

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Análise do Transcriptoma e da expressão gênica de importantes fases de desenvolvimento da broca do café (*Hypothenemus hampei*) e avaliação de potenciais genes alvos aplicados ao seu controle

Dissertação de Mestrado

Helena Ribeiro Barbosa

Porto Alegre

2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Análise do Transcriptoma e da expressão gênica de importantes fases de desenvolvimento da broca do café (*Hypothenemus hampei*) e avaliação de potenciais genes alvos aplicados ao seu controle

HELENA RIBEIRO BARBOSA

Dissertação submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Mestre em Biologia Celular e Molecular

Orientadora: Profa. Dra. Maria Fátima Grossi de Sá

Porto Alegre, março de 2019

Instituição e fonte financiadora

Este trabalho foi desenvolvido no Laboratório de Interação Molecular Planta Praga na Embrapa CENARGEN em Brasília e na Universidad de Caldas em Manizales, Colômbia.

As agências de financiamento deste trabalho foram CAPES, CNPq, FAPDF e Consórcio Café.

Agradecimentos

Em primeiro lugar agradeço a minha família, presente ao meu lado em todas as situações, segurando todas as barras que surgiram ao longo desses dois anos. Minha mãe, não apenas pela orientação acadêmica, mas apoio incansável e incondicional, pelo colo e renovação de fé quando muitas coisas pareciam perdidas. Aos meus irmãos, Rafael e Juliana por serem sempre minha base de ética e profissionalismo, meu norte. Ao meu cunhado Rodrigo pela camaradagem, pelo papo e pela cerveja no momento certo. Aos meus sobrinhos Amália, Lucas e Joel por provocarem em mim o desejo de sempre me superar para que eu seja o norte deles também. Ao Jorge por tornar muitas vezes as coisas mais leves, me ajudando a ver os problemas sob outra ótica. A “tia” Vanda, pelo chimarrão sincero, acolhimento em sua casa, risadas e compreensão. Minha gratidão eterna a vocês.

Agradeço a minha família em Porto Alegre, meus tios, primas, primo, Vera e Any, não apenas pela acolhida, mas por todo o convívio, amizade, conselhos acadêmicos, chimarrão, com muito apoio afetivo nesse período. Esse retorno a Porto Alegre não teria sido o mesmo sem vocês.

Agradeço ao Programa de Pós Graduação em Biologia Celular e Molecular e ao Centro de Biotecnologia da UFRGS por toda base acadêmica que levarei para a vida. Em especial à amiga Silvia Centeno, meu anjo da guarda em Porto Alegre. Sua amizade, atenção, carinho e gargalhadas (essas, principalmente) ficarão guardadas comigo para sempre.

Agradeço a toda equipe do Laboratório de Interação Molecular Planta Praga que esteve comigo nesses três anos me ajudando profissionalmente e tornando o dia a dia mais agradável. Em especial, Roberta, Dijair, Raquel e Leonardo (rostos conhecidos nesse retorno ao laboratório); Ana Gabriela, Ludovic, Rayssa, Hudson e Clidia (a nova geração que fez renovar minha fé no mundo acadêmico); Reneida, Elínea e Mayara (meu querido Supremo, tão importante para nossa saúde mental); Daniel (pelo conhecimento compartilhado sobre a broca), e Fabricio Arraes (pela ajuda inestimável desde o início dessa jornada, me trazendo de volta ao jogo). O cotidiano com vocês fez com que tudo tenha sido muito mais prazeroso e estimulante, e até mesmo desafiador.

Um agradecimento especial aos amigos Bruno Paes, por todo o apoio profissional, pelos *happy hours* na hora certa e por todo “*K’s fire*” que levaremos para a vida. E Alexandre Firmino que, nem nos meus melhores sonhos, imaginei ter a sua presença nessa reta final do mestrado.

Agradeço aos meus amigos, tão necessários e importantes, que me ajudaram a perceber que sempre existe mais do que nossos olhos conseguem ver. Thais, pelos incontáveis e inestimáveis cafés, tão necessários. Débora, a surpresa boa de Porto Alegre, pela ajuda nos momentos de amizade que passamos juntas, pela renovação de lembranças boas em Porto Alegre. Ao querido “Quinteto”, do qual sou parte, amigos de “infância” de Brasília, meu porto seguro nessa cidade, meu pé no chão quando essa loucura do ambiente acadêmico se tornava insuportável: Ângela, Carol, Nati, Galton. Obrigada!

Agradeço a equipe do IAPAR em Londrina/PR, à Dra. Ana Maria Meneguim e sua equipe pela acolhida, profissionalismo e todo o aprendizado sobre o manejo com a broca do café.

Agradeço à Dra. Maria Cristina Mattar e a Dra. Maria Fátima Grossi de Sá, pela oportunidade de fazer parte da equipe de pesquisa, pela orientação, apoio e ensinamentos ao longo desse período.

“The answer, my friend, is blowing in the wind”

(Bob Dylan)

Sumário

Lista de Figuras	9
Lista de Tabelas	11
Abreviaturas	12
Resumo	15
Abstract	16
1. Introdução Geral	17
1.1 Café: Características gerais e importância para a economia	17
1.2 A Broca do Café e estratégias de controle	19
1.3 A Biotecnologia como alternativa de controle de pragas	23
1.4 RNA interferente (RNAi) como ferramenta biotecnológica	24
2. Objetivos	28
2.1 Objetivo geral	28
2.2 Objetivos específicos	28
3. Sequenciamento de transcritomas	29
3.1 Sequenciamento de ácidos nucleicos	29
3.2 Sequenciamento de segunda geração	31
3.3 Sequenciamento de RNA	32
3.4 Transcritomas de insetos	34
3.5 Análise de expressão diferencial de genes	36
3.6 Estudo da expressão gênica em insetos	37
4. Material e Métodos	39
4.1 Extração de RNA e sequenciamento	39
4.2. Montagem De Novo e Anotação Funcional	40
4.3 Análise da Expressão Diferencial (DEGs)	41
4.4 Enriquecimento de Ontologia Gênica	42

4.5 Análise filogenética	42
4.6 Validação por PCR em tempo real (RT-PCR)	43
5. Resultados	44
5.1 Extração de RNA total	44
5.2 Análise do Transcritoma	45
5.2.1 Pré-processamento	45
5.2.2 Montagem De Novo e anotação funcional	47
5.2.3 Expressão diferencial de genes	49
5.2.4 Análise de enriquecimento por ontologia gênica	52
5.2.5 Enriquecimento enzimático	53
5.2.6 Análise de DEGs entre machos e fêmeas	54
5.2.7 DEGs entre fases de larva e adulta	55
5.2.8 Validação de dados de RNA-seq por PCR em tempo real	58
6. Discussão	60
7. Conclusões e Perspectivas	66
8. Referências bibliográficas	68

Lista de figuras

- Figura 1.** A broca do café (*H. hampei*). Fêmea de *H.hampei* sobre um grão de café, mostrando o tamanho do inseto adulto em relação ao grão de café (adaptado de: The Dow Chemical Company (1995- 2003) 20
- Figura 2.** Principais fases de desenvolvimento da broca da café (*H. hampei*). A) Ovos (Postura). B) Larva de segundo ínstar. C) Pupa. D) Adulto 21
- Figura 3.** Danos causados pelo inseto *H.hampei* em grãos de café. A esquerda, grãos saudáveis. A direita é possível notar perfurações nos grãos causados pelo inseto. Fonte: CaféPoint (2013) 22
- Figura 4.** Esquema de um experimento padrão de RNA-seq de uma levedura é mostrado na ilustração. Longas moléculas de RNAs são convertidas em uma biblioteca de fragmentos de cDNA, por meio de fragmentação do RNA ou do DNA. Adaptadores de sequenciamento (Azul), são adicionados em cada fragmento de cDNA, obtendo sequências curtas para cada cDNA. Os *reads* resultantes são alinhados com um genoma ou transcrito de referência, e classificados em três tipos: *reads* exônicos, *reads* de ligação e *reads* com cauda poli-A. Esses três tipos de *reads* são usados para gerar um perfil de expressão para cada gene, como representado na parte inferior; uma fase de leitura aberta (ORF: *Open Reading frame*). Wang *et al.* (2009)33
- Figura 5.** Géis de agarose com amostras de RNA total das fases de desenvolvimento da broca do café. 1º gel - amostras de larva L2 2º gel – amostras de adulto (fêmea) e 3º gel – amostras de adulto (macho)44
- Figura 6.** Espécies com principais *hits* após BLASTx. O maior número de *contigs* teve alta similaridade com o inseto coleoptera, *Dendroctonus ponderosae*.....48
- Figura 7.** Resumo da classificação dos transcritos de acordo com as bases de dados *Gene Ontology* (GO), Famílias de Proteínas (Pfam) e *Kyoto Encyclopedia of Genes e Genomes* (KEGG). (a) A classificação GO resume uma distribuição multinível dos dez principais termos anotados para cada uma das três principais categorias GO (componente celular, função molecular e processo biológico). (b) Os dez domínios mais abundantes da classificação Pfam. (c) Classificação enzimática mostrando as principais vias (eixo y) de quatro categorias funcionais (metabolismo, processamento de informações genéticas, processamento de informações ambientais e sistemas orgânicos) e a porcentagem de transcritos em referência ao número total de transcritos anotados usando o banco de dados KEGG (eixo x). Os números à direita das barras representam o número total de transcritos anotados para cada caminho. A) Metabolismo, B) processos de informação genética, C) processos de informação ambiental, e D) sistemas organizacionais.....49
- Figura 8.** Comparações de genes diferencialmente expressos entre os diferentes estágios de desenvolvimento de *H. hampei*. (a) Número de transcritos e unigenes diferencialmente expressos em larva L2 e macho em referência a fêmea. As setas vermelhas e azuis representam genes regulados positivamente e negativamente, respectivamente. A barra de escala mostrada em fotos de insetos corresponde a 500 µm..... 51

(b) *Heatmap* e agrupamento de todos os DEGs. As linhas representam transcrições únicas e as colunas representam as réplicas de cada estágio de desenvolvimento. Os genes foram agrupados com base na similaridade de expressão (distância euclidiana). As cores azul e vermelha no mapa de calor correspondem à expressão gênica relativa baixa e alta, respectivamente.....51

Figura 9. Ontologia gênica e análise de enriquecimento enzimático para cada conjunto de DEGs. Top 15 termos GO com os maiores valores de enriquecimento para cada comparação par a par.....53

Figura 10. Ontologia gênica e análise de enriquecimento enzimático para cada conjunto de DEGs. Taxas de descoberta para cada enzima e as vias associadas para cada comparação na análise de enriquecimento. Círculos pretos indicam transcritos regulados em larvas L2. Círculos brancos indicam transcritos regulados positivamente em machos. Triângulos pretos indicam transcritos regulados negativamente em larvas L2.....54

Figura 11. *Heatmap* mostrando as diferenças de expressão entre as bibliotecas para fêmeas, larvas e machos de *H.hampei* para genes que codificam proteínas relacionadas com TEs e proteínas relacionadas com reprodução. TEs: *transposable element-like proteins*, TSs: *transposase-like proteins*.....55

Figura 12. Número de unigenes (eixo x) *up-regulated* em larva L2 e macho em referencia a fêmea para cada grupo de genes relacionados ao metabolismo de cutícula e quitina (eixo y). CPAPs: *cuticular proteins analogous to peritrophins*.....57

Figura 13. Relação filogenética dos genes OBP em *H. hampei* e em outros insetos da ordem coleóptera.....57

Figura 14. Resultados de RT-qPCR para genes diferencialmente expressos em diferentes estágios de desenvolvimento do CBB. Letras maiúsculas diferentes indicam uma diferença significativa entre os estágios ($P < 0,05$). ETE: hormônio esterase juvenil, JHEH: hormônio juvenil hidrolase, CPAP: proteína cuticular análoga às peritrofinas, HexI: hexosaminidase I, Chi10: quitinase 10.....59

Lista de tabelas

Tabela 1. Número total de *reads* pareados nas bibliotecas de cada fase de desenvolvimento de *H.hampei* obtido pelo sequenciamento por Illumina.....**46**

Tabela 2. Resumo estatístico do transcrito de *H.hampei*.....**46**

Abreviaturas, símbolos e unidades

.fa – Formato fasta

°C – Graus celsius

µg/µl – Microgramas por microlitro

µg/cm³ – microgramas por centímetro cúbico

µM – Micromolar

Bt – *Bacillus thuringiensis*

CBB – *Coffee Berry Borer*

cDNA – DNA complementar

cm – Centrimetros

Contigs – Sequências contíguas

CTNBio – Comissão Técnica Nacional de Biossegurança

ddNTP –Dideoxynucleotídeo

DEGs – *Differential expressed genes*

dsRNA – RNA dupla fita

E.C – Código de classificação enzimática

E2F – Fator de alongação II

Fc – *Fold change*

FPKM – *Fragments Per Kilobase per Million mapped reads*

g – gramas

GAPDH – Gliceraldeído 3-fosfato desidrogenase

GC – Guanina-Citosina

GM – Geneticamente modificado

GO – *Gene ontology*

GSTs – Glutathionas S-transferases

h – Horas

KEGG – Kyoto Encyclopedia of Genes and Genomes

Km – constante de Michaelis

MCT – Média do comprimento total

mg – miligramas

MIP – Manejo Integrado de Pragas
miRNAs – Micro-RNAs
mL – Mililitros
mm – Milímetros
mM – Milimolar
MMol/L – Micromol por litro
MP – Membrana peritrófica
mRNA – RNA mensageiro
NCBI – *National Center for Biotechnology Information*
ncRNA – RNA não-codificante
ng – nanogramas
NGS – *Next Generation Sequencing*
nr – Não redundante
OGM – Organismo Geneticamente Modificado
ORF – *Open Reading Frame*
P450 – Citocromo P450
pb – pares de bases
PCR – *Polymerase Chain Reaction*
Primers – Oligonucleotídeos iniciadores
RT-PCR – PCR em tempo real
R² – Coeficiente de determinação
RdRP – RNA-polimerase dependente de RNA
Reads – Sequências de nucleotídeos curtas
RISC – *RNA-induced silencing complex*
RNAi – RNA interferente
RPKM – *Reads Per Kilobase per Million mapped reads*
RPS10 – Proteína ribossomal S10
RPS18 – Proteína ribossomal S18
rRNA – RNA ribossomal
SID – *Systemic RNA Interference Deficiency*
siRNA – *Small Interfering RNA*
U – Unidades
UNI – *Contigs únicos*
UV – Ultravioleta

μg – Microgramas

μL – Microlitros

Resumo

A produção de café (*Coffea spp.*) possui uma importância crucial na economia mundial estimada em cerca de US\$ 173 bilhões por ano. Neste cenário, o Brasil segue como o maior produtor de café, com colheita de 60 milhões de sacas/ano. Um dos principais desafios para os produtores é o manejo da broca-do-café (CBB), *Hypothenemus hampei*, que é o principal inseto praga das plantações em todo o mundo. Várias estratégias de controle, incluindo o uso de agentes de controle biológico, inseticidas microbianos e uma variedade de práticas de saneamento pós-colheita, têm sido usadas para diminuir os danos causados por esse inseto em todos os países produtores que são afetados. Devido ao comportamento endofítico da broca do café, os atuais métodos de manejo aplicados são considerados ineficientes. Alternativas biotecnológicas, como RNA interferente (RNAi), têm sido propostas nos últimos anos para controlar os danos causados por pragas nas plantações. No entanto, uma compreensão completa da genética dessas pragas é indispensável para o sucesso da aplicação de tecnologias emergentes. Neste trabalho, foi utilizada a tecnologia RNA-seq para obter a sequência do transcrito de três estágios de desenvolvimento da broca do café (larvas, fêmeas e machos), o que contribuiu para aumentar a compreensão das características moleculares do ciclo de vida do inseto. O sequenciamento foi executado usando uma plataforma Illumina, e um conjunto de dados foi gerado através da metodologia *de novo*. Análises de expressão diferencial foram realizadas para comparar os perfis de expressão nos estágios de desenvolvimento. A montagem final produziu mais de 50.000 transcritos agrupados em aproximadamente 29.000 unigenes, dos quais 4.664 transcritos foram diferencialmente expressos. Dentre os potenciais alvos biotecnológicos, foram identificados genes relacionados com processos fisiológicos fundamentais do inseto, como determinação sexual, funções quimiossensoriais, detoxificação, metabolismo de celulose, transporte de moléculas, proteínas cuticulares e enzimas envolvidas no metabolismo de quitina, entre outros. Além disso, alguns desses genes relacionados a funções cruciais demonstraram forte regulação entre as fases reprodutiva e não reprodutiva da broca-do-café. Da mesma forma, elementos transponíveis parecem ser induzidos nos machos. Juntos, esses dados sugerem o papel importante de vários genes como elementos reguladores no desenvolvimento desse inseto, podendo ser alvo de estratégias biotecnológicas de controle da broca do café.

Abstract

Coffee (*Coffea* spp.) production is a widespread industry estimated to be worth approximately US\$173 billion. In this context, Brazil continues to be the largest coffee producer, with a harvest of 60 million sacs/year. One of the main challenges for coffee producers is the management of the coffee berry borer (CBB), *Hypothenemus hampei*, which is the primary arthropod pest of coffee plantations worldwide. Several control strategies, including the use of biological control agents, microbial insecticides and a variety of post-harvest sanitation practices, have been used to reduce the damage caused by this insect in all producing countries that are affected. Current control strategies have been proved inefficient, due to the endophytic behavior of the CBB. Biotechnological alternatives, including RNA interference (RNAi) have been proposed in recent years to control insect pests. However, better understanding about genetics of the target pests is indispensable for the successful application of emerging technologies. In this work, the RNA-seq technology was used to obtain the transcriptome of three developmental stages of the CBB (larva, female and male), which contributes to increase the understanding about molecular features of the CBB life cycle. Transcriptome was sequenced using Illumina and assembled *De novo*. Differential expression analysis was performed between the developmental stages. The final assembly produced more than 50,000 transcripts grouped into 29,434 unigenes of which 4,664 transcripts were differentially expressed. Among the potential target genes, genes related to insect fundamental physiological processes, such as sex determination, chemosensory functions, detoxification, cellulose metabolism, transport of molecules, cuticular proteins and enzymes involved in the metabolism of chitin were identified. Besides that, genes related to crucial physiological functions such as digestion and detoxification were identified to be strongly regulated between the reproductive and non-reproductive phases of the CBB. Notably, transposable element-related proteins (TEs) and reproductive organ-related proteins were mainly up-regulated in males. Together, all data suggest the important roles of several genes as regulatory elements in the CBB's development.

1. Introdução

1.1 Café: características gerais e importância para a economia brasileira

O café, pertencente à família *Rubiaceae*, foi descrito pela primeira vez por Linnaeus, em 1753. Sua origem geográfica é atribuída ao nordeste africano, principalmente à região da Etiópia, onde o consumo era realizado originalmente pela ingestão do fruto, no preparo de refeições com a polpa macerada do grão, além da utilização das folhas para o preparo de chás (ABIC, 2018). A propagação do cultivo aconteceu com os árabes que dominaram rapidamente a técnica de plantio e preparação. No oriente, popularizou-se principalmente na Turquia, no século XVI. A partir disso, difundiu-se na Europa, a partir de Veneza, no século XVII (MARTINS, 2008). Foi na tentativa dos países europeus de conseguirem produzir o grão em suas colônias que o café foi inserido na América, na região da Guiana Francesa, já no século XVIII.

No Brasil, os primeiros relatos sobre o café datam do ano de 1727, quando foi introduzido pelo oficial português Francisco de Mello Palheta, vindo da Guiana Francesa com algumas mudas que recebera de presente do governador de Caiena. Apesar desses primeiros grãos terem sido cultivados sem dificuldades na região do Pará, foi no Rio de Janeiro, em 1781, que o café começou a ser produzido em maior escala, originando um novo ciclo econômico no Brasil. Com o fim do ciclo da mineração do ouro no país, o café ganhou espaço no Vale do Paraíba, em São Paulo, onde em um século, por volta de 1880, já era o principal produto nacional (TAUNAY, 1939; MARTINS, 2008)).

O café (*Coffea arabica*) sempre foi de grande importância para a economia brasileira, contribuindo decisivamente para a industrialização do país, graças ao capital acumulado com as exportações do produto. Devido às nossas condições climáticas, o cultivo de café, desde a sua introdução no estado do Maranhão em 1939, se espalhou rapidamente, com produção voltada, inicialmente, para o mercado doméstico. Em sua trajetória pelo Brasil, o cultivo dessa planta atravessou o país, sendo plantado em diversos estados brasileiros, nas regiões nordeste, sudeste e sul. Num espaço de tempo relativamente curto, o café passou de uma posição relativamente secundária para a de produto-base da economia brasileira (NEVES, 1974).

Hoje, o Brasil possui quinze regiões produtoras de café, que são favorecidas pelas suas condições climáticas. Nas regiões com altitudes acima de 800 metros, principalmente nos estados de Minas Gerais, São Paulo, Paraná, Bahia, Rio de Janeiro e parte do Espírito Santo predomina o cultivo de café arábica, que possui uma maior qualidade, sendo um produto mais fino, quando comparado com o café robusta ou conilon (*Coffea Canephora*). Este, por sua vez, é usado para a fabricação de cafés solúveis ou misturado com o arábica e apresenta uma menor acidez, mas um teor maior de cafeína. Seu cultivo é predominante nos estados do Espírito Santo, Rondônia e em parte da Bahia e de Minas Gerais (ABIC, 2017; MAPA, 2018; ICO, 2018).

Essas quinze regiões produtoras são responsáveis pelo Brasil ser hoje o maior produtor de café, com colheita superior a 60 milhões de sacas/ano. Além disso, é o maior exportador mundial e o segundo maior consumidor, ficando atrás apenas dos Estados Unidos (CONAB, 2018).

Apesar dessa importância, são inúmeros os problemas que afetam a produção dessa *commodity*. Dentre esses problemas, os insetos praga possuem uma importância

elevada devido ao seu impacto na produção de café. As principais pragas que afetam a cafeicultura são o bicho mineiro - *Perileucoptera coffeella* (Lepidoptera: Lyonetiidae), a broca do café - *Hypothenemus hampei* (Coleoptera: Scolitidae), o ácaro vermelho - *Oligonychus ilicis* (Acari: Tetranychidae) e as cigarrinhas (Hemiptera: Cicadellidae).

1.2 A Broca do café e estratégias de controle

A broca do café (*H.hampei* Ferrari, 1867) é um dos insetos praga mais importantes da cafeicultura, sendo responsável por perdas na economia mundial de US\$ 500 milhões por ano (INFANTE *et al.*, 2018). Endêmico da África, a broca do café entrou na Indonésia em 1908, e no Brasil em 1913 (CEJA-NAVARRO *et al.*, 2015). Entre 1913 e 1924, o inseto se alastrou pelos cafezais de Campinas, em São Paulo e por municípios vizinhos. Os primeiros prejuízos foram observados na colheita da safra de 1924 e a partir desse ano, a broca espalhou-se por todas as regiões cafeeiras do Brasil (BENASSI, 1995; MORAES, 1998; SOUZA e REIS, 1997). Para os dois maiores produtores mundiais, Brasil e Colômbia, o prejuízo anual causado pela broca do café chega a US\$ 300 milhões (INFANTE *et al.*, 2013; VEGA *et al.*, 2015), com a presença desse inseto em quase todas as regiões onde o café é cultivado.

Esse inseto pertence à ordem Coleoptera, família *Curculionidae*, possui coloração preta quando adulto, e diferença de tamanho entre o macho e a fêmea, medindo aproximadamente 0.99 - 1.3 mm e 1.6 - 1.9 mm, respectivamente (figura 1).



Figura 1. A broca do café (*H. hampei*). Fêmea de *H.hampei* sobre um grão de café, mostrando o tamanho do inseto adulto em relação ao grão de café (adaptado de: The Dow Chemical Company (1995- 2003))

O ciclo de vida da broca do café dura entre 20 a 30 dias (figura 2). Trata-se de um inseto holometábolo, com quatro estágios bem definidos (ovo, larva, pupa e adulto), e características bem distintas entre machos e fêmeas. Além da diferença de tamanho citada anteriormente, influenciada por diversos fatores, como temperatura e umidade, a duração média de vida do macho varia entre 30 a 40 dias, enquanto que da fêmea pode chegar a 156 dias. Além disso, a proporção encontrada de macho para fêmea em grãos de café é de 1 para 10 (VEGA *et al.*, 2009; GÓMEZ *et al.*, 2015; VEGA *et al.*, 2015).

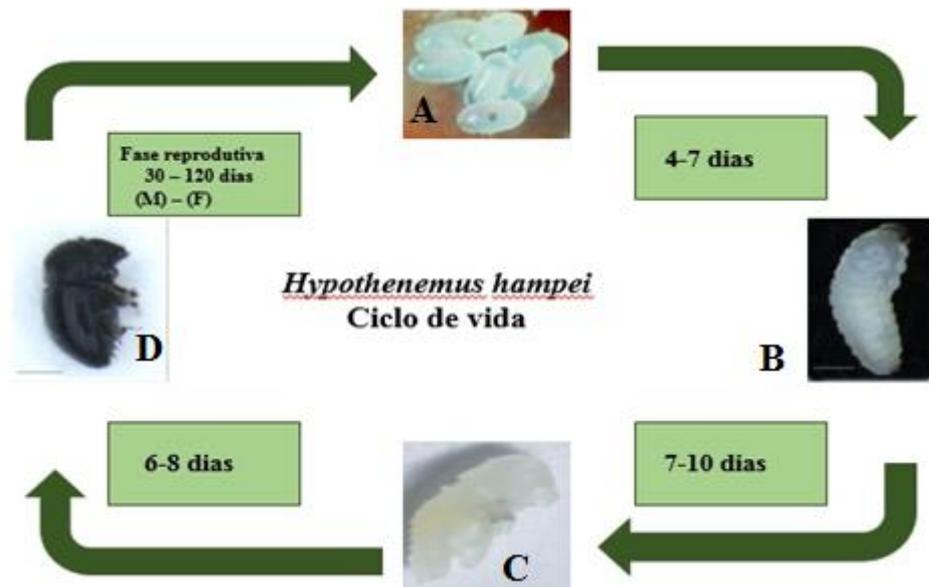


Figura 2. Principais fases de desenvolvimento da broca do café (*H. hampei*). A) Ovos (Postura). B) Larva de segundo ínstar. C) Pupa. D) Adulto. O inseto macho não possui o segundo instar da fase larval.

A broca do café é um inseto de hábito endofítico, que passa a maior parte da sua vida dentro do grão de café, o que dificulta a sua eliminação do cafeeiro (VEGA *et al.* 2009). O controle realizado por inseticida, além de oneroso é ineficaz, uma vez que, para surtir efeito, o produto precisa entrar em contato com o inseto. O inseto utiliza o grão de café tanto para a alimentação quanto para a reprodução, o que causa danos diretos e indiretos (figura 3). Cita-se a perda de peso do fruto, devido a sua destruição pelas larvas, perda da qualidade, pela depreciação do produto, queda prematura de frutos quando perfurados e o apodrecimento de sementes em frutos broqueados que apresentam maturação forçada, caindo precocemente no chão (GUHARAY e MONTERREY, 1997; SOUZA e REIS, 1997; BENASSI e CARVALHO, 1994).



Figura 3. Danos causados pelo inseto *H.hampei* em grãos de café. A esquerda, grãos saudáveis. A direita é possível notar perfurações nos grãos causados pelo inseto. Fonte: CaféPoint (2013)

São diversas as técnicas de manejo, aplicadas para tentar diminuir a presença do inseto nos cafezais. Hoje, a mais eficiente é o controle cultural. Nesse tipo de manejo, os cafezais precisam ser plantados com espaçamento suficiente para que ocorra presença de luz e arejamento, propiciando assim uma diminuição da umidade do ar em seu interior, tornando o ambiente desfavorável para o desenvolvimento do inseto (ABIC, 2017). Essa técnica permite também a circulação de tratores que possuem pulverizadores para aplicação de defensivos químicos.

Atualmente, existem mais de 10 inseticidas no mercado que atuam no controle da broca do café. Infelizmente, nenhum deles possui total eficiência, uma vez que para realizar o controle do inseto, o inseticida precisa entrar em contato com ele, o que é difícil devido ao comportamento da broca de completar a maior parte do seu ciclo de vida dentro do grão (VEGA *et al*, 2017). Além disso, o mercado consumidor de café, tanto nacional quanto mundial se encontra cada dia mais rigoroso quanto ao uso de defensivos agrícolas. Por esse motivo, se torna urgente o desenvolvimento de pesquisas que tenham como

objetivo principal a erradicação desse inseto praga dos cafezais, como, por exemplo, o uso da biotecnologia, que utiliza processos celulares e moleculares na realização de produtos e tecnologias que apresentam uma eficiência maior na proteção dos cafezais contra o ataque de pragas, sem o prejuízo financeiro e ambiental que o uso de defensivos químicos causam nessas e em outras culturas.

1.3 A Biotecnologia como alternativa de controle de pragas

Como alternativa ao controle cultural, químico e biológico da broca do café, surgiu o uso de ferramentas biotecnológicas para a produção de plantas resistentes a pragas. O uso da biotecnologia se torna cada vez mais necessária, pois a identificação e seleção de genes de interesse por meio do uso de marcadores moleculares aceleraram os processos convencionais de obtenção de indivíduos recombinantes. Os avanços na área de, Biologia Molecular, Bioinformática e Engenharia Genética permitiram o desenvolvimento de organismos geneticamente modificados (OGMs), dando origem a plantas resistentes a herbicidas, insetos e doenças, que acumulam sacarose e plantas adaptadas a ambientes adversos (GAZARRA *et al*, 2017).

Em insetos da ordem *Lepidoptera*, toxinas *Cry* de *Bacillus thuringiensis* (*Bt*) já demonstraram ser uma tecnologia eficiente e extremamente específica, (DAVOLOS *et al.*, 2015). Porém, apesar de ser uma tecnologia eficaz, já existem casos de insetos resistentes em outras culturas transgênicas que utilizam culturas *Bt* (DAVIES *et al*, 2014). Dessa maneira, o desenvolvimento de novas estratégias moleculares é importante para evitar a ocorrência de resistência por parte de pragas, e também permitir o controle

populacional de insetos de uma maneira segura tanto para a saúde humana, quanto para o meio ambiente (DAVOLOS *et al.*, 2015).

Em relação ao café, a produção de plantas transformadas com o gene de inibição de α -amilase (α AI-1), que codifica para um inibidor que mostrou ter atividade *in vitro* para a α -amilase do *H. hampei* (BARBOSA *et al.*, 2010; ALBUQUERQUE *et al.*, 2015), gerando plantas transgênicas resistentes a insetos, representou uma alternativa muito promissora. Além disso, Aguilera e colaboradores (2011), verificaram mortalidade de até 60% em larvas 1º e 2º instar de *H.hampei*, com a tecnologia de RNAi.

Nesse contexto, a prospecção e produção de moléculas entomotóxicas para superexpressão em plantas, assim como outras abordagens biotecnológicas, como a supressão funcional de genes por inativação, representam tecnologias com grande potencial para desenvolver culturas resistentes a insetos praga.

1.4 RNA interferente (RNAi) como ferramenta biotecnológica

Dentro do contexto de se criar técnicas altamente eficientes contra a broca do café, que não atinjam a qualidade do fruto nem a produção de grãos, a biotecnologia surge como uma ferramenta que se torna cada dia mais necessária. A identificação de genes de interesse por meio de marcadores moleculares, por exemplo, aceleraram os processos convencionais de obtenção de indivíduos recombinantes (DE ASSIS FONSECA, 2013). Esses genes de interesse também se tornam importantes para a técnica de RNA interferente. Por sua alta especificidade e pouco efeito em genes não alvos, essa tecnologia é tida como uma alternativa promissora no controle de insetos-praga (BAUM *et al.*, 2007).

A interferência mediada por RNA é um mecanismo de silenciamento gênico altamente conservado em diversos organismos. A diferença que ocorre entre eles é devido ao tipo de RNA ou de proteínas argonautas envolvidas no processo. A descrição desse mecanismo ocorreu pela primeira vez em *Caenorhabditis elegans*. Em um estudo realizado por Guo e Kemphues, em 1995, foi investigada a função do gene *par-1*. Esses animais ao serem tratados com RNA anti-sense para *par-1*, observou-se que o gene, codificante de uma quinase de Ser/Thr, não se mostrou expresso. Porém, os autores verificaram que tanto as moléculas de RNA anti-sense como de RNA sense eram capazes de silenciar o gene em estudo. Além disso, o mesmo efeito era observado quando ambas as fitas eram associadas. Nenhuma explicação plausível foi apresentada pelos autores na época, entretanto, uma das hipóteses levantadas referia-se à composição das moléculas de RNA usadas no experimento. Essa composição poderia conter moléculas de RNA dupla fita (dsRNA) que seriam capazes de induzir o silenciamento gênico percebido nos ensaios. Essa hipótese serviu de base para o estudo realizado por Fire *et al*, que viria a comprovar a ocorrência de um efeito sinérgico sobre a expressão de genes alvos quando moléculas de RNA *sense* e *anti-sense* eram introduzidas ao mesmo tempo em *C. elegans*.

Após algumas décadas, esse mecanismo de silenciamento é bem descrito em diversos organismos. Apesar de algumas diferenças, a absorção do RNA dupla fita é bem semelhante em todos eles. De maneira geral, esse RNA dupla fita (dsRNA) é reconhecido por um complexo RISC (*RNA Induced Silencing Complex*) formado por helicases, proteínas ligantes a dsRNA, e pela enzima Dicer que cliva o dsRNA em siRNA, moléculas com aproximadamente 21-28 nucleotídeos numa reação que é dependente de ATP. Após a separação da fita siRNA, uma delas, que garante a especificidade à sequência-alvo, é retida no complexo RISC (PRICE; GATEHOUSE, 2008). Essa molécula siRNA, que é fita simples, leva à ativação de outras proteínas da família das

argonautas, e é então direcionada para sequências complementares de RNA transcrito. Algumas proteínas argonautas do complexo possuem funções de RNase H e degradam o mRNA em pouco mais de 10 nucleotídeos a partir da extremidade 3' do siRNA. No momento em que as duas fitas dessa molécula são separadas pelas helicases no complexo RISC, a fita simples pode ser dissociada do complexo e servir de oligonucleotídeo iniciador da síntese de novas moléculas de dsRNA. A fita simples hibridiza com o RNAm complementar e uma RNA polimerase dependente de RNA (RDRP- *RNA- Dependent RNA Polymerase*) levando a uma reação de síntese de novos dsRNAs, que voltarão a ser substrato para a Dicer e podem ser exportados de uma célula à outra pela ação das proteínas transmembrânicas, SID, tornando o silenciamento sistêmico e hereditário, o que não ocorre em insetos (GHILDIYAL; ZAMORE; 2000). Acredita-se que os insetos possuam componentes da maquinaria de RNAi responsáveis por participarem da amplificação e da resposta sistêmica, porém a maquinaria de RNAi é pouco descrita ao longo da classe Insecta e a maioria das diferenças presentes entre as diferentes ordens ainda permanece desconhecida (GAZARRA *et al.*, 2017; SCOTT *et al.*, 2013).

Uma abordagem importante para identificar genes alvos que tenham potencial para serem validados por RNAi, é o sequenciamento de genomas e transcritomas dos organismos alvos. Essa estratégia tem por finalidade construir um banco de dados que contenha genes e informações importantes sobre a sobrevivência do inseto, como é o caso deste trabalho. Para esse processo é importante o uso de tecnologias de sequenciamento de segunda geração (NGS), como a técnica de RNA-seq. A análise de transcritomas por meio de RNA-seq permite comparar o nível de expressão de genes alvo em situações específicas ou fases de desenvolvimento, ajudando a compreender dessa maneira os

diferentes mecanismos moleculares relacionados com a sobrevivência do inseto (ZHANG e YUAN, 2013; KOCH e KOGEL, 2014; RODRIGUES et al., 2016; JOGA, 2016).

2. Objetivos

2.1 Objetivo geral

Obter e analisar o transcrito do inseto praga *H. hampei* visando à obtenção de genes essenciais para o seu desenvolvimento, e sua validação para o controle do inseto pelo uso da tecnologia de RNA interferente.

2.2 Objetivos específicos

Realizar o sequenciamento em plataforma Illumina para obter os transcritos de três importantes fases de desenvolvimento do inseto (larvas 2º instar, fêmeas e machos);

Analisar as sequências obtidas e construir um banco de dados de genes diferencialmente expressos nas diferentes fases do ciclo de vida da broca do café, a partir da anotação das sequências e montagem do transcrito;

Selecionar genes para validação do transcrito e para validação do silenciamento gênico, pela técnica de RNAi;

3. Sequenciamento de transcritomas

3.1 Sequenciamentos de ácidos nucléicos

Apesar da descoberta do DNA ter sido feita em 1953, foram necessários quase 20 anos para elucidação da primeira descrição da metodologia para o sequenciamento de ácidos nucléicos que permitiu o genoma completo do vírus bacteriófago MS2, que data do início dos anos 1970. Nessa década, vários métodos de sequenciamento foram desenvolvidos para sequenciar DNA de microorganismos e vírus. Porém, a técnica desenvolvida por Sanger e colaboradores em 1977 tornou-se a mais utilizada devido as suas características de baixa toxicidade, facilidade e repetibilidade em comparação com outras metodologias. A tecnologia desenvolvida por Sanger era baseada numa reação onde eram adicionadas DNA polimerase, iniciadores (*primers*), deoxinucleotídeos (dNTPs), molde de DNA e dideoxinucleotídeos (ddNTPs). A incorporação de ddNTP à cadeia de DNA interrompe o processo de alongação feito pela DNA polimerase (FIERS *et al.* 1976.).

Nessa metodologia, numa reação de sequenciamento comum, a amostra de DNA era dividida em quatro reações separadas, cada uma contendo todos os quatro nucleotídeos (dATP, dGTP, dCTP, dTTP), os *primers*, um dos ddNTPs (ddATP, ddCTP, ddGTP, ddTTP) e a polimerase. Após várias etapas, os fragmentos de DNA resultantes eram aquecidos e separados por eletroforese em gel desnaturante de poli-acrilamida resultando em diversos fragmentos de DNA, de tamanhos variados, que poderiam ser visualizados por autorradiografia, ou mesmo sob luz UV. Assim, a sequência de DNA podia ser determinada, tanto diretamente do filme radiográfico, como da imagem do gel. (SAIKI *et al.*, 1985).

A partir de então, a automatização do sequenciamento foi conseguida por meio de muitas inovações tecnológicas. A utilização dos quatro nucleotídeos em uma mesma reação foi possível graças ao uso de fluoróforos ligados às moléculas de ddNTPs, cada um emitindo um comprimento de onda específico. A utilização de fluoróforos permitiu também o abandono do uso da eletroforese e da radioatividade. Assim também, foi a partir do desenvolvimento de equipamentos de análise que emitem um feixe de laser sobre o gel, causando a excitação dos nucleotídeos marcados, o processo de leitura da sequência nucleotídica foi automatizado. Cada nucleotídeo passou a emitir uma fluorescência, detectada eletronicamente e convertida de forma digital para um arquivo de sequências. (ZHANG *et al.*, 2011.). A capacidade de leitura atingiu uma média de 1000 a 1200 pares de bases (pb). O sequenciamento de Sanger passou a requerer pequenas quantidades de amostras para realizar o sequenciamento necessário.

Ramos da ciência como Genômica Comparativa, Evolução, Genética Forense, Epidemiologia e Medicina Aplicada tiveram muitos ganhos com a utilização das tecnologias de sequenciamento, sendo o mais notório o sequenciamento do genoma humano. Nesse Projeto, que data dos anos 1990 (COLLINS *et al.*, 2003), geneticistas americanos criariam mapas genéticos e físicos de cada cromossomo humano. Esses mapas teriam que conter marcadores genéticos que seriam utilizados como guia para a montagem de um grande quebra-cabeça, onde o sequenciamento só ocorreria depois da montagem desses mapas. A expectativa de conclusão desse projeto era de pelo menos 15 anos, porém, Craig Venter, responsável pelo Projeto Genoma Humano, declarou em 1998, que poderia obter o genoma em dois anos, a partir do uso de uma técnica denominada por ele de sequenciamento *shotgun*, onde o genoma seria fragmentado, clonado em vetores e sequenciado aleatoriamente. Posteriormente, as sequências seriam reunidas em computadores que fariam o alinhamento a partir de sobreposições

encontradas em cada uma. O resultado final do Projeto Genoma Humano foi publicado em 2003, quando completou 50 anos da descoberta da estrutura do DNA (VENTER *et al.*, 2001; COLLINS *et al.*, 2003). Porém, o sequenciamento de genoma de diferentes organismos que incluindo animais superiores, insetos, plantas, vírus e bactérias, entre outros, serviram para adiantar as pesquisas no que se refere a diversas questões biológicas, levando a um desenvolvimento de estudos no sentido de ampliar as técnicas com ganho de tempo e baixa nos custos, como o sequenciamento de segunda geração. (NGS: *Next generation sequencing*).

3.2 Sequenciamento de segunda geração

São muitas as tecnologias de NGS desenvolvidas que permitem o sequenciamento de DNA em larga escala como as plataformas GS-FLX 454 *Genome Sequencer* (Roche), o *Genome Analyser* (Illumina), e o ABI SOLiD *analyser*, cada uma com diferentes aplicações, dependendo do tamanho e cobertura dos *reads* obtidos. (FU *et al.*, 2017). O sequenciamento tipo NGS lê a amostra de DNA aleatoriamente, o que é conseguido a partir da fragmentação do genoma alvo, seguido da ligação dos fragmentos a adaptadores para a leitura durante a análise (MARTIN e WANG, 2011). As leituras revelam sequências curtas de nucleotídeos (*reads*), com tamanhos entre 50 e 500 pares de bases, comparado com sequenciamento *shotgun* que pode gerar sequências de até 1.200 pares de bases. Uma das maneiras de driblar essa limitação é aumentar a cobertura do sequenciamento. Essa cobertura é o número de *reads* que se sobrepõem dentro de uma região específica do genoma. Graças às ferramentas de bioinformática, os *reads* são sobrepostos e agrupados por determinado grau de similaridade, gerando sequências

maiores e aumentando o tamanho das sequências e, conseqüentemente, a cobertura do sequenciamento (ZHANG *et al.*, 2011). A Plataforma Illumina possui diferentes variantes com finalidades específicas, como, por exemplo, o sequenciamento de mRNA e RNA total, uma tecnologia conhecida como RNA-seq.

3.3 Sequenciamento de RNA

Transcritoma é o conjunto de transcritos de um organismo observado em determinada circunstância. A sua compreensão é importante na interpretação do genoma em uma abordagem funcional, numa comparação entre diferentes condições e etapas de desenvolvimento. A técnica de RNA-seq possibilita sequenciar com exatidão o transcritoma de quase todos os organismos, desde que sejam obtidas amostras de RNA total de alta qualidade, ou seja, sem degradação e livre de contaminantes. É objetivo da transcriptômica a identificação de genes alvos específicos e a quantificação diferencial da expressão gênica, de modo comparativo (WANG *et al.*, 2009; ROBINSON *et al.*, 2010). Além disso, essa técnica permite mapear e quantificar transcritos em relação a outras técnicas como hibridização do RNA, micro-arranjos e sequenciamento tradicional de Sanger (MORTAZAVI *et al.*, 2008).

Os *reads* obtidos no sequenciamento podem ser alinhados com um genoma ou transcritoma de referência, ou montados através da metodologia *De Novo*, gerando dessa forma um mapa de transcrição formado pela estrutura transcricional dos genes e o nível de expressão de cada um deles (Figura 4) (ROBERTSON *et al.*, 2010.). A montagem *De novo* se refere a ausência de uma informação prévia, por exemplo, um genoma de referência. Ela é realizada através da identificação de sobreposição entre os *reads*, que formam *contigs* pela identificação de uma ou várias sequências consenso.

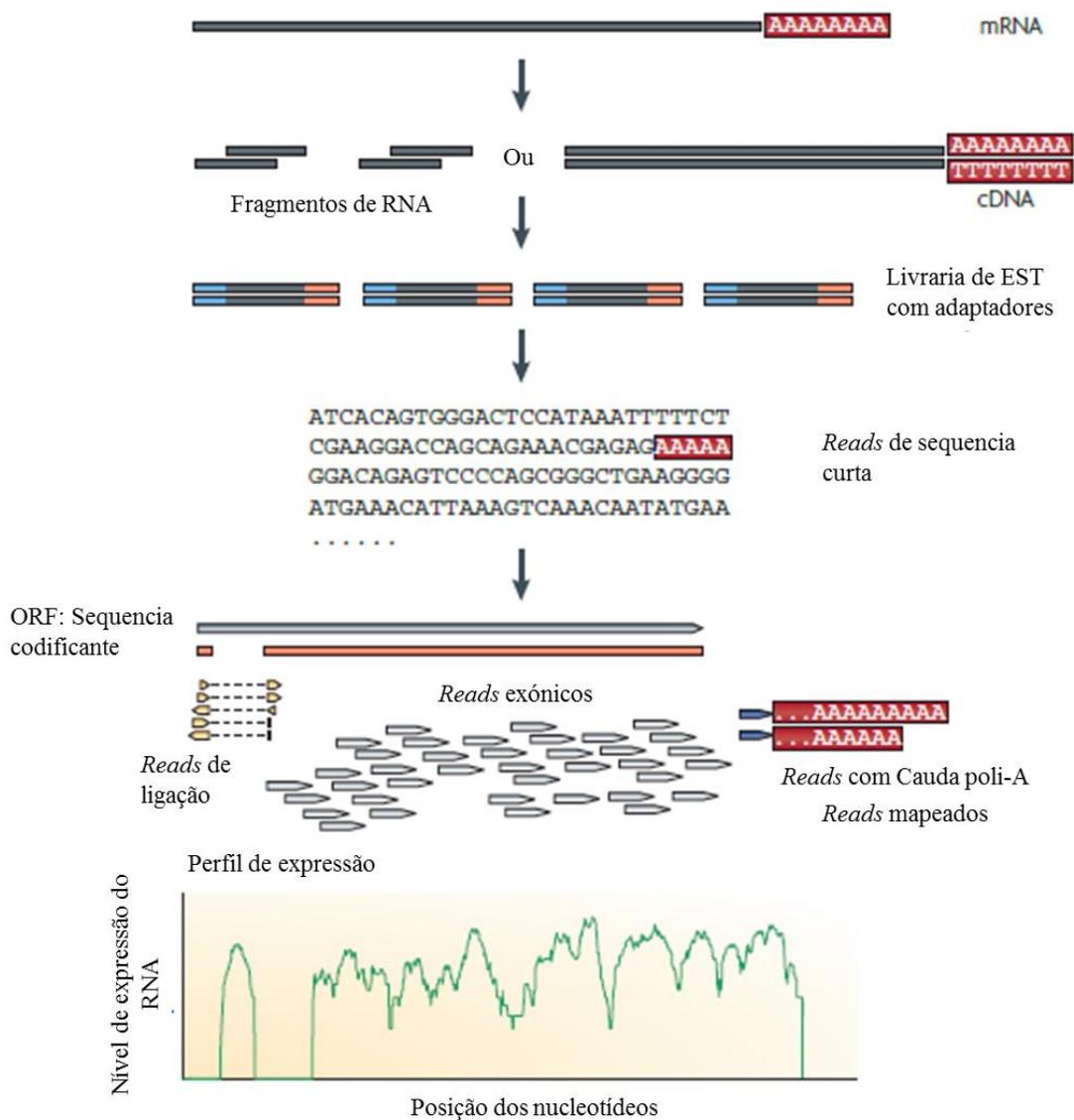


Figura 4. Esquema de um experimento padrão de RNA-seq de uma levedura é mostrado na ilustração. Longas moléculas de RNAs são convertidas em uma biblioteca de fragmentos de cDNA, por meio de fragmentação do RNA ou do DNA. Adaptadores de sequenciamento (Azul), são adicionados em cada fragmento de cDNA, obtendo seqüências curtas para cada cDNA. Os *reads* resultantes são alinhados com um genoma ou transcrito de referência, e classificados em três tipos: *reads* exônicos, *reads* de ligação e *reads* com cauda poli-A. Esses três tipos de *reads* são usados para gerar um perfil de expressão para cada gene, como representado na parte inferior; uma fase de leitura aberta (ORF: *Open Reading frame*). Wang *et al.* (2009).

Entre os muitos programas de bioinformática desenvolvidos para aplicação no processo de montagem de um transcrito, que usam algoritmos para finalidades específicas, um dos mais utilizados para montagens *De Novo*, é o pacote Trinity (GRABHERR *et al.*, 2011). Ele se divide em três componentes: 1. Inchworm, que monta os *reads* sequenciados a fim de gerar uma coleção de sequências consenso lineares (*contigs*), assinando a cada *contig* um *k-mer* (variantes associadas a um determinado padrão); 2. Chrysalis, cuja função é juntar os *contigs* que compartilhem pelo menos um *k-mer*, e gerar grafos os de De Bruijn para cada conjunto; 3. Butterfly, que compacta os grafos gerados anteriormente com os *reads* originais, para produzir uma sequência linear final para cada parágrafo ou isoforma.

A técnica de RNA-seq também tornou possível os estudos de transcrito em um grande número de organismos; além de ter um custo menor em relação ao microarranjo, permitindo estudar transcritos sem precisar do uso de genomas de referência, possibilita identificar isoformas e entrega informação sobre a conexão entre exons. (WANG *et al.*, 2009)

3.4 Transcritoma de insetos

A importância dos insetos está no fato de que interferem diretamente na vida de todos os seres vivos. Eles representam mais de um milhão de espécies descritas, constituindo um dos grupos mais diversos, podendo chegar a 75% dos metazoários do planeta (GRIMMELIKHUIJZEN *et al.*, 2007; GULLAN e CRANSTON, 2014). São ecologicamente importantes em todos os níveis da cadeia trófica, são altamente adaptáveis a qualquer ambiente, além de fazerem parte dos grupos de maior diversidade de organismos vivos (GULLAN e CRANSTON, 2014). Podem causar grandes prejuízos,

como no setor agrícola, por exemplo, atuando como pragas e vetores de doenças, causando prejuízos econômicos catastróficos, ou podem ajudar na polinização e proteção das lavouras, como os predadores de ovos e de larvas, além de fornecerem alimentos e produtos como o mel e a seda (SCUDDER *et al.*, 2009). Levando-se em conta a relação existente desde sempre entre o inseto e o homem, a pesquisa do genoma de várias espécies de insetos é de grande importância uma vez que possibilita estudos de genômica funcional e evolutivo, e análises comparativas associadas aos parâmetros de comportamento, na transmissão de doenças, desenvolvimento e reprodução (ZDOBNOV e BORK, 2007.).

Muitos foram os investimentos já realizados para a obtenção de genomas de diferentes insetos desde o primeiro genoma de *Drosophila melanogaster*, publicado há mais de 10 anos pelos mesmos cientistas do Projeto Genoma Humano, utilizando sequenciadores com capilares baseados na tecnologia de Sanger (ADAMS *et al.*, 2000). Com o desenvolvimento de tecnologias NGS, mais de 80 genomas de insetos foram sequenciados completamente e disponibilizados em bancos de dados. A maioria dos genomas sequenciados até hoje pode ser observada em bases de dados públicas como a InsectBase, que contém o genoma de 138 espécies de insetos e o transcrito de 116 espécies (YIN *et al.*, 2016).

De grande importância para o ser humano destacam-se os genomas de *Tribolium castaneum* (RICHARDS *et al.*, 2008), *Acyrtosiphon pisum* (RICHARDS *et al.*, 2010), *Apis mellifera* (WEINSTOCK *et al.*, 2006), *Bombyx mori* (Consortium, 2008), *Anopheles gambiae* (HOLT *et al.*, 2002) e *Aedes aegypti* (NENE *et al.*, 2007).

Graças aos estudos de transcritômica em insetos, já foi possível descrever as interações envolvidas entre diferentes sistemas de praga-hospedeiro, e também descrever diversos mecanismos moleculares de ação de insetos praga, auxiliando nas estratégias

para o controle das populações (ZHANG e YUAN, 2013). Alguns desses mecanismos estão relacionados com alimentação e desenvolvimento do inseto, assim como os seus mecanismos de proteção na planta hospedeira; o reconhecimento de substâncias moduladoras do comportamento, vias de digestão e detoxificação. (FIRMINO *et al.*, 2013; LI *et al.*, 2013; YIN *et al.*, 2014; DE ASSIS FONSECA *et al.*, 2015; PERERA *et al.*, 2015; CRAVA *et al.*, 2016; WANG, Y. *et al.*, 2016).

3.5 Análise de expressão diferencial de genes

A falta de uniformidade entre as amostras é uma das limitações mais comuns para a análise de dados de transcriptomas. O método mais usado é a análise de genes diferencialmente expressos (DEGs: *Differential expressed genes*), que mostra diferenças do nível de expressão dos transcritos, associadas a diferentes condições experimentais (SONESON E DELORENZI, 2013). A distribuição não uniforme de dados ocorre por razão de diferentes coberturas dos *reads* em relação à referência utilizada (genoma ou transcriptoma), e ao tamanho das bibliotecas de cDNA. Acrescenta-se o fato de que, em transcriptomas montados *De Novo*, a contagem de *reads* pode não ser diretamente comparável entre as amostras. A presença de poucos genes com contagens muito altas (alta expressão) impede uma resolução clara de genes com baixa contagem, causando uma interpretação errada de genes com aparente baixa expressão, identificados depois falsamente como DEGs (BULLARD *et al.*, 2010).

3.6 Estudo da expressão gênica em insetos

O estudo de transcritomas de insetos para a identificação de genes diferencialmente expressos tem sido realizado em diversas situações e condições experimentais. Atualmente, as condições mais estudadas pelos entomologistas são aquelas relacionadas com diferentes fases de desenvolvimentos (QI *et al.*, 2016; MA, L. *et al.*, 2016; SPARKS *et al.*, 2017), comportamento sexual (YANG *et al.*, 2015; WU *et al.*, 2016), adaptação a estresse abiótico (LIU *et al.*, 2017; TANG *et al.*, 2017), resposta imune (VOGEL *et al.*, 2011; WEI *et al.*, 2016), e resposta a diferentes dietas ou fontes de alimentação (LI *et al.*, 2016; ROY, A. W. *et al.*, 2016). Estudos relacionados com estágios de desenvolvimento têm sido de grande interesse nos últimos anos para análises transcricionais de diversos insetos praga. Dados relacionando genes e rotas metabólicas podem ser amplamente explorados para aplicação em estratégias biotecnológicas para o controle dos insetos (GOG *et al.*, 2014; HERDE e HOWE, 2014; SPIT *et al.*, 2014; ROY, A. *et al.*, 2016; SOUZA *et al.*, 2016; ZHANG *et al.*, 2016).

Os mecanismos moleculares por meio dos quais os insetos conseguem superar as barreiras químicas da planta, são pouco conhecidos. Porém, o sequenciamento massivo de transcritomas de insetos realizados nos últimos anos, tem permitido identificar um grande número de genes relacionados com detoxificação, reconhecimento sensorial de sequências voláteis, e resistência contra metabólitos secundários (VOGEL *et al.*, 2014). Genes que codificam proteínas de diferentes famílias, envolvidas em processo de detoxificação, foram caracterizados em várias espécies de insetos. Entre as famílias protéicas melhor descritas estão: citocromo P450 (P450), glutatona S-transferases (GSTs), Glucuronosiltransferases (UGTs), carboxilesterases (CEs) e os transportadores do tipo ABC (HECKEL, 2014).

Um grande número de genes que codificam estas proteínas apresenta uma expressão significativamente maior no intestino, em relação a outros tecidos do inseto. Isto se deve ao fato de que o sistema digestório é a primeira linha de defesa do inseto, além de ser o principal tecido onde ocorre o processo de absorção de moléculas vindas do meio externo (TELLAM, 1996). A obtenção do perfil de expressão de genes no intestino de diferentes insetos praga, submetidos a diferentes hospedeiros ou dietas, é de grande importância para entender o papel das diferentes proteínas envolvidas na defesa do inseto, frente às barreiras químicas da planta.

4. Material e métodos

4.1 Extração de RNA total e sequenciamento

Devido a parcerias existentes entre os pesquisadores brasileiros (Embrapa) e colombianos (Universidad de Caldas), amostras de larvas de 2º instar, insetos adultos fêmeas e machos de *H. hampei*, alimentados com grão seco de café, foram coletados de colônias bem estabelecidas do *Centro Nacional de Investigaciones de Café* – CENICAFÉ em Chinchiná, Caldas, Colômbia durante o ano de 2015. Os insetos foram coletados em tubos RNase-free com 1 mL de solução estabilizadora RNAlater (*Quiagen, Germany*) e armazenados a -80°C para extração posterior. No total, 20 mg de cada estágio de desenvolvimento foram utilizados para extração de RNA.

O RNA total de cada estágio analisado foi extraído separadamente de cada uma das fases de desenvolvimento com RNeasy Mini Kit (Qiagen, Alemanha) de acordo com as instruções do fabricante. A extração foi realizada com três réplicas biológicas. A integridade do RNA foi avaliada usando o sistema Bioanalyzer 2100 (*Agilent Technologies, EUA*). A concentração de RNA foi medida usando o NanoDrop 1000 (*Thermo-Fisher Scientific, Wilmington, DE*). As amostras específicas dos estágios de desenvolvimento que tiveram o RNA extraído foram reunidas em quantidades equimolares. As bibliotecas de cDNA foram construídas utilizando o Kit TruSeq RNA Library Prep v2 (Illumina Inc).

Finalmente as bibliotecas de cDNA foram purificadas com o kit de remoção de rRNA Ribo-zero e sequenciadas em uma corrida 125 PE da plataforma Illumina HiSeq™ 2500 no Centro de Genômica da Universidade de Minnesota (Minnesota, EUA). O RNA foi sequenciado em ambos os sentidos (*paired-end*).

4.2 Montagem De Novo e Anotação Funcional

Antes da montagem do transcrito, foi feita uma análise da qualidade dos *reads* com a ferramenta FastQC versão 0.11.5 para diagnosticar problemas relacionados com a construção das bibliotecas ou no sequenciamento (ANDREWS *et al.*, 2012). A verificação da quantidade total de sequências foi feita além da qualidade das bases, comprimento das sequências, porcentagem de GC, sequências duplicadas ou sobreexpressas. Após essa primeira análise, os *reads* totais foram aparados com o Programa Trimmomatic versão 0.3358 com o objetivo de eliminar as sequências de baixa qualidade (Phred <28), as sequências contendo nucleotídeos desconhecidos (N acima de 5%) e retirar os adaptadores do sistema Illumina (BOLGER *et al.*, 2014).

Para a montagem, os *reads* foram normalizados digitalmente usando o software KHMER v2.059 com uma cobertura máxima de 50 pb para cada *read*. O objetivo dessa etapa foi reduzir os custos computacionais, aumentando a cobertura dos *reads* e diminuindo o tamanho do conjunto total de dados. Os *reads* normalizados foram reunidos em sequências consenso (*contigs*), usando o *software* Trinity v2.0.660, utilizando os parâmetros padrão do programa (HAAS *et al.*, 2013), onde cada conjunto desses *contigs* foram designados como unigenes, em função da similaridade dos fragmentos de cDNA (EST: *expressed sequence tag*). Cada unigene contém um número variável de transcritos (isoformas).

Após a montagem, foi feita uma busca por homologia (BLASTx e BLASTn. *E-value* mínimo: 1×10^{-5}) usando o banco de dados não redundante (nr) do NCBI, para identificar ORFs preditas nas sequências montadas. Com o objetivo de se obter uma análise de expressão diferencial mais detalhada, as sequências que não apresentaram

nenhuma sequência homóloga referida no banco de dados nr e as que unicamente apresentaram homologia com microrganismos ou plantas, foram retiradas da montagem.

A anotação funcional dos transcritos foi feita com o software Blast2GO Basic v4.161 (www.blast2go.org), por meio de classificação por termos de ontologia gênica (GO; www.geneontology.org) e códigos de classificação enzimática (E.C). Para isso foi utilizadas, a base de dados de proteínas do InterPro (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>), e a base de dados de vias metabólicas KEGG (*Kyoto Encyclopedia of Genes and Genomes*; <http://www.genome.jp/kegg/>). Além disso, foi utilizado o programa BUSCO v3 com o objetivo de avaliar a completude dos transcritomas sequenciados (SIMÃO *et al.*, 2015).

4.3 Análise de expressão diferencial (DGEs)

A análise para identificar genes diferencialmente expressos nas diferentes fases de desenvolvimento foi realizada seguindo o pipeline do Trinity versão 2.0.6 (<https://github.com/trinityrnseq/trinityrnseq/wiki/Trinity-Differential-Expression>). O mapeamento foi feito através do pacote RSEM (LI e DEWEY, 2011) e a estimativa de abundância de genes diferencialmente expressos foi realizada seguindo o roteiro *align_and_estimate_abundance.pl* do pacote Trinity. A expressão foi analisada usando todas as isoformas dos transcritos, e os parâmetros usados para executar o edgeR foram: Valor $P < 0,01$ e *Fold change* ≥ 2 . As comparações dos genes diferencialmente expressos (DEGs) foram realizadas usando as bibliotecas de fêmea como referência.

4.4 Enriquecimento de termos GO

A análise de enriquecimento GO, feita para identificar quais termos de ontologia gênica que estavam estatisticamente enriquecidos no conjunto de genes diferencialmente expressos, foi realizada com o programa FUNC seguindo os parâmetros padrões (PRÜFER *et al.*, 2007). A remoção de termos GO redundantes foi feita com o *software* REVIGO (SUPEK *et al.*, 2011), com um valor de similaridade permitido de 0,5. A visualização dos perfis de expressão foi realizada com o *software* ClustVis. Os *Heatmaps* que mostram a expressão dos DEG's foram gerados usando o programa ClustVis (METSALU E VILO, 2015). A escala de expressão, foi determinada pelos valores de FPKM normalizados pelo edgeR.

4.5 Análise filogenética

Múltiplos alinhamentos foram feitos com as sequências de aminoácidos de genes relacionados a funções quimiossensoriais, pertencentes à família de *odorant binding protein* (OBP). Para isso, foram utilizadas sequências de dois genomas bem caracterizados de insetos praga pertencentes à ordem dos coleópteros (*Tribolium castaneum* e *Dendroctonus ponderosae*), e aquelas identificadas nesta montagem. Árvores filogenéticas foram construídas no *software* RAxML v8 (STAMATAKIS *et al.*, 2014) usando a máxima verossimilhança do método com o modelo LG e 1000 réplicas de *bootstrap*. Os genes candidatos codificadores de OBP da montagem do Transcritoma de *H. hampei*, juntamente com os seus resultados de BLAST e valores de FPKM, são ilustrados na figura 13.

4.6 Validação por PCR em tempo real (RT-PCR)

O cDNA do inseto (larva, macho e fêmea) foi sintetizado a partir de 1000 ng de RNA utilizando um kit de transcrição reversa QuantiTect (Qiagen, Valencia, CA) seguindo as instruções do fabricante. As reações de RT-qPCR incluíram 2 µL de cDNA diluído (1:50), 5 µL de *Fast SYBR® Green Master Mix* (Applied Biosystems, Foster City, CA), 0.2 µL de cada *primer* em 10 µM, e 2.6 µL de água *nuclease-free*, totalizando um volume total de 10 µL. Os *primers* foram desenhados usando o Programa Primer3 (<http://primer3.ut.ee/>) e foram validados por análise de suas eficiências de amplificação por PCR (E) e coeficientes de correlação (R²). Ambos os testes de eficiência de *primers* e RT-qPCR foram realizados em um sistema 7500 Fast RT-PCR (Applied Biosystems, Grand Island, NY). O programa RT-qPCR incluiu um estágio de espera de 95 ° C por 20 segundos, seguido por 40 ciclos de desnaturação a 95 ° C por 3 segundos e anelamento / extensão a 60 ° C por 30 segundos (estágio cíclico). Uma curva de *melting* foi gerada para confirmar a produção de um único pico e para descartar a possibilidade de dímeros de *primers* e formação inespecífica de produtos. A expressão dos genes foi calculada usando o método comparativo $2^{-\Delta\Delta CT}$ method (LIVAK *et al.*, 2001).

5. Resultados

5.1 Extração de RNA total

As amostras de RNA total extraído de *H.hampei*, das diferentes fases de desenvolvimento, apresentaram a integridade e a qualidade requeridas para a plataforma de sequenciamento Illumina. Isso foi confirmado por análise por meio de Bioanalyzer 2100 (*Agilent Technologies*, EUA) e eletroforese em gel de agarose (Figura 5). O rastro de RNA mostrado no gel com as bandas bem definidas representa o perfil de RNA total de insetos, descrito por Winneback e colaboradores em 2010. Esse perfil corresponde ao padrão de RNA que possui mais de duas bandas definidas, o que geralmente reflete os componentes endógenos presentes de RNA ribossômico e não à degradação de ácidos nucleicos. No geral, a ausência de bandas com tamanhos inespecíficos indica que não houve degradação suficiente do RNA total, para reduzir a qualidade do mínimo requerido para sequenciamento.

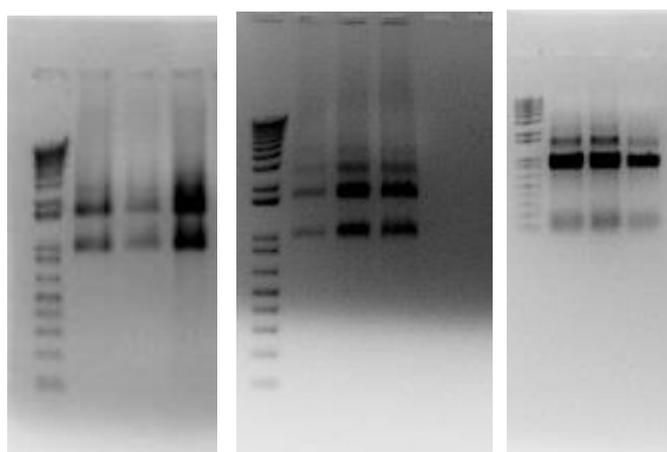


Figura 5. Géis de agarose com amostras de RNA total das fases de desenvolvimento da broca do café. 1º gel – Marcador 1Kb e amostras de larva L2 2º gel – Marcador 1Kb e amostras de adulto (fêmea) e 3º gel – Marcador 1Kb e amostras de adulto (macho).

5.2 Análise dos transcritomas

5.2.1 Sequenciamento e Montagem *De Novo*

O sequenciamento por Illumina das nove bibliotecas de *H. hampei* (fêmeas, machos e larva L2, sendo três réplicas biológicas por estágio de desenvolvimento) resultou em um total de 265,3 milhões de *reads* pareados (PE) contendo ~ 66,8 Gbp. O tamanho das bibliotecas variou de 16,3 a 49,2 milhões de *reads* PE. Os *reads* que apresentaram má qualidade de sequenciamento ou insuficiência de cobertura (~ 24% dos *reads* sequenciados) foram retirados das bibliotecas, e um total de 201.8 PE de alta qualidade foi usado para análise de expressão digital (Tabela 1). Além disso, apenas 95,1 milhões de PE de alta qualidade (~ 47%) foram usados para a montagem do transcritoma após a normalização digital com o *software* KHMER (CRUSOE *et al.*, 2015). Esse número de *reads* após o pré-processamento é o primeiro indicador de qualidade do RNA e das bibliotecas de cDNA. Uma vez que a porcentagem de leituras mapeadas no genoma de referência disponível era muito pequena (<40%) para uma análise de expressão robusta (VEGA *et al.*, 2015), o método *De novo* foi utilizado para a montagem do transcritoma. Os *reads*, após o pré-processamento, foram montados em 116.587 sequências contíguas (*contigs*) com um comprimento médio de 1.258,25 pb (N50 de 2.079 pb). Destes *contigs*, 80,046 unigenes (isoforma mais longa) foram identificados com um comprimento médio de 970,67 pb (N50 de 1,388 pb). Utilizando o alinhamento por BLAST contra o banco de dados não redundante (NR), foram identificados 56.738 (48,7%) *contigs* com supostos homólogos. Após a filtragem de sequências não codificantes e 2.760 sequências contaminantes (2.4%), obteve-se um conjunto final de ~ 106.6 Mbp contendo 53.978

contigs com um comprimento médio de 1.975.49 pb (N50 de 2.795 pb) e 29.434 unigenes com um comprimento médio de 1,609.92 pb (N50 de 2795 pb). A qualidade da montagem foi avaliada pelo programa BUSCO e apresentou 98,9% de completude. Dados gerais sobre as estatísticas da montagem do transcrito são apresentados na Tabela 2.

Estágio	Dados brutos <i>Reads</i> pareados	Pós-processamento <i>Reads</i> pareados
Fêmea	60,223,701	45,694,955
Macho	114,711,712	87,655,146
Larva (L2)	90,355,722	68,457,556

Tabela 1. Número total de reads pareados nas bibliotecas de cada fase de desenvolvimento de *H.hampei* obtido pelo sequenciamento por Illumina Hi-seq2500.

Dados estatísticos (Montagem do Transcritoma)	Montagem pré-filtrada * (dados brutos)	Montagem pós-filtrada (sequências limpas)		
Número de Unigenes	80,046	29,434		
Número de Contigs	116,587	53,978		
GC %	34.99	35.77		
Dados completos do BUSCOs		1,054 (98.9%)		
Cópias únicas BUSCOs		536 (50.3%)		
Cópias duplicadas BUSCOs		518 (48.6%)		
Contigs	Contigs	Unigenes	Contigs	Unigenes
<i>Median</i> (bp)	705	573	1,549	1,114
<i>Mean</i> (bp)	1,258.25	970.67	1,975.49	1,609.92
Total bases montadas:	146,695,858	77,697,918	106,632,933	47,386,471
Contig N10 (bp)	5,626	4,739	6250	5735
Contig N30 (bp)	3,288	2,546	3948	3551
Contig N50 (bp)	2,079	1,388	2795	2427

Tabela 2. Resumo estatístico do transcrito de *H.hampei*

* A montagem pré-filtrada inclui sequências não codificantes e contaminantes.

5.2.2 Anotação Funcional

Com o objetivo de analisar as sequências codificantes entre os *contigs* montados e realizar o estudo funcional dos genes relacionados, os transcritos da etapa final da montagem foram anotados utilizando as bases de dados NR, InterPro, GO, Pfam e KEGG. Como esperado, para os 53.978 *contigs* anotados na base de dados NR, mais de 75% foram similares às sequências de outras espécies da ordem coleóptera. O inseto *Dendroctonus ponderosae*, pertencente à mesma família da broca do café, *Scolytinae*, foi a espécie com maior número de sequências similares, com 29.502 (54,4%) *contigs* anotados, seguida por *Tribolium castaneum* com 11.245 (20,7%) *contigs* anotados. Surpreendentemente, *H. hampei* não estava entre as dez primeiras espécies com mais *hits*, com apenas 11 (0,02%) *contigs* anotados (Figura 6).

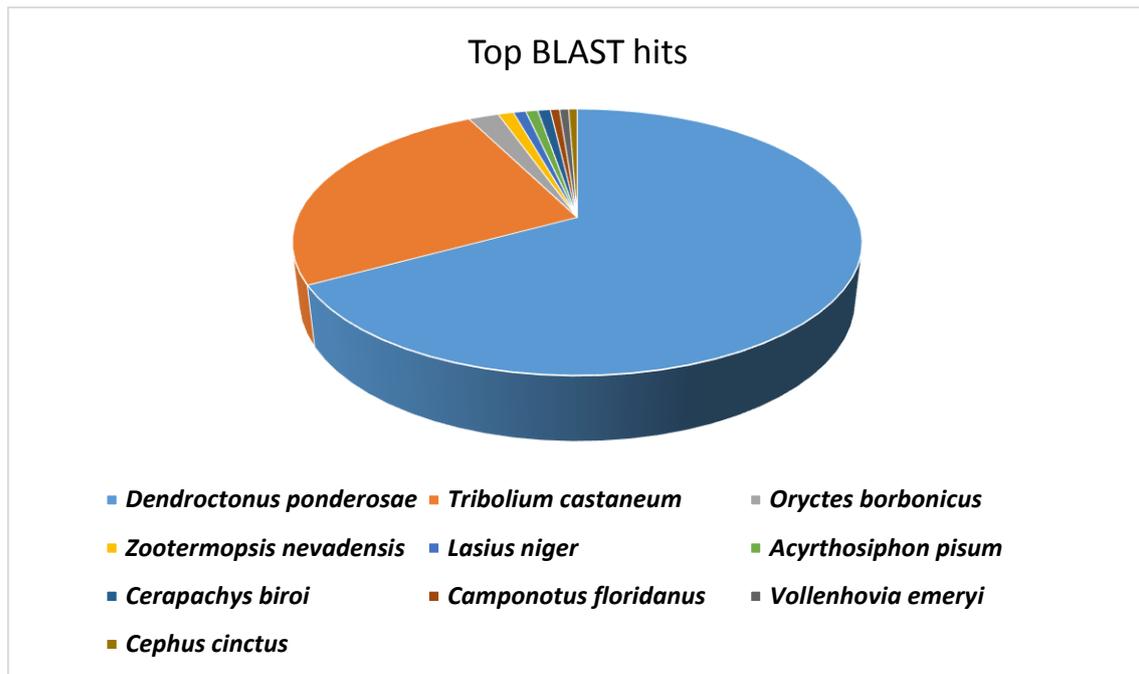


Figura 6. Espécies com principais hits após BLASTx. O maior número de contigs teve alta similaridade com o inseto coleóptero, *Dendroctonus ponderosae*.

Além disso, após a filtragem de sequências contaminantes, foi realizada a anotação funcional por ontologia gênica (GO). Para essa anotação, um total de 45.759 termos (1.176 GOs individuais e 1.888 GOs únicos) foram distribuídos em 98 categorias funcionais segundo a classificação GO, todas pertencentes às três principais categorias: Processos Biológicos (50 subcategorias), Funções Moleculares (21 subcategorias) e Componentes Celulares (27 subcategorias), que compreendiam 17.036, 20.101 e 14.739 *contigs*, respectivamente (figura 7A). Adicionalmente, um total de 19.027 (35,2%) *contigs* dentro de 8.611 (29,3%) unigenes foram anotados com pelo menos um código Pfam. Os três domínios da família de proteínas com a maioria das sequências associadas foram Proteína quinase (PF00069), com 1.683 *contigs* (8,8%), Citocromo P450 (PF00067), com 1.256 *contigs* (6,6%) e Receptores 7-transmembrana tipo 1 (PF00001), com 1.029 *contigs* (5,4%) (figura 7B). Finalmente, 2.685 *contigs* (1.499 unigenes) foram

associados com pelo menos um código de enzima após a anotação do banco de dados KEGG. No total, 344 enzimas foram identificadas e classificadas em 121 vias metabólicas (Figura 7C).

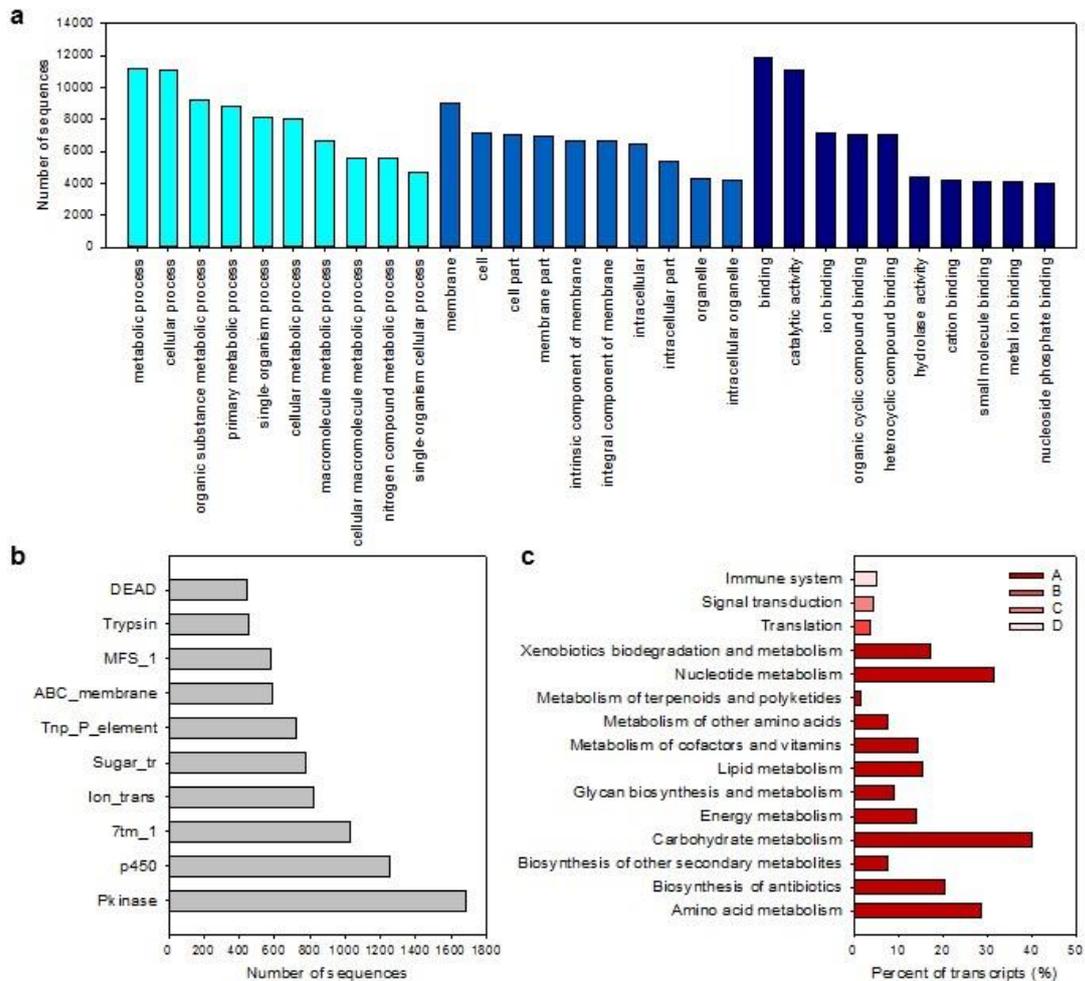


Figura 7. Resumo da classificação dos transcritos de acordo com as bases de dados *Gene Ontology* (GO), Famílias de Proteínas (Pfam) e *Kyoto Encyclopedia of Genes e Genomes* (KEGG). (a) A classificação GO resume uma distribuição multinível dos dez principais termos anotados para cada uma das três principais categorias GO (componente celular, função molecular e processo biológico). (b) Os dez domínios mais abundantes da classificação Pfam. (c) Classificação enzimática mostrando as principais vias (eixo y) de quatro categorias funcionais (metabolismo, processamento de informações genéticas, processamento de informações ambientais e sistemas orgânicos) e a porcentagem de transcritos em referência ao número total de transcritos anotados usando o banco de dados KEGG (eixo x). Os números à direita das barras representam o número total de transcritos anotados para cada caminho. A) Metabolismo, B) processos de informação genética, C) processos de informação ambiental, e D) sistemas organizacionais.

5.2.3 Expressão diferencial de genes (DEGs)

Para determinar os valores de expressão dos transcritos para cada estágio de desenvolvimento pelo método FPKM, uma média de 73,2% de *clean reads* (sequências limpas, livres de contaminantes), variando de 68,9% a 75,3% entre as bibliotecas, foram mapeados com sucesso na montagem final usando o Programa RSEM. Todas as três bibliotecas replicadas para cada estágio de desenvolvimento mostraram uma alta correlação (Pearson $r = 0,86-0,99$) e foram usadas para análise de expressão. Um total de 4.664 transcritos (3484 unigenes) foram diferencialmente expressos nos três estágios de desenvolvimento avaliados de *H. hampei* (Figura 8).

Usando a fêmea como referência, 676 transcritos foram regulados positivamente e 625 transcritos regulados negativamente na larva L2. Por outro lado, 3.069 e 294 transcritos foram regulados positivamente e negativamente, respectivamente, em insetos machos. Entre todos os DEGs, um total de 316 transcritos compreendendo 162 transcritos regulados positivamente e 154 regulados negativamente foram diferencialmente expressos tanto em macho como em larva L2 em comparação com a expressão observada em fêmea (Figura 8A). Quatro grupos principais de DEGs (larvas L2 reguladas positivamente, larvas L2 reguladas negativamente, machos regulados positivamente e machos regulados negativamente) foram gerados usando o método de agrupamento *EdgeR* (Fig. 8B).

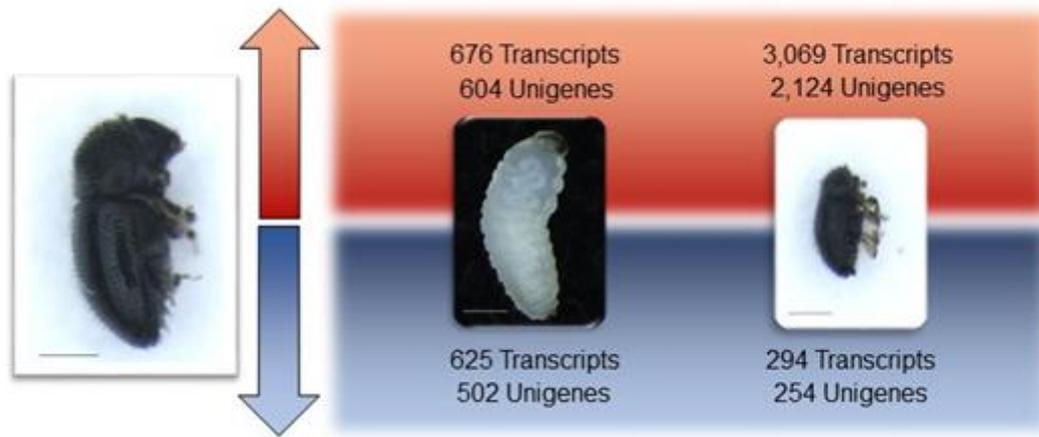


Figura 8. Comparações de genes diferencialmente expressos entre os diferentes estágios de desenvolvimento de *H. hampei*. (a) Número de transcritos e unigenes diferencialmente expressos em larva L2 e macho em referência a fêmea. As setas vermelhas e azuis representam genes regulados positivamente e negativamente, respectivamente. A barra de escala mostrada em fotos de insetos corresponde a 500 μ m.

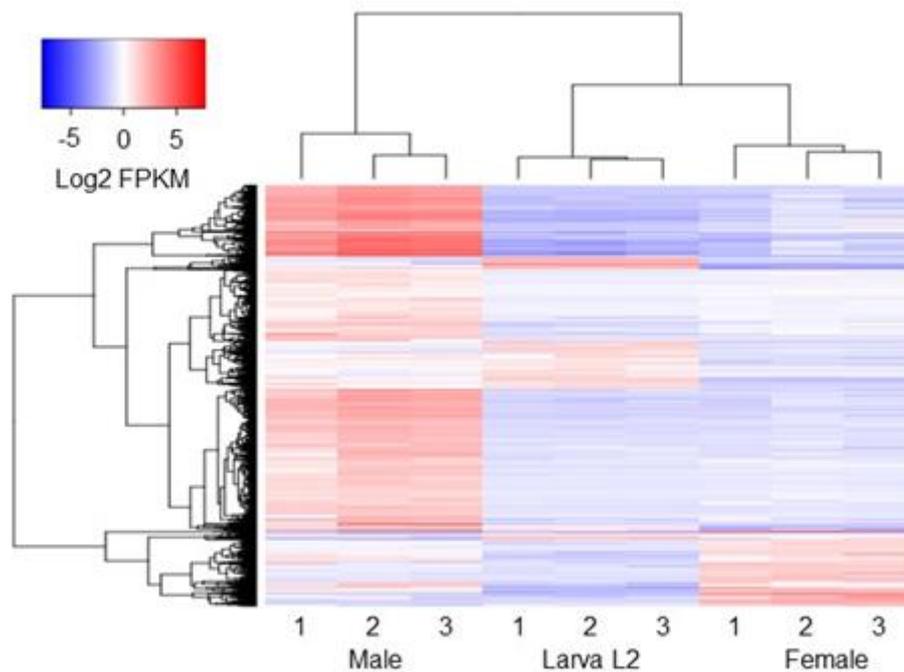


Figura 8. Comparações entre pares de genes diferencialmente expressos entre os diferentes estágios de desenvolvimento de *H. hampei*. (b) *Heatmap* e agrupamento de todos os DEGs. As linhas representam transcrições únicas e as colunas representam as réplicas de cada estágio de desenvolvimento. Os genes foram agrupados com base na similaridade de expressão (distância euclidiana). As cores azul e vermelha no mapa de calor correspondem à expressão gênica relativa baixa e alta, respectivamente.

É importante salientar que a maioria dos DEGs foram expressos com valores de *fold change* > 2 e $\leq 5,5$ (36,6%); em contraste, os DEG com valores $> 12,5$ e $\leq 15,6$ foram menos abundantes (1,9%). Surpreendentemente, 803 DEGs (17,2%) foram anotados de maneira incompleta, correspondendo apenas a genes associados a funções desconhecidas, proteínas hipotéticas e proteínas não caracterizadas, indicando a necessidade de melhor anotação e maior aprofundamento de estudos genéticos com a broca do café.

5.2.4 Análise de enriquecimento de ontologia gênica

A análise de enriquecimento de ontologia gênica (GO) foi realizada separadamente para todos os quatro conjuntos de DEGs. Depois de filtrar os termos GO redundantes, um total de 117 termos GO foram associados a todos os conjuntos de DEGs, dos quais 32, 24, 61 e 0 foram enriquecidos em larva L2 regulada positivamente, L2 regulada negativamente, macho regulado positivamente e macho regulado negativamente, respectivamente.

Além disso, foram identificados 8 termos GO que são compartilhados tanto em macho quanto em larva que são regulados positivamente. Os termos GO mais significativos para cada conjunto de DEGs nos diferentes estágios de desenvolvimento são mostrados na Figura 9. Curiosamente, os genes relacionados ao metabolismo de substâncias, metabolismo de quitina, proteólise e regeneração de cutículas foram regulados positivamente em ambos, larva L2 e macho.

Os genes relacionados aos processos de desenvolvimento, metabolismo de carboidratos, respostas de defesa e vias de sinalização celular foram regulados positivamente apenas em larva L2. Além disso, os genes relacionados aos processos de

transporte, atividade neuronal e comportamento foram regulados negativamente apenas na larva L2.

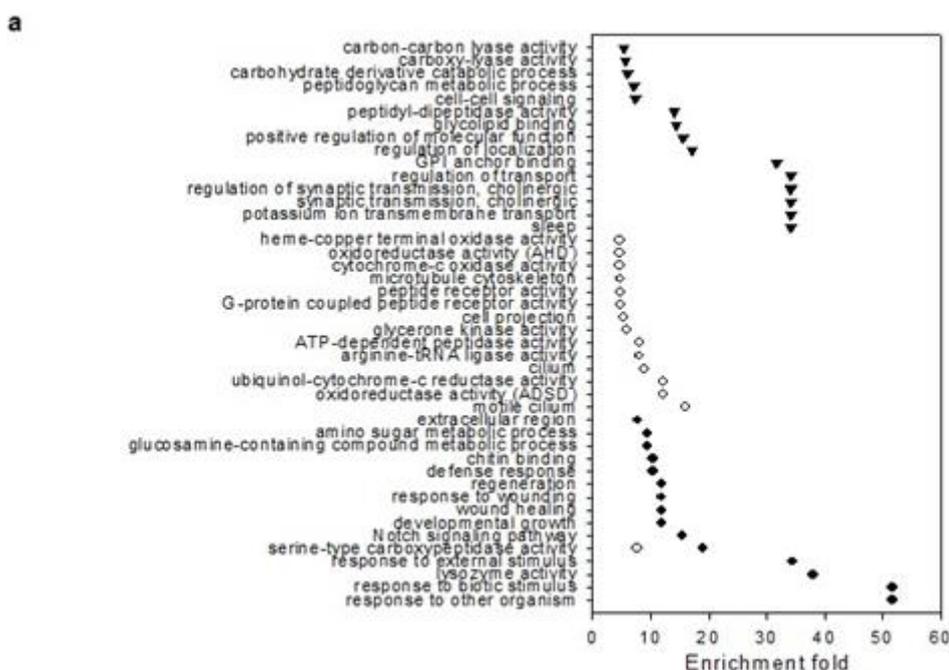


Figura 9. Ontologia gênica e análise de enriquecimento enzimático para cada conjunto de DEGs. Os 15 principais termos GO com os maiores valores de enriquecimento para cada comparação par a par.

5.2.5 Enriquecimento enzimático

Para identificar as vias metabólicas altamente específicas associadas a cada estágio de desenvolvimento do inseto, foi realizada uma análise de enriquecimento enzimático. No geral, 15 vias foram associadas com 12 enzimas que mostraram enriquecimento significativo ($FDR > 0,05$) no conjunto de genes que foram expressos diferencialmente. Surpreendentemente, a maioria das vias (14 no total) foi associada com o conjunto de macho regulado positivamente, enquanto a via metabólica restante foi associada com o grupo de larva L2 regulado positivamente (Figura 10). Por outro lado,

nenhum enriquecimento enzimático foi observado para o conjunto de macho regulado negativamente.

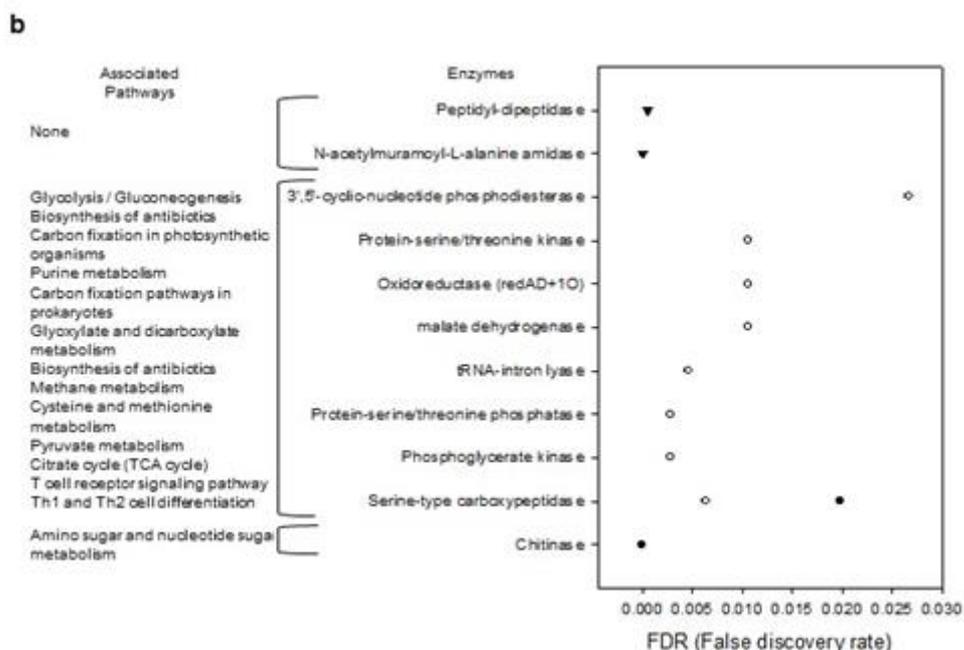


Figura 10. Ontologia gênica e análise de enriquecimento enzimático para cada conjunto de DEGs. Taxas de descoberta para cada enzima e as vias associadas para cada comparação na análise de enriquecimento. Círculos pretos indicam transcritos regulados em larvas L2. Círculos brancos indicam transcritos regulados positivamente em machos. Triângulos pretos indicam transcritos regulados negativamente em larvas L2.

5.2.6 Análise de genes diferencialmente expressos entre machos e fêmeas

Do total de unigenes que foram diferencialmente expressos, 2,124 (68,3% do total de DEGs) foram regulados positivamente em machos, dos quais 1,979 (56,8% do total de DEGs) foram regulados exclusivamente em machos e 145 também foram positivamente regulados em larva L2. Para uma análise mais detalhada desse grupo predominante de

DEGs, os unigenes associados aos termos/vias GO enriquecidos foram agrupados por funções gênicas (Figura 11).

Notavelmente, as proteínas relacionadas a elementos transponíveis (TEs) e as proteínas relacionadas à reprodução do inseto foram principalmente regulados positivamente nos machos. Entre as proteínas relacionadas a TE, os genes que codificam para 41 transposases (TSs) e 20 proteínas derivadas de TE foram identificados como DEGs. Entre os TEs e TSs, os pertencentes à superfamília *Tc1/mariner* foram os mais abundantes, com 19 e 26, respectivamente.

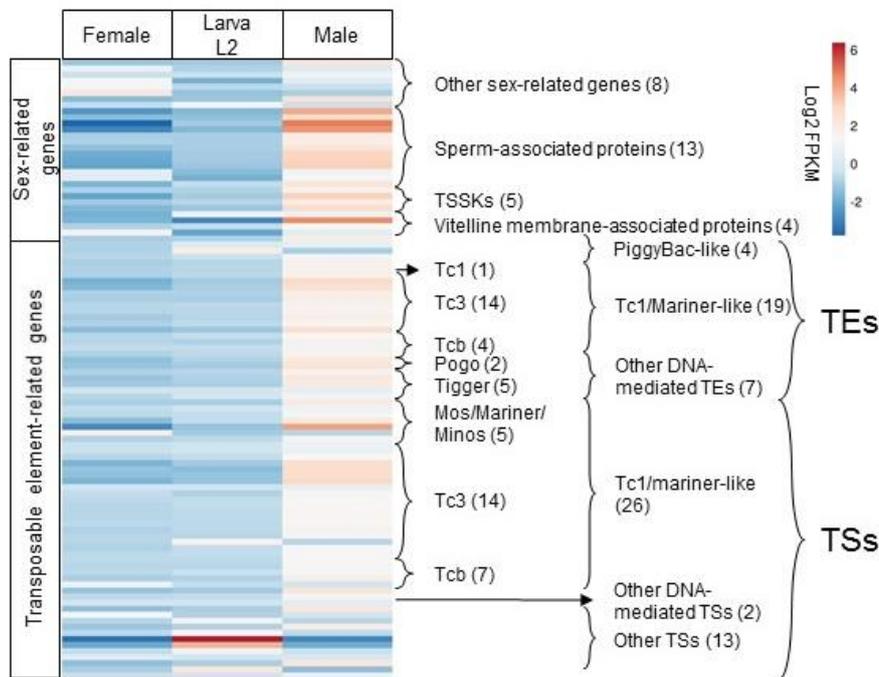


Figura 11. Heatmap mostrando as diferenças de expressão entre as bibliotecas para fêmeas, larvas e machos de *H.hampei* para genes que codificam proteínas relacionadas com TEs e proteínas relacionadas com reprodução. TEs: *transposable element-like proteins*, TSs: *transposase-like proteins*.

5.2.7 DEGs entre fases de larva e adulta

Ao comparar os estágios de larva L2 e fêmea, foram identificados vários genes diferencialmente expressos que poderiam ser elementos chave para a compreensão das diferenças fisiológicas e metabólicas entre os estágios reprodutivo e não reprodutivo da broca do café. Os genes anotados com os domínios da família de detoxificação do Citocromo p450 (CYP) foram ambos regulados positivamente (5 unigenes), e regulados negativamente (7 unigenes) em larva L2. Curiosamente, para genes relacionados ao metabolismo da cutícula, todos os DEGs foram positivamente regulados em larvas L2 e machos (25 unigenes cada). Os DEGs que foram anotados com funções relacionadas ao metabolismo de cutícula e quitina foram agrupados de acordo com suas proteínas preditas e são mostrados na Figura 12. Finalmente, genes com funções quimiossensoriais pertencentes à família OBP (Figura. 13) foram agrupados em classes diferentes por análise filogenética. A maioria dos genes OBPs foi regulada positivamente em adultos, com dois deles exclusivamente expressos em fêmeas e pertencentes à classe C-OBP. Adicionalmente, o único OBP regulado positivamente em larvas L2 foi agrupado nesta classe. Nenhum OBPs foi positivamente regulado em machos comparado com os que foram regulados dessa maneira em fêmeas.

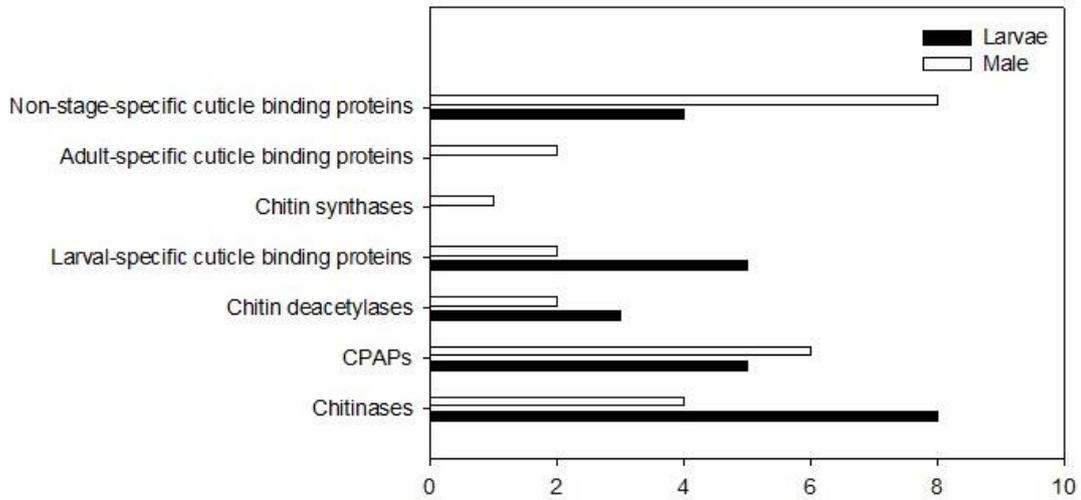


Figura 12. Número de unigenes (eixo x) regulados positivamente em larva L2 e macho em referência a fêmea para cada grupo de genes relacionados ao metabolismo de cutícula e quitina (eixo y). CPAPs: *cuticular proteins analogous to peritrophins*.

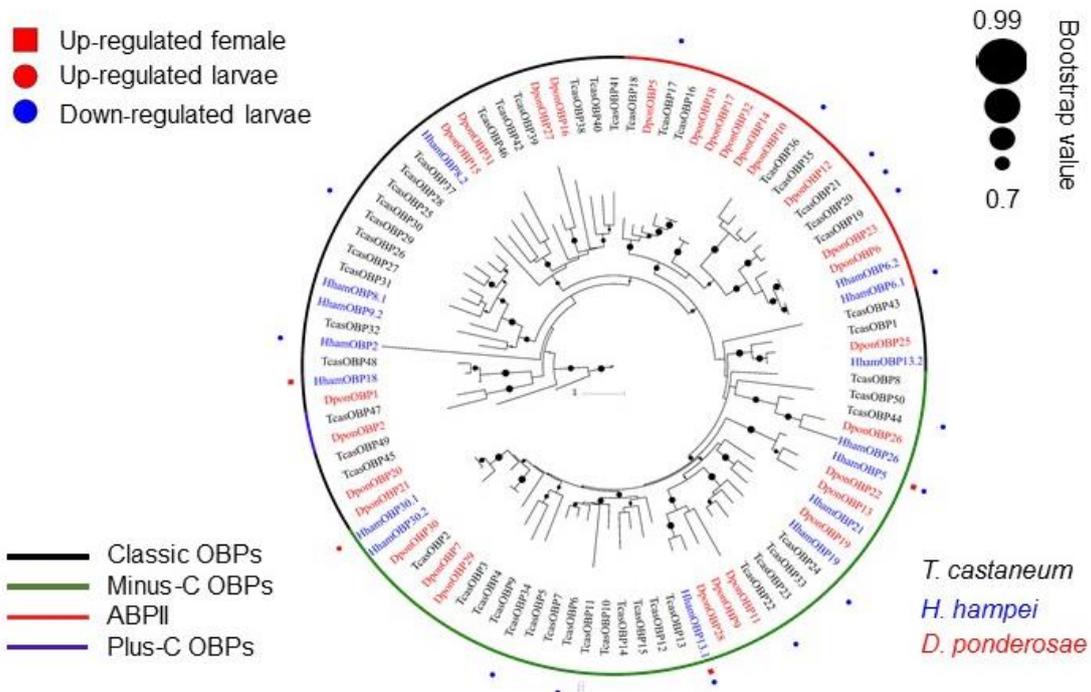
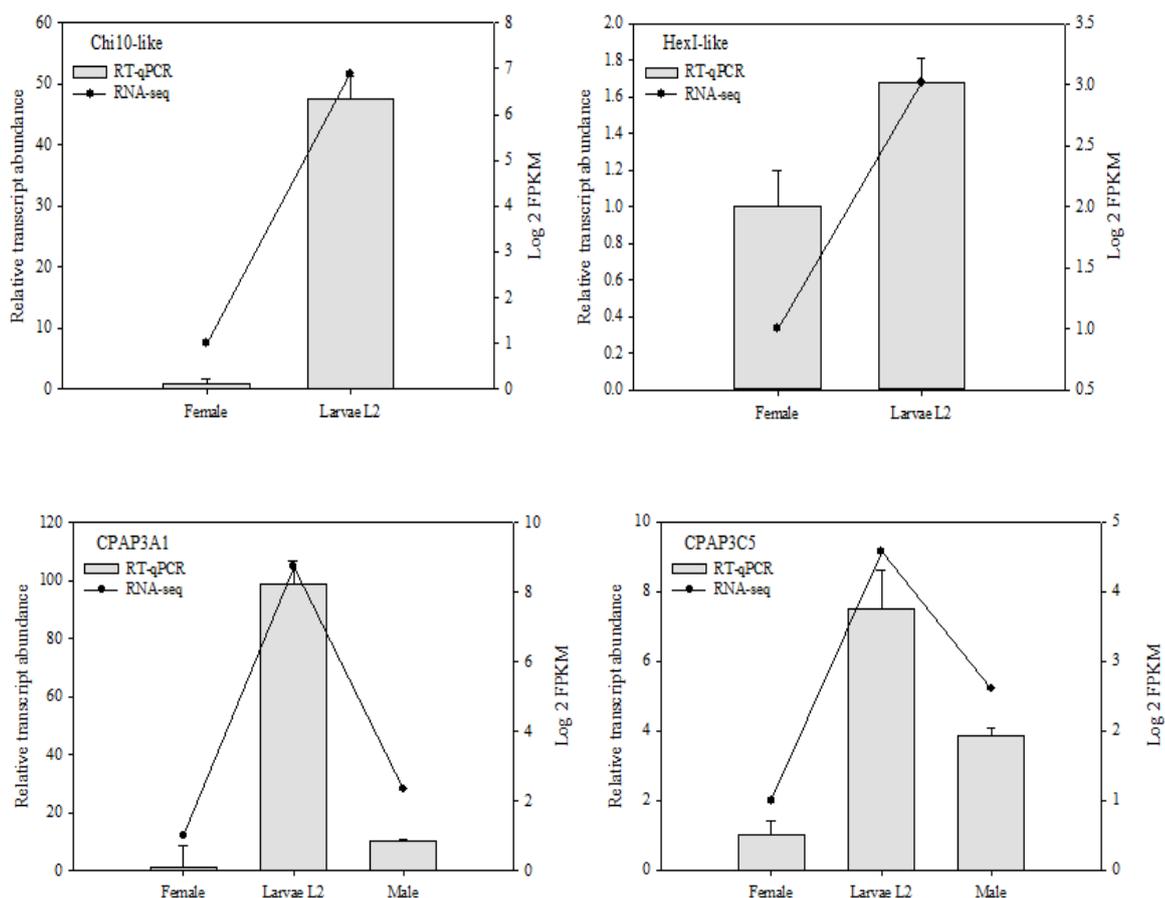


Figura 13. Relação filogenética dos genes OBP em *H. hampei* e em outros insetos da ordem coleóptera. *Up-regulated female*: Regulado positivamente em fêmeas; *Up-regulated larvae*: regulado positivamente em larvas L2 e *Down-regulated larvae*: regulado negativamente em larvas.

5.2.8 Validação de dados de RNA-seq por PCR em tempo real (RT-PCR)

Para validar os transcritos diferencialmente expressos identificados por análises de dados de RNA-seq (BROEDERS *et al.*, 2014), a RT-PCR foi conduzida com seis genes relacionados a processos de desenvolvimento do inseto: enzimas relacionadas ao hormônio juvenil (JH) (JH esterase e JH epóxido hidrolase), proteínas cuticulares análogas a peritrofinas (CPAPs) (CPAP3A1 e CPAP3C5), e proteínas envolvidas no metabolismo da quitina (hexosaminidase I e quitinase 10). No geral, os resultados da RT-PCR foram consistentes com os dados de RNA-seq, com um valor de correlação de 0,79 para todo o conjunto de genes que foram validados (Figura 14).



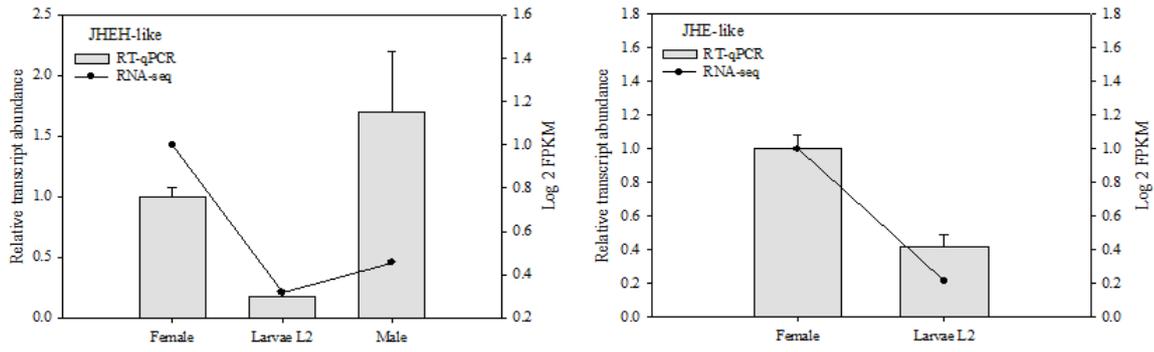


Figura 14. Resultados de RT-qPCR para genes diferencialmente expressos em diferentes estgios de desenvolvimento do CBB. Letras maisculas diferentes indicam uma diferena significativa entre os estgios ($P < 0,05$). ETE: hormnio esterase juvenil, JHEH: hormnio juvenil hidrolase, CPAP: protena cuticular anloga s peritrofinas, HexI: hexosaminidase I, Chi10: quitinase 10.

6. Discussão

Atualmente, é grande o número de pesquisas científicas que tem focado na determinação de métodos alternativos para o manejo de pragas na agricultura. Como já falado nesse trabalho, a broca do café causa prejuízos consideráveis na cafeicultura e, portanto, se torna cada vez mais importante, estudos que analisem suas bases moleculares “decifrando” suas características fisiológicas, ecológicas e comportamentais (DAVIES, 2014). Algumas pesquisas descrevem as características moleculares desse inseto, sua estrutura genômica, perfis de transcrição de genes ou proteoma. No genoma da broca do café, publicado em 2015, Vega e colaboradores avaliaram genes que estão envolvidos na digestão primária, bem como famílias de genes que possuem funções de desintoxicação de moléculas de defesa de plantas e inseticidas, tais como carboxilesterases, citocromo P450, glutathione S-transferases, transportadores de cassetes de ligação a ATP e um gene que confere resistência ao inseticida dieldrin. Além disso, outros estudos feitos com a broca do café comprovaram que a enzima α -amilase (*AmyHha*) está presente nos principais estágios de alimentação do inseto e é principalmente transcrita no trato intestinal de larvas de *H. hampei*, podendo ser considerada um alvo de grande valor para o controle de insetos em grãos de café por meio de estratégias biotecnológicas (BEZERRA et al., 2014). A transcritômica é uma abordagem útil nesse sentido porque fornece dados complementares para os genomas disponíveis, além de servir para analisar a expressão gênica de qualquer organismo em uma condição específica ou estágio de desenvolvimento, como foi no caso deste estudo.

Tanto a identificação de novos genes, como a caracterização funcional de grandes conjuntos de dados de genes e a estimativa de abundância de transcritos podem ser realizados e analisados por meio de sequenciamento de transcriptomas. Além dos transcriptomas de organismos modelos bem caracterizados como *Tribolium castaneum* e *Dendroctonus ponderosae* (DIPPEL,2014; KEELING,2012) os transcriptomas de diferentes estágios de desenvolvimento de outros coleópteros incluindo *Cylas formicarius*, *Batocera horsfieldi* e *Cytotrachelus buqueti* foram recentemente sequenciados (MA *et al.*,2016; YANG *et al.*,2016; YANG *et al.*,2016). Para este trabalho, foi feito um sequenciamento de alto rendimento para gerar um conjunto de 7,4Mbp de 29.434 unigenes, que é consideravelmente maior do que o genoma *draft* da broca (~28,4Mbp) com 19.222 genes previstos, dos quais apenas 15.546 foram expressos em fêmeas. Além disso, o tamanho previsto do genoma da broca, determinado por citometria de fluxo (~200Mbp) foi ligeiramente maior do que o relatado no artigo do genoma (~163 Mbp). No conjunto de dados produzidos neste trabalho, 3.7% e 8.1% dos genes foram diferencialmente expressos entre larva L2 e fêmea e entre fêmea e macho, respectivamente. O número menor de genes diferencialmente expressos entre fêmea e larva L2 pode ser explicado pelo fato de que macho não possui fase larval de 2º *instar*, portanto esse estágio analisado refere-se a futuros insetos fêmeas. De fato, encontramos diferenças maiores na expressão gênica quando comparamos machos e fêmeas ou machos e larva L1 do que quando comparamos machos e larva L2 (*instar* presente apenas em fêmeas). Curiosamente, estudos anteriores mostraram que a porcentagem de genes diferencialmente expressos entre os estágios não reprodutivos e reprodutivos (1% - 5.5%) é ligeiramente menor do que a porcentagem de genes diferencialmente expressos entre sexos na fase adulta (2.7% - 8.8%).

Dentro de uma perspectiva funcional, observamos que tanto para enriquecimento de termos GO, quanto para análise de vias metabólicas, os genes regulados positivamente nos machos foram enriquecidos com mais categorias funcionais (80 GOs e 14 KEGGs) do que qualquer outro conjunto de DEGs. Essa descoberta foi muito importante para entender melhor a determinação sexual no inseto, no qual a proporção sexual de machos é distorcida em relação as fêmeas com uma razão de 1:10. O conceito de haploidia masculina tem sido usado para explicar a determinação sexual na broca do café. Além disso, a ligação entre haplodiploidia e resistência a inseticidas tem sido proposta em algumas pesquisas. Brun e colaboradores comprovaram a haplodiploidia ‘funcional’ da broca do café em um exame citológico que revelou que os machos são na verdade diplóides, um conjunto de cromossomos foi condensado, e provavelmente não funcional, nas células somáticas de todos os machos examinados. Embora dois conjuntos de cromossomos estivessem presentes em espermatócitos primários, os cromossomos não conseguiram emparelhar antes da única divisão meiótica, e apenas um conjunto foi empacotado em espermatozóides, mostrando a haplodiploidia do inseto. Sua genética e história de vida podem, portanto, representar um interessante passo intermediário na evolução da verdadeira haplodiploidia. Entretanto, a genética associada à história de vida da broca do café é pouco compreendida e os genes específicos ou vias metabólicas envolvidas na determinação sexual permanecem desconhecidos. Neste estudo, foi mostrado que mais de 68% dos genes diferencialmente expressos eram regulados em machos e entre os genes positivamente regulados; há um destaque importante para aqueles associados com elementos transponíveis (TEs) e determinação sexual. Elementos transponíveis são seqüências de DNA intercaladas que se movem dentro do genoma e têm a capacidade de se propagar, atingindo finalmente altas proporções do genoma (HERNANDEZ-HERNANDEZ *et al.*, 2016). Este auxiliou para a caracterização dos TEs

presentes no genoma da broca do café, revelando que 11.51% deles pertencem a superfamília *Tc1/mariner*. Além disso, foi confirmado que pelo menos 8.2% do genoma da broca do café é composto por TEs. No presente estudo, foram identificados 724 transcritos anotados com a família TE (Pfam) e 1.367 unigenes com pelo menos uma correspondência com proteínas relacionadas ao TE na análise do BLAST. Dessa maneira, 2.5% das proteínas relacionadas a TEs são expressas no transcrito da broca do café, dos quais 4.7% são genes diferencialmente expressos. Como discutido em outros trabalhos, características reprodutivas podem estar relacionadas com essa baixa porcentagem de TEs. A haplodiploidia, que ocorre quando um conjunto de cromossomos dos machos não é transmitido para a geração seguinte, impede a disseminação de alguns genes de TEs. Esse fato pode explicar a maior abundância de genes que codificam proteínas relacionadas ao TEs em bibliotecas dos machos de RNAm encontradas nesse trabalho.

Além de comparar as diferenças de expressão entre machos e fêmeas, a comparação entre as fases reprodutiva e não reprodutiva revelou algumas características interessantes da variação da expressão gênica ao longo do ciclo de vida da broca do café. A transição larva-adulto e a regulação do desenvolvimento em insetos holometábolos são questões críticas, não apenas sob uma perspectiva evolutiva e ecológica, mas também para desenvolver métodos sustentáveis para o manejo de insetos. Este trabalho descreve diferenças de expressão entre os estágios larval e adulto. Além disso, identifica genes que foram previamente descritos no genoma deste inseto e que podem estar intimamente relacionados ao desenvolvimento da broca do café. Como foi mostrado nas análises de GO e enriquecimento de vias metabólicas, vários genes relacionados ao metabolismo de carboidratos e proteínas foram diferencialmente expressos em larvas em comparação a

fêmeas. No geral, a maioria das proteínas previstas envolvidas no metabolismo de celulose e hemicelulose (ex. endonucleases, celulasas (família das glicídes hidrolase) e xilanases, foram reguladas positivamente nas larvas. Tem sido relatado que várias espécies de coleópteros usam essas enzimas para digerir componentes da parede celular de plantas, especificamente celulose, hemicelulose e lignocelulose, como suas principais fontes de carboidrato. Também, os genes que codificam enzimas com atividade de protease foram regulados positivamente em larvas (serina protease e serol protease) e fêmeas (peptidil-dipeptidases e metaloproteases). Vários estudos tem mostrado maior expressão de genes codificadores de enzimas digestivas nos estágios larvais, provavelmente porque as fases não reprodutivas muitas vezes denominadas fases de alimentação envolvem comportamento alimentar mais ativo que na fase adulta. Enzimas digestivas são consideradas alvos importantes para o manejo de pragas, uma vez que possuem funções importantes na nutrição e sua inibição pode causar efeitos letais ou deletérios na fase larval. No entanto, tem sido relatado que genes de insetos rapidamente evoluem em resposta a inibidores direcionados a enzimas digestivas, significando que “variantes” desses genes podem ser cruciais no planejamento de estratégias mais precisas e eficientes para o controle de pragas.

Outros genes potencialmente úteis como alvos para RNAi são os que codificam proteínas cuticulares (CPs) e enzimas envolvidas no metabolismo de quitina. Em insetos, a estrutura da cutícula é determinada por proteínas cuticulares que desempenham papéis em muitos processos fisiológicos como proteção contra desidratação, inseticidas, patógenos e lesões físicas. Alguns genes que codificam proteínas de ligação cuticular e CPAPs foram diferencialmente expressos no conjunto de dados com maior expressão em larvas e machos do que em fêmeas. Da mesma forma, algumas quitoses, quitinases,

quetinas e desacetilases apresentaram maior expressão em larvas em machos do que em fêmeas. Curiosamente, os genes envolvidos na degradação da quitina aparecem mais regulados em machos do que em larvas ou fêmeas. O controle preciso da síntese de quitina e as taxas de degradação são essenciais para a homeostase do inseto. A determinação dos perfis de expressão de genes que regulam essas taxas pode ser útil para o planejamento de estratégia de controle de pragas (JASRAPURIA *et al.*, 2012).

Finalmente, o último grupo de genes estudados nesse trabalho foram as OBPs que possuem funções quimiossensoriais. Acredita-se que essa família multigênica esteja envolvida no reconhecimento químico que regula os comportamentos principais, incluindo escolha de hospedeiros, cópula e reprodução (YI *et al.*, 2017; ZHOU *et al.*, 2010; ANTHONY *et al.*, 2016). Em coleópteros, o olfato desempenha um importante papel na localização de parceiros ideais, locais para ovoposição, forrageamento, agregação e defesa a predadores. Apesar da importância do olfato, a base molecular da quimiorrecepção em coleópteros é pouco compreendida em comparação a outras ordens, como lepidópteros e dípteros (LI *et al.*, 2014; CUI *et al.*, 2018).

7. Conclusões e Perspectivas

O estudo aqui apresentado gerou, pela primeira vez, um banco de dados em conjunto com análise de expressão gênica para uma importante praga da cafeicultura, *H. hampei*. Uma vez publicada e depositada nos bancos de dados correspondentes, as sequências aqui obtidas enriquecerão o conhecimento relacionado com a biologia molecular, fisiologia e genética desse inseto, contribuindo enormemente para pesquisas futuras. A anotação das sequências e a quantificação da sua expressão em diferentes condições de alimentação, aportam informação acurada de genes que podem ser potencialmente explorados para desenhar estratégias de controle por técnicas como RNAi, CRISPR/Cas9 e toxinas *Bt* entre outras.

Dentre os potenciais alvos biotecnológicos foram identificados genes relacionados com processos fisiológicos fundamentais do inseto, como determinação sexual, funções quimiossensoriais, detoxificação, metabolismo de celulose, transporte de moléculas, proteínas cuticulares e enzimas envolvidas no metabolismo de quitina, entre outros. Outro aspecto a ser destacado é que foi possível descrever a resposta geral da expressão gênica que cada inseto exibe, em função do estágio de desenvolvimento em que ele se encontra. Isso é importante para identificar processos em comum, assim como particularidades desse inseto, que possam contribuir para aumentar o conhecimento sobre aspectos importantes da morfologia, comportamento e ecologia, e dessa maneira auxiliar em estratégias de custos relacionados ao controle da broca do café.

Finalmente é importante mencionar que é necessário desenvolver pesquisas futuras que complementem a informação gerada neste estudo, principalmente que visem

validar experimentalmente os dados obtidos, a partir de análises *in silico*. Alguns estudos deste tipo podem incluir a avaliação da expressão dos genes destacados neste estudo por RT-PCR, em diferentes fases de desenvolvimento de *H. hampei*.

Referências bibliográficas

- ADAMS, M. D. et al. The genome sequence of *Drosophila melanogaster*. **Science**, v. 287, n. 5461, p. 2185-2195, 2000.
- AGUILERA, C.G.; PADILLA, B.E.H., FLÓREZ, C.P.R., RUBIO, J.G., ACUNA, J.R.Z. ARN interferente: Potenciales usos en genómica funcional y control genético de *Hypothenemus hampei* (Coleoptera: Scolytinae) **Revista Colombiana de Entomología** 37(2); 167-172. 2011
- ALBUQUERQUE, É. et al. Seed-Specific Stable Expression of the α -AI1 Inhibitor in Coffee Grains and the *In Vivo*: Implications for the Development of the Coffee Berry Borer. *Trop Plant Biol.* v.8; p. 98-107. 2015
- ANDREWS, S. et al. FastQC A Quality Control tool for High Throughput Sequence Data. **Babraham Bioinformatics**. 2012 2016.
- ANTONY B, SOFFAN A, JAKŠE J, ABDELAZIM MM, ALDOSARI SA, ALDAWOOD AS, et al. Identification of the genes involved in odorant reception and detection in the palm weevil *Rhynchophorus ferrugineus*, an important quarantine pest, by antennal transcriptome analysis. **Bmc Genomics**. 17(1):69; 2016.
- ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ (ABIC). Disponível em www.abic.com.br em 10 de Dezembro de 2018.
- BARBOSA, A.; ALBUQUERQUE, É.; SILVA, M.; SOUZA, D.; OLIVEIRA-NETO.; VALENCIA, A., GROSSI-DE-SA, M. α -Amylase inhibitor-1 gene from *Phaseolus vulgaris* expressed in *Coffea arabica* plants inhibits α -amylases from the coffee berry borer pest. **BMC. Biotechnol.** 10(1):44. 2010.
- BAUM, J. A. et al. Control of coleopteran insect pests through RNA interference. **Nature biotechnology**, v. 25, n. 11, p. 1322-1326, 2007. ISSN 1087-0156.
- BENASSI, V. L. R. M; CARVALHO, C. H. S. Preferência de ataque a frutos de *Coffea arabica* e *Coffea canephora* pela broca-do-café (*Hypothenemus hampei* Ferrari, 1867 Coleoptera, Scolytidae). **Revista de Agricultura**, Piracicaba, v. 69, n. 1, p. 103-111. 1994.
- BENASSI, V.L.R.M. Levantamento dos inimigos naturais da broca do café *Hypothenemus hampei* (Ferr)(Coleoptera: Scolytidae) no norte do Espírito Santo. **EMCAPA**, p.635-638. 1995.
- BEZERRA C, MACEDO L, AMORIM T, SANTOS V, FRAGOSO RDR, LUCENA W, et al. Molecular cloning and characterization of an α -amylase cDNA highly expressed in major feeding stages of the coffee berry borer, *Hypothenemus hampei*. **Gene**. 553(1):7, 2014;
- BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014.

- BULLARD, J. H. et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. **BMC bioinformatics**, v. 11, n. 1, p. 94, 2010.
- CEJA-NAVARRO, J.A. et al. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. **Nature Communications**, 6:7618, 2015.
- COLLINS, F. S. et al. A vision for the future of genomics research. **Nature**, v. 422, n. 6934, p. 835-847, 2003.
- CONAB. Acompanhamento da safra brasileira: café, safra 2017/18. Brasília: Conab. 4: 1-57 p. 2017.
- CONSORTIUM, I. S. G. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. **Insect biochemistry and molecular biology**, v. 38, n. 12, p. 1036-1045, 2008.
- CRAVA, C. M.; BRÜTTING, C.; BALDWIN, I. T. Transcriptome profiling reveals differential gene expression of detoxification enzymes in a hemimetabolous tobacco pest after feeding on jasmonate-silenced *Nicotiana attenuata* plants. **BMC genomics**, v. 17, n. 1, p. 1005, 2016.
- CRUSOE, M. R. et al. The khmer software package: enabling efficient nucleotide sequence analysis. **F1000Research**, v. 4, 2015.
- CUI X, LIU D, SUN K, HE Y, SHI X. Expression Profiles and Functional Characterization of Two Odorant-Binding Proteins From the Apple Buprestid Beetle *Agilus mali* (Coleoptera: Buprestidae). **Journal of economic entomology**. 2018.
- DAVIES, S. A. et al. Cell signalling mechanisms for insect stress tolerance. **Journal of Experimental Biology**, v. 217, n. 1, p. 119-128, 2014.
- DAVOLOS, C. C. et al. Binding analysis of *Bacillus thuringiensis* Cry1 proteins in the sugarcane borer, *Diatraea saccharalis* (Lepidoptera: Crambidae). **Journal of invertebrate pathology**, v. 127, p. 32-34, 2015.
- DE ASSIS FONSECA, F. C. Aplicação de estratégias moleculares visando o controle da broca-gigante da cana-de-açúcar (*Telchin licus licus*, Drury 1770) (Lepidoptera: Castniidae). 2013. 234 (PhD). Departamento de Biologia Celular, Universidade de Brasília, Não Publicado.
- DE ASSIS FONSECA, F. C. et al. Sugarcane giant borer transcriptome analysis and identification of genes related to digestion. **PloS one**, v. 10, n. 2, p. e0118231, 2015.
- DIPPEL S, OBERHOFER G, KAHNT J, GERISCHER L, OPITZ L, SCHACHTNER J, et al. Tissue-specific transcriptomics, chromosomal localization, and phylogeny of chemosensory and odorant binding proteins from the red flour beetle *Tribolium castaneum* reveal subgroup specificities for olfaction or more general functions. **BMC genomics**.;15(1):1141.2014.
- FIERS, W. et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. **Nature**, v. 260, n. 5551, p. 500-507, 1976.

- FIRE, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. **Nature**, v. 391, n. 6669, p. 806-811, 1998.
- FIRMINO, A. A. P. et al. Transcriptome analysis in cotton boll weevil (*Anthonomus grandis*) and RNA interference in insect pests. **PLoS One**, v. 8, n. 12, p. e85079, 2013.
- FU, C. et al. next-generation sequencing. **Journal of Medical Genetics**, v. 54, n. 3, p. 190-195, 2017.
- GAZARA, R. K. et al. De novo transcriptome sequencing and comparative analysis of midgut tissues of four non-model insects pertaining to Hemiptera, Coleoptera, Diptera and Lepidoptera. **Gene**, v. 627, p. 85-93, 2017.
- GHILDIYAL, M.; ZAMORE, P.D. Small silencing RNAs: an expanding universe. **Nature reviews Genetics**, v. 10, n 2, p. 94-108, 2009.
- GOG, L. et al. Larval *Helicoverpa zea* transcriptional, growth and behavioral responses to nicotine and *Nicotiana tabacum*. **Insects**, v. 5, n. 3, p. 668-688, 2014.
- GOMÉZ, J. The Coffee Berry Borer (Coleoptera: Curculionidae): How many instars are there? **Annals of the Entomological Society of America** v.108, n.3, p.311-315. 2015
- GRABHERR, M. G. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature biotechnology**, v. 29, n. 7, p. 644, 2011.
- GRIMMELIKHUIJZEN, C. J.; CAZZAMALI, G.; WILLIAMSON, M.; HAUSER, F. The promise of insect genomics. **Pest Management Science**, v.63, n.5, p. 413-416, 2007.
- GUHARAY, J.; MONTERREY, J. Manejo ecologico de la broca del cafeto (*Hypothenemus hampei*) em America Central. **Manejo Integrado de Plagas**, n. 22, p. i-viii, set. 1997.
- GULLAN, P. J.; CRANSTON, P. S. **The insects: an outline of entomology**. 5. John Wiley & Sons. 624. 2014
- GUO, S.; KEMPHUES, K. J. par-1, a gene required for establishing polarity in *C.elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. **Cell**, v.81, n.4, p. 611-20, 1995.
- HAAS, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature protocols**, v. 8, n. 8, p. 1494-1512, 2013.
- HECKEL, D. G. Insect detoxification and sequestration strategies. In: VOELCKEL, C. e JANDER, G. (Ed.). **Insect-Plant Interactions**. Chichester, West Sussex, UK: John Wiley & Sons, v.47. cap. 3, p.77-114. 2014.
- HERDE, M.; HOWE, G. A. Host plant-specific remodeling of midgut physiology in the generalist insect herbivore *Trichoplusia ni*. **Insect biochemistry and molecular biology**, v. 50, p. 58-67, 2014.
- HERNANDEZ-HERNANDEZ, E.M. et al. Genome-wide analysis of transposable elements in coffee berry borer *Hypothenemus hampei* (Coleoptera: Curculionidae): description of novel families. **Molecular Genetics and Genomics** 292, 565-583. 2017

HOLT, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. **Science**, v. 298, n. 5591, p. 129-149, 2002.

ICO, **International Coffee Organization**. Disponível em <http://www.ico.org/> em 20 de Novembro de 2018.

INFANTE, F. , PÉREZ, J., VEGA, F.E. The coffee berry borer: the centenary of a biological invasion in Brasil. **Braz. J. Biol** v.74, n. 3 (suppl.); 125-126. 2013.

INFANTE, F. Pest Management Strategies Against the Coffee Berry Borer (Coleoptera: Curculionidae: Scolytinae). **J Agric Food Chem**. V.30, N.66(21; 5275-5280.2018.

JASRAPURIA S, SPECHT CA, KRAMER KJ, BEEMAN RW, MUTHUKRISHNAN S. Gene families of cuticular proteins analogous to peritrophins (CPAPs) in *Tribolium castaneum* have diverse functions. **PLoS One**. ;7(11):e49844; 2012.

JOGA, M. R. et al. RNAi efficiency, systemic properties, and novel delivery methods for pest insect control: what we know so far. **Frontiers in physiology**, v. 7, 2016.

KEELING CI, HENDERSON H, LI M, YUEN M, CLARK EL, FRASER JD, et al. Transcriptome and full-length cDNA resources for the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major insect pest of pine forests. **Insect biochemistry and molecular biology**. 42(8):525-36; 2012.

LEROY, T., HENRY, A.; ROYER, M.; ALTOSAAR, I.; FRUTOS, R.; DURIS, D.; PHILLIPE, R.. Genetically modified coffee plants expressing the *Bacillus thuringiensis* cry1Ac gene for resistance to leaf miner. **Plant Cell Reports**, 19:382-289. 2000

LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC bioinformatics**, v. 12, n. 1, p. 323, 2011.

LI, H. et al. Transcriptome analysis and screening for potential target genes for RNAi-mediated pest control of the beet armyworm, *Spodoptera exigua*. **PLoS One**, v. 8, n. 6, p. e65931, 2013.

LI H, ZHANG A, CHEN L-Z, ZHANG G, WANG M-Q, ZHANG Y. Construction and analysis of cDNA libraries from the antennae of *Batocera horsfieldi* and expression pattern of putative odorant binding proteins. **Journal of Insect Science**; 14(1); 2014.

LI, H.-S. et al. Variation in life history traits and transcriptome associated with adaptation to diet shifts in the ladybird *Cryptolaemus montrouzieri*. **BMC genomics**, v. 17, n. 1, p. 281, 2016.

LIU, Y. et al. Comparative transcriptome analysis of *Glyphodes pyloalis* Walker (Lepidoptera: Pyralidae) reveals novel insights into heat stress tolerance in insects. **BMC genomics**, v. 18, n. 1, p. 974, 2017.

LIVAK, K. & SCHMITTGEN, T. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(-Delta Delta C (T)). **CAS PubMed Article**, v.4, p.402-408. 2001.

MA, J. et al. Identification and comparative study of chemosensory genes related to host selection by legs transcriptome analysis in the tea geometrid *Ectropis obliqua*. **PloS one**, v. 11, n. 3, p. e0149591, 2016.

MA, L. et al. Transcriptome and gene expression analysis of *Cylas formicarius* (Coleoptera: Brentidae) during different development stages. **Journal of Insect Science**, v. 16, n. 1, p. 63, 2016.

MAPA - MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Vegetal: Café. Disponível em www.agricultura.gov.br/culturas/cafe/saiba-mais. Acesso em 30 de novembro de 2018.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, n. 10, p. 671-682, 2011.

MARTINS, A.L. *A História do Café*. 1º edição. Editora Contexto. 2008

METSALU, T.; VILO, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. **Nucleic acids research**, v. 43, n. W1, p. W566-W570, 2015.

MORAES, J.C. Pragas do cafeeiro: importância e métodos alternativos de controle. Lavras: UFLA-FAEPE, 74p. 1998.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature methods**, v. 5, n. 7, p. 621-628, 2008.

NENE, V. et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. **Science**, v. 316, n. 5832, p. 1718-1723, 2007.

NEVES, C. - A história do café. Rio de Janeiro. **Instituto Brasileiro do Café**. 52 p.1974.

NÓBREGA, J. D.; DORNELAS, M. Biotecnologia e melhoramento da cana-de-açúcar. In: SEGATO, S.;PINTO, A., *et al* (Ed.). **Atualização em produção de cana-de-açúcar**. São Paulo: Piracicaba, p.39-56. 2006.

PERERA, O. P. et al. Generation of a transcriptome in a model lepidopteran pest, *Heliothis virescens*, using multiple sequencing strategies for profiling midgut gene expression. **PloS one**, v. 10, n. 6, p. e0128563, 2015.

PRICE, D. R.; GATEHOUSE, J. A. RNAi-mediated crop protection against insects. **Trends in biotechnology**, v. 26, n. 7, p. 393-400, 2008.

PRÜFER, K. et al. FUNC: a package for detecting significant associations between gene sets and ontological annotations. **BMC bioinformatics**, v. 8, n. 1, p. 41, 2007.

QI, L. et al. De novo assembly and developmental transcriptome analysis of the small white butterfly *Pieris rapae*. **PloS one**, v. 11, n. 7, p. e0159258, 2016.

RICHARDS, S.; GIBBS, R. A.; GERARDO, N. M. Genome sequence of the pea aphid *Acyrtosiphon pisum*. **PLoS Biol**, v. 8, n. 2, p. e1000313, 2010.

RICHARDS, S. et al. The genome of the model beetle and pest *Tribolium castaneum*. **Nature**, v. 452, n. 7190, p. 949-955, 2008.

- ROBERTSON, G. et al. De novo assembly and analysis of RNA-seq data. **Nature methods**, v. 7, n. 11, p. 909-912, 2010.
- ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, 2010.
- ROBINSON, M. D.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome biology**, v. 11, n. 3, p. R25, 2010.
- RODRIGUES, T. B.; FIGUEIRA, A. Management of Insect Pest by RNAi—A New Tool for Crop Protection. In: (Ed.). **RNA Interference: InTech**, 2016.
- ROY, A. et al. Data set for diet specific differential gene expression analysis in three Spodoptera moths. **Data in brief**, v. 8, p. 448-455, 2016.
- ROY, A. W. et al. Diet dependent metabolic responses in three generalist insect herbivores Spodoptera spp. **Insect biochemistry and molecular biology**, v. 71, p. 91-105, 2016.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. Proceeding of the National Academy of Sciences of the United States of America, v. 74, n. 12, p. 5463-5467, 1977.
- SAIKI, R. et al. Polymerase chain reaction. **Science**, v. 230, p. 1350-1354, 1985.
- SCOTT, J. G. et al. Towards the elements of successful insect RNAi. **Journal of insect physiology**, v. 59, n. 12, p. 1212-1221, 2013.
- SCUDDER, G. The importance of insects. In: FOOTITT, R. G. e ADLER, P. H. (Ed.). **Insect biodiversity: science and society: John Wiley & Sons**, cap. 2, p.7-32. 2009.
- SIMÃO, F. A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics**, v. 31, n. 19, p. 3210-3212, 2015.
- SONESON, C.; DELORENZI, M. A comparison of methods for differential expression analysis of RNA-seq data. **BMC bioinformatics**, v. 14, n. 1, p. 91, 2013.
- SOUZA, J.; REIS, P.R. Broca do café: Histórico, reconhecimento, biologia, prejuízos, monitoramento e controle. **Boletim Técnico - EPAMIG**. v. 50, 40 p. 1997.
- SOUZA, T. P. et al. Comparative analysis of expression profiling of the trypsin and chymotrypsin genes from Lepidoptera species with different levels of sensitivity to soybean peptidase inhibitors. **Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology**, v. 196, p. 67-73, 2016.
- SPARKS, M. E. et al. A Transcriptome Survey Spanning Life Stages and Sexes of the Harlequin Bug, *Murgantia histrionica*. **Insects**, v. 8, n. 2, p. 55, 2017.
- SPIT, J. et al. Effects of different dietary conditions on the expression of trypsin-and chymotrypsin-like protease genes in the digestive system of the migratory locust, *Locusta migratoria*. **Insect biochemistry and molecular biology**, v. 48, p. 100-109, 2014.
- STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, v. 30, p. 1312-1313, 2014.

- SUPEK, F. et al. REVIGO summarizes and visualizes long lists of gene ontology terms. **PloS one**, v. 6, n. 7, p. e21800, 2011.
- TANG, B. et al. Transcriptome analysis and identification of induced genes in the response of *Harmonia axyridis* to cold hardiness. **Comparative Biochemistry and Physiology Part D: Genomics and Proteomics**, v. 22, p. 78-89, 2017.
- TAUNAY, A. de E. - História do café no Brasil: no Brasil Imperial 1822-1872. Rio de Janeiro, **Departamento Nacional do Café**, 1939.
- TELLAM, R. L. The peritrophic matrix. In: LEHANE, M. e BILLINGSLEY, P. (Ed.). **Biology of the insect midgut**. 1. Netherlands: Springer Science & Business Media, cap. 4, p.86-114. 1996.
- VEGA FE, INFANTE F, CASTILLO A, JARAMILLO J. The coffee berry borer, *Hypothenemus hampei* (Ferrari)(Coleoptera: Curculionidae): a short review, with recent findings and future research directions. **Terrestrial Arthropod Reviews**. v.2(2)129. 2009
- VEGA F.E., BROWN SM, CHEN H, SHEN E, NAIR MB, CEJA-NAVARRO JA, et al. Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer, *Hypothenemus hampei*. **Scientific reports**. 5:12525. 2015
- VEGA F.E., SIMPKINS A, MIRANDA J, HARNLY JM, INFANTE F, CASTILLO A, et al. A Potential Repellent Against the Coffee Berry Borer (Coleoptera: Curculionidae: Scolytinae). **Journal of Insect Science**. 17(6):122. 2017.
- VENTER, J. C. et al. The sequence of the human genome. **Science**, v. 291, n. 5507, p. 1304-1351, 2001.
- VOGEL, H. et al. A comprehensive transcriptome and immune-gene repertoire of the lepidopteran model host *Galleria mellonella*. **BMC genomics**, v. 12, n. 1, p. 308, 2011.
- VOGEL, H.; MUSSER, R. O.; CELORIO-MANCERA, M.-L. Transcriptome responses in herbivorous insects towards host plant and toxin feeding. **Annual plant reviews**, v. 47, p. 197-233, 2014.
- WANG, Y. et al. Identification and comparison of candidate olfactory genes in the olfactory and non-olfactory organs of elm pest *Ambrostoma quadriimpressum* (Coleoptera: Chrysomelidae) based on transcriptome analysis. **PloS one**, v. 11, n. 1, p. e0147144, 2016.
- WANG, Y. et al. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. **Nature**, v. 461, n. 7265, p. 754-761, 2009.
- WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews genetics**, v. 10, n. 1, p. 57-63, 2009.
- WEI, D. et al. Transcriptome analysis to identify genes for peptides and proteins involved in immunity and reproduction from male accessory glands and ejaculatory duct of *Bactrocera dorsalis*. **Peptides**, v. 80, p. 48-60, 2016.

- WEINSTOCK, G. M.; ROBINSON, G. E.; GIBBS, R. A. Insights into social insects from the genome of the honeybee *Apis mellifera*. **Nature**, v. 443, n. 7114, p. 931-949, 2006.
- WINNEBACK, E.C.; MILLAR, C.D.; WARMAN, G.R. Why does insect RNA look degraded? **Journal of Insect Sciences**, v. 10, p. 159-165. 2010.
- WU, C. et al. De Novo Transcriptome Analysis of the Common New Zealand Stick Insect *Clitarchus hookeri* (Phasmatodea) Reveals Genes Involved in Olfaction, Digestion and Sexual Reproduction. **PloS one**, v. 11, n. 6, p. e0157783, 2016.
- YANG, P. et al. Transcriptome analysis of sexually dimorphic Chinese white wax scale insects reveals key differences in developmental programs and transcription factor expression. **Scientific reports**, v. 5, 2015.
- YANG H, CAI Y, ZHUO Z, YANG W, YANG C, ZHANG J, et al. Transcriptome analysis in different developmental stages of *Batocera horsfieldi* (Coleoptera: Cerambycidae) and comparison of candidate olfactory genes. **PloS one**. v.13(2). 2018
- YANG H, SU T, YANG W, YANG C, LU L, CHEN Z. The developmental transcriptome of the bamboo snout beetle *Cyrtotrachelus buqueti* and insights into candidate pheromone-binding proteins. **PloS one**. v.12(6). 2017
- YI, X. et al. Differential expression of chemosensory-protein genes in midguts in response to diet of *Spodoptera litura*. **Scientific Reports**, v. 7, 2017.
- YIN, C. et al. ChiloDB: a genomic and transcriptome database for an important rice insect pest *Chilo suppressalis*. **Database**, v. 2014, 2014.
- YIN, C. et al. InsectBase: a resource for insect genomes and transcriptomes. **Nucleic acids research**, v. 44, n. D1, p. D801-D807, 2016.
- ZAMORE, P.D.; TUSCHL, T.; SHARP, P.A.; BARTEL, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21-23 nucleotide intervals. *Cell*, v.101, n.1, p. 25-33, 2000.
- ZDOBNOV, E. M.; BORK, P. Quantification of insect genome divergence. **Trends in Genetics**, v. 23, n. 1, p. 16-20, 2007.
- ZHANG, J. et al. The impact of next-generation sequencing on genomics. **Journal of genetics and genomics**, v. 38, n. 3, p. 95-109, 2011.
- ZHANG, Q.; YUAN, M. Progress in insect transcriptomics based on the next-generation sequencing technique. **Acta Entomologica Sinica**, v. 56, n. 12, p. 1489-1508, 2013.
- ZHANG, Y.-N. et al. Molecular identification and expression patterns of carboxylesterase genes based on transcriptome analysis of the common cutworm, *Spodoptera litura* (Lepidoptera: Noctuidae). **Journal of Asia-Pacific Entomology**, v. 19, n. 4, p. 989-994, 2016.
- ZHOU J-J. Odorant-binding proteins in insects. **Vitamins & hormones**. 83: Elsevier. p. 241-72, 2010

BARBOSA, H.R

DADOS PESSOAIS

Nome: Helena Ribeiro Barbosa

Local e data de nascimento: Porto Alegre, RS, 14.02.1984

Endereço Profissional: Embrapa Recursos Genéticos e Biotecnologia

Laboratório de Interação Molecular Planta-Praga

PqEB – Parque Estação Biológica Norte

AV. W5 Norte

70770-900 Brasília-DF

Telefone profissional: (61) 34484705

FORMAÇÃO

Graduação em Ciências Biológicas - Universidade Católica de Brasília (2007-2011)

ESTÁGIOS

Bolsista Iniciação Científica – CAPES (2009-2010)

Estagiária no Laboratório de Interação Molecular Planta-Praga

Embrapa Cenargen

Transformação de *Gossypium Hirsutum* via *Agrobacterium tumefaciens*

Estagio não remunerado (2010)

Projeto Tamar

Ubatuba – São Paulo

Manejo e reabilitação de tartarugas marinhas

Bolsista de DTI - Laboratório de Interação Molecular Planta-Praga (2015)

Embrapa Cenargen

Expressão do fator de transcrição CAHB12 em bactéria.

RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

Transcriptome analysis of main developmental stages of coffee berry borer *Hypothenemus hampei* (Coleoptera: Curculionidae) revealed candidate genes for pest management by RNAi

Helena R. Barbosa, Daniel D. Noriega, Fabricio B.M. Arraes, Roberta R. Coelho, Erika V.S. Albuquerque, Maria F. Grossi-de-Sa, Roberto C. Togawa, Priscila Grynberg, , Maria C.M da Silva, Arnubio V. Jimenez. In: 7th Brazilian Biotechnology Congress – Brasília-DF 18 a 21 de novembro de 2018.