT13_SOLLC	<i>M. perniciosa</i>	E2M162_MONPE	2.7.13.3	Histidine-kinase	DEG20070330	4,00E-36
<i>S. lycopersicum</i>	Q8GZD8_SOLLC	<i>M. perniciosa</i>	E2LYM3_MONPE	3.4.11.1	Leucyl aminopeptidase	—	—
<i>S. lycopersicum</i>	K4CJ01_SOLLC	<i>M. perniciosa</i>	E2LAS1_MONPE	5.4.2.8	Phosphomannomutase	DEG20020210	5,00E-30
<i>S. lycopersicum</i>	K4BVX3_SOLLC	<i>P. seryngae</i>	NP_794283.1	1.11.1.6*	Catalase	DEG10270348	0
<i>S. lycopersicum</i>	P15003 PER1_SOLLC	<i>P. seryngae</i>	NP_794565.1	1.11.1.7*	Peroxidase	DEG10180459	4,00E-10
<i>S. lycopersicum</i>	K4CN29_SOLLC	<i>P. seryngae</i>	NP_791387.1	3.1.3.2	Acid phosphatase	DEG10290292	1,00E-84
<i>S. lycopersicum</i>	Q05539 CHIA_SOLLC	<i>P. seryngae</i>	NP_794777.1	3.2.1.14	Chitinase	DEG10250423	5,00E-19
<i>S. lycopersicum</i>	P21568 CYPH_SOLLC	<i>P. seryngae</i>	NP_791005.1	5.2.1.8	Peptidylprolyl isomerase	DEG10470303	2,00E-59
<i>S. lycopersicum</i>	K4BVX3_SOLLC	<i>R. solanacearum</i>	AGH83314.1	1.11.1.6*	Catalase	DEG10270348	0
<i>S. lycopersicum</i>	P15003 PER1_SOLLC	<i>R. solanacearum</i>	AGH86619.1	1.11.1.7*	Peroxidase	DEG10350205	2,00E-08
<i>S. lycopersicum</i>	Q9FVN0 AMT13_SOLLC	<i>R. solanacearum</i>	AGH84344.1	2.7.13.3	Histidine kinase	DEG10330275	1,00E-65

<i>S. lycopersicum</i>	Q05539 CHIA_SOLLC	<i>R. solanacearum</i>	AGH83721.1	3.2.1.14	Chitinase	DEG10260021	1,00E-17
<i>S. lycopersicum</i>	K4C2F1_SOLLC	<i>R. solanacearum</i>	AGH86735.1	4.2.1.1	Carbonic anhydrase	DEG10050308	4,00E-38
<i>Z. mays</i>	NP_001304298.1	<i>A. flavus</i>	B8NGN0_ASPFN	1.10.2.2	Quinol-cytochrome-c reductase	DEG20091193	1,00E-54
<i>Z. mays</i>	XP_008660914.1	<i>A. flavus</i>	B8NX24_ASPFN	1.11.1.6*	Catalase	DEG10110209	2,00E-68
<i>Z. mays</i>	XP_008664058.1	<i>A. flavus</i>	B8NC39_ASPFN	1.11.1.7*	Peroxidase	—	—
<i>Z. mays</i>	NP_001145525.1	<i>A. flavus</i>	B8N164_ASPFN	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008664254.1	<i>A. flavus</i>	B8NB79_ASPFN	2.1.1.43	Histone-lysine N-methyltransferase	DEG20051547	7,00E-12
<i>Z. mays</i>	XP_008665261.1	<i>A. flavus</i>	B8N9N8_ASPFN	2.5.1.18	Glutathione transferase	—	—
<i>Z. mays</i>	XP_008660232.1	<i>A. flavus</i>	B8NQM9_ASPFN	2.6.1.1	Aspartate transaminase	—	—
<i>Z. mays</i>	XP_008663534.1	<i>A. flavus</i>	B8N9A7_ASPFN	2.7.11.22	Cyclin-dependent kinase	DEG20010254	6,00E-67
<i>Z. mays</i>	XP_008664470.1	<i>A. flavus</i>	B8NB93_ASPFN	3.1.3.2	Acid phosphatase	—	—
<i>Z. mays</i>	XP_008656307.1	<i>A. flavus</i>	B8NQT3_ASPFN	3.2.2.22	rRNA N-glycosylase	—	—
<i>Z. mays</i>	XP_008655471.1	<i>A. flavus</i>	B8NWM8_ASPFN	4.2.1.1	Carbonic anhydrase	DEG20101870	2,00E-11
<i>Z. mays</i>	NP_001148888.1	<i>A. flavus</i>	B8NT23_ASPFN	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001149850.1	<i>A. flavus</i>	B8N7U5_ASPFN	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	6,00E-110
<i>Z. mays</i>	X P_008644870.1	<i>A. flavus</i>	B8NFW5_ASPFN	5.3.1.6	Ribose-5-phosphate isomerase	DEG10140248	3,00E-12
<i>Z. mays</i>	XP_008657765.1	<i>E. turcicum</i>	XP_008026270.1	1.1.1.27	L-lactate dehydrogenase	DEG20010346	1,00E-86
<i>Z. mays</i>	NP_001105310.2	<i>E. turcicum</i>	XP_008029291.1	1.11.1.6*	Catalase	DEG10110209	0
<i>Z. mays</i>	XP_008664058.1	<i>E. turcicum</i>	XP_008030871.1	1.11.1.7*	Peroxidase	DEG10400636	4,00E-80
<i>Z. mays</i>	NP_001145525.1	<i>E. turcicum</i>	XP_008025877.1	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008664254.1	<i>E. turcicum</i>	XP_008025860.1	2.1.1.43	Histone-lysine N-methyltransferase	DEG20240496	3,00E-18
<i>Z. mays</i>	XP_008663534.1	<i>E. turcicum</i>	XP_008024068.1	2.7.11.22	Cyclin-dependent kinase	DEG20090883	2,00E-41
<i>Z. mays</i>	XP_008651541.1	<i>E. turcicum</i>	XP_008029497.1	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008664470.1	<i>E. turcicum</i>	XP_008024834.1	3.1.3.2	Acid phosphatase	DEG10390008	1,00E-63
<i>Z. mays</i>	NP_001148888.1	<i>E. turcicum</i>	XP_008026072.1	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001136955.1	<i>E. turcicum</i>	XP_008024266.1	4.6.1.1	Adenylate cyclase	DEG10030767	2,00E-10
<i>Z. mays</i>	NP_001149850.1	<i>E. turcicum</i>	XP_008028934.1	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	1,00E-108
<i>Z. mays</i>	XP_008644870.1	<i>E. turcicum</i>	XP_008028444.1	5.3.1.6	Ribose-5-phosphate isomerase	DEG10080091	8,00E-15
<i>Z. mays</i>	XP_008657765.1	<i>C. graminicola</i>	XP_008097388.1	1.1.1.27	L-lactate dehydrogenase	DEG20010346	1,00E-91
<i>Z. mays</i>	XP_008660914.1	<i>C. graminicola</i>	XP_008098502.1	1.11.1.6*	Catalase	DEG10110209	0
<i>Z. mays</i>	NP_008664058.1	<i>C. graminicola</i>	XP_008095952.1	1.11.1.7*	Peroxidase	DEG10400636	3,00E-79
<i>Z. mays</i>	NP_001145525.1	<i>C. graminicola</i>	XP_008093145.1	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008663534.1	<i>C. graminicola</i>	XP_008094831.1	2.7.11.22	Cyclin-dependent kinase	DEG20010254	6,00E-50
<i>Z. mays</i>	XP_008651541.1	<i>C. graminicola</i>	XP_008100128.1	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008675577.1	<i>C. graminicola</i>	XP_008100081.1	3.1.1.4	Phospholipase A2	DEG20240063	2,00E-26
<i>Z. mays</i>	XP_008664470.1	<i>C. graminicola</i>	XP_008094949.1	3.1.3.2	Acid phosphatase	—	—
<i>Z. mays</i>	XP_008658269.1	<i>C. graminicola</i>	XP_008092609.1	3.1.13.4	Poly(A)-specific ribonuclease	DEG20240339	7,00E-92
<i>Z. mays</i>	XP_008677367.1	<i>C. graminicola</i>	XP_008097450.1	3.1.3.3	Phosphoserine phosphatase	DEG20211963	6,00E-52
<i>Z. mays</i>	NP_001148888.1	<i>C. graminicola</i>	XP_008096879.1	4.4.1.5	Lactoylglutathione lyase	—	—

<i>Z. mays</i>	NP_001149850.1	<i>C. graminicola</i>	XP_008091175.1	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	8,00E-113
<i>Z. mays</i>	XP_008644870.1	<i>C. graminicola</i>	XP_008098210.1	5.3.1.6	Ribose-5-phosphate isomerase	DEG10080091	1,00E-15
<i>Z. mays</i>	NP_001145525.1	<i>G. moniliformis</i>	W7LPB7_GIBM7	1.11.1.15*	Peroxioredoxin	—	—
<i>Z. mays</i>	XP_008660232.1	<i>G. moniliformis</i>	W7MC41_GIBM7	2.6.1.1	Asparatase transaminase	—	—
<i>Z. mays</i>	XP_008663534.1	<i>G. moniliformis</i>	W7MSL6_GIBM7	2.7.11.22	Cyclin-dependent kinase	DEG20011066	2,00E-36
<i>Z. mays</i>	XP_008651541.1	<i>G. moniliformis</i>	W7M0K8_GIBM7	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008658269.1	<i>G. moniliformis</i>	W7M4G2_GIBM7	3.1.13.4	Poly(A)-specific ribonuclease	DEG20240339	1,00E-88
<i>Z. mays</i>	XP_008664470.1	<i>G. moniliformis</i>	W7NDR6_GIBM7	3.1.3.2	Acid phosphatase	DEG10390008	2,00E-13
<i>Z. mays</i>	XP_008655784.1	<i>G. moniliformis</i>	W7M5R3_GIBM7	3.2.1.4	Cellulase	—	—
<i>Z. mays</i>	NP_001148888.1	<i>G. moniliformis</i>	W7LNQ2_GIBM7	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001136955.1	<i>G. moniliformis</i>	W7MFF7_GIBM7	4.6.1.1	Adenylate cyclase	DEG20090256	1,00E-90
<i>Z. mays</i>	NP_001149850.1	<i>G. moniliformis</i>	W7M917_GIBM7	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	1,00E-107
<i>Z. mays</i>	NP_001105310.2	<i>P. ananatis</i>	D4GMF4_PANAM	1.11.1.6*	Catalase	—	—
<i>Z. mays</i>	XP_008667406.1	<i>P. ananatis</i>	D4GL47_PANAM	1.11.1.15*	Peroxioredoxin	DEG10030767	1,00E-06
<i>Z. mays</i>	XP_008672910.1	<i>P. ananatis</i>	D4GCI2_PANAM	1.16.3.1*	Ferroxidase	—	—
<i>Z. mays</i>	XP_008660532.1	<i>P. ananatis</i>	D4GJ68_PANAM	2.1.3.3	Ornithine carbamoyltransferase	DEG10350142	9,00E-55
<i>Z. mays</i>	XP_008657589.1	<i>P. ananatis</i>	D4GHC5_PANAM	2.3.1.51	1-acylglycerol-3-phosphate O-acyltransferase	DEG10480294	2,00E-93
<i>Z. mays</i>	XP_008656415.1	<i>P. ananatis</i>	D4GHA1_PANAM	2.7.2.3	Phosphoglycerate kinase	—	—
<i>Z. mays</i>	XP_008662013.1	<i>P. ananatis</i>	D4GMM0_PANAM	2.7.4.8	Guanylate kinase	DEG10030351	9,00E-64
<i>Z. mays</i>	XP_008672924.1	<i>P. ananatis</i>	D4GGT2_PANAM	3.1.1.5	Lysophospholipase	—	—
<i>Z. mays</i>	XP_008651541.1	<i>P. ananatis</i>	D4GFB8_PANAM	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008650400.1	<i>P. ananatis</i>	D4GCE1_PANAM	3.1.3.11	Fructose-bisphosphatase	DEG10480226	2,00E-90
<i>Z. mays</i>	XP_008672875.1	<i>P. ananatis</i>	D4GMQ4_PANAM	4.2.1.96	4a-hydroxytetrahydrobiopterin dehydratase	DEG10470424	3,00E-34
<i>Z. mays</i>	NP_001105425.1	<i>P. ananatis</i>	D4GK89_PANAM	4.3.3.7	4-hydroxy-tetrahydrodipicolinate synthase	DEG10180422	1,00E-20

*Enzimas do sistema antioxidante.

** Número de acesso do DEG.

No metabolismo de aminoácidos, muitas enzimas identificadas como essenciais e análogas, como a anidrase carbônica para *R. solanacearum* e *A. flavus* AF70; prolyl aminopeptidase, para *F. oxysporum* 4287; transaminase, para *A. flavus* AF70, *G. moniliformis*, *A. flavus* NRRL3357 e *F. oxysporum* Fo5176. (Tabela 6).

Enzimas análogas essenciais também foram encontradas no metabolismo de lipídios e na via de biossíntese de metabólitos secundários. Acetyl-CoA carboxylase foi identificada em *X. axonopodis* e a fosfolipase A2 em *C. graminicola*. A ornitina carbamoiltransferase, identificada em *P. ananatis*, participa do metabolismo dos aminoácidos (Tabela Suplementar 2). Algumas atividades enzimáticas encontradas essenciais para alguns agentes patogênicos não foram identificadas como essenciais em outros: esses casos são representados por enzimas codificadas por diferentes genes. Neste grupo, podemos citar enzimas pertencentes ao sistema antioxidante (AS), composto por enzimas envolvidas com a desintoxicação de espécies reativas de oxigênio (ROS), como catalase, peroxidase, superóxido dismutase, peroxiredoxina, dentre outras.

4.4 Enzimas análogas no sistema antioxidante

Um grupo de enzimas que se destacaram entre as NISEs validadas e essenciais, incluindo atividades não essenciais, foram as enzimas que compõem o sistema antioxidante (SA). Em todas as comparações feitas entre plantas e seus agentes patogênicos, exceto no caso de *B. cinerea*, para pelo menos uma das atividades funcionais do sistema antioxidante, a enzima do hospedeiro e sua contraparte no patógeno são estruturalmente diferentes (Tabela 4). No total, foram encontrados 27 casos de analogia para o sistema antioxidante, incluindo catalase (CAT), peroxidase (POX), superóxido dismutase (SOD), ferroxidase (HEPH) e peroxiredoxina (PRDX). Em nossos resultados, o CAT foi identificado como uma enzima essencial para 9 dos 14 patógenos estudados, e POX foi identificado como essencial em *E. turcicum*, *C. graminicola*, *P. seryngae* e *R. solanacearum*. SOD foi identificada como uma enzima essencial para *X. axonopodis*. Entre os patógenos analisados, existem duas espécies com cepas distintas, *A. flavus* (NRRL3357, AF70)

e *F. oxysporum* (Fo5176, 4287), no entanto não foram observadas diferenças entre as linhagens. É importante enfatizar que as atividades enzimáticas do SA estão presentes em todos os genomas incluídos no presente trabalho; no entanto, apenas os casos de NISEs validados foram mostrados, o que explica lacunas no padrão de ausência/presença observado para HEPH, PRDX e SOD (Tabela 7).

Tabela 7 Formas enzimáticas alternativas encontradas entre as enzimas do sistema antioxidante.

Organismos	Formas estruturais				
	CAT	POX	SOD	HEPH	PRDX
<i>G. max</i>	① ⑤	③ ⑥ ⑳	① ④ ⑥ ⑦		
<i>A. flavus</i>	②	⑫			
<i>F. oxysporum</i>	②			⑭	
<i>P. sojae</i>	②				
<i>S. sclerotiorum</i>		⑫			
<i>X. axonopodis</i>	③				
<i>S. lycopersicum</i>	① ⑤	③ ⑥	① ④ ⑥ ⑦		
<i>B. cinerea</i>					
<i>F. oysporum</i>	②			⑭	
<i>M. perniciosa</i>		⑫			
<i>P. syringae</i>	②	⑫ ⑮			
<i>R. solanacearum</i>		⑮			
<i>Z. mays</i>	① ⑤	③		② ⑥	① ⑩
<i>A. flavus</i>	②	⑫			⑨
<i>C. graminicola</i>	②	⑥			⑨
<i>E. turcicum</i>	②	⑥			⑨
<i>G. moniliformis</i>					⑨
<i>P. ananatis</i>	③			⑦	②

¹ *A. flavus* NRRL3357, ² *F. oxysporum* Fo5176, ³ *F. oxysporum* 4287, ⁴ *A. flavus* AF70.

* Os números representam os grupos em que uma sequência foi localizada. Apenas casos validados de analogia são mostrados. Círculos pretos indicam formas estruturais validadas encontradas apenas nos patógenos.

4.5 Formas estruturais específicas das enzimas análogas

Após a obtenção da lista final de NISEs validadas e essenciais entre os hospedeiros da planta e seus agentes patogênicos, a busca dessas atividades enzimáticas foi realizada nas proteínas preditas de *H. sapiens*, *A. mellifera*, *B.*

subtilis e *T. harzianum*. O objetivo desta comparação foi encontrar estruturas das enzimas específicas do patógeno em comparação aos genomas de espécies que não devem ser afetadas por um eventual inibidor visando essa forma estrutural particular, principalmente entre *H. sapiens* e *A. mellifera*. Das 97 NISEs validadas, foram encontradas 68 formas estruturais específicas do patógeno (em relação ao hospedeiro da planta, homens e abelhas) (Tabela 8). Elas estão distribuídas em 26 atividades enzimáticas (16 delas são essenciais). A partir destas 68 formas estruturais, 39 foram encontradas em *T. harzianum* e 17 em *B. subtilis*, o que é esperado, uma vez que esses organismos pertencem aos mesmos reinos dos fitopatógenos estudados neste trabalho (fungi e monera).

Tabela 8. Formas estruturais enzimáticas específicas dos fitopatógenos. Os números representam os grupos em que foram alocadas as enzimas dos organismos, as enzimas alocadas em um mesmo grupo são consideradas homólogas, as alocadas em diferentes grupos são consideradas análogas.

Comparação				Formas estruturais					
Planta**	Patógeno**	Número EC	ID Sequencia dos Patógenos	Patógenos	Planta	<i>H. sapiens</i>	<i>A. mellifera</i>	<i>T. harzianum</i>	<i>B. subtilis</i>
Gm	Af	1.11.1.6‡	XP_002384918.1	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Gm	Af	1.11.1.7	XP_002377297.1	3, 6, 12	3, 6, 20	1, 3	1, 3, 6	3, 6, 12*	7
Gm	Af	2.6.1.1	XP_002382374.1	1, 5	1	1	1	1, 5*	1
Gm	Fo	1.11.1.6‡	F9FP11_FUSOF	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Gm	Fo	1.15.1.1‡	F9FYF1_FUSOF	1, 4, 7, 14	1, 4, 6, 7	1, 4, 7	1, 4, 7	1, 4, 7, 14*	1, 4
Gm	Fo	2.6.1.1	F9G466_FUSOF	1, 5	1	1	1	1, 5*	1
Gm	Fo	4.4.1.5	F9G2J4_FUSOF	1, 3	1, 8	1	1	1, 3*	1, 3, 6, 7, 11
Gm	Ps	1.11.1.6‡	XP_009521283.1	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Gm	Ss	1.11.1.7	XP_001585507.1	3, 6, 12	3, 6, 20	1, 3	1, 3, 6	3, 6, 12*	7
Gm	Xa	1.11.1.6	WP_042823856.1	3	1, 5	1, 5	1, 5	1, 2	1, 3*, 8
Gm	Xa	6.4.1.2‡	WP_033483073.1	1, 6	1	1	2	1	1, 6*
SI	Bc	3.1.3.2	XP_001560519.1	2, 3, 7, 13	2, 6, 9, 11	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13*	—
SI	Bc	4.4.1.5	XP_001550649.1	1, 3	1, 8	1	1	1, 3*	1, 3*
SI	Fo	1.11.1.6‡	A0A0D2YKD1_FUSO4	1, 2, 5	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
SI	Fo	1.15.1.1	A0A0D2YE80_FUSO4	1, 4, 7, 14	1, 4, 6, 7	1, 4, 7	1, 4, 7	1, 4, 7, 14*	1, 4
SI	Fo	3.1.3.2	A0A0D2YGA3_FUSO4	1, 2, 3, 4, 7, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13	—
SI	Fo	3.2.1.4	A0A0D2XJE6_FUSO4	1, 6	1	—	1	1	1, 3
SI	Fo	4.4.1.5	A0A0D2XLV4_FUSO4	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
SI	Mp	1.11.1.7	E2LX62_MONPE	6, 12	3, 6	1, 3	1, 3, 6	3, 6, 12*	7
SI	Mp	3.4.11.1	E2LYM3_MONPE	1, 11	1	1	1	—	1
SI	Psy	1.11.1.6‡	NP_794283.1	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
SI	Psy	1.11.1.7‡	NP_794565.1	6, 16, 18, 19	3, 6	1, 3	1, 3, 6	3, 6, 12	7
SI	Psy	3.1.3.2‡	NP_791387.1	1, 3	2, 6, 9, 11	2, 4, 7	2, 4, 5, 7, 20	1*, 3, 4, 5, 7, 13	—
SI	Psy	3.2.1.14‡	NP_794777.1	1, 3	1	1, 10	1, 4	1	—
SI	Rs	1.11.1.6‡	AGH83314.1	6	1, 5	1, 5	1, 5	1, 2	1, 3, 8
SI	Rs	1.11.1.7‡	AGH86619.1	6, 18	3, 6	1, 3	1, 3, 6	3, 6, 12	7

SI	Rs	2.7.13.3‡	AGH84344.1	1, 21, 23, 24, 33, 36	1, 20	2, 12, 13, 20	12, 20	1	1
SI	Rs	3.2.1.14‡	AGH83721.1	3	1	1, 10	1, 4	1	—
SI	Rs	4.2.1.1‡	AGH86735.1	1, 3, 13	1, 2, 5	1, 2	1, 2, 3	1, 2	1, 3, 5, 12
Zm	Af	1.11.1.15	B8N164_ASPFN	1, 9	1, 10	1	1	1, 9*	1
Zm	Af	1.11.1.6‡	B8NX24_ASPFN	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Zm	Af	1.11.1.7	B8NC39_ASPFN	3, 6, 12	3	1, 3	1, 3, 6	3, 6, 12*	7
Zm	Af	2.6.1.1	B8NQM9_ASPFN	1, 5	1	1	1	1, 5*	1
Zm	Af	3.1.3.2	B8NB93_ASPFN	2, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13*	—
Zm	Af	3.2.2.22	B8NQT3_ASPFN	5	1, 7	—	—	—	—
Zm	Af	4.4.1.5	B8NT23_ASPFN	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Zm	Af	5.3.1.6‡	B8NFW5_ASPFN	1, 2	1	1	1	1, 2*	—
Zm	Cg	1.1.1.27‡	XP_008100733.1	2, 12	1, 12	1, 12	1, 12	1, 2*, 12	1, 11
Zm	Cg	1.11.1.15	XP_008093145.1	1, 9	1, 10	1	1	1, 9*	1, 9*
Zm	Cg	1.11.1.6‡	XP_008098502.1	1, 2, 5	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Zm	Cg	3.1.1.31	XP_008100128.1	1, 2	1	1, 4	1	1, 2*	2*
Zm	Cg	3.1.3.2	XP_008094949.1	2, 3, 5, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13*	—
Zm	Cg	4.4.1.5	XP_008096879.1	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Zm	Cg	5.3.1.6‡	XP_008098210.1	1, 2	1	1	1	1, 2*	2*
Zm	Et	1.11.1.15	XP_008025877.1	1, 9	1, 10	1	1	1, 9*	1
Zm	Et	1.11.1.6‡	XP_008029291.1	1, 2, 5	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Zm	Et	3.1.1.31	XP_008029497.1	1, 2	1	1, 4	2	1, 2*	2*
Zm	Et	3.1.3.2‡	XP_008024834.1	1, 2, 3, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	1*, 3, 4, 5, 7, 13	—
Zm	Et	4.4.1.5	XP_008026072.1	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Zm	Et	4.6.1.1‡	XP_008024266.1	2, 8, 10	2, 17, 18	2, 8	2, 6, 8, 13	2, 8	4
Zm	Et	5.3.1.6‡	XP_008028444.1	1, 2	1	1	1	1, 2*	—
Zm	Gm	1.11.1.15	W7LPB7_GIBM7	1, 2, 9	1, 10	1	1	1, 9*	1
Zm	Gm	2.6.1.1	W7MC41_GIBM7	1, 5	1	1	1	1, 5*	1
Zm	Gm	3.1.1.31	W7M0K8_GIBM7	1, 2	1	1, 4	1	1	2*
Zm	Gm	3.1.3.2‡	W7NDR6_GIBM7	1, 2, 3, 7, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2	—
Zm	Gm	3.2.1.4	W7M5R3_GIBM7	1, 6	1	—	1	1	1, 3
Zm	Gm	4.4.1.5	W7LNQ2_GIBM7	1, 3	1, 8	1	1	1	1, 3*, 6, 7, 11
Zm	Pa	1.11.1.15‡	D4GL47_PANAM	1, 2	1, 10	1	1	1	1
Zm	Pa	1.11.1.6	D4GMF4_PANAM	3, 5, 6	1, 5	1, 5	1, 5	1	1, 3*, 8
Zm	Pa	1.16.3.1	D4GCI2_PANAM	2, 7	2, 6	2, 4, 6	2, 6	6	1
Zm	Pa	2.1.3.3‡	D4GJ68_PANAM	2	1, 10	1, 10	10	1	1, 10
Zm	Pa	2.7.2.3	D4GHA1_PANAM	3	1	1	1	1	1, 3*, 9
Zm	Pa	2.7.4.8	D4GMM0_PANAM	1, 4, 7	1	1	1	1	1, 7

Zm	Pa	3.1.1.31	D4GFB8_PANAM	2, 6	1	1, 4	1	1	2*
Zm	Pa	3.1.1.5	D4GGT2_PANAM	2, 5	6, 7, 18	1, 4, 6, 7, 9, 10, 17, 18	1, 6, 7, 9, 17, 18	1	—
Zm	Pa	3.1.3.11‡	D4GCE1_PANAM	10, 12	1	1, 8	1	1	3, 11
Zm	Pa	4.2.1.96‡	D4GMQ4_PANAM	2	1	1	1	1	2*
Zm	Pa	4.3.3.7‡	D4GK89_PANAM	1, 2	1	—	—	—	1, 4

**Gm: *G. Max*, Af: *A. flavus*, Fo: *F. oxysporum*, Ps: *P. sojae*, Ss: *S. sclerotiorum*, Xa: *X. axonopodis*, Sl: *S. lycopersicum*, Rs: *R. solanacearum*, Psy: *P. syringae*, Mp: *M. perniciosus*. Bc: *B. cinerea*, Zm: *Z. mays*, Pa: *P. ananatis*, Gm: *G. moniliformis*, Et: *E. turcicum*, Gg: *C. graminicola*.

ΔOs números representam as diferentes estruturas. Os números em negrito são as formas estruturais enzimáticas específicas dos fitopatógenos.

‡ Enzimas essenciais.

— Atividade enzimática não encontrada.

* Forma estrutural homóloga ao patógeno.

5. DISCUSSÃO

As enzimas são um dos principais alvos para o desenvolvimento de inibidores de qualquer tipo; no entanto, espécies que compartilham funções enzimáticas essenciais podem ser inadvertidamente afetadas por produtos desenvolvidos com outras aplicações em mente [203]. Os agrotóxicos são comumente orientados para essas funções, e seus efeitos prejudiciais em várias espécies, incluindo o próprio homem e várias espécies vitais, como polinizadores e microorganismos benéficos, são motivo de grande preocupação [204-207].

5.1 Essencialidade enzimática

A partir da comparação de dados de estrutura primária, dados de estrutura terciária e dados de essencialidade, começando com 444198 sequências individuais, compreendendo 2096 atividades enzimáticas em 3 plantas e 15 fitopatógenos, descrevemos um subconjunto de sequências análogas em 29 atividades enzimáticas essenciais presentes tanto na planta como no patógeno. Estes pertencem a vários componentes do metabolismo central das plantas e agentes patogênicos, envolvidos no metabolismo dos carboidratos, metabolismo dos aminoácidos, detoxificação de espécies reativas de oxigênio e outros, oferecendo assim várias oportunidades como alvos. As enzimas selecionadas neste estudo constituem um grupo forte de candidatos pois sua maioria pertencem a rotas essenciais à manutenção do metabolismo de agentes patogênicos.

Algumas das diversidades encontradas para as enzimas do sistema antioxidante, tanto em termos de atividades enzimáticas como em formas estruturais, podem ser explicadas por pressões evolutivas: durante a co-evolução entre plantas e seus agentes patogênicos, é provável que diferentes enzimas antioxidantes das plantas se adaptaram para superar os mecanismos de virulência dos patógenos [208,209]. O papel dessas enzimas em mecanismos de virulência, susceptibilidade a infecções, desenvolvimento de alvos de drogas e avaliação de

efeitos de pesticidas foi estudado para SOD [39, 210-213], CAT [214-216] e POX [217].

As enzimas essenciais do metabolismo central também foram estudadas como potenciais alvos de fármacos em diversos organismos. A enzima glicose-6-fosfato isomerase foi estudada como alvo para infecções causadas por *Plasmodium falciparum* [218], *Trypanosoma* spp [219], *Toxoplasma gondii* [220] e *Leishmania* spp. [221], acetil-CoA carboxilase para *L. major* [222-223] e ribose-5-fosfato isomerase para *Escherichia coli* [224]. Dezoito das 29 atividades enzimáticas identificadas neste estudo como análogas e essenciais foram identificadas em bancos de dados de alvos de fármacos, como TDR Drug Targets (<http://tdrtargets.org/>), DrugBank (<https://www.drugbank.ca/>) e *Potential Drug Target Database* (PDTD) (<https://omictools.com/pdtd-tool>), o que significa que elas estão sendo estudadas ou empregadas como alvo de fármacos para pelo menos um patógeno. Entre elas, podemos citar enzimas do metabolismo de carboidratos e aminoácidos, tais como a lactoilglutamina liase, a acetil-CoA carboxilase, a anidrase carbônica e enzimas do SA como catalase, peroxidase, peroxiredoxina e superóxido dismutase. Uma vez que estas atividades enzimáticas apresentam múltiplas estruturas terciárias, não podemos dizer, a partir desses dados públicos, qual forma em particular está sendo estudada; no entanto, essas descobertas dão suporte direto às nossas análises, corroborando a ideia de que as enzimas análogas essenciais com formas estruturais específicas têm um grande potencial como alvos de drogas, conforme descrito em nosso estudo. As melhorias na anotação de genes e seus produtos e uma melhor caracterização experimental das atividades enzimáticas permitiriam o uso de critérios menos rigorosos em nossos procedimentos, principalmente na limpeza e filtragem de dados, mas também em agrupamento e validação estrutural, aumentando o número de enzimas essenciais e análogas que poderiam ser estudadas como potenciais alvos de fármacos.

5. 1. 2 Enzimas do Sistema Antioxidante

Um grupo que se destacou entre as enzimas essenciais foram as que compõem o Sistema Antioxidante (SA). Os organismos possuem múltiplos mecanismos de defesa contra espécies reativas de oxigênio (EROs), com um sistema enzimático muito eficiente, composto por diversas enzimas, tais como: superóxido dismutase (SOD), catalase (CAT), peroxidase (POX), glutathione redutase (GR), ferroxidase, dentre outros [225-227]. As EROs são geradas a partir de fontes ambientais ou por funções celulares normais do organismo, e incluem radicais livres (superóxido e radicais hidroxílicos), espécies não-radicaais de oxigênio (peróxido de hidrogênio e peroxinitrito) e lipídios reativos e carboidratos (cetoaldeídos, hidroxinonais) [228-229].

Durante uma invasão por agentes patogênicos, observa-se o mecanismo de explosão oxidativa, que consiste em uma produção rápida e transitória de EROs como peróxido de hidrogênio (H_2O_2) e superóxido (O_2); estes acumulam-se no apoplasto durante a explosão oxidativa e são extremamente tóxicos para os microorganismos invasores. Os danos causados à dupla fita de DNA podem resultar em modificações dos pares de bases de nucleotídeos causando mutações, envelhecimento celular e morte celular [53]. As EROs ainda podem produzir polímeros reticulados na parede celular para fortalecer as barreiras físicas contra os patógenos invasores ou atuar como um mensageiro em uma via de sinalização celular, levando à expressão de genes PR (PR - *pathogenesis related*) relacionados à patogênese provocando uma resposta rápida de hipersensibilidade [230-232].

As plantas desenvolveram um sistema complexo e elaborado para combater os danos causados pelas EROs. Como dito anteriormente, é provável que durante a co-evolução entre plantas e fitopatógenos, diferentes enzimas antioxidantes das plantas sofreram adaptações para superar o mecanismo de virulência dos patógenos [233,234]. Estes casos são particularmente interessantes, porque o sistema antioxidante apresenta-se com a primeira linha de defesa de ambas as partes. Este arsenal desempenha um papel crítico na detoxificação de espécies reativas de oxigênio durante as interações hospedeiro-patógeno, sendo, portanto, um importante fator de virulência [235,236]. Bloquear ou inibir essas enzimas, em

princípio, poderá diminuir a virulência do patógeno e / ou atrasará a resposta de defesa contra os radicais livres utilizados pela planta como mecanismos de defesa durante os ataques.

Dada a importância dessas enzimas, diversos estudos foram conduzidos a fim de elucidar mecanismos de virulência e susceptibilidade envolvidos na interação planta vs patógeno. A Superóxido Dismutase (SOD) é um exemplo de enzimas que possuem um papel crítico na defesa do organismo contra as EROs e é um importante fator de virulência. Estudos conduzidos com *Cryptococcus neoformans*, utilizando cepas mutantes para o gene *SOD1*, o qual codifica a SOD dependente de Cu / Zn, é facilmente atacado por EROs do hospedeiro e é significativamente menos virulento do que a cepa selvagem [211]. Em experimentos envolvendo o fungo patogênico *Isaria fumosorosea*, o uso de Cu + Zn provou ser o indutor mais ativo da atividade de SOD. O tratamento com Cu + Zn aumentou a tolerância fúngica ao estresse oxidativo gerado pela menadiona (Vitamina K) no meio (0-3 mM), conforme avaliado pelo crescimento de colônias [237].

Os agrotóxicos ciprodinil e fludioxonil, pertencentes ao grupo químico anilino pirimidina e cianopirrole, de classificação toxicológica nível I, são fungicidas de amplo espectro comumente utilizados nas lavouras de soja, milho, tomate e batata, dentre outros. Estudos anteriores reportaram que o modo de ação do ciprodinilactos inibem a biossíntese de metionina e outros aminoácidos tiônicos em fungos [238-240]. Devido às propriedades lipofílicas do ciprodinil e à ruptura das membranas celulares pelo fludioxonil, estes defensivos agrícolas penetram nos eritrócitos e afetam a enzima CuZnSOD em humanos. Experimentos envolvendo estes defensivos agrícolas revelaram que o fludioxonil inibiu o CuZnSOD não competitivo e o ciprodinil inibiu CuZnSOD competitivamente. Embora as moléculas de ciprodinil não se assemelhem a moléculas de superóxido, estas constituem radicais que competem com moléculas de superóxido para se ligar à região ativa da enzima [241]. Outro estudo, observou a inibição da atividade da CAT induzida por esses fungicidas. Os experimentos revelaram que o ciprodinil inibiu a CAT de forma competitiva e o fludioxonil inibiu a enzima de forma não competitiva [212].

Outros estudos envolvendo mutantes para *catB* (silenciamento do gene que codifica catalase) de *X. oryzae* e *X. oryzaicola* mostrou diminuição da atividade total da CAT e aumento da sensibilidade ao ácido fenazina-1-carboxílico, um produto

natural isolado de *Pseudomonas* spp. usado para controlar muitas doenças importantes do arroz [242]. Também, experimentos envolvendo a técnica de qPCR em tempo real revelaram que os níveis de expressão de *catB* e *oxyR* (um provável regulador de genes desintoxicantes de ROS) em *X. oryzae* foram significativamente induzidos por H_2O_2 e a ausência de um desses genes reduz drasticamente a viabilidade da bactéria na presença de H_2O_2 extracelular [101]. Em análises de cepas mutantes de *P. seryngae* pv tomato para o gene *OxyR*, observou-se a redução dos sintomas da doença e redução da população de bactérias em *S. lycopersicum* e *Arabidopsis thaliana* [216]. Estudos prévios com *Pseudomonas seryngae* demonstraram a relação de três isoformas de CAT (*Katb*, *KatE*, *KatG*) com a virulência, e como a manutenção das funções destas enzimas (que têm diferentes papéis biológicos no organismo) é essencial para o sucesso da infecção e colonização do hospedeiro [243].

As enzimas do sistema antioxidante, por serem enzimas de caráter ubíquo nas vias de desintoxicação enzimática, são encontradas em todos os organismos aeróbicos e podem catalisar a dismutação direta de peróxido de hidrogênio (H_2O_2) em água (H_2O) e oxigênio (O_2 modo catalítico) ou H_2O_2 para oxidar substratos, como o metanol, etanol, formaldeído, formato ou nitrito (modo peroxidático). Ainda participam de uma grande rede de sinalização altamente conservada entre organismos aeróbicos e controla uma ampla gama de processos biológicos, como crescimento, desenvolvimento e respostas a estímulos bióticos e / ou abióticos [244-246].

5.1.3 Enzimas do metabolismo central, metabolismo de carboidratos e metabolismo energético

As enzimas ribulose-fosfato 3-epimerase (*Rpe*) e ribose-5-fosfato isomerase (*Rpi*) participam das vias das pentoses-fosfato [247]. *Rpi* é necessário para a geração de NADPH, que é uma fonte de redução de energia e uma via chave do metabolismo da glicose na maioria dos organismos, sendo também necessária para a biossíntese de ácidos nucleicos e aminoácidos [248]. Esta via também tem sido

descrita como importante para a proteção de leveduras contra o estresse oxidativo, uma vez que o NADPH é um cofator essencial para enzimas dependentes de glutatona e tioredoxina, as quais defendem as células contra o dano oxidativo [249]. Em estudos com câncer em humanos, a via da pentoses-fosfato tem sido amplamente estudada, oferecendo potenciais alvos de fármacos contra a doença [250]. Existem duas formas da enzima Rpi, RpiA e RpiB, ambas sem similaridade detectável, tanto entre suas sequências como estruturas [251,252]. A essencialidade da enzima Rpi também foi observada em *Escherichia coli*, que possui ambas as formas (RpiA e RpiB). O duplo mutante da RpiA/RpiB resulta em disfunção severa do crescimento bacteriano [253]. Em *Leishmania donovani*, agente causal de um tipo de leishmaniose, esta enzima tem sido utilizada para direcionar o desenvolvimento de fármacos antileishmania [254]. Em *Trypanosoma brucei*, o silenciamento de Rpi por RNAi resultou em lentidão no crescimento e infectividade do patógeno, destacando a importância desta enzima como um alvo de fármacos [255].

A anidrase carbônica (CA) foi descrita como NISE na interação *S. lycopersicum* e *R. solanacearum*. Esta enzima desempenha um papel importante no transporte iônico e na regulação do pH em vários organismos. Estudos utilizando *A. mellifera* como modelo determinaram o risco de agrotóxicos na enzima CA, que foi exposta a pesticidas, incluindo tebuconazol, propoxur, carbaryl, carbofurano, simazina e atrazina. Os seis pesticidas inibiram, de uma forma dependente da dose, a atividade CA *in vitro* em concentrações micromolares. O mecanismo pelo qual estes compostos inibem a CA é desconhecido até o presente momento, mas é evidente os danos causados aos polinizadores pelo uso desses pesticidas [256].

A via da glioxalase envolve a participação de duas enzimas, Glyoxalase I e II (GLX-I e GLX-II). Estas enzimas desintoxicam o metilglioxal (MG), um 2-oxoaldeído tóxico que ocorre em todas as células vivas como subproduto da glicólise. Em altas concentrações, o metilglioxal pode danificar proteínas, ácidos nucleicos e lipídios através da modificação de arginina, lisina, cisteína, adenina e guanina, formando vários produtos finais de glicação avançada [257,258]. A via do glioxilato é funcional ao longo da vida do organismo e de caráter onipresente, sugerindo sua importância na função celular. Trabalhos anteriores mostraram as diferenças estruturais entre glioxilases de humanos e de tripanossomatídeos, e dada a essencialidade da via, essas enzimas foram identificadas como alvos potenciais para o tratamento de

leishmanioses [259,260]. Outros estudos indicam que o gene responsável pelo silenciamento da expressão de GLX-I em fungos patogênicos de plantas tem sido eficaz na diminuição da virulência; esta enzima tem sido relatada por possuir um papel importante de apoio envolvendo os processos de desintoxicação do fungo juntamente com o sistema antioxidante [261-264].

5.2 Enzimas não essenciais

Curiosamente, o subconjunto de NISEs não essenciais contém várias enzimas importantes no contexto das interações hospedeiro-patógeno, como celulasas, quitinases, glutathione transferase e lisofosfolipase. Bloquear ou inibir essas enzimas, em princípio, também diminuirá a virulência e/ou atrasará os mecanismos de defesa do patógeno [265]. A inibição de celulasas e quitinases também foi proposta como estratégia para o desenvolvimento de novos fármacos antifúngicos para a aspergilose em seres humanos [266]. A glutathione transferase desempenha um papel essencial na proteção de fungos necrótrofos contra metabólitos tóxicos derivados de plantas e espécies reativas de oxigênio [267], enquanto que a lisofosfolipase foi relacionada com a virulência em *C. neoformans* [262].

A enzima 6-fosfogluconolactonase, que participa do segundo passo da via das pentoses-fosfato, foi detectada em vários organismos neste estudo. Foi descrito em *E. coli* que mutantes para a 6-fosfogluconolactonase mostram um retardo expressivo do crescimento com a diminuição da eficiência do segundo passo da via das pentoses-fosfato. A enzima 6-fosfogluconolactonase catalisa a hidrólise de uma ligação éster, resultando na formação de 6-fosfogluconato a partir do 6-fosfato de D-gluconolactona. A enzima 6-fosfogluconato desidrogenase catalisa a oxidação do aldeído (hemiacetal) no C1 da glicose-6-fosfato, para um ácido carboxílico, numa ligação éster (lactona). Ao final, forma-se CO₂ e ocorre redução do NADP⁺ a NADPH. O NADPH é usado pela célula como um agente redutor em uma variedade de processos biossintéticos, bem como para a defesa contra as espécies de oxigênio, mantendo glutathione em seu estado reduzido [268].

5.3 Enzimas descartadas

Durante o processo de validação das enzimas análogas foi observada a presença de enzimas heteromultiméricas, que são formadas por diferentes subunidades e codificadas por genes não relacionados, com diferentes funções e origens evolutivas. Na maioria dos casos, as proteínas estavam anotadas com os termos “*subunit*”. A ocorrência destes casos diminuiu a precisão da detecção de NISEs e por isso foram excluídas da lista final de alvos [121,122]. Quanto às ocorrências de enzimas análogas classificadas como preditas pelo banco de dados SUPERFAMILY, estas possuíam informações incompletas em relação à sua superfamília e ao seu padrão de enovelamento. Para estas enzimas só foi possível obter a informação sobre as diferenças entre suas estruturas primárias, e inferir sua atividade enzimática. Com o aumento do número de estruturas resolvidas depositadas no PDB, logo será possível classificar estas enzimas quanto ao seu padrão de enovelamento. Estudos mais aprofundados dos casos de enzimas análogas preditas podem contribuir para enriquecer o panorama atual de enovelamentos proteicos.

6. PERSPECTIVAS

A metodologia empregada neste estudo pode ser aperfeiçoada e informações adicionais podem ser incorporadas, a fim de ampliar a listas de alvos validados com potencial biotecnológico para aplicação na agricultura. Com informações mais precisas de estudos futuros sobre dados de estrutura proteica, será possível refinar ainda mais os dados obtidos e as estratégias de comparação. As principais perspectivas são pontuadas a seguir:

- Realização de busca extensiva dos prováveis alvos apontados neste estudo nos genomas de outras espécies inadvertidamente afetadas pela aplicação de agrotóxicos, a fim de excluir casos de homologia;
- Validação experimental dos prováveis alvos apontados nos patógenos incluídos neste estudo;
- Identificação dos sítios ativos das enzimas para comprovação da essencialidade dos alvos propostos para cada patógenos;
- Ampliar a lista de culturas de plantas e patógenos estudados, priorizando culturas que detêm o maior consumo de agrotóxicos;

7. CONCLUSÕES

- A abordagem empregada neste estudo permitiu a elaboração de listas de enzimas essenciais e análogas, pertencendo ao metabolismo central e / ou envolvidas em interações hospedeiro-patógeno, com potencial para ser alvo de fungicidas e bactericidas;
- Essas enzimas proporcionam uma oportunidade para a descoberta de alvos com diferenças estruturais consideráveis em relação a suas contrapartes em organismos benéficos, como polinizadores, fungos e bactérias promotoras do crescimento em plantas;
- A inclusão de dados estruturais permite a divulgação de formas estruturais específicas, facilitando o desenvolvimento de inibidores enzimáticos amigáveis ao meio ambiente, que podem ser de grande importância para o uso agrícola;
- O tema abordado no estudo aponta para um problema de saúde pública por vezes negligenciado. A pesquisa se concentrou em apontar uma alternativa ao uso de agrotóxicos, evidenciando a necessidade de estudos e investimentos na em uma agricultura limpa e sustentável.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. MAPA, Ministério da Agricultura, Pecuária e Abastecimento - Projeções do Agronegócio: Brasil 2012/2013 a 2022/2023. Assessoria de Gestão Estratégica. Brasília, 2013. Disponível em: <http://www.agricultura.gov.br>.
2. MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Saldo da balança do agro em junho é de US\$ 8,17 bilhões. 2018. Disponível em: <http://www.agricultura.gov.br>.
3. CONAB. Companhia Nacional de Abastecimento. Supersafra de grãos aumenta e deve atingir 238 milhões de toneladas. 2017. Disponível em: <https://www.conab.gov.br>.
4. MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Relações Internacionais do Ministério da Agricultura, Pecuária e Abastecimento. 2018. Disponível em: <http://www.agricultura.gov.br>.
5. MAPA - Ministério da Agricultura, Pecuária e Abastecimento. Ministério da Agricultura, Pecuária e Abastecimento. Projeções do Agronegócio: Brasil 2013/14 a 2023/24. Brasília. 2014. Disponível em: <http://www.agricultura.gov.br>.
6. UN DESA. United Nations Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. 2015. Disponível em: <https://esa.un.org>.
7. MAPA. Ministério da Agricultura Pecuária e Abastecimento. Boas práticas agrícolas. 2017. Disponível em: <http://www.agricultura.gov.br>.
8. ANVISA. Agência Nacional De Vigilância Sanitária; Universidade Federal do Paraná. Seminário Mercado De agrotóxico e regulação, 2012. Brasília: Anvisa. 2012. Disponível em: <http://portal.anvisa.gov.br/>.
9. BRASIL. Lei nº 7.802, de 11 de julho de 1989. Dispõe sobre a pesquisa, a produção, a embalagem e rotulagem, o transporte, o armazenamento, a comercialização, a propaganda comercial, a utilização, a importação, a exportação, o destino final dos resíduos e embalagens, o registro, a classificação, o controle, a inspeção e a fiscalização de agrotóxicos, seus componentes e afins, e dá outras providências. In: Legislação federal de agrotóxicos e afins. Brasília (DF): Ministério da Agricultura e do Abastecimento. 1998: 7-13.
10. IBAMA. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. Produtos agrotóxicos e afins comercializados em 2009 no Brasil: uma abordagem ambiental. 2010. Disponível em: <https://www.ibama.gov.br>.

11. Oliveira E, Maggi MF, Matos E, Ramos MS, Vagner MW, Lopes EC. Technology of application of defensives and relations with the risk of contamination of the water and soil. *Pesquisa Aplicada & Agrotecnologia*. 2009;2(3):161-169.
12. EMBRAPA. Ministério da Agricultura, Pecuária e Abastecimento. Pragas mais resistentes a químicos exigem mudança nas estratégias de controle. 2014. Disponível em: <https://www.embrapa.br>.
13. Talamini V, Stadnik MJ. Extratos vegetais e de algas no controle de doenças de plantas. In: Stadnik MJ, Talamini V. (Eds.) *Manejo Ecológico de Doenças de Plantas*. Florianópolis, CCA/UFSC. 2004.
14. Dossiê ABRASCO: um alerta sobre os impactos dos agrotóxicos na saúde / Organização de Fernando Ferreira Carneiro, Lia Giraldo da Silva Augusto, Raquel Maria Rigotto, Karen Friedrich e André Campos Búrigo. - Rio de Janeiro: EPSJV; São Paulo. Expressão Popular. 2015: 624
15. Almeida VES, Friedrich K, Tygel AF, Melgarejo L, Carneiro FF. Uso de sementes geneticamente modificadas e agrotóxicos no Brasil: cultivando perigos. *Ciência & Saúde Coletiva*. 2017; 22(10): 3333-3339.
16. BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância em Saúde Ambiental e Saúde do Trabalhador. *Agrotóxicos na ótica do Sistema Único de Saúde / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância em Saúde Ambiental e Saúde do Trabalhador*. Brasília: Ministério da Saúde. 2016:(2).
17. IBGE. Instituto Brasileiro de Geografia e Estatística. Censo agropecuário 2017. Disponível em: <https://censos.ibge.gov.br>.
18. ANVISA. Agência Nacional De Vigilância Sanitária. Programa de análise de resíduos de agrotóxicos em alimentos (PARA). Relatório de atividades de 2011 e 2012. Disponível em: <http://portal.anvisa.gov.br/>. 2013
19. ANVISA. Agência Nacional De Vigilância Sanitária. Programa de análise de resíduos de agrotóxicos em alimentos (PARA). Relatório das análises de amostras monitoradas no período de 2013 a 2015. Disponível em: <http://portal.anvisa.gov.br/>. 2016.
20. SINITOX. Sistema Nacional de Informações Tóxico-Farmacológicas. Banco de Dados de óbitos. Disponível em: <https://sinitox.iciet.fiocruz.br/dados-nacionais>.
21. SINAN. Sistema de Informação de Agravos de Notificação Portal de Dados Abertos Sobre Agrotóxicos. 2017. Disponível em: <http://dados.contraosagrototoxicos.org>.
22. SINITOX. Sistema Nacional de informações tóxicos-farmacológicas. 2009. Disponível em: <https://sinitoxdados.iciet.fiocruz>.

23. Bochner R. Sistema Nacional de Informações Tóxico-Farmacológicas - SINITOX e as intoxicações humanas por agrotóxicos no Brasil. *Ciência & Saúde Coletiva*. 2007;12(1): 73-89.
24. Bochner R. Óbito ocupacional por exposição a agrotóxicos utilizado como evento sentinela: quando pouco significa muito. *Vigilância sanitária em debate*. 2015;3(4): 39-49.
25. Schmidt MLG, Godinho PH. Um breve estudo acerca do cotidiano do trabalho de produtores rurais: intoxicações por agrotóxicos e subnotificação. *Revista Brasileira de Saúde Ocupacional*. 2006;31(113):27-40.
26. Human Right Swatch. "Você não quer mais respirar veneno". 2018. Disponível em: <https://www.hrw.org/sites>.
27. INCA. Instituto Nacional do Câncer. Posicionamento público a respeito do uso de agrotóxicos. 2014. Disponível em: <http://www2.inca.gov.br>.
28. Londres, Flavia. *Agrotóxicos no Brasil: um guia para ação em defesa da vida*. – Rio de Janeiro: AS-PTA – Assessoria e Serviços a Projetos em Agricultura Alternativa, 2011.
29. Peres F, Moreira JC. Saúde e ambiente em sua relação com o consumo de agrotóxicos em um pólo agrícola do Estado do Rio de Janeiro, Brasil. *Cad. Saúde Pública*. 2007;4:612-621.
30. Copplestone JF. The development of the WHO Recommended Classification of Pesticides by Hazard. *Bulletin of the World Health Organization*. 1988;66(5): 545-551.
31. World Health Organization. The WHO recommended classification of pesticides by hazard and guidelines to classification 1990-1991. *International Programme on Chemical Safety* Disponível em: <http://www.who.int>.
32. Peres F, Moreira JC, Dunois GS. *É veneno ou é remédio? Agrotóxicos, saúde e ambiente*. Editora Fiocruz. 2003.
33. ANVISA. Agência Nacional de Vigilância Sanitária. *Agrotóxico, herbicida e pesticida*. Disponível em: <http://portal.anvisa.gov.br>.
34. Fridovich I. Superoxide dismutases. *Advances in Enzymology and Related Areas of Molecular Biology*. 1986;58: 61-97.
35. Cordeiro L, Dixon RA. The oxidative burst in plant disease resistance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 1997;48: 251–275.
36. Bolwell GP. Role of active oxygen species and NO in plant defence responses. *Current Opinion in Plant Biology*. 1999;2: 287–294.

37. Dat J, Vandenabeele S, Vranová E, Van Montagu M, Inzé D, Van Breusegem F. Dual action of the active oxygen species during plant stress responses. *Cell Mol Life Sci.* 2000;57: 779–795.
38. Hegedus A, Erdei S, Horváth G. Comparative studies of H₂O₂ detoxifying enzymes in green and greening barley seedlings under cadmium stress. *Plant Science.* 2001;160:1085-1093.
39. Yao SH, Guoa Y, Wanga YZ, Zhanga D, Xub L, Tang WH. A cytoplasmic Cu-Zn superoxide dismutase SOD1 contributes to hyphal growth and virulence of *Fusarium graminearum*. *Fungal Genetics and Biology.* 2016;91:32-42.
40. Noriega NC, Ostby J, Lambright C, Wilson VS, Gray JRLE. Late gestational exposure to the fungicide prochloraz delays the onset of parturition and causes reproductive malformations in male but not female rat offspring. *Biol Reprod.* 2005;72:1324–1335.
41. Kjærstad MB, Taxvig C, Nellesmann C, Vinggaard AM, Andersen HR. Endocrine disrupting effects in vitro of conazole antifungals used as pesticides and pharmaceuticals. *Reprod Toxicol.* 2010;30(4): 573-582.
42. Higley EB, Newsted JL, Zhang X, Giesy J P, Hecker M. Assessment of chemical effects on aromatase activity using the H295R cell line. *Environ Sci Pollut Res Int.* 2010;17(5):1137-1148.
43. Ohlsson A, Ullerås E, Oskarsson, A. A biphasic effect of the fungicide prochloraz on aldosterone, but not cortisol, secretion in human adrenal H295R cells--underlying mechanisms. *Toxicology letters, Amsterdam.* 2009;191(2-3):174-80.
44. Ohlsson A, Cedergreen N, Oskarsson A, Ullerås, E. Mixture effects of imidazole fungicides on cortisol and aldosterone secretion in human adrenocortical H295R cells. *Toxicology.* 2010;275(1-3): 21-8.
45. Laier P, Metzdorff SB, Borch J, Hagen ML, Hass U, Christiansen S, et al. Mechanisms of action underlying the antiandrogenic effects of the fungicide prochloraz. *Toxicol Appl Pharmacol.* 2006;213(2):160-71.
46. Vinggaard AM, Christiansen S, Laie P, Poulsen ME, Breinholt V, Jarfelt K, et al . Perinatal exposure to the fungicide prochloraz feminizes the male rat offspring. *Toxicological Sciences.* 2005;85(2):886-897.
47. Rakitsky VN, Koblyakov VA, Turuso VS. Nongenotoxic (epigenetic) carcinogens: Pesticides as an example. A critical review. *Teratogenesis, Carcinogenesis, and Mutagenesis.* 2000;20(4):229-240.
48. Kirsch-volders M, Vanhauwaert A, Eichenlaub-Ritter U, Decordier I. Indirect mechanisms of genotoxicity. *Toxicol Let.* 2003;11(140-141): 63-74.
49. MCcarroll NE, Protzel A, Ioannou Y, Frank Stack HF, Jackson M A, Waters MD, et al. A survey of EPA/OPP and open literature on selected pesticide chemicals.

- III. Mutagenicity and carcinogenicity of benomyl and carbendazim. *Mutat Res.* 2002;512(1): 1-35.
50. Hess RA, Nakai M. Histopathology of the male reproductive system induced by the fungicide benomyl. *Histol Histopathol.* 2000;15(1): 207- 224.
 51. Nakai M, Miller MG, Carnes K, Hess RA. Stage-specific effects of the fungicide carbendazim on Sertoli cell microtubules in rat testis. *Tissue Cell.* 2002;34(2): 73-80.
 52. Gray LE, Ostby J, Ferrell J, Sigmon R, Cooper R, Linder R, et al. Correlation of sperm and endocrine measures with reproductive success in rodents. *Prog Clin Biol Res.* 1989;302:193-206.
 53. Gracy RW, Talent JM, Kong Y, Conrad CC. Reactive oxygen species: the unavoidable environmental insult? *Mutat Res.* 1989;16;(1-2):17-22.
 54. Gray LE, Ostby J, Sigmon R, Ferrell J, Rehnberg G, Linder, et al. The development of a protocol to assess reproductive effects of toxicants in the rat. *Reprod Toxicol.* 1988;2(3-4):281-7.
 55. MAPA. Ministério da Agricultura Pecuária e Abastecimento. Mapa suspende registro de 63 fungicidas para controle da ferrugem da soja. 2017. Disponível em: <http://www.agricultura.gov.br>.
 56. Monod G, Rime H, Bobe J, Jalabert B. Agonistic effect of imidazole and triazole fungicides on in vitro oocyte maturation in rainbow trout (*Oncorhynchus mykiss*). *Environ Res.* 2004;58(2-5): 143-146.
 57. Taxvig C, Hass U, Axelstad M, Dalgaard M, Boberg J, Andeasen H, et al. Endocrine-disrupting activities in vivo of the fungicides tebuconazole and epoxiconazole. *Toxicological Sciences.* 2007;100(2): 464-73, 2007.
 58. Carneiro FF, Pignati W, Rigotto RM, Augusto LGS, Rizollo A, Muller NM, et al. Dossiê ABRASCO—Um alerta sobre os impactos dos agrotóxicos na saúde. Rio de Janeiro: ABRASCO; 2012. Disponível em: www.abrasco.org.br.
 59. Costanza R, D'arge R, De Groot R, Farber S, Grasso M, Hannon B, et al. The value of the world's ecosystem services and natural capital. *Nature.* 1997;387(6630): 253–260.
 60. MMA. Ministério do Meio Ambiente. Polinizadores em risco de extinção são ameaça à vida do ser humano. 2014. Disponível em: <http://www.mma.gov.br>.
 61. Nakasu EYT, Williamson SM, Edwards MG, Fitches EC, Gatehouse JA, Wright GA. Novel biopesticide based on a spider venom peptide shows no adverse effects on honeybees. *Proc. Biol. Sci.* 2014;281(1787); 20140619

62. Cham KO, Rebelo RM, Oliveira RP, Ferro AA, Vianasilva FEC, Borges LO, et al. Manual de avaliação de risco ambiental de agrotóxicos para abelhas. Brasília: Ibama/Diqua, 2017: 105.
63. Santos AB, Nascimento F S. Diversidade de visitantes florais e potenciais polinizadores de *Solanum lycopersicum* (Linnaeus) (Solanales: Solanaceae) em cultivos orgânicos e convencionais. *Neotropical Biology and Conservation*. 2016(3):162-169.
64. Deprà, S. M. abelhas polinizadoras de *Solanum lycopersicum* L. (solanaceae) em áreas de plantio inseridas em diferentes paisagens no município de são José de ubá, RJ. Dissertação de Mestrado, UENF. 2012.
65. Giannini TC, Cordeiro GD, Freitas BM, Saraiva AM, Imperatriz-Fonseca VL. The dependence of crops for pollinators and the economic value of pollination in Brazil. *Journal of Economic Entomology*. 2015: 1-9.
66. EFSA (European Food Safety Authority). Panel on Plant Protection Products and their Residues (PPR); Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). *EFSA Journal*. 2012;10(5): 275.
67. Gill KA, O'Neal ME. Survey of Soybean Insect Pollinators: Community Identification and Sampling Method Analysis. *Environ Entomol*. 2015;44(3):488-98.
68. Chiari WC, Toledo VAA, Ruvolo-Takasusuki MCC, Oliveira AJB, Sakaguti ES, Attencia V M, et al. Pollination of soybean (*Glycine max* L. Merrill) by honeybees (*Apis mellifera* L.). *Brazilian Archives of Biology and Technology*. 2005;48(1): 31-36.
69. EFSA. Guidance on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). *EFSA Journal* 2013;11(7):3295.
70. Faria Junior LRB, Bendini JN, Barreto LMRC. Eficiência polinizadora de *Apis mellifera* L. e polinização entomófila em pimentão 'Cascadura Ikeda'. *Bragantia* [online]. 2008;67(2): 261-266.
71. Raw A. Foraging Behaviour of Wild Bees at Hot Pepper Flowers (*Capsicum annuum*) and its possible Influence on Cross Pollination. *Annals of Botany* 2000;85:487-492.
72. Roselino A C. Polinização em culturas de pimentão - *Capsicum annuum* por *Melipona quadrifasciata anthidioides* e *Melipona scutellaris* e de morango - *Fragaria x ananassa* por *Scaptotrigona aff. depilis* e *Nannotrigona testaceicornis* (Hymenoptera, Apidae, Meliponini). Dissertação de Mestrado. USP. 2005.
73. Jordão AL, Noronha ACS. Visitantes florais e potenciais polinizadores da mandioca (*Manihot esculenta* Crantz) em localidades do estado do Amapá. In Congresso Brasileiro de Mandioca. 2011.

74. Silva RM, Bandel G, Faraldo MIF, Martins PS. Biologia reprodutiva de etnovariedades de mandioca. *Scientia Agricola*. 2001;58(1):101-107.
75. Malerbo-Souza DT. The corn pollen as a food source for honeybees. *Acta Scientiarum. Agronomy, Maringá*. 2011;33(4): 701-704.
76. ANVISA. Agência Nacional de Vigilância Sanitária. Contaminação dos corpos d'água. 2011. Disponível: <http://portal.anvisa.gov.br/agrotoxicos>.
77. BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Vigilância e controle da qualidade da água para consumo humano/ Ministério da Saúde, Secretaria de Vigilância em Saúde (Série B. Textos Básicos de Saúde). Brasília. Ministério da Saúde, 2006: 212.
78. IBGE. Instituto Brasileiro de Geografia e Estatística Atlas de Saneamento, 2011. Disponível em: <http://www.ibge.gov.br>.
79. MAPA .Ministério da Agricultura Pecuária e Abastecimento. Produção Agrícola. 2014. Disponível em: <http://www.agricultura.gov.br>.
80. FAO FAOSTAT. 2018. Acesso em: Available at <http://www.fao.org/>.
81. CONAB. Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de grãos Safra 2017/18. 2018;8: 145.
82. EMBRAPA. Embrapa Soja. Ministério da Agricultura, Pecuária e Abastecimento. Agência Embrapa de Informação tecnológica. 2016 Disponível em: <http://www.agencia.cnptia.embrapa.br>.
83. Guerreiro, L. Dossiê técnico: Produtos de Soja. Serviço Brasileiro de Respostas Técnicas. 2006. Disponível em: <http://www.respostatecnica.org.br/>
84. SINDAG. Sindicato Nacional das Indústrias de Defensivos Agrícolas; Anais do Workshop: Mercado Brasileiro de Fitossanitários; Avaliação da Exposição de Misturadores, Abastecedores e Aplicadores de Agrotóxicos. Brasília: 2009.
85. SINDAG. Sindicato Nacional das Indústrias de Defensivos Agrícolas. Dados de produção e consumo de agrotóxicos. 2011. Disponível em [www. sindag.com.br](http://www.sindag.com.br)
86. ANDA. Associação Nacional para Difusão de Adubos. Estatísticas 2011. Disponível em <http://www.anda.org.br>.
87. BGE/SIDRA. Brasil, série histórica de área plantada; série histórica de produção agrícola; safras 1998 a 2011. 2012. Disponível em: www.sidra.ibge.gov.br/bda/agric.
88. MAPA. Ministério da Agricultura Pecuária e Abastecimento Projeções do agronegócio de 2009/10 a 2019/2020. Brasília: MAPA/AGE/ACS, 2010. Disponível em: <http://www.agricultura.gov.br>.

89. IBGE/SIDRA. Instituto Brasileiro de Geografia e Estatística. Levantamento Sistemático da Produção Agrícola- 2015. Disponível em: <https://sidra.ibge.gov.br/home/lspa/brasil>.
90. EMBRAPA. Embrapa Milho e Sorgo. Ministério da Agricultura, Pecuária e Abastecimento. O produtor pergunta, a Embrapa responde. Coleção 500 perguntas, 500 respostas. Brasília. 2011. Disponível em: <http://www.agencia.cnptia.embrapa.br>.
91. IBGE/SIDRA. Instituto Brasileiro de Geografia e Estatística. Levantamento Sistemático da Produção Agrícola - Milho - agosto 2018. 2018. Disponível em: <https://sidra.ibge.gov.br/home/lspa/brasil>.
92. Pereira Filho IA, Borghi, E. Mercado de sementes de milho no Brasil: safra 2016/2017. Sete Lagoas: Embrapa Milho e Sorgo, 2016. Disponível em: <https://www.embrapa.br/milho-e-sorgo>.
93. EMBRAPA. Ministério da Agricultura, Pecuária e Abastecimento. Controle Químico de doenças. 2015. Disponível em: <http://www.agencia.cnptia.embrapa.br>.
94. Colariccio A, Eiras M, Chaves ALR, Lourenção AL, Melo AMT, Siqueira WJ. Detecção do 'Chysoanthemum stem necrosis virus' em tomateiro no Estado de São Paulo. Summa Phytopathologica. 2000;25: 252-254.
95. IBGE/SIDRA. Instituto Brasileiro de Geografia e Estatística. Levantamento Sistemático da Produção Agrícola - Tomate- agosto 2018. Disponível em: <https://sidra.ibge.gov.br/home/lspa/brasil>.
96. Borguini RG. Tomate (*Lycopersicum esculentum* Mill) Orgânico: O Conteúdo Nutricional e a Opinião do Consumidor. Dissertação de Mestrado, 2002.
97. Gidenne T, Carabano R, Garcia J, Blas C. Fibre Digestion in the nutrition of the rabbit. Anim. Feed Sci. Tech. 1998: 69-88.
98. Reboul E, Borel P, Mikail C, Abou L, Charbonnier M, Caris-Veyrat C et al. Enrichment of tomato paste with 6% tomato peel increases lycopene and -carotene bioavailability in men. J. Nutr. 2005;135:790–794.
99. Giovannucci, E.; Ascherio, A.; Rimm, E.B.; Stampfer, M.J.; Colditz, G.A. and Willett, W.C. (1995). Intake of carotenoids and retinol in relation to risk of prostate cancer. J. Natl. Cancer Inst., 87, p.1767-1776.
100. EMBRAPA. Ministério da Agricultura, Pecuária e Abastecimento. Cultivo de Tomate para Industrialização. Sistemas de Produção, 1 - 2ª Edição Versão Eletrônica. 2006. Disponível em: <https://sistemasdeproducao.cnptia.embrapa.br>.
101. Yu C, Wang N, Wu M, Tian F, Chen H, Yang F. OxyR-regulated catalase CatB promotes the virulence in rice via detoxifying hydrogen peroxide in *Xanthomonas oryzae* pv. *oryzae*. BMC Microbiology. 2016;16: 269.

102. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature*. 2003;422: 835–847.
103. Wu D, Wu M, Halpern A, Rusch DB, Yooseph S, Frazier M et al. Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. *PLoS One*. 2011;6:e18011.
104. Bayat A. Science, medicine, and the future: Bioinformatics. *BMJ*. 2002;22: 1018-1022.
105. Janssen P, Enright AJ, Audit B, Cases I, Goldovsky L, Harte N, et al. Complete Genome Tracking (COGENT): a flexible data environment for computational genomics. *Bioinformatics*. 2003;19:1451-1452.
106. Prodocimi, F. Curso Online: Introdução à Bioinformática. 2007. Disponível em: http://www2.bioqmed.ufrj.br/prodocimi/FProdocimi07_CursoBioinfo.pdf.
107. Dolovich MB, Dhand R. Aerosol drug delivery: developments in device design and clinical use. *Lancet*. 2011;377: 1032–1045.
108. Swinney DC, Anthony J. How were new medicines discovered? *Nature Reviews Drug Discovery*. 2011;10:507-519.
109. Gilbert IH. Target-based drug discovery for human African trypanosomiasis: selection of molecular target and chemical matter. *Parasitology*, v.141(1), p. 28–36, 2014.
110. Claros MG, Bautista R, Guerrero-Fernández D, Benzerki H, Seoane P, Fernández-Pozo N. Why assembly plant genome sequence is so challenging. *Biology*. 2012;1(2): 439-459.
111. Raven PH, Evert RF, Eichhorn SE. *Biologia vegetal*. Ed. Guanabara Koogan, 2007
112. Parreira DF, Neves WS, Zambolim L. Resistência de Fungos a Fungicidas Inibidores de Quinona. *Revista Trópica: Ciências Agrárias e Biológicas*. 2009;3(2).
113. Michelli F, Guiltinan MJ, Gramacho KP, Wilkinson MJ, Vargas de Oliveira Figueira A, De Mattos Cascardo J. Functional genomics of cacao. *Advances in Botanical Research*. 2010;55: 119-177.
114. Sánchez-Torres P, González-Candelas, L. Isolation and characterization of genes differentially expressed during the interaction between apple fruit and *Penicillium expansum*. *Plant Molecular Pathology*. 2003;4:447-457.
115. FDA, (CDER), Investigating Out-of-Specification (OOS) Test Results for Pharmaceutical. 2006

116. FDA/ICH, (CDER and CBER), Q8(R2) Pharmaceutical Development, guidance for industry. 2009.
117. Otto TD, Guimarães AC, Degraeve WM, Miranda AB. AnEnPi: identification and annotation of analogous enzymes. BMC Bioinformatics. 2008;9:544.
118. Gomes MR, Guimarães ACR, Miranda AB. Specific and Nonhomologous Isofunctional Enzymes of the Genetic Information Processing Pathways as Potential Therapeutic Targets for Trypsins. Enzyme Research. 2011; 2011:8.
119. Lehninger AL, Nelson DL, Cox MM. Principles of Biochemistry. W. H. Freeman, New York, 2001
120. Ouzounis A, Coulson RMR, Enright AJ, *et al.* Classification schemes for protein structure and function. Nature Reviews Genetics. 2003;4(7): 508–519.
121. Papin JA, Price ND, Wiback S J, *et al.* Metabolic pathways in the post-genome era. Trends in Biochemical Sciences. 2003;28:250-58.
122. Galperin MY, Walker DR, Koonin EV. Analogous enzymes: independent inventions in enzyme evolution. Genome Research. 1998;8(8):779-90.
123. Omelchenko MV, Galperin MY, Wolf YI, Koonin EV. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. Biology Direct. 2010;5:31.
124. Chollet R, Vidal J, O'Leary M. Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. Annu Rev Plant Physiol Plant Mol Biol. 1996; 47: 273–298
125. Nunez M, Mazzafera P, Mazorra LM, Siqueira WJ, Zullo MAT. Influence of brassinosteroid analogue on antioxidant enzymes in rice grown in culture medium with NaCl. Biol. Plant. 2003;47:67-70.
126. Capriles PV, Guimarães AC, Otto TD, Miranda AB, Dardenne LE, Degraeve WM. Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas' disease treatment. BMC Genomics. 2010; 11: 610.
127. Aardema MJ, Macgregor JT. Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. Mutation Research. 2002; 29: 13-25.
128. Perna NT, Plunkett G, Burland V, Mau B. *et al.* Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature. 2001;409: 529-533.
129. Ebersberger I, Metzler D, Schwarz CE, Paabo S. Genome wide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet. 2002;70: 1490-1497.

130. Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*. 2000;406:151-157.
131. Bhardwaj A, Dhar YV, HasanAsif M, Bag SK. In Silico identification of SNP diversity in cultivated and wild tomato species: insight from molecular simulations. *Scientific Reports*. 2016;6:38715.
132. Mehrotra S, Goyal V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics Proteomics Bioinformatics*. 2014;12:164–171.
133. Silva RA, Souza G, Lemos LSL, Lopes UV, Patrocínio NGRB, Alves RM, et al. Genome size, cytogenetic data and transferability of EST-SSRs markers in wild and cultivated species of the genus *Theobroma* L. (Byttnerioideae, Malvaceae). *PLoS ONE*. 2017;12(2): e0170799.
134. Vishwakarma H, Junaid A, Manjhi J, Singh GP, Gaikwad K, Padaria JC. Heat stress transcripts, differential expression, and profiling of heat stress tolerant gene TaHsp90 in Indian wheat (*Triticum aestivum* L.) cv C306. *PLoS ONE*. 2018;13(6): e0198293.
135. Alkharouf N, Khan R and Matthews B. Analysis of expressed sequence tags from roots of resistant soybean infected by the soybean cyst nematode. *Genome*. 2004;47:380-388.
136. Fuentes A, Ortiz J, Saavedra N, Salazar LA, Meneses C, Arriagada C. Reference gene selection for quantitative real-time PCR in *Solanum lycopersicum* L. inoculated with the mycorrhizal fungus *Rhizophagus irregularis*. *Plant Physiol Biochem*. 2016;101:124-131.
137. Bueno LCS, Mendes ANG, Carvalho SP. Melhoramento Genético de Plantas: Princípios e Procedimentos. Lavras. Disponível em: <http://www.dag.ufla.br/>. 2006.
138. Dekkers JCM, F. Hospital. The use of molecular genetics in improvement of agricultural populations. *Nat. Rev. Genet*. 2002;3: 22–32.
139. Dreher K, Khairallah M, Ribaut JM, Morri M. Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. *Mol. Breed*. 2003;11: 221-234.
140. Ferreira ME, Grattapaglia D. Introdução ao uso de marcadores moleculares em análise genética. Brasília: Embrapa-Cenargen, 1998;2: 220.
141. Brammer SP, Iorczeski EJ. Atualização em técnicas celulares e moleculares aplicadas ao melhoramento genético vegetal. Passo Fundo: Embrapa Trigo. 2002: 277-298.

142. Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, et al. Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in situ hybridization. *Genetics*. 2006;173: 349-362.
143. Bao JS, Corke H, Sun, M. Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*. 2006;113: 1185-1196.
144. Useche EJ, Gao J, Harafey M, Rafalski A. High-throughput identification, data base storage and analysis of SNPs in EST sequences. *Genome Informatics*. 2001;12:194.
145. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes *Nucleic Acids Res*. 2000; 28(1):27-30.
146. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012;40:D109-14.
147. Kanehisa M, Goto S, Sato Y, Kawashim M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*. 2014;42:199-205.
148. NC-IUBMB. International Union of Biochemistry And Molecular Biology Enzyme nomenclature. 2011. Disponível em: <http://www.sbc.sqmul.ac.uk/iubmb/>.
149. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol Biol*. 1995;247(4):536-40.
150. Madera, M., Vogel, C., Kummerfeld, S.K., Chothia, C., and Gough, J. The SUPERFAMILY database in 2004: Additions and improvements. *Nucleic Acids Res*. 2004;32: 235–D239.
151. Pethica RB, Levitt M, Gough J. Evolutionarily consistent families in scop: sequence, structure and function. *BMC Struct. Biol*. 2012;12: 27.
152. Schomburg I, Chang A, Hofmann O, Ehrentreich F, Schomburg D. BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem. Sci*. 2002;27:54-56.
153. Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res*. 2002;30:47-49.
154. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;1:28(1):235-42.

155. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, et al. The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.* 2013;41:475-482.
156. Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, et al. PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* 2014;42:285-291.
157. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, O'Donovan C, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Research.* 2004;32:115-119.
158. Farias AX. Contaminação Endogenica por *Aspergillus* spp. em milho Pós-colheita no estado do Paraná, *Pesq. agropec. bras.*, Brasília. 2000;35(3):617-621.
159. Killebrew JF, Roy K W, Abney TS. Fusaria and other fungi on soybean seedlings and roots of older plants and interrelationships among fungi, symptoms and soil characteristics. *Can. J. Plant Pathol.* 1993;15:139-146
160. Prathuangwong S, Amnuaykit K. Studies on tolerance and rate reducing bacterial pustule of soybean cultivars/lines. *Kasetsart Journal (Natural Science).* 1987;21:408-420.
161. Purdy LH. *Sclerotinia sclerotiorum*: history, diseases and symptomatology, host range, geographic distribution, and impact. *Phytopathology.* 1979;69: 875-880.
162. Dorrance AE, Mills D, Robertson AE, Draper MA, Giesler L, Tenuta A. *Phytophthora* root and stem rot of soybean. *The Plant Health Instructor.* 2007.
163. Williamson B, Duncan GH, Harrison JG, Harding LA, Elad Y, Zimand G, Effect of humidity on infection of rose petals by dry-inoculated conidia of *Botrytis cinerea*. *Mycological Research.* 1995;99:1303–1164.
165. Agrios GN. *Plant Pathology.* New York. Academic Press. 1997.
166. Aime MC, Phillips-Mora W. The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. *Mycol.* 2005;97(5):1012 - 1022.
167. Sivasithamparam K. 1998. Root cortex – the final frontier for the biocontrol of root-rot with fungal antagonists: a case study on a sterile red fungus. *Annual Review of Phytopathology.* 1998;36:439-452.
168. Denny TP, Brumbley SM, Carney BF, Clough SJ. Phenotype conversion of *Pseudomonas solanacearum*: its molecular basis and potential function. In: Hayward AC, Hartman GL, editors. *Bacterial wilt: the disease and its causative agent, Pseudomonas solanacearum.* Wallingford, UK: CAB International.1994;137.
169. Bennett J W, Klich MA. Mycotoxins. *Clin. Microbiol. Rev.* 2003;16: 497–516.

170. Hesseltine CW, Vandegrift EE, Fennell DI, Smith ML, Shotwell OL. Aspergilli as Ochratoxin Producers. *Mycologia*. 1997;64(3):539-50
171. Cota LV, Costa RV, Guimarães PE, Parentoni SN, Pacheco CAP, Silva DD, et al. Métodos de inoculação de *Colletotrichum graminicola* em colmo de milho. *Embrapa Milho e Sorgo*, 2010;137:7.
172. Munkvold GP. Epidemiology of *Fusarium* diseases and their mycotoxins in maize ears. *European Journal of Plant Pathology*. 2003;109:705-713.
173. Perkins JM, Pedersen W L. Disease development and yield losses associated with northern leaf blight on corn. *Plant Disease*. 1987;71(10): 940-943.
174. Fernandes FT, Oliveira E. Principais doenças na cultura do milho. *Embrapa-CNPMS*, 1997;27:80.
175. Paccola-Meirelles LD, Ferreira AS, Meirelles WF, Marriel IE, Casela CR. Detection of a bacterium associated with a leaf spot disease of maize in Brazil. *Journal of Phytopathology*. 2001;149:275-79
176. OECD. Perspectivas Agrícolas da OAA da OCDE: 2006-2015. 2015 Disponível em: <http://www.oecd.org/fr>.
177. FAO. Perspectivas Agrícolas 2015-2024. 2015. Disponível em: <http://www.fao.org>.
178. Schmutz J, Cannon S, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010;463:178–183.
179. Nierman WC, Yu J, Fedorova-Abrams ND, Losada L, Cleveland TE, Bhatnagar D, et al. Genome sequence of *Aspergillus flavus* NRRL 3357, a strain that causes aflatoxin contamination of food and feed. *Genome Announc*. 2015;3(2):e00168-15.
180. Thatcher LF, Gardiner DM, Kazan K, Manners JM. A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Mol Plant Microbe Interact*. 2012;25(2):180-90.
181. Tyler BM, Tripathy S, Zhang XM, Jiang RH, Aerts A, Arredondo FD. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science*. 2006;313:1261–1266
182. Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, et al. Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet*. 2011;7(8):e1002230.
183. Kim JG, Choi S, Oh J, Moon JS, Hwang I. Comparative analysis of three indigenous plasmids from *Xanthomonas axonopodis* pv. *glycines*. *Plasmid*. 2006;56(2):79-87.

184. Kahlau S, Aspinall S, Gray JC, Bock R. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J Mol Evol.* 2006;63(2):194-207.
185. Staats M, van Kan JA. Genome update of *Botrytis cinerea* strains B05.10 and T4. *Eukaryot Cell.* 2012;11(11):1413-4.
186. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 2010;18;464(7287):367-373.
187. Mondego JM, Carazzolle MF, Costa GG, Formighieri EF, Parizzi LP, Rincones J. A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao. *BMC Genomics.* 2008;18;9:548.
188. Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proc Natl Acad Sci U S A.* 2005;2;102(31):11064-9.
189. Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, et al. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature.* 2002;31;415(6871):497-502.
190. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S. The B73 maize genome: complexity, diversity, and dynamics. *Science.* 2009;20;326(5956):1112-5.
191. Faustinelli PC, Wang XM, Palencia ER, Arias RS. Genome sequences of eight *Aspergillus flavus* spp. and one *A. parasiticus* sp., isolated from peanut seeds in Georgia. *Genome Announc.* 2016;4(2):e00278-16.
192. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genetics.* 2012;44:1060–1065.
193. Cuomo C, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science.* 2007;317:1400–1402
194. Ohm R, Feu N, Henrissat B, Schoch CL, Horwitz BA, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen dothideomycetes fungi. *PLoS Pathog.* 2012;8: e1003037.
195. De Maayer P1, Chan WY, Venter SN, Toth IK, Birch PR, Joubert F, et al. Genome sequence of *Pantoea ananatis* LMG20103, the causative agent of Eucalyptus blight and dieback. *J Bacteriol.* 2010;192(11):2936-7.
196. Honey Bee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature.* 2006;443(7114):931-949.

197. Baroncelli R, Piaggieschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, et al. Draft whole-genome sequence of the biocontrol agent *Trichoderma harzianum* T6776. *Genome Announc.* 2015;3(3):e00647-15.
198. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology.* 2009;155(6):1758-75.
199. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Initial sequencing and analysis of the human genome. *Nature.* 2001;15:409(6822):860-921.
200. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al.. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res.* 25:3389-3402. 1997.
201. Guimaraes ACR, Otto TD, Alves-Ferreira M, Miranda AB, Degrave, W.M. . In silico reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*. *Genetics and Molecular Research.* 2008;7:872-882.
202. Alves-Ferreira M, Guimarães AC, Capriles PV, Dardenne LE, Degrave WM. A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information. *Memórias do Instituto Oswaldo Cruz.* 2009;104(8):1100-10.
202. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology.* 2001;313(4):903-19.
203. Guengerich FP. Cytochrome P450s and other enzymes in drug metabolism and toxicity. *The AAPS Journal.* 2008;8(1):101-11.
204. Aktar W, Sengupta D, Chowdhury A. Impact of pesticide use in agriculture: their benefits and hazards. *Interdisciplinary Toxicology.* 2009; 2(1): 1-12.
205. Fishel FM. Pesticides effects on nontarget organisms. PI-85. Pesticide information office, Florida Cooperative Extension Service, IFAS, University of Florida. 2011. Disponível em: <http://edis.ifas.ufl.edu/>.
206. Pan-Germany. Pesticide and health hazards. Facts and figures. 2012. Disponível em: www.pan-germany.org.
207. IBAMA Instituto do Meio Ambiente. Avaliação do risco de agrotóxicos para insetos polinizadores e lacunas de conhecimento. 2017. Disponível em: <https://www.ibama.gov.br>.
208. Torres MA, Jones JDJ, Dangl JL. Reactive oxygen species signaling in response to pathogens. *Plant Physiol.* 2006;141:37378.

209. Eaton CJ, Cox MP Scott B. What triggers grass endophytes to switch from mutualism to pathogenesis? *Plant Sci.* 2011;180:190–5.
210. Broxton CN, Culotta VC. SOD Enzymes and Microbial Pathogens: Surviving the Oxidative Storm of Infection. *PLoS Pathog.* 2016;12(1): e1005295.
211. Cox GM, Harrison TS, McDade HC, Taborda CP, Heinrich G, Casadevall A, et al. Superoxide Dismutase Influences the Virulence of *Cryptococcus neoformans* by Affecting Growth within Macrophages. *Infect. Immun.* 2003;71(1):173-180.
212. Karadag H, Ozhan F. Effect of cyprodinil and fludioxonil pesticides on bovine liver catalase activity. *Biotechnol Biotechnol Equip.* 2015; 29(1): 40-4.
213. Pratt AJ, DiDonato M, Shin DS, Cabelli DE, Bruns CK, Belzer CA, et al. Structural, Functional, and Immunogenic Insights on Cu,Zn Superoxide Dismutase Pathogenic Virulence Factors from *Neisseria meningitidis* and *Brucella abortus*. *J Bacteriol.* 2015;197(24): 3834–3847.
214. Heym B, Alzari PM, Honore N, Cole ST. Missense mutations in the catalase–peroxidase gene, *katG*, are associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Mol Microbiol.* 1995;15: 235–245.
215. Pym AS, Domenech P, Honoré N, Song J, Deretic V, Cole ST. Regulation of catalase-peroxidase (KatG) expression, isoniazid sensitivity and virulence by *furA* of *Mycobacterium tuberculosis*. *Mol Microbiol.* 2000;40(4): 879–889
216. Ishiga Y, Ichinose Y. *Pseudomonas syringae* pv. tomato OxyR Is Required for Virulence in Tomato and Arabidopsis. *Mol Plant Microbe Interact.* 2016;29(2):119-31.
217. Mir AA, Park SY, Sadat MA, Kim S, Choi J, Jeon J, et al. Systematic characterization of the peroxidase gene family provides new insights into fungal pathogenicity in *Magnaporthe oryzae*. *Sci Rep.* 2015;5:11831.
218. Barret MP. The Pentose Phosphate Pathway and Parasitic Protozoa. *Parasitology Today.*1997;13: 11–16.
219. Verlinde CLMJ, Hannaert V, Blonski C, Willson M, Périé JJ, Fothergill-Gilmore LA, et al. Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resist Update.* 2001;4:1-14.
220. Tomavo S. The differential expression of multiple isoenzyme forms during stage conversion of *Toxoplasma gondii*: an adaptive developmental strategy. *Internacional Journal Parasitol.* 2001;31: 1023-31.
221. Barrett MP, Gilbert IH. Perspectives for new drugs against trypanosomiasis and leishmaniasis. *Current Topics in Medicinal Chemistry.* 2002; 2: 471–482.
222. Tong L, Harwood HJJ. Acetyl-coenzyme A carboxylases: versatile targets for drug discovery. *J Cell Biochem.* 2006;99(6):1476–88.

223. Chawla B, Madhubala R. Drug targets in *Leishmania*. J Parasit Dis. 2010;34(1): 1–13.
224. Zhang J, Zhang Y, Inouye M. Thermotoga maritime MazG protein has both nucleoside triphosphate pyrophosphohydrolase and pyrophosphatase activity. J Biol Chem. 2003;278: 32300–32306.
225. Halliwell, B. Reactive species and antioxidants. Redox biology is a fundamental theme of aerobic life. Plant Physiol. 2006;141:312-22.
226. Roqueiro G, Maldonado S, Rios M, Del C, Maroder H. Fluctuation of oxidative stress indicators in *Salix nigra* seeds during priming. JOURNAL of Experimental Botany. 2012;63(10):3631-42.
227. Gil SS, Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiology and Biochemistry. 2010;48: 909-930.
228. Bartosz G. Generation of Reactive Oxygen Species in Biological Systems. Comments on Toxicology. 2003;1: 5-21.
229. Sharma P, Jha AB, Dubey RS, Pessarakli M. Reactive oxygen species, oxidative damage and anti-oxidative defense mechanism in plants under stressful conditions. J. Bot. 2012; 1–26.
230. Wojtaszek P. Oxidative burst: an early plant response to pathogen. Biochem J 1997;322:681–692.
231. Lamb C, Dixon RA. The oxidative burst in plant disease resistance. Annu Rev Plant Physiol Plant Mol Biol. 1997;48: 251–275.
232. Bradley D, Kjellbom P, Lamb C Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. Cell. 1992;70: 21–30.
233. Das K, Roychoudhury A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. Front. Environ. Sci. 2014;2(53).
234. Dowling DK, Simmons LW. Life-history evolution reactive oxygen species as universal constraints in life history. Proc. R. Soc. B. 2009;276: 1737–1745.
235. Tripathy BC, Oelmüller R. Reactive oxygen species generation and signaling in plants. Plant Signaling & Behavior. 2012;7(12):1621–1633.
236. Saed-Moucheshi, A., A. Shekoofa, & M. Pessarakli. Reactive Oxygen Species (ROS) Generation and Detoxifying in Plants. Journal of Plant Nutrition. 2014;37(10): 1573-1585

237. Ali S, Wang Z, Ren S, Huang Z. Superoxide dismutase production by *Isaria fumosorosea* on metals and its role in stress tolerance and fungal virulence. *Biocontrol Science and Technology*. 2011; 21(12): 1457-1469.
238. Masner P, Muster P, Schmid J. Possible methionine biosynthesis inhibition by pyrimidineamine fungicides. *Pestic Sci*. 1997;42:163–166.
239. Fritz R, Lanen C, Chapeland-Leclerc F, Leroux P. Effect of the anilinopyrimidine fungicide pyrimethanil on the cysthathionine beta-lyase of *Botrytis cinerea*. *Pestic Biochem Physiol*. 2003;77: 54-65.
240. Kanetis L, Forster H, Jones CA, Borkovich KA, Adaskaveg JE, Characterization of genetic and biochemical mechanisms of fludioxonil and pyrimethanil resistance in field isolates of *Penicillium digitatum*. *Phytopathology*. 2008;98:205-214.
241. Karadag H, Bilgin R. Effects of cyprodinil and fludioxonil pesticides on human superoxide dismutase. *Asian J Chem*. 2010;22:8147-8154.
242. Pan XY, Wu J, Xu S, Duan Y B, Zhou MG. CatB is critical for total catalase activity and reduces bactericidal effects of phenazine-1-carboxylic acid on *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*. *Phytopathology*. 2017;107:163-172.
243. Guo M, Block A, Bryan CD, Becker DF, Alfano JR. Pseudomonas syringae Catalases Are Collectively Required for Plant Pathogenesis. *J Bacteriol*. 2012;194(18):5054-64.
244. Mittler R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci*. 2002;7(9):405-10.
245. Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F. ROS signaling: the new wave? *Trends Plant Sci*. 2011;16(6):300-309.
246. Mittler R. ROS Are Good. *Trends Plant Science*. 2017;22(1):11-19.
247. Wamelink MMC, Struys EA, Jakobs C. The biochemistry, metabolism and inherited defects of the pentose phosphate pathway: a review. *Journal of Inherited Metabolic Disease*. 2008;31:703-717.
248. Miosga T, Schaaff-Gerstenschlager I, Franken E, Zimmermann FK. Lysine144 is essential for the catalytic activity of *Saccharomyces cerevisiae* transaldolase. *Yeast (Chichester, England)*. 1993;9:1241–1249.
249. Juhnke H, Krems B, Kotter P, Entian K-D.. Mutants that show increased sensitivity to hydrogen peroxide reveal an important role for the pentose phosphate pathway in protection of yeast against oxidative stress. *Molecular and General Genetics* 1996;252:456–64.

250. Li D, Zhu Y, Tang Q, Lu H, Li H, Yang Y, et al. A new G6PD knockdown tumor-cell line with reduced proliferation and increased susceptibility to oxidative stress. *Cancer Biother Radiopharm.* 2009;24: 81–90.
251. Essenberg MK, Cooper RA. Two ribose-5-phosphate isomerases from *Escherichia coli* K12: partial characterisation of the enzymes and consideration of their possible physiological roles. *Eur. J. Biochem.* 1975;55: 323-332.
252. Zhang RG, Andersson CE, Savchenko A, Skarina T, Evdokimova E, Beasley S. Structure of *Escherichia coli* Ribose-5-Phosphate Isomerase: A Ubiquitous Enzyme of the Pentose Phosphate Pathway and the Calvin Cycle. *Structure.* 2003;11(1):31-42.
253. Sørensen KI, Hove-Jensen B. Ribose catabolism of *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the *rpiR* gene, which is involved in regulation of *rpiB* expression. *J Bacteriol.* 1996;178(4):1003-11
254. Kaur PK, Tripathi N, Desale J, Neelagiri S, Yadav S, Bharatam PS, et al. Mutational and Structural Analysis of Conserved Residues in Ribose-5-Phosphate Isomerase B from *Leishmania donovani*: Role in Substrate Recognition and Conformational Stability. *PloS ONE.* 2016;11(3):e0150764.
255. Loureiro I, Faria J, Clayton C, Macedo-Ribeiro S, Santarem N, Roy N, et al. Ribose 5-phosphate isomerase B knockdown compromises *Trypanosoma brucei* blood stream form infectivity. *PLoS Negl Trop Dis.* 2015;9(1): e3430.
256. Soydan E, Güler A, Bıyık S, Şentürk M, Supuran CT, Ekinci D. Carbonic Anhydrase From *Apis Mellifera*: Purification and Inhibition by Pesticides. *J Enzyme Inhib Med Chem.* 2017;32(1):47-50.
257. Bucala R, Cerami A. Advanced glycosylation: chemistry, biology, and implications for diabetes and aging. *Advances in Pharmacology.* 1992;23: 1-34.
258. Ahmed MU, Thorpe SR, Baynes JW. Identification of N epsilon-carboxymethyllysine as a degradation product of fructoselysine in glycated protein. *Journal of Biological Chemistry.* 1986; 261: 4889-4894.
259. Chauhan SC, Padmanabhan PK, Madhubala R. Glyoxalase Pathway of Trypanosomatid Parasites: A Promising Chemotherapeutic Target. *Current Drug Targets.* 2008;9(11): 957–65.
260. Silva MS, Ferreira AEN, Gomes R, Tomás AM, Freire AP, Cordeiro C, et al. The glyoxalase pathway in protozoan parasites. *International Journal of Medical Microbiology.* 2012;302:225-29.
261. Chen ZY, Brown RL, Cleveland TE, Damann KE. Investigating the roles of an aflatoxin resistance-associated protein in maize using RNAi. *Phytopathology.* 2004; 94: 18.

262. Chen SC, Muller M, Zhou J Z, Wright LC, Sorrel TC. Phospholipase activity in *Cryptococcus neoformans*: a new virulence factor? *Journal Infectious Disease*. 1997;175: 414–420.
263. Fountain J C, Chen Z, Scully B T, Kemerait R. C, Lee R D, Guo B. Pathogenesis-related gene expressions in different maize genotypes under drought stressed conditions. *African Journal of Plant Science*. 2010;4(11): 433-440.
264. Fountain JC, Scully B, Ni X, Kemerait R, Lee D, Chen ZY, et al. Environmental influences on maize-*Aspergillus flavus* interactions and aflatoxin production. *Frontiers in Microbiology*. 2014;5: 40.
265. Kamoun S, Kado CL. A plant-inducible gene of *Xanthomonas campestris* pv. *campestris* encodes an exocellular component required for growth in the host and hypersensitivity on nonhosts. *J Bacteriol*. 1990;172(9):5165–5172.
266. Schüttelkopf AW, Gros L, Blair DE, Frearson JA, van Aalten DM, Gilbert IH. Acetazolamide-based fungal chitinase inhibitors. *Bioorg Med Chem*. 2010;18(23): 8334-40.
267. Calmes B, Morel-Rouhier M, Bataillé-Simoneau N, Gelhaye E, Guillemette T, Simoneau P. Characterization of glutathione transferases involved in the pathogenicity of *Alternaria brassicicola*. *BMC Microbiology*. 2015;15: 123.
268. Agledal L, Niere M, Ziegler M. The phosphate makes a difference: cellular functions of NADP. *Redox Reports*. 2010;15::2-10,

9. ARQUIVOS SUPLEMENTARES

Tabela Suplementar 1. Validação das sequências de acordo com seus enovelamentos usando o banco de dados SUPERFAMILY e SCOP.

Número EC	Nome da Enzima (Oficial)	Organismos	Clusters (Formas Estruturais)	Anotação original	Identificação da Sequência	Classificação de FOLD [SCOP ID]
1.11.1.6*	Catalase	<i>G. max</i>	1, 5	Catalase-3	351734412 ref NP_001235974.1	Heme-dependent catalase-like
		<i>A. flavus</i>	1, 2	Bifunctional catalase-peroxidase Cat2	238507433 ref XP_002384918.1	Heme-dependent peroxidases
1.11.1.7	Peroxidase	<i>G. max</i>	3, 6, 20	PREDICTED: cationic peroxidase 2-like, partial	571572675 ref XP_003557098.2	Heme-dependent peroxidases
		<i>A. flavus</i>	3, 6, 12	conserved hypothetical protein	238492121 ref XP_002377297.1	Dimeric alpha+beta barrel
1.2.1.3*	Aldehyde dehydrogenase (NAD+)	<i>G. max</i>	1	PREDICTED: delta-1-pyrroline-5-carboxylate dehydrogenase 12A1, mitochondrial-like isoform X4	571535254 ref XP_006600684.1	ALDH-like
		<i>A. flavus</i>	1, 6	Phosphotransferase enzyme family protein	238490122 ref XP_002376298.1	Protein kinase-like (PK-like)
2.6.1.1	Aspartate transaminase	<i>G. max</i>	1	PREDICTED: aspartate aminotransferase isoform X1	571532333 ref XP_006600243.1	PLP-dependent transferases
		<i>A. flavus</i>	1, 5	conserved hypothetical protein	238502281 ref XP_002382374.1	Dimeric alpha+beta barrel
1.11.1.6*	Catalase	<i>G. max</i>	1, 5	Catalase-3	351734412 ref NP_001235974.1	Heme-dependent catalase-like
		<i>F. oxysporum</i>	1, 2	catalase-peroxidase	F9FP11 F9FP11_FUSOF	Heme-dependent peroxidases
1.15.1.1	Superoxide dismutase	<i>G. max</i>	1, 4, 6, 7	PREDICTED: probable indole-3-pyruvate monooxygenase YUCCA3-like	571565271 ref XP_003555725.2	FAD/NAD(P)-binding domain/ FAD/NAD(P)-binding domain
		<i>F. oxysporum</i>	1, 4, 7, 14	Uncharacterized protein	F9FYF1 F9FYF1_FUSOF	Double-stranded beta-helix
2.6.1.1	Aspartate transaminase	<i>G. max</i>	1	PREDICTED: aspartate aminotransferase isoform X1	571532333 ref XP_006600243.1	PLP-dependent transferases
		<i>F. oxysporum</i>	1, 5	iron-dependent peroxidase	F9G466 F9G466_FUSOF	Dimeric alpha+beta barrel
4.4.1.5	Lactoylglutathione lyase	<i>G. max</i>	1, 8	PREDICTED: uncharacterized protein LOC102662470	571524316 ref XP_006598804.1	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase
		<i>F. oxysporum</i>	1, 3	Uncharacterized protein	F9G2J4 F9G2J4_FUSOF	Metallo-hydrolase/oxidoreductase
1.11.1.6*	Catalase	<i>G. max</i>	1, 5	Catalase-3	351734412 ref NP_001235974.1	Heme-dependent catalase-like
		<i>P. sojae</i>	1, 2	heme peroxidase	695395873 ref XP_009521283.1	Heme-dependent peroxidases/ Heme-dependent peroxidases
1.11.1.7	Peroxidase	<i>G. max</i>	3, 6, 20	PREDICTED: cationic peroxidase 2-like, partial	571572675 ref XP_003557098.2	Heme-dependent peroxidases
		<i>S. sclerotiorum</i>	3, 6, 12	hypothetical protein SS1G_13391	156033342 ref XP_001585507.1	Dimeric alpha+beta barrel
1.11.1.6	Catalase	<i>G. max</i>	1	catalase-3	351734412 ref NP_001235974.1	Heme-dependent catalase-like
		<i>X. axonopodis</i>	3	hypothetical protein	757568059 ref WP_042823856.1	Ferritin-like

1.15.1.1*	Superoxide dismutase	<i>G. max</i>	1, 4, 6, 7	PREDICTED: putative germin-like protein 9-2-like	571564643 ref XP_006605648.1	RmlC-like cupins
		<i>X. axonopodis</i>	1, 4	superoxide dismutase	933621128 ref WP_054320474.1	Cu,Zn superoxide dismutase-like
6.4.1.2*	Acetyl-CoA carboxylase	<i>G. max</i>	1	PREDICTED: alpha-carboxyltransferase aCT-1 precursor isoform X1	571541778 ref XP_006601861.1	ClpP/crotonase
		<i>X. axonopodis</i>	1, 6	nitrogen regulatory protein P-II 1	704483492 ref WP_033483073.1	GlnB-like
3.1.3.2	Acid phosphatase	<i>S. lycopersicum</i>	2, 6, 9, 11	PREDICTED: pentatricopeptide repeat-containing protein At5g64320, mitochondrial	K4CN29 K4CN29_SOLLC	TPR-like
		<i>B. cinerea</i>	2, 3, 7, 13	hypothetical protein BC1G_00547	XP_001560519.1	p53-like transcription factors
4.4.1.5	Lactoylglutathione lyase	<i>S. lycopersicum</i>	1, 8, 21	Lactoylglutathione lyase	Q42891 LGUL_SOLLC	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase
		<i>B. cinerea</i>	1, 3	hypothetical protein BC1G_11057	XP_001550649.1	Metallo-hydrolase/oxidoreductase
5.2.1.8*	Peptidylprolyl isomerase	<i>S. lycopersicum</i>	1, 2, 24	Peptide-proxy cis-trans isomerase	P21568 CYPH_SOLLC	Cyclophilin-like
		<i>B. cinerea</i>	1, 2, 7, 24	hypothetical protein BC1G_11305	XP_001549835.1	RING/U-box
1.11.1.6*	Catalase	<i>S. lycopersicum</i>	1, 5	Catalase	K4BVX3 K4BVX3_SOLLC	Heme-dependent catalase-like
		<i>F. oxysporum</i>	1, 2, 5	Catalase-peroxidase 2	A0A0D2YKD1 A0A0D2YKD1_FUSO4	Heme-dependent peroxidases / Heme-dependent peroxidases
1.15.1.1	Superoxide dismutase	<i>S. lycopersicum</i>	1, 4, 6, 7	Superoxide dismutase	Q7XAV2 Q7XAV2_SOLLC	Cu,Zn superoxide dismutase-like
		<i>F. oxysporum</i>	1, 4, 7, 14	Uncharacterized protein	A0A0D2YE80 A0A0D2YE80_FUSO4	Double-stranded beta-helix
3.1.3.2	Acid phosphatase	<i>S. lycopersicum</i>	2, 4, 6, 9, 11	PREDICTED: pentatricopeptide repeat-containing protein At5g64320, mitochondrial	K4CN29 K4CN29_SOLLC	TPR-like
		<i>F. oxysporum</i>	1, 2, 3, 4, 7, 13	Hypothetical protein FOXG_21887	A0A0D2YGA3 A0A0D2YGA3_FUSO4	SurE-like
3.2.1.4	Cellulase	<i>S. lycopersicum</i>	1	Endoglyucanase	Q42875 Q42875_SOLLC	Six-hairpin glycosidases
		<i>F. oxysporum</i>	1, 6	Hypothetical protein FOMG_03162	A0A0D2XJE6 A0A0D2XJE6_FUSO4	SGNH hydrolase
3.4.11.5*	Prolyl aminopeptidase	<i>S. lycopersicum</i>	1, 2	Leucine aminopeptidase 2	Q8GZD8 Q8GZD8_SOLLC	Zn-dependent exopeptidases
		<i>F. oxysporum</i>	1, 3	Uncharacterized protein	A0A0D2XCV3 A0A0D2XCV3_FUSO4	alpha/beta-Hydrolases
4.4.1.5	Lactoylglutathione lyase	<i>S. lycopersicum</i>	1, 8, 21	Lactoylglutathione lyase	Q42891 LGUL_SOLLC	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase
		<i>F. oxysporum</i>	1, 3	Hypothetical protein FOXG_04928	A0A0D2XLV4 A0A0D2XLV4_FUSO4	Metallo-hydrolase/oxidoreductase
1.11.1.7	Peroxidase	<i>S. lycopersicum</i>	3, 6	Suberization-associated anionic peroxidase 1+F1	P15003 PER1_SOLLC	Heme-dependent peroxidases
		<i>M. perniciososa</i>	6, 12	Hypothetical protein MPER_11863	E2LX62 E2LX62_MONPE	Dimeric alpha+beta barrel
2.7.13.3*	Histidine-kinase	<i>S. lycopersicum</i>	1, 20, 31, 33	Ammonium transporter 1 member	Q9FVN0 AMT13_SOLLC	Ammonium transporter
		<i>M. perniciososa</i>	1, 2	Nucleoside diphosphate kinase	E2M162 E2M162_MONPE	Nucleoside diphosphate kinase, NDK
3.4.11.1	Leucyl aminopeptidase	<i>S. lycopersicum</i>	1	Leucine aminopeptidase 2, chloroplastic	Q8GZD8 Q8GZD8_SOLLC	Zn-dependent exopeptidases / Macro domain-like
		<i>M. perniciososa</i>	1, 11	Urcharacterized protein	E2LYM3 E2LYM3_MONPE	alpha/beta-Hydrolases
5.4.2.8*	Phosphomannomutase	<i>S. lycopersicum</i>	2	Phosphomannomutase	K4CJ01 K4CJ01_SOLLC	HAD-like

1.11.1.6*	Catalase	<i>M. pernicios</i> <i>S. lycopersicum</i>	1 1, 5	Hypothetical protein MPER_03193 Catalase	E2LAS1 E2LAS1_MONPE K4BVX3 K4BVX3_SOLLC	Phosphoglucomutase, first 3 domains Heme-dependent catalase-like Heme-dependent peroxidases / Heme-dependent peroxidases
1.11.1.7*	Peroxidase	<i>P. seryngae</i> <i>S. lycopersicum</i>	1, 2 3, 6	Catalase/peroxidase HPI Suberization-associated anionic peroxidase 1 PAP2 superfamily protein/DedA family protein	NP_794283.1 P15003 PER1_SOLLC NP_794565.1	Heme-dependent peroxidases Acid phosphatase/Vanadium-dependent haloperoxidase
3.1.3.2*	Acid phosphatase	<i>P. seryngae</i> <i>S. lycopersicum</i>	6, 16, 18, 19 2, 6, 9, 11	Phospho-2-dehydro-3-deoxyheptonate aldolase 2, chloroplastic	K4CN29 K4CN29_SOLLC	TPR-like
3.2.1.14*	Chitinase	<i>P. seryngae</i> <i>S. lycopersicum</i>	1, 3 1	stationary-phase survival protein SurE Acidic 26 kDa endlchitinase	NP_791387.1 Q05539 CHIA_SOLLC	SurE-like Lysozyme-like
5.2.1.8*	Peptidylprolyl isomerase	<i>P. seryngae</i> <i>S. lycopersicum</i>	1, 3 1, 2, 24 1, 2, 14, 22, 23, 33, 34	NLP/P60 family protein Peptide-proxy cis-trans isomerase oxidoreductase FAD/FMN-binding protein	NP_794777.1 P21568 CYPH_SOLLC	Cysteine proteinases Cyclophilin-like
1.11.1.6*	Catalase	<i>P. seryngae</i> <i>S. lycopersicum</i> <i>R. solanacearum</i>	1, 5 6	Catalase	K4BVX3 K4BVX3_SOLLC 469773124 rb AGH83314.1	Heme-dependent catalase-like Heme-dependent peroxidases / Heme-dependent peroxidases
1.11.1.7*	Peroxidase	<i>S. lycopersicum</i> <i>R. solanacearum</i> <i>S. lycopersicum</i>	3, 6 6, 18	Suberization-associated anionic peroxidase 1 Type II/IV secretion system ATPase TadZ/CpaE, associated with Flp plasmid assembly (plasmid)	P15003 PER1_SOLLC 469776536 rb AGH86619.1	Heme-dependent peroxidases P-loop containing nucleoside triphosphate hydrolases
2.7.13.3*	Histidine kinase	<i>S. lycopersicum</i> <i>R. solanacearum</i> <i>S. lycopersicum</i>	1, 20, 31, 33 1, 21, 23, 24, 33, 46	Ammonium transporter 1 member 3 Phosphate transport system regulatory protein PhoU	Q9FVN0 AMT13_SOLLC 469774154 rb AGH84344.1	Ammonium transporter PhoU-like
3.2.1.14*	Chitinase	<i>S. lycopersicum</i> <i>R. solanacearum</i> <i>S. lycopersicum</i>	1 1, 3	Acidic 26 kDa endlchitinase Lipoprotein spr precursor	Q05539 CHIA_SOLLC 469773531 rb AGH83721.1	Lysozyme-like Cysteine proteinases
4.2.1.1*	Carbonic anhydrase	<i>S. lycopersicum</i> <i>R. solanacearum</i>	1, 2, 5 1, 3, 13	Sulfate transporter 4.1-like protein Molybdopterin-guanine dinucleotide biosynthesis protein B (MobB) (plasmid)	K4C2F1 K4C2F1_SOLLC 469776652 rb AGH86735.1	Spollaa-like P-loop containing nucleoside triphosphate hydrolases
1.10.2.2*	Quinol-cytochrome-c reductase	<i>Z. mays</i> <i>A. flavus</i>	2, 3, 4, 5, 8, 17 2, 3, 5, 8, 17	uncharacterized protein LOC100216756 Cytochrome c	959092556 ref NP_001304298.1 B8NGN0 B8NGN0_ASPFN	Non-heme 11 kDa protein of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase) Cytochrome c
1.11.1.15	Peroxioredoxin	<i>Z. mays</i> <i>A. flavus</i>	1, 10 1, 9	uncharacterized protein LOC100278941 4-carboxymuconolactone decarboxylase family protein	226533359 ref NP_001145525.1 B8N164 B8N164_ASPFN	Thioredoxin fold AhpD-like
1.11.1.6*	Catalase	<i>Z. mays</i> <i>A. flavus</i>	1, 5 1, 2	PREDICTED: aquaporin NIP1-4-like Bifunctional catalase-peroxidase Cat2	670441254 ref XP_008660914.1 B8NX24 B8NX24_ASPFN	Aquaporin-like Heme-dependent peroxidases
1.11.1.7	Peroxidase	<i>Z. mays</i> <i>A. flavus</i>	3 3, 6, 12	PREDICTED: peroxidase 5-like Uncharacterized protein	670447712 ref XP_008664058.1 B8NC39 B8NC39_ASPFN	Heme-dependent peroxidases Ferredoxin-like

2.1.1.43*	Histone-lysine N-methyltransferase	<i>Z. mays</i>	1, 2	PREDICTED: probable histone-lysine N-methyltransferase ATXR3	670448064 ref XP_008664254.1	F-box domain
		<i>A. flavus</i>	1, 2	SET and MYND domain protein, putative	B8NB79 B8NB79_ASPFN	beta-clip
2.5.1.18	Glutathione transferase	<i>Z. mays</i>	1, 4	PREDICTED: uncharacterized protein LOC100216831 isoform X1	670358514 ref XP_008665261.1	MAPEG domain-like
		<i>A. flavus</i>	1, 3	2-hydroxychromene-2-carboxylate isomerase, putative	B8N9N8 B8N9N8_ASPFN	Thioredoxin fold
2.6.1.1	Aspartate transaminase	<i>Z. mays</i>	1	PREDICTED: probable LL-diaminopimelate aminotransferase, chloroplastic	670439790 ref XP_008660232.1	PLP-dependent transferases
		<i>A. flavus</i>	1, 5	Uncharacterized protein	B8NQM9 B8NQM9_ASPFN	Ferredoxin-like
2.7.11.22*	Cyclin-dependent kinase	<i>Z. mays</i>	1, 7	PREDICTED: putative kinase-like protein TMKL1	670446665 ref XP_008663534.1	Protein kinase-like (PK-like)
		<i>A. flavus</i>	9	CDK-activating kinase assembly factor MAT1	B8N9A7 B8N9A7_ASPFN	RING/U-box
3.1.3.2	Acid phosphatase	<i>Z. mays</i>	2, 4, 6, 9	PREDICTED: pentatricopeptide repeat-containing protein At5g14770, mitochondrial-like	670448433 ref XP_008664470.1	alpha-alpha superhelix
		<i>A. flavus</i>	2, 13	Uncharacterized protein	B8NB93 B8NB93_ASPFN	p53-like transcription factors
3.2.2.22	rRNA N-glycosylase	<i>Z. mays</i>	1, 7	PREDICTED: ribosome-inactivating protein	670431358 ref XP_008656307.1	Ribosome inactivating proteins (RIP)
		<i>A. flavus</i>	5	GPI anchored protein, putative	B8NQT3 B8NQT3_ASPFN	alpha/alpha toroid
4.2.1.1*	Carbonic anhydrase	<i>Z. mays</i>	1, 2, 5	PREDICTED: gamma carbonic anhydrase 1, mitochondrial-like	670429179 ref XP_008655471.1	RRF/tRNA synthetase additional domain-like
		<i>A. flavus</i>	1, 2	Carbonic anhydrase, putative	B8NWM8 B8NWM8_ASPFN	Carbonic anhydrase
4.4.1.5	Lactoylglutathione lyase	<i>Z. mays</i>	1, 8, 21	lactoylglutathione lyase	226500150 ref NP_001148888.1	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase
		<i>A. flavus</i>	1, 3	Metallo-beta-lactamase family protein	B8NT23 B8NT23_ASPFN	Metallo-hydrolase/oxidoreductase
5.1.3.1*	Ribulose-phosphate 3-epimerase	<i>Z. mays</i>	1, 4	ribulose-phosphate 3-epimerase	226497150 ref NP_001149850.1	TIM beta/alpha-barrel
		<i>A. flavus</i>	1, 4, 15	Coronin	B8N7U5 B8N7U5_ASPFN	7-bladed beta-propeller
5.3.1.6*	Ribose-5-phosphate isomerase	<i>Z. mays</i>	1	PREDICTED: probable ribose-5-phosphate isomerase 2	670406652 ref XP_008644870.1	Ferredoxin-like
		<i>A. flavus</i>	1, 2	Uncharacterized protein	B8NFW5 B8NFW5_ASPFN	Ribose/Galactose isomerase RpiB/AlsB
1.1.1.27*	L-lactate dehydrogenase	<i>Z. mays</i>	1, 12	PREDICTED: L-lactate dehydrogenase B-like	670434471 ref XP_008657765.1	LDH C-terminal domain-like
		<i>C. graminicola</i>	2, 12	FMN-dependent dehydrogenase	827078716 ref XP_008100733.1	FMN-linked oxidoreductases
1.11.1.15	Peroxiredoxin	<i>Z. mays</i>	1, 10	uncharacterized protein LOC100278941	226533359 ref NP_001145525.1	Thioredoxin fold
		<i>C. graminicola</i>	1, 9	carboxymuconolactone decarboxylase	827063182 ref XP_008093145.1	AhpD-like
1.11.1.6*	Catalase	<i>Z. mays</i>	1, 5	PREDICTED: aquaporin NIP1-4-like	670441254 ref XP_008660914.1	Aquaporin-like
		<i>C. graminicola</i>	1, 2, 5	catalase/oxidase HPI	827074149 ref XP_008098502.1	Heme-dependent peroxidases // Heme-dependent peroxidases
1.11.1.7*	Peroxidase	<i>Z. mays</i>	3	PREDICTED: peroxidase 5-like	670447712 ref XP_008664058.1	Heme-dependent peroxidases
		<i>C. graminicola</i>	3, 6, 8	AhpC/TSA family protein	827068931 ref XP_008095952.1	Thioredoxin fold
2.7.11.22*	Cyclin-dependent kinase	<i>Z. mays</i>	1, 7	PREDICTED: putative kinase-like protein TMKL1	670446665 ref XP_008663534.1	Protein kinase-like (PK-like)

		<i>C. graminicola</i>	9	CDK-activating kinase assembly factor MAT1 PREDICTED: LOC100285062 isoform X1	827066639 ref XP_008094831.1	RING/U-box
3.1.1.31	6-phosphogluconolactonase	<i>Z. mays</i>	1		670421290 ref XP_008651541.1	NagB/RpiA/CoA transferase-like
		<i>C. graminicola</i>	1, 2	hypothetical protein GLRG_11252 PREDICTED: LOC100281846 isoform X1	827077465 ref XP_008100128.1	8-bladed beta-propeller
3.1.1.4*	Phospholipase A2	<i>Z. mays</i>	1, 3, 8, 10, 16, 17, 18		670359988 ref XP_008675577.1	Phospholipase A2, PLA2
		<i>C. graminicola</i>	1, 2, 17, 18	Cytosolic phospholipase A2 zeta PREDICTED: uncharacterized protein LOC100273611 isoform X1	827069253 ref XP_008096109.1	FabD/lysophospholipase-like
3.1.13.4*	Poly(A)-specific ribonuclease	<i>Z. mays</i>	1, 2, 3		670435395 ref XP_008658269.1	DNase I-like
		<i>C. graminicola</i>	2, 3, 5, 6, 7	Hypothetical protein GLRG_04615 PREDICTED: pentatricopeptide repeat-containing protein At5g14770, mitochondrial-like	827063896 ref XP_008093491.1	Protein kinase-like (PK-like)
3.1.3.2	Acid phosphatase	<i>Z. mays</i>	2, 4, 6, 9	NDT80/PhoG like DNA-binding family protein PREDICTED: LOC100285190 isoform X1	670448433 ref XP_008664470.1	alpha-alpha superhelix Common fold of diphtheria toxin/transcription factors/cytochrome f
		<i>C. graminicola</i>	2, 3, 5, 13		827066884 ref XP_008094949.1	
3.1.3.3*	Phosphoserine phosphatase	<i>Z. mays</i>	1, 8		670393897 ref XP_008677367.1	HAD-like
		<i>C. graminicola</i>	1, 8	MJ0936 family phosphodiesterase	827071996 ref XP_008097450.1	Metallo-dependent phosphatases Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase
4.4.1.5	Lactoylglutathione lyase	<i>Z. mays</i>	1, 8, 21	lactoylglutathione lyase metallo-beta-lactamase superfamily protein	226500150 ref NP_001148888.1	
		<i>C. graminicola</i>	1, 3		827070823 ref XP_008096879.1	Metallo-hydrolase/oxidoreductase
5.1.3.1*	Ribulose-phosphate 3-epimerase	<i>Z. mays</i>	1, 4	ribulose-phosphate 3-epimerase	226497150 ref NP_001149850.1	TIM beta/alpha-barrel
		<i>C. graminicola</i>	1, 4, 2, 15	WD domain-containing protein PREDICTED: probable ribose-5-phosphate isomerase 2	827059158 ref XP_008091175.1	7-bladed beta-propeller
5.3.1.6*	Ribose-5-phosphate isomerase	<i>Z. mays</i>	1		670406652 ref XP_008644870.1	Ferredoxin-like
		<i>C. graminicola</i>	1, 2	ribose 5-phosphate isomerase PREDICTED: L-lactate dehydrogenase B-like	827073555 ref XP_008098210.1	Ribose/Galactose isomerase RpiB/AlsB
1.1.1.27*	L-lactate dehydrogenase	<i>Z. mays</i>	1, 12	hypothetical protein SETTUDRAFT_184979	670434471 ref XP_008657765.1	LDH C-terminal domain-like
		<i>E. turcicum</i>	1, 12, 2		636590563 ref XP_008026620.1	FMN-linked oxidoreductases
1.11.1.15	Peroxioredoxin	<i>Z. mays</i>	1, 10	uncharacterized protein LOC100278941 hypothetical protein SETTUDRAFT_163404	226533359 ref NP_001145525.1	Thioredoxin fold
		<i>E. turcicum</i>	1, 9		636589077 ref XP_008025877.1	AhpD-like
1.11.1.6*	Catalase	<i>Z. mays</i>	1, 5	catalase isozyme 2 hypothetical protein SETTUDRAFT_164921	806776638 ref NP_001105310.2	Heme-dependent catalase-like Heme-dependent peroxidases / Heme-dependent peroxidases
		<i>E. turcicum</i>	1, 2, 5		636595905 ref XP_008029291.1	
1.11.1.7*	Peroxidase	<i>Z. mays</i>	3	PREDICTED: peroxidase 5-like hypothetical protein SETTUDRAFT_165877	670447712 ref XP_008664058.1	Heme-dependent peroxidases
		<i>E. turcicum</i>	3, 6		636599065 ref XP_008030871.1	Thioredoxin fold
2.1.1.43*	Histone-lysine N-methyltransferase	<i>Z. mays</i>	1, 2	PREDICTED: probable histone-lysine N-methyltransferase ATXR3 hypothetical protein SETTUDRAFT_163388	670448064 ref XP_008664254.1	F-box domain S-adenosyl-L-methionine-dependent methyltransferases
		<i>E. turcicum</i>	2, 3, 7		636589043 ref XP_008025860.1	
2.7.11.22*	Cyclin-dependent kinase	<i>Z. mays</i>	1, 7	PREDICTED: putative kinase-like protein TMKL1 hypothetical protein SETTUDRAFT_27723	670446665 ref XP_008663534.1	Protein kinase-like (PK-like)
		<i>E. turcicum</i>	9		636585459 ref XP_008024068.1	RING/U-box

3.1.1.31	6-phosphogluconolactonase	<i>Z. mays</i>	1	PREDICTED: LOC100285062 isoform X1	670421290 ref XP_008651541.1	NagB/RpiA/CoA transferase-like
		<i>E. turcicum</i>	1, 2	hypothetical protein SETTUDRAFT_43562	636596317 ref XP_008029497.1	8-bladed beta-propeller
3.1.3.2*	Acid phosphatase	<i>Z. mays</i>	2, 4, 6, 9	PREDICTED: pentatricopeptide repeat-containing protein At5g14770, mitochondrial-like	670448433 ref XP_008664470.1	alpha-alpha superhelix
		<i>E. turcicum</i>	1, 2, 3, 13	hypothetical protein SETTUDRAFT_169046	636586991 ref XP_008024834.1	SurE-like Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase
4.4.1.5	Lactoylglutathione lyase	<i>Z. mays</i>	1, 8, 21	lactoylglutathione lyase	226500150 ref NP_001148888.1	protein/Dihydroxybiphenyl dioxygenase
		<i>E. turcicum</i>	1, 3	hypothetical protein SETTUDRAFT_131356	636589467 ref XP_008026072.1	Metallo-hydrolase/oxidoreductase
4.6.1.1*	Adenylate cyclase	<i>Z. mays</i>	2, 8, 17	adenylate cyclase	219363285 ref NP_001136955.1	CYTH-like phosphatases
		<i>E. turcicum</i>	2, 8, 10	hypothetical protein SETTUDRAFT_149731	636585855 ref XP_008024266.1	OsmC-like
5.1.3.1*	Ribulose-phosphate 3-epimerase	<i>Z. mays</i>	1, 4	ribulose-phosphate 3-epimerase	226497150 ref NP_001149850.1	TIM beta/alpha-barrel
		<i>E. turcicum</i>	1, 4, 15	hypothetical protein SETTUDRAFT_164582	636595191 ref XP_008028934.1	7-bladed beta-propeller
5.3.1.6*	Ribose-5-phosphate isomerase	<i>Z. mays</i>	1	PREDICTED: probable ribose-5-phosphate isomerase 2	670406652 ref XP_008644870.1	Ferredoxin-like
		<i>E. turcicum</i>	1, 2	hypothetical protein SETTUDRAFT_92221	636594211 ref XP_008028444.1	Ribose/Galactose isomerase RpiB/AlsB
1.11.1.15	Peroxiredoxin	<i>Z. mays</i>	1, 10	uncharacterized protein LOC100278941	226533359 ref NP_001145525.1	Thioredoxin fold
		<i>G. moniliformis</i>	1, 2, 9	Uncharacterized protein	W7LPB7 W7LPB7_GIBM7	AhpD-like
2.6.1.1	Asparate transaminase	<i>Z. mays</i>	1	PREDICTED: probable LL-diaminopimelate aminotransferase, chloroplastic	670439790 ref XP_008660232.1	PLP-dependent transferases
		<i>G. moniliformis</i>	1, 5	Iron-dependent peroxidase	W7MC41 W7MC41_GIBM7	Ferredoxin-like
2.7.11.22*	Cyclin-dependent kinase	<i>Z. mays</i>	1, 7	PREDICTED: putative kinase-like protein TMKL1	670446665 ref XP_008663534.1	Protein kinase-like (PK-like)
		<i>G. moniliformis</i>	7, 9	Cyclin H	W7MSL6 W7MSL6_GIBM7	Cyclin-like // Cyclin-like
3.1.1.31	6-phosphogluconolactonase	<i>Z. mays</i>	1	PREDICTED: LOC100285062 isoform X1	670421290 ref XP_008651541.1	NagB/RpiA/CoA transferase-like
		<i>G. moniliformis</i>	1, 2	Uncharacterized protein	W7M0K8 W7M0K8_GIBM7	8-bladed beta-propeller
3.1.13.4*	Poly(A)-specific ribonuclease	<i>Z. mays</i>	1, 2, 3	PREDICTED: uncharacterized protein LOC100273611 isoform X1	670435395 ref XP_008658269.1	DNase I-like
		<i>G. moniliformis</i>	2, 3, 5, 6, 7	Uncharacterized protein	W7M4G2 W7M4G2_GIBM7	Ribonuclease H-like – motif
3.1.3.2*	Acid phosphatase	<i>Z. mays</i>	2, 4, 6, 9	PREDICTED: pentatricopeptide repeat-containing protein At5g14770, mitochondrial-like	670448433 ref XP_008664470.1	alpha-alpha superhelix
		<i>G. moniliformis</i>	1, 2, 3, 7, 13	5'/3'-nucleotidase SurE	W7NDR6 W7NDR6_GIBM7	SurE-like
3.2.1.4	Cellulase	<i>Z. mays</i>	1	PREDICTED: endoglucanase 2-like isoform X2	670429955 ref XP_008655784.1	alpha/alpha toroid
		<i>G. moniliformis</i>	1, 6	Uncharacterized protein	W7M5R3 W7M5R3_GIBM7	Flavodoxin-like

4.4.1.5	Lactoylglutathione lyase	<i>Z. mays</i> G. <i>moniliformis</i>	1, 8, 21 1, 3	lactoylglutathione lyase Metallo-hydrolase/oxidoreductase	226500150 ref NP_001148888.1 W7LNQ2 W7LNQ2_GIBM7	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase Metallo-hydrolase/oxidoreductase
4.6.1.1*	Adenylate cyclase	<i>Z. mays</i> G. <i>moniliformis</i>	2, 8, 17 2, 8	adenylate cyclase Adenylate cyclase	219363285 ref NP_001136955.1 W7MFF7 W7MFF7_GIBM7	CYTH-like phosphatases Leucine-rich repeat, LRR (right-handed beta-alpha superhelix)
5.1.3.1*	Ribulose-phosphate 3-epimerase	<i>Z. mays</i> G. <i>moniliformis</i>	1, 4 1, 15 , 4	ribulose-phosphate 3-epimerase Coronin PREDICTED: uncharacterized protein LOC100272300 isoform X1	226497150 ref NP_001149850.1 W7M917 W7M917_GIBM7 670372214 ref XP_008667406.1	TIM beta/alpha-barrel 7-bladed beta-propeller Thioredoxin fold
1.11.1.15*	Peroxiredoxin	<i>Z. mays</i> <i>P. ananatis</i>	1, 10 1, 2	PANAM OsmC	D4GL47 D4GL47_PANAM	OsmC-like
1.11.1.6	Catalase	<i>Z. mays</i> <i>P. ananatis</i>	1, 5 3, 5, 6	catalase isozyme 2 YciF PREDICTED: frataxin isoform 1 isoform X1	806776638 ref NP_001105310.2 D4GMF4 D4GMF4_PANAM 670359635 ref XP_008672910.1	Heme-dependent catalase-like Ferritin-like N domain of copper amine oxidase-like
1.16.3.1	Ferroxidase	<i>Z. mays</i> <i>P. ananatis</i>	2, 6 2, 7	Ferroxidase PREDICTED: ornithine carbamoyltransferase, chloroplastic-like	D4GC12 D4GC12_PANAM 670440524 ref XP_008660532.1	Ferritin-like ATC-like
2.1.3.3*	Ornithine carbamoyltransferase	<i>Z. mays</i> <i>P. ananatis</i>	1, 10 2	PotC PREDICTED: uncharacterized protein LOC103637104	D4GJ68 D4GJ68_PANAM 670434153 ref XP_008657589.1	MetI-like Glycerol-3-phosphate (1)-acyltransferase
2.3.1.51*	1-acylglycerol-3-phosphate O-acyltransferase	<i>Z. mays</i> <i>P. ananatis</i>	1, 6 1, 11	Single-stranded DNA-binding protein PREDICTED: phosphoglycerate kinase, chloroplastic-like	D4GHC5 D4GHC5_PANAM 670431622 ref XP_008656415.1	OB-fold Phosphoglycerate kinase
2.7.2.3	Phosphoglycerate kinase	<i>Z. mays</i> <i>P. ananatis</i>	1 3	YjbB PREDICTED: LOC100284281 isoform X1	D4GHA1 D4GHA1_PANAM 670443226 ref XP_008662013.1	Spectrin repeat-like P-loop containing nucleoside triphosphate hydrolases
2.7.4.8*	Guanylate kinase	<i>Z. mays</i> <i>P. ananatis</i>	1 1, 4, 7	Pal PREDICTED: LOC100285062 isoform X1	D4GMM0 D4GMM0_PANAM 670421290 ref XP_008651541.1	Bacillus chorismate mutase-like NagB/RpiA/CoA transferase-like
3.1.1.31	6-phosphogluconolactonase	<i>Z. mays</i> <i>P. ananatis</i>	1 2, 6	YkgB PREDICTED: acyl-protein thioesterase 2 isoform X1	D4GFB8 D4GFB8_PANAM 670384215 ref XP_008672924.1	8-bladed beta-propeller alpha/beta-Hydrolases
3.1.1.5	Lysophospholipase	<i>Z. mays</i> <i>P. ananatis</i>	2, 6, 7, 18, 22 2, 5	TesA OS PREDICTED: fructose-1,6- bisphosphatase, cytosolic-like, partial	D4GGT2 D4GGT2_PANAM 670419266 ref XP_008650400.1	Flavodoxin-like Carbohydrate phosphatase
3.1.3.11*	Fructose-bisphosphatase	<i>Z. mays</i> <i>P. ananatis</i>	1 10, 12	Threonylcarbamoyl-AMP synthase	D4GCE1 D4GCE1_PANAM	YrdC/RibB
4.2.1.96*	4a-hydroxytetrahydrobiopterin dehydratase	<i>Z. mays</i> <i>P. ananatis</i>	1 2	PREDICTED: pterin-4-alpha- carbinolamine dehydratase isoform X1 Molybdenum cofactor biosynthesis protein B	670384123 ref XP_008672875.1 D4GMQ4 D4GMQ4_PANAM	DCoH-like Molybdenum cofactor biosynthesis proteins
4.3.3.7*	4-hydroxy-tetrahydrodipicolinate synthase	<i>Z. mays</i>	1	4-hydroxy-tetrahydrodipicolinate synthase, chloroplastic precursor	162462912 ref NP_001105425.1	Aldolase

*Analogous essential enzymes

Tabela Suplementar 2. Enzimas selecionadas no presente estudo e suas respectivas vias metabólicas.

Hospedeiros	Patógenos	Número EC	Enzimas	Vias Metabólicas
<i>G. max</i>	<i>A. flavus</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>G. max</i>	<i>A. flavus</i>	1.2.1.3	Aldehyde dehydrogenase (NAD ⁺)	Carbohydrate metabolism / Amino acid metabolism / Xenobiotics biodegradation and metabolism/ Lipid metabolism
<i>G. max</i>	<i>F. oxysporum</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>G. max</i>	<i>P. sojae</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>G. max</i>	<i>X. axonopodis</i>	1.15.1.1	Superoxide dismutase	No such data*
<i>G. max</i>	<i>X. axonopodis</i>	6.4.1.2	Acetyl-CoA carboxylase	Carbohydrate metabolism / Biosynthesis of other secondary metabolites / Lipid metabolism
<i>S. lycopersicum</i>	<i>B. cinerea</i>	5.2.1.8	Peptidylprolyl isomerase	No such data
<i>S. lycopersicum</i>	<i>F. oxysporum</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>S. lycopersicum</i>	<i>F. oxysporum</i>	3.4.11.5	Prolyl aminopeptidase	Amino acid metabolism
<i>S. lycopersicum</i>	<i>M. perniciosa</i>	2.7.13.3	Histidine-kinase	No such data
<i>S. lycopersicum</i>	<i>M. perniciosa</i>	5.4.2.8	Phosphomannomutase	Carbohydrate metabolism
<i>S. lycopersicum</i>	<i>P. seryngae</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism

<i>S. lycopersicum</i>	<i>P. seryngae</i>	1.11.1.7	Peroxidase	Biosynthesis of other secondary metabolites
<i>S. lycopersicum</i>	<i>P. seryngae</i>	3.1.3.2	Acid phosphatase	Metabolism of cofactors and vitamins
<i>S. lycopersicum</i>	<i>P. seryngae</i>	3.2.1.14	Chitinase	Carbohydrate metabolism
<i>S. lycopersicum</i>	<i>P. seryngae</i>	5.2.1.8	Peptidylprolyl isomerase	No such data
<i>S. lycopersicum</i>	<i>solanacearum</i> <i>R.</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>S. lycopersicum</i>	<i>solanacearum</i> <i>R.</i>	1.11.1.7	Peroxidase	Biosynthesis of other secondary metabolites
<i>S. lycopersicum</i>	<i>solanacearum</i> <i>R.</i>	2.7.13.3	Histidine kinase	No such data
<i>S. lycopersicum</i>	<i>solanacearum</i> <i>R.</i>	3.2.1.14	Chitinase	Carbohydrate metabolism
<i>S. lycopersicum</i>	<i>solanacearum</i>	4.2.1.1	Carbonic anhydrase	Energy metabolism
<i>Z. mays</i>	<i>A. flavus</i>	1.10.2.2	Quinol-cytochrome-c reductase	Energy metabolism
<i>Z. mays</i>	<i>A. flavus</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>Z. mays</i>	<i>A. flavus</i>	2.1.1.43	Histone-lysine N-methyltransferase	Amino acid metabolism
<i>Z. mays</i>	<i>A. flavus</i>	2.7.11.22	Cyclin-dependent kinase	No such data
<i>Z. mays</i>	<i>A. flavus</i>	4.2.1.1	Carbonic anhydrase	Energy metabolism
<i>Z. mays</i>	<i>A. flavus</i>	5.1.3.1	Ribulose-phosphate 3-epimerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>A. flavus</i>	5.3.1.6	Ribose-5-phosphate isomerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>E. turcicum</i>	1.1.1.27	L-lactate dehydrogenase	Carbohydrate metabolism / Amino acid metabolism
<i>Z. mays</i>	<i>E. turcicum</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>Z. mays</i>	<i>E. turcicum</i>	1.11.1.7	Peroxidase	Biosynthesis of other secondary metabolites
<i>Z. mays</i>	<i>E. turcicum</i>	2.1.1.43	Histone-lysine N-methyltransferase	Amino acid metabolism
<i>Z. mays</i>	<i>E. turcicum</i>	2.7.11.22	Cyclin-dependent kinase	No such data
<i>Z. mays</i>	<i>E. turcicum</i>	3.1.3.2	Acid phosphatase	Metabolism of cofactors and vitamins
<i>Z. mays</i>	<i>E. turcicum</i>	4.6.1.1	Adenylate cyclase	Nucleotide metabolism
<i>Z. mays</i>	<i>E. turcicum</i>	5.1.3.1	Ribulose-phosphate 3-epimerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>E. turcicum</i>	5.3.1.6	Ribose-5-phosphate isomerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>C. graminicola</i>	1.1.1.27	L-lactate dehydrogenase	Carbohydrate metabolism / Amino acid metabolism
<i>Z. mays</i>	<i>C. graminicola</i>	1.11.1.6	Catalase	Carbohydrate metabolism / Amino acid metabolism
<i>Z. mays</i>	<i>C. graminicola</i>	1.11.1.7	Peroxidase	Biosynthesis of other secondary metabolites

<i>Z. mays</i>	<i>C. graminicola</i>	2.7.11.22	Cyclin-dependent kinase	No such data
<i>Z. mays</i>	<i>C. graminicola</i>	3.1.1.4	Phospholipase A2	Lipid metabolism
<i>Z. mays</i>	<i>C. graminicola</i>	3.1.13.4	Poly(A)-specific ribonuclease	No such data
<i>Z. mays</i>	<i>C. graminicola</i>	3.1.3.3	Phosphoserine phosphatase	Amino acid metabolism / Energy metabolism
<i>Z. mays</i>	<i>C. graminicola</i>	5.1.3.1	Ribulose-phosphate 3-epimerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>C. graminicola</i>	5.3.1.6	Ribose-5-phosphate isomerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>G. moniliformis</i>	2.7.11.22	Cyclin-dependent kinase	No such data
<i>Z. mays</i>	<i>G. moniliformis</i>	3.1.13.4	Poly(A)-specific ribonuclease	No such data
<i>Z. mays</i>	<i>G. moniliformis</i>	3.1.3.2	Acid phosphatase	Metabolism of cofactors and vitamins
<i>Z. mays</i>	<i>G. moniliformis</i>	4.6.1.1	Adenylate cyclase	Nucleotide metabolism
<i>Z. mays</i>	<i>G. moniliformis</i>	5.1.3.1	Ribulose-phosphate 3-epimerase	Carbohydrate metabolism / Energy metabolism
<i>Z. mays</i>	<i>P. ananatis</i>	1.11.1.15	Peroxiredoxin	Metabolism of other amino acids
<i>Z. mays</i>	<i>P. ananatis</i>	2.1.3.3	Ornithine carbamoyltransferase	Amino acid metabolism
<i>Z. mays</i>	<i>P. ananatis</i>	2.3.1.51	1-acylglycerol-3-phosphate O-acyltransferase	Lipid metabolism
<i>Z. mays</i>	<i>P. ananatis</i>	2.7.4.8	Guanylate kinase	Nucleotide metabolism
<i>Z. mays</i>	<i>P. ananatis</i>	3.1.3.11	Fructose-bisphosphatase	Energy metabolism / Carbohydrate metabolism
<i>Z. mays</i>	<i>P. ananatis</i>	4.2.1.96	4a-hydroxytetrahydrobiopterin dehydratase	Metabolism of cofactors and vitamins
<i>Z. mays</i>	<i>P. ananatis</i>	4.3.3.7	4-hydroxy-tetrahydrodipicolinate synthase	Biosynthesis of other secondary metabolites /Amino acid metabolism

10. ANEXOS

10.1 Artigo relacionado à tese publicado em periódico Qualis A1.

RESEARCH ARTICLE

Mining of potential drug targets through the identification of essential and analogous enzymes in the genomes of pathogens of *Glycine max*, *Zea mays* and *Solanum lycopersicum*

Rangeline Azevedo da Silva^{1*}, Leandro de Mattos Pereira², Melise Chaves Silveira¹, Rodrigo Jardim¹, Antonio Basilio de Miranda¹

1 Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Rio de Janeiro, Brazil, **2** Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

* rangeline.as@gmail.com



OPEN ACCESS

Citation: Silva RA, Pereira LdM, Silveira MC, Jardim R, Miranda ABd (2018) Mining of potential drug targets through the identification of essential and analogous enzymes in the genomes of pathogens of *Glycine max*, *Zea mays* and *Solanum lycopersicum*. PLoS ONE 13(5): e0197511. <https://doi.org/10.1371/journal.pone.0197511>

Editor: David A. Lightfoot, College of Agricultural Sciences, UNITED STATES

Received: November 3, 2017

Accepted: May 3, 2018

Published: May 25, 2018

Copyright: © 2018 Silva et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Coordination for the Improvement of Higher Education Personnel, Brazil (<http://www.capes.gov.br/>): (RAS, MCS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Pesticides are one of the most widely used pest and disease control measures in plant crops and their indiscriminate use poses a direct risk to the health of populations and environment around the world. As a result, there is a great need for the development of new, less toxic molecules to be employed against plant pathogens. In this work, we employed an *in silico* approach to study the genes coding for enzymes of the genomes of three commercially important plants, soybean (*Glycine max*), tomato (*Solanum lycopersicum*) and corn (*Zea mays*), as well as 15 plant pathogens (4 bacteria and 11 fungi), focusing on revealing a set of essential and non-homologous isofunctional enzymes (NISEs) that could be prioritized as drug targets. By combining sequence and structural data, we obtained an initial set of 568 cases of analogy, of which 97 were validated and further refined, revealing a subset of 29 essential enzymatic activities with a total of 119 different structural forms, most belonging to central metabolic routes, including the carbohydrate metabolism, the metabolism of amino acids, among others. Further, another subset of 26 enzymatic activities possess a tertiary structure specific for the pathogen, not present in plants, men and *Apis mellifera*, which may be of importance for the development of specific enzymatic inhibitors against plant diseases that are less harmful to humans and the environment.

Introduction

One of the major challenges for plant breeders is to maintain high levels of quality and production of cultures. Diseases caused by plant pathogens are one of the main factors limiting the productivity of large commodities, such as soybean (*Glycine max*), corn (*Zea mays*) and tomato (*Solanum lycopersicum*) [1,2]. Use of pesticides is one of the most commonly used alternatives to plant pathogens control, being used in a wide variety of crops [3].

Competing interests: The authors have declared that no competing interests exist.

Pesticides affect various population groups, including farm workers, residents in neighboring areas, consumers and wild animals [4,5]. Handling and consumption of these products are responsible for a series of conditions including acute intoxications [6], Parkinson's disease [7], skin diseases [8], congenital malformations [9] and the onset of cancer after long periods of exposure [10]. An increase of 93% in the world's consumption of pesticides was observed in the last two decades, while in Brazil, the largest consumer of pesticides in the world [11, 12], this increase was of 190%. New control alternatives are desired, where the new measures do not affect the development and production of the plant and present a lower risk of contamination for man and the environment [13,14].

Enzymes catalyze hundreds of successive reactions, consisting of highly coordinated processes indispensable for the maintenance of the life of an organism [15, 16]. Essential enzymes, which tend to be conserved between closely related organisms [17, 18] have been the subject of study as targets for diseases caused by a variety of organisms [19–25], including plant pathogens like *Pseudomonas syringae* [26] and *Xanthomonas* spp. [27]. Comparative genomic approaches, taking advantage of the huge amount of sequence data generated in the last decade, may contribute in several ways to the identification of key enzymes in the phytopathogens' genomes [28, 29].

Enzyme classification follows rules defined by the International Union of Biochemistry and Molecular Biology Nomenclature Committee (NC-IUBMB), in association with the International Union of Pure and Applied Chemistry (IUPAC). A four-digit classification scheme known as the Enzyme Commission Number (EC) was proposed by this committee [30]. The first three digits are those that define the catalyzed reaction, the second and third comprise the subclasses of the reactions, and the fourth digit is a unique identifier that corresponds to the catalytic activity itself. Enzymes can also be grouped into families based on sequence similarity, and families are organized into superfamilies according to the catalytic activity [31]. Sequence motifs and domain architecture are the main criteria employed, but other characteristics can be used [32]. This diversity may result in functional overlap: these cases are known as non-homologous isofunctional enzymes (NISEs), also known as functional analogous enzymes [33, 34]. Analogous enzymes perform the same biochemical function, but have different evolutionary origins, with distinct primary structures whose differences are reflected in their tertiary structures [35]. Convergent evolution, initially thought to be a rare phenomenon in enzyme evolution, has been demonstrated for several enzymes including superoxide dismutase [36–38] and proteases [39]. Later, cases of functional analogy were found in most biochemical pathways [40–42]. Most importantly, the structural differences found between analogous enzymes from the plant and the phytopathogen, a consequence of their different evolutionary origins, may be exploited for the design of specific molecules that will interact only with the form found in the phytopathogen, leaving the plant and other important species, particularly men itself and *Apis mellifera*, one of the most important pollinators [43,44], unharmed.

Thus, the objective of this study was to develop and implement a computational approach to i) identify and validate a set of NISEs, ii) reveal a subset of essential analogous enzymes and iii) disclose a subset of specific enzymatic structures, possessed only by the pathogens. To test our approach, we studied the genomes of three plants of great economic importance and worldwide distribution, *Glycine max*, *Zea mays* and *Solanum lycopersicum*, 15 bacterial and fungal plant pathogens, the genomes of *Homo sapiens*, *Apis mellifera* and two beneficial microorganisms, *Bacillus subtilis* and *Trichoderma harzianum*.

Material and methods

The analyzes were performed in four main stages: data preparation, clustering, functional inference, structural validation, and essentiality. A flowchart of the methodology is shown in Fig 1.

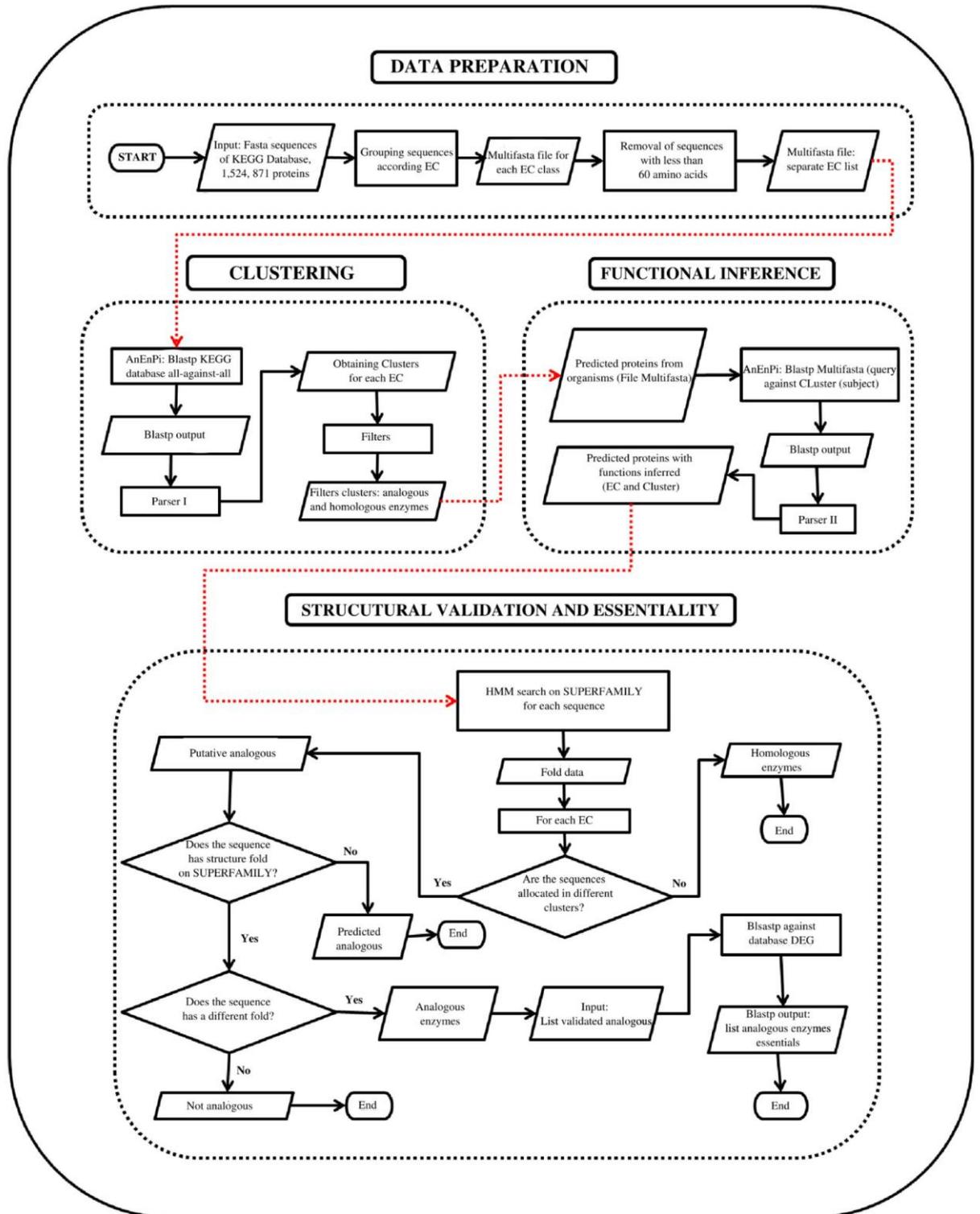


Fig 1. Identification of essential, non-homologous isofunctional enzymes.

<https://doi.org/10.1371/journal.pone.0197511.g001>

Datasets and clustering

The datasets of predicted proteins for each genome studied in this work were obtained from UniprotKB (version 2015_10 <http://www.uniprot.org/>) and RefSeq (Version 70, <http://www.ncbi.nlm.nih.gov/>). These datasets contained several proteins annotated as "uncharacterized", "hypothetical" and / or "putative". Three plant genomes were analyzed: *G. max*, *Z. mays* and *S. lycopersicum*. Pathogens were chosen according to the geographic distribution of the disease, most of them with a cosmopolitan occurrence. The pathogens analyzed comprise eleven fungal and four bacterial genomes, all pathogenic to one or more species of the plants studied. Also included were the genomes of *Homo sapiens*, *Apis mellifera* (pollinator), *Trichoderma harzianum* (soil fungus) and *Bacillus subtilis* (plant growth promoting bacteria) (Table 1).

The complete, annotated set of enzymes was extracted from KEGG (release 73.0, January 2015) and contained 1,524,871 protein sequences, from 298 Eukaryotes, 3014 Eubacteria and 175 Archaea genomes. Sequences with less than 60 amino acids were removed. To clusterize the sequences into groups based on sequence similarity, we used the AnEnPi pipeline [67]. A similarity score with a cut-off value of 120 was used for all BLASTp pairwise comparisons since this

Table 1. Description of the predicted proteins datasets of the organisms included in this study.

Organisms	Database	Accession NCBI	Reference	#Ptn	Unch.	Hyp.	Put.	Annot. (%)
<i>Glycine max</i>	RefSeq	NC_016088	[45]	59374	23618	—	1566	61
<i>Aspergillus flavus</i> ^{1*}	RefSeq	GCA_000006275.2	[46]	13287	5380	—	—	59
<i>Fusarium oxysporum</i> ^{2*}	Uniprot	GCA_000222805.1	[47]	17385	16,684	—	1	8
<i>Phytophthora sojae</i> [*]	RefSeq	AAQY00000000	[48]	26106	—	25279	125	2,8
<i>Sclerotinia sclerotiorum</i> [*]	RefSeq	AAGT00000000.1	[49]	12902	12,042	—	3	6,6
<i>Xanthomonas axonopodis</i> ^{**}	RefSeq	CP004399	[50]	4496	1413	—	35	67
<i>Solanum lycopersicum</i>	Uniprot	AEKE00000000	[51]	31683	28785	—	—	9,1
<i>Botrytis cinerea</i> [*]	RefSeq	NZ_AAID00000000.1	[52]	14687	—	8,696	—	40
<i>Fusarium oxysporum</i> ^{3*}	Uniprot	GCA_000149955.2	[53]	15811	15,148	—	—	4,3
<i>Monilophthora perniciosa</i> [*]	Uniprot	ABRE00000000	[54]	12915	12,741	—	—	1,3
<i>Pseudomonas syringae</i> ^{**}	RefSeq	NC_004578.1	[55]	5449	—	1446	—	73
<i>Ralstonia solanacearum</i> ^{**}	RefSeq	NC_003295.1	[56]	4400	696	135	1292	56
<i>Zea mays</i>	RefSeq	LPUQ00000000	[57]	59384	—	2363	2300	92
<i>Aspergillus flavus</i> ^{4*}	RefSeq	GCA_000952835.1	[58]	13561	—	5423	5884	16
<i>Colletotrichum graminicola</i> [*]	RefSeq	ACOD00000000	[59]	11910	—	5,381	—	54
<i>Gibberella moniliformis</i> [*]	Uniprot	AAIM00000000.2	[60]	17384	13,71	—	—	21
<i>Exserohilum turcicum</i> [*]	RefSeq	AIHT00000000	[61]	4248	—	11159	1	3,6
<i>Pantoea ananatis</i> ^{**}	RefSeq	CP001875	[62]	4302	707	—	14	83
<i>Apis mellifera</i>	Uniprot	AADG00000000	[63]	13514	12511	—	5	7,3
<i>Trichoderma harzianum</i> [*]	Uniprot	MRYK00000000	[64]	11480	7704	—	3	32
<i>Bacillus subtilis</i> ^{**}	Uniprot	NC_000964	[65]	26433	1299	—	301	93
<i>Homo sapiens</i>	Uniprot	CM000663	[66]	63487	1338	—	1071	96

—No proteins in this category

* Fungi

** Bacteria

¹ *A. flavus* NRRL3357

² *F. oxysporum* Fo5176

³ *F. oxysporum* 4287

⁴ *A. flavus* AF70.

#Ptn., total number of proteins; Unch., uncharacterized proteins; Hyp., hypothetical proteins; Put., putative proteins; Annot.%, annotation percentage

<https://doi.org/10.1371/journal.pone.0197511.t001>

score separates enzymes with different tertiary structures [34]. Results were parsed to obtain, for each enzymatic activity as defined by their Enzyme Commission (EC) number, files containing one or more groups of primary structures. If for a given enzymatic activity, only one group was produced at the end of the clusterization step, then all sequences would be considered homologous, and that enzymatic activity was removed from the analysis. On the other hand, if more than one group was produced, then sequences in the same group were considered homologous, with a score above 120, while sequences allocated in different groups were considered analogous (potential NISEs), with a score smaller than 120. In other words, sequences allocated in the same group have similar tertiary structures, while sequences allocated in different groups have different folding patterns, which reflects their different evolutionary origins [34, 35, 68].

Protein function inference

The groups of homologous sequences generated after the clustering step using the KEGG dataset were used for reannotation (with the pipeline AnEnPi) of the predicted proteins from the organisms in this study, which were compared, in a pairwise manner, to each primary protein structure within each protein functional group from KEGG. For the biochemical function inference, a cutoff value of 10^{-20} was used, a highly restrictive value that gives greater reliability to the results [67, 69–71]. Sequences with scores below this threshold were removed from the analysis.

NISEs: Identification, structural validation and essentiality

The search for cases of analogy (NISEs) between enzymes from plants and pathogens was performed through the analysis of the groups produced after the clustering step and functional inference. For this, one of the modules of AnEnPi was used together with in-house scripts to parse and filter the results. To validate the identified NISEs, that is, to verify if the enzymes found are cases of evolutionary convergence, we classified the sequences in accordance with their folds using the SUPERFAMILY database. The information in this database is based on a collection of Hidden Markov Models [72], which represent the structural domains of proteins classified by SCOP [73].

Heteromultimeric enzymes, enzymes annotated with the term "subunit" and sequences without an associated fold were excluded from the final list. Fused domains were maintained in our analysis, as in the case of the family "Dimeric alpha + beta barrel", which is an evolutionarily conserved group of protein families [73, 74]. Enzymes with the same EC number, but displaying different folds and, consequently, belonging to different superfamilies, were considered potential NISEs.

The Database of Essential Genes (DEG, 14.7, October/2016, <http://www.essentialgene.org/>) was used as a reference for the search for essential activities in the pathogens studied. A BLASTp search was performed between all enzymatic sequences identified as analogous against the DEG database. An e-value of 10^{-5} was used as threshold. Later, another BLASTp search was performed between all enzymatic sequences identified as analogues against the predicted proteins of organisms that should not be affected by an eventual inhibitor for the target identified in phytopathogen (*H. sapiens*, *A. mellifera*, *T. harzianum* and *B. subtilis*). An e-value of 10^{-5} was used as threshold.

Results

Data preparation, clustering and functional activity inference

After cleaning and preparation, the initial dataset obtained from KEGG was reduced to 1,225,682 protein sequences distributed over 3,893 enzymatic activities. After clusterization,

this dataset was used for the reannotation of the predicted proteins of the plants and phytopathogens, comprising 444198 individual sequences in 2096 enzymatic activities from the three plants and their 15 pathogens. Predicted proteins from *H. sapiens*, *A. mellifera*, *T. harzianum* and *B. subtilis* were also reannotated, comprising 114914 individual sequences in 2008 enzymatic activities. Annotation quality of the downloaded sets of predicted proteins varied greatly. Before the reannotation procedure, the best annotated organism among the plants was *Z. mays*, with approximately 90% of their proteins characterized, while *S. lycopersicum* presented only 9% of its proteins annotated. Among the pathogens, *P. ananatis* presented 83% of its entire conceptual proteome annotated and *M. perniciosa* had only 1.3% of its proteins characterized. After the functional inference step, where only enzymes were reannotated, on average 15% of the proteins of each organism were associated with an enzymatic activity (data not shown).

Potential NISEs: Identification and validation

Initially, a total of 568 cases of potential NISEs was identified, and from this set 97 cases were validated (Table 2, see S1 Table for more details). Sequences labeled with "subunit" or "chain" (324 cases), enzymes displaying the same fold (55 cases), and sequences without an associated fold in the SUPERFAMILY database (92 cases) were excluded. Cases of analogy were validated for all the pathogens studied: only one case was found for *P. sojae* and *S. sclerotiorum*, while 14 cases were found for *A. flavus* AF70. In total, 13 cases of analogy were found in the comparisons between *G. max* and its pathogens, 23 cases between *S. lycopersicum* and its pathogens, and 61 cases between *Z. mays* and its pathogens (Table 2).

The validated NISEs (97 cases), comprising 39 different enzymatic activities, participate in central metabolic pathways including the carbohydrate metabolism (13 enzymatic activities),

Table 2. Number of potential, validated, specific and essential NISEs. Numbers in parenthesis indicate the number of enzymatic activities identified.

Host	Pathogens	Potential NISEs	Validated	Specific*	Essential
<i>G. max</i>	<i>A. flavus</i> ¹	25	4	3	2
	<i>F. oxysporum</i> ²	21	4	4	1
	<i>P. sojae</i>	25	1	1	1
	<i>S. sclerotiorum</i>	21	1	1	0
	<i>X. axonopodis</i>	12	3	2	2
<i>S. lycopersicum</i>	<i>B. cinerea</i>	18	3	2	1
	<i>F. oxysporum</i> ³	30	6	5	2
	<i>M. perniciosa</i>	23	4	2	2
	<i>P. syringae</i>	38	5	4	5
	<i>R. solanacearum</i>	32	5	5	5
<i>Z. mays</i>	<i>A. flavus</i> ⁴	64	14	8	7
	<i>C. graminicola</i>	69	13	7	9
	<i>E. turcicum</i>	62	12	7	9
	<i>G. moniliformis</i>	65	10	6	5
	<i>P. ananatis</i>	63	12	11	7
Total		568	97 (39)	68 (26)	58 (29)

* Number of pathogen's specific tertiary structures

¹ *A. flavus* NRRL3357

² *F. oxysporum* Fo5176

³ *F. oxysporum* 4287

⁴ *A. flavus* AF70.

amino acid metabolism (8), energy metabolism (6), biosynthesis of secondary metabolites (4) and lipid metabolism (4). Eight enzymatic activities belong to other pathways such as xenobiotics degradation, metabolism of cofactors and vitamins, nucleotide metabolism and metabolism of other amino acids (Fig 2). It is important to remember that one enzymatic activity may participate in more than one pathway.

Essential NISEs

After the validation step a screening for essential enzymes was performed, revealing 58 cases of analogy (Table 3), involving 29 different essential enzymatic activities, corresponding to 119 different structures, for all organisms analyzed in this study. In the carbohydrate metabolism, the most frequent case was catalase, classified as essential for three pathogens of *G. max* (*A. flavus*, *F. oxysporum* and *P. sojae*), three pathogens of *S. lycopersicum* (*F. oxysporum*, *P. seryngae* and *R. solanacearum*) and three pathogens of *Z. mays* (*A. flavus*, *E. turcicum* and *C. graminicola*). Members of the pentoses pathway, like ribose 5-phosphate isomerase, ribulose-phosphate 3-epimerase and glyoxalase I, were identified in three *Z. mays*' pathogens (*A. flavus*, *G. moniliformis* and *C. graminicola*). Another frequent case, the enzyme cyclin-dependent kinase, was found for four of the five pathogens of *Z. mays* (*A. flavus*, *E. turcicum*, *C. graminicola* and *G. moniliformis*).

In the amino acid metabolism, several enzymes were identified as essential and analogous, like carbonic anhydrase for *R. solanacearum* and *A. flavus* AF70; prolyl aminopeptidase, for *F. oxysporum* 4287; transaminase, for *A. flavus* AF70, *G. moniliformis*, *A. flavus* NRRL3357 and *F. oxysporum* Fo5176. Chitinases were found as essential and analogous for *P. seryngae* and *R. solanacearum* (Table 3).

Analogous and essential enzymes were also found in the metabolism of lipids and biosynthesis of secondary metabolites pathways. Acetyl-CoA carboxylase was identified in *X. axonopodis* and phospholipase A2 in *C. graminicola*. Ornithine carbamoyltransferase, identified in *P. ananatis*, participates in the amino acid metabolism (S2 Table). Some enzymatic activities found to be essential for some pathogens have not been identified as essential in others: these cases are represented by enzymes encoded by different genes. In this group we can cite enzymes belonging to the antioxidant system (AS), composed of enzymes involved with the

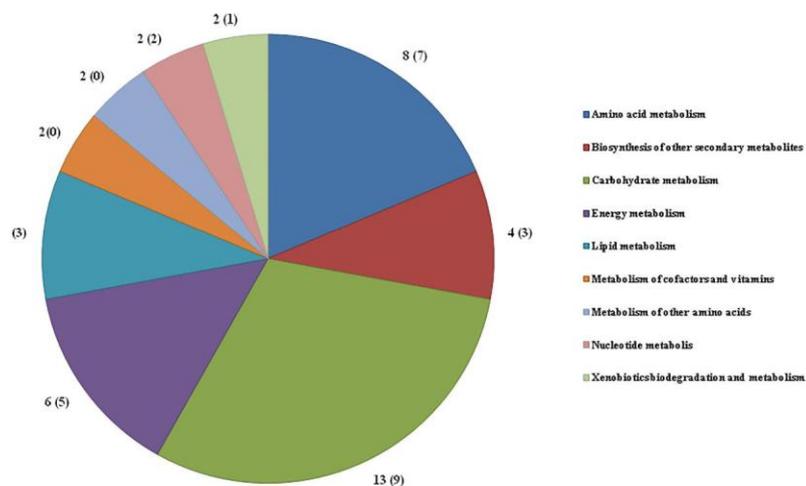


Fig 2. Functional classification of the validated NISEs. Numbers in parenthesis indicate the amount of essential enzymatic activities.

<https://doi.org/10.1371/journal.pone.0197511.g002>

Table 3. Essential and analogous enzymes.

Hosts	ID Sequence Host	Pathogens	NISEs			Essentiality data	
			ID sequence pathogens	EC number	Enzyme	ID DEG**	E-value
<i>G. max</i>	NP_001235974.1	<i>A. flavus</i>	XP_002384918.1	1.11.1.6*	Catalase	DEG10110209	2,00E-068
<i>G. max</i>	XP_003557098.2	<i>A. flavus</i>	XP_002377297.1	1.11.1.7*	Peroxidase	—	—
<i>G. max</i>	XP_006600684.1	<i>A. flavus</i>	XP_002376298.1	1.2.1.3	Aldehyde dehydrogenase (NAD+)	DEG20180006	1,00E-065
<i>G. max</i>	XP_006600243.1	<i>A. flavus</i>	XP_002382374.1	2.6.1.1	Aspartate transaminase	—	—
<i>G. max</i>	NP_001235974.1	<i>F. oxysporum</i>	9FP11 F9FP11_FUSOF	1.11.1.6*	Catalase	DEG10110209	0
<i>G. max</i>	XP_00355725.2	<i>F. oxysporum</i>	F9FYF1_FUSOF	1.15.1.1*	Superoxide dismutase	—	—
<i>G. max</i>	XP_006600243.1	<i>F. oxysporum</i>	F9G466_FUSOF	2.6.1.1	Aspartate transaminase	—	—
<i>G. max</i>	XP_006598804.1	<i>F. oxysporum</i>	F9G2J4_FUSOF	4.4.1.5	Lactoylglutathione lyase	—	—
<i>G. max</i>	NP_001235974.1	<i>P. sojae</i>	XP_009521283.1	1.11.1.6*	Catalase	DEG10110209	8,00E-115
<i>G. max</i>	XP_003557098.2	<i>S. sclerotiorum</i>	XP_001585507.1	1.11.1.7*	Peroxidase	—	—
<i>G. max</i>	NP_001235974.1	<i>X. axonopodis</i>	WP_042823856.1	1.11.1.6*	Catalase	—	—
<i>G. max</i>	XP_006605648.1	<i>X. axonopodis</i>	WP_054320474.1	1.15.1.1*	Superoxide dismutase	DEG20241649	6,00E-018
<i>G. max</i>	XP_006601861.1	<i>X. axonopodis</i>	WP_033483073.1	6.4.1.2	Acetyl-CoA carboxylase	DEG10030125	4,00E-057
<i>S. lycopersicum</i>	K4CN29_SOLLC	<i>B. cinerea</i>	XP_001560519.1	3.1.3.2	Acid phosphatase	—	—
<i>S. lycopersicum</i>	LGUL_SOLLC	<i>B. cinerea</i>	XP_001550649.1	4.4.1.5	Lactoylglutathione lyase	—	—
<i>S. lycopersicum</i>	P21568 CYPH_SOLLC	<i>B. cinerea</i>	XP_001545186.1	5.2.1.8	Peptidylprolyl isomerase	DEG20241291	1,00E-046
<i>S. lycopersicum</i>	K4BVX3_SOLLC	<i>F. oxysporum</i>	A0A0D2YKD1_FUSO4	1.11.1.6*	Catalase	DEG10110209	0
<i>S. lycopersicum</i>	Q7XAV2_SOLLC	<i>F. oxysporum</i>	A0A0D2YE80_FUSO4	1.15.1.1*	Superoxide dismutase	—	—
<i>S. lycopersicum</i>	K4CN29_SOLLC	<i>F. oxysporum</i>	A0A0D2YGA3_FUSO4	3.1.3.2	Acid phosphatase	—	—
<i>S. lycopersicum</i>	Q42875_SOLLC	<i>F. oxysporum</i>	A0A0D2XJE6_FUSO4	3.2.1.4	Cellulase	—	—
<i>S. lycopersicum</i>	Q8GZD8_SOLLC	<i>F. oxysporum</i>	A0A0D2XCV3_FUSO4	3.4.11.5	Prolyl aminopeptidase	DEG20210010	7,00E-014
<i>S. lycopersicum</i>	LGUL_SOLLC	<i>F. oxysporum</i>	A0A0D2XLV4_FUSO4	4.4.1.5	Lactoylglutathione lyase	—	—
<i>S. lycopersicum</i>	P15003 PER1_SOLLC	<i>M. perniciosa</i>	E2LX62_MONPE	1.11.1.7*	Peroxidase	—	—
<i>S. lycopersicum</i>	Q9FVN0 AMT13_SOLLC	<i>M. perniciosa</i>	E2M162_MONPE	2.7.13.3	Histidine-kinase	DEG20070330	4,00E-036
<i>S. lycopersicum</i>	Q8GZD8_SOLLC	<i>M. perniciosa</i>	E2LYM3_MONPE	3.4.11.1	Leucyl aminopeptidase	—	—
<i>S. lycopersicum</i>	K4CJ01_SOLLC	<i>M. perniciosa</i>	E2LAS1_MONPE	5.4.2.8	Phosphomannomutase	DEG20020210	5,00E-030
<i>S. lycopersicum</i>	K4BVX3_SOLLC	<i>P. seryngae</i>	NP_794283.1	1.11.1.6*	Catalase	DEG10270348	0
<i>S. lycopersicum</i>	P15003 PER1_SOLLC	<i>P. seryngae</i>	NP_794565.1	1.11.1.7*	Peroxidase	DEG10180459	4,00E-010
<i>S. lycopersicum</i>	K4CN29_SOLLC	<i>P. seryngae</i>	NP_791387.1	3.1.3.2	Acid phosphatase	DEG10290292	1,00E-084
<i>S. lycopersicum</i>	Q05539 CHIA_SOLLC	<i>P. seryngae</i>	NP_794777.1	3.2.1.14	Chitinase	DEG10250423	5,00E-019
<i>S. lycopersicum</i>	P21568 CYPH_SOLLC	<i>P. seryngae</i>	NP_791005.1	5.2.1.8	Peptidylprolyl isomerase	DEG10470303	2,00E-059
<i>S. lycopersicum</i>	K4BVX3_SOLLC	<i>R. solanacearum</i>	AGH83314.1	1.11.1.6*	Catalase	DEG10270348	0
<i>S. lycopersicum</i>	P15003 PER1_SOLLC	<i>R. solanacearum</i>	AGH86619.1	1.11.1.7*	Peroxidase	DEG10350205	2,00E-008
<i>S. lycopersicum</i>	Q9FVN0 AMT13_SOLLC	<i>R. solanacearum</i>	AGH84344.1	2.7.13.3	Histidine kinase	DEG10330275	1,00E-065
<i>S. lycopersicum</i>	Q05539 CHIA_SOLLC	<i>R. solanacearum</i>	AGH83721.1	3.2.1.14	Chitinase	DEG10260021	1,00E-017
<i>S. lycopersicum</i>	K4C2F1_SOLLC	<i>R. solanacearum</i>	AGH86735.1	4.2.1.1	Carbonic anhydrase	DEG10050308	4,00E-038
<i>Z. mays</i>	NP_001304298.1	<i>A. flavus</i>	B8NGN0_ASPFN	1.10.2.2	Quinol-cytochrome-c reductase	DEG20091193	1,00E-054
<i>Z. mays</i>	XP_008660914.1	<i>A. flavus</i>	B8NX24_ASPFN	1.11.1.6*	Catalase	DEG10110209	2,00E-068
<i>Z. mays</i>	XP_008664058.1	<i>A. flavus</i>	B8NC39_ASPFN	1.11.1.7*	Peroxidase	—	—
<i>Z. mays</i>	NP_001145525.1	<i>A. flavus</i>	B8N164_ASPFN	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008664254.1	<i>A. flavus</i>	B8NB79_ASPFN	2.1.1.43	Histone-lysine N-methyltransferase	DEG20051547	7,00E-012
<i>Z. mays</i>	XP_008665261.1	<i>A. flavus</i>	B8N9N8_ASPFN	2.5.1.18	Glutathione transferase	—	—

(Continued)

Table 3. (Continued)

Hosts	NISEs					Essentiality data	
	ID Sequence Host	Pathogens	ID sequence pathogens	EC number	Enzyme	ID DEG**	E-value
<i>Z. mays</i>	XP_008660232.1	<i>A. flavus</i>	B8NQM9_ASPFN	2.6.1.1	Aspartate transaminase	—	—
<i>Z. mays</i>	XP_008663534.1	<i>A. flavus</i>	B8N9A7_ASPFN	2.7.11.22	Cyclin-dependent kinase	DEG20010254	6,00E-067
<i>Z. mays</i>	XP_008664470.1	<i>A. flavus</i>	B8NB93_ASPFN	3.1.3.2	Acid phosphatase	—	—
<i>Z. mays</i>	XP_008656307.1	<i>A. flavus</i>	B8NQT3_ASPFN	3.2.2.22	rRNA N-glycosylase	—	—
<i>Z. mays</i>	XP_008655471.1	<i>A. flavus</i>	B8NWM8_ASPFN	4.2.1.1	Carbonic anhydrase	DEG20101870	2,00E-011
<i>Z. mays</i>	NP_001148888.1	<i>A. flavus</i>	B8NT23_ASPFN	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001149850.1	<i>A. flavus</i>	B8N7U5_ASPFN	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	6,00E-110
<i>Z. mays</i>	X P_008644870.1	<i>A. flavus</i>	B8NFW5_ASPFN	5.3.1.6	Ribose-5-phosphate isomerase	DEG10140248	3,00E-012
<i>Z. mays</i>	XP_008657765.1	<i>E. turcicum</i>	XP_008026270.1	1.1.1.27	L-lactate dehydrogenase	DEG20010346	1,00E-086
<i>Z. mays</i>	NP_001105310.2	<i>E. turcicum</i>	XP_008029291.1	1.11.1.6*	Catalase	DEG10110209	0
<i>Z. mays</i>	XP_008664058.1	<i>E. turcicum</i>	XP_008030871.1	1.11.1.7*	Peroxidase	DEG10400636	4,00E-080
<i>Z. mays</i>	NP_001145525.1	<i>E. turcicum</i>	XP_008025877.1	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008664254.1	<i>E. turcicum</i>	XP_008025860.1	2.1.1.43	Histone-lysine N-methyltransferase	DEG20240496	3,00E-018
<i>Z. mays</i>	XP_008663534.1	<i>E. turcicum</i>	XP_008024068.1	2.7.11.22	Cyclin-dependent kinase	DEG20090883	2,00E-041
<i>Z. mays</i>	XP_008651541.1	<i>E. turcicum</i>	XP_008029497.1	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008664470.1	<i>E. turcicum</i>	XP_008024834.1	3.1.3.2	Acid phosphatase	DEG10390008	1,00E-063
<i>Z. mays</i>	NP_001148888.1	<i>E. turcicum</i>	XP_008026072.1	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001136955.1	<i>E. turcicum</i>	XP_008024266.1	4.6.1.1	Adenylate cyclase	DEG10030767	2,00E-010
<i>Z. mays</i>	NP_001149850.1	<i>E. turcicum</i>	XP_008028934.1	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	1,00E-108
<i>Z. mays</i>	XP_008644870.1	<i>E. turcicum</i>	XP_008028444.1	5.3.1.6	Ribose-5-phosphate isomerase	DEG10080091	8,00E-015
<i>Z. mays</i>	XP_008657765.1	<i>C. graminicola</i>	XP_008097388.1	1.1.1.27	L-lactate dehydrogenase	DEG20010346	1,00E-091
<i>Z. mays</i>	XP_008660914.1	<i>C. graminicola</i>	XP_008098502.1	1.11.1.6*	Catalase	DEG10110209	0
<i>Z. mays</i>	XP_008664058.1	<i>C. graminicola</i>	XP_008095952.1	1.11.1.7*	Peroxidase	DEG10400636	3,00E-079
<i>Z. mays</i>	NP_001145525.1	<i>C. graminicola</i>	XP_008093145.1	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008663534.1	<i>C. graminicola</i>	XP_008094831.1	2.7.11.22	Cyclin-dependent kinase	DEG20010254	6,00E-050
<i>Z. mays</i>	XP_008651541.1	<i>C. graminicola</i>	XP_008100128.1	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008675577.1	<i>C. graminicola</i>	XP_008100081.1	3.1.1.4	Phospholipase A2	DEG20240063	2,00E-026
<i>Z. mays</i>	XP_008664470.1	<i>C. graminicola</i>	XP_008094949.1	3.1.3.2	Acid phosphatase	—	—
<i>Z. mays</i>	XP_008658269.1	<i>C. graminicola</i>	XP_008092609.1	3.1.13.4	Poly(A)-specific ribonuclease	DEG20240339	7,00E-092
<i>Z. mays</i>	XP_008677367.1	<i>C. graminicola</i>	XP_008097450.1	3.1.3.3	Phosphoserine phosphatase	DEG20211963	6,00E-052
<i>Z. mays</i>	NP_001148888.1	<i>C. graminicola</i>	XP_008096879.1	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001149850.1	<i>C. graminicola</i>	XP_008091175.1	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	8,00E-113
<i>Z. mays</i>	XP_008644870.1	<i>C. graminicola</i>	XP_008098210.1	5.3.1.6	Ribose-5-phosphate isomerase	DEG10080091	1,00E-015
<i>Z. mays</i>	NP_001145525.1	<i>G. moniliformis</i>	W7LPB7_GIBM7	1.11.1.15*	Peroxiredoxin	—	—
<i>Z. mays</i>	XP_008660232.1	<i>G. moniliformis</i>	W7MC41_GIBM7	2.6.1.1	Aspartate transaminase	—	—
<i>Z. mays</i>	XP_008663534.1	<i>G. moniliformis</i>	W7MSL6_GIBM7	2.7.11.22	Cyclin-dependent kinase	DEG20011066	2,00E-036
<i>Z. mays</i>	XP_008651541.1	<i>G. moniliformis</i>	W7M0K8_GIBM7	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008658269.1	<i>G. moniliformis</i>	W7M4G2_GIBM7	3.1.13.4	Poly(A)-specific ribonuclease	DEG20240339	1,00E-088
<i>Z. mays</i>	XP_008664470.1	<i>G. moniliformis</i>	W7NDR6_GIBM7	3.1.3.2	Acid phosphatase	DEG10390008	2,00E-013
<i>Z. mays</i>	XP_008655784.1	<i>G. moniliformis</i>	W7M5R3_GIBM7	3.2.1.4	Cellulase	—	—
<i>Z. mays</i>	NP_001148888.1	<i>G. moniliformis</i>	W7LNQ2_GIBM7	4.4.1.5	Lactoylglutathione lyase	—	—
<i>Z. mays</i>	NP_001136955.1	<i>G. moniliformis</i>	W7MFF7_GIBM7	4.6.1.1	Adenylate cyclase	DEG20090256	1,00E-090
<i>Z. mays</i>	NP_001149850.1	<i>G. moniliformis</i>	W7M917_GIBM7	5.1.3.1	Ribulose-phosphate 3-epimerase	DEG20210336	1,00E-107
<i>Z. mays</i>	NP_001105310.2	<i>P. ananatis</i>	D4GMF4_PANAM	1.11.1.6*	Catalase	—	—
<i>Z. mays</i>	XP_008667406.1	<i>P. ananatis</i>	D4GL47_PANAM	1.11.1.15*	Peroxiredoxin	DEG10030767	1,00E-006

(Continued)

Table 3. (Continued)

Hosts	ID Sequence Host	Pathogens	NISEs			Essentiality data	
			ID sequence pathogens	EC number	Enzyme	ID DEG**	E-value
<i>Z. mays</i>	XP_008672910.1	<i>P. ananatis</i>	D4GCI2_PANAM	1.16.3.1*	Ferroxidase	—	—
<i>Z. mays</i>	XP_008660532.1	<i>P. ananatis</i>	D4GJ68_PANAM	2.1.3.3	Ornithine carbamoyltransferase	DEG10350142	9,00E-055
<i>Z. mays</i>	XP_008657589.1	<i>P. ananatis</i>	D4GHCS_PANAM	2.3.1.51	1-acylglycerol-3-phosphate O-acyltransferase	DEG10480294	2,00E-093
<i>Z. mays</i>	XP_008656415.1	<i>P. ananatis</i>	D4GHA1_PANAM	2.7.2.3	Phosphoglycerate kinase	—	—
<i>Z. mays</i>	XP_008662013.1	<i>P. ananatis</i>	D4GMM0_PANAM	2.7.4.8	Guanylate kinase	DEG10030351	9,00E-064
<i>Z. mays</i>	XP_008672924.1	<i>P. ananatis</i>	D4GGT2_PANAM	3.1.1.5	Lysophospholipase	—	—
<i>Z. mays</i>	XP_008651541.1	<i>P. ananatis</i>	D4GFB8_PANAM	3.1.1.31	6-phosphogluconolactonase	—	—
<i>Z. mays</i>	XP_008650400.1	<i>P. ananatis</i>	D4GCE1_PANAM	3.1.3.11	Fructose-bisphosphatase	DEG10480226	2,00E-090
<i>Z. mays</i>	XP_008672875.1	<i>P. ananatis</i>	D4GMQ4_PANAM	4.2.1.96	4a-hydroxytetrahydrobiopterin dehydratase	DEG10470424	3,00E-034
<i>Z. mays</i>	NP_001105425.1	<i>P. ananatis</i>	D4GK89_PANAM	4.3.3.7	4-hydroxy-tetrahydrodipicolinate synthase	DEG10180422	1,00E-020

*Enzymes of the antioxidant system.

** Accession number in DEG.

<https://doi.org/10.1371/journal.pone.0197511.t003>

detoxification of reactive oxygen species (ROS) such as catalase, peroxidase, superoxide dismutase, peroxiredoxin, among others.

Analogous enzymes in the antioxidant system

One group of enzymes that stood out among the validated NISEs, including non-essential activities, were the enzymes that comprise the antioxidant system (AS). In all comparisons made between plants and their pathogens, except in the case of *B. cinerea*, for at least one of the functional activities of the antioxidant system, the host enzyme and its counterpart in the pathogen are structurally different (Table 4). In total, 27 cases of analogy were found for the antioxidant system, including catalase (CAT), peroxidase (POX), superoxide dismutase (SOD), ferroxidase (HEPH) and peroxiredoxin (PRDX). In our results, CAT was identified as an essential enzyme for 9 of the 14 pathogens studied, and POX was identified as essential in *E. turcicum*, *C. graminicola*, *P. seryngae* and *R. solanacearum*. SOD was identified as an essential enzyme for *X. axonopodis*. Among the pathogens analyzed, there are two species with distinct strains, *A. flavus* (NRRL3357, AF70) and *F. oxysporum* (Fo5176, 4287). No differences were observed between different lineages as in the case of *A. flavus* and *F. oxysporum*. It is important to emphasize that the AS enzymatic activities are present in all the genomes included in the present work; however, only the cases of validated NISEs have been shown, which explain gaps in the absence/presence pattern observed for HEPH, PRDX and SOD (Table 4).

Specific structural forms

After obtaining the final list of validated, essential NISEs between the plant hosts and their pathogens, a search for these enzymatic activities was performed on the predicted proteins of *H. sapiens*, *A. mellifera*, *B. subtilis* and *T. harzianum*. The objective of this comparison was to find specific structural enzymatic forms of the pathogen in the genomes of species that should not be affected by an eventual inhibitor targeting that particular structural form, mainly *H.*

Table 4. Alternative enzymatic forms found among the enzymes of the antioxidant system.

Organisms	Structural forms																		
	CAT				POX				SOD				HEPH		PRDX				
<i>G. max</i>	①*			⑤		③	⑥			⑳	①	④	⑥	⑦					
<i>A. flavus</i> ¹		②						⑫											
<i>F. oxysporum</i> ²		②													⑭				
<i>P. sojae</i>		②																	
<i>S. sclerotiorum</i>								⑫											
<i>X. axonopodis</i>			③																
<i>S. lycopersicum</i>	①			⑤		③	⑥				①	④	⑥	⑦					
<i>B. cinerea</i>																			
<i>F. oysporum</i> ³		②													⑭				
<i>M. perniciosa</i>								⑫											
<i>P. syringae</i>		②							⑮										
<i>R. solanacearum</i>					⑧					⑮									
<i>Z. mays</i>	①			⑤		③										②	⑥	①	⑩
<i>A. flavus</i> ⁴		②						⑫											⑨
<i>C. graminicola</i>		②					⑥												⑨
<i>E. turcicum</i>		②					⑥												⑨
<i>G. moniliformis</i>																			⑨
<i>P. ananatis</i>			③														⑦	②	

¹ *A. flavus* NRRL3357

² *F. oxysporum* Fo5176

³ *F. oxysporum* 4287

⁴ *A. flavus* AF70.

*Numbers represent the groups where a sequence was located. Only validated cases of analogy are shown. Black circles indicate structural forms validated found only on the pathogen.

<https://doi.org/10.1371/journal.pone.0197511.t004>

sapiens and *A. mellifera*. Of the 97 NISEs validated, 68 specific structural forms of the pathogen (in relation to the plant host, men and bee) were found (Table 5). They are distributed over 26 enzymatic activities (16 of them being essential). From these 68 structural forms, 39 were present in *T. harzianum* and 17 in *B. subtilis*, which is expected since these organisms belong to the same kingdoms of the phytopathogens studied in this work (Fungi and Bacteria).

Discussion

The correct description of the analogous enzymes is important for the practical tasks of metabolic reconstruction and enzymatic nomenclature. In addition to this practical importance, these enzymes represent important evolutionary phenomenon, existence shows that for various biochemical problems, evolutionarily independent solutions may appear [35]. The main works on the practical application of analogous enzymes describes studies of metabolic pathways and inhibitory targets for human pathogens [42, 69–70]. In the case of our study, we sought a practical application, focused on the solution of an agronomic problem.

Essential enzymes are one of the primary targets for the development of inhibitors of any kind; however, species that share essential enzymatic functions may inadvertently be affected by products developed with other applications in mind [75]. Pesticides are commonly targeted at these functions, and their damaging effects on several species including man himself and several vital species such as pollinators and beneficial microorganisms are reason for great concern [76–78]. In fact, it is estimated that approximately 35% of the crops are dependent on

Table 5. Phytopathogen specific enzymatic structural forms.

Comparison				Structural forms					
Plant**	Pathogen**	EC Number	ID Sequence Pathogens	Pathogens	Plant	<i>H. sapiens</i>	<i>A. mellifera</i>	<i>T. harzianum</i>	<i>B. subtilis</i>
Gm	Af	1.11.1.6‡	XP_002384918.1	1 ^A , 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Gm	Af	1.11.1.7	XP_002377297.1	3, 6, 12	3, 6, 20	1, 3	1, 3, 6	3, 6, 12*	7
Gm	Af	2.6.1.1	XP_002382374.1	1, 5	1	1	1	1, 5*	1
Gm	Fo	1.11.1.6‡	F9FP11_FUSOF	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Gm	Fo	1.15.1.1‡	F9FYF1_FUSOF	1, 4, 7, 14	1, 4, 6, 7	1, 4, 7	1, 4, 7	1, 4, 7, 14*	1, 4
Gm	Fo	2.6.1.1	F9G466_FUSOF	1, 5	1	1	1	1, 5*	1
Gm	Fo	4.4.1.5	F9G2J4_FUSOF	1, 3	1, 8	1	1	1, 3*	1, 3, 6, 7, 11
Gm	Ps	1.11.1.6‡	XP_009521283.1	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Gm	Ss	1.11.1.7	XP_001585507.1	3, 6, 12	3, 6, 20	1, 3	1, 3, 6	3, 6, 12*	7
Gm	Xa	1.11.1.6	WP_042823856.1	3	1, 5	1, 5	1, 5	1, 2	1, 3*, 8
Gm	Xa	6.4.1.2‡	WP_033483073.1	1, 6	1	1	2	1	1, 6*
Sl	Bc	3.1.3.2	XP_001560519.1	2, 3, 7, 13	2, 6, 9, 11	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13*	—
Sl	Bc	4.4.1.5	XP_001550649.1	1, 3	1, 8	1	1	1, 3*	1, 3*
Sl	Fo	1.11.1.6‡	A0A0D2YKD1_FUSO4	1, 2, 5	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Sl	Fo	1.15.1.1	A0A0D2YE80_FUSO4	1, 4, 7, 14	1, 4, 6, 7	1, 4, 7	1, 4, 7	1, 4, 7, 14*	1, 4
Sl	Fo	3.1.3.2	A0A0D2YGA3_FUSO4	1, 2, 3, 4, 7, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13	—
Sl	Fo	3.2.1.4	A0A0D2XJE6_FUSO4	1, 6	1	—	1	1	1, 3
Sl	Fo	4.4.1.5	A0A0D2XLV4_FUSO4	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Sl	Mp	1.11.1.7	E2LX62_MONPE	6, 12	3, 6	1, 3	1, 3, 6	3, 6, 12*	7
Sl	Mp	3.4.11.1	E2LYM3_MONPE	1, 11	1	1	1	—	1
Sl	Psy	1.11.1.6‡	NP_794283.1	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Sl	Psy	1.11.1.7‡	NP_794565.1	6, 16, 18, 19	3, 6	1, 3	1, 3, 6	3, 6, 12	7
Sl	Psy	3.1.3.2‡	NP_791387.1	1, 3	2, 6, 9, 11	2, 4, 7	2, 4, 5, 7, 20	1*, 3, 4, 5, 7, 13	—
Sl	Psy	3.2.1.14‡	NP_794777.1	1, 3	1	1, 10	1, 4	1	—
Sl	Rs	1.11.1.6‡	AGH83314.1	6	1, 5	1, 5	1, 5	1, 2	1, 3, 8
Sl	Rs	1.11.1.7‡	AGH86619.1	6, 18	3, 6	1, 3	1, 3, 6	3, 6, 12	7
Sl	Rs	2.7.13.3‡	AGH84344.1	1, 21, 23, 24, 33, 36	1, 20	2, 12, 13, 20	12, 20	1	1
Sl	Rs	3.2.1.14‡	AGH83721.1	3	1	1, 10	1, 4	1	—
Sl	Rs	4.2.1.1‡	AGH86735.1	1, 3, 13	1, 2, 5	1, 2	1, 2, 3	1, 2	1, 3, 5, 12
Zm	Af	1.11.1.15	B8N164_ASPFN	1, 9	1, 10	1	1	1, 9*	1
Zm	Af	1.11.1.6‡	B8NX24_ASPFN	1, 2	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Zm	Af	1.11.1.7	B8NC39_ASPFN	3, 6, 12	3	1, 3	1, 3, 6	3, 6, 12*	7
Zm	Af	2.6.1.1	B8NQM9_ASPFN	1, 5	1	1	1	1, 5*	1
Zm	Af	3.1.3.2	B8NB93_ASPFN	2, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13*	—
Zm	Af	3.2.2.22	B8NQT3_ASPFN	5	1, 7	—	—	—	—
Zm	Af	4.4.1.5	B8NT23_ASPFN	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Zm	Af	5.3.1.6‡	B8NFW5_ASPFN	1, 2	1	1	1	1, 2*	—
Zm	Cg	1.1.1.27‡	XP_008100733.1	2, 12	1, 12	1, 12	1, 12	1, 2*, 12	1, 11
Zm	Cg	1.11.1.15	XP_008093145.1	1, 9	1, 10	1	1	1, 9*	1, 9*
Zm	Cg	1.11.1.6‡	XP_008098502.1	1, 2, 5	1, 5	1, 5	1, 5	1, 2*	1, 3, 8
Zm	Cg	3.1.1.31	XP_008100128.1	1, 2	1	1, 4	1	1, 2*	2*
Zm	Cg	3.1.3.2	XP_008094949.1	2, 3, 5, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2, 3, 4, 5, 7, 13*	—
Zm	Cg	4.4.1.5	XP_008096879.1	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Zm	Cg	5.3.1.6‡	XP_008098210.1	1, 2	1	1	1	1, 2*	2*
Zm	Et	1.11.1.15	XP_008025877.1	1, 9	1, 10	1	1	1, 9*	1
Zm	Et	1.11.1.6‡	XP_008029291.1	1, 2, 5	1, 5	1, 5	1, 5	1, 2*	1, 3, 8

(Continued)

Table 5. (Continued)

Comparison				Structural forms					
Plant**	Pathogen**	EC Number	ID Sequence Pathogens	Pathogens	Plant	<i>H. sapiens</i>	<i>A. mellifera</i>	<i>T. harzianum</i>	<i>B. subtilis</i>
Zm	Et	3.1.1.31	XP_008029497.1	1, 2	1	1, 4	2	1, 2*	2*
Zm	Et	3.1.3.2‡	XP_008024834.1	1, 2, 3, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	1*, 3, 4, 5, 7, 13	—
Zm	Et	4.4.1.5	XP_008026072.1	1, 3	1, 8	1	1	1, 3*	1, 3*, 6, 7, 11
Zm	Et	4.6.1.1‡	XP_008024266.1	2, 8, 10	2, 17, 18	2, 8	2, 6, 8, 13	2, 8	4
Zm	Et	5.3.1.6‡	XP_008028444.1	1, 2	1	1	1	1, 2*	—
Zm	Gm	1.11.1.15	W7LPB7_GIBM7	1, 2, 9	1, 10	1	1	1, 9*	1
Zm	Gm	2.6.1.1	W7MC41_GIBM7	1, 5	1	1	1	1, 5*	1
Zm	Gm	3.1.1.31	W7M0K8_GIBM7	1, 2	1	1, 4	1	1	2*
Zm	Gm	3.1.3.2‡	W7NDR6_GIBM7	1, 2, 3, 7, 13	2, 4, 6, 9	2, 4, 7	2, 4, 5, 7, 20	2	—
Zm	Gm	3.2.1.4	W7M5R3_GIBM7	1, 6	1	—	1	1	1, 3
Zm	Gm	4.4.1.5	W7LNQ2_GIBM7	1, 3	1, 8	1	1	1	1, 3*, 6, 7, 11
Zm	Pa	1.11.1.15‡	D4GL47_PANAM	1, 2	1, 10	1	1	1	1
Zm	Pa	1.11.1.6	D4GMF4_PANAM	3, 5, 6	1, 5	1, 5	1, 5	1	1, 3*, 8
Zm	Pa	1.16.3.1	D4GCI2_PANAM	2, 7	2, 6	2, 4, 6	2, 6	6	1
Zm	Pa	2.1.3.3‡	D4GJ68_PANAM	2	1, 10	1, 10	10	1	1, 10
Zm	Pa	2.7.2.3	D4GHA1_PANAM	3	1	1	1	1	1, 3*, 9
Zm	Pa	2.7.4.8	D4GMM0_PANAM	1, 4, 7	1	1	1	1	1, 7
Zm	Pa	3.1.1.31	D4GFB8_PANAM	2, 6	1	1, 4	1	1	2*
Zm	Pa	3.1.1.5	D4GGT2_PANAM	2, 5	6, 7, 18	1, 4, 6, 7, 9, 10, 17, 18	1, 6, 7, 9, 17, 18	1	—
Zm	Pa	3.1.3.11‡	D4GCE1_PANAM	10, 12	1	1, 8	1	1	3, 11
Zm	Pa	4.2.1.96‡	D4GMQ4_PANAM	2	1	1	1	1	2*
Zm	Pa	4.3.3.7‡	D4GK89_PANAM	1, 2	1	—	—	—	1, 4

**Gm: *G. Max*, Af: *A. flavus*, Fo: *F. oxysporum*, Ps: *P. sojae*, Ss: *S. sclerotiorum*, Xa: *X. axonopodis*, Sl: *S. lycopersicum*, Rs: *R. solanacearum*, Psy: *P. syringae*, Mp: *M. perniciosa*, Bc: *B. cinerea*, Zm: *Z. mays*, Pa: *P. ananatis*, Gm: *G. moniliformis*, Et: *E. turcicum*, Gg: *C. graminicola*.

^ Numbers represent the different structures. Numbers in bold are the specific phytopathogen enzymatic structural forms.

‡ Essential enzymes.

— Enzymatic activity not found.

*Structural form homologous to the pathogen.

<https://doi.org/10.1371/journal.pone.0197511.t005>

pollinators for sexual reproduction, and pesticides are the main factor contributing to the current decrease of the pollinator population [44, 79].

Through the joint use of primary structure data, tertiary structure data and essentiality data, beginning with 444198 individual sequences, comprising 2096 enzymatic activities in 3 plants and 15 phytopathogens, we have disclosed a subset of analogous sequences in 29 essential enzymatic activities present both in the plant and the pathogen. These belong to several components of the central metabolism of plant and pathogens, being involved in the carbohydrate metabolism, the metabolism of amino acids, the detoxification of reactive oxygen species and others, thus offering several opportunities as targets.

Interestingly, the subset of non-essential NISEs contains several enzymes important in the context of host-pathogen interactions, such as cellulases, chitinases, glutathione transferase and lysophospholipase. Blocking or inhibiting these enzymes would, in principle, decrease virulence and / or delay the defense mechanisms of the pathogen [80, 81]. Inhibition of cellulases and chitinases has also been proposed as a strategy for the development of new antifungal drugs for aspergillosis in humans [22]. Glutathione transferase play an essential role in the protection of necrotrophic fungi against toxic metabolites derived from plants and reactive oxygen

species [82], while lysophospholipase has been implicated with virulence in *Cryptococcus neoformans* [83].

Some of the diversity found for the enzymes of the antioxidant system, both in terms of enzymatic activities and in structural forms, may be explained by evolutionary pressures: during the co-evolution between plants and their pathogens, it is likely that different antioxidant enzymes of plants have adapted to overcome the pathogen virulence mechanisms [84, 85]. The role of these enzymes in mechanisms of virulence, susceptibility to infections, development of drug targets and evaluation of pesticide effects has been studied for SOD [86–90], CAT [91–94] and POX [95].

Essential enzymes from the central metabolism have also been studied as potential drug targets in several organisms. Glucose-6-phosphate isomerase has been studied as a target for infections caused by *Plasmodium falciparum* [96], *Trypanosoma* spp [97], *Toxoplasma gondii* [98], and *Leishmania* spp [99], acetyl-CoA carboxylase for *L. major* [100, 101], and ribose 5-phosphate isomerase in other organisms [102]. Deletion of these genes usually results in a severe reduction in growth rates and virulence [103–105], and they have been studied as drug targets in other organisms [106–109].

Eighteen of the 29 enzymatic activities identified in this study as analogous and essential were identified in databases of drug targets such as TDR Drug Targets (<http://tdrtargets.org/>), DrugBank (<https://www.drugbank.ca/>) and Potential Drug Target Database (<http://www.dddc.ac.cn/pdtd/>), meaning they are being studied or employed as a drug target for at least one pathogen. Among them we can mention enzymes from the carbohydrate and amino acids metabolism such as lactoylglutathione lyase, acetyl-CoA carboxylase, carbonic anhydrase, and enzymes of the AS like catalase, peroxidase, peroxiredoxin and superoxide dismutase. Since these enzymatic activities present multiple tertiary structures, we are not able to tell, from this data, which one is under study; nonetheless, these findings give indirect support to our analyzes, corroborating the idea that essential enzymes with specific structural forms have great potential as drug targets as described in our study. Improvements in the annotation of genes and their products, and a better experimental characterization of enzymatic activities, would allow the use of less-stringent criteria in our procedures, mainly in data cleaning and filtering, but also in clustering and structural validation, increasing the number of essential and analogous enzymes that could be further studied as potential drug targets.

Conclusions

The approach employed in this study enabled the elaboration of lists of essential and analogous enzymes, most belonging to the central metabolism and/or involved in host-pathogen interactions, with potential to be a drug target. These enzymes provide an opportunity for the discovery of targets with considerable structural differences over their counterpart in beneficial organisms such as pollinators. Inclusion of structural data allows the disclosure of specific structural forms, facilitating the development of environment-friendly enzyme inhibitors, which may be of great importance for agricultural use.

Supporting information

S1 Table. Non-homologous isofunctional enzymes found in this study.
(XLS)

S2 Table. Distribution of metabolic pathways in essential analogous enzymes.
(XLS)

Acknowledgments

RAS recognizes CAPES (Coordination for the Improvement of Higher Education Personnel, Brazil) for supporting her with a scholarship during her DSc program. The authors also thank the staff of the Laboratory of Systems and Computational Biology in conducting this study and to Dr. Fábio Motta, Dr. Marcos Catanho and Dr. Monete Rajão for their help in the discussions.

Author Contributions

Conceptualization: Rangeline Azevedo da Silva, Leandro de Mattos Pereira, Antonio Basilio de Miranda.

Data curation: Rangeline Azevedo da Silva.

Formal analysis: Rangeline Azevedo da Silva, Leandro de Mattos Pereira, Melise Chaves Silveira.

Investigation: Rangeline Azevedo da Silva.

Methodology: Rangeline Azevedo da Silva, Rodrigo Jardim.

Project administration: Antonio Basilio de Miranda.

Software: Rangeline Azevedo da Silva.

Supervision: Antonio Basilio de Miranda.

Validation: Rangeline Azevedo da Silva.

Visualization: Rangeline Azevedo da Silva.

Writing – original draft: Rangeline Azevedo da Silva.

Writing – review & editing: Rangeline Azevedo da Silva, Leandro de Mattos Pereira, Melise Chaves Silveira, Rodrigo Jardim, Antonio Basilio de Miranda.

References

- Oerke E. C. Crop losses to pests. *J Agric Sci.* 2006; 144: 31–43.
- Sadras VO, Villalobos FJ, Fereres E. Limitations to Crop Productivity. In: Villalobos F, Fereres E. *Principles of Agronomy for Sustainable Agriculture.* Springer International Publishing; 2016.
- FAO International Code of Conduct on Pesticide Management—Guidance on Pest and Pesticide Management Policy Development. 2013. Available from: <http://www.fao.org>.
- Berry P. Pesticides and the intoxication of wild animals. *Journal of Veterinary Pharmacol Ther.* 2007; 30: 93–100.
- Asogwa EU, Dongo LN. Problems associated with pesticide usage and application in Nigerian cocoa production: A review. *Afr J Agric Res.* 2009; 4: 675–683.
- Nigatu AW, Brätveit M, Moen BE. Self-reported acute pesticide intoxications in Ethiopia. *BMC public health.* 2016; 16 (1): 1–8.
- Hancock DB, Martin ER, Mayhew GM, Stajich JM, Jewett R, Stacy MA, et al. Pesticide exposure and risk of Parkinson's disease: A family-based case-control study. *BMC Neurology.* 2008; 8: 6. <https://doi.org/10.1186/1471-2377-8-6> PMID: 18373838
- Spiewak R. Pesticides as a cause of occupational skin diseases in farmers. *Ann Agric Environ Med.* 2001; 8(1): 1–5. PMID: 11426918
- Ueker ME, Silva VM, Moi GP, Pignati WA, Mattos IE, Silva AGC. Parenteral exposure to pesticides and occurrence of congenital malformations: hospital-based case-control study. *BMC Pediatrics.* 2016; 16: 125. <https://doi.org/10.1186/s12887-016-0667-x> PMID: 27520287

10. Parrón T, Requena M, Hernández AF, Alarcón R. Environmental exposure to pesticides and cancer risk in multiple human organ systems. *Toxicol Lett.* 2014; 230(2): 157–65. <https://doi.org/10.1016/j.toxlet.2013.11.009> PMID: 24269242
11. Carneiro FF, Pignati W, Rigotto RM, Augusto LGS, Rizollo A, Muller NM, et al. Dossiê ABRASCO—Um alerta sobre os impactos dos agrotóxicos na saúde. Rio de Janeiro: ABRASCO; 2012. Available from: www.abrasco.org.br.
12. ANVISA. Agência Nacional de Vigilância Sanitária. Programa de Análise de Resíduos de Agrotóxicos em Alimentos (PARA). Relatório de Atividades de 2011 e 2012. Brasília: Agência Nacional de Vigilância Sanitária. 2013. Available from: <http://portal.anvisa.gov.br/documents>.
13. ABRASCO—Um alerta sobre os impactos dos agrotóxicos na saúde. Parte 2—Agrotóxicos, Saúde, Ambiente e Sustentabilidade. 2015. Available from: www.abrasco.org.br.
14. ABRASCO—Um alerta sobre os impactos dos agrotóxicos na saúde. EPSJV- Expressão Popular. 2015. Available from: www.abrasco.org.br.
15. Papin JA, Price ND, Wibakc S J, Fell DA, Palsson BO. Metabolic pathways in the post-genome era. *Trends Pharmacol Sci.* 2003; 28: 250–58.
16. Lehninger AL, Nelson DL, Cox MM. *Principles of Biochemistry In: Enzymes.* New York, 2008. pp. 191–225.
17. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* 2002; 12: 962–968. <https://doi.org/10.1101/gr.87702> PMID: 12045149
18. Silander OK, Ackermann M. The constancy of gene conservation across divergent bacterial orders. *BMC Res Notes.* 2009; 2: 2. <https://doi.org/10.1186/1756-0500-2-2> PMID: 19128452
19. Pancholi V, Chhatwal GS. Housekeeping enzymes as virulence factors for pathogens. *J Med Microbiol.* 2003; 293: 391–401.
20. Ouaiissi M, Ouaiissi A. Histone Deacetylase Enzymes as Potential Drug Targets in Cancer and Parasitic Diseases. *J Biomed Biotechnol.* 2006; 2006: 1–10.
21. Fu ZQ, Guo M, Jeong B-J, Tian F, Elthon TE, Cerny RL, et al. A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature.* 2007; 447: 284–289. <https://doi.org/10.1038/nature05737> PMID: 17450127
22. Schüttelkopf AW, Gros L, Blair DE, Frearson JA, van Aalten DM, Gilbert IH. Acetazolamide-based fungal chitinase inhibitors. *Bioorg Med Chem.* 2010; 18(23): 8334–40. <https://doi.org/10.1016/j.bmc.2010.09.062> PMID: 21044846
23. Wyatt PG, Gilbert IH, Read KD, Fairlamb AH. Target Validation: Linking Target and Chemical Properties to Desired Product Profile. *Curr Top Med Chem.* 2011; 11: 1275–83. <https://doi.org/10.2174/156802611795429185> PMID: 21401506
24. Kappes B, Tews I, Binter A, Macheroux P. PLP-dependent enzymes as potential drug targets for protozoan diseases. *Biochim Biophys Acta.* 2011; 1814(11): 567–76.
25. Vassar R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimers Res Ther.* 2014; 6: 89. <https://doi.org/10.1186/s13195-014-0089-7> PMID: 25621019
26. Katara P, Grover A, Sharma V. In silico prediction of drug targets in phytopathogenic *Pseudomonas syringae* pv. phaseolicola: charting a course for agrigenomics translation research. *OMICS.* 2012; 16(12): 700–6. <https://doi.org/10.1089/omi.2011.0141> PMID: 23215808
27. Hotson A, Chosed R, Shu H, Orth K, Mudgett MB. *Xanthomonas* type III effector XopD targets SUMO-conjugated proteins in planta. *Mol Microbiol.* 2003; 50(2): 377–89. PMID: 14617166
28. Sintchenko V, Roper MP. Pathogen genome bioinformatics. *Methods Mol Biol.* 2014; 1168: 173–193. https://doi.org/10.1007/978-1-4939-0847-9_10 PMID: 24870136
29. Fields FR, Lee SW, McConnell MJ. Using bacterial genomes and essential genes for the development of new antibiotics. *Biochem Pharmacol.* 2017; 134: 74–86. <https://doi.org/10.1016/j.bcp.2016.12.002> PMID: 27940263
30. IUBMB Nomenclature Commission. *Enzyme nomenclature.* Academic Press, San Diego, CA. 1992.
31. Gough J, Chothia C. SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res.* 2002; 30(1): 268–72. PMID: 11752312
32. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 2009; 37(1): 380–386.
33. Fitch WM. Distinguishing Homologous from Analogous Proteins. *Syst Biol.* 1970; 19(2): 99–113.

34. Galperin MY, Walker DR, Koonin EV. Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* 1998; 8(8): 779–90. PMID: [9724324](#)
35. Omelchenko MV, Galperin MY, Wolf YI, Koonin EV. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol Direct.* 2010; 5: 31. <https://doi.org/10.1186/1745-6150-5-31> PMID: [20433725](#)
36. Stallings WC, Powers TB, Patridge KA, Fee JA, Ludwig ML. Iron superoxide dismutase from *Escherichia coli* at 3.1-Å resolution: A structure unlike that of copper/zinc protein at both monomer and dimer levels. *Proc Natl Acad Sci USA.* 1983; 80: 3884–3888. PMID: [6346322](#)
37. Lobkovsky E, Moews PC, Liu H, Zhao H, Frere JM, Knox JR. Evolution of an enzyme activity: Crystallographic structure at 2-Å resolution of cephalosporinase from the ampC gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proc Natl Acad Sci USA.* 1993; 90(11): 11257–11261.
38. Carfi A, Pares S, Duee E, Galleni M, Duez C, Frere JM, et al. The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO Journal.* 1995; 14: 4914–4921. PMID: [7588620](#)
39. Buller AR; Townsend CA. Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc Natl Acad Sci USA* 2013; 110 (8): 653–61.
40. Koonin EV, Mushegian AR, Bork P. Non-orthologous gene displacement. *Trends Genet.* 1996; 12: 334–336. PMID: [8855656](#)
41. Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE. 2007. Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol.* 372:817–845. <https://doi.org/10.1016/j.jmb.2007.06.017> PMID: [17681532](#)
42. Piergiorgio RF, Miranda AB, Guimarães ACG, Catanho M. Functional Analogy in Human Metabolism: Enzymes with Different Biological Roles or Functional Redundancy? *Genome Biol Evol.* 2017; 9(6): 1624–1636. <https://doi.org/10.1093/gbe/evx119> PMID: [28854631](#)
43. Goulson D. Conserving wild bees for crop pollination. *J Food Agr Environ.* 2003; 1: 142–144.
44. Klein AM, Vaissiere JH, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, et al. Importance of pollinators in changing landscapes for world crops. *Proceedings Proc R Soc Lond [Biol].* 2007; 274: 303–313.
45. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature.* 2010; 463: 178–183. <https://doi.org/10.1038/nature08670> PMID: [20075913](#)
46. Nierman WC, Yu J, Fedorova-Abrams ND, Losada L, Cleveland TE, Bhatnagar D, et al. Genome Sequence of *Aspergillus flavus* NRRL 3357, a Strain That Causes Aflatoxin Contamination of Food and Feed. *Genome Announc.* 2015; 6(3): e00168–15.
47. Thatcher LF, Gardiner DM, Kazan K, Manners JM. A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Mol Plant Microbe Interact.* 2012; 25: 180–190. <https://doi.org/10.1094/MPMI-08-11-0212> PMID: [21942452](#)
48. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, et al. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science.* 2006; 313(5791):1261–1266. <https://doi.org/10.1126/science.1128796> PMID: [16946064](#)
49. Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, et al. Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics.* 2011; 7(8): e1002230. <https://doi.org/10.1371/journal.pgen.1002230> PMID: [21876677](#)
50. Kim JG1, Choi S, Oh J, Moon JS, Hwang I. Comparative analysis of three indigenous plasmids from *Xanthomonas axonopodis* pv. *glycines*. *Plasmid.* 2006; 56(2): 79–87. <https://doi.org/10.1016/j.plasmid.2006.03.001> PMID: [16697042](#)
51. Kahlau S, Aspinall S, Gray JC, Bock R. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J Mol Evol.* 2006; 63: 194–207. <https://doi.org/10.1007/s00239-005-0254-5> PMID: [16830097](#)
52. Staats M, van Kan JA. Genome update of *Botrytis cinerea* strains B05.10 and T4. *Eukaryot Cell.* 2012; 11: 1413–141 <https://doi.org/10.1128/EC.00164-12> PMID: [23104368](#)
53. Ma LJ, Does HCV, Borkovich KA, Coleman JJ, Daboussi MJ, Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 2010; 18: 464(7287): 367–373. <https://doi.org/10.1038/nature08850> PMID: [20237561](#)
54. Mondego JM, Carazzolle MF, Costa GG, Formighieri EF, Parizzi LP, Rincones J. A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao. *BMC Genomics.* 2008; 18(9): 548.

96. Barret MP. The Pentose Phosphate Pathway and Parasitic Protozoa. *Parasitol Today*. 1997; 13: 11–16. PMID: [15275160](#)
97. Verlinde CLMJ, Hannaert V, Blonski C, Willson M, Périé JJ, Fothergill-Gilmore LA, et al. Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resist Update*. 2001; 4: 1–14.
98. Tomavo S. The differential expression of multiple isoenzyme forms during stage conversion of *Toxoplasma gondii*: an adaptive developmental strategy. *Int J Parasitol*. 2001; 31: 1023–31. PMID: [11429165](#)
99. Barrett MP, Gilbert IH. Perspectives for new drugs against trypanosomiasis and leishmaniasis. *Curr Top Med Chem*. 2002; 2: 471–482. PMID: [11966468](#)
100. Tong L, Harwood HJJ. Acetyl-coenzyme A carboxylases: versatile targets for drug discovery. *J Cell Biochem*. 2006; 99(6): 1476–88. <https://doi.org/10.1002/jcb.21077> PMID: [16983687](#)
101. Chawla B, Madhubala R. Drug targets in *Leishmania*. *J Parasit Dis*. 2010; 34(1): 1–13. <https://doi.org/10.1007/s12639-010-0006-3> PMID: [21526026](#)
102. Zhang RG, Andersson CE, Savchenko A, Skarina T, Evdokimova E, Beasley S. Structure of *Escherichia coli* Ribose-5-Phosphate Isomerase: A Ubiquitous Enzyme of the Pentose Phosphate Pathway and the Calvin Cycle. *Structure*. 2003; 11(1): 31–42. PMID: [12517338](#)
103. Sørensen KI, Hove-Jensen B. Ribose catabolism of *Escherichia coli*: characterization of the rpiB gene encoding ribose phosphate isomerase B and of the rpiR gene, which is involved in regulation of rpiB expression. *J Bacteriol*. 1996; 178(4):1003–11. PMID: [8576032](#)
104. Loureiro I, Faria J, Clayton C, Macedo-Ribeiro S, Santarem N, Roy N, et al. Ribose 5-phosphate isomerase B knockdown compromises *Trypanosoma brucei* blood stream form infectivity. *PLoS Negl Trop Dis*. 2015; 9(1): e3430. <https://doi.org/10.1371/journal.pntd.0003430> PMID: [25568941](#)
105. Kaur PK, Tripathi N, Desale J, Neelagiri S, Yadav S, Bharatam PS, et al. Mutational and Structural Analysis of Conserved Residues in Ribose-5-Phosphate Isomerase B from *Leishmania donovani*: Role in Substrate Recognition and Conformational Stability. *PLoS ONE*. 2016; 11(3):e0150764. <https://doi.org/10.1371/journal.pone.0150764> PMID: [26953696](#)
106. Juhnke H, Krems B, Kotter P, Entian KD. Mutants that show increased sensitivity to hydrogen peroxide reveal an important role for the pentose-phosphate pathway in protection of yeast against oxidative stress. *Mol Gen Genet*. 1996; 252: 456–64. PMID: [8879247](#)
107. Li D, Zhu Y, Tang Q, Lu H, Li H, Yang Y, et al. A new G6PD knockdown tumor-cell line with reduced proliferation and increased susceptibility to oxidative stress. *Cancer Biother Radiopharm*. 2009; 24: 81–90. <https://doi.org/10.1089/cbr.2008.0494> PMID: [19243250](#)
108. Chauhan SC, Padmanabhan PK, Madhubala R. Glyoxalase Pathway of Trypanosomatid Parasites: A Promising Chemotherapeutic Target. *Curr Drug Targets*. 2008; 9(11): 957–65. PMID: [18991608](#)
109. Silva MS, Ferreira AEN, Gomes R, Tomás AM, Freire AP, Cordeiro C, et al. The glyoxalase pathway in protozoan parasites. *Int J Med Microbiol*. 2012; 302: 225–29. <https://doi.org/10.1016/j.ijmm.2012.07.005> PMID: [22901378](#)

10.2 Artigo publicado em colaboração durante o doutorado.

Systematic Identification and Classification of β -Lactamases Based on Sequence Similarity Criteria: β -Lactamase Annotation

Melise Chaves Silveira¹, Rangeline Azevedo da Silva¹, Fábio Faria da Mota¹, Marcos Catanho², Rodrigo Jardim¹, Ana Carolina R Guimarães² and Antonio B de Miranda¹

¹Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil. ²Laboratório de Genômica Funcional e Bioinformática, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil.

Evolutionary Bioinformatics
Volume 14: 1–11
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934318797351



ABSTRACT: β -lactamases, the enzymes responsible for resistance to β -lactam antibiotics, are widespread among prokaryotic genera. However, current β -lactamase classification schemes do not represent their present diversity. Here, we propose a workflow to identify and classify β -lactamases. Initially, a set of curated sequences was used as a model for the construction of profiles Hidden Markov Models (HMM), specific for each β -lactamase class. An extensive, nonredundant set of β -lactamase sequences was constructed from 7 different resistance proteins databases to test the methodology. The profiles HMM were improved for their specificity and sensitivity and then applied to fully assembled genomes. Five hierarchical classification levels are described, and a new class of β -lactamases with fused domains is proposed. Our profiles HMM provide a better annotation of β -lactamases, with classes and subclasses defined by objective criteria such as sequence similarity. This classification offers a solid base to the elaboration of studies on the diversity, dispersion, prevalence, and evolution of the different classes and subclasses of this critical enzymatic activity.

KEYWORDS: β -lactamase, class, subclass, identification, sequence similarity

RECEIVED: March 15, 2018. **ACCEPTED:** August 8, 2018.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

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: Melise Chaves Silveira, Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fiocruz, Av. Brasil 4365, Manguinhos, Rio de Janeiro RJ 21040-900, Brazil. Email: melisechaves@gmail.com

Introduction

The increasing amounts of genomic data produced by next-generation sequencing technologies made β -lactamases (BLs), the enzymes responsible for the irreversible inactivation of β -lactam antibiotics, one of the most numerous families of proteins studied to date.¹ First described by Abraham and Chain,² BLs can be found in pathogenic or commensal bacteria, isolated from humans or varied environments.³ Due to its great medical importance and clinical impact, several groups directed efforts in the development of a proper BL classification scheme, usually based on functional or structural criteria.^{4–6}

Function-based classification is achieved using experimental data to link the enzyme to its clinical role.^{6,7} The determination of these parameters for a large number of BLs, however, may be relatively costly and time-consuming. Also, they do not generate sequence data, essential for studies involving molecular evolution.⁸

Currently, the most widely used classification scheme for BLs is the Ambler structural classification, which is based on sequence similarity, and separates BLs into 4 classes: the classes A, C, and D of serine- β -lactamases (SBLs) and the class B of metallo- β -lactamases (MBLs).^{4,9,10} Class B is further divided into subclasses B1, B2, and B3, using sequence conservation data.¹¹

Despite SBLs and MBLs are able to break amide and ester bonds (EC 3.5.2.6), they belong to 2 distinct protein super-families that do not share a common ancestor.¹² Considering SBLs tertiary structures, they are similar enough among

themselves to be considered homologous,⁵ whereas the differences between their primary structures and catalytic mechanisms justify their division into classes A, C, and D.¹³

Experimental data indicate that the 3 subclasses of MBL should not be treated as equally separated groups. Subclasses B1 and B2 have detectable sequence similarity between them but not with B3,¹⁴ and structural evidence strongly suggest different Most Recent Common Ancestor between the group formed by subclasses B1 and B2 and the subclass B3.¹⁵

Based on this structural information, a reorganization of BL at 4 hierarchical levels was proposed by Hall and Barlow.⁵ In this scheme, the former subclasses B1 and B2 were merged and renamed as class MB, whereas subclass B3 was renamed as class ME. Thus, the 5 BL classes (third classification level) are SA, SC, SD, MB, and ME, and subclasses MB1 and MB2 represent the fourth and last hierarchical level.⁵

Moreover, recent studies employing phylogeny and amino acid-based sequence similarity networks showed that classes SA and SD could be further divided into 2 different groups, suggesting new BL subclasses.^{1,16} Indeed, the current molecular classification of BLs does not represent its actual sequence diversity, and there is yet no precise definition of various classes and subclasses.^{1,5,16}

Profiles Hidden Markov Models (HMM) and sequence clustering using similarity underlie the workflow proposed here. It is based on a previous hierarchical scheme which



reflects the distinct evolutionary origins of SBLs and MBLs.⁵ Application of this scheme to 2774 assembled bacterial genomes provided improvements in BL annotation and in the knowledge about BLs distribution among bacteria phyla, confirming previous studies suggesting new BL subclasses.^{1,16} Finally, our results propose the existence of a new BL class with fused domains and extended action spectrum.

Methodology

Data collection and preparation

On March 22, 2016, an online search for BL structures using the EC number (3.5.2.6) in the National Center for Biotechnology Information (NCBI) Structure Database¹⁷ returned 516 entries. Records with the word “mutant” in the description were excluded. The PDB (Protein Data Bank) IDs were retrieved and used for downloading PDB files and their corresponding FASTA files from RCSB PDB.¹⁸ The data were filtered using the following criteria: duplicate atoms positions were removed, the resolution of the structure should be less than 3 Å, and only monomers or the chain A from homomultimers were used.

A Non-Redundant Beta-Lactamase Dataset (NRBLD) was constructed with protein sequences retrieved from 7 different antibiotic resistance databases. Four databases are specific for BLs: (1) Comprehensive Beta-lactamase Molecular Annotation Resource (CBMAR, downloaded on August 2015),¹⁹ (2) The Institute Pasteur Database (downloaded on October 2015),²⁰ (3) DLact Antimicrobial Resistance Gene Database (downloaded on October 2015),²¹ and (4) Lactamase Engineering Database (LacED, downloaded on August 2015).²² The Metallo-Beta-Lactamase Engineering Database (MBLED, release 1.0) is the only specific for MBLs.²³ The remaining 2 databases, Comprehensive Antibiotic Resistance Database v1.0.0 (CARD)²⁴ and Resfams v1.2,³ have protein sequences related to antibiotic resistance in general. To recover only BL sequences, we search for those with the term “bla” and “beta” in the headers. Identical sequences have been removed using CD-HIT²⁵ at a 100% identity threshold. The following methodology steps are summarized in Figure 1.

Clustering curated BL structures and sequences

The clustering assays were made using BL structures and their corresponding amino acid sequences from PDB, in an attempt to reproduce the BL classes and subclasses proposed by previous works. The MaxCluster program²⁶ was used to cluster the structures based on root mean-squared deviation and hierarchical clustering, applying 3 tests: single, average, and maximum linkage. The BLASTClust v2.2.26 program²⁷ was used for the hierarchical clustering of the amino acid sequences, which performs a single linkage type clustering based on pairwise matches found by BLAST. Different threshold values of “BLAST score density” (BLAST score divided by the alignment length) and “minimum length coverage” were tested. An E value of 1E-05 was used.

Building a profile HMM for each BL class

The profiles HMM were constructed using the clusters of sequences corresponding to the 5 BL classes.⁵ Because these sequences came from proteins with a resolved structure, they were considered reliable to be used to construct the profiles. Sequences from each cluster were saved in separate multi-FASTA files. Identical sequences were removed using CD-HIT²⁵ at a 100% identity threshold. Sequences in each file were aligned using MUSCLE.²⁸ Profiles HMM were built from each alignment using the program *hmmbuild* from the HMMER package v3.1b2.²⁹

Calibration and validation of the profiles HMM

The profiles HMM generated were used against the protein sequences at NRBLD and superfamilies 3.40.710.10 (DD-peptidase/BL like) and 3.60.15.10 (Metallo-BL like) from Protein Structure Classification Database (CATH) (downloaded on March 17, 2016).³⁰ *Hmmsearch* from the HMMER package v3.1b2 with an E value of 1E-05 was employed.²⁹

The Class profile Specificity index (CpSp) evaluates whether each profile identifies a unique group of sequences. CpSp was calculated by dividing the number of NRBLD sequences identified exclusively with a given profile (Ne) by the total number of NRBLD sequences recovered by it (T), including intersections with others profiles results [$CpSp = (Ne/T) * 100$]. Profiles with CpSp below 100% were calibrated. Sequences from other classes that should not be identified were used as “negative training sequences,” following the HMM-ModE protocol.³¹ After this, the *hmmsearch* Gathering Threshold (GA) parameter was used, substituting the E value.

A total of 851 amino acid sequences of DD-peptidase/BL-like superfamily downloaded from CATH³⁰ were used to construct an unrooted Maximum Likelihood phylogenetic tree in MEGA-CC v7.0.18,³² using the Jones-Taylor-Thornton model, partial deletion for gaps/missing data treatment (95% site coverage cutoff) and 500 bootstrap replicates. Using in-house Perl scripts, the BL sequences were manually labeled with their respective class, and the clade node of each BL class was identified. The sequences in these clades were allocated in corresponding multi-FASTA files.

Graphics were created with all the sequences retrieved by each profile against the DD-peptidase/BL superfamily with their respective HMM bit score. These results were used to calibrate the profiles. A new *hmmsearch* parameter of HMM bit score threshold was established to separate true BL from other sequences in the superfamily. The Function Specificity (FSp) index verifies whether all sequences retrieved by a profile have the BL activity already described. For this, the number of superfamily sequences identified by a profile that has BL activity (Nbl) was divided by the total number of superfamily sequences identified by the profile (T) [$FSp = (Nbl/T) * 100$].

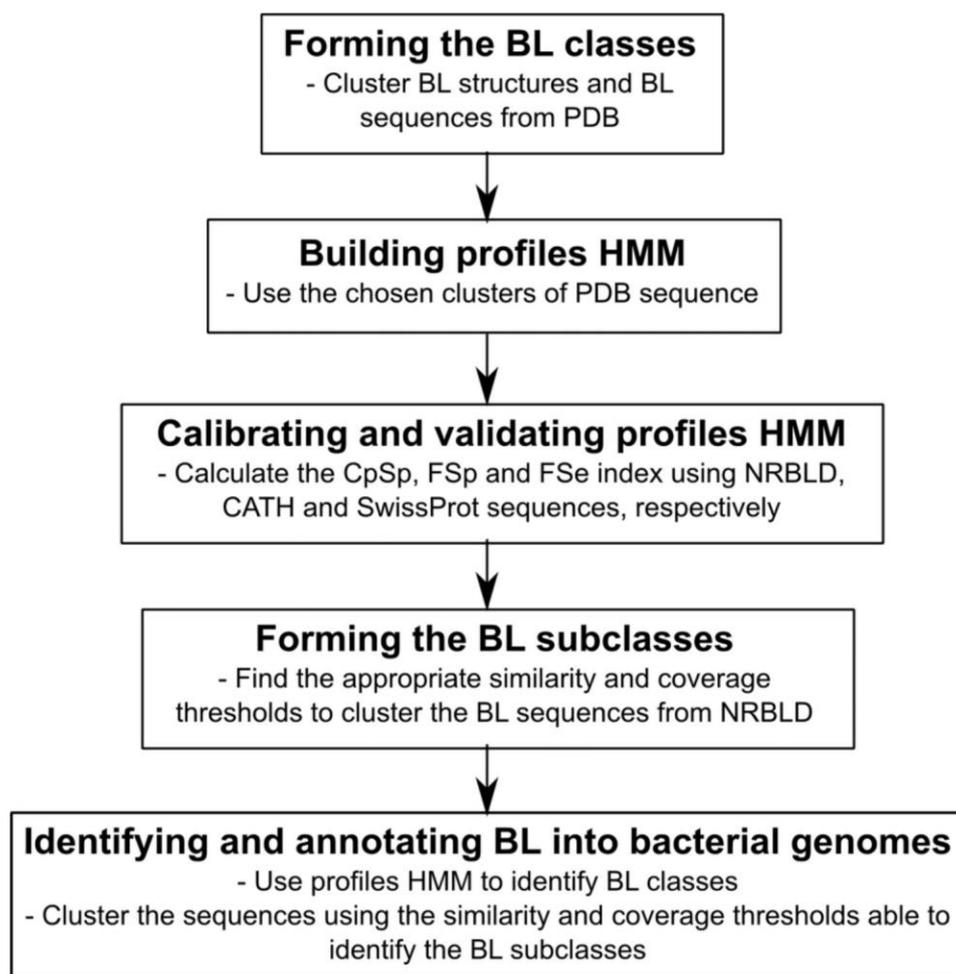


Figure 1. Main conceptual steps of the workflow.

BL indicates β -lactamase; CATH, Protein Structure Classification Database; PDB: Protein Data Bank; CpSp, Class profile Specificity; FSe, Function Sensibility; FSp, Function Specificity; NRBLD, Non-Redundant Beta-Lactamase Dataset.

A total of 144 sequences are attributed to the Gene Ontology (GO) molecular function “BL activity” (GO:0008800) in the SwissProt database (downloaded on March 2, 2016).³³ The validation index Function Sensitivity (FSe) evaluates the ability of all profiles together to identify the sequences annotated with the BL function in SwissProt. To calculate this index, the number of identified SwissProt sequences attributed to the BL GO term (N_{go}) was divided by 144 [$FSe = (N_{go}/144) * 100$].

Validation of thresholds used to form BL subclasses

The results obtained with the curated PDB data set were used as a basis to group the NRBLD sequences into subclasses. The BLs from PDB were added to NRBLD, and the profiles HMM were used against them.

In order to reproduce BL subclasses, different thresholds of “minimum length coverage” were tested to cluster the sequences in each class, together with the “BLAST score density” thresholds previously established using the PDB sequences. The sequences in each cluster were annotated using the BLASTP³⁴ best hit (v2.2.28) against the nonredundant NCBI protein database.¹⁷

The size of specific domains of BLs was used to stipulate a minimum size necessary for the enzyme to be functional.³⁵ The Pfam³⁶ models most common in MBL (PF00753), class SA (PF13354), and class SC (PF00144) have a length of 197, 202, and 330 residues, respectively. A total of 214 residues domain have been attributed to class SD enzymes.³⁷

The clusters formed were categorized as follows: (1) clusters corresponding to those obtained in the hierarchical classification of functionally characterized BLs (PDB sequences) and (2) clusters containing sequences that do not fit into the previous established similarity and coverage thresholds.

Comparison of the improved profiles HMM with Pfam profiles and BL motifs

To test the efficiency of our profiles HMM in identifying and classifying BL sequences, these were compared with profiles available in the Pfam database³⁶ and with motifs specific to BL classes from different sources.

The profiles HMM from Pfam³⁶ were used to search for BL sequences from PDB, and the CpSp indexes were calculated accordingly. Seven Pfam profiles (PF00144.22, PF13354.4,

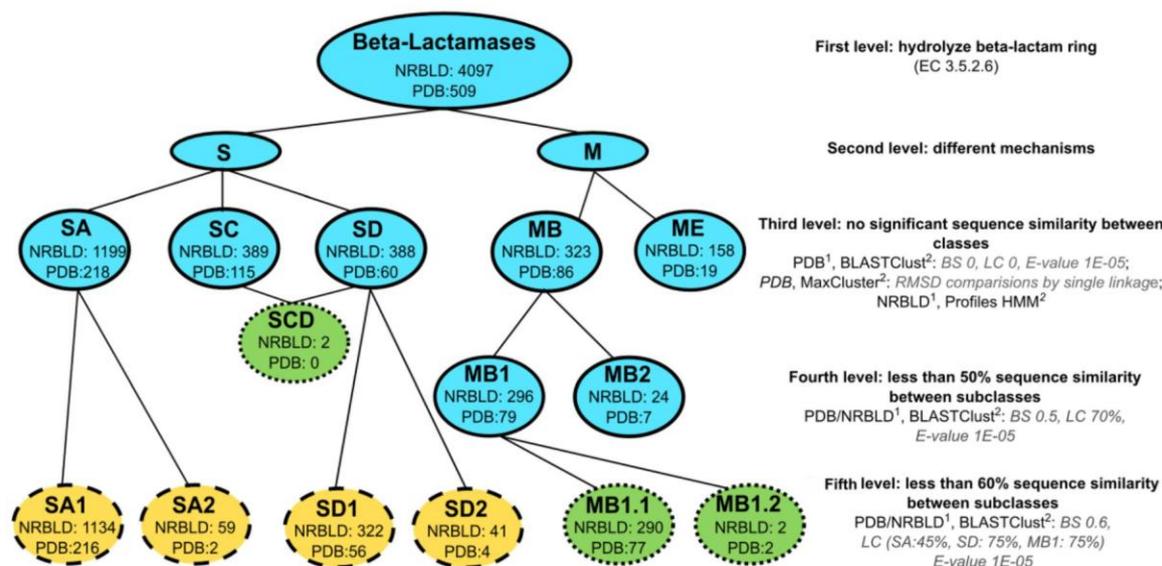


Figure 2. Hierarchical classification of β -lactamases.

The number of Protein Data Bank (PDB) and Non-Redundant Beta-Lactamase Dataset (NRBLD) sequences in each cluster after clustering is shown (PDB identifiers can be obtained in Table S2). 1 indicates the sequence dataset used for clustering. 2 indicates the program used for clustering, followed by the parameters used. Blue: Ambler's classification with real relationships as shown by Hall and Barlow⁵; green and dotted: new groups proposed for the first time in this study; yellow and dashed: groups described recently and confirmed in this work^{1,16}; BS: BLAST score density; LC: minimum length coverage; BL: Beta-lactamase; S: Serine-BL; M: Metallo-BL; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD; MB: Metallo-BL class MB (former subclasses B1 and B2); ME: Metallo-BL class ME (former subclass B3); SCD: Serine BL class SC-class SD; MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; MB1.1: Metallo-BL subclass MB1.1; MB1.2: Metallo-BL subclass MB1.2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2.

PF00753.25, PF12706.5, PF13483.4, PF14597.4, and PF16661.3) were downloaded on December 2016.

Motifs specific for BL classes obtained from the literature were evaluated for their ability to distinguish between classes and subclasses. An in-house Perl script was developed to identify the motifs in the BL sequences from PDB.

Identification and classification of BLs in fully assembled genomes

The workflow proposed here was applied to identify and classify BL sequences present in 2774 bacterial strains with fully assembled genomes that were deposited in NCBI.¹⁷ A genome was defined as the complete set of chromosome and plasmids of each strain. The protein sequences present in each genome were downloaded (June 2016). Using the profiles HMM, the BL classes were formed, which were then separated into their respective subclasses applying the BLASTClust program²⁷ with the "BLAST score density" and "minimum length coverage" thresholds set in the previous step. The taxonomic information of each sequence was determined using the Genome Online Database (GOLD)³⁸ and in-house Perl scripts.

The scripts, profiles HMM, and instructions required to apply the workflow presented here, in addition to the data used for the searches, are available at <https://github.com/melisesilveira/betaLactamase-classification.git>.

Results

Clustering curated BLs structures and sequences

In total, 509 PDB structures and their respective sequences were used in the clustering and in the construction of profiles

HMM. Of these, 208 were monomers and 301 were homomultimeric proteins.

Clustering of the curated set of structures using single linkage was consistent with the BL classification as proposed by Hall and Barlow, producing the same 5 classes (Figure 2).⁵ Average and maximum linkage resulted in 7 and 18 clusters, respectively, and were not further used.

The PDB sequences clustering applying "BLAST score density" thresholds of 0 and 0.5 were also consistent with the Hall and Barlow's classification scheme,⁵ producing the 5 BL classes and 2 MB subclasses, respectively. Using a "BLAST score density" threshold equal to 0.6, the subclasses proposed in previous works for classes SA and SD were reproduced,^{1,16} as well as the division of MB1 into 2 groups (Figure 2). The length coverage did not influence these results, and therefore a "minimum length coverage" of zero was chosen.

Separate clusters, one composed by 2 BL TEM-1 fused to a maltose-binding protein and another with 5 *Escherichia coli* penicillin-binding protein 5 (PBP5), were created in all clustering assays and were excluded because they are not true BLs, being wrongly associated with EC 3.5.2.6 in PDB. All sequence and structural clustering results are presented in Tables S1 to S3.

Building, calibration, and validation of the profiles HMM

Profiles HMM were built for each of the 5 clusters created using curated PDB sequences. These profiles were calibrated and validated using 3 sets of sequences: (1) NRBLD, our non-redundant protein set containing 4097 sequences; (2) the

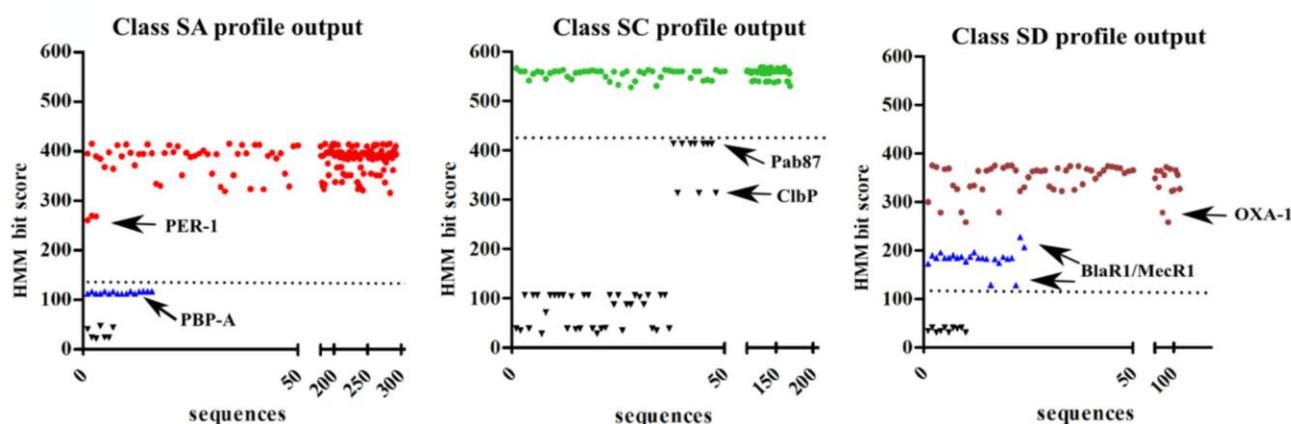


Figure 3. Plot of the 851 DD-peptidase/β-lactamase-like superfamily (CATH 3.40.710.10) sequences.

Each dot represents a sequence identified by the profile with its HMM bit score result. The HMM bit score reflects whether the sequence is a better match to the profile model (positive score) or the null model of nonhomologous sequences (negative score). The dashed lines in the graphs indicate the established HMM bit score threshold for each profile. X: numbered sequence order, Y: HMM bit score. Red circle: 295 sequences from BL class SA, green circle: 177 sequences from BL class SC, brown circle: 103 sequences from BL class SD, blue triangle: proteins inserted within the BL clades that are not functionally characterized as BLs, and black inverted triangle: other non-BL sequences from DD-peptidase/β-lactamase-like superfamily. Arrows indicate particular proteins quoted in the text; BL: Beta-lactamase.

DD-peptidase/BL- and MBL-like superfamilies, with 851 and 399 protein sequences, respectively; and (3) SwissProt/UniProt³³ database, with 550 552 sequences. The profiles for the SBL and MBL classes were analyzed separately as these 2 groups belong to different CATH superfamilies. The data and profiles HMM are available at <https://github.com/melisesilveira/betaLactamase-classification.git>.

The profiles for classes SA, SC, and SD were searched against NRBLD recovering 1292, 1121, and 396 sequences, respectively. The profile of the class SA was 100% specific for the class (CpSp), whereas the profiles of the classes SC and SD were 99% specific. Two sequences were identified by both profiles, BL LRA-13 (ACH58991.1) and an enzyme annotated as “class C BL” of *Janthinobacterium* sp. (WP_008451281.1). Both have domains of SC and SD classes, which led us to propose a new class, SCD, composed of BLs with fused domains (Figure 2).

The profiles for MBL classes recovered 527 (MB, CpSp=84%) and 612 (ME, CpSp=86%) NRBLD protein sequences, 84 of which were identified by both profiles. Therefore, the profiles were optimized, and after the calibration they recovered 323 (MB) and 159 (ME) sequences, without intersections, reaching 100% CpSp.

To compare the calibrated and noncalibrated MBL profiles, they were tested against the 598 MBL protein sequences from the MBLED database.²³ Initial profiles identified 440 (MB, 85% CpSp) and 222 (ME, 70% CpSp) sequences, which represents 99.8% of the database. However, calibrated profiles identified 424 (MB, 100% CpSp) and 164 (ME, 100% CpSp) sequences, representing 98.3% of the database. The remaining 1.7% was composed of protein fragments ranging from 75 to 131 amino acids, meaning that the calibrated profiles were able to identify all the complete BL sequences.

The phylogenetic relationships of the SBL superfamily proteins showed a few sequences with no BL activity inserted into BL clades (Figure S1). The penicillin-binding proteins A (PBP-A) are in the inner branches of class SA and have a

structure very similar to the BL PER-1 (subclass SA2) but do not have BL activity.³⁹ The regulatory proteins BlaR1 (and its cognate MecR1) are in inner branches of the class SD. Their extracellular domains are phosphorylated by β-lactams and, consequently, these proteins regulate resistance to these antibiotics in *Staphylococcus aureus*.⁴⁰

Initially, the profiles HMM for SBLs identified these sequences and others outside the BLs clades (all non-BLs). However, based on their HMM bit score values, a total of 75 non-BL sequences could be excluded (Figure 3). The separation of true SD2 BLs from BlaR1 proteins in this step was not possible because the HMM bit score of the single structure available of this subclass present in the CATH database³⁰ (OXA-1) was very close to some BlaR1 proteins (Figure 3). Additional tests have shown that when other variants of this subclass are included, their scores are smaller than those of some BlaR1 sequences. This can be explained by the structural homology of the extracellular sensor domain of BlaR1 to BLs from subclass SD2.⁴¹

The enzymes ClbP and Pab87 are associated with the BL EC number in the PDB and share a significant similarity between their active sites with the BLs from class SC.⁴² However, they can be separated from the class SC BLs by both phylogeny (Figure S1) and HMM bit score (Figure 3). Sequences in each BL clade and their respective HMM bit score are shown in Table S4.

HMM bit score thresholds of 120, 430, and 120 were defined for classes SA, SC, and SD, respectively (Figure 3). After the utilization of these bit scores, the FSp was equal to 100%, 100%, and 81%, for classes SA, SC, and SD, respectively. These thresholds retrieved 1199, 389, and 388 sequences from NRBD, respectively.

After the calibrations, all BL profiles together retrieved 132 proteins from SwissProt³³: 82, 10, 24, 15, and 1 sequences with the profiles for classes SA, SC, SD, MB, and ME, respectively. In all, 125 out of the 144 sequences annotated with the BL GO term in this database were identified by the profiles (87% FSe).

The remaining 19 sequences not identified as BLs presented one of the following annotations: BL fragments, BL-like protein, DacA carboxypeptidase, hydroxyacylglutathione hydrolases, and ribonuclease. Also unidentified were Hcp family proteins and a sequence described as a class SA BL PenA.⁴³ Hcp proteins, a family of cysteine-rich PBP, do not have a significant sequence or structural similarity to known BLs.⁴⁴ A BLASTP³⁴ search in the NCBI protein database¹⁷ with the putative PenA returned as best-hit a class SC protein with only 13% coverage, meaning that it is probably not a BL and certainly not a PenA. Seven BlaR1/MecR1 proteins were also identified by the profiles but are regulatory proteins not associated with the GO term for BL.

Validation of BLASTClust thresholds to form BLs subclasses

The “BLAST score density” values applied to the curated BL sequences to form the subclasses were validated in the PDB sequence set plus NRBLD. A significant length variation was observed between the BL sequences in PDB and those in NRBLD. The sequences in PDB range from 219 to 447 amino acids, whereas in NRBLD, they range from 96 to 619 amino acids. Therefore, in addition to the previous values of “BLAST score density,” different “minimum length coverage” thresholds were chosen to cluster NRBLD sequences. Clustering results are available in Tables S5 and S6 and the thresholds in Figure 2.

Application of the similarity and coverage thresholds stipulated for clustering resulted in the separation of true BLs from other sequences such as partial domains and the regulatory proteins BlaR1/MecR1. For instance, in the case of class SA, most non-BL sequences have a larger average size (345–637 amino acids) than BLs in subclasses SA1 and SA2 (285 and 300 amino acids, respectively). One cluster contains a functionally characterized BL (LRA-5, non-BL9, Table S5). No non-BL sequence was observed for the class SC. All clusters containing non-BL sequences from class SD have one sequence, which is similar in size to BlaR1/MecR1 proteins (~585 amino acids) or partial domains (<214 amino acids). Only one of them (YP_612206, non-BL13, Table S5) is similar in size to class SD BL (274 amino acids) and its best hit in the NCBI nonredundant protein database¹⁷ shows 43% identity with a sequence annotated as “class D BL” from *Oceanicaulis alexandrii* (E value of 1E-67). All non-BL sequences from class MB are partial domains (<197 amino acids). Among the 2 clusters containing non-BL sequences formed from MB1, one possesses partial domain sequences and the other has a 340-amino acid sequence from *Stigmatella aurantiaca*, considerably larger than BL sequences (~250 amino acids). The ME2 subclasses were not maintained after clustering of NRBLD sequences, even with higher minimum coverage thresholds (90%), not corroborating what was observed when only sequences from PDB were clustered.

Comparison of the improved profiles HMM with Pfam profiles and BL motifs

Seven Pfam³⁶ profiles were tested against the curated data set of BLs from PDB, showing low specificity for BL classes. The 2 Pfam profiles for SBL (PF00144.22 and PF13354.4) identified 339 and 227 enzymes out of a total of 399 SBLs, with an intersection of 220 sequences (35% and 3% CpSp, respectively). Three profiles for MBL (PF00753.25, PF12706.5, and PF16661.3) identified 98, 35, and 12 enzymes out of a total of 105 MBLs available (64%, 0%, and 0% CpSp, respectively). The other 2 MBL profiles did not identify any enzyme (PF13483.4 and PF14597.4).

Different sources were used to select 13 motifs related to the various BL classes, which were used to search among the 509 PDB sequences. About 11 specific motifs for the third classification level (SA, SC, and SD classes) and 2 motifs specific only for the second level (MBL) were used (Tables 1 and 2). The efficiency of these motifs was tested to separate BL sequences into subclasses (fourth and fifth levels).

No motifs developed for MBL nor motifs for class SA are present in all the sequences allocated to their respective groups. The KxxS motif⁴⁷ was found in all class SC sequences, whereas the motif SxV⁶ is present in all sequences allocated in the class SD. None of the motifs analyzed was specific to BL subclasses (MB1 or MB2, SA1 or SA2, SD1 or SD2; Tables 1 and 2).

Identification and classification of BLs in fully assembled genomes

A total of 1476 BL sequences were identified in 2774 prokaryotic genomes. SA, SC, SD, MB, and ME profiles recovered 616, 280, 366, 103, and 111 sequences, respectively. No SCD class members were found in the genomes surveyed.

After the clustering and annotation process, 123 (8.3%) sequences were considered non-BLs (Table S7). The remaining 1352 sequences (91.7%) were distributed among 12 phyla and classified according to the BL subclasses to which they belong (Table 3). All subclasses were found in the *Proteobacteria* phylum, excepted for MB1.2. SD1 is the most disseminated subclass among the phyla analyzed, whereas SC, SD2, and MB2 were mostly restricted to *Proteobacteria*. A clear difference can be observed between the phyla where BL sequences of the subclasses SA1 (*Proteobacteria*, *Firmicutes*, and *Actinobacteria*) and SA2 (*Bacteroidetes*, *Cyanobacteria*, and *Spirochaetes*) were identified. It should be noted that 49% of all analyzed strains belong to the phyla *Proteobacteria*. Regarding MBL subclasses, 51% (46) of MB1.1 sequences are in *Firmicutes*, whereas 82% (91) of ME are in *Proteobacteria* (Table 3). A single sequence of the MB1.2 class was identified, isolated from *Sediminispirochaeta smaragdinae* (WP_013255389.1), a bacteria belonging to the *Spirochaetes* phyla, with a specific domain of SPM-1 (cd16286).

Table 1. Efficiency of serine-β-lactamase motifs.

MOTIF	TARGET	CLASS	SUBCLASS 1*	SUBCLASS 2**
ExxLN ^a	SA	174/218 (80%)	174/216 (81%)	0/2 (0%)
SDN ^a	SA	183/218 (84%)	181/216 (83%)	2/2 (100%)
KTG ^a	SA	169/218 (78%)	167/216 (77%)	2/2 (100%)
S-[DG]-N-x(1,2)-A-[ACGNST]-x(2)-[ILMV]-x(4)-[AGSTV] ^b	SA	107/218 (49%)	105/216 (48%)	2/2 (100%)
[FY]-x-[LIVMFY]-{E}-S-[TV]-x-K-x(3)-{T}-[AGLM]-{D}-{KA}-[LC] ^c	SA	192/218 (88%)	190/216 (87%)	2/2 (100%)
YxN ^a	SC	116/119 (97%)	—	—
KxxS ^d	SC	119/119 (100%)	—	—
[FY]-E-[LIVM]-G-S-[LIVMG]-[SA]-K ^e	SC	118/119 (99%)	—	—
SxV ^a	SD	60/60 (100%)	56/56 (100%)	4/4 (100%)
SxxxxS ^d	SD	50/60 (83%)	46/56 (82%)	4/4 (100%)
[PA]-x-S-[ST]-F-K-[LIV]-[PALV]-x-[STA]-[LI] ^c	SD	43/60 (72%)	41/56 (73%)	2/4 (50%)

^aBush.⁶^bSingh et al.⁴⁵^cPROSITE.⁴⁶^dMACiE.⁴⁷ Target: BL class for which the motif was developed; BL: Beta-lactamase; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD.

*Subclasses SA1 and SD1; **subclasses SA2 and SD2; — represents classes that have no subclass.

Table 2. Efficiency of metallo-β-lactamase (MBL) motifs.

MOTIF	TARGET	MBL	ME	MB	MB1	MB2
[LI]-x-[STN]-[HN]-x-H-[GSTAD]-D-x(2)-G-[GP]-x(7,8)-[GS] ^a	MBL	55/105 (52%)	12/19 (63%)	43/86 (51%)	36/79 (46%)	7/7 (100%)
P-x(3)-[LIVM](2)-x-G-x-C-[LIVMF](2)-K ^a	MBL	46/105 (44%)	0	46/86 (54%)	39/79 (50%)	7/7 (100%)

^aPROSITE.⁴⁶ Target: BL class for which the motif was developed; BL: Beta-lactamase; MBL: Metallo-BL; MB: Metallo-BL class MB; ME: Metallo-BL class ME; MB1: Metallo-BL subclass 1; MB2: Metallo-BL subclass MB2.

A higher absolute number of BL sequences were observed in *Proteobacteria*. The distribution of BLs was also analyzed at the taxonomic class level for this phylum. *Gammaproteobacteria* class has 60% (545) of the BL sequences, divided among all BL subclasses. SD1 is the only subclass found in all classes, and the unique BL found in *Epsilonproteobacteria* (Table 4).

A total of 100% of the genomic sequences retrieved by the profiles (after the exclusion of non-BL sequences) were allocated to some subclass of BL. About 70% of their original annotations were designated only as BL (first level), 24% had information on the second or third level of classification or the gene name, and there were still 6% with erroneous or imprecise annotations, such as “hypothetical protein” (Figure 4).

Discussion

Classification schemes for BLs are of utmost importance due to the diversity of these enzymes and their importance in the scenario of bacterial resistance to antibiotics.^{1,4,5,16} In general, the identification of new sequences is most often done by sequence comparison methods,⁴⁸ such as the BLASTP

program.³⁴ Profiles HMM and other profile-sequence comparison methods led to a significant improvement in sensitivity over sequence comparison approaches and are already used in the identification of antibiotic resistance proteins.^{3,49}

The workflow developed here systematizes the annotation of BLs based mainly on 2 steps: searches using profiles HMM, followed by clustering the resulting sequences. The calibrated profiles HMM can assign a sequence to a specific class. They also discriminate functionally characterized BLs from proteins with other biochemical functions that belong to the same superfamily and therefore share fold signals that make this separation difficult.³¹ In addition, the calibrated profiles allowed the recognition of sequences erroneously attributed to the BL GO term (“BL activity”) in SwissProt³³ and also enable the identification of sequences imprecisely described as BL in different antibiotic resistance databases. In the clustering step, the established thresholds of similarity and coverage allowed the clearing of non-BL sequences, providing coherent BL subclasses.

Table 3. β -lactamase sequences identified in bacterial genomes.

PHYLA	GENOMES	SA		SC	SD		MB			ME	TOTAL
		SA1	SA2		SD1	SD2	MB1.1	MB1.2	MB2		
<i>Proteobacteria</i>	1176	302	5	276	132	68	24	—	5	91	903
<i>Firmicutes</i>	583	148	—	—	30	—	46	—	—	3	227
<i>Actinobacteria</i>	283	107	—	4	4	—	—	—	—	1	116
<i>Bacteroidetes</i>	88	—	16	—	14	—	18	—	—	3	51
<i>Cyanobacteria</i>	73	—	2	—	17	—	—	—	—	—	19
<i>Spirochaetes</i>	60	—	1	—	3	—	2	1	—	7	14
Other	135	2	4	—	9	1	—	—	—	6	22
Total	2398	559	28	280	209	69	90	1	5	111	1352

Genomes: number of strains analyzed by phylum. — represents phyla where no sequences were found. Other: Verrucomicrobia, Acidobacteria, Fusobacteria, Gemmatimonadetes, Chlorobi, and Chlamydiae; BL: Beta-lactamase; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD; MB: Metallo-BL class MB (former subclasses B1 and B2); ME: Metallo-BL class ME (former subclass B3); MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2; MB1.1: Metallo-BL subclass MB1.1; MB1.2: Metallo-BL subclass MB1.2.

Table 4. β -lactamase sequences identified in bacterial genomes from *Proteobacteria* phylum.

CLASS	GENOMES	SA		SC	SD		MB		ME	TOTAL
		SA1	SA2		SD1	SD2	MB1.1	MB2		
<i>Gammaproteobacteria</i>	546	148	3	216	77	20	13	4	64	545
<i>Alphaproteobacteria</i>	321	75	2	29	15	15	2	—	22	160
<i>Betaproteobacteria</i>	146	76	—	31	13	29	—	1	4	154
<i>Epsilonproteobacteria</i>	104	—	—	—	21	—	—	—	—	21
<i>Deltaproteobacteria</i>	59	3	—	—	6	4	9	—	1	23
Total	1176	302	5	276	132	68	24	5	91	903

Genomes: number of strains analyzed by phylum. — represents phyla where no sequences were found BL: Beta-lactamase; SA: Serine-BL class SA; SC: Serine-BL class SC; SD: Serine-BL class SD; MB: Metallo-BL class MB (former subclasses B1 and B2); ME: Metallo-BL class ME (former subclass B3); MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2; MB1.1: Metallo-BL subclass MB1.1.

Among the sequences from BL databases used to construct NRBLD, some non-BL sequences presented partial domains, suggesting assembly/sequencing artefacts or annotation errors (eg, non-BL3, Table S5).³⁵ Others had much larger sizes than BLs, such as the BlaR1 protein, homologous to the BLs of the class SD, but with different functions. However, 2 sequences classified as non-BL did not cluster within any BL subclass despite being similar in size to them. The LRA-5 protein, described and functionally characterized as a class SA BL, has low similarity and is a distant relative to functionally characterized BLs and their ancestors.⁵⁰ As the experimental and HMM profile data indicate that LRA-5 is a BL of class SA, we speculate that it may belong to a third subclass (SA3) considering its low similarity to other sequences. Availability of new sequences in the future will help confirm the existence of this new. The second exception (YP_612206) is 43% identical to a sequence annotated as

“class D BL” of the dimorphic rods *O alexandrii*, although the activity of this protein has not been demonstrated.⁵¹

It has been shown that the use of Pfam³⁶ profiles to identify sequences from the “Ser-BL-like superfamily” may capture unrelated sequences.¹⁶ In our comparisons, the improved profiles HMM displayed higher specificity when compared with Pfam profiles and BL motifs from literature. Recently, individualized subgroups of the class SA have been demonstrated, such as LSBL or TEM/SHV and CARB clusters. Characteristic residues have already been attributed to each of them, but in this work we have chosen to test motifs attributed to BL classes.¹ However, subclass-specific motifs, such for SA1 and SA2, should also be tested and compared in future studies.

Some BL subclasses that were previously described based on phylogenetic criteria were identified here using sequence similarity criteria. The sequences in subclasses MB1 and MB2 correspond to the subclasses “B1” and “B2” in the work of

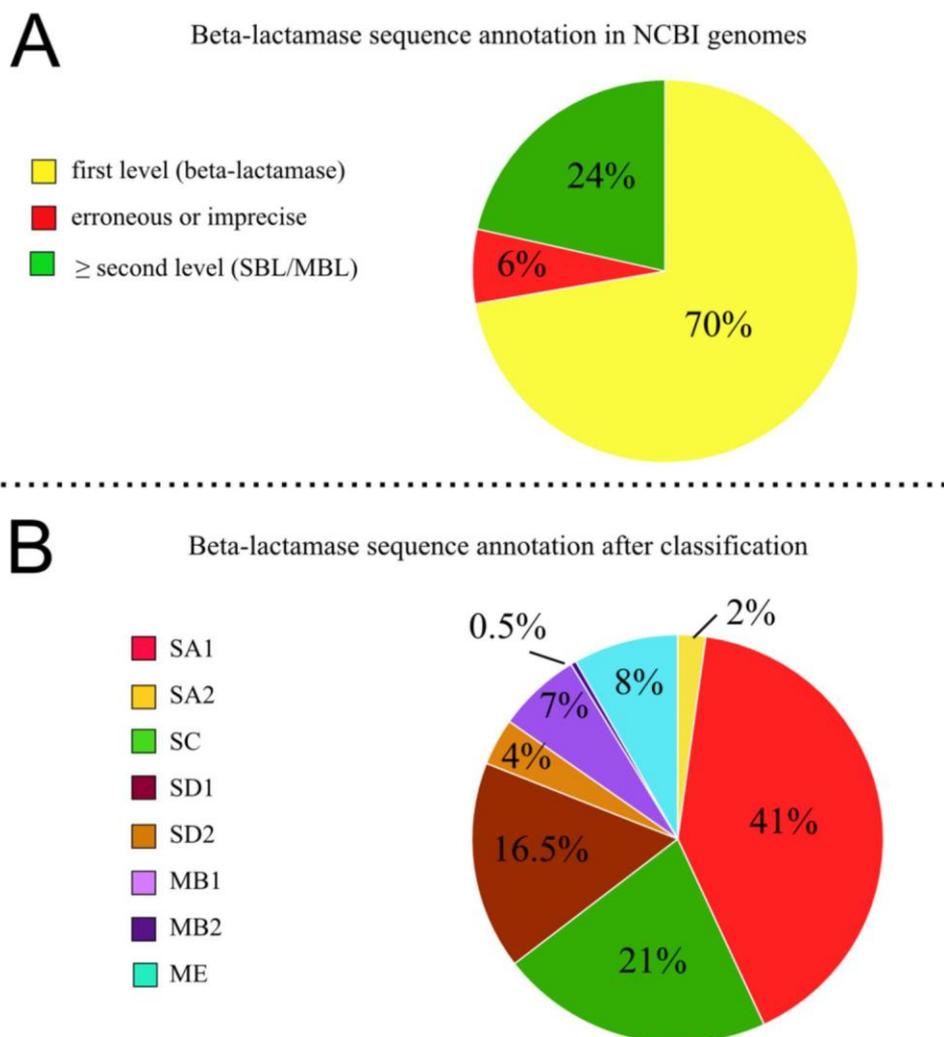


Figure 4. Original (A) annotation and (B) reannotation of the 1352 sequences according to the hierarchical levels of BL classification. BL indicates β -lactamase; MBL: metallo- β -lactamase; SBL: serine- β -lactamase; MB1: Metallo-BL subclass MB1; MB2: Metallo-BL subclass MB2; SA1: Serine-BL subclass SA1; SA2: Serine-BL subclass SA2; SD1: Serine-BL subclass SD1; SD2: Serine-BL subclass SD2; ME: Metallo-BL class ME.

Galleni et al.⁵² The sequences in the subclass SA2 correspond to BLs isolated mainly from the group Cytophagales-Flavobacteriales-Bacteroidales.^{1,16,53} The OXA alleles in subclass SD2 are the same as those found in the class called “D2” by Brandt et al.¹⁶

A new class of BLs with fused domains, the class SCD, is proposed. Two NRBLD proteins were captured by both SC and SD profiles without using the HMM bit score threshold. One of them, LRA-13, isolated from a noncultivated soil bacterium in Alaska, was confirmed experimentally as a BL, displaying a broad hydrolytic profile.⁵⁰

Subclass MB1.2 was first presented here, formed by members of the BL SPM family. It has been suggested that SPM-1 may be a structural hybrid between MB1 and MB2 subclasses.⁵⁴ SPM-1 genes were found only in isolates of *Pseudomonas aeruginosa*, a *Proteobacteria*, and its chromosomal location may have contributed to its isolation to other BL families of subclass MB1.⁵⁵ Curiously, we found that *S. smaragdinae*, a spirochaete, carries a protein with the SPM-1 domain

(WP_013255389.1). Further studies are needed to establish the evolutionary relationships between these proteins.

The distribution of BL sequences obtained from fully assembled genomes in different subclasses confirmed and complemented previous observations. For instance, the observed enrichment of BLs in *Actinobacteria* relative to other phyla³ is caused mainly by members of subclass SA1. The BLs of subclass SA2, related to the *Bacteroidetes* phylum,^{1,16} were also found in other phyla. The majority presence of class SC in the *Proteobacteria* phylum is a consensus, and the few sequences found in *Actinobacteria* confirm the occasional isolation of this class.¹⁶ The wide distribution of the subclass SD1 between the phyla analyzed confirms that class SD, more precisely the subclass SD1, has been underestimated.¹⁶ The recently described presence of this subclass in Gram-positive bacteria has also been confirmed here.⁵⁶ The association of BL sequences from subclass SD2 to *Proteobacteria* may be related to the fact that they are naturally occurring intrinsic genes, most likely chromosomally located.¹⁶ The occurrence of MB1 in *Bacteroidetes*

and *Firmicutes* phyla has been associated with chromosome, whereas the *Proteobacteria* MB1 enzymes are mostly mobile.⁵⁷ The association of class ME with soil bacteria may be related to its distribution among different phyla, which can share the same environment.³ New BLs of class ME have already been described in metagenomes, and the majority diverge deeply from other known enzymes of this class,⁵⁰ which reinforces the idea that class ME is widely distributed and diverse.

Acidobacteria are abundant mainly in the soil, but their cultivation is difficult.⁵⁸ Although few genomes of this phylum were available in 2016, the number of BLs identified was significant, suggesting that this is an important reservoir of BLs in nature.

The class *Gammaproteobacteria* includes common human pathogens such as *Enterobacteriaceae* and *Pseudomonadaceae*.¹⁶ These pathogens are exposed continuously to evolutionary pressure exerted by antibiotics, favoring the acquisition of resistance genes, which may be related to the presence of different BL subclasses in this group. *Epsilonproteobacteria*, which are widely distributed bacteria including genera with pathogenic species for humans such as *Helicobacter* and *Campylobacter*, presented BLs of only the SD1 subclass. Indeed, the production of 2 major BLs (OXA-61 and OXA-184) was detected in 85% of the *Campylobacter* strains.⁵⁹

Conclusions

The workflow developed in this study presents better specificity when compared with available BL motifs and Pfam profiles. Application of the improved profiles HMM and sequence similarity clustering parameters resulted in a 5-level hierarchical classification, consistent with previous BL classification scheme and recent proposed subclasses based on phylogeny. We also emphasize that the number of amino acids may be an important criterion for characterizing BLs, although there may be exceptions, such as the new class of fused domain enzymes (class SCD), proposed here. The workflow presented here, which can be further improved by the addition of functional and phylogenetic data, will be of great help in studies on the prevalence, distribution, and evolution of this critical enzymatic activity.

Acknowledgements

The authors would like to thank Dr Reema Singh and Dr Harpreet Singh for providing the DLact database sequences and Dr Alex Herbert for giving us the script to remove alternative positions of the atoms. MCS recognizes CAPES (Coordination for the Improvement of Higher Education Personnel, Brazil) for supporting her with a scholarship during her DSc program.

Author Contributions

MCS and ABM conceived of the manuscript outline; MCS, FFM, MC, RJ, ACRG, and ABM jointly developed the methodology; MCS, RAS, FFM, MC, and ABM wrote, edited and revised the manuscript.

REFERENCES

- Philippou A, Slama P, Dény P, Labia R. A structure-based classification of class A β -lactamases, a broadly diverse family of enzymes. *Clin Microbiol Rev.* 2016;29:29–57.
- Abraham EP, Chain E. An enzyme from bacteria able to destroy penicillin. *Nature.* 1940;146:837.
- Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 2014;9:207–216.
- Ambler RP. The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci.* 1980;289:321–331.
- Hall BG, Barlow M. Revised Ambler classification of β -lactamases. *J Antimicrob Chemother.* 2005;55:1050–1051.
- Bush K. The ABCD's of β -lactamase nomenclature. *J Infect Chemother.* 2013;19:549–559.
- Bush K, Jacoby GA, Medeiros AA. A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob Agents Chemother.* 1995;39:1211–1233.
- Ouzounis CA, Coulson RMR, Enright AJ, Kunin V, Pereira-Leal JB. Classification schemes for protein structure and function. *Nat Rev Genet.* 2003;4:508–519.
- Jaurin B, Grundström T. AmpC cephalosporinase of *Escherichia coli* K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type. *Proc Natl Acad Sci U S A.* 1981;78:4897–4901.
- Ouellette M, Bissonnette L, Roy PH. Precise insertion of antibiotic resistance determinants into Tn21-like transposons: nucleotide sequence of the OXA-1 beta-lactamase gene. *Proc Natl Acad Sci U S A.* 1987;84:7378–7382.
- Rasmussen BA, Bush K. Carbapenem-hydrolyzing beta-lactamases. *Antimicrob Agents Chemother.* 1997;41:223–232.
- Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE. Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol.* 2007;372:817–845.
- Frère JM, Galleni M, Bush K, Dideberg O. Is it necessary to change the classification of β -lactamases? *J Antimicrob Chemother.* 2005;55:1051–1053.
- Hall BG, Salipante SJ, Barlow M. The metallo- β -lactamases fall into two distinct phylogenetic groups. *J Mol Evol.* 2003;57:249–254.
- Alderson RG, Barker D, Mitchell JBO. One origin for metallo- β -lactamase activity, or two? An investigation assessing a diverse set of reconstructed ancestral sequences based on a sample of phylogenetic trees. *J Mol Evol.* 2014;79:117–129.
- Brandt C, Braun SD, Stein C, et al. In silico serine β -lactamases analysis reveals a huge potential resistome in environmental and pathogenic species. *Sci Rep.* 2017;7:43232. doi:10.1038/srep43232.
- NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2015;43:D6–D17.
- Burley SK, Berman HM, Christie C, et al. RCSB Protein Data Bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* 2017;27:316–330.
- Srivastava A, Singhal N, Goel M, Viridi JS, Kumar M. CBMAR: a comprehensive β -lactamase molecular annotation resource. *Database.* 2014;2014:bau111.
- Bialek-Davenet S, Criscuolo A, Ailloud F, et al. Genomic definition of hyper-virulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. *Emerg Infect Dis.* 2014;20:1812–1820.
- Singh R, Singh H. DLact: an antimicrobial resistance gene database. *J Comput Intell Bioinform.* 2008;1:93–108.
- Thai QK, Bös F, Pleiss J. The lactamase engineering database: a critical survey of TEM sequences in public databases. *BMC Genomics.* 2009;10:390.
- Widmann M, Pleiss J, Oelschlaeger P. Systematic analysis of metallo- β -lactamases using an automated database. *Antimicrob Agents Chemother.* 2012;56:3481–3491.
- McArthur AG, Waglechner N, Nizam F, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother.* 2013;57:3348–3357.
- Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26:680–682.
- MaxCluster: a tool for protein structure comparison and clustering. <http://www.sbg.bio.ic.ac.uk/maxcluster/index.html>. Accessed March, 2016.
- Wei D, Jiang Q, Wei Y, Wang S. A novel hierarchical clustering algorithm for gene sequences. *BMC Bioinformatics.* 2012;13:174.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–1797.
- Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7:e1002195.
- Sillitoe I, Lewis TE, Cuff A, et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* 2015;43:D376–D381.
- Sinha S, Lynn AM. HMM-Mode: implementation, benchmarking and validation with HMMER3. *BMC Res Notes.* 2014;7:483.

32. Kumar S, Stecher G, Peterson D, Tamura K. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*. 2012;28:2685–2686.
33. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–D169.
34. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
35. Triant DA, Pearson WR. Most partial domains in proteins are alignment and annotation artifacts. *Genome Biol*. 2015;16:99.
36. Finn RD, Coghill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279–D285.
37. Pratap S, Katiki M, Gill P, Kumar P, Golemi-Kotra D. Active-site plasticity is essential to carbapenem hydrolysis by OXA-58 class D β -lactamase of *Acinetobacter baumannii*. *Antimicrob Agents Chemother*. 2016;60:75–86.
38. Mukherjee S, Stamatis D, Bertsch J, et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res*. 2017;45:D446–D456.
39. Urbach C, Evrard C, Pudzaitis V, Fastrez J, Soumillon P, Declercq JP. Structure of PBP-A from *Thermosynechococcus elongatus*, a penicillin-binding protein closely related to class A beta-lactamases. *J Mol Biol*. 2009;386:109–120.
40. Boudreau MA, Fishovitz J, Llarrull LI, Xiao Q, Mobashery S. Phosphorylation of BlaR1 in manifestation of antibiotic resistance in methicillin-resistant *Staphylococcus aureus* and its abrogation by small molecules. *ACS Infect Dis*. 2016;1:454–459.
41. Wilke MS, Hills TL, Zhang HZ, Chambers HF, Strymadka NCJ. Crystal structures of the Apo and penicillin-acylated forms of the BlaR1 beta-lactamase sensor of *Staphylococcus aureus*. *J Biol Chem*. 2004;279:47278–47287.
42. Dubois D, Baron O, Cougnoux A, et al. ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *J Biol Chem*. 2011;286:35562–35570.
43. Proenca R, Niu WW, Cacalano G, Prince A. The *Pseudomonas cepacia* 249 chromosomal penicillinase is a member of the AmpC family of chromosomal beta-lactamases. *Antimicrob Agents Chemother*. 1993;37:667–674.
44. Lüthy L, Grütter MG, Mittl PRE. The crystal structure of *Helicobacter pylori* cysteine-rich protein B reveals a novel fold for a penicillin-binding protein. *J Biol Chem*. 2002;277:10187–10193.
45. Singh R, Saxena A, Singh H. Identification of group specific motifs in beta-lactamase family of proteins. *J Biomed Sci*. 2009;16:109.
46. Sigrist CJA, De Castro E, Cerutti L, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013;41:344–347.
47. Holliday GL, Andreini C, Fischer JD, et al. MACiE: exploring the diversity of biochemical reactions. *Nucleic Acids Res*. 2012;40:783–789.
48. Feuermann M, Gaudet P, Mi H, Lewis SE, Thomas PD. Large-scale inference of gene function through phylogenetic annotation of gene ontology terms: case study of the apoptosis and autophagy cellular processes. *Database*. 2016;2016:baw155.
49. Fróes AM, da Mota FF, Cuadrat RRC, D'Avila AMR. Distribution and classification of serine β -lactamases in Brazilian hospital sewage and other environmental metagenomes deposited in public databases. *Front Microbiol*. 2016;7:1790.
50. Allen HK, Moe LA, Rodbummer J, Gaarder A, Handelsman J. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J*. 2009;3:243–251.
51. Oh HM, Kang I, Vergin KL, Lee K, Giovannoni SJ, Cho JC. Genome sequence of *Oceanicaulis* sp. strain HTCC2633, isolated from the western Sargasso Sea. *J Bacteriol*. 2011;193:317–318.
52. Galleni M, Lamotte-brasseur J, Rossolini GM. Standard numbering scheme for class B beta-lactamases. *Society*. 2001;45:660–663.
53. Hall BG, Barlow M. Evolution of the serine beta-lactamases: past, present and future. *Drug Resist Updat*. 2004;7:111–123.
54. Bebrone C. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem Pharmacol*. 2007;74:1686–1701.
55. Silveira MC, Albano RM, Asensi MD, Carvalho-Assef AP. Description of genomic islands associated to the multidrug-resistant *Pseudomonas aeruginosa* clone ST277. *Infect Genet Evol*. 2016;42:60–65.
56. Toth M, Antunes NT, Stewart NK, et al. Class D β -lactamases do exist in Gram-positive bacteria. *Nat Chem Biol*. 2016;12:9–14.
57. Berglund F, Marathe NP, Österlund T, et al. Identification of 76 novel B1 metallo- β -lactamases through large-scale screening of genomic and metagenomic data. *Microbiome*. 2017;5:134.
58. Kielak AM, Barreto CC, Kowalchuk GA, Van Veen JA, Karumae EE. The ecology of Acidobacteria: moving beyond genes and genomes. *Front Microbiol*. 2016;7:744.
59. Weis AM, Storey DB, Taff CC, et al. Genomic comparison of *Campylobacter* spp. and their potential for zoonotic transmission between birds, primates, and livestock. *Appl Environ Microbiol*. 2016;82:7165–7175.