



UNIVERSIDADE FEDERAL DO SUL E SUDESTE DO PARÁ
INSTITUTO DE ESTUDOS EM SAÚDE E BIOLÓGICAS
FACULDADE DE BIOLOGIA

WENDERSON FELIPE COSTA RODRIGUES

ANÁLISE *IN SILICO* DE GENES DIFERENCIALMENTE EXPRESSOS
NA REGULAÇÃO DE VIAS METABÓLICAS EM PIMENTA-DO-
REINO ACOMETIDA POR FUSARIOSE

Marabá, PA

2018

WENDERSON FELIPE COSTA RODRIGUES

**ANÁLISE *IN SILICO* DE GENES DIFERENCIALMENTE EXPRESSOS
NA REGULAÇÃO DE VIAS METABÓLICAS EM PIMENTA-DO-
REINO ACOMETIDA POR FUSARIOSE**

Trabalho de conclusão de curso apresentado a Faculdade de Biologia da Universidade Federal do Sul e Sudeste do Pará, como requisito parcial para a obtenção de grau de Bacharel em Ciências Biológicas.

Orientador(a): Prof.^a Dr.^a Edith Cibelle de Oliveira Moreira.

Co-orientador(a): Prof.^o Me. Jersey Heitor da Silva Maués.

Marabá, PA

2018

Dados Internacionais de Catalogação-na-Publicação (CIP)
Biblioteca Setorial Campus do Tauarizinho

Rodrigues, Wenderson Felipe Costa

Análise *in silico* de genes diferencialmente expressos na regulação de vias metabólicas em pimenta-do-reino acometida por fusariose / Wenderson Felipe Costa Rodrigues ; orientadora, Edith Cibelle de Oliveira Moreira ; Coorientadora, Jersey Heitor da Silva Maués. — Marabá : [s. n.], 2018.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal do Sul e Sudeste do Pará, Campus Universitário de Marabá, Instituto de Estudos em Saúde e Biológicas, Faculdade de Biologia, Curso de Bacharel em Ciências Biológicas, 2018.

1. Pimenta-do-reino. 2. *Fusarium solani*. 3. Biologia computacional. I. Moreira, Edith Cibelle de Oliveira, orient. II. Maués, Jersey Heitor da Silva. coorient. III. Título.

CDD: 22. ed.: 633.84

AGRADECIMENTOS

Em primeiro lugar, sem dúvida alguma, os meus primeiros agradecimentos são dedicados aos responsáveis por esse projeto ter sido realizado, os meus orientadores: Prof.^a Dr.^a Cibelle Moreira e Prof.^o Me. Jersey Maués. Além de agradecido, me sinto lisonjeado em receber a orientação de pessoas incríveis em termos de personalidade, caráter e inteligência. Prof. Cibelle, eu agradeço por me receber de braços abertos para esse desafio e ter me ensinado parte da bioinformática. Nunca fui o aluno exemplar, mas sempre respeitei tudo o que você representa para mim e me sinto feliz em ter absorvido parte do seu grandioso conhecimento. Prof. Jersey, eu agradeço *principalmente* pela paciência e pelo exemplo de profissional que mostrou ser durante minha orientação. Me vi diante da bioinformática faz nove meses e ainda tenho muito o que aprender, portanto, me desculpo imensamente para ambos por qualquer decepção.

Agradeço imensamente ao Dr.^o Luis Willian Pacheco Arge e ao Laboratório Nacional de Computação Científica (LNCC) pelo processamento dos dados analisados nesse trabalho.

Não poderia deixar de agradecer a orientação que recebi durante dois anos da minha vida dentro do grupo de Biologia Vegetal da UNIFESSPA. Prof.^a Zanderluce Gomes, obrigado por todo conhecimento em botânica que você me proporcionou e, acima de tudo, do conhecimento de vida que me passou pela sua grande amizade.

Não tem como deixar de agradecer a todos os professores da Faculdade de Biologia, em especial ao Prof. Diógenes Silva e Prof. Sidnei Cerqueira por confiarem em mim.

Agradeço aos meus familiares, pai Edson, mãe Daniele e irmão Wadson, por me apoiarem em qualquer situação ou decisão que eu faça. Pai e Mãe, sem vocês eu não estaria escrevendo esse trabalho, vocês são responsáveis pelas minhas conquistas. Obrigado.

Esse último parágrafo, eu dedico em agradecimento a todos os meus amigos, começando por aqueles que me acompanham desde os dez anos de idade: Aline Lima, Fernanda Carolina, Ângelo Rafael e Brenda Carolline, vocês são minha saudade, meu laço eterno. Aline, você é minha estrela, nem sempre te vejo ou te converso, mas sei que sempre estará lá. Agradeço também ao meu eterno companheiro: João Luiz, que me acompanhou e foi minha família durante muito tempo, sem você talvez eu estivesse na matemática. Agradeço também ao CALICY: Marciely Lira e Camila Rodrigues, pela incrível amizade que pretendo cultivar pela eternidade, obrigado por me aturarem, de coração. Por fim, agradeço a família “*bio*” que fiz nesses 4 anos: Juan David, Elis Rejaine, Jhennifer Gomes, Thalyta Silva, Lorena Paz e Danielly Castro, obrigado por serem motivo da minha felicidade.

RESUMO

A pimenta-do-reino (*Piper nigrum* L.) é uma especiaria de grande importância econômica para o estado do Pará, porém sua produção vem sendo reduzida em função da doença denominada fusariose, causada pelo fungo *Fusarium solani* f. sp. *piperis*. Análises de bioinformática *in silico* de sequências de transcriptomas da pimenta-do-reino estão sendo estudados em larga escala para fornecer conhecimentos importantes do ponto de vista biotecnológico sobre a relação planta-patógeno, afim de fornecer apoio a programas de melhoramento genético. Dessa forma, o presente trabalho tem como objetivo analisar os perfis de genes diferencialmente expressos (DEG) do transcriptoma da pimenta-do-reino acometida por fusariose e relacionar o papel desses na regulação de vias metabólicas de interesse biotecnológico. Para isso foram realizadas análises *in silico* utilizando programas computacionais principalmente de expressão diferencial e análise funcional. Por meio da análise da expressão diferencial foi possível identificar 2.108 DEGs com a expressão diminuído (*downregulated*) e outros 1.642 DEGs com o perfil de expressão aumentado (*upregulated*). Esses DEGs foram usados para uma anotação funcional a parte com a ferramenta Blast2GO que resultou em um enriquecimento de 2.535 sequências anotadas com GO e 94 vias metabólicas identificadas na base de dados *Kyoto Encyclopedia of Gene and Genome* (KEGG), que foram mapeadas para 550 sequências do transcriptoma da planta. Foi identificado um gene *downregulated* responsável pela defesa da planta contra o fungo: o *PCK*, que regula vias metabólicas da gliconeogênese, fixação do carbono, ciclo de Krebs, metabolismo do piruvato e biossíntese de antibióticos. Análises do transcriptoma da pimenta-do-reino tornam-se necessária, pois visam elucidar o processo de interação planta-patógeno respondendo questões sobre a defesa dessa planta em resposta a infecção.

Palavras-chave: *Piper nigrum* L.; *Fusarium solani* f. sp. *piperis*; bioinformática; Blast2GO.

ABSTRACT

The black pepper (*Piper nigrum* L.) is a spice of great economic importance for the state of Pará, but its production has been reduced because of the disease called fusariosis, caused by the fungus *Fusarium solani* f. sp. *piperis*. *In silico* bioinformatics analyzes of black pepper transcriptome sequences are being studied on a large scale to provide important biotechnological knowledge about the plant-pathogen relationship in order to provide support to genetic breeding programs. Thus, the present work aims to analyze the profiles of differentially expressed genes (DEG) of the transcriptome of the black pepper affected by fusariosis and to relate their role in the regulation of metabolic pathways of biotechnological interest. For this, *in silico* analyzes were performed using computer programs mainly for differential expression and functional analysis. By means of differential expression analysis it was possible to identify 2,108 DEGs with downregulated expression and another 1,642 DEGs with upregulated expression profile. These DEGs were used for a separate functional annotation with the Blast2GO tool, that resulted in an enrichment of 2,535 GO annotated sequences and 94 metabolic pathways identified in the Kyoto Encyclopedia of Gene and Genome (KEGG) database, which were mapped to 550 sequences of the plant transcriptome. A downregulated gene responsible for plant defense against fungi was identified: PCK, which regulates metabolic pathways of gluconeogenesis, carbon fixation, citrate cycle, pyruvate metabolism and antibiotic biosynthesis. Black pepper transcriptome analysis are necessary because they aim to elucidate the process of plant-pathogen interaction by answering questions about the defense of this plant in response to infection.

Keywords: *Piper nigrum* L.; *Fusarium solani* f. sp. *piperis*; bioinformatics; Blast2GO.

LISTA DE ILUSTRAÇÕES

Figura 1 – Custo do sequenciamento por genoma no período de 2001 a 2017.....	15
Figura 2 – Crescimento de dados em Terabases armazenados no banco de dados <i>Sequence Read Archive</i> (SRA) do <i>National Center for Biotechnology Information</i> (NCBI), no período de 2009 a 2018. O SRA é referência para busca de sequências brutas geradas pelas NGS.....	16
Figura 3 – Pimenta-do-reino (<i>P. nigrum</i> L.)	19
Figura 4 – Fluxograma descritivo das sequências de análises realizadas para esse trabalho. Caixas retangulares com bordas arredondadas representam as análises realizadas e as caixas retangulares com bordas pontiagudas representam os resultados das análises.....	26
Figura 5 – Gráfico de dispersão <i>Volcano plot</i> dos genes diferencialmente expressos de pimenta-do-reino. No eixo y está representado o valor de <i>Fold Change</i> e no eixo x o valor de p. Pontos azuis são genes regulados negativamente e vermelhos genes regulados positivamente. Pontos presentes no topo do gráfico e distantes do centro representam os genes de interesse que foram filtrados.....	27
Figura 6 – Distribuição de dados após consulta de sequências (<i>hits</i>) em bancos de dados.	29
Figura 7 – Distribuição de similaridade das sequências contra o banco de dados.....	30
Figura 8 – Distribuição da abundância de <i>top-hits</i> por espécie na análise de BlastX para o transcriptoma da pimenta-do-reino.....	30
Figura 9 – Distribuição do número de anotações (termos-GO) recuperadas dos diferentes bancos de dados.....	31
Figura 10 – Distribuição da quantidade de termos candidatos à ontologia de gene atribuídos a cada sequência durante a etapa de mapeamento de GO.....	32
Figura 11 – Classificação dos transcritos no GO quanto a Processos Biológico em pimenta-do-reino.....	32
Figura 12 – Classificação dos transcritos no GO quanto a Componentes Celulares em pimenta-do-reino.....	33
Figura 13 – Classificação dos transcritos no GO quanto a Função Molecular em pimenta-do-reino.....	34
Figura 14 – Distribuição da abundância de classes enzimáticas de acordo com banco de dados disponível em KEGG.....	36
Figura 15 – Nível de expressão das sequências envolvidas em vias de biossíntese de antibióticos da pimenta-do-reino.....	37
Figura 16 – Via metabólica da gliconeogênese em pimenta-do-reino obtida no Blast2GO. Retângulos coloridos representam a diversidade enzimática obtida a partir dos transcritos da pimenta-do-reino, cada cor representa uma enzima diferente. O quadrado na cor verde representa a enzima PEPCCK na via.....	40

Figura 17 - Via metabólica da fixação do carbono em organismos fotossintéticos de pimenta-do-reino obtida no Blast2GO. Retângulos coloridos representam a diversidade enzimática obtida a partir dos transcritos da pimenta-do-reino, cada cor representa uma enzima diferente. O quadrado na cor verde representa a enzima PEPCK na via..... 41

Figura 18 – Mapa global da via metabólica da biossíntese de antibióticos obtida no Blast2GO. Linhas representam enzimas e reações e círculos representam metabólitos. Cores representam categorias funcionais do KEGG, como processos celulares em verde. A linha preta representa a enzima PEPCK no mapa global..... 43

LISTA DE TABELAS

Tabela 1 – Diferenças no método de sequenciamento em relação ao tamanho das leituras (pb), custo por base, número de leituras por corrida e número de dados gerados por corrida.....	14
Tabela 2 – Sumário das análises realizadas dos DEGs.....	28
Tabela 3 – Resultado do mapeamento das vias enzimáticas do transcriptoma da pimenta-do-reino, através do Blast2GO-KEGG.....	35
Tabela 4 – Sumário de algumas vias metabólicas identificadas no transcriptoma de pimenta-do-reino, através do Blast2GO-KEGG.....	38
Tabela 5 – Resultado dos DEGs downregulated e upregulated identificados pelo KEGG.....	38

LISTA DE ABREVIATURAS E SIGLAS

AR – Regra de Anotação

AS – Escore de Anotação

AT – Termo de Abstração

BLAST – Alinhamento Local Básico

CC – Componente Celular

DEGs – Genes Diferencialmente Expressos

DNA – Ácido Desoxirribonucleico

DT – Termo Direto

EC – Código de Evidência da Anotação

EC – Código Enzimático

FC – *Fold Change*

FM – Função Molecular

FPKM – Fragmentos por Quilobase de Exon por Milhões de Fragmentos Mapeados

Gb - *Gigabyte*

GO – *Gene Ontology*

ID – Identidade

Kb - *Kilobyte*

KEGG – Enciclopédia de Kyoto de Genes e Genomas

LSPA – Levantamento Sistemático da Produção Agrícola

mRNA – RNA mensageiro

NCBI – *National Center for Biotechnology Information*

NGS – Sequenciamento de Nova Geração

NHGRI – *National Human Genome Research Institute*

nr – não redundante

pb – pares de bases

PB – Processo Biológico

PCA – Análise de Componentes Principais

PCK – Gene Fosfoenolpiruvato

PDB – *Protein Data Bank*

PEPCK – Enzima Fosfoenolpiruvato Carboxiquinase

PRINSEQ - *Preprocessing and Information of Sequences*

RNA – *Ácido Ribonucleico*

SOLiD – *Sequencing by Oligonucleotide Ligation and Detection*

SRA – *Sequence Read Archive*

TAIR – *The Arabidopsis Information Resource*

UniProt – *Universal Protein*

UniProtKB – *The UniProt Knowledgebase*

SUMÁRIO

1. INTRODUÇÃO	12
2. REVISÃO BIBLIOGRÁFICA	14
2.1. EVOLUÇÃO DOS MÉTODOS DE SEQUENCIAMENTO.....	14
2.2. ANÁLISE BIOINFORMÁTICA E POSSÍVEIS RESULTADOS	16
2.3. PIMENTA-DO-REINO (<i>Piper Nigrum</i> L).....	18
3. OBJETIVOS	21
3.1. OBJETIVO GERAL.....	21
3.2. OBJETIVOS ESPECÍFICOS	21
4. MATERIAL E MÉTODOS	22
4.1. IDENTIFICAÇÃO <i>IN SILICO</i> E <i>SCREEN</i> DE DEGS	22
4.2. ENRIQUECIMENTO DE DEGS.....	23
4.3. CÓDIGO ENZIMÁTICO E KEGG	24
4.4. ESTATÍSTICAS.....	25
4.5. SELEÇÃO DE DEG RELATADO NA DEFESA/RESISTÊNCIA E ANÁLISE DE VIAS METABÓLICAS.....	25
5. RESULTADOS E DISCUSSÃO	27
5.1. <i>SCREEN</i> E COMPARAÇÃO DE DEGS APÓS A INFECÇÃO	27
5.2. ANÁLISE DOS PERFIS DE EXPRESSÃO DE DEGS	28
5.3. ANÁLISE QUANTITATIVA DA DISTRIBUIÇÃO DOS DADOS DE ANOTAÇÃO COM O BLAST2GO.....	29
5.4. ONTOLOGIA GENÉTICA DOS DEGS	32
5.5. RECUPERAÇÃO DAS VIAS METABÓLICAS	35
5.6. SELEÇÃO DE DEGS RELATADOS NA DEFESA/RESISTÊNCIA E ANÁLISE DE VIAS METABÓLICAS	38
5.5.1. Fosfoenolpiruvato (<i>PCK</i>)	39
6. CONSIDERAÇÕES FINAIS	44
REFERÊNCIAS	45

1. INTRODUÇÃO

Após o sequenciamento do genoma humano utilizando metodologia de Sanger, houve um aumento na demanda por métodos que fossem mais rápidos e baratos para obtenção de genomas completos. Essa demanda estimulou o desenvolvimento de plataformas de sequenciamento de nova geração, também chamadas NGS (*Next Generation Sequencing*) ou de segunda geração, que executariam o sequenciamento em menos tempo e com um custo menor (GRADA; WEINBRECHT, 2013).

Atualmente, usando plataformas de nova geração, os estudos de sequenciamento de genomas utilizam duas estratégias: a genômica estrutural e a funcional. A genômica estrutural visa sequenciar e montar o genoma, consistindo em fragmentar aleatoriamente o DNA genômico (*shotgun*), gerando uma grande lista de sequências nucleotídicas (*reads*) que são posteriormente montadas pelas suas sobreposições, afim de identificar regiões comuns para a construção de fragmentos contíguos (*contigs*), que são sequências contínuas e únicas (STEINDORFF, 2016; VERLI, 2014; POPIN, 2017). A genômica funcional ou transcriptômica, consiste no sequenciamento e mapeamento dos transcritos gênicos. Alguns objetivos desse tipo de sequenciamento incluem catalogar mRNAs, RNA não codificantes e pequenos RNAs e quantificar níveis de expressão gênica (WANG; GERSTEIN; SNYDER, 2009).

Vários trabalhos utilizam a metodologias NGS para estudo de transcriptoma, permitindo verificar a função dos transcritos, suas isoformas, analisar alterações dos níveis de expressões e a identificação das vias metabólicas que são alteradas em resposta a situações fisiológicas na qual as células ou tecidos se encontram (VERLI, 2014).

Os trabalhos de genômica e transcriptômica utilizando NGS geram um grande volume de dados, que precisam ser analisados para permitir a compreensão das especificidades moleculares dos organismos sequenciados. Dentro desse contexto, ferramentas de bioinformática são utilizadas para explorar esses dados, permitindo identificar mecanismos moleculares complexos (LESK, 2008; VERLI, 2014).

Em plantas de interesse biotecnológico, a análise de transcriptoma tem sido usada para verificar a resposta fisiológica e comportamental de uma possível infecção por patógeno, o que permite entender como os mecanismos de defesa das plantas são desencadeados a partir das mudanças no perfil de transcritos observados (GIACHETTO; HIGA, 2014; GANESH et al., 2006).

Os dados gerados por esse tipo de estudos são processados por meio de análises de bioinformática, afim de se obter informações de interesse biotecnológico importantes que apoiem programas de melhoramento genético (CARVALHO; SILVA, 2010).

No Brasil, trabalhos de melhoramento genético de plantas utilizando tecnologias NGS e análises de bioinformática estão sendo realizados continuamente. Exemplo disso são os estudos desenvolvidos através da análise do genoma e do transcriptoma de espécies como o café, a soja e a pimenta-do-reino (VARON et al., 2017; SOUZA, 2018; MOREIRA et al., 2017).

A pimenta-do-reino (*Piper nigrum* L.), possui grande importância econômica para o estado do Pará. Essa espécie, originária da Índia, adaptou-se facilmente aos solos amazônicos e tem grande relevância para o agronegócio paraense, pois gera lucro para a região (GORDO, 2012; LEMOS et al., 2011).

A produção de pimenta-do-reino tem sido reduzida em função de uma doença denominada fusariose, ocasionada pelo fungo *Fusarium solani* f. sp. *piperis*, que afeta a parte aérea da planta e o seu sistema radicular levando ao apodrecimento e conseqüentemente a morte do organismo (FILGUEIRAS, HOMMA, SANTOS, 2007; CHU et al., 1997).

O uso de estratégias de melhoramento clássico, como o uso de fungicidas e a erradicação de plantas doentes apresentam apenas soluções paliativas no combate a fusariose. Assim, objetivando apoiar programa de melhoramento genético da pimenta-do-reino, foram realizados trabalhos com RNA-Seq a partir de NGS para sequenciar o transcriptoma da planta, resultando na geração de um grande volume de dados e a identificação de numerosos genes e isoformas (OLIVEIRA, 2015; MOREIRA et al., 2017).

Esses dados necessitam ser explorados por meio de análises de bioinformática, afim de se obter um catálogo (*screen*) dos Genes Diferencialmente Expressos (DEGs) que possam revelar o papel destes na regulação de vias metabólicas da pimenta-do-reino, uma vez que essas vias são representações das transformações químicas que ocorrem nas células e a diferenciação na expressão gênica fornece o entendimento sobre a função dos genes e dos processos biológicos que participam (MOREIRA et al., 2017; NELSON; COX, 2014).

Diante disso, o presente estudo pretende analisar *in silico* os perfis de DEGs do transcriptoma da pimenta-do-reino acometida por fusariose e relacionar o papel desses genes na regulação de vias metabólicas, fornecendo conhecimentos importantes do ponto de vista biotecnológico para apoiar programas de melhoramento genético da espécie.

2. REVISÃO BIBLIOGRÁFICA

2.1. EVOLUÇÃO DOS MÉTODOS DE SEQUENCIAMENTO

O sequenciamento do DNA foi proposto por Frederick Sanger e colaboradores em 1977 e esse método revolucionou as pesquisas científicas, pois permitia determinar a ordem dos nucleotídeos que codificam a informação genética (MARGULIES et al., 2005).

A técnica de sequenciamento foi automatizada em 1986, utilizando o mesmo princípio de sequenciamento de Sanger, porém de forma mais rápida e simples e com o auxílio de um computador (THAMPI, 2009).

O método automatizado de sequenciamento foi de grande importância para o surgimento do Projeto Genoma Humano, em 1990. Esse projeto possuía como objetivo a identificação, na ordem correta, dos 3 bilhões de nucleotídeos que compõe o genoma humano e durou mais de uma década, sendo finalizado apenas em 2003 (CORRÊA, 2002; WIECZOREK; LEAL, 2002).

Com a publicação das sequências do genoma humano e o sequenciamento paralelo de milhares de genomas de diferentes espécies, as tecnologias de sequenciamento avançaram rapidamente culminando no surgimento dos sequenciadores NGS (*Next Generation Sequencing*). Esses sequenciadores são capazes de gerar milhares de informações em uma única corrida, com elevada cobertura, acurácia, redução de custo e tempo, em relação ao método de Sanger (Tabela 1) (PAVLOPOULOS et al., 2013).

Tabela 1 – Diferenças no método de sequenciamento em relação ao tamanho das leituras (pb), custo por base, número de leituras por corrida e número de dados gerados por corrida.

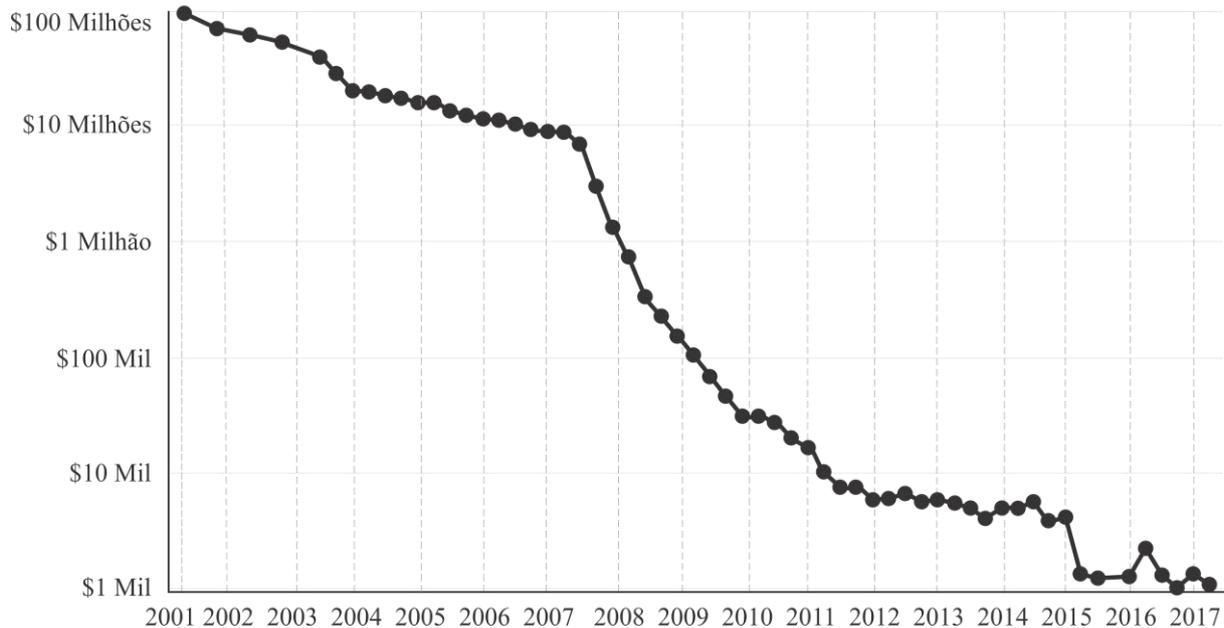
	Sequenciamento de Sanger	Plataformas NGS
Tamanho das Leituras (pares de bases)	~650 pb	35–100 pb
Custo por base	\$ 1.915,65	\$ 2,17
Nº de leituras por corrida	384	8x 10.000.000
Nº de dados gerados por corrida	290 Kb	~7 Gb

Fonte: Adaptado de Weiss, 2010.

Desde 2001 o custo do sequenciamento por genoma está decaindo de forma acelerada graças aos avanços do sequenciamento NGS. Nesse mesmo ano, segundo o *National Human Genome Research Institute* – NHGRI (2017), o preço equivalente ao sequenciamento de um

genoma custava em média 100 milhões de dólares, porém esse preço diminuiu para quase mil dólares até 2017 (Figura 1) (NHGRI, 2017).

Figura 1 – Custo do sequenciamento por genoma no período de 2001 a 2017.

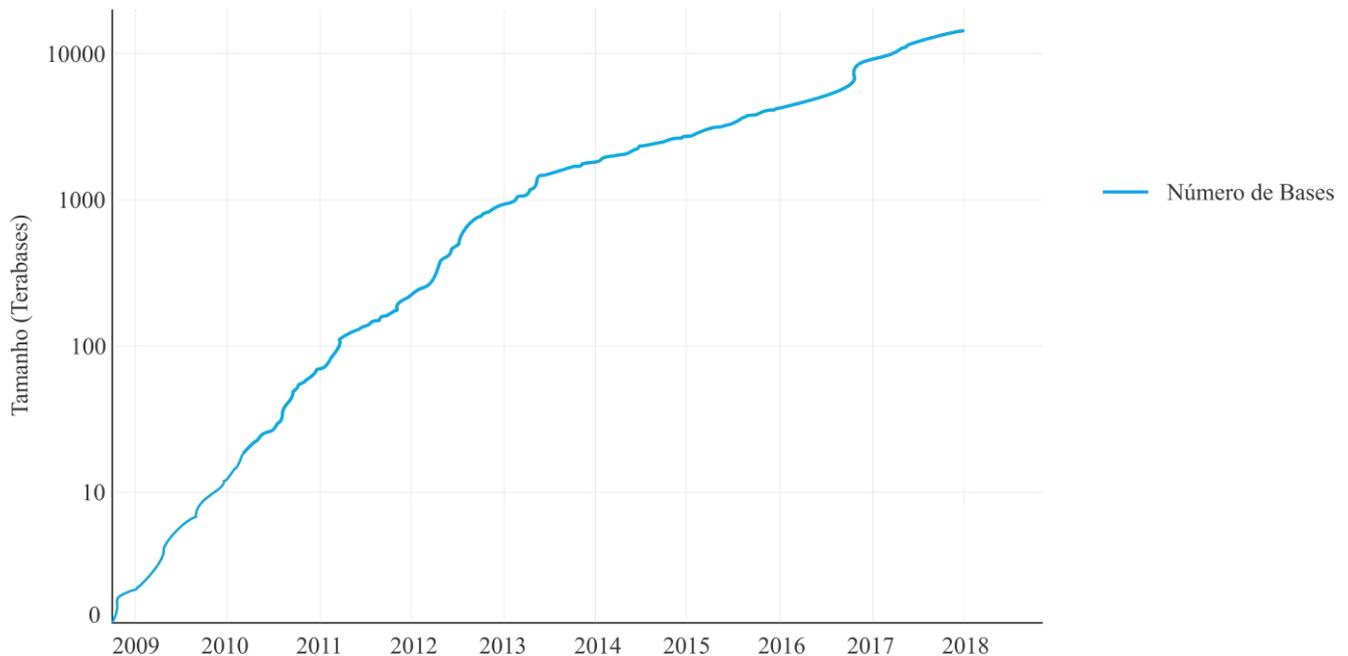


Fonte: Adaptado de NHGRI, 2017.

Os sequenciadores NGS revolucionaram o campo da biologia molecular, pois permitiu o estudo de genoma, transcriptoma, proteoma e metaboloma em larga escala, dando espaço para o aparecimento do conjunto de técnicas biológicas denominado ômicas, que estudam os processos celulares e moleculares do DNA, RNA's, das proteínas e dos metabólitos. Causaram, portanto, uma mudança de paradigma, onde a geração de dados deixa de ser o fator limitante para se estudar amplamente os organismos a níveis moleculares e o processamento e a análise dos dados tornaram-se o maior desafio das pesquisas (CARVALHO; SILVA, 2010).

A grande quantidade de dados e sequências depositadas nos bancos de dados biológicos aumentaram drasticamente nos últimos anos (Figura 2) e impulsionaram a difusão da bioinformática no meio científico, uma ciência interdisciplinar que une conhecimentos de biologia molecular, matemática e informática. Essa ciência realiza a análise e interpretação de dados através do desenvolvimento de algoritmos com comandos específicos, capazes de analisar sequências, realizar sua montagem, anotação funcional, entre outras funcionalidades (THAMPI, 2009).

Figura 2 – Crescimento de dados em Terabases armazenados no banco de dados *Sequence Read Archive (SRA)* do *National Center for Biotechnology Information (NCBI)*, no período de 2009 a 2018. O SRA é referência para busca de sequências brutas geradas pelas NGS.



Fonte: Adaptado de NCBI, 2018.

Por meio da bioinformática é possível compreender o mundo molecular e entender os processos biológicos com mais eficiência, tais como mudanças na expressão gênica, interações entre proteínas e genes, estrutura de peptídeos, vias metabólicas, vias de sinalização, entre outros processos moleculares (CAN, 2014; VASSILEV et al., 2005).

Com a difusão dos sequenciadores NGS, a bioinformática tornou-se fundamental nos estudos genéticos de melhoramento de animais e vegetais de interesse econômico (GIACHETTO; HIGA, 2014).

No caso dos vegetais, a bioinformática tem sido empregada como uma ferramenta capaz de revelar as diversas funções desempenhadas pelos genes das plantas, entender o funcionamento de seu metabolismo e compreender as respostas fisiológicas em relação a uma infecção por patógenos (VASSILEV et al., 2005).

2.2. ANÁLISE BIOINFORMÁTICA E POSSÍVEIS RESULTADOS

O uso de dados genômicos, transcriptômicos e das outras ômicas no melhoramento genético de plantas possui grande importância econômica e biológica, na medida que esses

possibilitam um aumento na produtividade e a melhoria de várias espécies (GIACHETTO; HIGA, 2014).

O primeiro vegetal a ter seu genoma sequenciado foi a planta-modelo *Arabidopsis thaliana*, no ano 2000. Após isso, várias outras espécies tiveram seus genomas descobertos pelas plataformas NGS, tais como o a batata, a banana, a melancia e a soja (XU et al., 2011; D'HONT et al., 2012; XU et al., 2013; XIA et al., 2012). As sequências obtidas dessas espécies e de muitas outras foram interpretadas, após sua montagem, através do processo da bioinformática chamado anotação funcional. Essa metodologia consiste na identificação e caracterização de genes, promotores e outros elementos funcionais do DNA por meio de softwares, que predizem as regiões gênicas no genoma e atribuem funções a essas (BOLGER et al., 2017).

Outra maneira de se obter os elementos funcionais e seus constituintes moleculares, sem precisar saber a sequência do genoma inteiro, é através da análise de transcriptomas. O transcriptoma é o conjunto de todos os transcritos gênicos (mRNA) presentes em uma célula, em um estágio de desenvolvimento e em condições fisiológicas específicas e, atualmente, vem sendo analisado por meio das tecnologias amplamente utilizadas como as de microarranjo de DNA (*microarrays*), que possibilita a avaliação simultânea da expressão de vários genes em diferentes tecidos de um determinado organismo, e do RNA-Seq, metodologia que utiliza o sequenciamento NGS para obtenção de transcritos de uma célula ou tecido (JÚNIOR et al., 2005; WALSH; HENDERSON, 2004; WANG; GERSTEIN; SNYDER, 2009).

Após a recuperação das sequências de transcritos gênicos a anotação funcional é uma etapa importante para o entendimento do funcionamento do organismo vivo, pois é nesse processo que programas computacionais atribuem funções aos genes identificados (BOLGER et al., 2017).

Um método de anotação é a partir da utilização do *software* Blast2GO (CONESA; GOTZ, 2008), uma ferramenta com interface Java, desenvolvida para a análise funcional de sequências gênicas. Esse *software* utiliza o sistema de ontologias *Gene Ontology* (*THE GENE ONTOLOGY CONSORTIUM*, 2000), afim de padronizar a representação dos genes para todos os seres vivos em três categorias: Processo Biológico (atividade biológica com o qual o gene contribui), Função Molecular (atividade bioquímica do gene) e Componente Celular (local na célula onde o gene é ativo). Além disso, o Blast2GO permite fazer buscas por sequências similares, códigos de enzimas e visualizar mapas metabólicos onde os genes e seus produtos atuam através do KEGG – Enciclopédia de Kyoto de Genes e Genomas (KANEHISA; GOTO,

2000) para, a partir das sequências estudadas, determinar suas possíveis funções e ontologias (NODA; DAMASCENO; SOUSA, 2005).

A anotação de genes é realizada automaticamente e fornece informações importantes do organismo vivo. As sequências funcionais da batata por exemplo, revelaram fatores de transcrição que regulam a maturidade da planta e seu ciclo de vida (KLOOSTERMAN et al., 2013). No caso da soja, através da anotação foi descoberto o *locus* E1, região que controla o tempo de floração em adaptação a diferentes fotoperíodos (XIA et al., 2012). Essas informações de modo geral, são de extrema significância para o melhoramento dessas espécies e de tantas outras, uma vez que o entendimento do desenvolvimento da planta é revelado (JÚNIOR et al., 2005).

Além da anotação funcional, análises de transcriptomas fornecem informações dos níveis de regulação da expressão de genes do organismo vivo. Nesse caso, o sequenciamento de transcritos apresenta precisão em suas alterações nos níveis de expressões sob diferentes condições, promovendo assim, monitoramento de mudanças dos níveis de expressão gênica e a identificação das vias metabólicas que são alteradas por ação externa, como uma infecção por patógeno, devastação por parasitas e outros tipos de estresse (WANG; GERSTEIN; SNYDER, 2009; GIACHETTO; HIGA, 2014).

Por esses motivos, a análise do transcriptoma vem sendo utilizada especialmente em espécies de plantas comercializadas no Brasil, como a cana-de-açúcar, o café e a pimenta-do-reino, pois garantem melhoria nos seus manejos, em função da produtividade e da qualidade do produto (CARDOSO-SILVA et al., 2014; VARON et al., 2017; MOREIRA, 2012).

2.3. PIMENTA-DO-REINO (*Piper Nigrum* L).

A pimenta-do-reino (*Piper nigrum* L.) (Figura 3) é uma especiaria de grande importância econômica para o Brasil. Originária da Índia, essa espécie se adaptou facilmente ao solo ácido amazônico e ao clima da região norte do país (GORDO, 2012; LEMOS et al., 2011).

Figura 3 – Pimenta-do-reino (*P. nigrum* L.)



Fonte: Jahmu PBC ©, 2018.

No Brasil, o estado que mais produz pimenta-do-reino é o Pará. Segundo dados do Relatório do Levantamento Sistemático da Produção Agrícola do Estado do Pará – LSPA (2016), o estado dispõe de 90% da produção brasileira de pimenta-do-reino, possuindo cerca de 23 mil hectares de área plantada e produção de aproximadamente 34 mil toneladas de pimenta-do-reino por ano.

Desde a década de 50 os produtores locais passaram a ter seus cultivos limitados devido a Fusariose, doença causada pelo fungo *Fusarium solani* f. sp. *piperis*. Nessa mesma época, a doença se alastrou rapidamente na região, destruindo em pouco tempo, grande parte dos cultivos (FILGUEIRAS; HOMMA; SANTOS, 2007; ANDO et al., 1996).

Com o passar dos anos, a fusariose ainda é uma das doenças mais prejudiciais ao cultivo de pimenta-do-reino, sendo cerca de 3% da área cultivada reduzida, ocasionando prejuízo anual de 10 milhões de dólares e diminuição da vida útil da plantação, de 12 para 5 ou 6 anos (BENCHIMOL et al., 2017; LEMOS; TREMACOLDI; POLTRONIERI, 2014).

O conhecimento molecular, juntamente com as técnicas de melhoramento genético e a identificação de genes, através da biotecnologia e bioinformática, são ferramentas fundamentais para resolver a questão da fusariose (BENCHIMOL et al., 2017).

A Embrapa Amazônia Oriental e a Universidade Federal do Pará realizaram um projeto que visava identificar sequências gênicas do transcriptoma da espécie envolvidas no mecanismo de interação entre a pimenta-do-reino e o patógeno *F. solani* f. sp. *piperis*. Essas sequências foram obtidas via plataforma SOLiD de sequenciamento NGS e analisadas via bioinformática (GORDO et al. 2012, MOREIRA et al., 2017).

O alto número de dados gerados por esse projeto ainda necessitam ser explorados com o auxílio da bioinformática, onde essa área promissora tem contribuído muito na avaliação da relação dos genes em resposta a infecção do patógeno responsável pela fusariose, uma vez que métodos como algoritmos genéticos, métodos de agrupamentos, análise de componentes principais (PCA) e outras metodologias, auxiliam no desenvolvimento de estratégias mais efetivas para resolver tais problemáticas biológicas (OLIVEIRA, 2015).

Diante disso, analisar os dados gerados da pimenta-do-reino pelas tecnologias NGS por meio da bioinformática é de grande importância para ajudar a entender o envolvimento de determinados genes nos mecanismos moleculares de defesa da espécie e fornecer suporte na erradicação do fungo que afeta a produção da pimenta-do-reino.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Analisar os perfis de DEGs (Genes Diferencialmente Expressos) do transcriptoma da pimenta-do-reino (*P. nigrum* L.), acometida por fusariose (*F. solani* f. sp. *piperis*) e relacionar o papel desses genes na regulação das vias metabólicas mais importantes na resposta de defesa e resistência da planta.

3.2. OBJETIVOS ESPECÍFICOS

- Avaliar o perfil de expressão dos genes envolvidos em mecanismos moleculares da pimenta-do-reino acometida por fusariose
- Identificar o perfil de anotação funcional de DEGs;
- Investigar as principais vias metabólicas identificadas e reguladas por DEGs;
- Selecionar DEGS envolvidos no mecanismo de resistência.

4. MATERIAL E MÉTODOS

4.1. IDENTIFICAÇÃO *IN SILICO* E *SCREEN* DE DEGS

Os dados das sequências brutas da pimenta-do-reino analisados nesse trabalho, passaram por um pré-processamento descrito por Gordo et al. (2012). Um experimento de sequenciamento de RNA foi usado para estudar o transcriptoma da raiz 60 dias após a infecção pelo patógeno *F. solani* f. sp. *piperis* (MOREIRA et al., 2017).

Os dados das sequências obtidos pelo SOLiD foram gerados de duas bibliotecas contendo as *reads* (leituras) da amostra infectada e da outra amostra controle, os quais foram previamente processadas para limpeza e o controle de qualidade com os *softwares* CutAdapt (0.95) (MARTIN, 2011) e PRINSEQ (0.14.4) (SCHIMIEDER; EDWARDS, 2011), respectivamente (MOREIRA et al., 2017).

Em seguida, as leituras de pimenta-do-reino foram alinhadas contra o genoma de referência com o programa *Bowtie* (LANGMEAD et al., 2009) e os dados subsequentes foram utilizados como entrada no pacote *Cufflinks* (TAPNELL et al., 2010) para estimar a abundância de transcritos que foram normalizados com FPKM (Fragmentos por quilobase de exon por milhões de fragmentos mapeados) (MOREIRA et al., 2017).

Posteriormente, para obtenção da expressão diferencial, utilizou-se o *software* DEGSeq (WANG et al., 2010) do programa R/*Bioconductor*, onde foram considerados apenas os genes que apresentaram *Fold Change* (FC) ≥ 2 e valor de $p \leq 0.001$. O $\text{Log}_2 \text{FC} = (\text{Infectada}/\text{Controle})$ foi usado para estimar a expressão diferencial dos genes (MOREIRA et al., 2017).

Os DEGs obtidos, foram plotados usando o Multiplot v2 (<http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/multiplot/2>), que exibiram os perfis de expressão gênica personalizados, para uma análise de dispersão.

4.2. ENRIQUECIMENTO DE DEGS

As sequências obtidas da pimenta-do-reino foram usadas em formato *fasta* como entrada no *software* Blast2GO (5.0.13) (CONESA; GOTZ, 2008). Essa ferramenta de pesquisa de Alinhamento Local Básico - BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>) foi usada para encontrar sequências semelhantes ao seu conjunto de dados de consultas. Nesse estudo as sequências da pimenta-do-reino foram anotadas com base na execução de um BlastX (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), com um *E-value* de $1e-3$, contra um banco de dados *nr-viridiplantae*. O BlastX busca e identifica potenciais produtos de proteínas a partir da consulta com uma sequência de nucleotídeos.

As anotações de ontologia genética foram feitas usando o GO (*Gene Ontology*) (*THE GENE ONTOLOGY CONSORTIUM*, 2000), os quais foram identificados atuantes em Processo Biológico, Função Molecular e Componente Celular. A anotação GO é executada aplicando uma regra de anotação (*Annotation Rule* – AR) durante os termos da ontologia encontrada. A regra procura encontrar as anotações mais específicas com um certo nível de confiabilidade. Este processo é ajustável em especificidade e rigor. Para cada candidato, um escore de anotação (*Annotation Score* – AS) é calculado. O AS é composto de dois termos aditivos. O primeiro, *Direct Term* (Termo Direto – DT), representa a maior similaridade de acertos deste GO ponderado por um fator correspondente ao seu código de evidência (EC), que pode ser interpretado como um índice de confiabilidade da anotação GO. O segundo termo do AS fornece a possibilidade de abstração (AT). Isso é definido como uma anotação para um nó pai quando vários nós filhos estão presentes na coleção de candidatos GO. Este termo multiplica o número total de GOs unificados no nó por um fator de peso GO definido pelo usuário, que controla a possibilidade e a força da abstração. Quando o peso GO é definido como 0, nenhuma abstração é feita. Finalmente, o AR seleciona o menor termo por ramificação, que fica acima de um limite definido pelo usuário (CONESA et al., 2005). DT, AT e os termos AR são definidos como mostrado abaixo:

$$DT = \max(\text{similarity} \times EC_{\text{weight}}) \quad (1)$$

$$AT = (\#GO - 1) \times GO_{\text{weight}} \quad (2)$$

$$AR : \text{lowest.node}(AS(DT + AT)) \geq \text{threshold} \quad (3)$$

Os parâmetros definidos na configuração de anotação do Blast2GO foram: limite de anotação padrão = 55, onde todos os menores termos por ramificação acima desse limite são

selecionados; *GO-Weight* padrão = 5, esse parâmetro representa o peso dado a contribuição de termos filhos mapeados para a anotação de um termo pai; foi usado ID de taxonomia, onde os termos GO que não estão na taxonomia indicada foram removidos; *E-Value-Hit-Filter* padrão = 1.0E-6, pré-filtro que seleciona apenas os termos GO obtidos de ocorrências com *E-value* maior que o dado; *HSP-HitCoverage CutOff* (limite de cobertura mínima) = 80, estabelecido para definir a cobertura mínima de um *hit* contra a referência, apenas aquelas anotações do *hit* que atenderam a esse critério foram consideradas para transferência de anotação; *Hit Filter* = 500, que permite considerar apenas os primeiros 500 *hits* durante a anotação.

Os resultados de anotação para cada sequência foram visualizados no GO DAG selecionando *Draw Graph of GO-Mapping* com *Score de Anotação* no menu de contexto. Além disso, as opções *Change Annotation and Description* (Alterar Anotação e Descrição Manualmente) deste menu permitiu ajustar aquelas anotações especificamente para uma única sequência. Esta função edita a anotação da sequência selecionada e permite digitar e excluir a anotação ou descrição da sequência.

4.3. CÓDIGO ENZIMÁTICO E KEGG

A etapa de anotação das vias metabólicas do KEGG – *Kyoto Encyclopedia of Gene and Genome* (KANEHISA; GOTO, 2000) da pimenta-do-reino foi feita com Blast2GO o qual forneceu anotação de EC (*Enzyme Code* - Código Enzimático) através do arquivo de mapeamento GO>EC direto disponível no site do GO. Isso significa que apenas sequências com anotações GO eventualmente exibirão também números EC e que a precisão da anotação GO pode ser extensa para as anotações da enzima. Além disso, o módulo de mapa KEGG permite a exibição de funções enzimáticas no contexto das vias metabólicas das quais participam.

As sequências de interesse foram analisadas pelo código enzimático e KEGG > *Load Pathway-Maps* do KEGG, online. Dessa forma, foram pesquisados todos os possíveis mapas KEGG contendo os números EC das sequências selecionadas, que foram disponibilizadas na guia *Mapas KEGG*. Os arquivos do KEGG foram exportados em três formatos diferentes usados pelo Blast2GO: *One Sequence per Row*, *Onde Enzyme per Row* e *One Pathway per row*.

4.4. ESTATÍSTICAS

Os resultados de um modo geral foram plotados em gráficos de distribuição de dados obedecendo as seguintes etapas de opções estatísticas da ferramenta Blast2GO: O projeto (*Data distribution* e *Sequence length statistic*), BLAST (*E-value distribution*, *Similarity distribution*, *Species Distribution* e *Top-hit Species Distribution*), InterPro (*Statistics related to the InterProScan results*), Mapeamento (*GO Mapping distribution*, *EC Distribution for sequences*, *EC Distribution for BLAST hits* e *DB-Resources of Mapping*) e Anotação (*Annotation distribution* e *GO Annot. Level Distribution*).

Distribuição de sequência por nível de GO foram gerados em gráficos de pizza para representar o número de sequências para cada termo de GO a um determinado nível. Gráficos de barras também foram gerados para representar os termos GO de acordo com o número de sequências anotadas para o Top 50 dos termos GO.

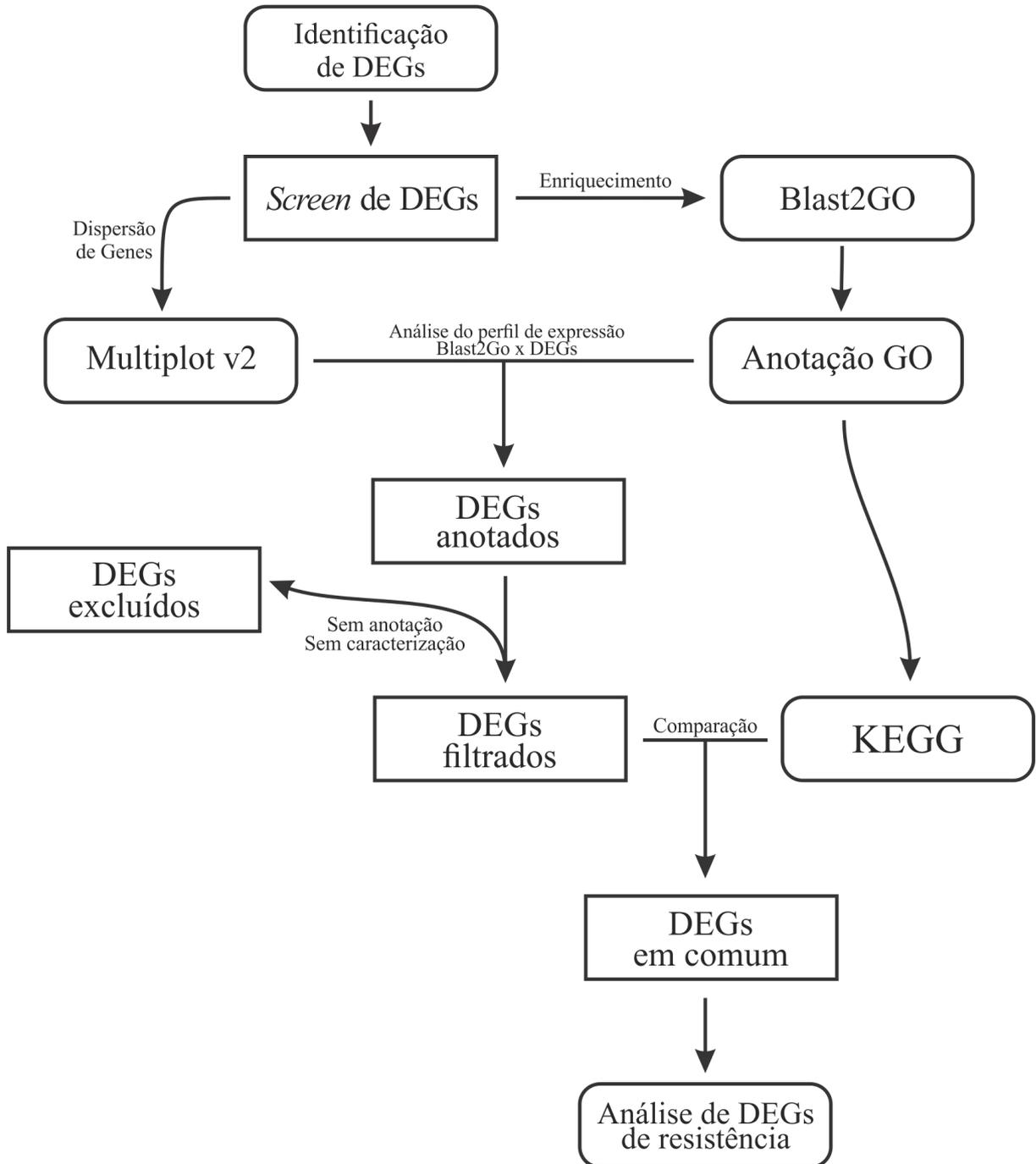
Os parâmetros foram estabelecidos no filtro de *Sequência* da ferramenta em *Make Combined Graph*, conforme o tutorial de análises quantitativas estabelecido (<http://docs.blast2go.com/user-manual/quantitative-analysis/>). De acordo com os resultados gerados foram feitos ajustes sobre o valor total de sequências, sendo estabelecido um valor de 10% do seu número total de sequências para os três processos do GO.

4.5. SELEÇÃO DE DEG RELATADO NA DEFESA/RESISTÊNCIA E ANÁLISE DE VIAS METABÓLICAS

Após a recuperação das vias metabólicas, esses dados foram contrastados com os da anotação funcional feita pelo Blast2GO por meio de fórmulas de formatação condicional, que buscam valores duplicados. Dos genes em comum, foram analisados os principais envolvidos apenas na defesa e resistência da planta, para isso foram filtrados apenas os que apresentaram os termos *disease* (doença), *resistance* (resistência) e *defense* (defesa) nos termos GO de anotação funcional.

A metodologia desse trabalho e toda análise realizada encontra-se sumarizada no fluxograma da figura 4.

Figura 4 – Fluxograma descritivo das seqüências de análises realizadas para esse trabalho. Caixas retangulares com bordas arredondadas representam as análises realizadas e as caixas retangulares com bordas pontiagudas representam os resultados das análises.



Fonte: Autor da pesquisa, 2018.

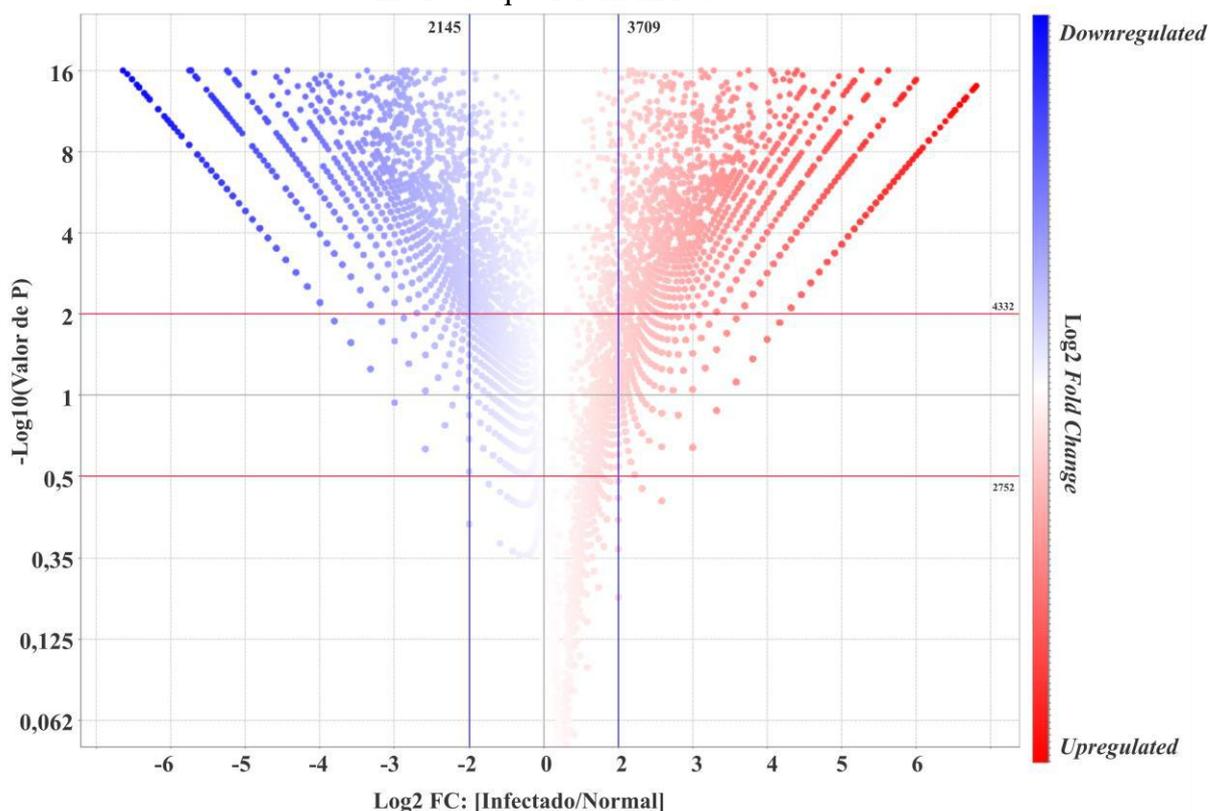
5. RESULTADOS E DISCUSSÃO

5.1. SCREEN E COMPARAÇÃO DE DEGS APÓS A INFECÇÃO

O sequenciamento do transcriptoma da raiz de pimenta-do-reino permitiu a identificação de 16.941 DEGs através do DEGSeq (MOREIRA et al., 2017).

A análise da expressão diferencial usando o $\text{Log}_2 \text{FC} = (\text{Infectada}/\text{Controle})$ permitiu identificar, entre as duas condições, um total de 2.145 DEGs com um perfil *downregulated* (regulados negativamente). Após o filtro com o ponto de corte ($\text{FC} \leq 2$ e valor de $p \leq 0.001$) restaram apenas 2.108 DEGs com esse perfil. Da mesma forma, um total de 3.709 DEGs com um perfil *upregulated* (regulados positivamente) foram identificados e após serem filtrados pelo ponto de corte ($\text{FC} \geq 2$ e valor de $p \leq 0.001$), resultaram em 1.642 DEGs com esse perfil oposto. O gráfico da figura 5 mostra um *screen* desses DEGs gerados pela dispersão dos dados.

Figura 5 – Gráfico de dispersão *Volcano plot* dos genes diferencialmente expressos de pimenta-do-reino. No eixo y está representado o valor de *Fold Change* e no eixo x o valor de p. Pontos azuis são genes regulados negativamente e vermelhos genes regulados positivamente. Pontos presentes no topo do gráfico e distantes do centro representam os genes de interesse que foram filtrados.



Fonte: Adaptado de Multiplot, 2018.

5.2. ANÁLISE DOS PERFIS DE EXPRESSÃO DE DEGS

Após a anotação funcional feita pelo Blast2GO para mais de 16 mil sequências dos genes da pimenta-do-reino, apenas 7.835 DEGs tiveram uma anotação funcional bem-sucedida. Em seguida, essas sequências anotadas pelo Blast2GO foram avaliadas pelo perfil de expressão diferencial conforme Wang et al. (2010), onde foram identificadas 3.732 DEGs enriquecidos com a ferramenta Blast2GO. Outras sequências que não exibiram faixa alguma de anotação coberta durante o enriquecimento totalizam cerca de 4.103 sequências.

De fato, as 3.732 sequências enriquecidas com o Blast2GO corresponderam aqueles DEGs que exibiram distintos perfis de expressão diferencial: regulados negativamente e regulados positivamente observados no gráfico de dispersão (Figura 5). Os resultados das análises de anotação e perfis de expressão realizadas estão mostrados na tabela 2.

Tabela 2 – Sumário das análises realizadas dos DEGs.

Descrição	Quantitativo de Análise (n)
DEGs identificados com perfil de expressão	3.750
DEGs regulados negativamente	2.108
DEGs regulados positivamente	1.642
Anotação Funcional (Blast2GO)	7.835
Blast2GO vs. DEGs (WANG et al., 2010)	3.732
DEGs sem anotação	4.103
GO (DEGs regulados negativamente)	819
Processo Biológico (PB)	212
Função Molecular (FM)	413
Componente Celular (CC)	194
GO (DEGs regulados positivamente)	570
Processo Biológico (PB)	164
Função Molecular (FM)	286
Componente Celular (CC)	120
Vias identificadas pelo KEGG	94
Sequências identificadas pelo KEGG	550
Enzimas identificadas pelo KEGG	123

Fonte: Autor da Pesquisa, 2018.

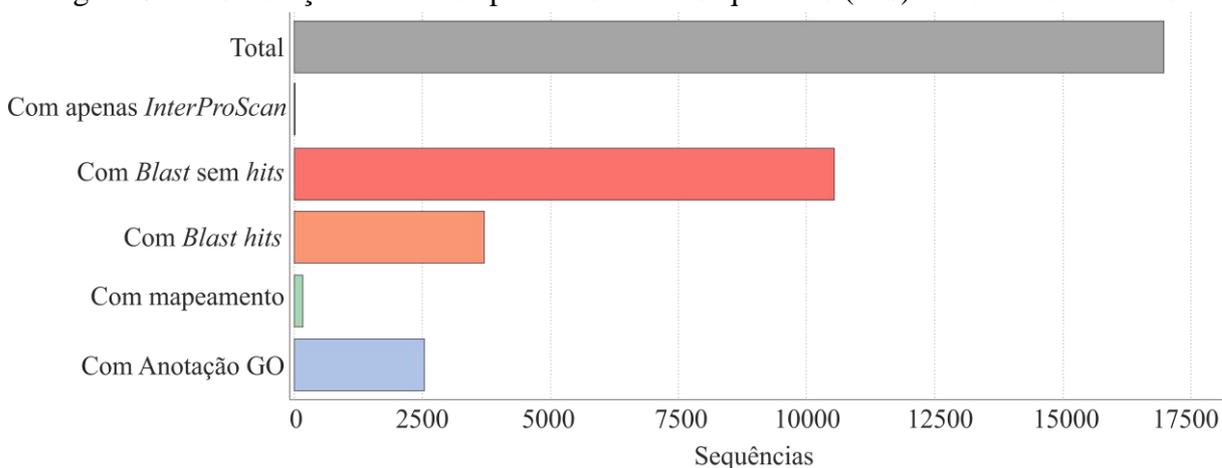
A maioria dos genes identificados em todas as análises estão regulados negativamente, sugerindo que grande parte dos transcritos não estejam sendo produzidos de forma adequada para desencadear uma resposta eficaz na planta. Segundo Moreira et al. (2017), após o sequenciamento do transcriptoma da raiz infectada da pimenta-do-reino, os genes relacionados a resistência da planta encontraram-se hipoexpressos, ou seja, regulados negativamente, o que corrobora com a alta taxa de genes com perfil de expressão negativo nessa nova análise.

5.3. ANÁLISE QUANTITATIVA DA DISTRIBUIÇÃO DOS DADOS DE ANOTAÇÃO COM O BLAST2GO

A primeira parte da análise quantitativa da distribuição dos dados de anotação obtidos com a ferramenta Blast2GO é demonstrada no gráfico de distribuição da figura 6.

Os resultados mostram cerca de 3.740 sequências com *hits* combinados significativamente com sequências do banco de dados *nr* após a etapa do BLAST. Além disso, 2.535 sequências foram anotadas com GO, 61 mapeadas e mais de 10.605 não obtiveram semelhantes após a etapa do BLAST (Figura 6).

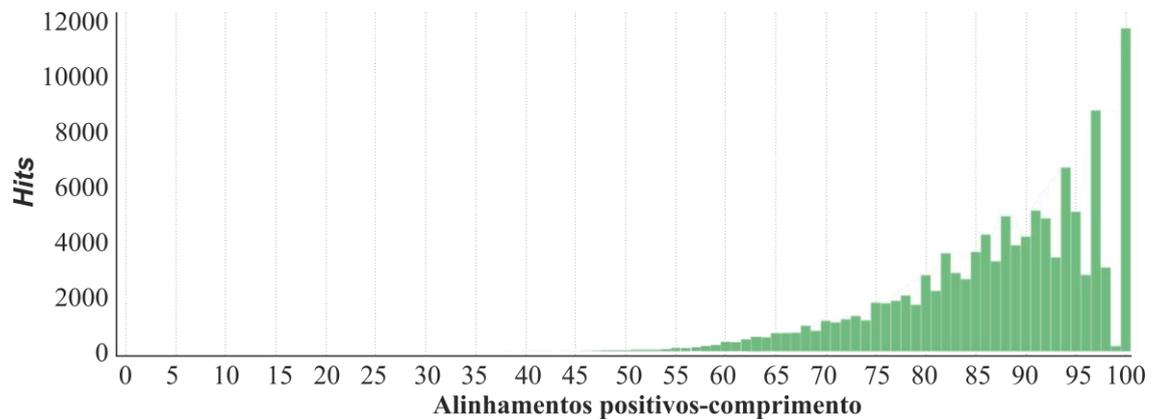
Figura 6 – Distribuição de dados após consulta de sequências (*hits*) em bancos de dados.



Fonte: Adaptado de Blast2GO, 2018.

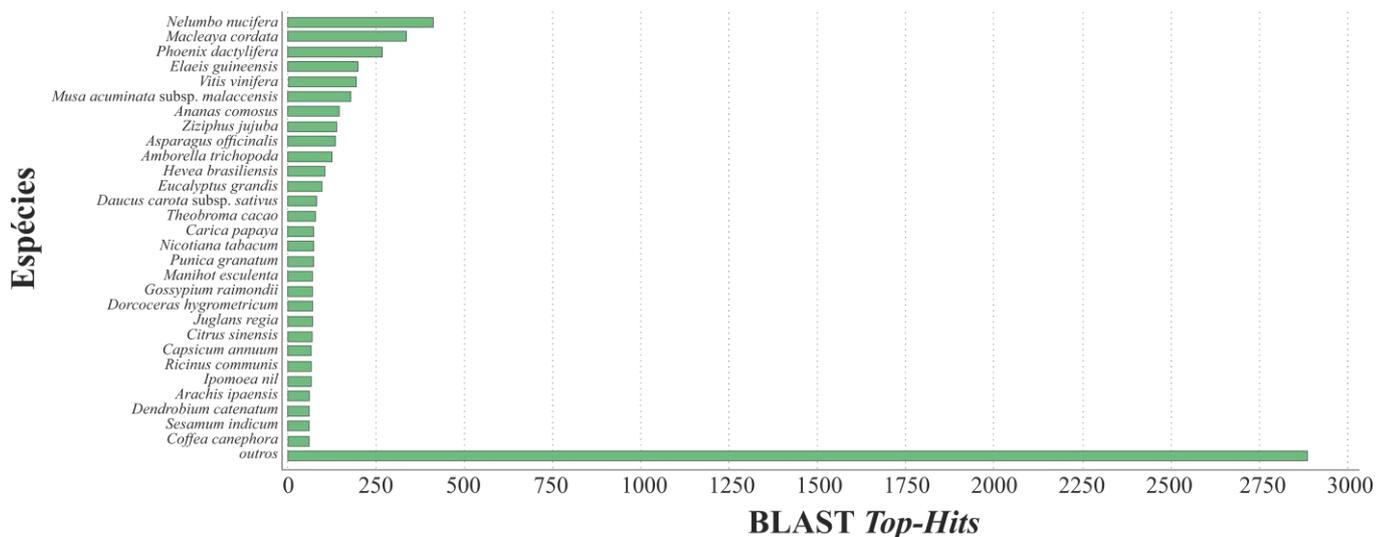
O resultado do BlastX contra o banco de dados *nr-viridiplantae* mostrou que a distribuição de similaridade das sequências apresenta em sua maioria 100 pares de bases (pb) similares a sequências do banco de dados, como mostra a figura 7.

Figura 7 – Distribuição de similaridade das sequências contra o banco de dados.



A distribuição de espécies entre as sequências de maior similaridade com o transcriptoma da pimenta-do-reino mostrou que *Nelumbo nucifera*, foi a espécie que apresentou maior número de sequências similares, seguida por *Macleaya cordata* e *Phoenix dactylifera*, com mais de 250 hits contra a referência, conforme mostra o gráfico da figura 8.

Figura 8 – Distribuição da abundância de *top-hits* por espécie na análise de BlastX para o transcriptoma da pimenta-do-reino.

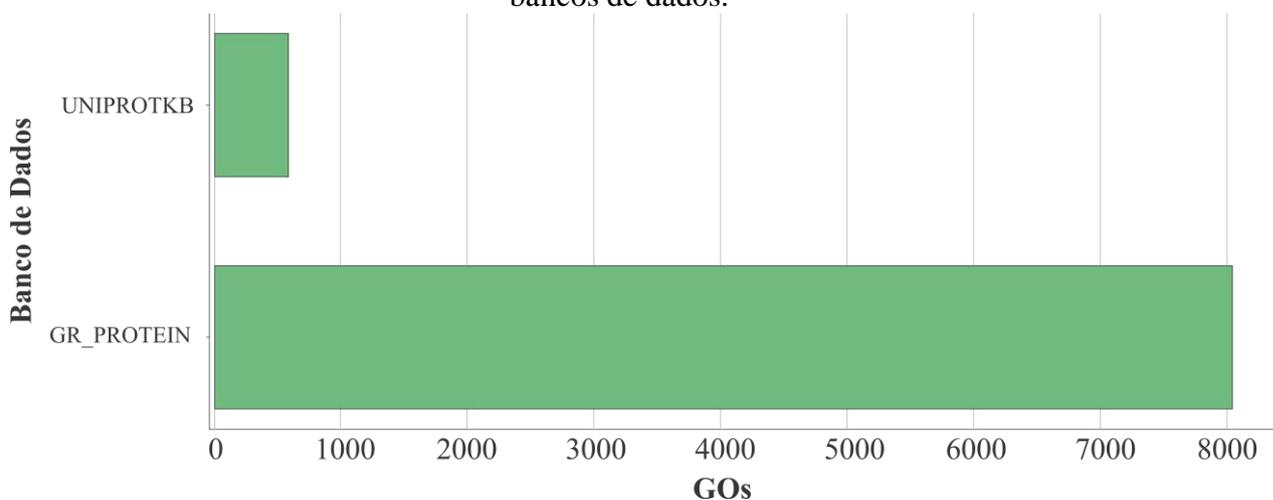


A pimenta-do-reino e a *N. nucifera* (Flor de Lótus) são mais próximas filogeneticamente que as outras duas espécies com mais de 250 sequências compatíveis (RAVEN; EICHHORN; EVERT, 2014). Ambas pertencem a classe Magnoliopsida, mais conhecidas como

dicotiledôneas, plantas com flor, cujo embrião (semente) contém dois ou mais cotilédones. Enquanto que a *M. cordata* e a *P. dactylifera* são pertencentes a classe Liliopsida (monocotiledôneas) (RAVEN; EICHHORN; EVERT, 2014).

O mapeamento de ontologia genética utilizado para recuperar os termos GO associados aos acertos obtidos pela pesquisa do BLAST foi usado nesse estudo. O resultado é exibido no gráfico da figura 9, que mostra uma distribuição do número de anotações (termos-GO) recuperadas dos diferentes bancos de dados de origem, que podem ser, por exemplo, o UniProt (*Universal Protein*), o PDB (*Protein Data Bank*), o TAIR (*The Arabidopsis Information Resource*), etc. Como resultados dessa análise, a maioria das anotações realizadas são derivadas do banco de dados GR_PROTEIN, seguido do UniProtKB, totalizando 8.975 anotações para os 16.941 genes de pimenta-do-reino (Figura 9).

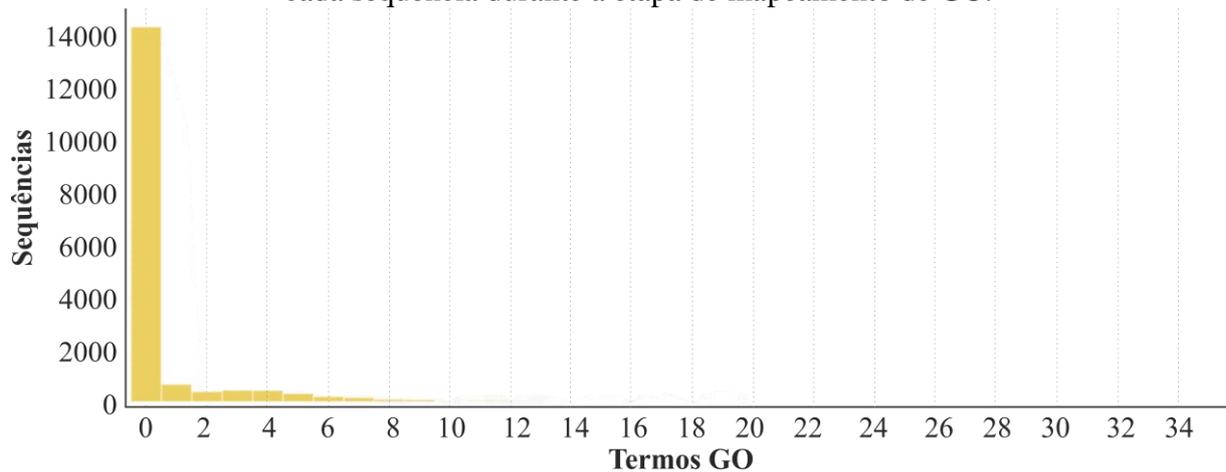
Figura 9 – Distribuição do número de anotações (termos-GO) recuperadas dos diferentes bancos de dados.



Fonte: Adaptado de Blast2GO, 2018.

Como uma visão geral da extensão e intensidade da anotação GO, o gráfico de distribuição de anotações (Figura 10) mostra mais de 14 mil sequências sem atribuição de termos GO, condizendo com o gráfico de distribuição de dados realizado no enriquecimento (Figura 6).

Figura 10 – Distribuição da quantidade de termos candidatos à ontologia de gene atribuídos a cada sequência durante a etapa de mapeamento de GO.



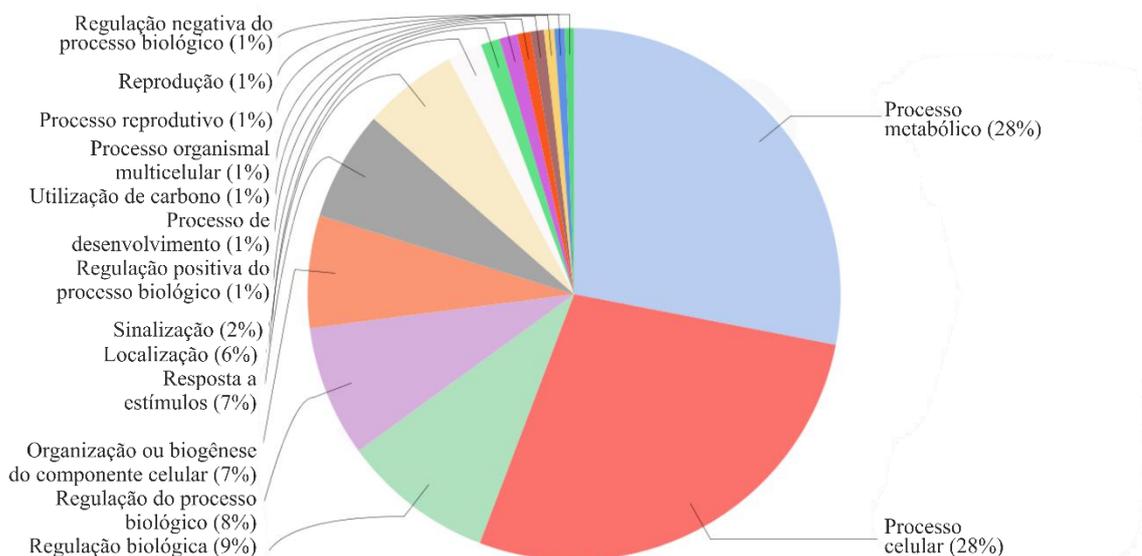
Fonte: Adaptado de Blast2GO, 2018.

5.4. ONTOLOGIA GENÉTICA DOS DEGS

Os resultados da anotação funcional mostraram que a maioria das sequências tiveram anotações para Processo Biológico (39,49%), seguido de Componente Celular (36,98%) e Função Molecular (23,53%).

Na categoria Processos Biológico, os transcritos foram classificados em sua grande maioria em processos metabólicos (28%), seguido por processos celulares (28%) e regulação biológica (9%) (Figura 11).

Figura 11 – Classificação dos transcritos no GO quanto a Processos Biológico em pimenta-do-reino.

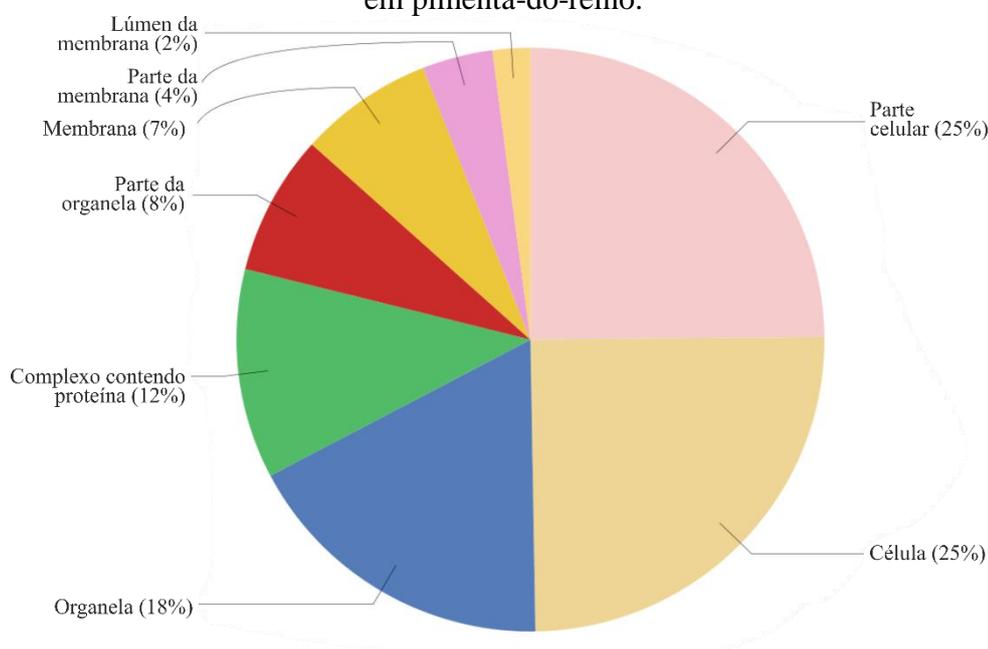


Fonte: Adaptado de Blast2GO, 2018.

As categorias observadas são, em sua maioria, compatíveis com resultados de anotações funcionais de outras espécies, com exceção da subcategoria regulação biológica. Isso ocorre, pois, os processos metabólicos são um conjunto amplo de procedimentos responsáveis pela síntese e degradação de moléculas das células e as vias metabólicas dessa síntese ou degradação ocorrem constantemente para suprir o funcionamento da célula (SMITH, 2000). Os processos celulares são abundantes uma vez que fazem referência as atividades celulares de crescimento, manutenção e comunicação entre células (JUNG et al., 2014). Ambos são processos importantes para o funcionamento e manutenção dos organismos vivos.

Na categoria Componente Celular a grande maioria das sequências estão relacionadas à célula (25%) e parte celular (25%), seguido por organelas (18%) (Figura 12).

Figura 12 – Classificação dos transcritos no GO quanto a Componentes Celulares em pimenta-do-reino.

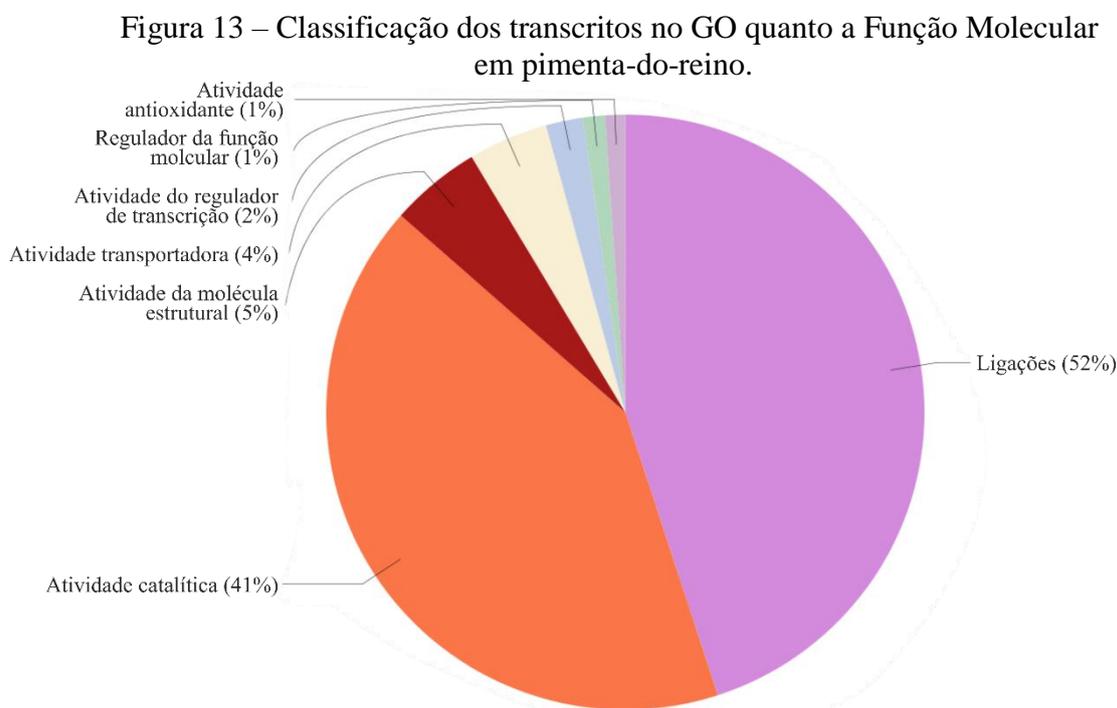


Fonte: Adaptado de Blast2GO, 2018.

As sequências ligadas a célula ou parte celular são correspondentes a genes que exercem sua função no citoplasma e não estão agregados na membrana plasmática ou em organelas (CARBON et al., 2009). É esperado que 50% dos genes atuem no citoplasma, pois é nesse local da célula onde ocorrem a grande maioria dos processos metabólicos. As sequências correspondentes a organelas estão associadas aos elementos constitutivos da célula ou aos que atuam no núcleo, mitocôndria, plastídio, vacúolos, vesículas, ribossomos e citoesqueleto. Essas sequências estão entre as mais abundantes pois é nas organelas que ocorrem processos metabólicos fundamentais para o funcionamento da célula como a respiração, fotossíntese,

estruturaco celular e sntese de biomolculas, alm de comporem aproximadamente 40% do volume celular (NELSON; COX, 2014).

Por fim, na categoria Funo Molecular, mais da metade das sequncias esto relacionadas a ligaoes (52%), seguido por sequncias ligadas a atividade cataltica (41%) (Figura 13).



Fonte: Adaptado de Blast2GO, 2018.

O fato das sequncias estarem amplamente relacionadas a ligaoes  referente a responsabilidade dessas sequncias em formar ligaoes ocorrentes entre molculas como protena-protena, protena-DNA, cofator-enzima, entre outras, sendo essas ligaoes fundamentais para o controle e funcionamento das molculas envolvidas no metabolismo celular. Sequncias relacionadas  atividade cataltica esto em abundncia, pois grande parte das reaoes qumicas ocorre com a presena de enzimas (catalisadores), molculas capazes de aumentar a velocidade das reaoes qumicas sem serem consumidas no processo (CARBON et al., 2009; NELSON; COX, 2014). Alm desses motivos para a apario em grandes quantidades desses termos GO na anotao da pimenta-do-reino, plantas que esto sendo afetadas por um patgeno acabam induzindo uma extensa reprogramao em seu metabolismo, envolvendo mudanas na atividade gnica, ou seja, a sntese e degradao de molculas que vo auxiliar no retardamento do patgeno esto sendo realizadas em grandes quantidades diante da infeco, envolvendo papis dos genes responsveis pelas ligaoes de molculas de defesa e pela

atividade catalítica para aceleração da síntese de moléculas (WALTERS; NEWTON; LYON, 2007; STANGARLIN et al., 2011).

5.5. RECUPERAÇÃO DAS VIAS METABÓLICAS

Usando o KEGG, as vias enzimáticas foram mapeadas via Blast2GO-KEGG. A partir disso, foram identificadas 94 vias metabólicas que possuem 550 sequências do transcriptoma da pimenta-do-reino, sendo dessas, 123 sequências enzimáticas (Tabela 3).

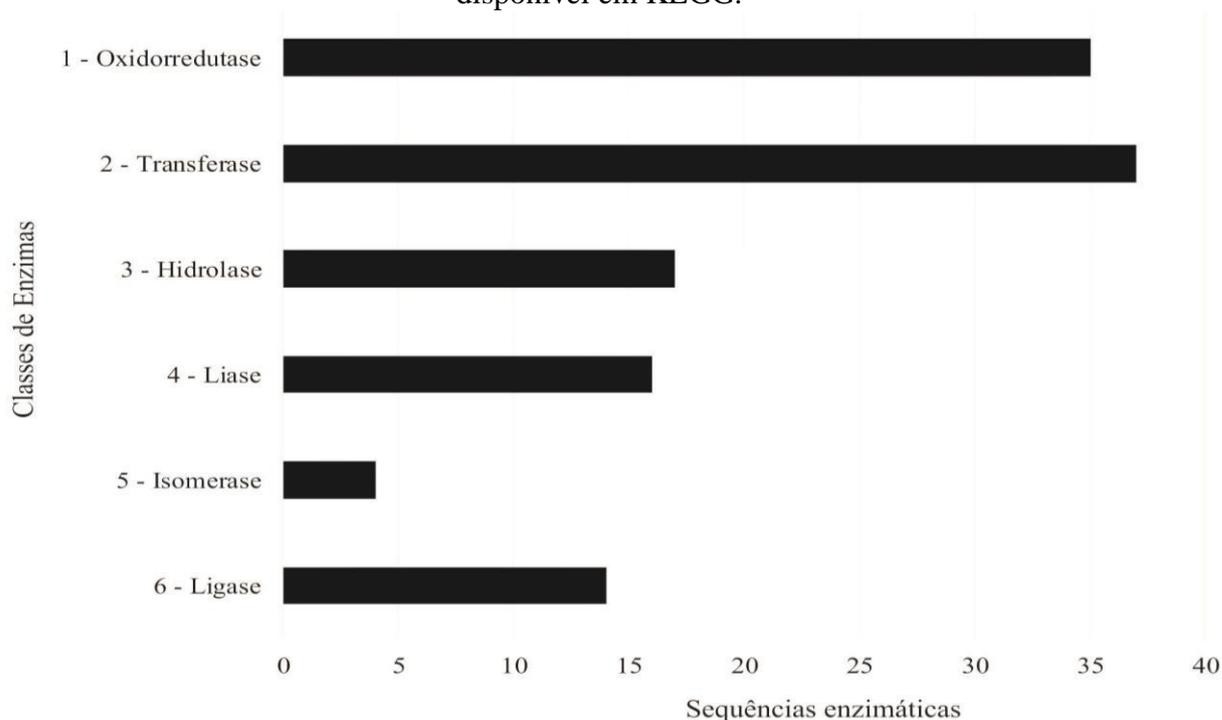
Tabela 3 – Resultado do mapeamento das vias enzimáticas do transcriptoma da pimenta-do-reino, através do Blast2GO-KEGG.

Mapeamento	Quantidade (n)
Vias metabólicas identificadas	94
Sequências do transcriptoma identificadas nas vias metabólicas	550
Sequências enzimáticas identificadas no transcriptoma da pimenta-do-reino envolvidas nas vias metabólicas	123

Fonte: Autor da Pesquisa, 2018.

A classe mais abundante das enzimas identificadas foi a transferase, com um total de 37 sequências enzimáticas, seguidas pelas classes Oxidorredutase e Hidrolase, com 34 e 17 sequências enzimáticas, respectivamente (Figura 14).

Figura 14 – Distribuição da abundância de classes enzimáticas de acordo com banco de dados disponível em KEGG.



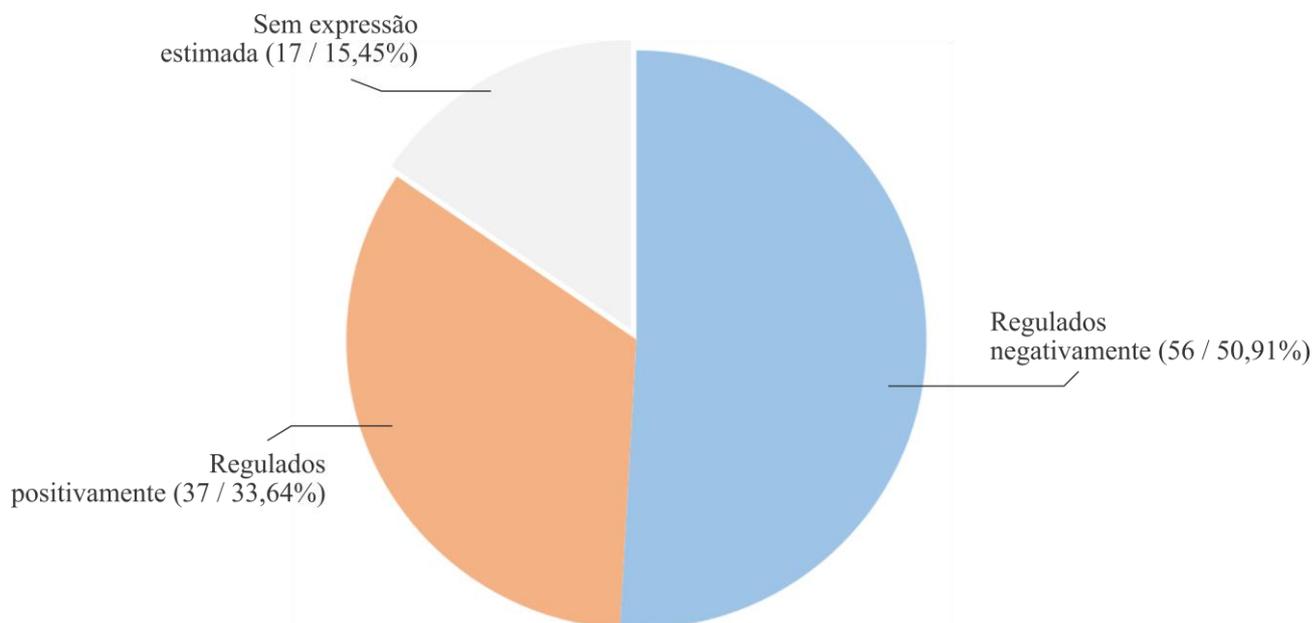
Fonte: Autor da Pesquisa, 2018.

Das 550 sequências do transcriptoma identificadas nas vias metabólicas, 110 estão envolvidas em vias de biossíntese de antibióticos da pimenta-do-reino. Destas, 37 são sequências enzimáticas. De modo geral, as plantas produzem antibióticos como parte de seu programa normal de crescimento e desenvolvimento em resposta a estresse ou patógenos. Entretanto, em resposta ao ataque de patógenos que passam por esses antibióticos, a planta induz compostos antimicrobianos conhecidos como fitoalexinas (MATIELLO et al., 2016).

Segundo Cavalcanti et al. (2005), as fitoalexinas agem em fungos na granulação citoplasmática, desorganização de conteúdos celulares, rompimento da membrana plasmática e inibição de enzimas fúngicas, provocando a redução ou anulação do crescimento micelial.

Logo, o fato da maioria das enzimas identificadas pelo KEGG estarem relacionadas com a biossíntese de antibióticos naturais ou induzidos, pode indicar que o transcriptoma da pimenta-do-reino contém uma grande quantidade de informações de defesa sendo geradas em resposta a infecção do *F. solani* f. sp. *piperis*. Apesar disso, das 110 sequências observadas, 56 (50,90%) estão reguladas negativamente após infecção do fungo (Figura 15) dando um indicativo da possível susceptibilidade da planta.

Figura 15 – Nível de expressão das sequências envolvidas em vias de biossíntese de antibióticos da pimenta-do-reino.



Fonte: Autor da Pesquisa, 2018.

Doze enzimas identificadas estão envolvidas no metabolismo do amido e da sacarose. A sacarose, como produto final da fotossíntese, desempenha papel fundamental no crescimento de plantas e funciona como regulador direto ou indireto da expressão gênica, sendo de grande importância na sobrevivência da planta (WINTER; HUBER, 2000).

Na análise foram identificadas 34 sequências do transcriptoma ligadas a via do metabolismo da glutathione (35,35% reguladas negativamente), sendo 7 delas sequências enzimáticas. A glutathione é um dos mais importantes antioxidantes dos vegetais. Em plantas, esse antioxidante é fundamental no controle do estresse biótico e abiótico, sendo necessária para uma defesa eficiente contra patógenos (SANTOS et al., 2011; MAURIZI; POINSSOT, 2012).

Estudos com mutantes de *Arabidopsis thaliana* deficientes em glutathione apresentaram alta suscetibilidade a uma ampla gama de invasores patogênicos, revelando o papel desse composto na sinalização de plantas sob estresse biótico (PARISY et al., 2007; MAURIZI; POINSSOT, 2012).

Além desses resultados, vias comuns para o desenvolvimento de plantas foram sinalizadas, tais como da glicólise, gliconeogênese, metabolismo de aminoácidos fundamentais como a cisteína e a metionina, metabolismo da purina, entre outras. Algumas vias metabólicas obtidas com o Blast2GO-KEGG estão sumarizadas na Tabela 4.

Tabela 4 – Sumário de algumas vias metabólicas identificadas no transcriptoma de pimenta-do-reino, através do Blast2GO-KEGG.

Via metabólica	Identificação da via	Enzimas identificadas envolvidas (n)
Biossíntese de Antibióticos	map01130	37
Metabolismo de amido e sacarose	map00500	12
Glicólise/Gliconeogênese	map00010	12
Metabolismo de cisteína e metionina	map00270	10
Biossíntese de aminoacil-tRNA	map00970	10
Metabolismo de purina	map00230	8
Metabolismo de piruvato	map00620	7
Metabolismo de glutatona	map00480	7
Fosforilação oxidativa	map00190	6
Metabolismo de alanina, aspartato e glutamato	map00250	6

Fonte: Adaptado de Blast2GO-KEGG, 2018.

Para uma análise mais profunda dessas vias é fundamental verificar se essas sequências enzimáticas e de genes da pimenta-do-reino estão sendo regulados adequadamente ou estão desativados frente à infecção do fungo. Por esse motivo, esses dados gerados pelo KEGG foram contrastados com o resultado da anotação funcional dos DEGs regulados negativamente e positivamente pelo Blast2GO.

5.6. SELEÇÃO DE DEGS RELATADOS NA DEFESA/RESISTÊNCIA E ANÁLISE DE VIAS METABÓLICAS

Como resultado da busca por valores duplicados dos dados da anotação e do KEGG, foram identificados 133 genes em comum nas duas análises, sendo 78 regulados negativamente e 55 regulados positivamente (Tabela 5).

Tabela 5 – Resultado dos DEGs downregulated e upregulated identificados pelo KEGG.

Descrição	Quantitativo de Análise (n)
KEGG <i>Enzyme Pathways</i> DEGs (<i>Downregulated</i>)	78
KEGG <i>Enzyme Pathways</i> DEGs (<i>Upregulated</i>)	55
Total	133

Fonte: Autor da Pesquisa, 2018.

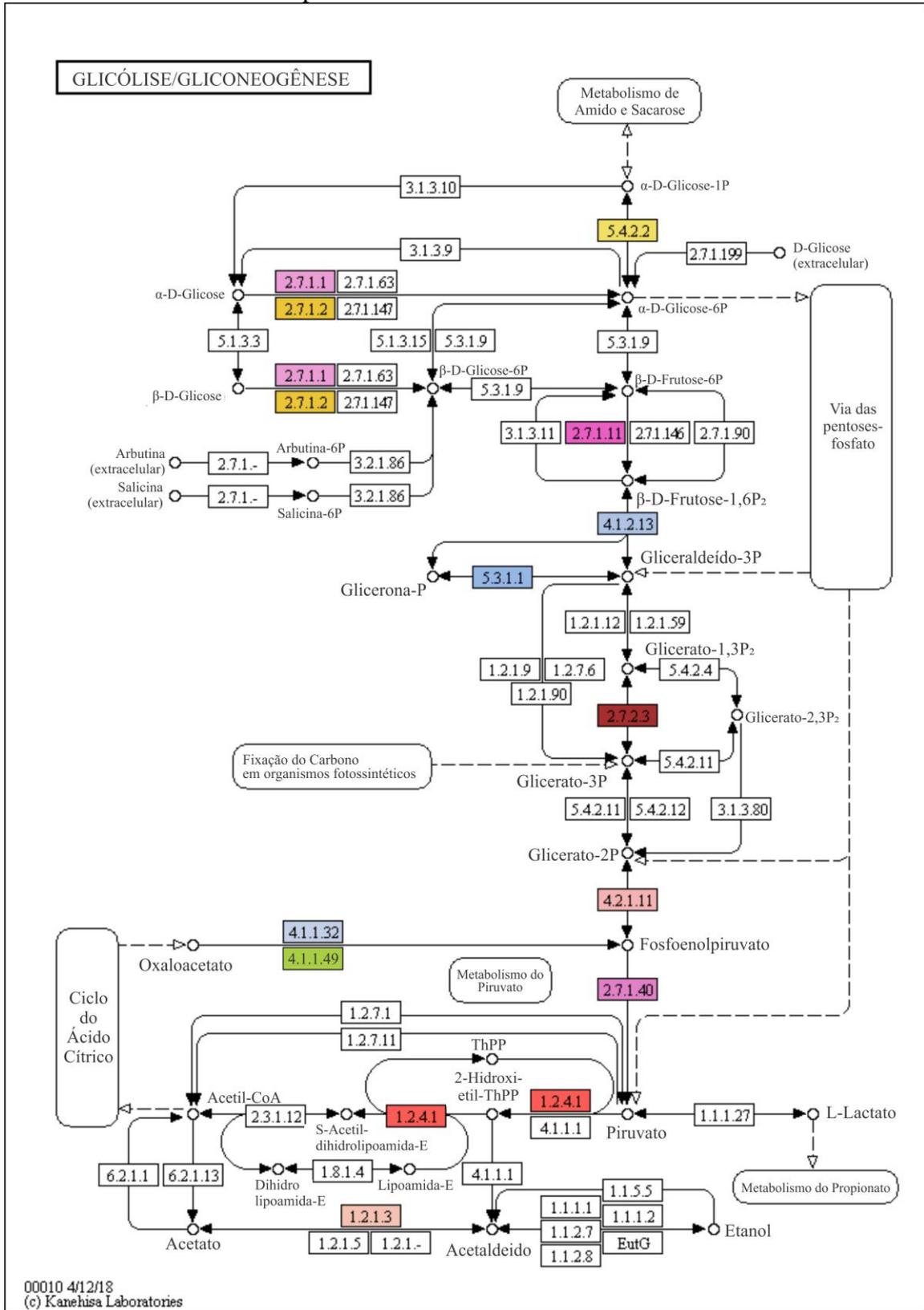
Dos DEGs identificados, selecionamos um gene denominado *Phosphoenolpyruvate* (*PCK*), regulado negativamente em pimenta-do-reino, para uma análise mais aprofundada das vias metabólicas relacionadas e sua anotação funcional.

5.5.1. Fosfoenolpiruvato (*PCK*)

A sequência denominada Fosfoenolpiruvato (*PCK*) representa o gene que codifica a enzima fosfoenolpiruvato carboxiquinase (PEPCK) (IZUI et al., 2004). Esse gene apresentou 11 termos GO distribuídos em Função Molecular, Componente Celular e Processo Biológico no Blast2GO. Um termo GO atribuído a esse gene e que parece ser de extrema relevância, foi o termo filho *Resposta de Defesa ao Fungo, Interação Incompatível* (GO: 0009817) da categoria Processo Biológico, que, segundo Carbon et al. (2009), significa a resposta da planta a um fungo no sentido de impedir a propagação ou ocorrência de uma doença. Isso implica na hipótese de que, o fato do gene *PCK* está sendo regulado negativamente, faz com que a pimenta-do-reino não desenvolva respostas necessárias para impedir que a fusariose se propague em seu organismo.

A enzima PEPCK (EC 4.1.1.49), codificada pelo gene *PCK*, está envolvida diretamente na via metabólica da gliconeogênese (Figura 16), na produção de fosfoenolpiruvato a partir do oxaloacetato do ciclo do ácido cítrico (ciclo de Krebs). A via metabólica da gliconeogênese é responsável pela produção de glicose a partir de compostos moleculares presente em plantas, animais, fungos e outros microrganismos, sendo que nas plantas, a glicose é fundamental para o seu desenvolvimento e manutenção, pois fornece energia para a fotossíntese (NELSON; COX, 2014).

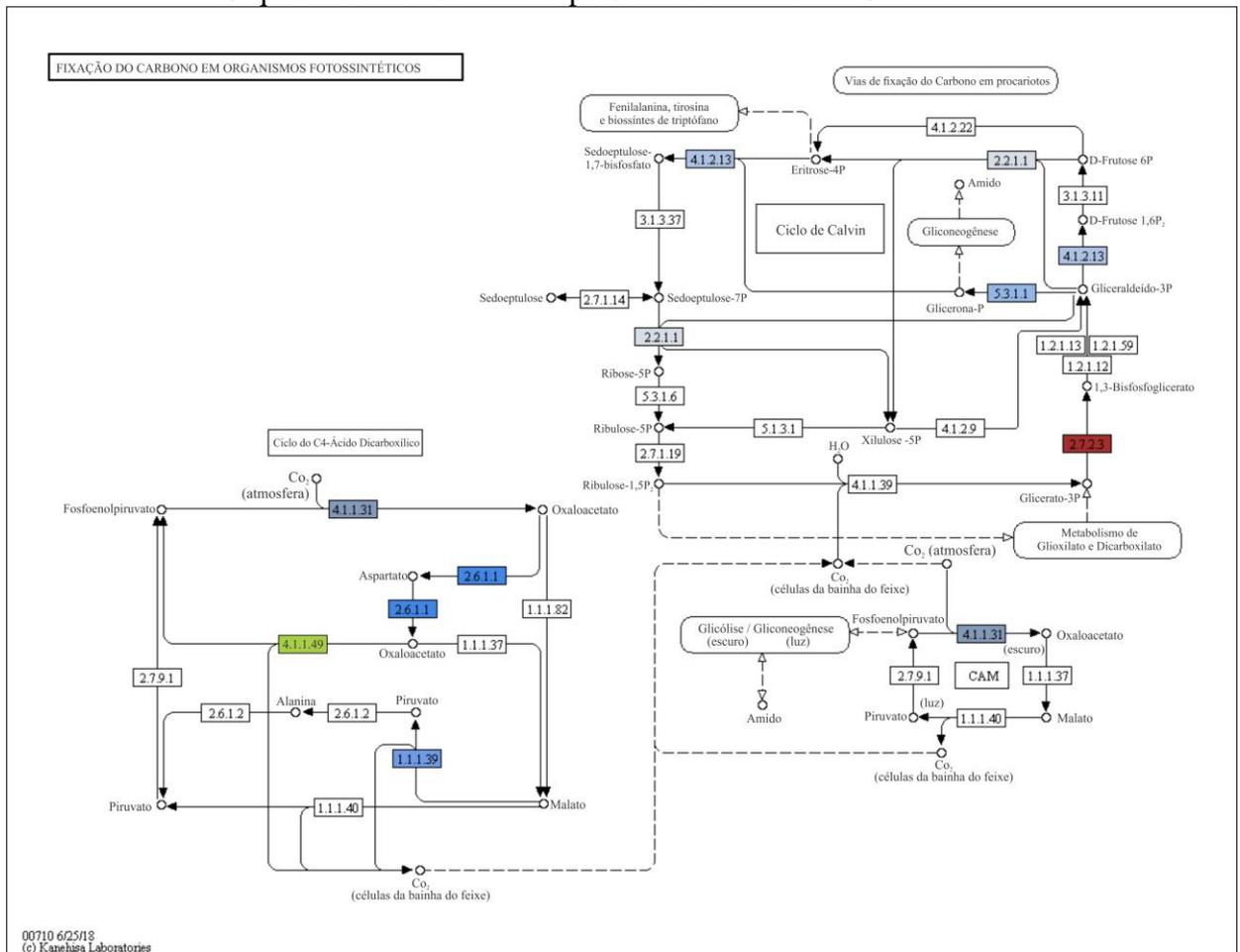
Figura 16 – Via metabólica da gliconeogênese em pimenta-do-reino obtida no Blast2GO. Retângulos coloridos representam a diversidade enzimática obtida a partir dos transcritos da pimenta-do-reino, cada cor representa uma enzima diferente. O quadrado na cor verde representa a enzima PEPCK na via.



Fonte: Adaptado de Blast2GO-KEGG, 2018.

Outra via metabólica regulada pela PEPCK é a de fixação do carbono em organismos fotossintéticos (Figura 17). Essa via também atua na fotossíntese, sendo fundamental para sua realização. Esse processo ocorre por meio do ciclo de Calvin, conhecido como “fase escura” da fotossíntese, e realiza a coleta de moléculas de CO₂ simples para formar moléculas maiores e mais complexas como aminoácidos, ácidos graxos e carboidratos que serão utilizados pelas plantas (GUERRA, 2010).

Figura 17 – Via metabólica da fixação do carbono em organismos fotossintéticos de pimenta-do-reino obtida no Blast2GO. Retângulos coloridos representam a diversidade enzimática obtida a partir dos transcritos da pimenta-do-reino, cada cor representa uma enzima diferente. O quadrado na cor verde representa a enzima PEPCK na via.



Fonte: Adaptado de Blast2GO-KEGG, 2018.

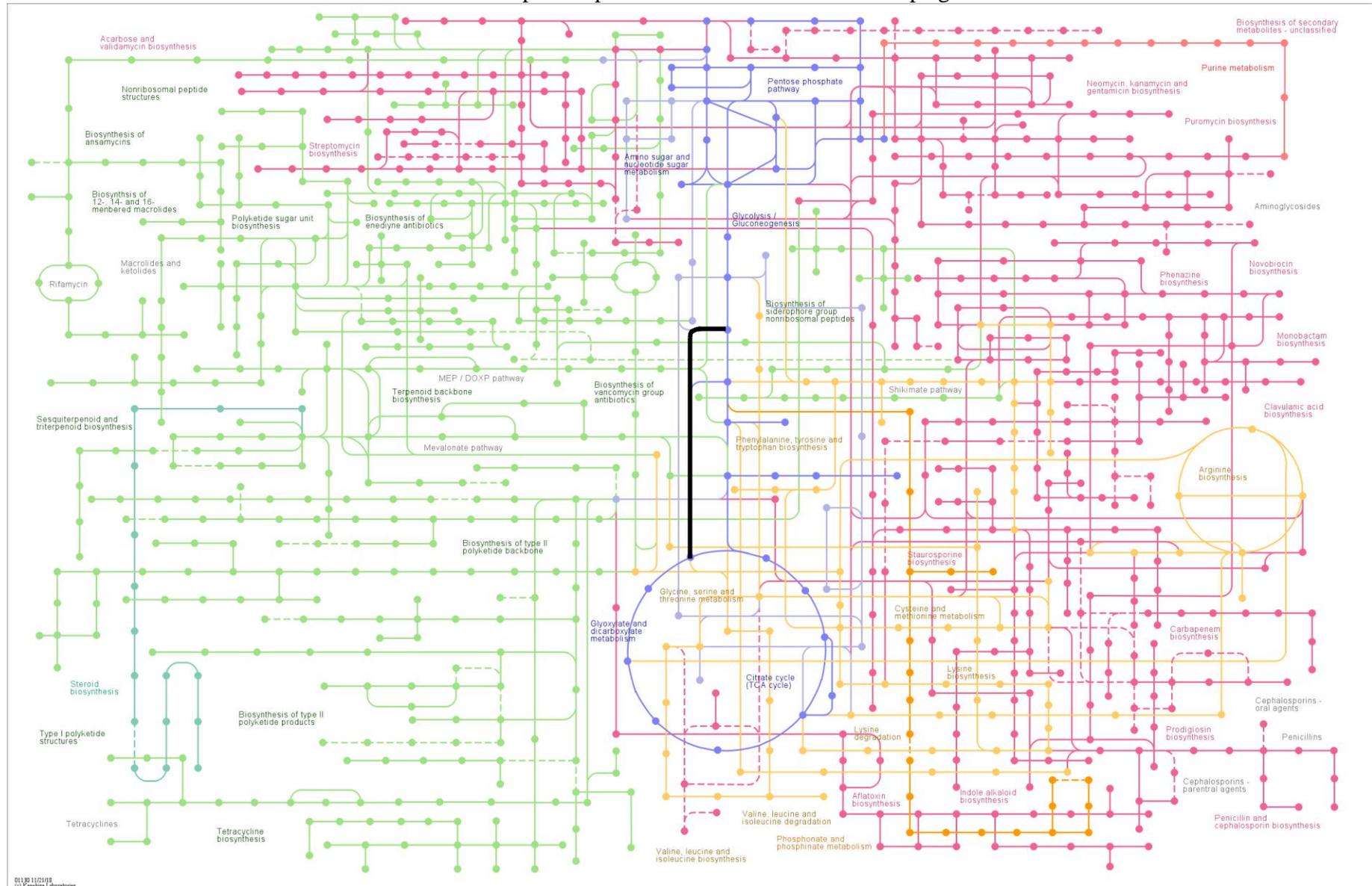
Outras duas vias com envolvimento da enzima PEPCK que estão relacionadas com a fotossíntese da planta, são a via do ciclo de Krebs e do metabolismo do piruvato. Ambas as vias possuem importância significativa no crescimento da planta, pois acabam sendo bases energéticas para o organismo, formando moléculas biológicas fundamentais para a sobrevivência do vegetal (TAIZ, 2017).

Muito se sabe sobre a função da PEPCK em vias metabólicas relacionadas a fotossíntese, como a gliconeogênese, porém não está muito claro a função deste na imunidade da planta em mecanismos de defesa a infecções. Estudos realizados com uma espécie de pimenta (*Capsicum annuum*) infectados pela bactéria *Xanthomonas campestris* pv. *vesicatoria* resultaram na verificação de baixa atividade de PEPCK em folhas saudáveis e aumento drástico em folhas infectadas pela bactéria. Ou seja, em termos gerais, a PEPCK contribui positivamente para a imunidade inata de plantas contra agentes patogênicos (CHOI; KIM; HWANG, 2015).

Outro estudo em relação a resposta a patógenos de plantas inclui aumento na atividade enzimática da PEPCK nas folhas de repolho chinês infectado pelo vírus do mosaico amarelo. Esse resultado se deu devido as mudanças na fixação de CO₂ sofridas pela planta após infecção (TAKACS et al., 2014).

As anotações do KEGG mostraram ainda a relação da enzima PEPCK na via metabólica de biossíntese de antibióticos. Essa via, já discutida nesse trabalho, é de grande importância para a defesa e resistência da planta e encontra-se envolvida com a enzima PEPCK (Figura 18).

Figura 18 – Mapa global da via metabólica da biossíntese de antibióticos obtida no Blast2GO. Linhas representam enzimas e reações e círculos representam metabólitos. Cores representam categorias funcionais do KEGG, como processos celulares em verde. A linha preta representa a enzima PEPCK no mapa global.



Fonte: Adaptado de Blast2GO-KEGG, 2018

6. CONSIDERAÇÕES FINAIS

O uso da ferramenta Blast2GO permitiu identificar transcritos de interesse biotecnológico em pimenta-do-reino. A distribuição das anotações nas diferentes categorias funcionais mostrou grande diversidade de transcritos anotados, evidenciando a alta representatividade do transcriptoma estudado. Após o refinamento dos dados, foi possível identificar 1.389 DEGs com anotação funcional e expressão estimada, sendo desses, 819 regulados negativamente.

Na era da transcriptômica, onde começamos a tentar compreender melhor a complexidade dos sistemas, torna-se necessária abordagens em larga escala como a realizada, visando elucidar os processos celulares que o organismo desenvolve, principalmente em relação a uma infecção por patógeno.

Da análise funcional detalhada foi possível observar um gene responsável pela defesa da planta contra a fungos, o *PCK*, regulado negativamente. A caracterização desse e de outros 133 DEGs anotados pelo GO e pelo KEGG, geram respostas valiosas para o entendimento do processo de defesa da pimenta-do-reino à fusariose e aprimorar futuros programas de melhoramento genético vegetal.

REFERÊNCIAS

- ANDO, A. et al. Obtenção de Mutantes Resistentes à Fusariose (*Fusarium solani* f.sp. *piperis*) em Pimenta-do-reino (*Piper nigrum* L.) Através de Irradiação Gama. Belém: **Embrapa Amazônia Oriental**, p. 440, 1996.
- ASHBURNER, et al. Gene ontology: tool for the unification of Biology. **Nat Genet**, n. 2, p. 25-9, 2000.
- BENCHIMOL, R. L. et al. By-products of *Piper aduncum* in the control of fusariosis in black pepper plant. **Brazilian Journal of Agricultural Sciences**, v. 12, n. 3, p. 303–308, 2017.
- BOLGER, M. et al. From plant genomes to phenotypes. **Journal of Biotechnology**, v. 261, n. Fevereiro, p. 46–52, 2017.
- CAN, T. Introduction to Bioinformatics. In: YOUSEF, M.; ALLMER, J. miRNomics: MicroRNA Biology and Computational Analysis. Totowa: **Humana Press**, p. 51–71, 2014.
- CARBON, S. AmiGO: online access to ontology and annotation data. **Bioinformatics Applications Note**, v. 25, n. 2, p. 288–289. 2009.
- CARDOSO-SILVA, C. B. et al. *De novo* assembly and transcriptome analysis of contrasting sugarcane varieties. San Francisco: **PloSONE**, v. 9, n. 2, 2014.
- CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. Santa Maria: **Cienc. Rural**, v. 40, n. 3, p. 735-744, 2010.
- CAVALCANTI, L. S. et al. Aspectos bioquímicos e moleculares da resistência induzida. Piracicaba: **FEALQ**, v.81, p.124, 2005.
- CHOI, D. S.; KIM, N. H.; HWANG, B. K. The pepper phosphoenolpyruvate carboxykinase CaPEPCK1 is involved in plant immunity against bacterial and oomycete pathogens. **Plant Mol Biol.**, n. 89, p. 99-111, 2015.
- CHU, E. et al. Avaliação da inoculação de fungos micorrízicos arbusculares sobre a incidência da fusariose da pimenta-do-reino. **Fitopatologia Brasileira**, v. 22, n. 2, p. 205–208, 1997.
- CONESA, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, p. 3674–3676, 2005.
- CONESA, A.; GOTZ S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. **Int J Plant Genomics**, 2008.
- CORRÊA, M. O admirável Projeto Genoma Humano. Physis: **Revista de Saúde Coletiva**, v. 12, n. 2, p. 277–299, 2002.
- D'HONT, A. et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. **Nature**, v. 488, n. 7410, p. 213–217, 2012.

FILGUEIRAS, G. C.; HOMMA, A. K. O.; SANTOS, M. A. S. Conjuntura Do Mercado Da Pimenta-do-reino. **Embrapa**, p. 1–23, 2007.

GANESH, D. et al. Monitoring of the early molecular resistance responses of coffee (*Coffea arabica* L.) to the rust fungus (*Hemileia vastatrix*) using real-time quantitative RT-PCR. **Plant Sci.**, p. 1045–1051, 2006.

GIACHETTO, P. F.; HIGA, R. H. Bioinformática aplicada à agricultura. In: MASSRUHÁ, S. M. F. S. et al. Tecnologias da informação e comunicação e suas relações com a agricultura. Brasília: **Embrapa**, p. 67-83, 2014.

GORDO, S. M. C. Análise do Transcriptoma da Raiz de *Piper nigrum* L. (Piperaceae) utilizando sequenciamento com Tecnologia de RNA-Seq. **Tese: Doutorado em Genética e Biologia Molecular – Universidade Federal do Pará**, 90 f., 2012.

GORDO, S. M. C. et al. High-throughput sequencing of black pepper root transcriptome. **BMC Plant Biology**, p. 12-168, 2012.

GRADA, A.; WEINBRECHT, K. Next-generation sequencing: methodology and application. **The Journal of investigative dermatology**, v. 133, n. 8, p. 11, 2013.

GUERRA, R. A. T. **Cadernos Cb Virtual 5: Fisiologia Vegetal**. João Pessoa: Ed. Universitária, 2010.

IZUI K., et al. Phosphoenolpyruvate Carboxylase: A New Era of Structural Biology. **Annu Rev. Plant Biol.**, n. 55, p. 69–84, 2004.

JAHMU PBC. The King Of Spices – Black Pepper. 2018. Disponível em: <<https://jahmu.com/black-pepper/>>; Acesso em: 01 dez 2018.

JUNG, S. et al. The Genome Database for Rosaceae (GDR): 10 year update. **Nucl. Acids Res.**, v. 42, n. 1, p. 1237-1244, 2014.

JÚNIOR, M. T. S. et al. Transcriptoma de *Musa acuminata* no data Musa. Brasília: **Embrapa Recursos Genéticos e Biotecnologia**, p. 21, 2005.

KANEHISA, M.; GOTO, S. KEGG: Kyoto encyclopedia of genes and genomes. **Nucleic Acids Res**, n. 28, p. 27 –30, 2000.

KLOOSTERMAN, B. et al. Naturally occurring allele diversity allows potato cultivation in northern latitudes. **Nature**, v. 495, p. 246–250, 2013.

LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome biology**, 2009.

LEMOS, O. F. et al. Conservação e Melhoramento Genético da Pimenteira-do-reino (*Piper nigrum* L.) em associação com as técnicas de Biotecnologia. Belém: **Embrapa – Amazônia Oriental**, p.45, 2011.

LEMOS, O. F.; TREMACOLDI, C. R.; POLTRONIERI, M. C. Boas práticas agrícolas para aumento da produtividade e qualidade da pimenta-do-reino no Estado do Pará. Brasília: **Embrapa**, 2014.

LESK, A. M. **Introdução à Bioinformática**. Porto Alegre: Artmed, 2008.

LSPA – Levantamento Sistemático da Produção Agrícola. 2016. Disponível em: <<https://sidra.ibge.gov.br/home/lspa/brasil>>. Acesso em: 05 mar 2018.

MARGULIES, M et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437, p. 376-80, 2005.

MARTIN, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. **EMBnet.journal**, n. 17, p. 10-12, 2011.

MATIELLO, J. et al. Synthesis of phytoalexins in soybean and sorghum for extracts and tinctures belonging to three forest species. **Rev. Agro. Amb.**, v.9, n.3, p. 617-633, 2016.

MAURIZI, C. D.; POINSSOT, B. Role of glutathione in plant signaling under biotic stress. *Plant Signaling & Behavior*, n. 7, v. 2, p. 210–212, 2012.

MOREIRA, E. C. O. Análise da Expressão diferencial do transcriptoma de raiz de *Piper nigrum* L. em resposta a infecção por *Fusarium solani* f. sp. *piperis*. **Tese**: Doutorado em Genética e Biologia Molecular – Universidade Federal do Pará, 95 f., 2012.

MOREIRA, L. M. **Ciências Genômicas: fundamentos e aplicações**. Ribeirão Preto: Sociedade Brasileira de Genética, 2015.

MOREIRA, E. C. O. et al. Transcriptional profiling by RNA sequencing of black pepper (*Piper nigrum* L.) roots infected by *Fusarium solani* f. sp. *piperis*. **Acta Physiol Plant**, p. 39-239, 2017.

NCBI – National Center for Biotechnology Information. SRA: Sequence Read Archive, Collection of sequence data from next-generation sequencing technology for different organisms. 2018. Disponível em: <<https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/>>. Acesso em: 07 nov 2018.

NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Lehninger**. Porto Alegre: Artmed, 2014.

NHGRI – National Human Genome Research Institute. The Cost of Sequencing a Human Genome. 2018. Disponível em: <<https://www.genome.gov/sequencingcostsdata/>>. Acesso em: 07 nov 2018.

NODA, R. W.; DAMASCENO, C. M. B.; SOUSA, S. M. Anotação Funcional de Sequências com BLAST2GO. Sete Lagoas: **Embrapa Milho e Sorgo**, 2010.

OLIVEIRA, L. S. Método de identificação de Genes taxonomicamente restritos em dados de RNA-Seq em organismo não modelo. **Dissertação**: Programa de Pós-Graduação em Biotecnologia – Universidade Federal do Pará, 2015.

PARISY, V. et al. Identification of PAD2 as a c-glutamylcysteine synthetase highlights the importance of glutathione in disease resistance of *Arabidopsis*. *The Plant Journal*, n. 49, p. 159–172, 2007.

PAVLOPOULOS, G. A. et al. Unraveling genomic variation from next generation sequencing data. **BioData mining**, v. 6, n. 1, p. 13, 2013.

POPIN, R. V. Análise genômica e funcional da *Nodularia spumigena* CENA596 formadora de florações em tanques de produção de camarões. **Tese** (Mestrado em Ciências) – Universidade de São Paulo, Piracicaba, 86 f., 2017.

RAVEN, P. H.; EICHHORN, S. E.; EVERT, R. F. **Biologia Vegetal**. Rio de Janeiro: Editora Guanabara Koogan S. A., 8ª ed., 2014.

REICH, M. et al. GenePattern 2.0. **Nat. Genet.** n. 38, p. 500–501, 2006.

SANTOS, F. S. et al. Resposta antioxidante, formação de fitoquelatinas e composição de pigmentos fotoprotetores em *Brachiaria decumbens* Stapf submetida à contaminação com Cd e Zn. São Paulo: **Quím. Nova**, v. 34, n. 1, p. 16-20, 2011.

SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics**, n. 27, p. 863–864, 2011.

SMITH, A. **Oxford Dictionary of Biochemistry and Molecular Biology**. Oxford University Press, 2000.

SOUZA, G. B. Caracterização funcional dos genes BZIP105 e ARP (auxin repressed protein) de soja. **Tese**: Doutorado em Bioquímica Aplicada – Universidade Federal de Viçosa, f. 77, 2018.

STANGARLIN, J. R. A defesa vegetal contra fitopatógenos. **Scientia Agraria Paranaensis**, v. 10, n. 1, p. 18-46, 2011.

STEINDORFF, A. S. Genômica estrutural e funcional de fungos do gênero *Trichoderma*. **Tese** (Doutorado em Biologia Molecular) — Universidade de Brasília, Brasília, 96 f., 2016.

TAIZ, L. et al. **Fisiologia e Desenvolvimento Vegetal**. Porto Alegre: Artmed, ed. 6, 2017.

TAKACS, A. et al. Chapter 20: Virus-induced physiologic changes in plants. **Plant Virus–Host Interaction: Molecular Approaches and Viral Evolution**, p. 373-384, 2014.

THAMPI, S. M. Introduction to Bioinformatics. **Computational Engineering, Finance and Science**, 2009.

THE GENE ONTOLOGY CONSORTIUM. Gene ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25-29, 2000.

TRAPNELL, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, n. 28, p. 511–515, 2010.

VARON, J. C. F. et al. Transcriptoma do cafeeiro (*Coffea arabica* L.) durante a interação com *Hemileia vastatrix* Berk. & Br. **Tese**: Doutorado em Genética e Melhoramento - Universidade Federal de Viçosa, 56 f., 2017.

VASSILEV, D. et al. Application of bioinformatics in plant breeding. **Biotechnology and Biotechnological Equipment**, v. 19, p. 139–152, 2005.

VERLI, H. **Bioinformática**: da Biologia à flexibilidade molecular. 1ª ed., São Paulo: SBPq, 2014.

WALSH, B.; HENDERSON, D. Microarrays and beyond: what potential do current and future genomics tools have for breeders?. **Journal of Animal Science**, v.82, p. 292–299, 2004.

WALTERS, D.; NEWTON, A.; LYON, G. Induced resistance for plant defence – a sustainable approach to crop protection. Oxford: **Blackwell**, p. 258, 2007.

WANG, L. et al. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. **Bioinformatics**, n. 26, p. 136–138, 2010.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. London: **Nature Review Genetics**, v. 10, n. 1, p. 57-63, 2009.

WEISS, V. A. Estratégias de Finalização da Montagem do Genoma da Bactéria Diazotrófica Endofítica *Herbaspirillum seropedicae* SmR1. **Dissertação**: Mestrado em Ciências/Bioquímica – Universidade Federal do Paraná. 72 f., 2010.

WIECZOREK, E. M.; LEAL, E. Caminhos e Tendências do uso de Banco de Dados em Bioinformática. Palmas: **Curso de Sistemas de Informação – Centro Universitário Luterano de Palmas (CEULP)**, p. 1–10, 2002.

WINTER, H.; HUBER, S. C. Regulation of Sucrose Metabolism in Higher Plants: Localization and Regulation of Activity of Key Enzymes. **Critical Reviews in Biochemistry and Molecular Biology**, n. 35, p. 253-289, 2000.

XIA, Z. et al. Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. **Proceedings of the National Academy of Sciences**, v. 109, n. 32, p. 2155–2164, 2012.

XU, Y. et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. **Nature Genetics**, v. 45, n. 1, p. 51–58, 2013.

XU, X. et al. Genome sequence and analysis of the tuber crop potato. **Nature**, v. 475, n. 7355, p. 189–195, 2011.