



**LORRANA VERDI FLORES**

**PROSPECÇÃO DE TRANSCRITOS CODIFICADORES DE  
QUITINASES EM DADOS DE RNA-SEQ DA FORMIGA  
CORTADEIRA *Atta sexdens***

**LAVRAS-MG**

**2023**

**LORRANA VERDI FLORES**

**PROSPECÇÃO DE TRANSCRITOS CODIFICADORES DE QUITINASES EM  
DADOS DE RNA-SEQ DA FORMIGA CORTADEIRA *Atta sexdens***

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Biotecnologia Vegetal, para a obtenção do título de Mestre.

Prof. Dr. Luciano Vilela Paiva

Orientador

Dra. Nathaly Lara Castellanos

Coorientadora

Dr. Wesley Pires Flausino Máximo

Coorientador

**LAVRAS-MG**

**2023**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Flores, Lorrana Verdi.

Prospecção de transcritos codificadores de quitinases em dados  
de RNA-Seq da formiga cortadeira *Atta sexdens* / Lorrana Verdi  
Flores. - 2023.

66 p.

Orientador(a): Luciano Vilela Paiva.

Coorientador(a): Nathaly Lara Castellanos, Wesley Pires  
Flausino Máximo.

Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2023.

Bibliografia.

1. Formigas cultivadoras de fungo. 2. Transcritoma. 3. 18-  
Glicosil-hidrolases. I. Paiva, Luciano Vilela. II. Castellanos,  
Nathaly Lara. III. Máximo, Wesley Pires Flausino. IV. Título.

**LORRANA VERDI FLORES**

**PROSPECÇÃO DE TRANSCRITOS CODIFICADORES DE QUITINASES EM  
DADOS DE RNA-SEQ DA FORMIGA CORTADEIRA *Atta sexdens***

**PROSPECTING CHITINASE CODING TRANSCRIPTS IN LEAF-CUTTING  
ANT *Atta sexdens* RNA-SEQ DATA.**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Biotecnologia Vegetal, para a obtenção do título de Mestre.

APROVADA em 17 de janeiro de 2023.  
Dr. Filippe Elias de Freitas Soares UFLA  
Dr. Leonardo Lima Pepino de Macedo EMBRAPA

Prof. Dr. Luciano Vilela Paiva

Orientador

Dra. Nathaly Lara Castellanos

Coorientadora

Dr. Wesley Pires Flausino Máximo

Coorientador

**LAVRAS-MG**

**2023**

## **AGRADECIMENTOS**

Agradeço à minha família pelo apoio emocional.

Ao meu orientador e coorientadores pelo apoio intelectual e emocional, às vezes tão necessários no mundo acadêmico.

Ao programa de pós-graduação em Biotecnologia Vegetal e à Universidade Federal de Lavras (UFLA) pela infraestrutura.

À Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) pelo apoio financeiro.

Muito obrigada!

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

*“O começo de todas as ciências é o espanto de as coisas serem o que são”.*

*Aristóteles*

## RESUMO

A formiga cortadeira *Atta sexdens* apresenta ampla distribuição no Brasil e é considerada um importante inseto-praga de diversas culturas agrícolas devido à sua intensa atividade de forrageio de material vegetal fresco. A intensa desfolhação dessas formigas ocorre em virtude da necessidade de fornecerem alimentos para cultivo do fungo (*Leucoagaricus gongylophorus*) com o qual vivem em simbiose e do qual também se alimentam. Embora muito se saiba sobre a bioecologia destes insetos, ainda são poucos os estudos de validação da função gênica de famílias proteicas que desempenham funções importantes no desenvolvimento das formigas cortadeiras, uma vez que são quase inexistentes os dados moleculares disponíveis para estas espécies. O objetivo deste trabalho consistiu em obter o transcrito da formiga cortadeira *A. sexdens* proveniente do sequenciamento de RNA de cinco amostras biológicas (larva inteira, intestino de larva, formiga forrageadora inteira, cabeça e intestino de formiga forrageadora), visando identificar transcritos codificadores de quitinases, que são enzimas envolvidas na rota de degradação da quitina - um importante polissacarídeo componente de matrizes extracelulares e que desempenha diversas funções nos insetos. A partir dos dados brutos do sequenciamento das amostras foi conduzida análise de qualidade e, posteriormente, as leituras processadas foram utilizadas na montagem *de novo* do transcrito por meio da ferramenta Trinity. A predição de ORFs (do inglês - Open Reading Frames) ocorreu a partir de dados de similaridade e identificação de domínios proteicos, com identificação de 45.253 sequências candidatas a transcritos codificadores. Deste total, 36.639 sequências apresentaram ao menos um domínio depositado no banco de dados Pfam, sendo que foi possível anotar 24.392 sequências contendo ortólogos diretos nos bancos de dados disponíveis no eggNOG. De maneira geral, os transcritos presentes em maior quantidade no transcrito de *A. sexdens* se enquadraram na categoria de mecanismos de transdução de sinais. Após a análise filogenética foram identificados 23 transcritos codificadores de quitinases, divididos em 11 grupos de quitinases comumente encontradas em insetos. De modo geral, quinze das quitinases identificadas em *A. sexdens* apresentaram todos os domínios esperados. Portanto, com a identificação *in silico* de potenciais candidatas a quitinases no transcrito de *A. sexdens* será possível ampliar o conhecimento molecular e fisiológico acerca da espécie e contribuir com estudos posteriores que visem elucidar o papel destas enzimas no desenvolvimento da formiga cortadeira *Atta sexdens*.

**Palavras-chave:** Formigas cultivadoras de fungo. Transcrito. 18-glicosil-hidrolases.

## ABSTRACT

The leaf-cutting ant *Atta sexdens* is widely distributed in Brazil and is considered an important insect pest of several crops due to its intense foraging activity on fresh plant material. The intense defoliation performed by ants takes place to grow the symbiotic fungus (*Leucoagaricus gongylophorus*) on which ants feed. Although much is known about the ants bioecology, there are still few studies to validate the gene function of protein families that play important roles in the development of leaf-cutting ants, since little molecular data is currently available for these species. This work aimed to obtain the leaf-cutting ant *A. sexdens* transcriptome by RNA sequencing of five biological samples (whole larva, larval gut, whole foraging ant, head, and gut of foraging ant), aiming to identify sequences encoding chitinases – enzymes involved in the degradation route of chitin, a vital polysaccharide for building extracellular matrices and for accomplishing other functions in insects. After sequencing, *de novo* assembly of the transcripts was performed using the Trinity tool. The prediction of Open Reading Frames occurred from similarity data and identification of protein domains, with identification of 45,253 protein-coding sequences. Of this total, 36,639 sequences had at least one domain deposited in the Pfam database, and it was possible to annotate 24,392 sequences containing direct orthologs in databases available in eggNOG. In general, the transcripts present in larger quantities in *A. sexdens* transcriptome fell into the category signal transduction mechanisms. From the phylogenetic analysis, twenty-three sequences encoding chitinases were identified, divided into eleven groups of chitinases commonly found in insects, of which fifteen presented all the expected domains. Therefore, with the *in silico* identification of potential candidates for chitinases in the *A. sexdens* transcriptome it will be possible to expand molecular and physiological knowledge about the species and contribute to further studies aimed at elucidating the role of these enzymes in the development of leaf-cutting ants.

Keywords: Fungus-Farming Ants. Transcriptome. 18 glycosyl hydrolases.

## LISTA DE TABELAS

<b>Tabela 1</b> - Sequências utilizadas como referência para identificação das candidatas a quitinases no transcrito de <i>A. sexdens</i> .....	44
<b>Tabela 2</b> - Resumo estatístico por amostra biológica de <i>A. sexdens</i> obtido após análise dos dados de sequenciamento de RNA e taxa de alinhamento das leituras ao transcrito montado via abordagem <i>de novo</i> utilizando o programa Trinity.....	46
<b>Tabela 3</b> - Resumo estatístico do transcrito da formiga cortadeira <i>A. sexdens</i> obtido após a utilização de análises nativas da ferramenta Trinity.....	47
<b>Tabela 4</b> - Análises de cobertura de proteínas do banco de dados UniprotKB/Swiss-ProtKB e completude do transcrito de <i>A. sexdens</i> obtidas utilizando os programas Blast-X, análises nativas da ferramenta Trinity e BUSCO.....	48
<b>Tabela 5</b> - Transcritos codificadores em <i>A. sexdens</i> preditos a partir da ferramenta TransDecoder e identificados como candidatos a quitinases através de análise filogenética. ....	51

## LISTA DE ILUSTRAÇÕES

- Figura 1** - Representação do ciclo de vida e das castas em uma sociedade de formigas cortadeiras.....27
- Figura 2** - Visão geral das etapas do metabolismo de quitina em insetos. .... 33
- Figura 3** - Representação intestino de formiga cortadeira adulta e larva em relação a presença de quitinases em estruturas específicas, as quais estão identificadas em negrito e cor vermelha. .... 39
- Figura 4** - Contabilização das transcritos codificadores por categorias das ontologias gênicas (GO) no transcrito de *A. sexdens*. Os dados foram obtidos a partir da anotação funcional das sequências de peptídeos utilizando a ferramenta eggNOG.....49
- Figura 5** - Análise filogenética das proteínas envolvidas na degradação de quitina em *Atta sexdens*. A análise foi realizada utilizando o software IQTREE2, empregando o método de máxima verossimilhança, com bootstrap de 1000 replicatas. As sequências utilizadas como referência foram obtidas das espécies *Drosophila melanogaster*, *Tribolium castaneum*, *Bombyx mori*, *Apis mellifera* e *Agrotis ipsilon* (Tabela 1). Os motivos presentes em cada sequência foram preditos usando a ferramenta MemeTool. As sequências putativas em *Atta sexdens* foram preditas a partir da ferramenta TransDecoder..... 50
- Figura 6** - Análise de motivos conservados comumente presentes nos domínios catalíticos (GH18). Os transcritos codificadores de quitinases em *Atta sexdens* foram alinhados usando o programa MAFFT. .... 52
- Figura 7** - Identificação dos domínios presentes nos candidatos a transcritos codificadores de quitinases em *Atta sexdens*. ....53
- Figura 8** - Identificação de domínios utilizando o programa SMART a partir da sequência consenso (As35323+As35324) candidata a quitinase Cht5 em *Atta sexdens*. ..... 54
- Figura 9** - Análise de identificação de domínios utilizando o programa SMART na sequência consenso (As22589+As22591) candidata à quitinase Cht2 em *Atta sexdens*. .....56

## SUMÁRIO

1	INTRODUÇÃO .....	12
2	REFERENCIAL TEÓRICO .....	15
2.1.	Análise genômica e funcional.....	15
2.1.1.	Sequenciamento de RNA via Plataforma Illumina .....	16
2.1.2.	Bioinformática aplicada ao estudo de transcriptomas .....	19
2.1.3.	Anotação gênica funcional.....	20
2.1.4.	Filogenia .....	22
2.2.	Origem e bioecologia das formigas cortadeiras. ....	24
2.2.1.	Estrutura da colônia de formigas cortadeiras .....	26
2.2.2.	Análise genômica e funcional de formigas cortadeiras.....	30
2.3.	Quitina em insetos e o papel das quitinases .....	31
2.3.1.	Identificação de quitinases em formigas cortadeiras .....	36
3	METODOLOGIA.....	40
3.1.	Amostra biológica e extração de RNA .....	Erro! Indicador não definido.
3.2.	Construção das bibliotecas de cDNA e sequenciamento via Illumina HiSeq 2500 .....	40
3.3.	Montagem <i>de novo</i> do transcrito.....	41
3.4.	Anotação funcional dos transcritos codificadores .....	42
3.5.	Identificação <i>in silico</i> de transcritos codificadores de quitinases no transcrito da formiga cortadeira <i>A. sexdens</i> .....	42
4	RESULTADOS E DISCUSSÃO.....	46
4.1.	Montagem e anotação do transcrito da formiga cortadeira <i>A. sexdens</i> ....	46
4.2.	Identificação <i>in silico</i> de transcritos codificadores de quitinases no transcrito de <i>A. sexdens</i> .....	49
5	CONCLUSÃO.....	59
	REFERÊNCIAS.....	60
	APÊNDICE A - VALORES DE INTEGRIDADE DE RNA .....	65

## 1 INTRODUÇÃO

As formigas cortadeiras (ordem *Hymenoptera*, família *Formicidae*, subfamília *Myrmicinae*, tribo *Attini*, subtribo *Attina*) são insetos sociais que desempenham um importante papel ecológico nos ecossistemas em que habitam (HÖLLDOBLER; WILSON, 1990; LOFGREN, C.S., & VANDER MEER, 1986). A sociedade das formigas cortadeiras apresenta castas bem definidas, e cada uma desempenha um papel importante dentro da colônia. Nessa organização, os machos emergem somente no período do acasalamento, quando também emergem as jovens rainhas. As rainhas, além da cópula e ovoposição, são necessárias para o fundação e manutenção da colônia. Outra casta muito importante é a das operárias, as quais são divididas em subcastas: jardineiras, generalistas, forrageadoras e as soldadas. As operárias são as responsáveis por executar atividades essenciais para a coesão do formigueiro tais como o cuidado da prole, limpeza, forrageio de material vegetal e proteção da colônia (DELLA LUCIA, 2011; SÉRGIO DE BRITTO *et al.*, 2016).

Estes insetos apresentam comportamentos característicos, como o cultivo do fungo, *Leucoagaricus gongylophorus* (divisão *Basidiomycota*, ordem *Agaricales*, família *Agaricaceae*) do qual se alimentam (SCHULTZ, 2021). Na verdade, esses dois organismos vivem em um sistema de simbiose obrigatória, onde a formiga possui diversas adaptações comportamentais e fisiológicas que viabilizam o crescimento e o sucesso ecológico do fungo no interior de seus ninhos (WORSLEY *et al.*, 2018). A vantagem desta relação mutualística para o fungo é a obtenção dos nutrientes necessários à sua sobrevivência, além de se beneficiar da assepsia feita pelas formigas a fim de evitar a contaminação do jardim fúngico. Em contrapartida, as formigas alimentam-se de hifas especializadas produzidas pelo fungo, os chamados gongilídeos. As paredes celulares desta estrutura são ricas em proteínas, sendo usada para alimentar as larvas, e o conteúdo líquido, rico em carboidratos, serve como alimento para as adultas (DELLA LUCIA, 2011; SÉRGIO DE BRITTO *et al.*, 2016).

Essa relação ecológica muito bem estabelecida tem um custo para os sistemas de produção agrícolas. As formigas cortadeiras, oriundas do neotrópico, são as maiores desfolhadoras conhecidas na natureza. A sociedade destas formigas é muito grande, sendo que as do gênero *Atta* podem ter milhões de indivíduos. Esta característica faz com que elas coletem grandes quantidades de material vegetal para o cultivo do fungo. O problema é que este comportamento causa danos à diversas culturas, podendo levar a perdas que variam de 10 a 100% dependendo do estágio de desenvolvimento das plantas. Dessa forma, estas formigas são

consideradas importantes pragas agrícolas na América Central e América do Sul (DELLA LUCIA, 2011; SCHULTZ, 2021; SÉRGIO DE BRITTO *et al.*, 2016).

Nesse estudo, a espécie alvo é a formiga cortadeira *Atta sexdens* (ordem *Hymenoptera*, família *Formicidae*), com ampla distribuição no Brasil (<https://antmaps.org/?mode=species&species=Atta.sexdens>) e popularmente conhecida como saúva-limão (SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014). Como demonstrado anteriormente, as formigas cortadeiras são importantes alvos tanto do ponto de vista ecológico quanto do controle de pragas, os quais poderiam beneficiar-se de dados moleculares da espécie. Existem muitos dados disponíveis sobre a bioecologia das formigas cortadeiras, mas quando se trata de dados genômicos e transcritômicos ocorre uma escassez de informação. A formiga *A. sexdens*, por exemplo, não contém genoma de referência publicado, o que dificulta a compreensão evolutiva e genética da espécie. Neste contexto, pouca informação está disponível sobre quais são os genes e proteínas diretamente associados com as interações com o fungo e com o crescimento e desenvolvimento deste inseto. Portanto, considerando a importância da espécie e a escassez de informação sobre as formigas, há uma crescente demanda por esforços da comunidade científica e do setor privado na produção de dados biológicos, bioquímicos e moleculares para suportar a geração de conhecimento sobre as formigas cortadeiras.

Através de dados transcritômicos é possível prospectar genes essenciais ao desenvolvimento dos insetos em condições específicas, como em um determinado estágio do desenvolvimento, em tecidos e/ou órgãos e até mesmo em resposta a fatores externos sejam eles bióticos ou abióticos. Isso é possível porque com o surgimento do sequenciamento de ácidos nucleicos surgiram, por conseguinte, bancos de dados de sequências biológicas. Esses bancos de dados são acessados por biólogos que procuram determinar a função biológica das sequências depositadas em organismos modelos. Assim, quando uma sequência biológica tem sua função validada experimentalmente, ela pode ser usada como referência para anotação de sequências similares em outras espécies. A partir disso, mesmo na ausência de um genoma de referência, é possível, por meio do sequenciamento do ácido ribonucleico (RNA-Seq), empregar diversas técnicas de bioinformática atuais para a obtenção do transcrito de uma espécie, a fim de identificar potenciais candidatos de uma família gênica que se deseja estudar (BONO *et al.*, 2022; OPPENHEIM *et al.*, 2015).

Dentre os genes que podem ser estudados em insetos devido à sua importância morfológica, são aqueles envolvidos na rota do metabolismo de quitina - um polissacarídeo constituinte de matrizes extracelulares - desempenham importante função como proteção a

fatores bióticos, abióticos e mecânicos. A quitina é principalmente conhecida por ser o principal componente do exoesqueleto dos insetos, mas também está presente em órgãos como o intestino e as asas, afetando tanto a digestão quanto a locomoção. Dentre as enzimas que participam no metabolismo de quitina, as quitinases (E.C. 3.2.1.14) possuem importante função, pois são requeridas nas fases de mudança de estágio do desenvolvimento dos insetos, quando é necessário que a cutícula velha seja degradada e uma nova seja sintetizada a fim de viabilizar a metamorfose e/ou crescimento do inseto (LIU; ZHANG; ZHU, 2019). Além disso, as quitinases participam nos processos de digestão e fluxo de nutrientes das colônias das formigas cortadeiras, pois foram detectadas tanto nos jardins de fungos como no sistema digestório destes insetos (DELLA LUCIA, 2011).

Isto posto, as quitinases são potenciais alvos de estudos para compreender o desenvolvimento das formigas cortadeiras. Neste contexto, objetivou-se empregar ferramentas de bioinformática atualizadas para montagem de transcritos, predição de transcritos codificadores de proteínas baseadas em dados de similaridade e anotação de domínios proteicos, e identificar potenciais transcritos codificadores de quitinases no transcrito da formiga cortadeira *A. sexdens* a partir do sequenciamento de RNA de cinco amostras biológicas: larva inteira, intestino de larva, formiga forrageadora inteira, cabeça e intestino de formiga forrageadora; tendo estes dados sido gerados primeiramente por Máximo (2018).

## 2 REFERENCIAL TEÓRICO

### 2.1. Análise genômica e funcional

O ácido desoxirribonucleico (DNA) é uma molécula de dupla fita carreadora da informação genética dos seres vivos, cuja estrutura foi elucidada em meados de 1950 por James Watson e Francis Crick. O DNA é um polímero de nucleotídeos compostos por uma unidade de açúcar (desoxirribose), uma base nitrogenada - adenina (A), timina (T), guanina (G) ou citosina (C) - e um grupo fosfato. Devido a interação por pontes de hidrogênio entre as bases (A-T e C-G), as moléculas de DNA são estáveis como hélices dupla fita e antiparalelas (GAUTHIER *et al.*, 2019; VERLI, 2014).

O fluxo da informação biológica, conhecido como dogma central da biologia molecular, estabelece que a informação contida no DNA deve ser transcrita para uma molécula de ácido ribonucleico (RNA) podendo este ser ou não traduzido em uma proteína. A transcrição gênica por sua vez depende da necessidade da célula em determinado momento e condição. O RNA é composto por uma unidade de açúcar (ribose), uma base nitrogenada – A, C, G e Uracila (U) - e um grupo fosfato. A molécula de RNA é a intermediária no fluxo da informação genética entre DNA e proteína. Com base nas suas propriedades físico-químicas e em sua estrutura de fita simples, o RNA é menos estável que o DNA, o que o torna mais suscetível à degradação. A tradução da informação da molécula de RNA a proteína (polímero de aminoácidos) é crucial para o processo biológico, devido à diversidade de funções que as proteínas, efetoras da informação gênica, desempenham na célula (GAUTHIER *et al.*, 2019; VERLI, 2014).

Desde a elucidação da estrutura do DNA diversos avanços ocorreram no campo da biologia molecular. Com o surgimento de técnicas de sequenciamento de ácidos nucleicos, os cientistas passaram a ter acesso a infinidades de sequências biológicas: genomas (DNA nuclear), mitogenomas (DNA mitocondrial), genomas cloroplastidiais (DNA cloroplastidial), transcritomas (conjunto genes expressos em RNA), entre outros, como sequenciamento de elementos cis-regulatórios. Atualmente estas sequências estão amplamente difundidas em bancos de dados ao redor do mundo e disponíveis para análises de dados biológicos. Margareth Dayhoff foi a pioneira no uso de computadores para estudos de moléculas, desde a sistematização da nomenclatura dos aminoácidos e a trabalhos filogenéticos até os estudos da estrutura tridimensional de proteínas por métodos de cristalografia de raios-X. Desde então, vivenciamos uma revolução científica acompanhada pelo alto poder computacional e pelo

grande volume de dados provenientes dos projetos genoma (GAUTHIER *et al.*, 2019; VERLI, 2014).

### 2.1.1. Sequenciamento de RNA via Plataforma Illumina

Desde a publicação do primeiro genoma humano, diversas tecnologias foram lançadas no mercado, perfazendo 3 gerações de sequenciadores. O sequenciamento de primeira geração, também conhecido como sequenciamento didesóxi, método de terminação de cadeia ou sequenciamento de Sanger, baseia-se no processo de amplificação de uma sequência de DNA empregando uma DNA polimerase em presença de nucleotídeos modificados, os chamados didesoxirribonucleosídeos trifosfato (ddNTP - nucleotídeos terminadores de cadeia) que não apresentam a hidroxila no carbono 3 da desoxirribose. O efeito dos ddNTPs está envolvido com o bloqueio da amplificação da molécula após sua incorporação na fita de DNA gerando fragmentos de DNA de tamanhos diferentes. Atualmente, o sequenciamento de Sanger baseia-se na utilização de um sistema automatizado, onde cada ddNTP (ddATP, ddTTP, ddCTP, ddGTP) é marcado com fluoróforos específicos, e o produto do sequenciamento submetido à eletroforese capilar para identificação dos nucleotídeos da sequência alvo por ordem de tamanho, sendo que os fragmentos sequenciados apresentam em média 600-1000 bases (HU *et al.*, 2021).

O sequenciamento de segunda ou de próxima geração (NGS) consiste em um método de sequenciamento massivo de sequências curtas (100-600 pares de base - pb) de DNA por meio de síntese química, de modo que cada corrida gera milhões de cópias de um fragmento. O sequenciamento NGS revolucionou o mercado porque barateou o sequenciamento, além de gerar grande volume de dados, facilitando o sequenciamento de genomas e transcritomas a um menor tempo e menor custo quando comparado ao sequenciamento de primeira geração (HU *et al.*, 2021)..

Por fim, o sequenciamento de terceira geração também se baseia no sequenciamento massivo de moléculas DNA ou RNA, porém este método gera poucas sequências se comparado ao NGS e os fragmentos gerados são maiores que 1500 bases. Tecnologias como Pacbio (do inglês - *Pacific Biosciences*) e ONT (do inglês - *Oxford Nanopore Technology*) são exemplos de tecnologias de sequenciamento de terceira geração que estão revolucionando o mercado devido à possibilidade de obtenção de leituras longas que cobrem grandes porções do genoma. No caso dos transcritomas, estas tecnologias têm possibilitado o sequenciamento de isoformas

completas, o que é sua grande vantagem quando comparadas aos métodos de segunda geração (HU *et al.*, 2021).

Para o estudo de transcritomas é necessário o sequenciamento de moléculas de RNA. O Sequenciamento de RNA (RNA-Seq) é uma tecnologia usada para determinar a identidade e abundância de sequências de RNA (transcritos) em uma amostra biológica, uma vez que moléculas de RNA são as efetoras intermediárias da informação biológica e essenciais para todas as células. O sequenciamento da informação biológica existente em um dado momento em uma determinada amostra biológica tem uma série de aplicações que envolvem a descoberta de novos transcritos, processamento alternativo de transcritos, ontologia gênica e polimorfismo gênico (SNP). Porém, uma das grandes contribuições do RNA-Seq é a possibilidade de quantificar a expressão gênica e compará-la entre bibliotecas. Ou seja, verificar o efeito de uma condição específica (temperatura, estágio do desenvolvimento, exposição a agentes biológicos, diferentes tecidos) sobre a expressão gênica em determinada amostra biológica (WANG; GERSTEIN; SNYDER, 2009; HAAS; ZODY, 2010; MCGETTIGAN, 2013; HRDLICKOVA; TOLOUE; TIAN, 2017).

Os primeiros estudos de sequências de RNA em larga escala datam da década de 90, por meio das técnicas de Sequências Expressas Marcadas (EST, do inglês – Expressed Sequence Tags), Análise Serial da Expressão Gênica (SAGE, do inglês – Serial Analysis of Gene Expression) e Microarranjos de DNA (DNA Microarrays). Todas estas análises demandam um conhecimento prévio do genoma/transcritoma, porém, não permitem a obtenção massiva de dados. Neste contexto, o sequenciamento de 2ª geração revolucionou esse campo de estudo, tanto pela obtenção massiva de sequências quanto pela possibilidade de abordagens *de novo*, as quais não requerem dados genômicos e/ou transcritômicos de referência previamente disponíveis. As etapas essenciais desta metodologia envolvem o isolamento do RNA da amostra biológica (células, tecidos, organismos), em alguns casos conversão em DNA complementar (cDNA), preparação das bibliotecas, sequenciamento dos nucleotídeos das bibliotecas e análises dos dados biológicos (bioinformática) (HRDLICKOVA; TOLOUE; TIAN, 2017).

O sequenciamento de RNA por meio da plataforma Illumina baseia-se na obtenção da molécula de cDNA (DNA complementar obtido por de uma reação de síntese de DNA catalisada por uma enzima DNA polimerase dependente de RNA). O cDNA é sintetizado após a extração do RNA da amostra biológica e nesta metodologia a construção da biblioteca de cDNA é uma etapa crucial. No entanto, existem outras considerações importantes antes da construção da biblioteca: definir quais as moléculas de RNA são interessantes para o estudo;

qual o tamanho da biblioteca de cDNA de interesse; qual o método de ligação de adaptadores utilizar (CONESA *et al.*, 2016; HRDLICKOVA; TOLOUE; TIAN, 2017; PARK *et al.*, 2019).

Organismos eucariotos contêm RNAs codificadores de proteínas (mRNA) e RNAs longos não codificadores (lncRNA – maiores que 200 nt) contendo cauda poli-A. As moléculas de mRNA representam apenas 1-5% do conteúdo de RNA total na célula. Logo, existem duas formas básicas de seleção destas moléculas: transcrição reversa utilizando iniciadores oligo-dT, ou seleção via esferas ligadas a iniciadores oligo-dT (CONESA *et al.*, 2016; HRDLICKOVA; TOLOUE; TIAN, 2017; PARK *et al.*, 2019).

Como em muitos dos estudos o RNA ribossomal (rRNA) não é interessante, este é removido por ser o mais abundante na célula. Existem diversas metodologias disponíveis, desde a utilização de sondas que se hibridizam no rRNA seguido por digestão via RNase H, até a utilização de iniciadores não randômicos durante a etapa de transcrição reversa o que evita rRNAs. Apesar da grande diversidade de métodos de enriquecimento das sequências de RNA de interesse e depleção de rRNA, na maior parte dos estudos prefere-se a utilização de métodos baseados na purificação da cauda poli-a devido à facilidade do protocolo e custo comparado às outras metodologias (CONESA *et al.*, 2016; HRDLICKOVA; TOLOUE; TIAN, 2017; PARK *et al.*, 2019).

Após a seleção das moléculas de RNA de interesse, as bibliotecas precisam passar por um processo de fragmentação devido à limitação do método de sequenciamento empregado, que no caso da 2ª geração prevê fragmentos menores que 600pb. Os métodos que podem ser empregados no processo de fragmentação são: digestão alcalina, através do uso de cátions divalentes ( $Mg^{2+}$ ,  $Zn^{2+}$ ); ou enzimas (RNase III – endoribonucleases de moléculas dupla fita RNA). Todas as metodologias podem gerar fragmentos não randômicos, no entanto, como geralmente incorpora-se a etapa de síntese de cDNA na obtenção das bibliotecas e a enzima RNase III tem preferência por moléculas dupla fita, todas as moléculas de cDNA podem ser fragmentadas (CONESA *et al.*, 2016; HRDLICKOVA; TOLOUE; TIAN, 2017; PARK *et al.*, 2019).

Após a fragmentação, todas as moléculas de cDNA são ligadas a adaptadores antes do processo de amplificação e sequenciamento. Para evitar que se perca a informação de direcionalidade da molécula de RNA, diferentes metodologias foram desenvolvidas para se manter esta informação. Apesar da diversidade de métodos disponíveis, geralmente emprega-se a ligação de adaptadores diferentes a cada terminação (5' e 3') pelo uso de RNA ligase tipo I (3') e RNA ligase tipo II (5'). Outra possibilidade é a depleção de uma das fitas através da

incorporação de moléculas dUTP na fita complementar e posterior digestão desta fita pela utilização de uma enzima chamada Uracil-DNA Glycosylase (enzima que cliva a base uracil na fita complementar) (CONESA *et al.*, 2016; HRDLICKOVA; TOLOUE; TIAN, 2017; PARK *et al.*, 2019).

Após as etapas de seleção, fragmentação e ligação de adaptadores às moléculas de interesse, estas precisam passar por uma etapa de amplificação seguida pelo sequenciamento. Em posse das leituras geradas pelo sequenciador, a etapa seguinte consiste na montagem das sequências e interpretação dos dados biológicos.

### **2.1.2. Bioinformática aplicada ao estudo de transcritomas**

A bioinformática pode ser dividida em dois campos de estudo: clássica e estrutural. A bioinformática clássica compreende a análise de dados de sequências de nucleotídeos e aminoácidos. E os estudos de sequências de biomoléculas incluem: comparação por meio de alinhamentos; identificação de padrões (domínios, motivos estruturais); caracterização das relações evolutivas; construção e anotação de genomas, transcritomas, proteomas (conjunto de proteínas de um organismo); e, construção de redes de expressão (ou seja, como a expressão de um gene A está correlacionada à expressão de um gene B). Por outro lado, a bioinformática estrutural compreende a análise dos dados biológicos a partir de uma perspectiva tridimensional, demandando maior poderio computacional em relação à bioinformática clássica. Alguns dos principais estudos da bioinformática estrutural envolvem a modelagem da estrutura molecular de biomoléculas, modelagem da interação entre biomoléculas e caracterização da flexibilidade molecular (dinâmica molecular) (GAUTHIER *et al.*, 2019; VERLI, 2014).

A bioinformática clássica evoluiu a partir do escalonamento de projetos de sequenciamento e estes têm revolucionado tudo o que sabemos sobre a origem, evolução e biologia das espécies. No entanto, embora continuemos a desenvolver novas tecnologias de sequenciamento, e mais genomas sejam sequenciados, ainda há um gargalo no conhecimento disponível sobre a função de genes. O sequenciamento em si não representa o fim, mas apenas o meio para auxiliar biólogos a responderem suas perguntas biológicas. E muitas das perguntas ainda não respondidas são sobre a anotação gênica, processo onde atribui-se a função das sequências descobertas (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Após o sequenciamento, os dados brutos devem ser convertidos em informação que dará respaldo para o desenvolvimento de soluções e novas perguntas. Portanto, a interpretação

biológica e a anotação gênica são as etapas mais complexas, e por isso, existem diversos programas e fluxos de análises que ajudam a interpretar os dados obtidos após o sequenciamento (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Após obter as leituras provenientes do sequenciamento, o processo de análise de dados envolve as seguintes etapas: (i) pré-processamento das leituras por meio da remoção de leituras de baixa qualidade e de adaptadores; (ii) obtenção do transcrito de referência por montagem *De novo*, quando da ausência de um genoma de referência; (iii) alinhamento ao genoma ou transcrito de referência; (iv) anotação gênica funcional (análises de similaridade, busca por domínios conservados, domínios intermembranares, peptídeos sinal); (v) análise de expressão gênica diferencial; (vi) análise de coexpressão; (vii) análises de agrupamentos; e, (viii) análises filogenéticas (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Após o processamento das leituras e análises do controle de qualidade, inicia-se a montagem dos transcritos das bibliotecas usando ferramentas de bioinformática que fusionam sequências que se sobrepõem para gerar contigs (GRABHERR *et al.*, 2011). Além de existirem programas específicos para montagem *de novo* dos transcritos da amostra biológica de interesse, como o Trinity, existem programas específicos cujo propósito é reduzir a redundância da montagem, através de análise de similaridade e agrupamento de sequências por meio de um limiar pré-determinado. Esses transcritos que são agrupados são chamados de unigenes. Essa etapa é necessária porque é gerado um número tão grande de transcritos na etapa de montagem que isso pode dificultar as análises posteriores (FU *et al.*, 2012; LI Ñ; GODZIK, 2006;). Esse unigenes serão então utilizados nas análises posteriores para anotação da função gênica e predição de ORFs (do inglês - Open Reading Frames), que são sequências de nucleotídeos delimitadas por um códon de início e um códon de parada da tradução do mRNA em proteína (VERLI, 2014).

### **2.1.3. Anotação gênica funcional**

As etapas básicas seguidas em um pipeline de anotação gênica são:

1. busca por genes similares em bancos de dados como o NCBI (do inglês - *National Center for Biotechnology Information*) ou UniprotKB/Swiss-Prot e de espécies próximas;
2. predição de ORFs candidatas a transcritos codificadores nos transcritos;

3. e a partir destas ORFs buscas por domínios conservados, domínios intermembranares, peptídeos sinal e análises de enriquecimento (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

A partir dos transcritos, e a depender da biblioteca, também é possível anotar RNA não codificante (ncRNA), determinar se os transcritos são expressos constitutivamente ou tecido-específico, e, ainda fazer análises de expressão gênica diferencial entre bibliotecas. Muitas vezes são incorporadas ao pipeline de anotação gênica as análises filogenéticas com o objetivo de estudar proteínas ou famílias proteicas específicas, para determinar o melhor candidato para validação da função gênica ou comparar genes ortólogos entre si, ou seja, genes homólogos, que possuem um ancestral comum, mas acabaram divergindo por especiação (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Uma vez que um genoma é publicado gera-se um banco de dados de sequências de nucleotídeos e proteínas utilizáveis por pesquisadores a fim de comparar o genoma do organismo em estudo com aquele publicado (referência), podendo este ser da mesma espécie ou de uma espécie próxima. Essa comparação de sequências pode ser feita por meio do alinhamento. Em suma, o processo de alinhamento emprega cálculos que visam encontrar a melhor posição de correspondência entre os caracteres em diferentes sequências (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Alinhamentos resultam em uma estimativa de similaridade entre sequências. A análises de similaridade podem ser empregadas para inferir a estrutura e função de sequências próximas assim como o relacionamento evolutivo entre elas. De fato, a análise de similaridade pode indicar que exista alta identidade entre sequências, mas não determina que elas são de fatos homólogas. A identidade refere-se a porcentagem da sequência de referência que é idêntica à sequência que se deseja anotar. Por outro lado, homologia não requer que as sequências sejam idênticas, pois estas podem acumular diferentes taxas de evolução ao longo do tempo (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Sequências homólogas são aquelas que compartilharam em algum momento no passado, um ancestral comum, portanto, elas são evolutivamente relacionadas ((KAPLI; YANG; TELFORD, 2020). No entanto, podem ocorrer casos em que a pressão do ambiente pode fazer com que sequências convirjam evolutivamente tornando-se altamente relacionadas, mas não são homólogos (análogos) (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Portanto, a análise de similaridade é importante em estudos de anotação gênica, mas não conclusiva. Dentre, as ferramentas mais utilizadas atualmente em análises de similaridade, a

mais conhecida é o programa BLAST (do inglês - Basic Local Alignment Search Tool). Por meio desta ferramenta é possível comparar sequências de nucleotídeos contra nucleotídeos, nucleotídeos contra proteínas, e vice-versa, assim como proteínas contra proteínas. Para cada alinhamento gerado de uma sequência contra um banco de dados é gerado um valor estatístico (e-value) que indica a chance de aquele alinhamento ter ocorrido ao acaso, quanto mais próximo de zero, mais significativo é o alinhamento. O BLAST é um algoritmo de busca rápida, isso porque para alinhar sequências busca por palavras de tamanho fixo, i.g., 3 para aminoácidos e 11 para nucleotídeos. Assim, a busca por sequências similares torna-se mais objetiva, eficiente e significativa (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Outra análise empregada na identificação da função de um transcrito codificador, são as análises de busca por domínios proteicos, os quais representam partes da cadeia polipeptídica que se enovela independentemente formando uma estrutura compacta e estável (VERLI, 2014). A cada domínio é associado uma função seja ela bioquímica ou estrutural. Nas análises de bioinformática, os domínios são representados por sequências conservadas, que podem ser chamadas também de motivos estruturais, sendo que um domínio pode apresentar mais de um motivo estrutural. A identificação dos produtos gênicos com base em bancos de dados de proteínas curados manualmente (UniprotKB/Swiss-Prot) e identificação de domínios proteicos (Pfam) gera grupos distintos de genes anotados: ortólogos diretos; proteína contendo domínio (ortólogo desconhecido, domínio conhecido), proteína hipotética (ortólogo e domínio desconhecidos); e, proteína hipotética conservada (ortólogo direto, domínio desconhecido) (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

Uma vez que uma proteína tenha sido predita, torna-se possível prever a localização subcelular utilizando ferramentas específicas, *i.e.*, proteínas com regiões hidrofóbicas, podem ser classificadas como proteínas de membrana. Enquanto outras, que são endereçadas a organelas específicas podem apresentar peptídeos sinal (BONO *et al.*, 2022; CONESA *et al.*, 2016; VERLI, 2014).

#### **2.1.4. Filogenia**

Por meio da sistematização do conhecimento biológico das espécies foi possível a classificação dos seres vivos em grupos. Após a publicação de Darwin “A origem das espécies”, em 1859, incorporou-se à classificação os dados de ancestralidade. Esses dados, quando representados graficamente geram as árvores filogenéticas. As primeiras árvores filogenéticas

foram construídas baseadas em dados fósseis, e características morfológicas e fenotípicas, mas com o advento da biologia molecular e do sequenciamento, passou a ser possível comparar sequências nucleicas e inferir a distância evolutiva entre elas. Sendo que, inicialmente as filogenias moleculares eram dependentes de dados gerados via PCR (do inglês – Polimerase Chain Reaction). Foi somente a partir do sequenciamento de genomas e transcritomas que as análises filogenéticas passaram a utilizar conjunto de dados de sequências muito maiores (KAPLI; YANG; TELFORD, 2020; VERLI, 2014).

A inferência filogenética tem expandido seu campo de atuação desde então, sendo empregada tanto em análises de relações evolutivas, como em estudos de prospecção de genes e famílias gênicas em espécies relacionadas. A análise filogenética é obtida após cálculos de distância genética entre as sequências, e essa medida representa a dissimilaridade entre as unidades taxonômicas (sequências comparadas), sendo obtidas a partir de diferentes modelos matemáticos. Por meio de dados de distribuição geográfica, a filogeografia estabelece a relação evolutiva entre espécies a partir de sua distribuição geográfica, sendo possível incorporar também dados moleculares. A partir de sequências de famílias gênicas, genomas completos e mitogenomas é possível também inferir a filogenia das espécies (KAPLI; YANG; TELFORD, 2020; VERLI, 2014).

A primeira etapa na reconstrução da relação filogenética entre sequências homólogas pode ser conduzida por meio de um alinhamento local, como o BLAST, a fim de se encontrar sequências similares àquela que se deseja identificar. A análise de similaridade pode retornar sequências que não contém as características esperadas, como a presença de um domínio específico, logo, é comum o emprego de métodos de seleção das sequências candidatas. Posteriormente, as sequências similares selecionadas e as utilizadas como referência são submetidas a um método de alinhamento global, de modo que se possa obter as distâncias entre todas as sequências. Como a qualidade do alinhamento depende da taxa de divergência entre as sequências, faz-se necessário empregar métodos de filtragem de regiões com um alinhamento de baixa qualidade. Outrora, essas “limpezas” eram feitas manualmente, mas hoje existem softwares próprios que reduzem o viés do cientista (KAPLI; YANG; TELFORD, 2020; VERLI, 2014).

Posteriormente o alinhamento é submetido a um método de estimação da filogenia, podendo este ser qualitativo ou quantitativo. Métodos quantitativos baseiam em medidas de distância e geram apenas uma filogenia, sendo o mais conhecido o método NJ (do inglês – Neighbour Joining). Dentre todos é o que computa a filogenia mais rápido, no entanto, perde

eficiência quando existe alta divergência entre as sequências. Já os métodos qualitativos empregam parâmetros de otimização, e segundo um critério estabelecido são geradas diversas árvores filogenéticas, ao final, escolhe-se a filogenia que melhor explica o alinhamento, ou seja, a que melhor representa a relação entre as sequências. Algoritmos qualitativos conhecidos incluem, método da máxima parcimônia, método da máxima verossimilhança e o método de inferência bayesiana. Embora mais eficientes em computar distâncias filogenéticas entre sequências mais divergentes, estes métodos demandam mais recursos computacionais e mais tempo para processar as análises (KAPLI; YANG; TELFORD, 2020; VERLI, 2014). Entretanto, tem surgido algoritmos mais eficientes e rápidos na implementação destes métodos (MINH *et al.*, 2020).

O principal objetivo da análise filogenética é identificar o relacionamento entre as sequências analisadas, de modo que esta represente verdadeiramente as reais distâncias entre elas. Para atingir alta confiabilidade são empregadas algumas análises estatísticas: bootstrap (método NJ, máxima parcimônia e máxima verossimilhança) e distribuição posterior (método de inferência bayesiana). O método bootstrap consiste em amostragens a partir do mesmo conjunto de dados, onde são gerados novos alinhamentos e conseqüente uma nova filogenia. A confiabilidade de um clado representará, portanto, o número de vezes que ele se repetiu, sendo o número de amostragens definida pelo cientista (geralmente 1000). Por outro lado, no método de inferência bayesiana a geração da estatística é inerente ao algoritmo utilizado e a confiabilidade é representada pela probabilidade de formação de cada clado (KAPLI; YANG; TELFORD, 2020; VERLI, 2014).

## **2.2. Origem e bioecologia das formigas cortadeiras**

As formigas (ordem *Hymenoptera*, família *Formicidae*) apareceram durante o período Cretáceo aproximadamente há 135 milhões de anos, e são divididas atualmente em 23 subfamílias contendo 24000 espécies (HÖLLDOBLER; WILSON, 1990). Em relação ao seu desenvolvimento, as formigas são insetos holometábolos, ou seja, passam por uma metamorfose completa. Até atingir a fase adulta, esses insetos precisam passar pelas fases de ovo, larva, pré-pupa, pupa e adultos. Dentre os diversos grupos de formigas destacam-se as formigas cultivadoras de fungos (subfamília *Myrmicinae*, tribo Attini, subtribo Attina). Diferente das demais que são caçadoras, as formigas da tribo Attini evoluíram para obter seu

alimento a partir de fungos cultivados por elas (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986).

As formigas cultivadoras de fungo pertencem a um grupo de 245 espécies e ao todo 19 gêneros. Divergiram evolutivamente há aproximadamente 60 milhões de anos de seu ancestral comum e são espécies características do Neotrópico. As Attini distribuem-se do norte da Argentina ao sul dos Estados Unidos, ocorrendo em diversos biomas, desde a floresta amazônica até ambientes extremos como o deserto (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986). Os gêneros dominantes das Attini são: *Acromyrmex* e *Atta*, popularmente conhecidas como formigas cortadeiras, ou Attini superiores. Estes gêneros originaram-se somente há 16 milhões de anos adotando o uso de material vegetal fresco como substrato para o cultivo do fungo, do qual se alimentam, o que representou uma vantagem evolutiva às espécies destes gêneros, já que as Attinis inferiores ao utilizarem restos vegetais e animais ficam limitadas na obtenção de fontes de proteínas (SCHULTZ, 2021).

As formigas cortadeiras são uns dos insetos-praga mais importantes da América Central e América do Sul. Como resultado do uso de substrato fresco e o grande montante de vegetação retirada de florestas tropicais, as formigas cortadeiras são consideradas invertebrados dominantes e herbívoros prevalentes na maioria dos ecossistemas neotropicais (SÉRGIO DE BRITTO *et al.*, 2016). Diversas espécies agrícolas são suscetíveis ao ataque das formigas cortadeiras - citros, cacau, café, milho, algodão e, em particular, florestas de eucalipto e pinus, podendo atacar também reservas de cereais, farinhas e pastos (SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014). Os danos causados pelo forrageamento destas formigas podem chegar à cifra de milhões de dólares. Além do mais, quando atacam culturas jovens, os danos podem chegar a 100%. Após defoliação intensa podem provocar perda total da produção de madeira, redução do diâmetro e altura das árvores, sendo que, após desfolhamento completo de plantas de eucalipto e pinus as taxas de mortalidade podem chegar a 10 e 25%, respectivamente (SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014)

Embora existam diversas técnicas empregadas no manejo integrado das formigas cortadeiras, uma das formas mais eficientes no controle destes insetos são as iscas formicidas a base de sulfluramida. Este é o único método que apesar de agressivo ao meio ambiente e tóxico aos seres humanos e animais não-alvos, oferece vantagens econômicas, operacionais e técnicas que justificam seu uso (SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014).

Notadamente, as Attini cultivam diversas espécies de fungos da tribo *Leucocoprineae* (ordem *Agaricales*, família *Agaricaceae*), a qual abrange os gêneros *Leucoprinus*,

*Leucoagaricus* e *Lepiota*. Enquanto a maioria das Attini estabeleceu uma simbiose facultativa, as formigas cortadeiras desenvolveram uma relação simbiótica obrigatória com o fungo *Leucoagaricus gongylophorus* (*Lepiotacea: Agaricales: Basidiomycota*) (WORSLEY *et al.*, 2018). Ao cultivarem o fungo, as formigas visam à produção de hifas com as extremidades infladas chamadas de gongilídeos, as quais são utilizadas como fonte de nutrientes (lipídeos e carboidratos) e de enzimas essenciais para o crescimento dos jardins de fungo (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986; SCHULTZ, 2021).

A dominar técnicas de cultivo fúngico, as formigas cortadeiras beneficiam-se ao obter alimento processado livre de terpenoides, alcaloides e outros metabólitos secundários que servem como defesa das plantas a insetos herbívoros. O fungo, por sua vez, beneficia-se de uma espécie altamente organizada que provê seu sustento, sua proteção e gerencia o local ocupado por ambos (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986; SCHULTZ, 2021; SÉRGIO DE BRITTO *et al.*, 2016; WORSLEY *et al.*, 2018)

### **2.2.1. Estrutura da colônia de formigas cortadeiras**

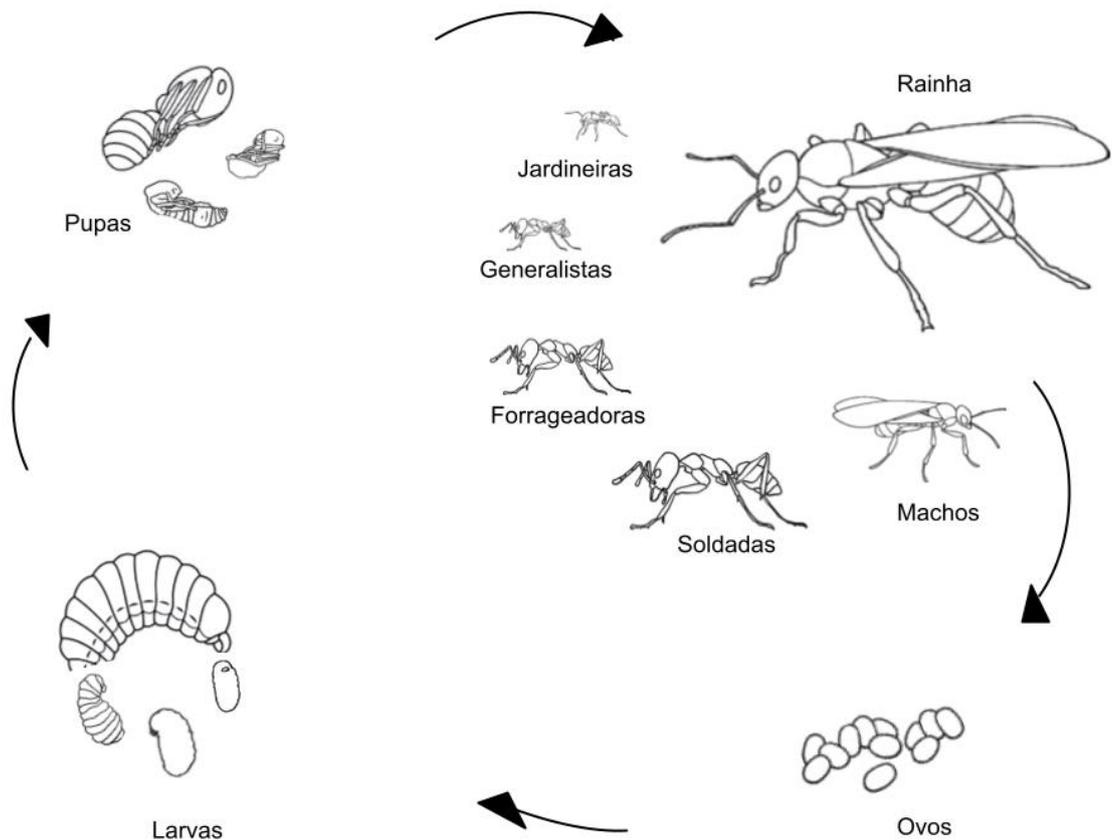
As rainhas acasalam-se com múltiplos machos durante a revoada, então, armazenam o esperma por toda sua vida útil (8 a 15 anos) e assim conseguem gerar uma descendência de trabalhadoras altamente diversa e polimórfica. A polimorfia é muito importante para a colônia, uma vez que assim é garantido a divisão de trabalho, uma característica importante para os organismos sociais (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986; SÉRGIO DE BRITTO *et al.*, 2016).

A polimorfia é observada nos ninhos das formigas cortadeiras por meio das castas (Figura 1). As castas são símbolos não apenas da distinção comportamental, mas também morfológica (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986). As castas compreendem: machos, rainha e operárias (jardineiras, generalistas, forrageadoras e soldados). Os machos emergem somente na época do acasalamento, período em que também emergem as jovens rainhas aladas. As rainhas são responsáveis por fundar, gerar indivíduos e manter a coesão social da colônia.

O período de acasalamento das castas sexuais aladas (conhecidas como tanajuras e bitus) ocorre após o início das chuvas na estação de calor. Após a revoada o macho morre e a jovem rainha inicia o estabelecimento da colônia no solo. O sucesso da sua empreitada depende que ela sobreviva ao possível ataque de predadores, às chuvas e microrganismos presentes neste

ambiente. Após a escavação, regurgitam o fungo guardado na câmara infrabucal cultivando-o com secreções e fezes e iniciam a ovoposição (SCHULTZ, 2021; SÉRGIO DE BRITTO *et al.*, 2016).

**Figura 1** - Representação do ciclo de vida e das castas em uma sociedade de formigas cortadeiras.



Fonte: Adaptado (PINTEREST-a, b)

Após poucos meses, inicia-se a emergência das primeiras operárias aptas a forragear, assim, a rainha não mais dependerá de sua reserva energética e da ovoposição de ovos tróficos, o que permite então que a colônia comece a desabrochar. Esse período é o mais crítico para a colônia uma vez que a taxa de sucesso (sobrevivência) é baixíssimo - menos de 1 % - só atingindo a maturidade aproximadamente 3 anos após a sua fundação (SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014). A rainha e as operárias (fêmeas estéreis) são as únicas castas permanentes da colônia. As operárias executam diversas tarefas no formigueiro e diferenciam-se pela sua morfologia. As operárias menores (jardineiras e generalistas) executam suas funções no interior da colônia, enquanto as operárias maiores (forrageadoras e soldados)

trabalham no exterior da colônia (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986; SOLIS *et al.*, 2012).

As jardineiras são responsáveis por manipular as larvas, processar o substrato do fungo e limpar o formigueiro (DELLA LUCIA, 2011; DELLA LUCIA *et al.*, 2003; LOFGREN, C.S., & VANDER MEER, 1986). A coleta de material vegetal cabe às forrageadoras, uma vez que estas possuem mandíbulas adaptadas ao corte de material vegetal; já a defesa da colônia fica a cargo dos soldados, sendo estas muito maiores que as demais operárias. Outra forma de polimorfismo que pode existir dentro da colônia é o polietismo etário, onde operárias jovens realizam tarefas no interior da colônia enquanto as mais velhas desempenham funções externamente à colônia. A vantagem, seria manter o interior da colônia livre de contaminantes externos e garantir a sobrevivência dos indivíduos mais jovens e da rainha prolongando a vida útil do formigueiro (SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014).

Observa-se que há uma divisão clara de tarefas e as atividades desempenhadas pelas formigas cortadeiras envolvem comportamentos complexos adquiridos ao longo de milhares de anos de coevolução entre elas e o fungo, i. e., elas processam o material vegetal por meio da mastigação visando tornar o conteúdo foliar mais facilmente metabolizável pelo fungo (WORSLEY *et al.*, 2018; ZANETTI *et al.*, 2014). Após o processamento do material vegetal as formigas o inoculam com micélio e constantemente patrulham os jardins de fungo removendo contaminantes e resíduos destinando-os às câmaras de lixo. Evitam a contaminação aplicando substâncias produzidas nas glândulas metapleurais, exibindo comportamentos chamados de “autogrooming” e “allogrooming” - ato de lamber a si mesma e as companheiras, respectivamente (DELLA LUCIA, 2011; DELLA LUCIA *et al.*, 2003; LOFGREN, C.S., & VANDER MEER, 1986; SÉRGIO DE BRITTO *et al.*, 2016; ZANETTI *et al.*, 2014).

As colônias de formigas cortadeiras são altamente organizadas e resistentes a doenças (WORSLEY *et al.*, 2018). A resistência a doenças se deve tanto ao comportamento das formigas quanto à rainha. Sabe-se que as formigas do gênero *Acromyrmex* possuem associação com actinobactérias, as quais produzem antibióticos que impedem a proliferação de microrganismos indesejados na colônia. Além disso, as formigas cortadeiras possuem a incrível habilidade de reconhecer fontes vegetais que afetam negativamente o desenvolvimento do fungo parando de forragear esta fonte em particular por semanas (DELLA LUCIA, 2011; DELLA LUCIA *et al.*, 2003; LOFGREN, C.S., & VANDER MEER, 1986; SCHULTZ, 2021; SÉRGIO DE BRITTO *et al.*, 2016).

As formigas cortadeiras são uns dos insetos sociais mais bem-sucedidos devido à alta integração da sociedade, obtida por meio da comunicação química. A comunicação química permite que as formigas cortadeiras possuam a capacidade de reconhecer fontes seguras para o forrageio, principalmente após um evento adverso, assim como reconhecer membros da colônia e comunicar entre si, graças à presença de lobos olfativos superdesenvolvidos compostos de centenas de glomérulos, os quais permitem a comunicação entre os neurônios sensoriais e os neurônios do sistema nervoso central. Além disso, as formigas são capazes de sintetizar feromônios desencadeadores de alarme, territorialidade, marcação de trilha e reconhecimento das companheiras de ninho. Essas substâncias são produzidas e armazenadas em diversas glândulas espalhadas pelo seu corpo e a presença e tamanho dessas glândulas também é afetada pela função que a operária desempenha na colônia (DELLA LUCIA, 2011; LOFGREN, C.S., & VANDER MEER, 1986; SCHULTZ, 2021; SÉRGIO DE BRITTO *et al.*, 2016; WORSLEY *et al.*, 2018; ZANETTI *et al.*, 2014).

Como mencionado, diversos são os processos ocorrendo dentro da colônia - processamento do alimento, limpeza do jardim de fungo, troca de fluidos e alimentos (DELLA LUCIA, 2011; SÉRGIO DE BRITTO *et al.*, 2016). As hifas (alto teor proteico) são a principal fonte de alimentação das larvas, que são capazes de digeri-las por completo, enquanto as adultas se alimentam de líquidos obtendo sua fonte de energia do suco dos gongilídeos e da seiva de plantas. As larvas são capazes de digerir as hifas pois possuem mandíbulas com espinhos capazes de perfurar o micélio e enzimas que catalisam reações de degradação das paredes celulares das hifas. Por outro lado, as adultas possuem uma região denominada câmara infrabucal, onde ocorre a retenção de partículas sólidas, as quais serão descartadas na forma de pellets pelas formigas nas câmaras de lixo (SOLIS *et al.*, 2012; WORSLEY *et al.*, 2018; ZANETTI *et al.*, 2014)

Um comportamento peculiar das formigas consiste na diversidade de comportamentos associados à dispersão de nutrientes pela colônia. Esses comportamentos específicos são chamados de trofalaxia. A trofalaxia envolve a troca de material excretado, secretado e regurgitado e é um comportamento importante dos insetos sociais. As larvas dependem das jardineiras para alimentarem-se pois são imóveis, mas existem registros de alimentos digeridos por larvas que são ingeridos pelas adultas (trofalaxia anobucal), sugerindo que as larvas também desempenham uma função importante na colônia, como casta digestiva. Assim, as larvas garantem o cuidado por parte das operárias fortalecendo a ligação social entre elas. Outra forma

de troca de nutrientes se dá pela troca de líquidos estocados no papo das adultas (trofalaxia oral) (DELLA LUCIA, 2011; MEURVILLE; LEBOEUF, 2021).

### **2.2.2. Análise genômica e funcional de formigas cortadeiras**

Muitos estudos vêm sendo feitos nas últimas décadas visando elucidar os aspectos moleculares da relação simbiótica entre o fungo *L. gongylophorus* e as formigas do gênero *Acromyrmex* e *Atta*. Com o surgimento das ômicas, a geração de genomas e transcritomas tem nos ajudado a compreender as relações evolutivas e genéticas entre essas espécies. Além disso, estudos genômicos permitem a caracterização de genes envolvidos em rotas metabólicas essenciais ao desenvolvimento biológico das formigas cortadeiras. Em estudos mais recentes descobriu-se que na verdade as espécies do gênero *Atta* surgiram entre 16 e 8 milhões de anos e que sua expansão biogeográfica para áreas do Cerrado na América do Sul coincide com um período de 4 milhões de anos e maior diversificação das espécies deste gênero (BARRERA *et al.*, 2021).

A publicação dos genomas de duas espécies de formigas cortadeiras (*Atta cephalotes* e *Acromyrmex echinator*) confirmou a hipótese da dependência nutritiva das Attini superiores em relação ao fungo simbiote. Esses estudos evidenciam que não foi possível encontrar 2 genes necessários para a biossíntese de arginina (EC 6.3.4.5, EC 4.3.2.1) no genoma destas espécies, enquanto o fungo é capaz de sintetizá-la (NYGAARD *et al.*, 2011; SUEN *et al.*, 2011). Estudos mais recentes sugerem que a perda destes genes coincide com a origem das Attini uma vez que também não foram encontradas em outras classes de formigas cultivadoras de fungo (Attini inferiores) (JEŠOVNIK; GONZÁLEZ; SCHULTZ, 2016).

Estudos genômicos também são importantes para a detecção de genes alvos ao controle de pragas baseadas em estratégias moleculares. Um estudo em *Atta laevigata* através da técnica de marcação de sequências expressas gerou 2006 sequências únicas, tendo caracterizado e selecionado 146 como potenciais candidatas para o controle, sendo estas envolvidas em diversas funções: desenvolvimento, longevidade, imunidade, sinalização celular, comportamento (RODOVALHO *et al.*, 2011).

Devido à grande diversidade de funções e comportamentos na colônia, espera-se que as diferentes castas apresentem diferentes padrões de expressão gênica, principalmente relacionadas às atividades envolvendo a proteção da colônia e comunicação entre os membros. Sabendo disso, um estudo baseado no transcritoma das antenas de machos, rainhas e de duas subcastas de operárias identificou genes diferencialmente expressos de diferentes classes de

proteínas ligadas ao sistema olfativo e quimio sensorial em *Atta vollenweideri* (KOCH *et al.*, 2013).

Análises filogenéticas a partir do mitogenoma de 5 espécies de *Atta* mostram que neste gênero há uma alta organização gênica, ordem gênica muito conservada e regiões intergênicas gigantes. Além disso, os autores concluem que a análise filogenética para estas espécies foi mais informativa ao se utilizar o mitogenoma completo (BARBOSA *et al.*, 2020; RODOVALHO *et al.*, 2014).

Ampliar o conhecimento sobre a biologia das formigas cultivadoras de fungo por meio de dados moleculares é imprescindível para: elucidação da relação simbiótica das Attini com seu simbiote, estudo de outras interações presentes no interior da colônia e compreensão das funções de genes essenciais à mudança de estádios de desenvolvimento. Muito tem-se descoberto sobre as interações simbióticas das formigas cortadeiras e as vantagens dessas descobertas para a comunidade científica são inestimáveis, pois apesar de serem consideradas insetos-praga, as formigas desempenham diversas funções ecológicas importantes para a manutenção do ecossistema em que habitam. No entanto, ainda são escassos os estudos baseados em dados moleculares para validação da função gênica de enzimas importantes, como aquelas relacionadas ao metabolismo de quitina, que são importantes constituintes do exoesqueleto dos insetos e das matrizes peritróficas presentes no intestino. Além de que, as formigas cortadeiras possuem como principal fonte de alimento o fungo, cuja parede celular é composta de quitina.

### 2.3. Quitina em insetos e o papel das quitinases

O exoesqueleto e a matriz peritrófica (camada externa às células epiteliais do intestino médio) são estruturas essenciais durante o desenvolvimento dos insetos sendo estes formados essencialmente por quitina – um polissacarídeo ( $\beta$ -1,4 N-acetilglucosamina – GlcNAc) componente de matrizes extracelulares de diversos organismos. O exoesqueleto cumpre diversas funções nos insetos: homeostase dos fluidos corporais; manutenção do formato corporal; facilitação da locomoção e do voo; proteção contra danos mecânicos e radiação; proteção contra patógenos; entre outros. A matriz peritrófica atua como uma barreira protetora de compostos digeridos pelos insetos, tais como, partículas de comida, patógenos e toxinas (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019).

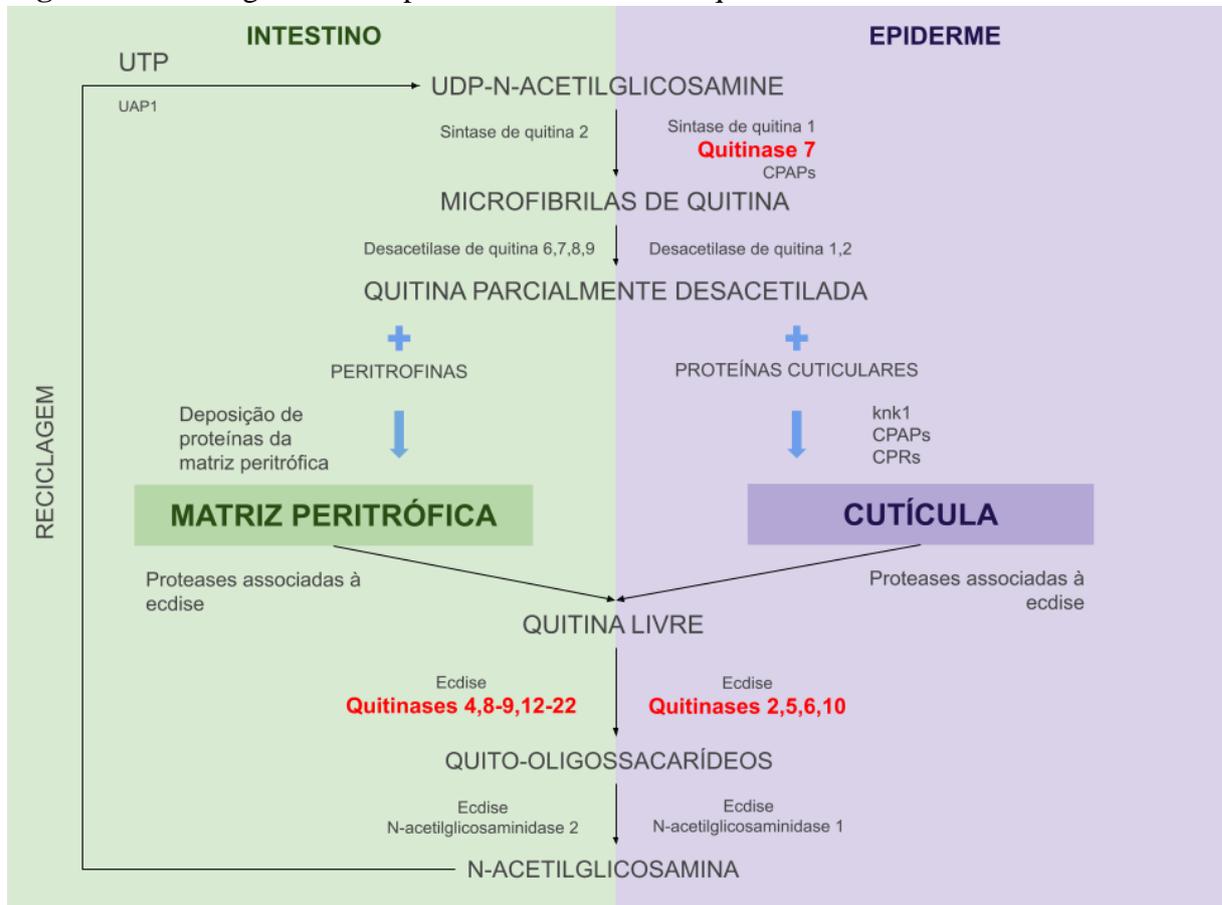
O exoesqueleto é constituído pela cutícula, epiderme e membrana basal. A cutícula é dividida em uma camada mais externa chamada de epicutícula (altamente esclerotizada, i.e., rígida e escura) e a procutícula (compostas por camadas de quitina mais flexíveis) sendo esta subdividida em exocutícula e endocutícula, e em alguns casos interna a estas duas há a mesocutícula. Como os insetos devem passar por diferentes estádios de desenvolvimento, e conseqüente modificação do exoesqueleto (ecdises), eles devem digerir a cutícula velha e sintetizar uma nova. Para que isso ocorra a cutícula velha deve ser separada da epiderme (apólise). Nesse processo a epiderme secreta fluidos, ou seja, componentes cuticulares, assim como enzimas do metabolismo de quitina (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019).

A matriz peritrófica (MP) é uma matriz extracelular semipermeável presente no intestino médio de muitos insetos, cuja função se assemelha às das secreções das mucosas do trato digestivo dos vertebrados. A constituição da matriz peritrófica varia em torno de 3% a 13% de microfibras flexíveis de quitina e 20% a 55% de proteínas, glicoproteínas e proteoglicanos. Como existem diversos comportamentos alimentares entre os insetos a estrutura da matriz peritrófica também apresenta grande diversidade, sendo este um reflexo da adaptação evolutiva, sendo que isto tornou a quitina desta estrutura mais hidratada, flexível, e que o espessamento e o número de camadas de quitina pode variar entre diferentes espécies (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019).

O metabolismo de quitina se divide em processos altamente regulados de síntese, remodelação, desacetilação e degradação (Figura 2). As quitinases (E.C. 3.2.1.14) são enzimas que catalisam a hidrólise de quitina (polímeros de N-acetil-glicosamina) através da clivagem de ligações glicosídicas ( $\beta$ - [1-4]), cujo tamanho varia de 20 kDa a 90 kDa, e estão presentes em diversos organismos – bactérias, fungos, plantas, artrópodes e humanos. As quitinases também participam do processo de hidrólise de polissacarídeos de paredes celulares constituídas de resíduos ácidos de N-acetil-glicosamina e N-acetilmurâmico (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019). As quitinases são classificadas em 2 famílias: 18-glicosil-hidrolases (GH18) e 19-glicosil-hidrolases. Enquanto as enzimas pertencentes à família 18 estão amplamente distribuídas em todos os reinos, as da família 19 são encontradas em sua maioria em plantas. Em geral, as enzimas da família 18 são exclusivamente endoquitinases e apresentam pouca ou nenhuma atividade exoquitinase. Podem atuar com grupos de enzimas da família 20-glicosil-hidrolases

e combinadas com enzimas que apresentam domínios de ligação a quitina (CBD) (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019).

**Figura 2** - Visão geral das etapas do metabolismo de quitina em insetos.



Fonte: Adaptado (MUTHUKRISHNAN *et al.*, 2019).

De modo geral, as quitinases podem exercer as seguintes funções em insetos: modificação de matrizes extracelulares, como cutícula e tecido peritrófico; atuar na digestão de insetos que se alimentam de fungos; e, ainda, como fatores de crescimento. Essas enzimas foram encontradas em diferentes ordens de insetos, tendo sido purificadas a partir de diferentes tecidos: intestino; corpo gorduroso; hemolinfa; asas; entre outros (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019).

A partir de estudos comparativos, determinou-se que essas enzimas podem apresentar os seguintes domínios: o GH18, catalítico; domínio de ligação a quitina rico em cisteína (peritrophin-a/CBM-14 - CBD); domínio de ligação a serina/treonina altamente glicosilados; e, em alguns casos, um domínio transmembranar e ou peptídeo sinal. Baseado nas diferentes arquiteturas que essas enzimas podem apresentar, foram então divididas em 11 grupos nos

insetos. Há fortes evidências corroborando que os grupos de quitinases desempenham diferentes funções no organismo, embora nem todos tenham sido ainda bem caracterizados. (ARAKANE; MUTHUKRISHNAN, 2010; LIU; ZHANG; ZHU, 2019; MUTHUKRISHNAN *et al.*, 2019)

Estudos empregando técnicas de silenciamento gênico pós-transcricional permitiram a determinação da função de alguns destes genes durante as diferentes ecdises pelas quais os insetos passam (larva-larva, larva-pupa, pupa-adulto):

- Grupo I: quitinase-5 (Cht5) apresenta um domínio catalítico, um domínio CBM-14 e um peptídeo sinal; Cht5 é expressa em todos os estádios, mas há indícios que seja essencial somente para a ecdise do estágio pupal-adulto (ZHU *et al.*, 2008);
- Grupo II: quitinase-10 (Cht10), de alto peso molecular, apresenta 4-5 domínios catalíticos, 4-7 CBD; Cht10 quando suprimida afetou a eclosão da larva, muda da larva e metamorfose do adulto. Geralmente é expressa em todos os estádios do desenvolvimento e no tegumento, não nos intestinos, pressupõe-se que esteja relacionada à renovação da quitina cuticular (DONG, 2020);
- Grupo III: quitinase-7 (Cht7) apresenta 2 domínios catalíticos, 1 CBD e 1 domínio transmembranar na região terminal N; Cht7 é essencial para a formação das asas (ZHU *et al.*, 2008);
- Grupo IV: quitinases-4/8-9/12-22 (Cht4, Cht8-9, Cht12-22), apresentam um peptídeo sinal, 1 domínio catalítico e geralmente não apresentam CBD; Quitinases do grupo IV (Cht4, Cht8-9, Cht12-22) quando silenciadas não induziram mudanças significativas, mas sabe-se que são expressas no intestino (ARAKANE; MUTHUKRISHNAN, 2010);
- Grupo V: fator de crescimento do disco imaginal (IDGF), esse grupo apesar de apresentar um domínio catalítico GH18, perdeu a atividade de hidrolase; IDGFs atuam na regulação da proliferação celular, são expressas nas células vitelinas, nos corpos gordurosos e são transportadas pela hemolinfa. Atuam na proliferação dos discos imaginiais (precursoras das estruturas cuticulares) e são necessárias para a emergência do adulto (ZHU *et al.*, 2008);
- Grupo VI: quitinase-6 (Cht6), similar ao grupo I, no entanto, apresenta na extremidade C uma região com muitos resíduos Ser/Thr que a torna altamente glicosilada; Cht6 é necessária para formação correta das asas, desde a permeabilidade até o espessamento desta, atua aumentando o sinal de apoptose, e também afetou o desenvolvimento abdominal; além disso, é altamente expressa nos estádios pupais iniciais, coincidente

com quitina sintase – CHS. Após o experimento de supressão dos transcritos codificadores desta enzima, os autores observaram uma modificação na via de síntese de quitina pela modulação dos fatores Gfat1 e Ter, precursores no processo de síntese de quitina. Os autores do estudo hipotetizam que Cht6, Cht2, Cht10 e CHS são necessárias para formação da cutícula nas asas (DONG *et al.*, 2022);

- Grupo VII: quitinase-2 (Cht2) apresenta uma estrutura similar à do grupo IV; Cht2 promove espessamento da cutícula e formação na matriz cuticular; essencial para organização de quitina em larva e na formação de asas em adultos. Um estudo recente em *D. melanogaster* empregando ensaios imunofluorescentes e RNAi demonstrou que Cht2 é expressa em órgãos formadores de cutícula durante a embriogênese e na epiderme, traqueia, espiráculos, intestinos e nas glândulas salivares de larvas e quando sua expressão foi suprimida levou ao afinamento da camada de procutícula e à deformação dos poros da cutícula (PESCH; RIEDEL; BEHR, 2017);
- Grupo VIII: quitinase-11 (Cht11) apresenta um domínio catalítico e um domínio transmembranar na região terminal N. Não existem evidências de sua função específica em insetos (ARAKANE; MUTHUKRISHNAN, 2010).
- Grupo IX: quitinase 1 (Cht1) apresenta um domínio catalítico e um grupo peptídeo sinal; foi encontrada em órgãos reprodutivos masculino e feminino em *Manduca sexta* (ordem *Lepidoptera*, família *Sphingidae*) (TETREAU *et al.*, 2015).
- Grupo X: quitinase 3 (Cht3) apresenta um domínio catalítico, peptídeo sinal e 2-3 CBDs; foi encontrada em cabeça e abdômen de larvas do 4º instar de *M. sexta* (TETREAU *et al.*, 2015).
- Grupo h: quitinase h (Cht-h) apresenta um domínio catalítico e um domínio PKD1; somente foi encontrada em lepidópteros sendo expressas em todos os estádios do desenvolvimento e em diversos tecidos (TETREAU *et al.*, 2015).

Sendo assim, embora nem todas as quitinases tenham sido caracterizadas, com base nos dados disponíveis atualmente, pode-se inferir que as quitinases são ótimas candidatas de estudos em insetos sociais como a formiga cortadeira *A. sexdens* que apresenta diferentes classes morfológicamente distinguíveis, onde o metabolismo de quitina deve sofrer muitos processos de regulação da expressão gênica para garantir o desenvolvimento das diferentes castas. Devendo-se considerar também os diversos processos de troca de nutrientes entre as formigas e o fungo, os quais serão abordados no tópico posterior.

### 2.3.1. Identificação de quitinases em formigas cortadeiras

Até o momento, a maior parte dos estudos envolvendo a identificação de quitinases em formigas cortadeiras estão relacionados aos processos de digestão e fluxo de nutrientes pela colônia. Por isso, faz-se necessário uma breve descrição desses processos tanto envolvendo larvas, formigas adultas e o próprio fungo simbiote.

Formigas cortadeiras adultas possuem a alimentação bem diversificada. As rainhas no momento de fundação da colônia alimentam a si, as larvas e as primeiras operárias com ovos tróficos; após o estabelecimento do jardim de fungo, a rainha e as operárias alimentam-se do conteúdo líquido dos gongilídeos rico em carboidratos, das secreções labiais das larvas e secreções anais que são ricas em fontes proteicas; além disso, as operárias também ingerem exsudatos líquidos no momento do forrageamento (DELLA LUCIA, 2011; ERTHAL; PERES SILVA; IAN SAMUELS, 2004). Observa-se, portanto, que os processos de trofalaxia são especialmente importantes na alimentação das adultas.

Na primeira etapa da digestão, as partículas sólidas ficam retidas na câmara infrabucal (Figura 3), onde foi identificada intensa atividade quitinolítica, o que se acredita esteja relacionado à digestão inicial do fungo. O conteúdo líquido ingerido passa à glândula pós-faringiana onde as substâncias hidrossolúveis seguem para o papo; em contrapartida, as substâncias lipossolúveis entram nos dutos das glândulas pós-faringeana e são transferidos para a hemolinfa sendo dispersados por todo o corpo da formiga. O papo exerce uma função social muito ligada a trofalaxia, pois o conteúdo líquido armazenado nesta estrutura é usado nas trocas orais de conteúdo enzimático e nutrientes. Inclusive acredita-se que as enzimas sintetizadas na glândula labial sejam também armazenadas no papo, e é importante notar que nas glândulas labiais foi encontrada intensa atividade quitinolítica (DELLA LUCIA, 2011; ERTHAL; PERES SILVA; IAN SAMUELS, 2004).

O conteúdo alimentar pré-digerido, segue para o intestino médio, o íleo e o ânus, onde também foram encontradas atividades quitinolíticas, remontando à presença de quitinases. Outro fato interessante, é que as operárias utilizam o líquido fecal para inocular o material vegetal, possuindo enzimas importantes para esse processo, incluindo as quitinases. Alguns autores identificaram que essas enzimas sejam de origem do fungo, as quais se mantêm íntegras durante todo o processo de digestão da formiga (SCHIØTT; BOOMSMA, 2021). Estes estudos sugerem que as quitinases sejam importantes tanto na digestão do fungo, e que por estarem presentes no conteúdo fecal sejam utilizadas junto de outras enzimas em processos de

detoxificação da colônia (DELLA LUCIA, 2011; ERTHAL; PERES SILVA; IAN SAMUELS, 2004).

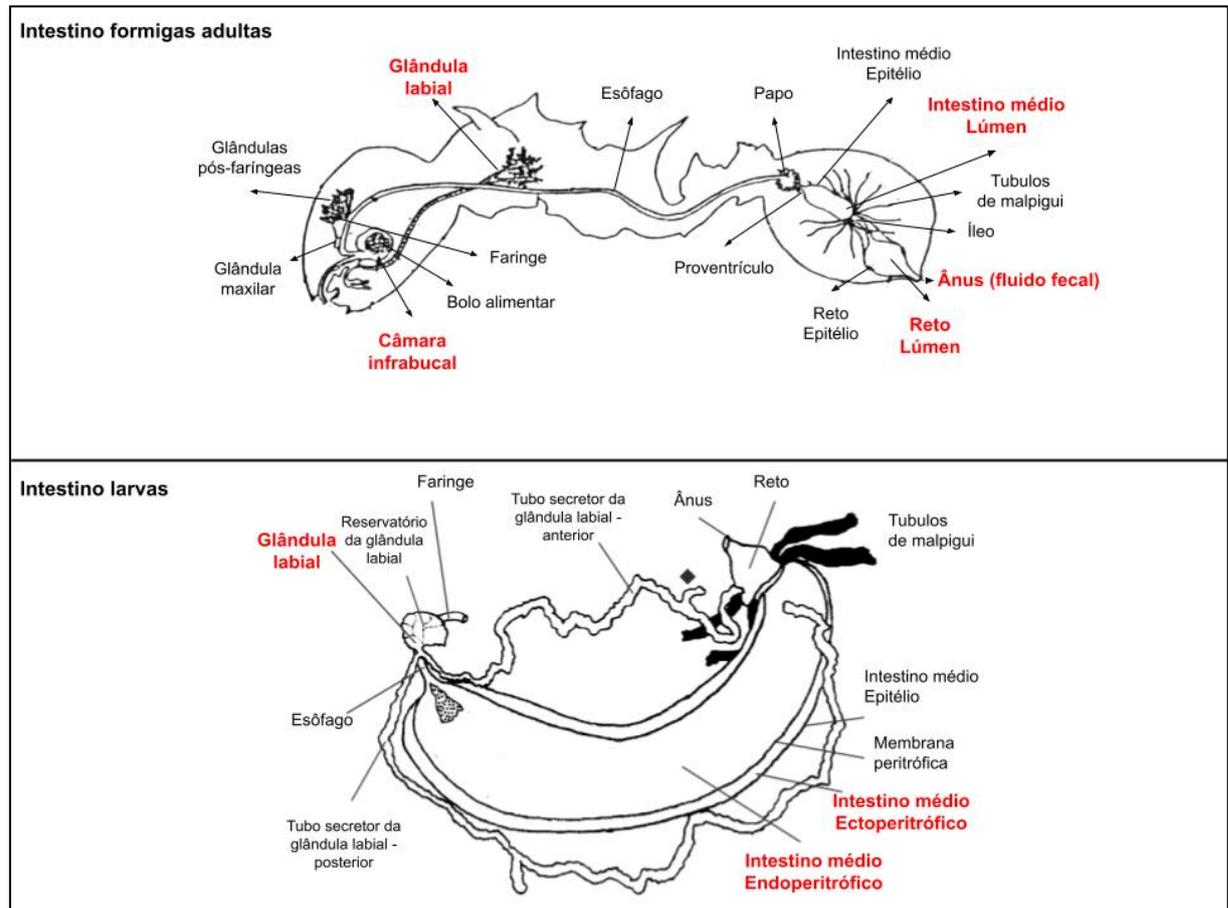
Essa pressuposição de que as quitinases possam ser usadas em mecanismos de defesa, relaciona-se com um estudo recente de isolamento de uma quitinase do grupo II (Cht10) em formigas da espécie *A. sexdens*. Neste estudo os autores isolaram uma porção contendo um domínio catalítico GH18 ativo e um domínio CBD, a qual nomearam como AsChtII-C4B1, expressaram em levedura e isolaram a enzima, a qual foi testada contra larvas de *Spodoptera frugiperda* (ordem *Lepdoptera*, família *Noctuidae*) e 2 espécies de fungos, *Aspergillus fumigatus* (ordem *Eurotiales*, família *Trichocomaceae*) filamentoso e *Candida albicans* (ordem *Saccharomycetales*, família *Saccharomycetaceae*) levedura. Os resultados observados foram: a quitinase do grupo II afetou a motilidade dos estádios larva-pupa, afetou negativamente o crescimento do fungo filamentoso, mas o mesmo resultado não foi observado contra a levedura. A partir desses resultados os autores concluem que a enzima apresentou importante atividade inseticida e fungicida e que mais estudos seriam necessários para determinar o efeito de quitinases exógenas no desenvolvimento dos insetos e que devido à proximidade de formigas e fungos filamentosos que vivem no solo a atividade fungicida se justifique (MICOCCI *et al.*, 2023).

Esse último trabalho evidencia a importância dos estudos que identificaram atividade quitinolítica em diferentes amostras biológicas da formiga relacionadas ao sistema digestivo, mas que também são necessários estudos posteriores de identificação de quais são as enzimas específicas presentes nestas amostras e a função delas. Além do fato, de existirem supostas quitinases fúngicas, que fogem ao escopo deste trabalho, mas que será necessário estudá-las a fim de elucidar as suas funções nos formigueiros.

Semelhantemente ao que ocorre nas adultas, as larvas apresentam quitinases tanto na glândula labial, como nas regiões endo e ectoperitróficas do intestino médio (Figura 3). As larvas alimentam-se principalmente das paredes celulares dos gongilídeos, isso por que possuem mandíbulas contendo espinhos que são capazes de desestruturar as paredes celulares do fungo, além disso possuem uma maior diversidade de enzimas no lúmen intestinal, sendo capazes de digerir substratos que as formigas não são. Interessantemente, quitinases não foram detectadas no epitélio do intestino e por isso os autores sugerem que as quitinases nesta porção do sistema digestivo possam ser oriundas do fungo (DELLA LUCIA, 2011; ERTHAL; PERES SILVA; IAN SAMUELS, 2007).

Quitinases foram detectadas tanto nas hifas como nos jardins de fungo, inclusive quitinases de origem fúngica foram detectadas nas secreções fecais de formigas adultas (SCHIØTT; BOOMSMA, 2021). Como o fungo alimenta-se do material vegetal coletado e processado pelas formigas, as quitinases acabam não tendo uma função evidente na digestão do fungo, o qual produz outras enzimas com este propósito. Logo, pressupõe-se que a presença de quitinases esteja relacionado na lise de fungos concorrentes, pois as quitinases seriam armazenadas no papo e utilizadas pelas formigas naqueles atos de lambedura de si mesmas e das companheiras, visando evitar contaminações, além de auxiliar na digestão (DELLA LUCIA, 2011; ERTHAL *et al.*, 2009).

**Figura 3** - Representação intestino de formiga cortadeira adulta e larva em relação a presença de quitinases em estruturas específicas, as quais estão identificadas em negrito e cor vermelha.



Fonte: Adaptado (DELLA LUCIA, 2011).

### 3 METODOLOGIA

#### 3.1. Amostra biológica e extração de RNA

Formigas cortadeiras da casta das Rainhas, pertencentes à espécie *A. sexdens*, foram coletadas durante o período de revoada - entre os meses de novembro e dezembro de 2015 - no campus universitário da Universidade Federal de Lavras (UFLA), em Lavras, Minas Gerais. Para o estabelecimento de colônias *in vitro*, as rainhas foram mantidas em sala de criação ( $23 \pm 2$  °C, 0 lux, 70% Umidade Relativa) no Laboratório de Manejo Integrado de Pragas Florestais (LMIP), Departamento de Entomologia da UFLA (MAXIMO, 2018).

A partir de três colônias estabelecidas *in vitro* obteve-se o RNA total de cinco amostras biológicas: larva inteira (LARVAE, n=225), intestino de larva (LMG, n=600), formiga forrageadora inteira (ANT, n=75), cabeça de formiga forrageadora (HEAD, n=225), intestino de formiga forrageadora (AMG, n=600). Conforme instruções do fabricante, o RNA total foi extraído utilizando o reagente TRizol® (Invitrogen), seguido pela etapa de tratamento com o kit Turbo DNase (Invitrogen). A qualidade das amostras foi avaliada utilizando o equipamento Bioanalyzer (Agilent, USA) e por meio de análise do perfil do RNA total em gel de agarose 1,0% após eletroforese (MAXIMO, 2018).

#### 3.2. Construção das bibliotecas de cDNA e sequenciamento via Illumina HiSeq 2500

As amostras de RNA foram enviadas para o Centro de Genômica Funcional, Departamento de Zootecnia, Escola Superior de Agricultura "Luiz de Queiroz" da Universidade de São Paulo (ESALQ/USP), em Piracicaba, São Paulo (MAXIMO, 2018).

As bibliotecas de mRNA foram obtidas empregando-se o protocolo Illumina TruSeq Stranded mRNA Sample Prep LS. Primeiramente, as moléculas de RNA contendo cauda poli-A foram isoladas utilizando esferas magnéticas ligadas a oligopoli-T. Após uma segunda etapa de purificação, as amostras de mRNA foram fragmentadas, e os insertos de cada biblioteca continham em média 150pb. Para a síntese de cDNA foram usados iniciadores randômicos. Após a reação de transcrição reversa, o RNA molde foi removido da reação, e as fitas complementares de cDNA foram sintetizadas. As sequências obtidas foram hibridizadas na "flow-cell", suas terminações ligadas à adaptadores e enriquecidas via PCR (MAXIMO, 2018).

O sequenciamento das bibliotecas foi realizado na plataforma Illumina HiSeq2500 (Illumina, Inc., USA), especificamente, HiSeq Flow Cell v4, com o kit HiSeq SBS kit v4, obtendo-se leituras pareadas de aproximadamente 100 pb. Os dados brutos do sequenciamento foram armazenados em arquivos no formato fastq. (MAXIMO, 2018).

### 3.3. Montagem *de novo* do transcrito

As leituras brutas foram processadas para remoção das sequências adaptadoras e das leituras de baixa qualidade (BOLGER; LOHSE; USADEL, 2014); em seguida utilizou-se o programa Trinity v2.14 para montagem *de novo* do transcrito da formiga cortadeira *A. sexdens*. O tamanho do k-mer utilizado foi de 25, com fragmentos sobrepostos esperados de no mínimo 105 pb (GRABHERR *et al.*, 2011). Após a montagem dos contigs pelo programa Trinity, o programa CD-HIT-EST v4.8.1 foi usado para reduzir a redundância e para montagem dos unigenes (transcritos agrupados por similaridade) empregando um limiar de identidade entre as sequências de 90% (FU *et al.*, 2012; LI Ñ; GODZIK, 2006). Posteriormente, os unigenes foram submetidos à ferramenta TransDecoder v5.5.0 para predição dos transcritos codificadores (<https://github.com/TransDecoder/TransDecoder/wiki>).

A predição dos transcritos codificadores seguiu-se pelas seguintes etapas: primeiro foram extraídas as fases abertas de leitura (ORFs, do inglês - Open reading frames) que codificavam, no mínimo 100 aminoácidos, utilizando o módulo TransDecoder.LongOrfs. Posteriormente, essas ORFs foram comparadas utilizando a ferramenta BlastP com o banco de dados UniprotKB/Swiss-Prot; elas também foram submetidas à uma busca de domínios proteicos, utilizando a ferramenta hmmscan (EDDY, 2009) e o banco de dados Pfam (FINN *et al.*, 2010). Por último, utilizando o módulo TransDecoder.Predict, os dados de similaridade e os domínios identificados foram incorporados ao processo de identificação das candidatas a transcritos codificadores. Três arquivos foram gerados após análises do TransDecoder: fasta de nucleotídeos (CDS), fasta de aminoácidos (PEP), arquivos de anotação gênica (GFF3 e BED).

As estatísticas de montagem nas diferentes etapas (contigs, unigenes, CDS) foram obtidas utilizando ferramentas nativas do Trinity (TRINITYRNASEQ-a); os dados de cobertura de proteínas do banco de dados UniprotKB/Swiss-Prot, foram obtidos utilizando BlastX v2.12 e ferramentas nativas do Trinity (TRINITYRNASEQ-b).

As taxas de alinhamento de leituras ao transcrito foram obtidas por meio dos softwares Bowtie2 (LANGMEAD; SALZBERG, 2012) e Samtools (LI *et al.*, 2009), a

completude do transcrito foi obtida através da ferramenta BUSCO (modo “transcriptome”, linhagem “hymenoptera\_odb10”) (MANNI *et al.*, 2021).

### 3.4. Anotação funcional dos transcritos codificadores

As sequências de peptídeos (arquivo PEP), oriundas das ORFs preditas, foram comparadas a quatro bancos de dados: UniprotKB/Swiss-ProtKB (UNIPROT-DATABASE); KOG (do inglês - Eukaryotic Orthologous Groups) (KOG-DATABASE); sequências de peptídeos oriundas do genoma da abelha *Apis mellifera* (RefSeq GCF\_003254395.2); sequências de peptídeos oriundas do genoma da formiga cortadeira *Atta cephalotes* (RefSeq GCF\_000143395.1).

A análise de similaridade foi realizada pelo programa BlastP v2.12 (e-value 1e-5) (ALTSCHUL *et al.*, 1990). E a anotação gênica funcional das sequências de peptídeos oriundas das ORFs preditas foi atribuída pelo programa eggNOG (CANTALAPIEDRA *et al.*, 2021). Além disso foram usados os programas: hmmscan, para busca de domínios proteicos (EDDY, 2009); TMHMM v2.0, para busca de domínios intermembranares (KROGH *et al.*, 2001); e, SignalP v4.1, para busca por peptídeos sinal (PETERSEN *et al.*, 2011).

### 3.5. Identificação *in silico* de transcritos codificadores de quitinases no transcrito da formiga cortadeira *A. sexdens*

Proteínas pertencentes à família das quitinases (Tabela 1) foram utilizadas para busca de sequências similares nas sequências de peptídeos oriundas das ORFs preditas pelo TransDecoder (arquivo PEP) a partir do transcrito da formiga *A. sexdens* utilizando o programa BlastP (e-value 1e-5 max\_hsps) (ALTSCHUL *et al.*, 1990). As sequências de peptídeos de *Atta sexdens* similares às quitinases utilizadas como referência foram filtradas baseadas na presença do domínio GH18 porque todas as quitinases devem apresentar pelo menos o domínio catalítico. Posteriormente, as sequências utilizadas como referência e similares foram alinhadas utilizando o programa MAFFT v7.490 (default) (KATOHI *et al.*, 2002). O alinhamento foi otimizado utilizando-se CIALign v1.0.18 (--remove\_insertions --crop\_ends) (TUMESCHEIT; FIRTH; BROWN, 2022).

A inferência da árvore filogenética foi obtida através do método de máxima verossimilhança implementada pelo programa IQ-TREE v2.0.7 (MINH *et al.*, 2020). Os valores de suporte de ramificação da árvore filogenética foram obtidos utilizando-se o método

UFBoot (Bootstrapping - 1000 replicatas) implementado pelo IQ-TREE (HOANG *et al.*, 2018). A visualização e edição da árvore foi gerada pelo programa iTOL v.6.6 (LETUNIC; BORK, 2021). Ademais, foi incorporada à árvore filogenética a predição de motivos estruturais nas sequências proteicas utilizando-se MEME v5.5.0 (BAILEY *et al.*, 2015). Os números de motivos foram determinados pela contagem de domínios presentes nas sequências através do programa hmmscan (EDDY, 2009).

A fim de identificar os motivos conservados do domínio GH18 (KxxxxxGGW, FDGxDLDWEYP, MxYDxxG, GxxxWxxDxD), as candidatas foram alinhadas usando o programa MAFFT v7.490 (default) (KATOHI *et al.*, 2002) e o alinhamento visualizado através do programa Aliview (LARSSON, 2014).

Além disso, as sequências proteicas identificadas como candidatas à quitinases em *A. sexdens* foram analisadas por meio do programa SMART (<http://smart.embl-heidelberg.de/>) para verificação dos domínios básicos referentes a cada proteína. Quando necessário, ou seja, se em algum grupo não foi possível encontrar uma candidata que tivesse todos os domínios esperados, as sequências candidatas foram alinhadas e agrupadas para gerar uma sequência consenso utilizando o programa Aliview (LARSSON, 2014).

**Tabela 1** - Sequências utilizadas como referência para identificação das candidatas a quitinasas no transcrito de *A. sexdens* (continua).

<b>Espécie</b>	<b>Código NCBI</b>	<b>Comprimento (aa)</b>	<b>Proteína</b>	<b>Grupo</b>
<i>Drosophila melanogaster</i>	NP_650314	595	Cht5	I
<i>Drosophila melanogaster</i>	NP_001036422	2286	Cht10A	II
<i>Drosophila melanogaster</i>	NP_647768	1013	Cht7	III
<i>Drosophila melanogaster</i>	NP_524962	462	Cht4	IV
<i>Drosophila melanogaster</i>	NP_611542	476	Cht8	IV
<i>Drosophila melanogaster</i>	NP_611543	368	Cht9	IV
<i>Drosophila melanogaster</i>	NP_477258	439	IDGF1	V
<i>Drosophila melanogaster</i>	NP_477257	440	IDGF-2A	V
<i>Drosophila melanogaster</i>	NP_477256	441	IDGF-3A	V
<i>Drosophila melanogaster</i>	NP_511101	442	IDGF-4A	V
<i>Drosophila melanogaster</i>	NP_611321	444	IDGF-5	V
<i>Drosophila melanogaster</i>	NP_477081	452	IDGF-6A	V
<i>Drosophila melanogaster</i>	NP_572598	4519	Cht6J	VI
<i>Drosophila melanogaster</i>	NP_477298	484	Cht2A	VII
<i>Drosophila melanogaster</i>	NP_572361	432	Cht11	VIII
<i>Tribolium castaneum</i>	NP_001034524.1	533	Cht5	I
<i>Tribolium castaneum</i>	XP_008198138.1	2712	Cht10A	II
<i>Tribolium castaneum</i>	NP_001036067.1	2700	Cht10B	II
<i>Tribolium castaneum</i>	NP_001036035.1	980	Cht7	III
<i>Tribolium castaneum</i>	NP_001038094.1	496	Cht8	IV
<i>Tribolium castaneum</i>	XP_015835197.1	444	IDGF-4A	V
<i>Tribolium castaneum</i>	NP_001038091.1	431	IDGF-4B	V
<i>Tribolium castaneum</i>	NP_001038092.1	439	IDGF-2	V
<i>Tribolium castaneum</i>	XP_971647.1	388	Cht1	IX
<i>Tribolium castaneum</i>	XP_008197064.1	2031	Cht3	X
<i>Bombix mori</i>	AAB47538.1	565	Cht5	I
<i>Bombix mori</i>	XP_004922005.1	987	Cht7	III
<i>Bombix mori</i>	XP_004933352.2	513	Cht2	VII
<i>Bombix mori</i>	XP_004926923.1	429	Cht11	VIII
<i>Bombix mori</i>	XP_004931749.1	394	Cht1	IX
<i>Bombix mori</i>	XP_012551241.2	2281	Cht3	X
<i>Bombix mori</i>	BAC67246.1	555	Cht-h	h
<i>Apis mellifera</i>	XP_026299805.1	2659	Cht10A	II
<i>Apis mellifera</i>	XP_026299806.1	2659	Cht10B	II
<i>Apis mellifera</i>	XP_006570346.2	2658	Cht10C	II
<i>Apis mellifera</i>	XP_396925.4	968	Cht7	III
<i>Apis mellifera</i>	XP_016769016.1	438	IDGF4	V
<i>Apis mellifera</i>	XP_026300868.1	451	Cht2	VII
<i>Apis mellifera</i>	XP_006572082.1	468	Cht11	VIII
<i>Apis mellifera</i>	XP_026299646.1	393	Cht1	IX
<i>Agrotis ipsilon</i>	UTR35847.1	578	Cht5	I
<i>Agrotis ipsilon</i>	UTR35851.1	2987	Cht10	II
<i>Agrotis ipsilon</i>	UTR35849.1	988	Cht7	III

**Tabela 1** - Sequências utilizadas como referência para identificação das candidatas a quitinases no transcrito de *A. sexdens* (conclusão).

<b>Espécie</b>	<b>Código NCBI</b>	<b>Comprimento (aa)</b>	<b>Proteína</b>	<b>Grupo</b>
<i>Agrotis ipsilon</i>	UTR35850.1	377	Cht8	IV
<i>Agrotis ipsilon</i>	UTR35854.1	433	IDGF	V
<i>Agrotis ipsilon</i>	UTR35848.1	1927	Cht6	VI
<i>Agrotis ipsilon</i>	UTR35845.1	507	Cht2	VII
<i>Agrotis ipsilon</i>	UTR35852.1	446	Cht11	VIII
<i>Agrotis ipsilon</i>	UTR35844.1	394	Cht1	IX
<i>Agrotis ipsilon</i>	UTR35846.1	2456	Cht3	X
<i>Agrotis ipsilon</i>	UTR35853.1	555	Cht-h	h

Fator de Crescimento do Disco Imaginal (IDGF, do inglês - *Imaginal Disc Growth Factor*)

Fonte: Do autor (2023).

## 4 RESULTADOS E DISCUSSÃO

### 4.1. Montagem e anotação do transcrito da formiga cortadeira *A. sexdens*

Após a extração de RNA, construção das bibliotecas de cDNA, sequenciamento e processamento das 5 amostras biológicas de *A. sexdens* (LARVAE, LMG, ANT, AMG e HEAD) foram obtidas ao todo 194 milhões de leituras pareadas de 100 pb (Tabela 2). Os dados de qualidade do RNA extraído encontram-se no Apêndice.

**Tabela 2** - Resumo estatístico por amostra biológica de *A. sexdens* obtido após análise dos dados de sequenciamento de RNA e taxa de alinhamento das leituras ao transcrito montado via abordagem *de novo* utilizando o programa Trinity.

Bibliotecas	Conteúdo	Comprimento	Nº	Alinhamento
	GC (%)	Leituras (pb)	Leituras*	(%)
HEAD	40	100	36,7	99,18
ANT	39	100	40,9	99,15
AMG	38	100	43,5	99,28
LMG	39	100	35,7	99,41
LARVAE	42	100	37,2	99,10

As bibliotecas de *A. sexdens* correspondem a: HEAD, cabeça de formiga forrageadora; ANT, formiga forrageadora inteira; AMG, intestino de formiga forrageadora; LMG, intestino de larva; e, LARVAE, larva inteira. \* Milhões de leituras pareadas.

Fonte: Do autor (2023).

O programa Trinity foi utilizado para montagem *de novo* visando à obtenção do transcrito de *A. sexdens*, sendo que foram gerados 322.651 contigs correspondentes a 245.494 genes. A fim de diminuir a redundância dos contigs, a ferramenta CD-HIT-EST realizou o agrupamento das sequências e gerou 267.811 unigenes reunidos em 235.586 genes. Após a predição de transcritos codificadores nos unigenes utilizando a ferramenta TransDecoder, foi possível identificar 45.253 transcritos do tipo CDS (do inglês - *Coding DNA Sequence*) compreendidos em 17.451 genes (Tabela 3).

Diferentes resultados estatísticos (conteúdo GC, N50, Cobertura de proteínas do Uniprot/Swissprot, completude gênica) foram usados para avaliar o perfil transcritômico de *A. sexdens* (Tabela 3 e 4). Em relação ao alinhamento, a biblioteca LARVAE foi a que exibiu a menor cobertura de leituras (99,10%), demonstrando que menos de 1% das leituras não tiveram

correspondência no transcrito montado com a ferramenta Trinity (Tabela 3). E como o limiar de fragmentos montados foi de 105 pares de base, pode-se supor que este percentual corresponda a reads únicas que podem ser oriundas de fragmentos de RNA muito pequenos ou erros aleatórios durante o processo de sequenciamento.

Metade dos contigs, unigenes e CDSs foram montados com tamanhos de pelo menos 2.443 pb, 1.658 pb, 1.458 pb (N50), respectivamente. Enquanto o N50 diminuiu ao longo das análises de redução da redundância e predição de transcritos codificadores, o N50 das maiores isoformas por genes aumentou (Tabela 3).

**Tabela 3** - Resumo estatístico do transcrito da formiga cortadeira *A. sexdens* obtido após a utilização de análises nativas da ferramenta Trinity.

<b>Montagem</b>	<b>Nº Genes</b>	<b>Nº Transcritos</b>	<b>Nº Bases</b>	<b>Conteúdo GC (%)</b>	<b>N50 (pb)</b>	<b>N50 maior isoforma (pb)</b>
Contigs	24.549	322.651	107.958.999	37,40	2.443	742
Unigenes	23.558	267.811	105.961.666	37,10	1.658	760
CDS	17.451	45.253	18.941.316	43,10	1.458	1.797

Fonte: Do autor (2023).

Para avaliar a qualidade da montagem dos transcritos, foram conduzidas análises de: cobertura de proteínas do banco de dados UniprotKB/Swiss-ProtKB pelos transcritos codificadores de peptídeos e a completude dos transcritos em termos de conteúdo gênico esperado para a ordem Hymenoptera à qual *A. sexdens* pertence (Tabela 4). Após a análise de cobertura identificou-se que 3.759, 3.547 e 3.227 proteínas foram cobertas em mais de 90% de seus comprimentos, pelos contigs, unigenes e CDSs, respectivamente. A partir da análise BUSCO observou-se alta taxa de completude gênica, porém também com alta taxa de duplicidade (Tabela 4) o que pode ter ocorrido devido à presença de diferentes isoformas de cada gene.

**Tabela 4** - Análises de cobertura de proteínas do banco de dados UniprotKB/Swiss-ProtKB e completude do transcrito de *A. sexdens* obtidas utilizando os programas Blast-X, análises nativas da ferramenta Trinity e BUSCO.

Montagem	Cobertura UniprotKB/ Swiss-Prot*	BUSCO			
		C	F	M	N
Contigs	3.759	96.7% [S:16,0%, D:80,7%]	1,8%	1,5%	5.991
Unigenes	3.547	96.4% [S:36,7%, D:59,7%]	2,1%	1,5%	5.991
CDS	3.227	91.0% [S:53,4%, D:37,6%]	4,4%	4,6%	5.991

\*A coluna Cobertura corresponde ao número de proteínas do banco de dados do UniprotKB/Swiss-Prot que tiveram cobertura pelos transcritos montados em *A. sexdens* maior que 90%. Na coluna BUSCO, as subcolunas representam: C (genes completos; S (genes únicos) e D (genes duplicados)), F (genes fragmentados), M (genes faltantes), N (número de genes totais pesquisados).

Fonte: Do autor (2023).

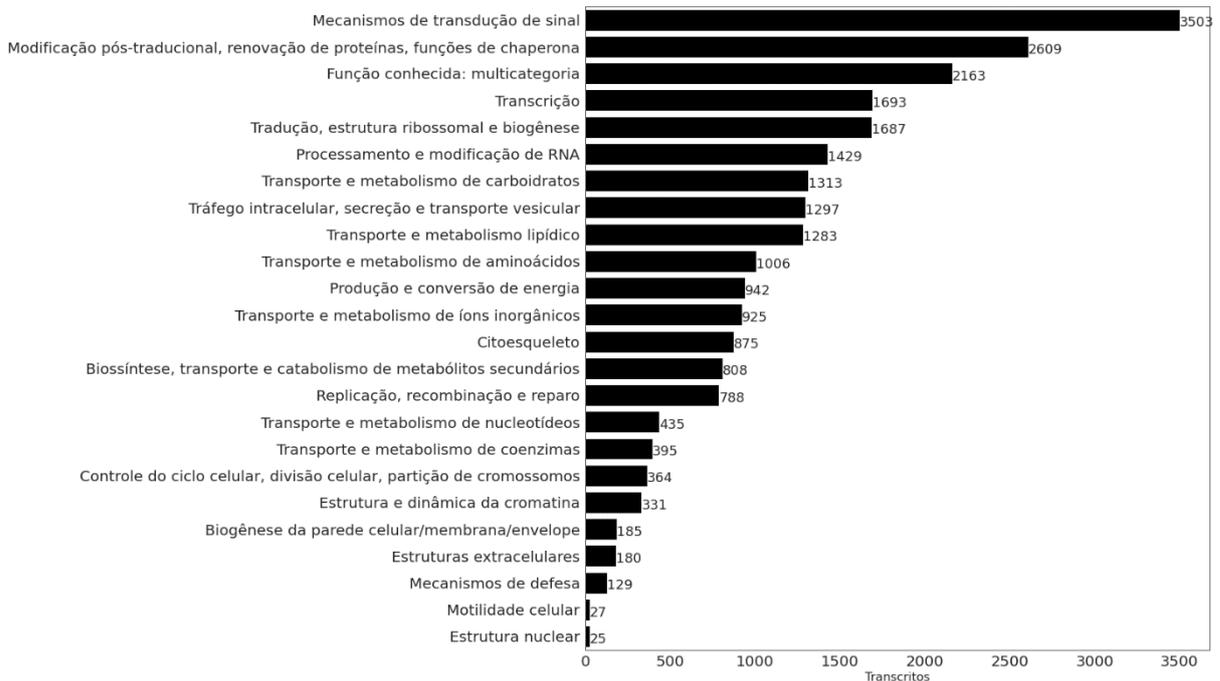
Um total de 45.253 transcritos codificadores foram preditas (arquivo de peptídeos) e submetidas a diferentes pipelines de anotação gênica funcional: 1) análise de similaridade comparada a 4 bancos de dados; 2) predição de domínios proteicos (geral, intermembranares e peptídeos sinal); 3) identificação de ortólogos e anotação da categoria dos Grupos de Genes Ortólogos (COGs, do inglês - *Clusters of Orthologous Genes*).

O número de proteínas preditas que apresentaram ao menos uma sequência similar contra os bancos de dados de *A. cephalotes*, *A. melifera*, KOG e UniprotKB/Swiss-Prot, respectivamente, corresponde a: 34.802, 33.806, 31.109 e 29.995 proteínas. Os bancos de dados que retornaram mais sequências anotadas foram o de *A. cephalotes* e *A. melifera*, uma vez que pertencem à ordem Hymenoptera, e, conseqüentemente, possuem um maior nível de proximidade filogenética com *A. sexdens*. O menor número de sequências anotadas ocorreu com o banco de dados do UniprotKB/Swiss-Prot, pois neste banco são depositadas somente sequências que tiveram sua função validada experimentalmente.

Ao todo, 36.639 sequências apresentaram pelo menos um domínio presente no banco de dados Pfam, com 9.681 sequências exibindo domínios intermembranares e outras 3.098 sequências com peptídeos-sinal. Ao empregar a anotação funcional via eggNOG foram computadas 24.392 sequências com ortólogos diretos. Essa anotação acaba sendo mais restritiva em relação à anotação por domínios e similaridade porque evita a anotação de parálogos próximos e transfere somente as informações de ortólogos refinados, uma vez que aplica métodos de identificação de ortólogos, como hmmer ou DIAMOND, e filogenias do banco de dados do próprio eggNOG.

A partir desta anotação também foi possível extrair as categorias COG. As 5 categorias mais representativas nos dados anotados foram: 1) mecanismos de transdução de sinal; 2) modificação pós-traducional, renovação de proteínas e funções de chaperona; 3) CDSs pertencentes a múltiplas categorias; 4) transcrição; e 5) tradução, estrutura ribossomal e biogênese (Figura 4).

**Figura 4 -** Contabilização das transcritos codificadores por categorias das ontologias gênicas (GO) no transcrito de *A. sexdens*. Os dados foram obtidos a partir da anotação funcional das sequências de peptídeos utilizando a ferramenta eggNOG.

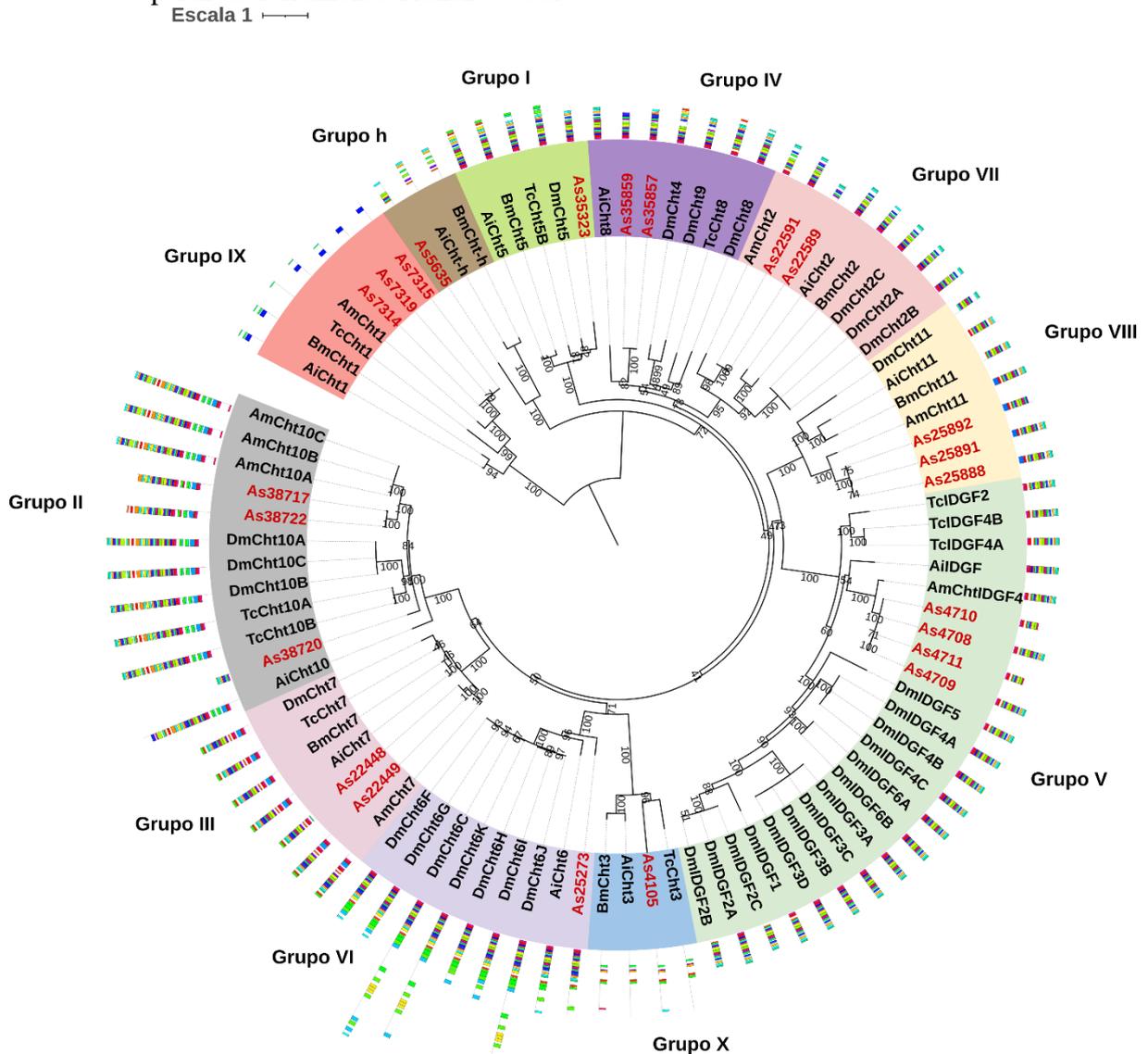


Fonte: Do autor (2023).

#### 4.2. Identificação *in silico* de transcritos codificadores de quitinases no transcrito de *A. sexdens*

As quitinases (família 18-glicosil-hidrolases) são importantes famílias de enzimas envolvidas nos processos de degradação de quitina. Após análise filogenética foram encontradas sequências relacionadas às enzimas hidrolisadoras de quitina em *A. sexdens* pertencentes a todos os grupos aqui descritos, embora algumas não tenham apresentados todos os motivos esperados com base nas sequências utilizadas como referência (Figura 5).

**Figura 5** - Análise filogenética das proteínas envolvidas na degradação de quitina em *Atta sexdens*. A análise foi realizada utilizando o software IQTREE2, empregando o método de máxima verossimilhança, com bootstrap de 1000 replicatas. As sequências utilizadas como referência foram obtidas das espécies *Drosophila melanogaster*, *Tribolium castaneum*, *Bombyx mori*, *Apis mellifera* e *Agrotis ipsilon* (Tabela 1). Os motivos presentes em cada sequência foram preditos usando a ferramenta MemeTool. As sequências putativas em *Atta sexdens* foram preditas a partir da ferramenta TransDecoder.



Abreviaturas: Tc (*Tribolium castaneum*); Dm (*Drosophila melanogaster*); Bm (*Bombyx mori*); Am (*Apis mellifera*); Ai (*Agrotis ipsilon*); e, As (*Atta sexdens*). A partir da análise de predição de domínios, cada cor representa um motivo identificado. O código das sequências utilizadas na construção da árvore filogenética encontra-se na Tabela 1 e Tabela 5.

Fonte: Do autor (2023).

Neste estudo foram identificadas um total de 23 transcritos compreendidos em 11 genes envolvidos na degradação de quitina em *A. sexdens* (Tabela 5).

**Tabela 5** - Transcritos codificadores em *A. sexdens* preditos a partir da ferramenta TransDecoder e identificados como candidatos a quitinases através de análise filogenética.

Sequência	Gene	Tamanho (aa)	Proteína	Grupo
As35323	DN5500_c0_g1	500	Cht-5	I
As38717	DN673_c1_g1	2358	Cht-10	II
As38720	DN673_c1_g1	183	Cht-10	II
As38722	DN673_c1_g1	2463	Cht-10	II
As22448	DN2847_c0_g1	699	Cht-7	III
As22449	DN2847_c0_g1	970	Cht-7	III
As35857	DN567_c0_g1	385	Cht-4,8-9	IV
As35859	DN567_c0_g1	339	Cht-4,8-9	IV
As4708	DN1285_c1_g1	444	IDGF	V
As4709	DN1285_c1_g1	437	IDGF	V
As4710	DN1285_c1_g1	281	IDGF	V
As4711	DN1285_c1_g1	283	IDGF	V
As25273	DN3305_c0_g1	1816	Cht-6	VI
As22589	DN2870_c0_g1	242	Cht-2	VII
As22591	DN2870_c0_g1	439	Cht-2	VII
As25888	DN3422_c2_g1	499	Cht-11	VIII
As25891	DN3422_c2_g1	472	Cht-11	VIII
As25892	DN3422_c2_g1	442	Cht-11	VIII
As7314	DN1470_c0_g1	259	Cht-1	IX
As7315	DN1470_c0_g1	349	Cht-1	IX
As7319	DN1470_c0_g1	266	Cht-1	IX
As4105	DN12528_c0_g1	1893	Cht-3	X
As5635	DN13424_c0_g1	402	Cht-h	h

Fonte: Do autor (2023).

Os transcritos codificadores de quitinases identificados em *A. Sexdens* foram alinhados a fim de se identificar os quatro grupos de motivos conservados (KxxxxxGGW, FDGxDLDWEYP, MxYDxxG, GxxxWxxDxD) presentes nas sequências respectivas ao domínio GH18. Observa-se que o transcrito As38720 não apresenta os motivos I, II, e III comumente presentes nas quitinases do grupo II, evidenciando os achados do MEME tool (figura 5) e indicando que este transcrito não está completo como esperado. Também é possível

observar que os transcritos do grupo V de quitinases, em conformidade com a literatura, apresentam algumas mutações nos motivos II e III. E pode-se notar também que os grupos de quitinases IX e X apresentam deleções em alguns motivos conservados.

**Figura 6** - Análise de motivos conservados comumente presentes nos domínios catalíticos (GH18). Os transcritos codificadores de quitinases em *Atta sexdens* foram alinhados usando o programa MAFFT e visualizado por meio do programa Aliview.

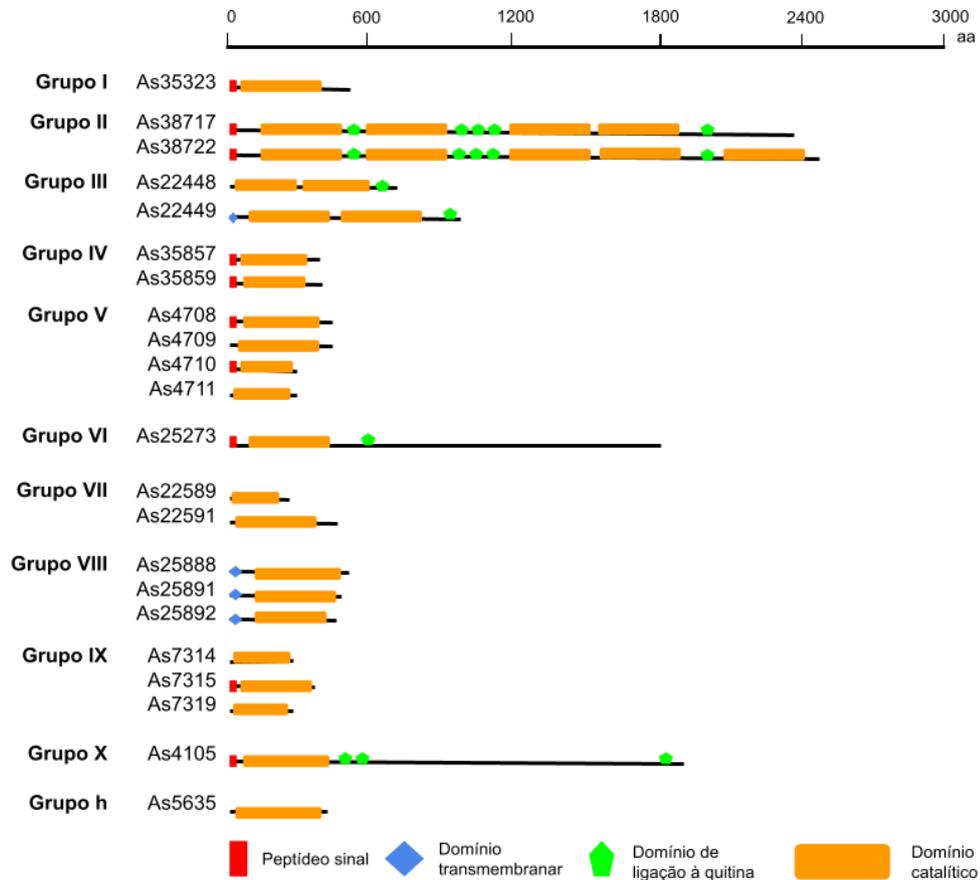
		I	II	III	IV
		KxxxxGGW	FDGxDLDWEYP	MxYDxxG	GxxxWxxDxD
Grupo I	As35323	K T Q V A I G G W	F D G F D L D W E Y P	M T Y D L R G	G A M V W A C D M D
	As38717	K V L L A L G G W	F D G L D M D W E Y P	M A Y D Y H G	G G M I W A L D L D
Grupo II	As38720	- - - - -	- - - - -	- - - - -	G G M I W S L D L D
	As38722	K V L L A L G G W	F D G L D M D W E Y P	M A Y D Y H G	G G M I W A L D L D
Grupo III	As22448	K I L L A I G G W	F D G L D V D W E Y P	M A Y D F H G	G V M I W S V D M D
	As22449	K I L L A I G G W	F D G L D V D W E Y P	M A Y D F H G	G V M I W S V D M D
Grupo IV	As35857	K M M V A V G G W	F D G F D I D W E Y P	M T Y D L H G	G L M T W S I D T D
	As35859	K M M V A V G G W	F D G F D I D W E Y P	M T Y D L H G	- - - - -
Grupo V	As4708	K I Y L S V G G N	F D G I D L A W Q F P	F A F D Q K T	G I A I V D L S L D
	As4709	K I Y L S V G G N	F D G I D L A W Q F P	F A F D Q K T	G I A I V D L S L D
	As4710	K I Y L S V G G N	F D G I D L A W Q F P	F A F D Q K T	- - - - -
	As4711	K I Y L S V G G N	F D G I D L A W Q F P	F A F D Q K T	- - - - -
Grupo VI	As25273	K T M L A I G G W	F D G L D L D W E Y P	L S Y D Y H S	G I M F W S I D N D
Grupo VII	As22589	- - - - -	- - - - -	- - - - -	G V M V W S I D T D
	As22591	K V S L G I G G W	F D G L D L D W E F P	M A Y D Y H G	G V M V W S I D T D
Grupo VIII	As25888	K V M I S A G G N	L D G F D V D W E F P	M S Y D Y H F	G A M I L S L N V D
	As25891	K V M I S A G G N	L D G F D V D W E F P	M S Y D Y H F	G A M I L S L N V D
	As25892	K V M I S A G G N	L D G F D V D W E F P	M S Y D Y H F	- - - - -
	As7314	K L I P R V - - -	F D G Y V L E - - - -	M T Y D Y Q -	G I S I W E L - - -
Grupo IX	As7315	K L I P R V - - -	F D G Y V L E - - - -	M T Y D Y Q -	- - - - -
	As7319	K L I P R V - - -	F D G Y V L E - - - -	M T Y D Y Q -	- - - - -
Grupo X	As4105	Q F V I S I - - -	V D G V E L G - - - -	S T D Y M D G	G L A L R N I E - N
Grupo h	As5635	K N I L S I G G W	F D G L D I D Y E Y P	M A Y D F S G	G S M F W E L S G D

Os resíduos de aminoácidos presentes em mais de 50% das sequências foram marcados com cores de fundo. Cada grupo (I, II, III e IV) refere-se um conjunto de sequências conservadas.

Fonte: Do autor (2023).

A fim de determinar se as sequências candidatas possuem os domínios esperados para cada grupo de quitinases, também foi realizada uma análise de identificação dos domínios utilizando a plataforma SMART.

**Figura 7** - Identificação dos domínios presentes nas candidatas a transcritos codificadores de quitinases em *Atta sexdens*.



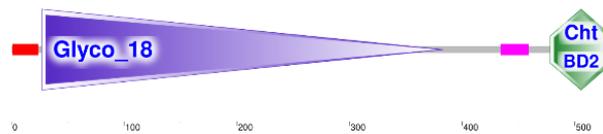
Fonte: Do autor (2023).

As quitinases Cht5 pertencem ao grupo I e possuem um grupo peptídeo sinal, um domínio catalítico (GH18) e um domínio de ligação à quitina (CBD). A sequência candidata ao grupo I (As35323) não apresenta o domínio de ligação de quitina (CBD). No entanto, essa sequência é uma das isoformas do gene DN5500\_c0\_g1 que contém uma isoforma curta (As35324) de 140 aminoácidos que corresponde à essa porção faltante em As35323. Dessa forma, com a formação de uma sequência consenso a partir das sequências As35323 e As35324 obtém-se uma sequência completa codificadora de Cht5 com os domínios esperados, ou seja, peptídeo sinal, um GH18 e um CDB (Figura 7) e de 535 aminoácidos. Nesse caso, há duas possibilidades de interpretação. A primeira poderia ser em função da montagem que não conseguiu refletir a estrutura completa da sequência. A segunda interpretação é de que exista duas isoformas do gene codificador de Cht5 como ocorre em outras espécies (Li *et al.*, 2015).

Li e colaboradores (2015) identificaram duas isoformas do gene de Cht5 (Cht5-1 e Cht5-2) no genoma de *Locusta migratoria* (ordem *Orthoptera*), onde somente a Cht5-1 apresenta a porção corresponde ao domínio de ligação à quitina. Após estudar a função desses genes no

desenvolvimento do gafanhoto eles observaram que somente após o silenciamento da isoforma Cht5-1 induziu a mortalidade do inseto adulto, demonstrando a importância do domínio CBD para que a enzima exerça sua função. A Cht5 foi detectada em todos os estágios do desenvolvimento do besouro *T. castaneum*. No entanto, o silenciamento desta quitinase não afetou os estágios larvais e pupais, somente causando mortalidade em adultos (ZHU *et al.*, 2008). Logo, pressupõe-se que esta enzima seja necessária somente para a metamorfose do adulto.

**Figura 8** - Identificação de domínios utilizando o programa SMART a partir da sequência consenso (As35323+As35324) candidata a quitinase Cht5 em *Atta sexdens*.



Representação: retângulo vermelho (peptídeo sinal), triângulo roxo (domínio catalítico), retângulo rosa (sequência de baixa complexidade) e hexágono verde (domínio de ligação a quitina).

Fonte: Do autor (2023).

As quitinases do grupo II (Cht10) possuem um peptídeo sinal, 4-5 domínios catalíticos e 4-7 CBDs. Duas candidatas a transcritos codificadores do grupo II foram identificadas em *A. sexdens*: As38717 e As38722. Ambas foram identificadas como isoformas do gene DN673\_c1\_g1 e apresentaram no mínimo 4 domínios catalíticos e 4 domínios CBD, além do peptídeo sinal (Figura 6). Estudos em *T. castaneum* têm demonstrado que o silenciamento de transcritos da Cht10 afeta tanto as ecdises larva-larva, quanto larva-pupa, levando a mortalidade dos insetos. Adicionalmente, embora não afete o desenvolvimento dos adultos, as fêmeas postam ovos que não eclodem (ZHU *et al.*, 2008). Recentemente, Dong e colaboradores (2022) demonstraram a importância dessa enzima na formação de asas em *D. melanogaster*, e que o silenciamento dos transcritos codificadores de Cht10 foi específico, pois não afetou a expressão de outras quitinases. Li e colaboradores (2022), também silenciaram os transcritos de Cht10 em *Agrotis ipsilon* (ordem *Lepidoptera*, família *Noctuidae*) e observaram a ocorrência de mortalidade em larvas e defeitos morfológicos em pupas deste inseto.

As quitinases do grupo III (Cht7) apresentam um domínio transmembranar, dois domínios catalíticos e um domínio CBD. Dentre as candidatas do grupo III, a sequência As22449 apresenta todos os domínios esperados para as quitinases deste grupo (Figura 6). Embora não tenha ocorrido mortalidade após silenciamento dos transcritos codificadores de

Cht7 em *T. castaneum*, observou-se deformações morfológicas que afetaram o desenvolvimento abdominal e das asas nos adultos (ZHU *et al.*, 2008). Em *A. ipsilon*, por outro lado, não foram observados efeitos morfológicos e/ou mortalidade em larvas e pupas (LI *et al.*, 2022a). O silenciamento de Cht7 em *Sogatella furcifera* (ordem *Hemiptera*), levou a uma série de deformações fenotípicas e até mesmo a letalidade de ninfas (CHEN *et al.*, 2017). Esses dados demonstram que o silenciamento de chitinases do mesmo grupo podem ter efeitos diferentes dependendo das ordens de insetos.

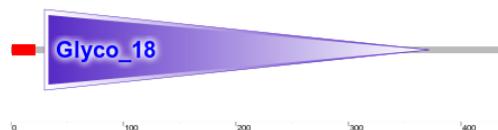
O grupo IV contém diversas quitinases (Cht4, Cht8, Cht9, Cht12-22), as quais são mais abundantes em Coleópteros, enquanto em outras, há somente 1 ou 2 genes identificados em outras ordens. Todas as candidatas em *A. sexdens* (As35857 e As35859) pertencentes ao grupo IV das quitinases apresentaram os domínios esperados, ou seja, um grupo peptídeo sinal e um domínio catalítico (Figura 6). Em *T. castaneum*, essas quitinases foram identificadas no intestino (ZHU *et al.*, 2008), por isso, especula-se que estejam envolvidas na digestão. Estudos de expressão gênica em diferentes espécies como *Tenebrio molitor* (ordem *Coleoptera*, família *Tenebrionidae*), *M. sexta* e *A. ipsilon* corroboram essa informação (LI *et al.*, 2022a, 2022b; TETREAU *et al.*, 2015). No entanto, quando transcritos de candidatos deste grupo foram silenciados em *T. castaneum* e *A. ipsilon* não foram observados efeitos morfológicos ou letais.

IDGFs são consideradas quitinases, no entanto, o domínio catalítico neste grupo, não tem a atividade hidrolase (ARAKANE; MUTHUKRISHNAN, 2010). Embora a análise de predição de motivos estruturais das candidatas codificadoras a IDGF (grupo V) tenha retornado que a sequência As4710 tenha uma pequena porção faltante em relação às demais proteínas desse grupo, assim como a candidata As4708, As4710 apresentou todos os domínios esperados – peptídeo sinal e domínio catalítico – na análise de identificação de domínios através do programa SMART (Figura 6). Até o momento, estudos adicionais são necessários para elucidar a função das quitinases do grupo V. Acredita-se que elas estejam envolvidas na proliferação celular e diferenciação celular, porém nenhum efeito letal ou morfológico expressivo foi detectado quando um gene deste grupo foi silenciado em *T. castaneum* (ZHU *et al.*, 2008). No entanto, estudos mais recentes, revelam efeitos surpreendentes quando foram conduzidos ensaios de silenciamento de transcritos de um candidato a IDGF em *A. ipsilon* conduziu a alta mortalidade de larvas (LI *et al.*, 2022a). Os autores deste estudo também observaram, um enegrecimento das larvas, efeito chamado de melanização e que está relacionado a resposta a agentes patológicos como bactérias. Logo, as enzimas deste grupo poderiam estar relacionadas à resposta imune em *A. ipsilon*.

As quitinases do grupo VI (Cht6) apresentam um grupo peptídeo sinal, um domínio catalítico e um domínio CBD. Em *A. sexdens*, a sequência As25273 se assemelha às isoformas I, J e H de Cht6 de *D. melanogaster* utilizadas como referência. Todas essas sequências apresentam um peptídeo sinal, um domínio catalítico, um domínio de ligação de quitina e uma longa porção contendo sequências de baixa complexidade (Figura 6). A função dessa enzima foi validada em *D. melanogaster*, onde observou-se que o silenciamento deste gene causou redução do tamanho das asas, modificação do formato das asas e afinamento da cutícula (DONG *et al.*, 2022).

As quitinases Cht2, pertencentes ao grupo VII, apresentam os seguintes domínios: um grupo peptídeo sinal e um domínio catalítico. Neste estudo, foram identificadas duas candidatas a transcritos codificadores do grupo Cht2 do grupo VII, as sequências As22589 e As22591 (Figura 6); porém, foi preciso gerar uma sequência consenso entre ambas as candidatas para que fosse possível identificar os domínios esperados para essa quitinase (Figura 8). A quitinase Cht2 é importante para a formação da matriz extracelular de quitina e estudos em *Drosophila* envolvendo o silenciamento de Cht2, têm demonstrado a ocorrência de uma alta mortalidade de larvas, além de prejudicar a motilidade do estágio larval-pupal. O silenciamento deste gene provoca o afinamento da camada de cutícula da matriz extracelular, compatível com as deformidades observadas em larvas, pupas e asas de adultos (PESCH *et al.*, 2016; PESCH; RIEDEL; BEHR, 2017).

**Figura 9** - Análise de identificação de domínios utilizando o programa SMART na sequência consenso (As22589+As22591) candidata à quitinase Cht2 em *Atta sexdens*.



Representação: retângulo vermelho (peptídeo sinal) e triângulo roxo (domínio catalítico).

Fonte: Do autor (2023).

A partir da análise das quitinases Cht11 (grupo VIII) foi possível identificar 3 sequências candidatas (As25888, As25891 e As25892) em *A. sexdens*. Todas apresentaram os domínios esperados: um domínio transmembranar e um catalítico. Não foram encontrados estudos relacionados a validação da função de Cht11. No entanto, dados de expressão gênica são bem variáveis entre espécies. Em *T. molitor*, esses transcritos de Cht11 foram detectados nos Túbulos de Malpighi de larvas e pernas e abdomens de adultos (LI *et al.*, 2022b). Em *A.*

*ípsilon*, por outro lado, foram detectados transcritos de Cht11em todos os estádios do desenvolvimento (LI *et al.*, 2022a). E em *M. sexta*, houve expressão no intestino, corpo gorduroso e testículos de pupas, enquanto em adultos somente no corpo gorduroso (TETREAU *et al.*, 2015).

Pertencentes ao grupo IX (Cht1), as supostas quitinases identificadas em *A. sexdens* foram as sequências As7314, As7315 e As7319. Somente a sequência As7315 apresentou os domínios esperados (peptídeo sinal e domínio catalítico). Os genes deste grupo foram expressos especificamente em testículos e ovário de pupas e adultos de *M. sexta* (TETREAU *et al.*, 2015); e, em testículos de *T. molitor* (LI *et al.*, 2022b).

O grupo X (Cht3) foi identificado como contendo unicamente a candidata As4105. Esta sequência apresentou todos os domínios esperados para este grupo: um peptídeo sinal, um domínio catalítico e três domínios de ligação à quitina. Análises de expressão gênica conduzidas em *M. sexta* identificaram transcritos na cabeça e abdômen de larvas (TETREAU *et al.*, 2015); e em *T. molitor*, foram identificados transcritos de Cht3 no tecido muscular de larvas e nas pernas de adultos (LI *et al.*, 2022b).

O grupo h (Cht-h), que até o momento foi identificado somente em insetos da ordem lepidóptera teve apenas uma sequência agrupada no mesmo clado da árvore filogenética (As5635). No entanto, a análise estrutural de domínio realizada na plataforma SMART não identificou o domínio que caracteriza este grupo (PKD1), assim como, não foram identificadas sequências pertencentes ao mesmo grupo gênico da candidata que contenham esse domínio. As Cht-h apresentam alta similaridade com quitinases homólogas de bactérias (DAIMON *et al.*, 2003), além disso, o silenciamento de transcritos do gene deste grupo induziu mortalidade de *Spodoptera exigua* (ordem *Lepidoptera*, família *Noctuidae*) (ZHANG *et al.*, 2012).

A partir do emprego de diversas ferramentas de bioinformática foi possível encontrar sequências candidatas a pertencer à família das quitinases no transcrito da formiga cortadeira *A. sexdens*. Essas análises são importantes porque permitem a identificação de candidatos potenciais de validação da função destas enzimas da espécie-alvo. Além disso, tais dados poderão ser complementados com dados de expressão gênica em diferentes estádios do desenvolvimento, uma vez que as quitinases são essenciais durante as ecdises dos insetos. Com base em dados da literatura todas as quitinases são ótimas candidatas de estudo em *Atta sexdens*, embora pouco se saiba sobre a função específica de algumas, uma vez que foram observados diferentes efeitos para o mesmo grupo de quitinases em diferentes insetos. Ou seja, embora tenhamos evidências das funções gerais de algumas quitinases, ao se estudar essas enzimas

especificamente em *A. sexdens* pode ser que venhamos a fazer observações diferentes daquelas encontradas na literatura.

## 5 CONCLUSÃO

A nova era da informação biológica baseada em dados moleculares obtidos em escala permite a maior compreensão do desenvolvimento de diversos organismos, entre eles os insetos. *A. sexdens* é um inseto complexo para o qual poucos dados moleculares estão disponíveis. Assim, a obtenção de dados transcritômicos para esta espécie tornou possível este trabalho.

Após a montagem do transcritoma da formiga cortadeira *A. sexdens*, a partir de 5 amostras biológicas por meio da ferramenta Trinity, foram obtidos 322.651 transcritos. Estes transcritos foram agrupados por similaridade para gerar unigenes, totalizando 267.811 sequências, as quais foram utilizadas para identificação de transcritos codificadores de proteínas similares às proteínas do banco de dados Pfam e Uniprot, por meio da ferramenta TransDecoder. Ao final, foram identificadas 45.253 sequências, sendo que após análise filogenética foram identificados 23 candidatos codificadores de quitinases.

Este é um estudo pioneiro na identificação ampla de quitinases em *A. sexdens* a partir da predição de transcritos codificadores empregando diversas ferramentas de bioinformática a partir de dados de RNA-Seq e isso permitirá que novos estudos sejam conduzidos visando a elucidação do papel destas enzimas no desenvolvimento da formiga cortadeira *A. sexdens*, por exemplo, por meio de ensaios de expressão gênica e silenciamento gênico pós-transcricional via RNA interferente.

## REFERÊNCIAS

- ALTSCHUL, S. F. *et al.* **Basic local alignment search tool.** *Journal of molecular biology*, v. 215, n. 3, p. 403–410, 1990.
- ARAKANE, Y.; MUTHUKRISHNAN, S. **Insect chitinase and chitinase-like proteins.** *Cellular and Molecular Life Sciences*, jan. 2010.
- BAILEY, T. L. *et al.* **The MEME Suite.** *Nucleic Acids Research*, v. 43, p. 39–49, 2015.
- BARBOSA, J. T. V. *et al.* **Mitochondrial genomes of genus *Atta* (Formicidae: Myrmicinae) reveal high gene organization and giant intergenic spacers.** *Genetics and Molecular Biology*, v. 42, n. 4, 13 jan. 2020.
- BARRERA, C. A. *et al.* **Phylogenomic reconstruction reveals new insights into the evolution and biogeography of *Atta* leaf-cutting ants (Hymenoptera: Formicidae).** *Systematic Entomology*, p. syen.12513, 3 set. 2021.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* (Oxford, England), v. 30, n. 15, p. 2114–2120, 1 ago. 2014.
- BONO, H. *et al.* **Systematic Functional Annotation Workflow for Insects.** *Insects*, v. 13, n. 7, 1 jul. 2022.
- CANTALAPIEDRA, C. P. *et al.* **eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale.** *bioRxiv*, p. 2021.06.03.446934, 4 jun. 2021.
- CHEN, C. *et al.* **Identification and functional analysis of chitinase 7 gene in white-backed planthopper, *Sogatella furcifera*.** *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, v. 208–209, p. 19–28, 1 jun. 2017.
- CONESA, A. *et al.* **A survey of best practices for RNA-seq data analysis.** *Genome BiologyBioMed Central Ltd.*, , 26 jan. 2016.
- DAIMON, T. *et al.* **A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases.** *Insect Biochemistry and Molecular Biology*, v. 33, n. 8, p. 749–759, 1 ago. 2003.
- DELLA LUCIA, T. **Formigas cortadeiras: da bioecologia ao manejo.** Viçosa, MG: [s.n.].
- DELLA LUCIA, T. M. C. *et al.* **Colony behavior of *Atta sexdens rubropilosa* (Hymenoptera: Formicidae) in the absence of the queen under laboratory conditions.** *Behavioural Processes*, v. 64, n. 1, p. 49–55, 29 ago. 2003.
- DONG, W. *et al.* **Chitinase 10 controls chitin amounts and organization in the wing cuticle of *Drosophila*.** *Insect Science*, v. 27, n. 6, p. 1198–1207, 1 dez. 2020.
- DONG, W. *et al.* **Chitinase 6 is required for procuticle thickening and organ shape in *Drosophila* wing.** *Insect Science*, 2022.
- EDDY, S. R. **A new generation of homology search tools based on probabilistic inference.** *Genome informatics. International Conference on Genome Informatics*, v. 23, n. 1, p. 205–211, 2009.

- ERTHAL, M. *et al.* **Hydrolytic enzymes of leaf-cutting ant fungi.** *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, v. 152, n. 1, p. 54–59, 1 jan. 2009.
- ERTHAL, M.; PERES SILVA, C.; IAN SAMUELS, R. **Digestive enzymes of leaf-cutting ants, *Acromyrmex subterraneus* (Hymenoptera: Formicidae: Attini): distribution in the gut of adult workers and partial characterization.** *Journal of Insect Physiology*, v. 50, n. 10, p. 881–891, 1 out. 2004.
- ERTHAL, M.; PERES SILVA, C.; IAN SAMUELS, R. **Digestive enzymes in larvae of the leaf cutting ant, *Acromyrmex subterraneus* (Hymenoptera: Formicidae: Attini).** *Journal of Insect Physiology*, v. 53, n. 11, p. 1101–1111, 1 nov. 2007.
- FINN, R. D. *et al.* **The Pfam protein families database.** *Nucleic Acids Research*, v. 38, n. suppl\_1, p. D211–D222, 1 jan. 2010.
- FU, L. *et al.* **CD-HIT: accelerated for clustering the next-generation sequencing data.** v. 28, n. 23, p. 3150–3152, 2012.
- GAUTHIER, J. *et al.* **A brief history of bioinformatics.** *Briefings in Bioinformatics*, v. 20, n. 6, p. 1981–1996, 27 nov. 2019.
- 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, jul. 2011.
- HAAS, B. J.; ZODY, M. C. **Advancing RNA-Seq analysis.** *Nature Biotechnology*, maio 2010.
- HOANG, D. T. *et al.* **UFBoot2: Improving the Ultrafast Bootstrap Approximation.** *Molecular Biology and Evolution*, v. 35, n. 2, p. 518–522, 1 fev. 2018.
- HÖLLDOBLER, B.; WILSON, E. O. **The ants.** p. 732, 1990.
- HRDLICKOVA, R.; TOLOUE, M.; TIAN, B. **RNA-Seq methods for transcriptome analysis.** *Wiley Interdisciplinary Reviews: RNABlackwell Publishing Ltd*, , 1 jan. 2017.
- HU, T. *et al.* **Next-generation sequencing technologies: An overview.** *Human Immunology*, v. 82, n. 11, p. 801–811, 1 nov. 2021.
- JEŠOVNIK, A.; GONZÁLEZ, V. L.; SCHULTZ, T. R. **Phylogenomics and Divergence Dating of Fungus-Farming Ants (Hymenoptera: Formicidae) of the Genera *Sericomyrmex* and *Apterostigma*.** *PLOS ONE*, v. 11, n. 7, p. e0151059, 1 jul. 2016.
- KAPLI, P.; YANG, Z.; TELFORD, M. J. **Phylogenetic tree building in the genomic age.** *Nature Reviews Genetics* *Nature Research*, , 1 jul. 2020.
- KATOH, K. *et al.* **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Research*, v. 30, n. 14, p. 3059–3066, 15 jul. 2002.
- KOCH, S. I. *et al.* **Caste-Specific Expression Patterns of Immune Response and Chemosensory Related Genes in the Leaf-Cutting Ant, *Atta vollenweideri*.** *PLOS ONE*, v. 8, n. 11, p. e81518, 15 nov. 2013.
- KOG-DATABASE. **Index of KOG database.** Disponível em: <https://ftp.ncbi.nih.gov/pub/COG/KOG/>

- KROGH, A. *et al.* **Predicting transmembrane protein topology with a hidden markov model: application to complete genomes.** *Journal of Molecular Biology*, v. 305, n. 3, p. 567–580, 19 jan. 2001.
- LANGMEAD, B.; SALZBERG, S. L. **Fast gapped-read alignment with Bowtie 2.** *Nature Methods* 2012 9:4, v. 9, n. 4, p. 357–359, 4 mar. 2012.
- LARSSON, A. **AliView: a fast and lightweight alignment viewer and editor for large datasets.** *Bioinformatics*, v. 30, n. 22, p. 3276–3278, 15 nov. 2014.
- LETUNIC, I.; BORK, P. **Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.** *Nucleic Acids Research*, v. 49, n. W1, p. W293–W296, 2 jul. 2021.
- LI, W.; GODZIK, A. **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** v. 22, n. 13, p. 1658–1659, 2006.
- LI, H. *et al.* **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics*, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.
- LI, D. *et al.* **Two chitinase 5 genes from *Locusta migratoria*: Molecular characteristics and functional differentiation.** *Insect Biochemistry and Molecular Biology*, v. 58, p. 46–54, 1 mar. 2015.
- LI, D. *et al.* **Identification and RNAi-based functional analysis of chitinase family genes in *Agrotis ipsilon*.** *Pest Management Science*, v. 78, n. 10, p. 4278–4287, 1 out. 2022a.
- LI, L. *et al.* **Genomic and transcriptomic analyses of chitin metabolism enzymes in *Tenebrio molitor*.** *Archives of Insect Biochemistry and Physiology*, v. 111, n. 3, 1 nov. 2022b.
- LIU, X.; ZHANG, J.; ZHU, K. Y. **Chitin in arthropods: Biosynthesis, modification, and metabolism.** *Advances in Experimental Medicine and Biology*, v. 1142, p. 169–207, 2019.
- LOFGREN, C.S., & VANDER MEER, R. K. (EDS.). **Fire Ants and Leaf-Cutting Ants: Biology and Management.** [s.l.] CRC Press, 1986. v. 1st ed.
- MANNI, M. *et al.* **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Molecular Biology and Evolution*, v. 38, n. 10, p. 4647–4654, 27 set. 2021.
- MAXIMO, W. P. F. **Análise de RNA-Seq e RNA de interferência em formigas cortadeiras *Atta sexdens*.** p. 151, 2018.
- MCGETTIGAN, P. A. **Transcriptomics in the RNA-seq era.** *Current Opinion in Chemical Biology*, fev. 2013.
- MEURVILLE, M. P.; LEBOEUF, A. C. **Trophallaxis: the functions and evolution of social fluid exchange in ant colonies (Hymenoptera: Formicidae).** *Myrmecological News* Austrian Society of Entomofaunistics, , 2021.
- MICOCCI, K. C. *et al.* **Identification, cloning, and characterization of a novel chitinase from leaf-cutting ant *Atta sexdens*: An enzyme with antifungal and insecticidal activity.** *Biochimica et Biophysica Acta - General Subjects*, v. 1867, n. 1, 1 jan. 2023.
- MINH, B. Q. *et al.* **IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.** *Molecular Biology and Evolution*, v. 37, n. 5, p. 1530–1534, 1 maio 2020.

MUTHUKRISHNAN, S. *et al.* **Chitin organizing and modifying enzymes and proteins involved in remodeling of the insect cuticle.** *Advances in Experimental Medicine and Biology*, v. 1142, p. 83–114, 2019.

NYGAARD, S. *et al.* **The genome of the leaf-cutting ant *Acromyrmex echinator* suggests key adaptations to advanced social life and fungus farming.** *Genome Research*, v. 21, n. 8, p. 1339–1348, ago. 2011.

OPPENHEIM, S. J. *et al.* **We can't all be supermodels: The value of comparative transcriptomics to the study of non-model insects.** *Insect Molecular Biology*, v. 24, n. 2, p. 139–154, 1 abr. 2015.

PARK, Y. S. *et al.* **Comparison of library construction kits for mRNA sequencing in the Illumina platform.** *Genes and Genomics*, v. 41, n. 10, p. 1233–1240, 1 out. 2019.

PESCH, Y. Y. *et al.* **Chitinases and Imaginal disc growth factors organize the extracellular matrix formation at barrier tissues in insects.** *Scientific Reports* 2016 6:1, v. 6, n. 1, p. 1–14, 3 fev. 2016.

PESCH, Y. Y.; RIEDEL, D.; BEHR, M. ***Drosophila* Chitinase 2 is expressed in chitin producing organs for cuticle formation.** *Arthropod Structure and Development*, v. 46, n. 1, p. 4–12, 1 jan. 2017.

PETERSEN, T. N. *et al.* **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nature Methods* 2011 8:10, v. 8, n. 10, p. 785–786, 29 set. 2011.

PINTEREST-a. **Life Cycle of an Ant.** Disponível em:

<https://www.pinterest.pt/pin/518265869632336795/>

PINTEREST-b. **Dezembro 2020 – Jana Blog Infantil.** Disponível em:

<https://www.pinterest.pt/pin/632896553884026135/>

RODOVALHO, C. DE M. *et al.* **The Mitochondrial Genome of the Leaf-Cutter Ant *Atta laevigata*: A Mitogenome with a Large Number of Intergenic Spacers.** *PLOS ONE*, v. 9, n. 5, p. e97117, 14 maio 2014.

RODOVALHO, C. M. *et al.* **Expressed sequence tags from *Atta laevigata* and identification of candidate genes for the control of pest leaf-cutting ants.** *BMC Research Notes*, v. 4, n. 1, p. 203, 2011.

SCHIØTT, M.; BOOMSMA, J. J. **Proteomics reveals synergy between biomass degradation enzymes and inorganic fenton chemistry in leaf-cutting ant colonies.** *eLife*, v. 10, p. 1–39, 2021.

SCHULTZ, T. R. **Fungus-Farming Ants (*Attini* in Part).** Em: *Encyclopedia of Social Insects*. Cham: Springer International Publishing, 2021. p. 404–411.

SÉRGIO DE BRITTO, J. *et al.* **Use of alternatives to PFOS, its salts and PFOSF for the control of leaf-cutting ants *Atta* and *Acromyrmex*.** *Ijres*, v. 3, n. May, p. 11–92, 2016.

SCHULTZ, T. R. **Fungus-Farming Ants (*Attini* in Part).** Em: *Encyclopedia of Social Insects*. Cham: Springer International Publishing, 2021. p. 404–411.

SOLIS, D. R. *et al.* **On the morphology of the worker immatures of the leafcutter ant *Atta sexdens linnaeus* (Hymenoptera: Formicidae).** *Microscopy Research and Technique*, v. 75, n. 8, p. 1059–1065, ago. 2012.

SUEN, G. *et al.* **The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle.** *PLoS Genetics*, v. 7, n. 2, 1 fev. 2011.

TETREAU, G. *et al.* **Overview of chitin metabolism enzymes in *Manduca sexta*: Identification, domain organization, phylogenetic analysis, and gene expression.** *Insect Biochemistry and Molecular Biology*, v. 62, p. 114–126, 1 jul. 2015.

TRINITYRNASEQ-a. **Transcriptome Contig Nx and ExN50 Stats.** Disponível em: <https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Contig-Nx-and-ExN50-stats>

TRINITYRNASEQ-b. **Counting Full Length Trinity Transcripts.** Disponível em: <https://github.com/trinityrnaseq/trinityrnaseq/wiki/Counting-Full-Length-Trinity-Transcripts>

TUMESCHEIT, C.; FIRTH, A. E.; BROWN, K. **CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments.** *PeerJ*, v. 10, p. e12983, 15 mar. 2022.

UNIPROT-DATABASE. **Index of Uniprot databases.** Disponível em: [https://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/complete/](https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/)

VERLI, H. **Bioinformática da Biologia à flexibilidade molecular.** 1. ed. [s.l: s.n.].

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

WORSLEY, S. F. *et al.* **Symbiotic partnerships and their chemical interactions in the leafcutter ants (Hymenoptera: Formicidae).** 2018.

ZANETTI, R. *et al.* **An overview of integrated management of leaf-cutting ants (Hymenoptera: Formicidae) in Brazilian forest plantations.** *Forests*, v. 5, n. 3, p. 439–454, 2014.

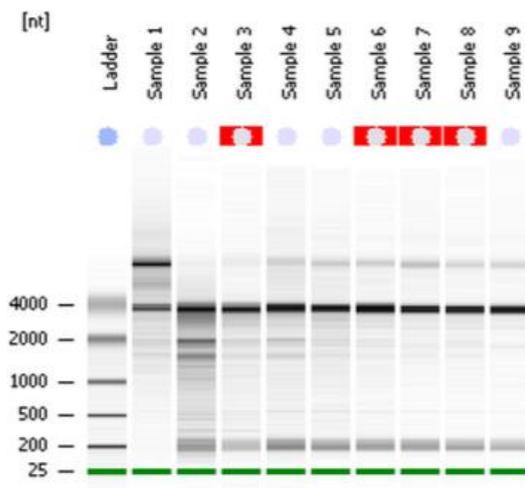
ZHANG, D. *et al.* **Functional analysis of two chitinase genes during the pupation and eclosion stages of the beet armyworm *Spodoptera exigua* by RNA interference.** *Archives of Insect Biochemistry and Physiology*, v. 79, n. 4–5, p. 220–234, 1 abr. 2012.

ZHU, Q. *et al.* **Functional specialization among insect chitinase family genes revealed by RNA interference.** *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 18, p. 6650–6655, 6 maio 2008.

## APÊNDICE A - VALORES DE INTEGRIDADE DE RNA

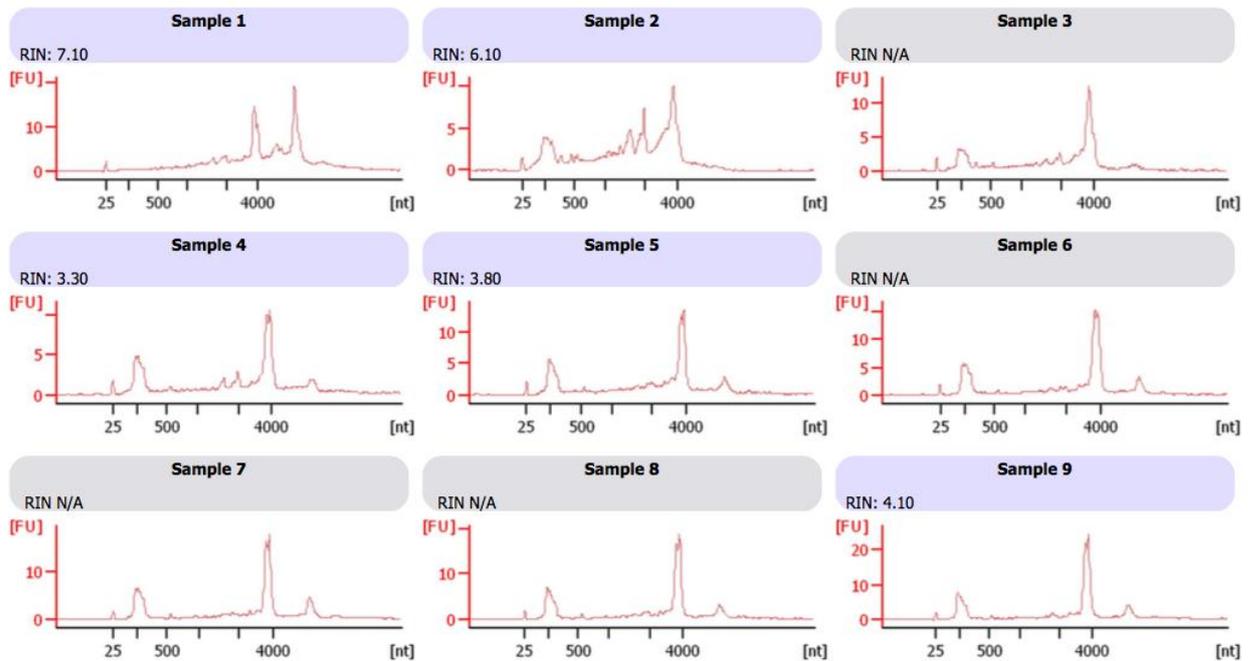
A análise de qualidade do RNA foi realizada pelo método de eletroforese microfluídica com o equipamento Bioanalyzer 2100 (Agilent).

### Análise 01 – Amostras formiga forrageadora

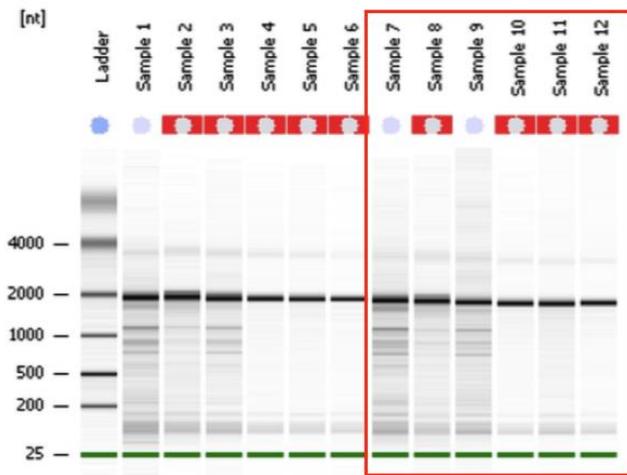


#### Amostras Tratadas:

- 01 - Intestino Formiga 01
- 02 - Intestino Formiga 02
- 03 - Intestino Formiga 03
- 04 - Formiga 01
- 05 - Formiga 02
- 06 - Formiga 03



### Análise 02 – Amostras larvas



**Amostras Tratadas:**

- 07 - Intestino Larva 01
- 08 - Intestino Larva 02
- 09 - Intestino Larva 03
- 10 - Larva 01
- 11 - Larva 02
- 12 - Larva 03

