Luciana Moreira Martins

Análise do transcriptoma de *Podalia orsilochus* (Cramer, 1775)

Tese apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia USP/Instituto Butantan/IPT, para obtenção do Título de Doutor em Biotecnologia.

São Paulo
2016

Luciana Moreira Martins

Análise do transcriptoma de *Podalia orsilochus* (Cramer, 1775)

Tese apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia USP/Instituto Butan tan/IPT, para obtenção do Título de Doutor em Biotecnologia.

Área de concentração: Biotecno logia

Orientador: Prof. Dr. Ronaldo Zucatelli Mendonça

Versão original

São Paulo
2016

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e Informação Biomédica do
Instituto de Ciências Biomédicas da Universidade de São Paulo

reprodução não autorizada pelo autor

Martins, Luciana Moreira.
Análise do transcriptoma de *Podalia orsilochus* (Cramer, 1775) / Luciana Moreira Martins. -- São Paulo, 2016.

Orientador: Prof. Dr. Ronaldo Zucatelli Mendonça.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan. Área de concentração: Biotecnologia. Linha de pesquisa: Transcriptômica.

Versão do título para o inglês: Transcriptome analysis of *Podalia orsilochus* (Cramer, 1775).

1. Lepidoptera 2. Tegumento 3. Transcriptoma I. Mendonça, Prof. Dr. Ronaldo Zucatelli II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan III. Título.

ICB/SBIB018/2016

UNIVERSIDADE DE SÃO PAULO

Programa de Pós-Graduação Interunidades em Biotecnologia

Universidade de São Paulo, Instituto Butantan, Instituto de Pesquisas Tecnológicas

---

Candidato(a): Luciana Moreira Martins.

Título da Tese: Análise do transcriptoma de Podalia orsilochus (Cramer, 1775).

Orientador(a): Prof. Dr. Ronaldo Zucatelli Mendonça.

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a ...../...../....., considerou

( ) Aprovado(a) **( ) Reprovado(a)**

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome:.....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Presidente: Assinatura: .....
Nome: .....
Instituição: .....

INSTITUTO BUTANTAN

COMISSÃO DE ÉTICA NO USO DE ANIMAIS
INSTITUTO BUTANTAN

Av. Vital Brazil, 1500, CEP 05503-900, São Paulo, SP, Brazil
Telefone: (55) (011) 2627-9585 - Fax: (55) (011) 2627-9505
ceuaib@butantan.gov.br

São Paulo, 22 de junho de 2012.

Protocolo: I-934/12

Dr. Ronaldo Zucatelli Mendonça
Laboratório de Parasitologia
Instituto Butantan

Referente Projeto: "Construção de biblioteca de cDNA de lagartas da família Megalopygidae".

A Comissão de Ética no Uso de Animais do Instituto Butantan (CEUAIB) informa que o projeto acima referido é isento de análise por parte desta Comissão, pois não faz uso de animais vertebrados em sua execução.

Cordialmente,

Dr. Marcelo Larami Santoro
Coordenador da CEUAIB

DEDICATÓRIA

Esse trabalho é dedicado principalmente ao meu amado filho Heitor. Espero que um dia saiba que todas as minhas lutas e conquistas após a sua chegada foram incentivadas, motivadas, dedicadas e, com certeza, alcançadas por sua causa. Todo sentimento, palavra e atitude boa que nasce em mim provêm de você. Meu amor por você é incondicional e eterno.

A minha mãe Vanda por ser exatamente tudo o que eu gostaria de ser também como mãe para meu filho. Uma mulher guerreira, paciente, encantadora, companheira, dedicada, atenciosa, carinhosa, um coração gigantesco com todos e que sempre arranja um espaço no seu colo quando preciso. Você sempre foi e sempre será o meu maior exemplo e tenho um imenso orgulho de ser sua filha.

Ao meu esposo, amigo, amante, confidente e companheiro de todas as horas, Abner. Deus não poderia ter escolhido uma pessoa melhor para partilhar minha vida e, por isso, agradeço todos os minutos por ter você ao meu lado. Você é meu porto seguro, o lugar pra onde sempre quero voltar e estar.... Obrigado por me apoiar e me incentivar em todas as minhas decisões, pois sem você nada disso estaria acontecendo. Amo você mais do que possa imaginar.

AGRADECIMENTOS

Meu maior agradecimento ao meu orientador Ronaldo Zucatelli Mendonça, principalmente pela confiança que sempre depositou em mim e no meu trabalho. Obrigado por ser sempre essa pessoa cheia de luz e pensamentos positivos, que nos invade a alma e nos deixa prontos a realizar qualquer tarefa, independente do grau de dificuldade.

Agradecimentos especiais aos amigos que de certa forma foram uma espécie de "co-orientadores" Carol, Enéas e Fernando Luiz Kamitani, pela ajuda, paciência, pelo tempo dispendido, pelas correções e por me acalmar em muitos momentos. Sem a ajuda de vocês este trabalho não existiria, obrigada por tudo.

Agradeço aos meus irmãos que sempre estão desejando minhas melhores conquistas, me apoiando nos momentos de dificuldade e me alegrando com suas presenças sempre que possível. Obrigado por fazerem parte da minha vida. Amo muito vocês.

Aos meus sobrinhos Victor e Marcelo, que são uma dádiva em nossa família, enchendo a casa de alegria, de diversão e também de loucuras e preocupação, rsrsrs... Espero poder sempre tê-los próximos, mesmo que a distância acabe nos afastando. Sou uma tia apaixonada por vocês, meus anjinhos.

As melhores amigas, irmãs, primas e companheiras que alguém poderia ter Ana Luiza, Alana, Monise, Paola, Greta, Bianca e, em especial, minha prima Juliana Martins e minha filhinha de coração Juliana Badari. Obrigado pelas risadas, brincadeiras, carinho e por estarem sempre ao meu lado, me compreenderem e me apoiarem. Vocês são muito mais do que especiais para mim.

Agradeço à Fundação de Amparo a Pesquisa do Estado de São Paulo pelo apoio financeiro.

Agradeço a todos do laboratório Especial de toxinologia aplicada por toda a cooperação e atenção que tiveram no início desse trabalho. Ele nunca teria acontecido sem a participação de vocês.

Agradeço todos os companheiros e amigos que fiz durante meus anos de trabalho no Instituto Butantan, que sempre foram como uma família para mim e tenho certeza que essa amizade permanecerá durante a vida toda.

E finalmente agradeço a Deus, por me dar força para terminar todos os projetos iniados e por ter me dado o Heitor, que é o maior incentivo do mundo para querer sempre ser e fazer o melhor.

“Suba o primeiro degrau com fé. Não é necessário que você veja toda a escada. Apenas dê o primeiro passo.”

Martin Luther King

RESUMO

MARTINS, L. M. **Análise do transcriptoma de *Podalia orsilochus* (Cramer, 1775).** 2016. 131 f. Tese (Doutorado em Biotecnologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2016.

A classe dos insetos representa 55 % de toda a biodiversidade existente na terra e aproximadamente 85 % de todos os animais. Os insetos são capazes de sobreviver em diversos ecossistemas do planeta, com exceção ao ambiente marinho profundo e as regiões polares. Assim como todos os seres vivos, os insetos estão constantemente expostos à ameaça de infecção microbiana, contudo, permanecem livres de infecções na maior parte do tempo. Essa capacidade de sobrevivência aliada à larga distribuição dos insetos em regiões com características totalmente diferentes tem estimulado as pesquisas de novos agentes terapêuticos nesta classe de artrópodos, uma vez que já foram encontrados diversos componentes de mecanismos inespecíficos de combate à infecção, sendo possível sua aplicação no controle de diversas doenças. Todavia, apesar de um grande número de moléculas envolvidas em respostas de defesa ter sido identificado a partir de diferentes espécies de insetos, pouca informação sobre suas aplicações está disponível. Desta maneira, conhecer os constituintes moleculares que são vitais para sobrevivência dos insetos, tornou-se meta de diversos trabalhos que buscam elucidar e descobrir novas moléculas bioativas de interesse, visando à produção de biofármacos mais eficazes e de menores custos, e de potencial para a elaboração de novos métodos de controle de doenças. Desta forma, o presente trabalho elucida o perfil transcriptômico geral da espécie *Podalia orsilochus* (Cramer, 1775), que representa um lepidóptero presente na fauna brasileira e neotropical, que há tempos é conhecido por sua importância médica em casos de erucismo, mas pouco se sabe sobre sua biologia e ecologia, não existindo estudos que visam conhecer os mecanismos biológicos envolvidos na sua resposta de defesa contra os diversos microrganismos que podem causar infecções em seu habitat natural. Para tanto, estudamos os principais transcritos do tegumento desta espécie através de uma análise transcriptômica dos genes expressos nesse tecido durante sua fase larval. Para a realização deste trabalho o mRNA do tegumento da *P. orsilochus* foi extraído e, a partir desta amostra, foram realizados diversos protocolos com a finalidade da obtenção do transcriptoma. O pirosequenciamento da amostra foi realizado, seguido pelas montagens dos *contigs* através do programa CLC-Genomics e pelas análises das sequências geradas através do programa Blast2GO. Com os dados gerados através das análises dos *contigs* foi possível inferir as sequências predominantes, os números de sequências com funções moleculares e localizações relacionadas, assim como a porcentagem de sequências com a presença de peptídeo sinal para secreção de proteína, sinal de proteína transmembrânica e as classes de enzimas encontradas nos transcritos provenientes de *P. orsilochus*. A partir dessas informações, as sequências encontradas foram categorizadas e alguns transcritos foram selecionados para uma avaliação mais aprofundada. Os dados preliminares obtidos nesse trabalho sugerem a presença de diversos fatores envolvidos em processos ligados à imunidade, tendo sido encontrados transcritos regulatórios da resposta imune, representantes de famílias de AMPs, inclusive um AMP encontrado apenas em espécies de Lepidópteros. Como consequência, os transcritos obtidos permitirão o auxílio em pesquisas posteriores, seja para comparação, citação ou conhecimento biológico e relações de filogenia do animal.

Palavras-chave: Podalia. Transcriptoma. Proteínas. Imunidade inata.

ABSTRACT

MARTINS, L. M. **Transcriptome analysis of *Podalia orsilochus* (Cramer, 1775).** 2016. 131 p. Masters thesis (Biotechnology) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2016.

The class of insects is 55% of the existing biodiversity on land and about 85% of all animals. The insects are able to survive in diverse ecosystems on Earth, except the deep marine environment and the polar regions. As with all living beings, the insects are constantly exposed to the threat of microbial infection, however, remain free of infection for most of the time. This survivability combined with the wide distribution of insects in areas with totally different characteristics has stimulated research into new therapeutic agents this arthropod class, as have been found various components of nonspecific mechanisms to fight infection, making it possible its application in control various diseases. However, despite a large number of molecules involved in defense responses have been identified from various species of insects, little information is available on their applications. In this way, knowing the molecular constituents which are vital for survival of insects, it has become target of several works that explain and discover new bioactive molecules of interest, aimed at producing more effective biopharmaceuticals and lower costs, and potential for development of new disease control methods. Thus, this study elucidates the general transcriptomic profile of the kind *Podalia orsilochus* (Cramer, 1775), which is a gift lepidopteran the Brazilian and Neotropical fauna, which has long been known for its medical importance in cases of erucism, but little is known about their biology and ecology, there are no studies that aim to understand the biological mechanisms involved in its defense response against various microorganisms that can cause infections in their natural habitat. We studied the major transcripts of the integument of this species through a transcriptomic analysis of genes expressed in that body during its larval stage. For this work the mRNA of the seed coat of *P. orsilochus* was extracted and, from this sample were conducted various protocols for the purpose of obtaining the transcriptome. The sample pyrosequencing was carried out, followed by the assembly of contigs using CLC-Genomics program and the analysis of sequences generated through Blast2GO program. With the data generated through the analysis of the contigs was possible to infer the predominant sequences, the number of sequences with molecular related functions and locations, as well as the percentage of sequence with the presence of signal peptide for protein secretion transmembrane protein signal and classes of enzymes found in the transcripts from *P. orsilochus*. From this information, the sequences found were categorized and some transcripts were selected for further evaluation. Preliminary data obtained in this work suggest the presence of several factors involved in processes related to immunity, having been found regulatory transcripts of the immune response, representatives of MPAs families, including an AMP found only in species of Lepidoptera. As a result, transcripts obtained will aid in later searches, either for comparison or reference biological knowledge and animal phylogenetic relationships.

Keywords: Podalia. Transcriptome. Proteins. Innate immunity.

LISTA DE FIGURAS

Figura 1 - Regiões biogeográficas onde se distribuem as espécies da ordem Lepidoptera..... 20
Figura 2 - Cladograma do grupo Limacodidae (Zygaenoidea) baseado na análise de parcimônia de caracteres de adultos e imaturos, evidenciando a família Megalopygidae..... 26
Figura 3 - Exemplos da variação cromática das asas de algumas espécies membros da família Megalopygidae..... 28
Figura 4 - Morfologia externa das lagartas. 30
Figura 5 - Lagartas da família Megalopygidae..... 31
Figura 6 - Representante de cada fase do ciclo da espécie Podalia orsilochus (Cramer, 1775). 34
Figura 7 - Representação esquemática das respostas imunológicas dos insetos..... 36
Figura 8 - Intersecção de uma lagarta, mostrando as estruturas internas. Em destaque o corpo gorduroso..... 37
Figura 9 – Representação das quatro famílias estruturais de peptídeos antimicrobianos (AMPs)..... 38
Figura 10 - Fontes de peptídeos antimicrobianos (AMPs). 38
Figura 11 - Agrupamento de sequências..... 59
Figura 12 - Alinhamento global e local entre diferentes sequências..... 61
Figura 13 - Categorias funcionais dos COGs..... 73
Figura 14 - Resultados das avaliações de tamanho e fragmentação da amostra de mRNA através do Bioanalyzer..... 76
Figura 15 - Resultado da avaliação do cDNA através do Bioanalyzer. 76
Figura 16 - Imagens captadas pelo 454 GS-Junior, mostrando a microplaca inteira e ampliada. 77
Figura 17 - Imagens captadas pelo 454 GS-Junior, mostrando a microplaca inteira e ampliada. 78
Figura 18 - Gráfico de barras representando o resultado do BLAST (BLASTx hits) e da anotação (GO hits), realizado pelo programa Blast2GO. 80
Figura 19 – Gráfico da distribuição de similaridades entre as sequências de Podalia orsilochus e as sequências alinhadas através do BLAST..... 81
Figura 20 - Gráfico com a distribuição das 12 espécies que mais vezes apresentaram similaridade com as sequências de *Podalia orsilochus*. 81
Figura 21 - Gráfico com a distribuição das espécies que apresentaram os melhores hits em comparação com as sequências de *Podalia orsilochus*..... 82
Figura 22 – Termos GO mais encontrados do domínio componente celular e a quantidade de sequências de Podalia orsilochus associadas a esses termos..... 84
Figura 23 – Termos GO mais encontrados do domínio componente celular e a quantidade de sequências de Podalia orsilochus associadas a esses termos..... 85

Figura 24 - Gráfico de distribuição das sequências associadas aos termos GO de funções moleculares. 86
Figura 25 - Gráfico de distribuição das sequências associadas aos termos GO do domínio Processos biológicos. 87
Figura 26 - Gráfico demonstrando a porcentagem de sequências que apresentaram predições da presença de peptídeo sinal de secreção de proteína e da presença de sequências características de proteína transmembrana através do recurso InterPro. 88
Figura 27 - Gráfico dos dados da porcentagem de predição de sinal de proteína secretada entre os 6.659 contigs analisados. 88
Figura 28 - Gráfico dos dados de porcentagem de predição de sinal de proteína transmembrânica entre os 6.659 contigs analisados. 89
Figura 29 – Termos gerais dos códigos de classificação das enzimas (EC). 90
Figura 30 – Gráfico com a porcentagem de distribuição de contigs correspondente a cada categoria funcional. 92
Figura 31 – Gráficos das porcentagens de contigs incluídos em cada subcategoria dentro de cada categoria geral. 93
Figura 32 - Distribuição dos transcritos relacionados à imunidade em *Podalia orsilochus*. 96
Figura 33 - Porcentagem dos peptídeos antimicrobianos encontrados nas sequências de *Podalia orsilochus*. 99

LISTAS DE TABELAS

Tabela 1 - Distribuição regional da diversidade de espécies de Lepidoptera. 21
Tabela 2 - Lista das superfamílias da ordem Lepidoptera 25
Tabela 3 - Subprogramas BLAST utilizados de acordo com o formato de entrada de sequência e banco de dados desejados. 63
Tabela 4 - Dados das contagens de poços obtidos no sequenciamento. 78
Tabela 5 - Análise das sequências e montagem dos contigs do tegumento de *Podalia orsilochus*. 79
Tabela 6 – Classificação dos contigs em categorias funcionais. 91
Tabela 7 - Transcritos do tegumento de Podalia orsilochus que possuem altos níveis de expressão gênica (maior frequências em número de reads*), totalizando os contigs 50 mais expressos. 94
Tabela 8 – Dados do subconjunto de contigs da Imunidade (totalizando 39 contigs). 96
Tabela 9 - Abundância em reads dos prováveis contigs de peptídeos antimicrobianos e os números de isômeros encontrados em *Podalia orsilochus*. 98

LISTA DE ABREVIATURAS E SIGLAS

AA = aminoácido

AMP = peptídeo antimicrobiano

APD = The Antimicrobial Peptide Database

BLAST = Basic Local Alignment Search Tool

CDD = Conserved Domains Database

cDNA = DNA complementar

DNA = ácido desoxirribonucléico

ESTs = Expressed Sequence Tags (marcadores de sequências expressas)

GO = Gene Ontology Consortium

InterPro = Protein Sequence Analysis & Classification

KEGG = Kyoto Encyclopedia of Genes and Genomes

mRNA = RNA mensageiro

NCBI = National Center for Biotechnology Information

pb = pares de base

PBS = phosphate buffered saline solution (solução salina tamponada com fosfato)

PCR = Polymerase Chain Reaction (reação em cadeia da polimerase)

PPi = composto de pirofosfato

RNA = ácido ribonucléico

RNA*later* = *RNA Stabilization Solution* (solução de estabilização do RNA)

rRNA = RNA ribossomal

UniprotKB = Universe Protein Resource Knowledgebase

SUMÁRIO

1 INTRODUÇÃO 18

2 REVISÃO DE LITERATURA 20

2.1 A ordem Lepidoptera 20

2.1.1 Um breve histórico sobre filogenia, taxonomia e classificações atuais da ordem Lepidoptera 24

2.2 A família Megalopygidae 27

2.2.1 Ecologia e características morfológicas dos Megalopygideos 28

2.3 A espécie *Podalia orsilochus* (Cramer, 1775) 33

2.5 A imunidade nos insetos 34

2.5.1 Peptídeos antimicrobianos de insetos 38

2.5.1.1 Classe dos AMPs α-hélices e privados de resíduos de cisteína 40

2.5.1.2 Classe dos AMPs cíclicos contendo resíduos de cisteína 41

2.5.1.3 Classe dos AMPs com rica representação de prolina 43

2.5.1.4 Classe dos AMPs com rica representação de glicina 45

2.5.2 Potenciais aplicações de AMPs de insetos 47

2.5.3 Exploração da biodiversidade em busca de moléculas biologicamente ativas 50

2.6 A pesquisa genômica 51

2.6.1 Sequenciamento de última geração – NGS 53

2.6.1.1 Considerações sobre o método de pirosequenciamento 54

2.6.2 Transcriptoma de lepidópteros 55

2.7 Uma visão global da bioinformática 58

2.7.1 Agrupamento de sequências 58

2.7.2 Bancos de dados em biologia molecular 59

2.7.3 Alinhamento de sequências 60

2.7.4 Anotação de genomas 63

3 OBJETIVO 66

3.1 Objetivo Geral 66

3.2 Objetivos específicos 66

4 MATERIAL E MÉTODOS 67

4.1 Obtenção e preparo do lisado do tegumento da lagarta *Podalia orsilochus.* 67

4.2 Isolamento e purificação do RNA mensageiro 67

4.2.1 Preparação de Dynabeads Oligo (dT)25 68
4.2.2 Isolamento do RNA mensageiro 68
4.2.3 Eliminação da contaminação com RNA ribossomal 69
4.3 Avaliação da quantidade e qualidade do RNA extraído 69
4.4 Pirosequenciamento 69
4.5. Bioinformática 70
4.5.1 Remoção das sequências de RNA ribossomal 71
4.5.2 Montagem de novo dos contigs 71
4.5.3 Comparação dos contigs através do BLAST 71
4.5.4 Anotação funcional 72
4.5.4.1 Classificação dos contigs em categorias funcionais 73
4.5.4.2 Análise dos contigs da subcategoria funcional Imunidade 74
5 RESULTADOS 74
5.1 Obtenção do RNA mensageiro do tegumento da *Podalia orsilochus* 75
5.1.1 Avaliação da qualidade do mRNA 75
5.2 Pirosequenciamento 76
5.2.1 PCR em emulsão, montagem da microplaca e sequenciamento 77
5.3 Tratamento das sequências e montagem dos *contigs* 78
5.4 Anotação automatizada: função, ontologia gênica e sublocalização celular 79
5.4.1 Identificação dos transcritos e dados da anotação 79
5.4.2 Ontologia gênica (GO) 83
5.5 Predição da presença de peptídeo sinal e de segmentos transmembrânicos 87
5.6 Classificação enzimática 89
5.7 Classificação dos *contigs* em categorias funcionais 90
5.7.1 Análise contigs com maior número de reads. 94
5.8 Análise dos contigs da subcategoria Imunidade 95
5.8.1 Grupos funcionais dos contigs da subcategoria Imunidade 98
5.8.1.1 Grupo I- Peptídeo antimicrobiano (AMP) 98
5.8.1.2 Grupo II - Receptores de reconhecimento de padrões (PRRs) e de vias de sinalização 100
5.8.1.3 Grupo III - Proteases e inibidores de proteases 100
5.8.1.4 Grupo IV – Prováveis proteínas de defesa com função desconhecida 101

6 DISCUSSÃO 101
7 CONCLUSÕES 113
REFERÊNCIAS* 115

1 INTRODUÇÃO

Os seres humanos e outros seres vivos estão constantemente expostos à ameaça de infecção microbiana, acreditava-se que grande parte dessa proteção contra essas infecções era fornecida exclusivamente pelo sistema imune adaptativo. No entanto, levantou-se a questão sobre como as plantas e insetos, que carecem de um sistema imunitário adaptativo, também permanecem livres de infecções na maior parte do tempo. A resposta para essa pergunta pode estar em mecanismos de ação e moléculas bioativas que atuam em uma resposta de defesa inespecífica, realizada pelo sistema imune inato (PHOENIX; DENNISON; HARRIS, 2013).

Devido ao rápido desenvolvimento de bactérias patogênicas resistentes aos antibióticos clássicos nos últimos anos (MOELLERING, 1998), estudos intensivos foram realizados em busca de antimicrobianos mais eficazes (ANDREU; RIVAS, 1999; LÖWENBERGER, 2001).

Os insetos atraem a atenção dos cientistas tanto para a compreensão da biologia básica do sistema imunológico como para a procura de modelos moleculares para desenho de novas drogas antibióticas, uma vez que já foram encontrados diversos componentes de mecanismos inespecíficos de combate à infecção, sendo possível sua aplicação no controle de diversas doenças. Contudo, apesar de um grande número de moléculas envolvidas em respostas de defesa ter sido identificado a partir de diferentes espécies de insetos, pouca informação sobre suas aplicações está disponível, assim como os fatores envolvidos nas suas atividades e o local específico de ação (TIAN et al., 2010; ZHAO et al., 2010).

Desta maneira, conhecer os constituintes moleculares, como genes e/ou proteínas, que são vitais para sobrevivência dos insetos, tornou-se meta de diversos trabalhos. Trabalhos esses que buscam elucidar e descobrir novas moléculas de interesse biotecnológico e farmacológico, visando à produção de biofármacos mais eficazes e de menores custos, e de potencial para a elaboração de novos métodos de controle de doenças (PAUCHET et al., 2010; POIVET et al., 2013; VOGEL et al., 2011).

Tecidos de diversos insetos são coletados e processados por métodos moleculares e bioquímicos, e analisados em abordagens genômica, transcriptômica e proteômica. Todos esses estudos visam descobrir moléculas de interesse biotecnológico ou simplesmente genes com histórias evolutivas que poderiam solucionar determinadas definições sistemáticas,

filogenéticas e taxômicas de um grupo específico (PAUCHET et al., 2010; POIVET et al., 2013; VOGEL et al., 2011).

Desta maneira, o presente trabalho elucida o perfil transcriptômico geral de uma espécie de Lepidóptero presente na fauna brasileira e neotropical, que é há tempos conhecido por sua importância médica em casos de erucismo, mas pouco se sabe sobre sua biologia e ecologia, não existindo estudos que visam conhecer os mecanismos biológicos envolvidos na sua resposta de defesa contra os diversos microrganismos que podem causar infecções em seu habitat natural. Para tanto, estudamos os principais transcritos do tegumento da espécie de *Podalia orsilochus* (Cramer, 1775) através de uma análise transcriptômica dos genes expressos nesse tecido durante sua fase larval.

2 REVISÃO DE LITERATURA

2.1 A ordem Lepidoptera

A ordem de insetos Lepidoptera representa a segunda maior ordem em número de espécies de insetos, com uma estimativa de cerca de 160.000 espécies descritas desde 1758 (CHAPMAN, 2009; MALLET, 2007). Contudo, a estimativa total mundial varia de 255.000 (HEPPNER, 1991) a 500.000 espécies (KRISTENSEN; SKALSKI, 1999). De acordo com algumas estimativas mais recentes, o número de espécies de Lepidoptera descritas só pode ser excedido por besouros e plantas com flores (VAN NIEUKERKEN et al., 2011).

Os lepidópteros possuem a segunda maior diversidade de insetos do planeta e podem ser encontrados em quase todas as regiões do mundo e em grande variedade de habitats, principalmente em regiões tropicais (Figura 1) (ISERHARD; ROMANOWSKI, 2004). A Tabela 1 mostra os números globais das espécies descritas em cada região biogeográfica e a média de espécies encontradas por área. A região Neotropical conta com número aproximado de 45.000 espécies (Tabela 1), porém, estima-se que a diversidade real de espécies nessa região biogeográfica seja entre 71.000 e 79.000 espécies, pois o número de espécies descritas certamente é provisório, devido a espécies frequentemente encontradas ainda não inventariadas e a pequena representatividade de áreas amostradas (SILVA et al., 2011). Apesar das pesquisas sobre a diversidade biológica serem consideradas incipientes, somente no Brasil, mais de 5.000 espécies já foi descrita (CARDOSO; HADDAD JR, 2005; ISERHARD; ROMANOWSKI, 2004; SILVA et al., 2011).

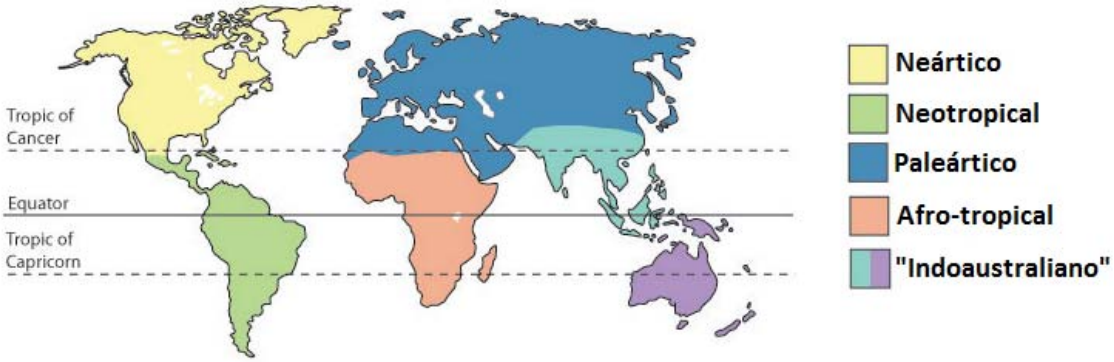


Figura 1 - Regiões biogeográficas onde se distribuem as espécies da ordem Lepidoptera.
Fonte: http://www.nature.com/scitable/knowledge/library/factors-affecting-global-climate-17079163.

Acesso em: 13/11/2015.

Tabela 1 - Distribuição regional da diversidade de espécies de Lepidoptera.

	PAL	NEA	NET	AFT	INA
Espécies	22,465	11,532	44,791	20,491	47,286
Espécies/área	1.2	1.3	6.3	2.1	6.3

PAL: Palaeártico; NEA: Neártico; NET: Neotropical; AFT: Afro-tropical; INA: 'Indoaustraliano'.
Fonte: Adaptação de Heppner, J. B. (1991).

O grupo inclui insetos com dois pares de asas membranosas cobertas de escamas e peças bucais adaptadas à sucção. Existe grande diversidade de formas e comportamentos dentre os lepidópteros. É comum em várias espécies, tanto na fase adulta quanto na larval, o uso de aposematismo, mimetismo e camuflagem, como meios de defesa contra predadores (LEPESQUEUR, 2012).

Muitos Lepidópteros são voadores eficazes, em comum com libélulas e alguns grupos de moscas, vespas e abelhas, mas diferentemente da maioria dos grupos de insetos voadores, que são voadores menos eficazes ou voam pouco. Assim, o poder de dispersão desses insetos varia, bem como a capacidade dos Lepidópteros e outros insetos voadores colonizarem novos habitats ou ilhas ser qualitativamente maior do que para outros grupos de insetos (COCK, 2003).

Informalmente a ordem é dividida em dois grupos: Rhophalocera, que compreende o grupo de insetos conhecidos como borboletas, e Heterocera, que compreende o grupo de insetos conhecidos como mariposas (CARDOSO; HADDAD JR, 2005). A divisão da ordem Lepidoptera em Rhophalocera e Heterocera é baseada em diferentes critérios.

O grupo Rhophalocera ou borboletas apresenta como principais características a atividade diurna, o corpo fino e alongado, as antenas retilíneas, as asas geralmente coloridas, com as marcações sendo muitas vezes em ambas as superfícies superior e inferior das asas, o dobramento das asas para cima quando em repouso e a metamorfose ocorrer dentro de uma estrutura rígida denominada crisálida (CARDOSO; HADDAD JR, 2005; COCK, 2003).

O grupo Heterocera ou mariposas apresenta como principais características a atividade maioritariamente noturna ou crepuscular, o corpo arredondado e robusto, as antenas bastante variadas conforme a espécie, mas normalmente em forma de plumas, as

asas com coloração monótona, com as marcações em grande parte restritas a superfície superior da asa, pouso com as asas abertas e a metamorfose do grupo ocorrer dentro de um casulo mole, geralmente formado pelas cerdas presentes nas larvas (CARDOSO; HADDAD JR, 2005; COCK, 2003).

A proporção de espécies de lepidópteros é de uma borboleta para 10 mariposas, mas apesar de toda a diversidade de mariposas, o grupo apresenta inúmeros problemas relacionados à taxonomia, à biologia dos imaturos e suas relações ecológicas com as características das plantas hospedeiras (LEPESQUEUR, 2012).

Todos os lepidópteros têm o mesmo tipo de desenvolvimento, passando pelas fases de ovo, larva ou lagarta, pupa ou crisálida e adulto (imago), ou seja, apresentam evolução completa ou holometabólica variando apenas os períodos conforme a espécie. As lagartas são os representantes dos lepidópteros que se encontram na fase larval (CARDOSO; HADDAD JR, 2005) e iniciam seu ciclo biológico através da eclosão dos ovos provenientes do acasalamento dos adultos (borboletas e mariposas machos e fêmeas). Após a cópula fazem as posturas de seus ovos nas folhas e galhos de árvores frutíferas comestíveis, como o abacateiro, ameixeira, pessegueiro, comuns em pomares ou mesmo em árvores nativas dentro das matas. Após um tempo variado de acordo com a espécie, as larvas eclodem dos ovos, passando a se alimentar das folhas da planta hospedeira (LEPESQUEUR, 2012).

Os lepidópteros, durante a fase larval, constituem a maior radiação entre os insetos fitófagos (SCOBLE, 1995) e podem ser utilizados para estudos ambientais sobre a biodiversidade pela importância na dinâmica de ecossistemas, agindo como desfolhadores, decompositores, presas, hospedeiros e polinizadores. Além disso, os lepidópteros possuem o *status* de bioindicadores de qualidade de vegetação em diferentes tipos de habitats, respondendo rapidamente às alterações ambientais (KITCHING et al., 2000).

Esses animais geralmente vivem em grupos, podendo ser encontrados nos troncos das árvores durante o dia. A maioria dos lepidópteros não causa prejuízos ao homem. As lagartas podem por sua vez causar danos em lavouras para se alimentar, mas fertilizam o solo com as fezes, e algumas são importantes comercialmente, como a *Bombix mori*, que é conhecida como bicho-da-seda. Durante o período larval, que varia de acordo com a espécie, as lagartas trocam de pele (ecdise) várias vezes, aumentando consideravelmente de tamanho (CARDOSO; HADDAD JR, 2005).

Após a última ecdise, elas se transformam em pupas, mantendo-se nesta fase imóveis, por período variado conforme a espécie e alojando-se preferencialmente na base das árvores sob o húmus, onde a umidade é alta. Esta é uma fase importantíssima no desenvolvimento do lepidóptero devido às mudanças morfológicas e fisiológicas que ocorrem. Após o período de pupa, emergem os adultos machos e fêmeas, alados, reiniciando o ciclo biológico (CARDOSO; HADDAD JR, 2005).

Muitos dos lepidópteros adultos possuem um tempo de vida extremamente curto, vivendo em média 15 dias, sendo em muitos casos apenas o período necessário para a reprodução. Geralmente nestes casos, o animal adulto possui o aparelho bucal atrofiado, por isso não se alimentam, sendo necessário acumular toda a energia para viver essa fase durante o estágio larval. As espécies que se alimentam na fase adulta o fazem através da sucção de líquidos, geralmente o néctar de flores ou o sumo de frutos em decomposição. Muitas espécies tem papel fundamental na polinização de diversas plantas (KITCHING et al., 2000).

Muitas famílias formam a ordem Lepidoptera e poucas efetivamente causam problemas ao homem. Os acidentes causados por lepidópteros são divididos em dois tipos, lepidoterismo e erucismo. A palavra Lepidoptera provém do grego, sendo que *lepis*, *idos* e *ptera*, significa asa escamosa, por isso, o termo lepdopterismo se destina aos acidentes desencadeados pelo contato com as formas adultas aladas de borboletas e mariposas. As dermatoses provocadas pelo contato com os exemplares alados são mais raras, sendo escassa a literatura a respeito (CARDOSO; HADDAD JR, 2005).

O erucismo (do latim eruca = larva) trata-se de acidente mais comum e mais grave, desencadeados pelo contato com algumas lagartas, que como forma de defesa para evitar a predação da espécie, apresentam toxinas nas cerdas que revestem seu corpo. Com os seres humanos os contatos são acidentais ou por motivos profissionais. Embora esse tipo de acidente seja subnotificado, é sem dúvida o acidente por animal peçonhento mais comum no país (CARDOSO; HADDAD JR, 2005; NETO, 2003).

O erucismo está associado principalmente a três famílias de mariposas na fase larval: Arctiidae, Saturniidae e Megalopygidae e o formato das cerdas permite o reconhecimento de duas das principais famílias envolvidas com acidentes: a família Saturniidae, que apresenta cerdas esparsas cujo aspecto lembra pequenos pinheiros, fator característico e decisivo para a identificação da família. Os gêneros importantes são: *Automeris*, *Dirphia,*

Hylesia e Lonomia; já a família Megalopygidae apresenta cerdas finas e abundantes em todo o corpo. Os gêneros mais importantes da família Megalopygidae são *Podalia* e *Megalopyge* e entre as espécies de interesse desta família podemos citar: *Podalia orsilochus*, conhecida como Taturana-gatinho ou Taturana-cachorinho, *Megalopyge lanata* e *Megalopyge albicollis* (CARDOSO; HADDAD JR, 2005; MORAES, 2003).

2.1.1 Um breve histórico sobre filogenia, taxonomia e classificações atuais da ordem Lepidoptera

A ordem Lepidoptera constitui um grupo monofilético originado de um ancestral do período Jurássico, há cerca de 200 milhões de anos (GRIMALDI; ENGEL 2005). A monofilia do grupo está firmemente estabelecida por um conjunto de sinapomorfias de suas linhagens basais constituintes. Várias apomorfias indiscutivelmente exclusivas dos lepidópteros já foram identificadas e a grande maioria dessas características em questão ocorre no inseto adulto. A posição do grupo dentro da hierarquia de insetos é igualmente bem estabelecida (KRISTENSEN et al., 2007; KRISTENSEN; SKALSKI 1999).

Apesar da hierarquia e da monofilia do grupo ser bem estabelecida, os representantes dessa ordem são muito mais homogêneos, estruturalmente e ecologicamente, do que outros grupos de insetos. A combinação da pouca diversidade estrutural e o grande número de espécies são fatos que podem impedir o avanço no tratamento taxonômico de um grupo animal (KRISTENSEN et al., 2007, LEPESQUEUR, 2012).

Alguns pontos-chave na história da classificação dos lepidópteros foram dados por Scoble (1992), porém, o mais proeminente marco no início da classificação foi dado por Linnaeus, com a publicação da 10ª edição do *Linnaeu's Systema Naturae* em 1758, que representa o início da nomenclatura zoológica aceita pelo Código Internacional de Nomenclatura Zoológica. Embora trabalhos publicados sobre a história natural dos lepidópteros remontam a Aristóteles (384-322 a.C.) e a outros autores pré-Linnaeus, entre eles, Albertus Magnus, Marcel Malpighius, Jan Swammerdam, René Réaumur, e Charles De Geer, que fizeram contribuições significativas, foi Carolus Linnaeus que forneceu uma classificação abrangente para a época (KRISTENSEN et al., 2007).

Linnaeus baseou sua classificação nas características das antenas e asas, em particular a posição das asas em repouso, e sobre se os insetos voavam de dia ou de noite.

Ele também levou em conta as peças bucais e observou que as antenas das borboletas eram associadas, e que algumas espécies, que ele chamou de “Bombyces”, carecia de uma língua (KRISTENSEN et al., 2007).

Linnaeus reconheceu três principais divisões de Lepidoptera - Papilio, Sphinx e Phalaena. Phaleana foi dividida em sete subgrupos, que ainda são usados por nove superfamílias. Seguindo a sequência proposta por Linnaeus estas superfamílias são: Papilionoidea, Sphingoidea, Bombycoidea, Noctuoidea, Geometroidea, Tortricoidea, Pyraloidea, Tineoidea, e Alucitoidea. A contribuição de Linnaeus foi notável e, evidentemente, seu sistema é incorporado em nossa classificação ainda hoje (KRISTENSEN et al., 2007).

Sistematicamente, a ordem atualmente é dividida em quatro subordens: Zeugloptera, Dachnonypha, Monotrysia e Ditrysia. Anualmente, são descritas em média, pouco mais de mil novas espécies, com isso, a classificação dos lepidópteros vem sofrendo constantes modificações. A classificação mais utilizada atualmente é a adotada pela edição do *Handbook of Zoology* (1998) (Tabela 2), onde constam 47 superfamílias, nas quais estão reunidas 124 famílias, sendo 52 delas com classificação de subfamílias (COCK, 2003; KRISTENSEN et al*.,* 2007; KRISTENSEN, 1998).

Tabela 2 - Lista das superfamílias da ordem Lepidoptera

BORBOLETAS	Bombycoidea	Geometroidea	Micropterigoidea	Sesioidea
Hesperioidea	Calliduloidea	Gracillarioidea	Mimallonoidea	Simaethistoidea
Papilionoidea	Choreutoida	Hedyloidea	Mnesarchaeoidea	Thyridoidea
	Copromorphoidea	Hepialoidea	Neopseustoidea	Tineoidea
MARIPOSAS	Cossoidea	Heterobathmioidea	Nepticuloidea	Tischerioidea
Acanthopteroctetoidea	Drepanoidea	Hyblaeoidea	Noctuoidea	Tortricoidea
Agathiphagoidea	Epermenioidea	Immoidea	Palaephatoidea	Urodoidea
Alucitoidea	Eriocranioidea	Incurvarioidea	Pterophoroidea	Whalleyanoidea
Andesianoidea	Galacticoidea	Lasiocampoidea	Pyraloidea	Yponomeutoidea
Axioidea	Gelechioidea	Lophocoronoidea	Schreckensteinioidea	Zygaenoidea

Lista das 47 superfamílias da ordem Lepidoptera, destacando em cinza e sublinhado a superfamília Zygaenoidea.
Fonte: Adaptação de Kristensen, N. P. (1998).

Das 47 superfamílias de lepidópteros, apenas duas, Hesperioidea e Papilionoidea, são consideradas borboletas e, estão distribuídas nas seguintes famílias: Hesperiidae, Papilionidae, Pieridae, Nymphalidae, Lycaenidae e Riodinidae. As superfamílias Hesperioidea

e Papilionoidea juntas correspondem a 12% do total de espécies da ordem Lepidoptera (cerca de 22.000 espécies). As superfamílias e suas respectivas famílias restantes são designadas para mariposas (COCK, 2003).

A superfamília Zygaenoidea é composta por 16 famílias, que são amplamente distribuídas pelo mundo. Os exemplares incluídos nessa superfamília têm por características a presença de uma nervura mediana visível dentro das células discoidais em suas duas asas, além de apresentarem na asa posterior três nervuras livres na região anal. Em geral, os palpos maxilares são rudimentares e os palpos labiais são moderados. Ocelos muitas vezes bem pequenos, porém presentes, e espiritromba bem desenvolvida (EPSTEIN et al., 1999).

Dentro da superfamília Zygaenoidea encontra-se um grupo monofilético denominado "Grupo Limacodidae", que representa o grupo mais rico em espécies e o mais difundido entre os grupos monofiléticos de mariposas. O grupo Limacodidae abrange cinco famílias, Zygaenidae, Megalopygidae, Aididae, Dalceridae e Limacodidae, que se encontram distribuídas em todo o mundo, mas principalmente no Novo Mundo (EPSTEIN, 1996, LEPESQUEUR, 2012).

O cladograma ilustrado na figura 2 apresenta as cinco famílias pertencentes ao grupo Limacodidae, onde se encontra em destaque a família Megalopygidae, a qual pertence à espécie objeto deste trabalho.

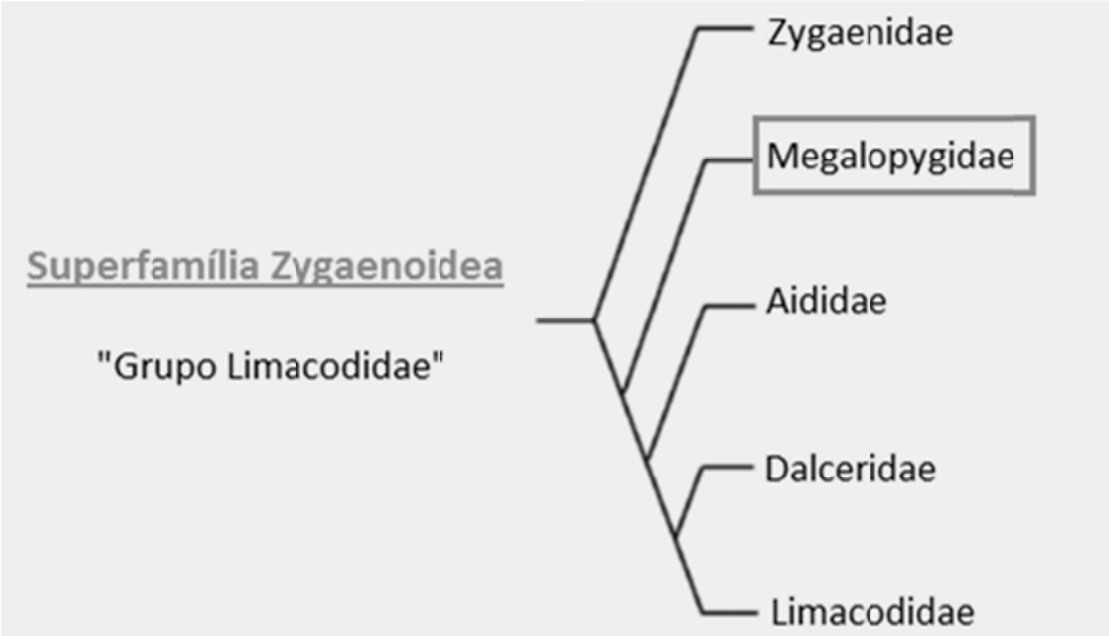


Figura 2 - Cladograma do grupo Limacodidae (Zygaenoidea) baseado na análise de parcimônia de caracteres de adultos e imaturos, evidenciando a família Megalopygidae.
Fonte: Adaptação de Epstein, M. E. (1996).

O grupo Limacodidae é caracterizado pela presença de cabeça retrátil, presença de falsas pernas nos segmentos abdominais A2-A7 e por apresentar superfície ventral flexível durante o estágio larval. Existe uma grande diversidade de formas larvais no grupo Limacodidae, embora as larvas pertencentes ao grupo sejam conhecidas por lagartas espinhosas que podem causar dermatite. As lagartas das famílias Limacodidae, Megalopygidae e Dalceridae são pragas econômicas de palmas e outras culturas de plantação nos trópicos (EPSTEIN, 1996).

O relacionamento entre as famílias do grupo Limacodidae ainda é mal compreendido, devido, em grande parte, a um conhecimento incompleto dos estágios imaturos e a escassez de estudos morfológicos detalhados tanto dos estágios imaturos quanto dos adultos (EPSTEIN, 1996; LEPESQUEUR, 2012).

2.2 A família Megalopygidae

Megalopygidae (do grego *Mega* = grande e *pyge* = nádega, parte posterior do corpo) é o nome dado a uma família representada por mariposas que são, atualmente, consideradas limitadas ao Novo Mundo, e encontradas primariamente na região neotropical, com poucos representantes na América do Norte e África (EPSTEIN et al*.,* 1999; HEPPNER, 1991). Apenas duas subfamílias (Megalopyginae e Trosiinae) são reconhecidas atualmente (EPSTEIN, 1996; LEPESQUEUR, 2012).

A família Megalopygidae possui 232 espécies descritas (BECKER, 1995; EPSTEIN, 1999, LEPESQUEUR, 2012). Provavelmente, acima de 50% do total das espécies descritas são encontradas no Brasil, porém, essas informações ainda são incipientes e bastante fragmentadas, não existindo dados completos sobre a diversidade do grupo em qualquer região ou bioma brasileiro (LEPESQUEUR, 2012). Apenas na cidade de São Paulo, estima-se a ocorrência de 57 espécies, em Itatiaia são encontradas 54 espécies e no estado do Rio Grande do Sul, 23 espécies (BROWN JR; FREITAS, 1999; SPECHT et al., 2004).

As primeiras espécies de Megalopygidae foram descritas por Cramer, em 1775, como *Phalaena dimas* e *Phalaena orsilochus.* Como a família Megalopygidae foi estabelecida somente em 1855, por Herrich-Schäffer, estas espécies pertenceram a diferentes famílias e somente cerca de 150 anos depois, essas espécies foram revistas por Dyar (1910) e Hopp

(1934), respectivamente, e incluídas em Megalopygidae como *Trosia dimas* e *Podalia orsilochus* (BECKER, 1995; LEPESQUEUR, 2012).

Os séculos XIX e XX foram marcados por grandes lepidopterologistas e taxônomos, e mais de 94% das espécies de Megalopygidae atuais foram descritas entre 1850 e 1950 (BECKER, 1995; LEPESQUEUR, 2012).

2.2.1 Ecologia e características morfológicas dos Megalopygideos

O ciclo biológico e desenvolvimento da família Megalopygidae, assim como outros lepidópteros, apresentam evolução completa ou holometabólica, variando apenas os períodos conforme a espécie (KAWAMOTO; KUMADA, 1984).

Os adultos de Megalopygidae compreendem mariposas de corpo robusto, densamente piloso, com envergadura alar variando de dois a 11 cm, sendo a maioria de tamanho mediano (COSTA LIMA, 1945). Possuem asas anteriores e posteriores apicalmente arredondadas, acetinadas ou aveludadas com áreas revestidas por cerdas crespas ou onduladas. As asas podem apresentar variadas colorações como: totalmente branca, róseas, vermelhas e, ainda, podem apresentar tons de laranja e amarelo, marrom ou cinza com manchas pretas ou pardas (EPSTEIN et al*.,* 1999; LEPESQUEUR, 2012). Exemplos dessa variação cromática podem ser observados na Figura 3.

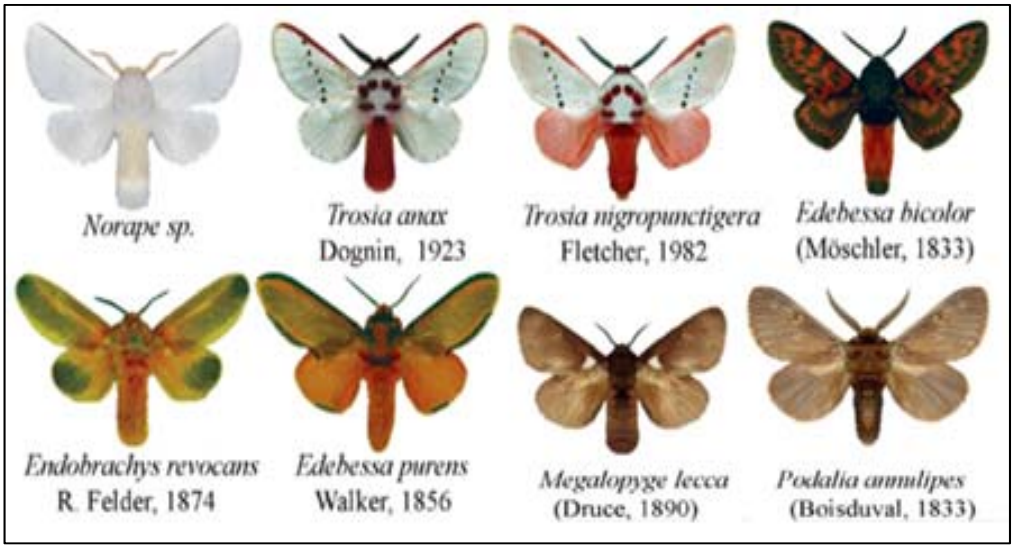


Figura 3 - Exemplos da variação cromática das asas de algumas espécies membros da família Megalopygidae.
Fonte: Adaptação da figura de Lepesqueur, C. (2012).

Pode ser identificado o dimorfismo sexual do grupo através de diferenças no tamanho corporal, onde os machos são menores que as fêmeas, assim como através das antenas, onde a dos machos são bipectinadas e a das fêmeas são serreadas ou pectinadas com dentes curtos (EPSTEIN et al*.,* 1999; LEPESQUEUR, 2012).

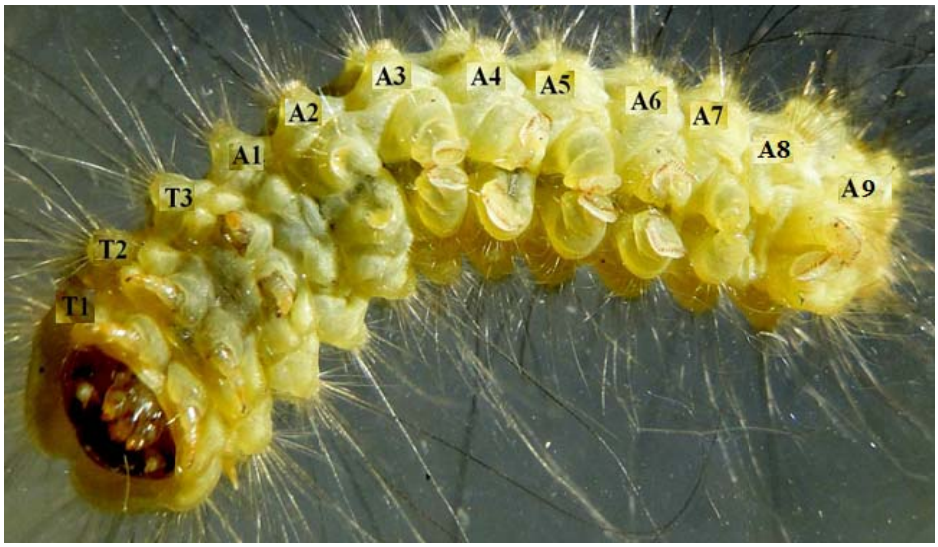
Em geral, as mariposas acasalam durante a noite ou ao anoitecer (NATION, 2002). Machos e fêmeas podem acasalar mais de uma vez resultando em um número maior de fecundações (WARD; LANDOLF, 1995). A idade dos adultos pode influenciar a fertilidade de ambos os sexos, assim como a temperatura, que pode afetar indiretamente a reprodução por alterar a frequência de cópulas, maturação ovariana, metabolismo e nutrição (CREMA; CASTELO BRANCO, 2004). Os adultos vivem em média 15 dias, a maioria das espécies apresenta aparelho bucal rudimentar ou ausente e, por isso, nesta fase do ciclo, não se alimentam (SCOBLE, 1995).

Os ovos provenientes desse grupo medem 1,2 x 0,6 x 0,6 mm, são cilíndricos ou elipsoides, amarelados e cobertos com cerdas provenientes do final do abdome da fêmea. A oviposição apresenta entre 30 a 50 ovos, dependendo da espécie e ocorre geralmente, nos meses de janeiro a março, onde os ovos são dispostos em fileiras únicas ou duplas diretamente nas plantas que servirão de alimento para as lagartas, eclodindo em média, uma semana após a postura (COSTA LIMA, 1945; PEREIRA; HALFELD-VIEIRA; NECHET*,* 2003; SCOBLE, 1995).

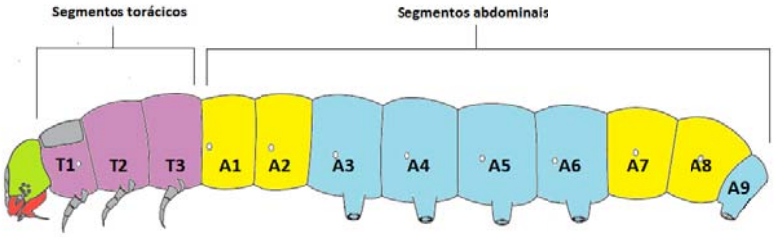
Em uma análise sistemática, a morfologia das lagartas da família Megalopygidae difere das demais famílias da ordem Lepidoptera, por apresentarem nos segmentos abdominais, um número maior de falsas pernas (n=7), enquanto lagartas de outras famílias apresentam menos pares de pernas nesses segmentos (n=5). Essas falsas pernas, denominadas larvópodes, possuem ganchos e estão situadas nos segmentos abdominais A3 a A6 e há mais dois pares localizados nos segmentos A2 e A7, podendo, porém, serem desprovidos de ganchos (Figura 4) (BARTH, 1954; DUARTE et al*.,* 2012). O segmento anal (A9) também apresenta larvópodes, com ganchos uniordinais (COSTA LIMA, 1945; EPSTEIN et al*.,* 1999; SCOBLE, 1995).

A presença de grandes almofadas membranosas nos larvópodes e de sensilas digitiformes na região proximal dos espiráculos foi considerada como característica peculiar (apomórfica) dos Megalopygidae (EPSTEIN et al*.,* 1999; SPECHT; CORSEUIL, 2008).

A)

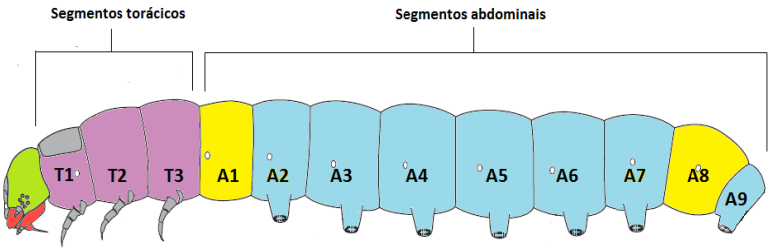


B)



Lagartas de outras famílias de Lepidoptera

C)



Lagartas da família Megalopygidae

Figura 4 - Morfologia externa das lagartas.
Legenda: (**A**) Lagarta de *Norape sp*. (família Megalopygidae) (visão ventral) evidenciando os setes pares de larvópodes, sendo A2 e A7 desprovidos de pernas. T1-3 = segmentos torácicos; A1-9 = segmentos abdominais. (**B**) Esquema representando a morfologia externa de lagartas de outras famílias. (**C**) Esquema representando a morfologia externa das lagartas da família Megalopygidae.
Fonte: (**A**) Lepesqueur, C. (2012); (**B** e **C**) Adaptação de Barth, R. (1954).

As lagartas dessa família são polífagas, ou seja, alimentam-se de folhas de plantas de famílias taxonomicamente diversas (LEPESQUEUR, 2012), podendo trocar de alimento durante o desenvolvimento da lagarta (COSTA LIMA, 1945). Algumas espécies têm sido

referidas como pragas em eucaliptoculturas de Belo Oriente, MG (ZANUNCIO et al., 1993), de *Acacia mangium*. (Fabaceae), em Roraima (NECHET, 2003; PEREIRA; HALFELD-VIEIRA) e de várias palmáceas, cacaueiro, goiabeira e de samambaias (EPSTEIN et al*.,* 1999).

As lagartas da família Megalopygidae são geralmente solitárias, não agressivas, com 1 a 8 cm de comprimento. Como podemos observar na Figura 5, as lagartas desta família são providas de longas e sedosas cerdas não urticantes, mais ou menos densas, com colorido que varia do cinza ao vermelho, róseos, amarelos, marrons, negros, daí os nomes populares de "lagartas cabeludas", "lagartas de fogo" ou "taturanas" (de origem tupi-guarani que significa *"aquilo que arde como fogo"*). Alguns nomes vulgares como taturana-cachorrinho, taturana-de-flanela, taturana-gatinho, chapéu-armado e bicho-cabeludo são dados para estas lagartas devido à presença dos pelos sedosos (DUARTE et al*.,* 2012; LEPESQUEUR, 2012).

Figura 5 - Lagartas da família Megalopygidae.
Legenda: (**A**) *Podalia annulipes;* (**B**) *Podalia orsilochus*; (**C**) *Megalopyge lanata* e (**D**) *Edebessa purens*.
Fonte: Lepesqueur, C. (2012).

Essa pilosidade densa é formada por fileiras de verrugas dorsolateral, lateral e subespiracular (SCOBLE, 1995). Conforme a disposição das cerdas, estas lagartas podem ser divididas em dois grupos (SPECHT; CORSEUIL, 2008): as arctiidiformes, com cerdas curtas, robustas e eretas, como a *Megalopyge urens* (BERG, 1882) e as "lagartas gatinho" ou *puss*

caterpillars com cerdas longas, macias, sedosas e coloridas, como a *M. albicollis* (WALKER, 1855).

A maioria das cerdas longas e sedosas que se assemelha à pelos, é inofensiva, porém, camuflam cerdas pontiagudas, denominadas espículas, que são estruturas produtoras e inoculadoras de compostos químicos com ação urticante (MORAES, 2003).

Nos estudos sobre a biologia das lagartas, toda a inserção na epiderme é considerada cerda, entretanto é comumente adotada a nomenclatura "espícula", apenas para diferenciar as cerdas peçonhentas, que contém veneno, das não peçonhentas (BARTH, 1954; CARDOSO; HADDAD JR., 2005; MORAES, 2003). As espículas das lagartas peçonhentas são cerdas modificadas que apresentam seu interior oco até a extremidade e são preenchidas por um líquido com compostos químicos de ação urticante secretado por células situadas na base do mesmo, denominadas células tricógenas (MORAES, 2003).

As espículas das lagartas de Megalopygidae apresentam um número bastante reduzido em relação às cerdas (aproximadamente 1:8-10), e localizam-se imperceptíveis de forma regular entre elas. Essas espículas inserem-se diferentemente das cerdas, possuindo largos canais que atravessam a cutícula, penetram a hipoderme e entram nas células glandulares em cavidades denominadas bombas de peçonha. Possuindo articulação completamente diferente das cerdas, as espículas apresentam dois cilindros, um interno e outro externo e forma um fino canal que entra no interior da bomba de peçonha. A secreção da célula glandular passa pelo canal e preenche a espícula até a sua extremidade (BARTH, 1954).

Acidentes envolvendo os seres humanos e lagartas urticantes ocorrem quando as espículas do animal são pressionadas e, devido à pressão do contato no momento do acidente, a espícula é rompida em qualquer ponto, o que é suficiente para que ocorra a injeção do veneno, entrando em contato com a pele humana e provocando ação irritante (CARDOSO; HADDAD JR., 2005). Após o contato, a bomba de peçonha é capaz de retornar à sua posição inicial, porém, não se sabe se a espícula se reconstitui totalmente. (BARTH, 1954). A gravidade corresponderá ao número de espículas envolvidas no acidente. Algumas espécies de taturanas são especialmente bonitas e atrativas, o que faz com que as pessoas, principalmente as crianças, passem a mão sobre a lagarta (CARDOSO; HADDAD JR., 2005; LEPESQUEUR, 2012).

A fase larval dos megalopygideos é longa e, durante este período, as lagartas sofrem ecdise e passam por seis instares. Nos primeiros instares, as lagartas apresentam hábitos gregários, se alimentam constantemente e crescem consideravelmente sem mudar sua morfologia durante toda a fase que pode variar de dois (COSTA LIMA, 1945) a doze meses. Após a última ecdise, elas se transformam em pupas (MORAES, 2003).

O casulo é construído por pelos provenientes da própria lagarta formando um tecido fibroso e resistente para a proteção da pupa durante a metamorfose. Internamente, o casulo apresenta um espaço com tecidos menos espessos para facilitar a imersão do Lepidóptero na forma adulta. O casulo é produzido e fixado em uma superfície escolhida pela lagarta, de tal forma que se torna quase impossível removê-lo sem danificá-lo. Do casulo emergem as mariposas adultas e recomeçam o ciclo (LORDELLO, 1951).

Assim como para a maioria das outras famílias de mariposas, as informações sobre a biologia, ecologia, comportamento e distribuição das espécies da família Megalopygidae são escassas. Até o momento, trabalhos publicados se referem à importância médica das lagartas (BARTH, 1954; CARDOSO; HADDAD JR, 2005; SPECHT et al*.,* 2004; SPECHT; CORSEUIL, 2008), ou apenas incluem uma ou outra espécie de Megalopygidae em listagens de coletas de adultos (HAMBLETON, FORBES, 1935) ou de lagartas associadas a determinadas plantas hospedeiras, sem focar exclusivamente sobre a família (DINIZ; MORAIS*,* 2010; PEREIRA; HALFELD-VIEIRA; NECHET, 2003; SCHERRER; SILVA et al., 2011; ZANUNCIO et al*.,* 1993).

2.3 A espécie *Podalia orsilochus* (Cramer, 1775)

A espécie *Podalia orsilochus* foi originalmente descrita por Cramer em 1775, pertencendo à família Megalopygidae e subfamília Megalopyginae. Os adultos da espécie *P. orsilochus* possuem tórax e abdome marrom, com os anéis dos segmentos intercalados com branco. A borda externa das asas anteriores apresenta coloração marrom clara, com tonalidade mais escura na porção basal até a metade e com uma mancha circular marrom na porção distal da célula discal. Os machos são menores e menos robustos que as fêmeas e possuem asas posteriores mais esbranquiçadas (LEPESQUEUR, 2012; ZORZENON, 2010).

A lagarta possui o tegumento densamente coberto por cerdas de coloração marrom, porém bastante variável entre indivíduos e ao longo do desenvolvimento (escurecendo

durante seu desenvolvimento). Reage ao contato físico mantendo o corpo completamente enrolado e é caracterizada pelo “topete” sobre o tórax e no final do abdome. Suas características peculiares a torna popularmente conhecida como “lagarta cachorrinho”. Assim como as demais lagartas da família Megalopygidade, a *P. orsilochus* possui uma dieta polífaga. O casulo pupal é construído por seda e cerdas, podendo apresentar-se em diferentes tons de marrom (LEPESQUEUR, 2012; ZORZENON, 2010).

Os representantes de cada fase do ciclo da *P. orsilochus* podem ser observados na figura 6 logo abaixo:

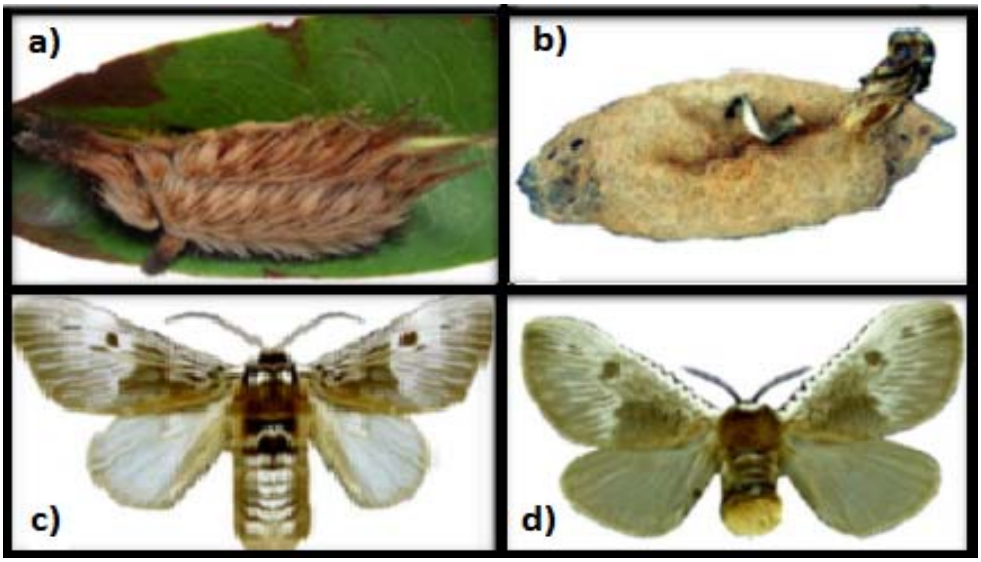


Figura 6 - Representante de cada fase do ciclo da espécie Podalia orsilochus (Cramer, 1775).
Legenda: (**A**) Lagarta solitária, exposta no limbo foliar caracterizada pelo “topete” sobre o tórax e no final do abdome; (**B**) casulo pupal construído com seda e cerdas; (**C**) macho adulto; (**D**) fêmea adulta.
Fonte: Adaptação de Zorzenon, F. J. (2010); Lepesqueur, C. (2012).

2.5 A imunidade nos insetos

A Classe Insecta, pertencente ao filo Arthropoda (do grego *arthros*: articulado e *podos*: pés, patas, apêndices), representa a classe com maior abundância em espécies do grupo animal, representando cerca de 60% de toda a biodiversidade existente na terra e aproximadamente 85% de todos os animais (CHERNYSH et al., 2002). A classe dos insetos possui mais de um milhão de espécies, com um número equivalente a 30 milhões de espécies estimadas (RAVI; JEYASHREE; RENUKA DEVI, 2011), que se encontram organizadas em 32 ordens, divididas em quatro subordens (Zeugloptera, Aglossata, Heterobathmiina e

Glossata), nas quais estão distribuídas aproximadamente 800 famílias distribuídas em todo o mundo (BULET; STOCKLIN, 2005; GRIMALDI; ENGEL, 2005; VAN NIEUKERKEN et al., 2011).

Acredita-se que os insetos tenham surgido no período Devoniano, há mais de 350 milhões de anos. Durante este período, os insetos tiveram um grande sucesso evolutivo devido a características importantes, como a vida relativamente curta em comparação com a maioria dos vertebrados, a sua capacidade de colonizar a maioria dos nichos biológicos, com exceção ao ambiente marinho profundo e as regiões polares e, a capacidade de se alimentar de quase todas as espécies de plantas e animais (BULET et al., 1999; BULET; STOCKLIN, 2005).

Além de sua larga distribuição nos diferentes ecossistemas do planeta, os insetos conseguiram sua sobrevivência em ambientes contaminados ou extremamente prejudiciais ao desenvolvimento de vida (CHERNYSH et al., 2002), sendo necessário desenvolver mecanismos de proteção e defesa para cada situação adversa, com a finalidade de superar de forma eficaz os impactos causados pelos diversos patógenos que possam infectá-los (LOWENBERGER, 2001). O desenvolvimento da capacidade dos insetos em montar uma resposta imune resitente aos patógenos certamente contribuiu para a sua proliferação e diversidade (BULET et al., 1999; BULET; STOCKLIN, 2005).

Nos insetos, a primeira barreira contra a invasão de patógenos e parasitas é física, ou seja, a cutícula. No entanto, uma vez que esta barreira é rompida, ocorre uma interação complexa de reações imunes. A imunidade inata é um mecanismo evolutivamente conservado em eucariotos e é a primeira linha de defesa, que permite ao organismo sobreviver a diferentes agentes infecciosos, tais como bactérias, fungos, protozoários e vírus, de uma forma não específica (MONTAÑO et al., 2011).

Em organismos superiores, a imunidade inata coexiste com a imunidade adaptativa, que em oposição à imunidade inata possui memória e tem especificidade. Embora evidências sugiram a presença de respostas adaptativas (DONG; TAYLOR; DIMOPOULOS, 2006; SADD; SCHMID-HEMPEL, 2006), os insetos dependem exclusivamente da imunidade inata (sem memória), que é prontamente induzida nos tecidos e hemocele através de mecanismos de defesa inata e humoral, que se manifestam de três maneiras: a primeira, uma resposta humoral que gera rápida síntese e circulação de diferentes tipos de peptídeos antimicrobianos (AMPs) na hemolinfa, em segundo, uma resposta celular, resultando na fagocitose ou encapsulamento de patógenos por células do sangue, plasmatócitos e,

terceiro, a ativação da via fenoloxidase que cria depósitos de melanina em torno de feridas e partículas estranhas (RAVI; JEYASHREE; RENUKA DEVI, 2011; TAPADIA; VERMA, 2012). Um esquema resumido das respostas imunológicas que ocorrem nos insetos encontra-se na figura 7.

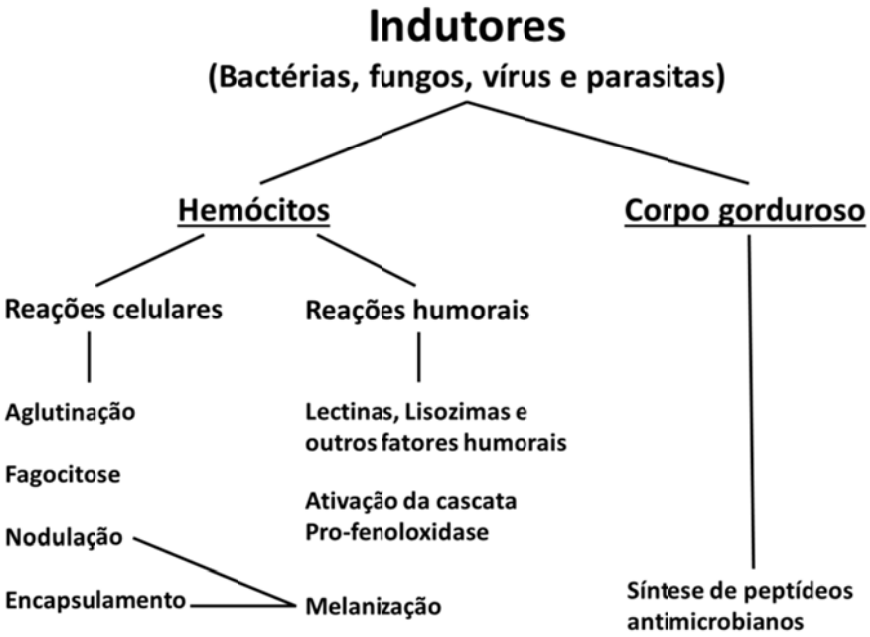


Figura 7 - Representação esquemática das respostas imunológicas dos insetos.
Legenda: A ligação de moléculas receptoras na superfície de um microrganismo pode induzir diferentes tipos de reações tais como aglutinação, fagocitose, formação de nódulos e encapsulamento pelos hemócitos. A ativação da cascata de pro-fenoloxidase é necessária para o processo de melanização e morte dos microrganismos isolados nos nódulos ou cápsulas. A síntese de peptídios antimicrobianos constitui a última fase no combate a um microrganismo invasor.
Fonte: Silva, 2012.

Coletivamente, os AMPs são conhecidos por montar uma defesa imunológica eficaz em resposta à invasão sistemica. Em insetos com metamorfose incompleta (heterometabólica), os AMPs são produzidos pelos hemócitos (células do sangue) no animal saudável e secretados na hemolinfa sobre infecção. Em contraste, em insetos com metamorfose completa (holometabólica), que é o caso dos Lepidópteros, os AMPs são rapidamente e transitoriamente sintetizados pelo corpo gorduroso (tecido equivalente funtional ao fígado de mamíferos), e por vários epitélios (Figura 8). Quando produzido pelo corpo gorduroso, os AMPs são rapidamente secretados para a hemolinfa (sangue), de onde podem facilmente difundir e agir em todo o animal (BULET; STOCKLIN, 2005; SADD; SCHMID-HEMPEL, 2006; TAPADIA; VERMA, 2012).

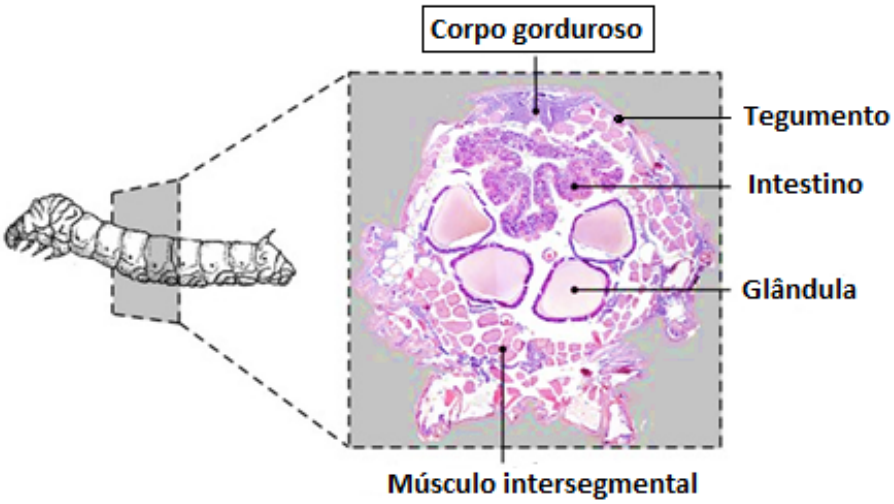


Figura 8 - Intersecção de uma lagarta, evidenciando suas estruturas internas. Em destaque o corpo gorduroso.
Fonte: Adaptação de Romanelli, D. et al. (2014).

A maioria dos AMPs é produzida maciçamente, muitos deles atingindo altas concentrações micromolares (BULET; STOCKLIN, 2005). Como regra geral, um dado inseto produz um repertório de AMPs que têm características estruturais que se sobrepõem, mas frequentemente são alvos de microrganismos específicos. Além disso, a presença simultânea de vários AMPs que atuam em sinergia pode proporcionar aos insetos uma defesa mais eficaz contra invasores nocivos (RAVI; JEYASHREE; RENUKA DEVI, 2011). Os AMPs também ajudam na manutenção de um estado de equilíbrio da microflora natural para o bom funcionamento do sistema (TAPADIA; VERMA, 2012).

Insetos, mamíferos, répteis, plantas e seres vivos de todos os reinos produzem peptídeos antimicrobianos (AMPs) para se proteger contra infecção microbiana e garantir a sobrevivência em ambientes em constante mudança. Centenas de peptídeos antimicrobianos foram descobertos e classificados com base em sua relação estrutura/atividade (SOARES; MELLO, 2004).

Na base de dados *APD- The Antimicrobial Peptide Database* (ps.unmc.edu/AP/main.php) (WANG; WANG, 2004) é possível encontrar os diversos AMPs classificados em quatro famílias: alfa (α), beta (β), alphabeta (αβ), e não-alphabeta (não-αβ) com base nos tipos de estruturas secundárias (Figura 9). A família α consiste de AMPs com estruturas helicoidais. A família β é composta de AMPs com cadeias do tipo beta. A família αβ apresenta tanto domínios estruturais helicoidais do tipo alfa quanto do tipo beta e, a família não-αβ, que não apresenta domínios do tipo alfa nem beta (WANG; LI; WANG, 2016).

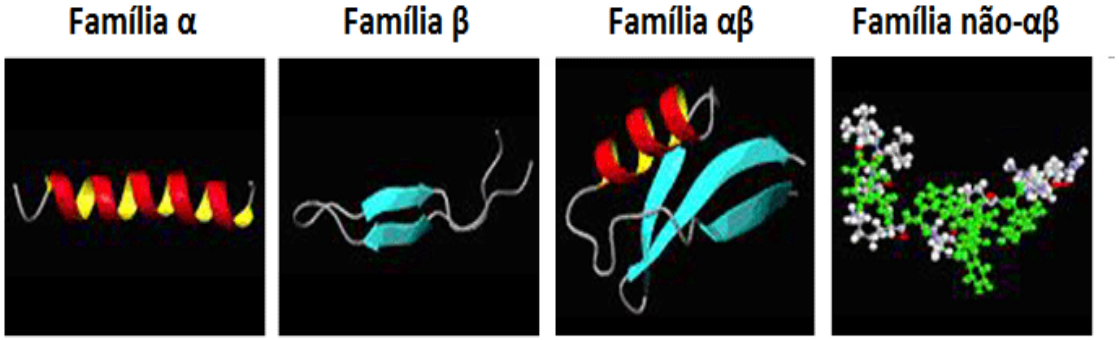


Figura 9 – Representação das quatro famílias estruturais de peptídeos antimicrobianos (AMPs).
Fonte: http://aps.unmc.edu/AP/main.php
Acesso em: 14/01/2016.

2.5.1 Peptídeos antimicrobianos de insetos

Do número de estruturas antimicrobianas relatadas até 2012, 75.73 % foram identificadas em animais (Figura 10), e destes, pelo menos 50 % foram identificados em invertebrados, com grande predominância em insetos (RAVI; JEYASHREE; RENUKA DEVI, 2011). Os insetos produzem uma grande variedade de proteínas/peptídeos antimicrobianos (AMPs) que partilham características comuns, tais como, baixo peso molecular (inferior a 5 kDa), uma carga líquida positiva em pH fisiológico, são estáveis ao calor, tipicamente catiônicos e geralmente são formados por sequências de aminoácidos menores do que 100 resíduos (WANG; LAI, 2010).

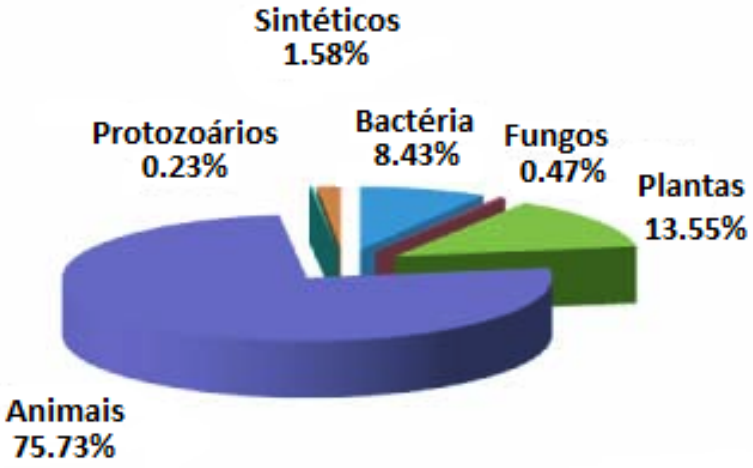


Figura 10 - Fontes de peptídeos antimicrobianos (AMPs).
Fonte: http://aps.unmc.edu/AP/facts.php
Acesso em: 14/01/2016.

A maioria dos AMPs é pequena, apresentando entre 20 a 50 resíduos de aminoácidos, e são sintetizados como precursores inativos de proteínas ou pró-proteínas, tornando-se peptídeos ativos somente após proteólise. Porém, existem algumas grandes proteínas antimicrobianas, como as gloverinas (~ 14 kDa) e atacinas (~ 20 kDa) (YI et al., 2014).

Os AMPs produzidos por insetos podem apresentar atividades contra bactérias e/ou fungos e, alguns apresentam atividade contra parasitas e vírus (YI et al., 2014). A atividade antibacteriana em insetos foi observada pela primeira vez em pupas dos lepidópteros *Samia cynthia* e *Hyalophora cecropia* (família Saturniidae), após indução com bactérias (FAYE et al., 1975) e, mais tarde, em adultos do díptero *Drosophila melanogaster* também induzidos por bactérias (ROBERTSON; POSTLETHWAIT, 1986). O primeiro peptídeo antibacteriano isolado e caracterizado em insetos foi cecropina, a partir da hemolinfa da pupa de *H. cecropia* em 1981 (HULTMARK et al., 1980; STEINER et al., 1981), e desde então têm sido identificados mais de 170 AMPs de insetos (BULET, et al., 1999; WANG; LI; WANG, 2016; YI et al., 2014).

A maioria dos AMPs de insetos, tais como as defensinas de insetos, cecropinas, peptídeos ricos em prolina e atacinas, foram isolados de diferentes espécies, sendo encontrados em mais de duas ordens, mas moricina e gloverina até o momento foram identificadas apenas na ordem Lepidoptera (COCIANCICH et al., 1994).

Em *D. melanogaster*, sete grupos distintos de AMPs (cecropinas, drosocinas, atacinas, diptericinas, defensinas, drosomicinas e metchnikowinas) têm sido identificados e caracterizados através da análise do genoma total, além da regulação desses genes e suas vias de sinalização que também têm sido bem estudadas (BULET et al., 1999; YI et al., 2014).

Os AMPs de insetos podem adotar certas estruturas ou conter sequências únicas e, assim, podem ser classificados em quatro grandes classes: (I) peptídeos α-hélices e privados de resíduos de cisteína, como cecropina e moricina; (II) peptídeos cíclicos contendo resíduos de cisteína, como defensina e drosomicina; (III) peptídeos com rica representação de prolina, como apidaecina, drosocina e lebocina; e (IV) peptídeos com rica representação de glicina, como atacina e gloverina (BULET et al., 1999; BULET; STOCKLIN, 2005; OTVOS, 2000; YI et al., 2014).

2.5.1.1 Classe dos AMPs α-hélices e privados de resíduos de cisteína

Os principais representantes da classe de peptídeos antimicrobianos isolados de insetos que apresentam em sua estrutura uma α-hélice e são privados de resíduos do aminoácido cisteína são as cecropinas, as moricinas e seus variantes (BULET et al., 1999; BULET; STOCKLIN, 2005).

Cecropinas são representantes de uma família de peptídeos antimicrobianos catiônicos que possuem de 31 a 39 resíduos e, desde seu descobrimento em 1981, têm sido identificadas em insetos lepidópteros, dípteros e coleópteros (STEINER et al., 1981; YI et al., 2014). Os membros da família cecropina também incluem aqueles com diferentes nomes, tais como sarcotoxina-I (OKADA; NATORI, 1985), papillona (KIM et al., 2010), stomoxina (LANDON et al., 2006), hinnavina (YOE et al., 2006), o SB-37 e Shiva (derivados sintéticos de cecropinas) (JAYNES et al., 1988).

Cecropinas são sintetizadas como proteínas secretadas e se tornam ativas somente após a remoção do peptídeo sinal. Possuem um amplo espectro de atividade contra bactérias Gram-negativas e Gram-positivas, bem como contra fungos (HULTMARK et al., 1980; LEMAITRE; HOFFMANN, 2007; MOORE et al., 1996; VIZIOLI et al., 2000).

Cecropinas adotam uma estrutura em espiral aleatória em solução aquosa, mas convertem para uma estrutura α-helicoidal em ambientes hidrofóbicos. Possuem duas α-hélices que proporcionam maior eficiência na atuação sobre bactérias Gram-negativas (TANIAI et al., 1996; YI et al., 2014).

Além de sua atividade antimicrobiana, cecropinas e seus derivados, SB-37 e Shiva, são também ativos contra parasitas, incluindo *Plasmodium* e *Trypanossoma* (JAYNES et al., 1988), e podem inibir a replicação do vírus HIV-1 (WACHINGER et al., 1998). Papiliocina também apresenta atividade antinflamatória (KIM et al., 2011) e pode induzir apoptose em *C. albicans* (HWANG et al., 2011). Sarcotoxin-AI pode desempenhar um papel no desenvolvimento *S. peregrina* (NANBU et al., 1988; YI et al., 2014).

Cecropinas não são letais para a linhagem celular de fígado humano (*Chang liver cells*), células vermelhas do sangue de ovelhas, hemácias humanas, ou linfócitos normais, mas são capazes de matar seletivamente as células bacterianas. Além disso, cecropinas podem inibir a proliferação de células cancerosas, exibem a capacidade de lisar células cancerosas de leucemia e demonstrou potencial como um sistema de entrega de drogas

anticâncer, mascarando os efeitos tóxicos de drogas atuais, enquanto ainda fornece as drogas para as células cancerosas (SOARES; MELLO, 2004).

Um grande desafio na análise de AMPs desta classe é a diversidade de mecanismos de interação entre os peptídeos e as membranas celulares dos microrganismos e, a maioria das análises tem sido focada apenas em cecropinas de inseto isoladas das pupas da mariposa *Hyalophora cecropia* (SOARES; MELLO, 2004).

Moricina foi primeiramente isolada a partir da hemolinfa de pupas de *Bombyx mori,* induzidas por *E. coli*, como um peptídeo altamente básico de 42 resíduos e são sintetizadas como proteínas secretadas após a clivagem do peptídeo sinal (YI et al., 2014). Até o momento moricina foi encontrada apenas em Lepidópteros, e os cDNAs que as codificam foram identificados em *M. sexta* (ZHU et al., 2003), *S. litura* (OIZUMI et al., 2005), *G. mellonella* (BROWN et al., 2008), *H. armigera* (WANG et al., 2010), *S. exigua, H. virescens, e H. puera* (YI et al., 2014).

Moricinas apresentam atividade contra bactérias Gram-negativas e Gram-positivas, e as moricinas de *G. mellonella* também mostram alta atividade contra fungos filamentosos e leveduras (BROWN et al., 2008). As estruturas terciárias de moricinas mostram uma longa α-hélice com 8 voltas ao longo de praticamente o peptídeo inteiro, com exceção de alguns resíduos nas regiões N- e C-terminal (OIZUMI et al., 2005). O segmento N-terminal da α-hélice (resíduos 5-22) é anfipático e é o responsável pelo aumento da permeabilidade da membrana das bactérias, e o segmento C- terminal da α-hélice (resíduos 23-36) é hidrofóbico e é crítico para a atividade antimicrobiana de moricina (HEMMI et al., 2002).

2.5.1.2 Classe dos AMPs cíclicos contendo resíduos de cisteína

Classe composta por defensinas e seus derivados. As defensinas são pequenos AMPs (~ 4 kDa), catiônicos/básicos, que apresentam seis resíduos conservados de cisteína que formam três pontes dissulfeto intramoleculares, e que foram identificados em quase todos os organismos vivos (YI et al., 2014). Com base nas características estruturais, as defensinas podem ser classificadas em três grupos: defensinas "clássicas", β-defensinas e defensinas de insetos (GANZ; LEHRER, 1994).

Há diversos estudos sobre defensinas, incluindo defensinas de vertebrados que atuam na imunidade inata (ZHAO; LU, 2014), defensinas vegetais (CARVALHO ADE; GOMES,

2011), o modo de ação e estrutura de defensinas de diferentes reinos (WILMES et al., 2011), assim como o modo de ação de defensinas de plantas, insetos e humanas em resposta antifúngica (AERTS et al., 2008).

Assim como as defensinas, muitos peptídeos ricos em cisteína com nomes diferentes, tais como sapecinas, roialisina, tenecina-1, holotricina-1, heliomicina, spodoptericina, gallerimicina, coprisina e lucifensina, podem todos pertencer à família de defensina de insetos (YI et al., 2014).

As defensinas de insetos, ao contrário de outros AMPs, foram isoladas a partir de várias ordens de insetos, tais como Lepidoptera, Diptera, Himenoptera, Coleoptera, Trichoptera, Hemiptera e também está presente na antiga ordem de insetos Odonata (BULET et al., 1999; COCIANCICH et al, 1994; RAVI; JEYASHREE; RENUKA DEVI, 2011), sugerindo que as defensinas de insetos podem derivar de um gene ancestral comum (YI et al., 2014).

As defensinas de insetos foram relatadas pela primeira vez como sapecinas nos dípteros *Sarcophaga peregrina* (MATSUYAMA; NATORI, 1988), e *Phormia terranovae*, contendo 40 resíduos e 6 cisteínas (LAMBERT et al., 1989). Como esses peptídeos apresentavam significativa homologia com defensinas de mamíferos, foram nomeados como defensinas de insetos (YI et al., 2014).

Sapecina de *S. peregrina* e Defensina-A de *P. terranovae* são sintetizadas como pré-pró-proteínas compostas por um peptídeo sinal, um pré-peptídeo e um peptídeo de defensina maduro de 40 resíduos apresentando três ligações do tipo ponte dissulfeto (MATSUYAMA E NATORI, 1988). Pré-pró-defensinas foram confirmadas em *Aedes aegypti, Stomoxys calcitrans, Spodoptera frugiperda* (YI et al., 2014), *Bombyx mori* (KANEKO et al., 2008), e em *Spodoptera littoralis* (SEUFI; HAFEZ; GALAL, 2011).

A estrutura de defensinas de insetos é composta de um anel na porção N-terminal, uma α-hélice, seguido por uma β-folha antiparalela. A α-hélice e a β-folha são ligadas por duas pontes dissulfeto intramoleculares, formando uma estrutura estabilizada denominada “cisteína alfa-beta” (CSαβ) (CORNET et al., 1995). O motivo CSαβ foi confirmado nos outros membros de defensinas de inseto, tenecina-1 e heliomicina (LAMBERTY et al., 1989), e está também presente no peptídeo antimicrobiano termicina de *Pseudacanthotermes spiniger,* e no peptídeo antifûngico drosomicina de *D. melanogaster* (YI et al., 2014).

Entretanto, o peptídeo drosomicina contém oito cisteínas e uma β-folha extra no resíduo N-terminal, o qual se encontra ligado ao resíduo C-terminal pela ligação dissulfeto adicional. Assim, drosomicina e termicina também podem pertencer à família de defensinas de insetos (YI et al., 2014). Uma nova família de proteínas chamada X-Tox, que foi identificada apenas na ordem Lepidoptera, contém 11 motivos CSαβ imperfeitamente conservados, mas perde atividade antimicrobiana ao comparar com demais proteínas (D'ALENÇON et al, 2013).

A maioria das defensinas de insetos são catiônicas e contém cerca de 40 resíduos, mas duas defensinas de abelhas, defensinas de *Bombus pascuorum* e roialisina de *Apis melífera,* apresentam 51 resíduos totais (YI et al., 2014). Defensina de *B. mori* (36 resíduos) e spodoptericina de *S. frugiperda* (36 resíduos) são aniônicos (pI 4,12 e 4,35, respectivamente) (KANEKO et al., 2008), enquanto que a defensina de *S. littoralis* também é aniônico, mas apresenta 50 resíduos de comprimento (SEUFI; HAFEZ; GALAL, 2011).

Defensinas de insetos são ativas principalmente contra bactérias Gram-positivas, incluindo *Micrococcus luteus*, *Aerococcus viridian*, *Bacillus megaterium*, *B. subtilis*, *B. thuringiensis*, e incluindo bactérias patogénicas humanas, tais como *Staphylococcus aureus.* Algumas defensinas de inseto são também ativas contra bactérias Gram-negativas *E. coli* e alguns fungos (SEUFI; HAFEZ; GALAL, 2011; VIZIOLI et al., 2001).

As defensinas de inseto podem matar as bactérias através da formação de canais na membrana citoplasmática de bactérias (COCIANCICH et al., 1994), assim como no caso da sapecina, que apresenta elevada afinidade para cardiolipina, um grande fosfolipídeo de *S. aureus* (MATSUYAMA; NATORI, 1988) e, desse modo, as defensinas podem interagir com o fosfolipídeo na membrana lipídica, o que pode estar relacionado com formação de canais responsáveis pela atividade biológica (YI et al., 2014).

2.5.1.3 Classe dos AMPs com rica representação de prolina

Lebocinas foram inicialmente isoladas a partir da hemolinfa do bicho-da-seda *B. mori* após indução por *E. coli*, como peptídeos de 32 resíduos ricos em prolina O-glicosilados (HARA; YAMAKAWA, 1995). Alguns peptídeos ricos em prolina de insetos são O-glicosilados, enquanto outros não o são, mas em todos os casos, a O-glicosilação aumenta claramente a atividade antimicrobiana, assim como sua atividade se torna muito mais elevada na

presença de cecropina (YI et al., 2014). Um cDNA de lebocina de *B. mori* mostrou que a mesma é inicialmente sintetizada como uma proteína precursora de 179 resíduos, e o peptídeo ativo de 32 resíduos situa-se próximo a extremidade C-terminal do precursor. Estes resultados indicam que lebocinas ativas são geradas pela clivagem proteolítica das proteínas precursoras. Nos precursores de lebocina, existem vários motivos RXXR conservados que podem ser reconhecidos pelas enzimas tipo furina (HARA; YAMAKAWA, 1995).

Lebocinas são ativas contra bactérias Gram-negativas e Gram-positivas e alguns fungos. Lebocinas de *B. mori* são ativas contra bactérias Gram-negativas *Acinetobacter sp*. e *E. coli* (HARA; YAMAKAWA, 1995). Peptídeos sintéticos ricos em prolina lebocina-B de *M. sexta* (28 resíduos) e lebocina-C (27 resíduos) são ativos contra bactérias Gram-negativas *Serratia arcescens* e *Salmonella typhimurium*, e bactérias Gram-positivas *S. aureus* e *B. cereus*, bem como o fungo *Cryptococcus neoformans* (RAO; XU; YU, 2012).

Pequenos peptídeos antimicrobianos ricos em prolina de 16-34, ativos contra bactérias Gram-negativas, Gram-positivas e alguns fungos, também foram identificados nas ordens Diptera, Hymenoptera, Hemiptera e Lepidoptera, apresentando resíduos e nomes diferentes, tais como drosocina e metchnikowina a partir de *D. melanogaster*, pyrrhocoricina de *Pyrrhocoris apterus*, formaecina de *Myrmecia gulosa*, apidaecina e abaecina de abelhas (YI et al., 2014).

O nome comum, PR-AMPS, para esta família de peptídeos antimicrobianos de insetos ricos em prolina foi proposto por Scocchi, Tossi e Gennaro (2011), uma vez que esses peptídeos foram identificados em, pelo menos 4 ordens de insetos, não apenas em Lepidoptera. PR-AMPS de diferentes ordens de insetos podem ser diferentes nas sequências de aminoácidos dos precursores e/ou em seus processos de transformação, assim como alguns precursores são mais curtos e não contêm os motivos conservados RXXR como os encontrados em precursores de lebocina, mas todos os peptídeos ativos contêm altos teores de prolina. O fato de um membro da família PR-AMP ser ativo ou não contra bactérias Gram-negativas, Gram-positivas ou fungos pode depender da ligação deste PR-AMP à superfície microbiana e de suas alterações estruturais e/ou conformacionais (YI et al., 2014).

2.5.1.4 Classe dos AMPs com rica representação de glicina

Atacinas foram inicialmente purificadas a partir da hemolinfa de pupas de *H. Cecropia* induzidas com bactérias, apresentando massas moleculares de 20 a 23 kDa e pontos isoelétricos (pI) entre 5,7 e 8,3. Estas isoformas de atacinas podem ser divididas em dois grupos, as atacinas básicas (A-D) e atacinas ácidas (E e F) (HULTMARK et al., 1980). Atacinas básicas e ácidas são altamente semelhantes em relação as suas sequências de aminoácidos, apresentando apenas maiores teores de resíduos do aminoácido aspartato em atacinas ácidas, mas elas são codificadas por dois genes separados (KOCKUM et al, 1984).

Atacinas são sintetizadas como pré-pró-proteínas que contêm um peptídeo sinal, um pró-peptídeo (P domínio), um domínio N-terminal, seguido por dois domínios ricos em glicina (domínios G1 e G2). Um motivo conservado RXXR, que pode ser reconhecidos pelas enzimas tipo furina, está presente no N-terminal pró-peptídeo de atacinas (KOCKUM et al., 1984), indicando que atacinas maduras são produzidas por processamento de pro-atacinas pelas enzimas tipo furina. O pro-peptídeo de atacina é necessário para a secreção de pró-atacina, que não apresenta atividade biológica, e é removida durante ou após a passagem pelo compartimento do Golgi (YI et al., 2014).

Atacinas já foram identificadas em muitas espécies de lepidópteros, incluindo *H. cecropia* (HULTMARK et al., 1980), *Heliothis virescens* (Ourth, 2004), *B. mori* (HARA; YAMAKAWA, 1995), *Trichoplusia ni* (LUNDSTROM et al, 2002), *Hyphantria cunea* (KWON et al., 2008), *S. cynthia* (FAYE et al, 1975), *Manduca sexta* (ZHU et al., 2003), *Helicoverpa armigera* (WANG et al., 2010), e *Spodoptera exígua* (HWANG; KIM, 2011), bem como em algumas espécies de dípteros como *D. melanogaster* (LEMAITRE; HOFFMANN, 2007) e *Glossina morsitans* (WANG; LAI, 2010).

A maioria das atacinas é ativa contra *E. coli* e algumas bactérias Gram-negativas selecionadas (HULTMARK et al., 1980). Atacina-A1 de *G. morsitans* apresenta atividade contra *E. coli* e contra o protozoário *Trypanosoma brucei in vitro* (YI et al., 2014). Atacina-B de *H. cunea* é ativa contra *E. coli* Gram-negativas e *Citrobacter freundii*, bem como contra o fungo *C. albicans* (KWON et al., 2008). Atacina de *S. exigua* é ativa contra bactérias *E. coli* Gram-negativas, *Pseudomonas cichorii*, *B. subtilis* Gram-positivas e *Listeria monocytogenes* (HWANG; KIM, 2011).

Atacinas podem inibir o crescimento de bactérias *E. coli* aumentando a permeabilidade da membrana externa bacteriana e podem inibir a síntese de várias proteínas de membrana externa bacteriana, incluindo OmpC, OmpF, OmpA, e LamB através da ligação ao LPS bacteriano, sem mesmo entrar na membrana interna ou no citoplasma (CARLSSON et al., 1998).

Geralmente atacinas adotam uma estrutura em espiral aleatória, que em solução aquosa é convertida para uma estrutura helicoidal em ambiente hidrofóbico, que é semelhante à conversão estrutural realizada por cecropinas e gloverinas. Assim, a estrutura helicoidal de atacinas pode ser a responsável pelas atividades antimicrobianas. Atacinas contêm dois domínios ricos em glicina (domínios G1 e G2) na extremidade C-terminal. Quando atacinas adotam uma estrutura helicoidal aleatória na hemolinfa, se torna mais suscetível a ação de proteases. Assim, existe uma probabilidade de encontrar pequenos peptídeos antimicrobianos ricos em glicina (7-10 kDa) semelhantes a atacinas, que correspondem a produtos da proteólise da proteína atacina (YI et al., 2014).

Gloverina é uma proteína antibacteriana de ~ 14 kDa, purificada inicialmente a partir da hemolinfa de pupas de *Hyalophora gloveri*, que contêm alto teor (> 18%) de resíduos de glicina, é estável ao calor, e a grande maioria são básicas ou altamente básicas (pI ~ 8.3 ou pI> 9,0), mas algumas poucas gloverinas podem ser neutras ou ácidas (AXEN et al., 1997).

Gloverinas são sintetizadas como pré-pró-proteínas e a região N-terminal de pró-gloverinas contêm um motivo RXXR conservado (XU et al., 2012). Assim, é provável que gloverinas maduras sejam produzidas após a remoção da pró-região N-terminal (~ resíduo 22-26) por enzimas do tipo furina. Porém, a pró-gloverina recombinante de *T. ni* também é ativa contra *E. coli*, com atividade comparável à de gloverina madura de *H. gloveri* (LUNDSTROM et al., 2002), sugerindo que a remoção da pró-região pode não ser necessária para tornar algumas pró-gloverinas ativas.

Gloverinas foram identificadas apenas em Lepidoptera, incluindo *H. armigera* (MACKINTOSH et al., 1998), *T. ni* (LUNDSTROM et al., 2002), *G. mellonella* (SEITZ et al., 2003), *Antheraea Mylitta* (Gandhe et al., 2006), *M. sexta* (XU et al, 2012), *Diatraea saccharialis* (ETEBARI et al., 2011), *Plutella xylostella* (SILVA et al., 2010), *S. exigua* (HWANG; KIM, 2011), e *B. mori* (KAWAOKA et al., 2008) e apresentam atividade principalmente contra *E. coli* (AXEN et al, 1997; KAWAOKA et al., 2008; MACKINTOSH et al., 1998). Mas gloverinas de *T. ni* também são ativas contra vírus(LUNDSTROM et al., 2002), gloverinas de *S. exigua*

são ativas contra a bactéria Gram-positiva *Flavobacterium sp.*, e gloverinas de *M. sexta* são ativas contra a bactéria Gram-positiva *B. cereus* e os fungos *Saccharomyces cerevisiae* e *C. neoformans* (XU et al., 2012).

Não está claro como as gloverinas de diferentes espécies da ordem Lepidoptera têm atividade contra *E. coli*, bactérias Gram-positivas, fungos ou mesmo o vírus, mas um estudo realizado por Yi e colaboradores (2014) sugere que as gloverinas possam interagir com LPS das bactérias através de ligações a diferentes porções do mesmo.

As gloverinas adotam uma estrutura espiral aleatória em solução aquosa, mas sofrem transição conformacional da estrutura em α-helicoidal em ambiente hidrofóbico (XU et al., 2012). Estudos sugerem que a ligação de gloverinas ao LPS bacteriano, ou outras moléculas de superfície microbiana, seja o pré-requisito para que ocorra a conversão conformacional das gloverinas de estrutura aleatória para α-helicoidal, sendo, neste caso, a ligação à superfície microbiana a chave para a atividade de gloverinas contra *E. coli* ou outros microrganismos. Este mecanismo pode também aplicar-se a cecropinas, moricinas e atacinas, uma vez que todas elas sofrem alterações conformacionais após a ligação ao ambiente hidrofóbico da superfície microbiana (YI et al., 2014).

Gloverinas apresentam uma elevada quota de semelhanças entre elas, como por exemplo, a gloverina de *M. sexta*, é de 48 a 74% idêntica a gloverinas de outras espécies de lepidópteros (XU et al., 2012). Outros estudos com as gloverinas sugerem que, além de ter atividade antimicrobiana, elas podem também desempenhar um papel no desenvolvimento e/ou reprodução nos lepdópteros (HWANG; KIM, 2011; XU et al., 2012).

2.5.2 Potenciais aplicações de AMPs de insetos

AMPs de insetos podem ter potenciais aplicações na agricultura, no controle de vetores de doenças e na medicina. AMPs foram projetados em plantas para conferir resistência a patógenos bacterianos e fúngicos, assim como as defensinas vegetais que têm sido expressas em arroz transgênico (JHA; CHATTOO, 2010), trigo (LI et al., 2011), banana (GHAG; SHEKHAWAT; GANAPATHI, 2012), tomate (ABDALLAH et al., 2010), melão (NTUI et al., 2010), amendoim , tabaco (SWATHI ANURADHA et al., 2008), e *Arabidopsis* (KAUR et al., 2012).

A expressão das defensinas de inseto galerimicina (*G. mellonella*) e cecropina (sarcotoxin-IA) no tabaco transgênico também confere resistência a fungos patogênicos (MITSUHARA et al., 2000). Expressão de cecropinas em plantas transgênicas, incluindo arroz e tomate, pode conferir resistência a patógenos bacterianos e fúngicos (JAN; HUANG; CHEN, 2010). Metchnikowina (peptídeo rico em prolina) foi expressa em cevada transgênica para aumentar a resistência (RAHNAMAEIAN; VILCINSKAS, 2012). Peptídeos quiméricos formados através da combinação de regiões ativas de dois AMPs também têm sido utilizados em plantas transgênicas para aumentar o espectro de resistência a patógenos (YEVTUSHENKO et al., 2005).

AMPs de inseto mostraram atividade contra alguns parasitas, incluindo *Plasmodium* e *Trypanosoma*, em duas pesquisas recentes que descrevem o atual progresso sobre peptídeos antimaláricos e antiparasitários (BELL, 2011; PRETZEL et al., 2013). Cecropinas e defensinas mostraram-se ativas contra parasitas (FIECK et al, 2010).

A expressão de AMPs em vetores transgênicos, como mosquitos é uma nova abordagem para matar os parasitas ou bloquear a transmissão dos mesmos. A expressão de cecropina-A em *Anopheles gambiae* transgênico, um vetor de parasitas da malária humana, pode reduzir o número de oocistos de *Plasmodium berghei* (KIM et al., 2010), e co-expressão de cecropina-A e defensina-A em *Aedes aegypti* transgênico pode bloquear a transmissão de *Plasmodium* (KOKOZA et al., 2010).

Diversos estudos têm demonstrado o efeito antiviral de peptídeos obtidos a partir de insetos (CHERNYSH et al., 2002). Hiraki e colaboradores (1997) demonstraram a presença de uma molécula com atividade antiviral obtida das fezes de lagartas capaz de combater o HVJ (vírus Sendai), HSV (virus herpes simplex tipe-1) e HIV (vírus da imunodeficiência adquirida). Cueto e colaboradores demonstraram a atividade antiviral do extrato de própolis de abelhas contra o calicivírus felino (FCV), adenovírus canino 2 (CAV-2) e vírus da diarréia viral bovina (BVDV).

Chernysh e colaboradores (2002) purificaram uma proteína com 12,65 kDa da hemolinfa de *Calliphora vicina* que controla a infecção viral intracelular e inibe a replicação do vírus da influenza A e B, desencadeando respostas intracelulares quando adicionada antes da infecção do vírus de forma semelhante ao interferons de vertebrados.

Em 2004, Popham e colaboradores relataram uma redução significativa nos títulos do baculovírus HzSNPV (Heliothis zea nuclear polyhedrosis vírus) devido à ação de uma proteína

antiviral presente na hemolinfa de larvas de *H. virescens*. Neste mesmo ano foi encontrada atividade antiviral na hemolinfa da lagarta *H. virescens* contra sete vírus de DNA e RNA, incluindo HIV-1. A atividade antiviral foi atribuída a um pequeno peptídeo de peso molecular de 916 Da e apenas seis aminoácidos (OURTH, 2004).

Uma proteína de aproximadamente 20 kDa presente na hemolinfa de *Lonomia obliqua* foi identificada, isolada e caracterizada, apresentando atividade antiviral contra vírus humanos e animais. Os vírus Pólio (Sabin 1) e Sarampo (Edmonton) foram usados para determinar a atividade antiviral da proteína purificada que levou a uma redução de 157 vezes na produção do vírus do sarampo e uma redução de 61 vezes na produção de vírus vacinal Sabin 1 (GRECO et al., 2009). Neste trabalho foi sugerido que essa proteína possa atuar em diferentes estágios da replicação do vírus, semelhante ao alloferon, ou em fases tardias da infecção por vírus, como demonstrado por Popham e colaboradores (2004) com o peptídeo extraído de *H. Virescens*.

Esta sequência foi clonada e expressa em sistema baculovírus/célula de inseto, sendo que a proteína recombinante foi capaz de ocasionar uma redução com valores próximos a 1000 vezes na produção do picornavirus, vírus da encefalomiocardite (EMC), em células L929 (CARMO et al., 2012).

O desenvolvimento de agentes antivirais, especialmente para o surgimento recente de novas cepas variantes altamente agressivas de vírus humanos (como o coronavírus, causador da Síndrome respiratória aguda grave - SARS e os vírus da gripe H5N1 e H1N1) são de grande importância, uma vez que afetam centenas de milhões de pessoas no mundo. Em 2010, cerca de 40 compostos antivirais estavam em fase de teste (DELCROIX; RILEY, 2011).

Só nos Estados Unidos podem ser encontrados quatro agentes antivirais para influenza. O zanamivir e oseltamivir são agentes pertencentes à classe de inibidores da neuromidase e ativos contra os vírus influenza A e B. Já a amantadina e a rimantadina são medicamentos da classe conhecida como adamantanes, sendo eficazes apenas contra vírus influenza A, demonstrando a necessidade de desenvolvimentos de novos agentes virais (FIORE et al., 2011).

2.5.3 Exploração da biodiversidade em busca de moléculas biologicamente ativas

Uma das maneiras de se extrair valor econômico da biodiversidade é a bioprospecção, termo que pode ser entendido como a busca sistemática por organismos, genes, enzimas, compostos, processos e partes provenientes de seres vivos, que tenham potencial econômico e, eventualmente, levam ao desenvolvimento de um produto (SACCARO JR, 2011).

A bioprospecção tem se apoiado em todo um conjunto de conhecimentos e práticas que emergem com o desenvolvimento do capitalismo e do conhecimento científico-tecnológico. Isto tem sido bem documentado na literatura médica, ao descrever as primeiras drogas obtidas a partir do conhecimento das propriedades químicas de determinados componentes biológicos. Apenas para citar um exemplo: o uso do veneno de cobras para o tratamento de doenças relacionadas a distúrbios circulatórios, como a chamada hipertensão (ARTUSO, 2002).

Com o avanço da engenharia genética, da biologia molecular e de todo um conjunto de novas tecnologias de processamento de dados e de informação, o interesse pela descoberta e conhecimento das inúmeras possibilidades disponíveis na natureza para uso na obtenção de novos produtos e processos farmacêuticos, agroquímicos, cosméticos e vários outros ligados à indústria de alimentos, vem crescendo acentuadamente (ARTUSO, 2002).

Diversas técnicas são utilizadas para extraír informações moleculares presentes na biodiversidade, isolando e analisando as moléculas com potenciais biológicos e caracterizando as moléculas bioativas com potencial comercial (ARTUSO, 2002). As áreas de genômica, pós-genômica e proteômica, são consideradas as mais impactantes para os setores de medicina e saúde e biofármacos. Os biofármacos e biomarcadores baseados na genômica constituem uma das grandes apostas do setor biofarmacêutico, especialmente os biomarcadores para câncer (MORALES, 2006).

A bioprospecção aparece como um tópico recorrente em estudos internacionais de mercado da biotecnologia, como por exemplo, os estudos relacionados ao sequenciamento genético em massa de microrganismos do Mar de Sargaços (região alongada no meio do Atlântico Norte, cercado por correntes oceânicas), além dos estudos que utilizam a técnica de metagenômica, que consiste no sequenciamento do DNA de todo um ambiente, com o qual os pesquisadores encontraram cerca de 1,2 milhões de novos genes, a partir de

um número estimado de 1800 espécies até então desconhecidas (DEPARTMENT OF ENERGY – US, 2006; KALYUZHNAYA et al., 2008).

Em 2003, estimava-se que cerca de 40 % dos medicamentos disponíveis na terapêutica eram desenvolvidos a partir de fontes naturais, sendo 25 % de plantas, 13 % de microrganismos e 3 % de animais (CALIXTO, 2003). Atualmente, esse valor já cresceu para cerca de 50 % (SACCARO JR, 2011). Considerando a grande quantidade de fármacos atuais desenvolvidos a partir de moléculas biológicas, torna-se evidente o papel da bioprospecção no setor farmacêutico.

A larga distribuição dos insetos em regiões com características totalmente diferentes tem estimulado as pesquisas de novos agentes terapêuticos nesta classe de artrópodos (CHERNYSH et al., 2002). Vários estudos realizados demonstraram a presença de substâncias farmacologicamente ativas na hemolinfa de insetos com ação imune antiviral, antibacteriana e antifungica, que são representados, principalmente por AMPs (LOWENBERGER et al., 1999; ZHU et al., 2000).

Como uma rica fonte de AMPs, os insetos atraem a atenção dos cientistas tanto para a compreensão da biologia básica do sistema imunológico como para a procura de modelos moleculares para desenho de novas drogas antibióticas mais eficazes, devido ao rápido desenvolvimento de bactérias patogênicas resistentes aos antibióticos clássicos nos últimos anos (MOELLERING, 1998). Além da necessidade do desenvolvimento de agentes antivirais, especialmente para aqueles que possam atuar sobre novas cepas variantes e altamente agressivas de vírus humanos (como o coronavírus, causador da Síndrome Respiratória Aguda Grave - SARS e os vírus da gripe H5N1 e H1N1), uma vez que afetam centenas de milhões de pessoas no mundo (DELCROIX; RILEY, 2011). No contexto do desenvolvimento de novas drogas, os AMPs apresentam várias propriedades desejáveis, tais como tolerância ao calor, amplo espectro antimicrobiano e baixa toxicidade para as células eucarióticas (TIAN et al., 2010).

2.6 A pesquisa genômica

A pesquisa genômica consiste no conjunto de informações genéticas presentes no DNA de um determinado organismo. A molécula de DNA é estática e está presente, com a mesma constituição, em todas as células do organismo. A decifração desse conteúdo

estático de DNA é a tarefa da genômica estrutural (DAVIES, 2001). Se o genoma consiste no sequenciamento da molécula de DNA de uma célula, o transcriptoma consiste no sequenciamento do conteúdo dos RNAs mensageiros (mRNAs) produzidos em uma célula específica e sujeita a determinada condição (DAVIES, 2001; LANDER et al., 2001).

Enquanto uma célula apresenta apenas um genoma estático, a mesma pode apresentar milhares de diferentes regiões a serem transcritas, já que a expressão de genes depende do tempo e das diversas condições à qual a célula está sendo submetida, como o grau de maturação, a temperatura à qual ela está sujeita, os nutrientes presentes no meio, a presença de algum agente mutagênico específico e mais milhares de outros fatores (PROSDOCIMI et al., 2002).

O transcriptoma representa apenas as regiões do genoma que estão sendo utilizado num determinado momento. E essa parte do genoma expresso é diferente para cada tipo celular. Existem genes que são expressos apenas na pele, outros no cérebro e alguns nos testículos. Alguns genes são ainda mais expressos quando a célula é submetida a um choque térmico, à restrição calórica ou à falta de oxigênio. Assim, os estudos de transcriptoma podem mostrar a adaptação da célula a uma determinada condição e podemos estudar os genes que ficam ativos quando dessa condição (DAVIES, 2001; PROSDOCIMI et al., 2002).

Ao discutirmos as diferenças entre as análises de genomas e transcriptomas, vale notar primeiramente que uma análise não exclui a outra e são estudos que, apesar de relacionados, provém respostas para perguntas diferentes. Enquanto o genoma é apenas um, existem vários transcriptomas possíveis para uma mesma espécie. Algumas perguntas, entretanto, só podem ser obtidas quando se observa o genoma expresso, enquanto outras, apenas quando se observa o genoma estático. Por exemplo, por mais que se obtenham sequências geradas de diferentes tecidos de um organismo, nunca é possível dizer que ele não apresenta um determinado gene através de análises transcriptômicas. De forma contrária, quando se obtém toda a sequência do genoma de um organismo, não é possível saber todo o repertório de genes que é super-expresso numa determinada condição, como, por exemplo, quando a célula se torna cancerígena. Dessa forma, as análises de genoma e transcriptoma são complementares e ambas são importantes para responder perguntas específicas (DAVIES, 2001; PROSDOCIMI et al., 2002; VENTER et al., 2001).

2.6.1 Sequenciamento de última geração – NGS

Durante os últimos 30 anos, o método de Sanger tem sido a abordagem padrão dominante para sequenciamento de DNA. A partir de 1977, foram desenvolvidos diferentes métodos e várias empresas desenvolveram sequenciadores capilares de larga escala, que são vendidos ainda hoje e o número de sequências de DNA produzidas vem aumentando exponencialmente até o presente momento (GUŽVIĆ, 2013; POP; SALZBERG, 2008).

O lançamento comercial da primeira plataforma de pirosequenciamento em 2005 inaugurou a nova era de análise genômica e transcriptômica de alto rendimento, agora conhecida como "sequenciamento de última geração" (NGS). Num espaço relativamente curto de tempo, NGS alterou fundamentalmente a pesquisa envolvendo sequenciamento de DNA, permitindo aos investigadores realizar experiências que antes eram tecnicamente inviáveis (GUŽVIĆ, 2013; VOELKERDING; DAMES; DURTSCHI, 2009).

Plataformas NGS compartilham uma característica tecnológica comum - sequenciamento em massa de moléculas de DNA amplificados por clonagem ou individuais que são separadas espacialmente em células. A maior eficiência de NGS advém do uso da clonagem *in vitro* e de sistemas de suporte sólido para as unidades de sequenciamento, não precisando mais do intensivo trabalho laboratorial de produção de clones bacterianos, da montagem das placas de sequenciamento e da separação dos fragmentos em géis. A clonagem *in vitro* em suporte sólido permite que milhares de leituras possam ser produzidas de uma só vez em um único instrumento, dependendo da plataforma (CARVALHO; SILVA, 2010; POP; SALZBERG, 2008).

As novas plataformas de sequenciamento apresentam a grande vantagem de permitir um sequenciamento altamente representativo de genomas e/ou transcriptomas e seu emprego tem revolucionado, principalmente, a transcriptômica com a geração de dados altamente reprodutíveis e informativos e com precisão na quantificação de transcritos (CARVALHO; SILVA, 2010).

Atualmente, os equipamentos conhecidos como sequenciadores de nova geração, são capazes de suprir novas exigências, como a demanda de um maior número de pares de bases sequenciados em menor espaço de tempo, somado à necessidade da redução de custos, gerando uma quantidade muito maior de dados, a um custo reduzido quando

comparado com métodos usados anteriormente, como o método de Sanger (GUŽVIĆ, 2013; POP; SALZBERG, 2008).

2.6.1.1 Considerações sobre o método de pirosequenciamento

Em 1993, Nyren e colaboradores descreveram uma abordagem de sequenciamento baseado na detecção quimioluminescente do pirofosfato liberado durante a incorporação de desoxinucleotídeo trifosfato (dNTP) mediada pela polimerase. Um refinamento desta abordagem realizado por Ronaghi e colaboradores em 1996 e 1998, serviu de base para o desenvolvimento comercial do pirosequenciamento.

Em 2000, Jonathan Rothberg funda 454 Life Sciences, que desenvolveu a primeira plataforma NGS disponível comercialmente, o GS20, lançado em 2005. Combinando as técnicas de PCR em emulsão de uma única molécula com o pirosequenciamento, foi relizado na plataforma 454 Life Sciences GS20 o sequenciamento shotgun de todos os 580.069 pb do genoma do *Mycoplasma genitália* com uma cobertura de 96 % e uma acurácia de 99,96 % em uma única corrida (MARGULIES et al., 2005).

Utilizando a metodologia de pirosequenciamento para análises de transcriptoma, o RNA mensageiro é o molde e, a partir deste molde, ocorrem as reações de transcrição reversa realizadas pela enzima Transcriptase reversa, que adiciona os nucleotídeos e monta o DNA unifilamentar, que é formado fora do organismo, ou seja, sinteticamente. Através do pareamento das duas fitas, uma de DNA e outra de RNA, é formada uma molécula bifilamentar chamada de DNA complementar ou simplesmente cDNA (GUŽVIĆ, 2013; SCHUSTER, 2008; VOELKERDING; DAMES; DURTSCHI, 2009).

No método de pirosequenciamento, o sequenciamento é baseado em síntese e a leitura da sequência nesse sistema é realizada a partir de uma combinação de reações enzimáticas que se inicia em um evento de incorporação de um desoxinucleotídeo à cadeia que, produz a liberação de um pirofosfato inorgânico (PPi). Em seguida, esse pirofosfato é utilizado pela enzima ATP sulfurilase para produzir ATP, sendo este utilizado pela luciferase para oxidar a luciferina, produzindo um sinal de luz. A intensidade luminosa do sinal, é detectada e gravada de acordo com o fluxo da reação de cada dNTP, possibilitando a determinação do nucleotídeo que foi incorporado à cadeia de DNA (HUSE et al., 2007; RONAGHI et al., 1998).

O registro da emissão da luz detectada é traduzido por algoritmos em formato de picos, que em conjunto formam o pirograma, que corresponde exatamente a sequência linear final de nucleotídeos. A enzima Apirase é responsável por degradar os dNTPs que não foram utilizados durante a reação, evitando assim, falsos resultados nas próximas etapas que utilizam outros nucleotídeos (HUSE et at., 2007).

2.6.2 Transcriptoma de lepidópteros

Os lepidópteros (borboletas e mariposas) são notavelmente diversificados e a sua perceptibilidade resultante da grande variedade de padrões de cores atraentes presentes em muitas espécies explica grande parte da atenção que esses insetos têm recebido (GRIMALDI; ENGEL, 2005; VAN NIEUKERKEN et al., 2011). Há uma longa tradição de pesquisa e uma série de disciplinas que usam modelos de lepidópteros para investigar fenômenos biológicos fundamentais, incluindo o desenvolvimento e regulação dos genes, processos genéticos de populações (fluxo gênico, colonização e extinção), adaptação morfológica, especiação e processos co-evolutivos, tais como interações entre a planta hospedeira e o inseto parasita (PAPANICOLAOU et al., 2008).

No estudo de Lepidópteros, a lagarta *Bombyx mori*, conhecida como bicho-da-seda, é um modelo utilizado para estudos de fisiologia e biologia molecular de insetos, assim como a lagarta do tabaco *Manduca sexta*, que tem desempenhado um papel central nos estudos das funções comportamentais, bioquímicas e fisiológicas dos insetos, obtendo muitos avanços fundamentais incluindo, por exemplo, estudos sobre a cutícula, o sistema nervoso, a muda, a regulação hormonal e a imunidade nos insetos (HOPKINS et al, 2000; HIRUMA; RIDDIFORD, 2009; SHIELDS; HILDEBRAND, 2001; KANOST; JIANG; YU, 2004; RIDDIFORD et al., 2003). Apesar destes importantes avanços, a falta de dados transcriptômicos dificultava a clonagem dos genes responsáveis por tais atividades (PAUCHET et al., 2010).

Até junho de 2009, os dados de genômica e transcriptômica de Lepidópteros eram completamente incipientes, onde apenas 401 sequências de codificação de DNA (*CDS – Coding sequence*) e 3.317 marcadores de sequências expressas (*ESTs – Expressed Sequence Tags*) estavam presentes no *GenBank*. Estes conjuntos de dados de ESTs foram gerados a partir de estudos relacionados com proteínas de ligação olfatória (ROBERTSON et al., 1999),

relacionadas à resposta de defesa (ZHU et al., 2000) e uma coleção não publicada de ESTs isolados a partir de hemócitos no ano de 2006 (PAUCHET et al., 2010).

A partir desta data, houve um aumento considerável na quantidade de dados de *ESTs* em um número crescente de espécies de lepidópteros. Diversos conjuntos de *ESTs* de tamanhos médios e moderados vêm sendo produzidos atualmente e podem ser uma fonte poderosa de marcadores para mapeamento comparativo, análise genética populacional e estudos de evolução adaptativa. Por exemplo, há grandes conjuntos de dados genômicos públicos para o bicho-da-seda *Bombyx mori*, a traça *Spodoptera frugiperda*, as borboletas *Bicyclus anynana*, *Heliconius melpomene* e *Heliconius erato*, entre outras. A geração de sequências para estas e outras espécies levou à descoberta de que cerca de metade dos genes sequenciados em Lepidoptera apresentam pouca ou nenhuma semelhança de sequência com proteínas de outros níveis taxonômicos (PAPANICOLAOU et al., 2008).

A mudança drástica de situação ocorreu como resultado da introdução de novas tecnologias de sequenciamento, capazes de fornecer de maneira rápida e eficaz grandes quantidades de dados genômicos e transcriptômicos, aumentando a cada ano as informações disponíveis sobre a arquitetura genética de um número crescente de espécies não-modelo (PAUCHET et al., 2010). O pirosequencimento de última geração tornou-se uma importante ferramenta em estudos transcriptômicos que permitem a identificação de genes alvos que são diferencialmente expressos em diferentes tecidos ou células, durante o desenvolvimento ou após a ativação de respostas imunes. Esta tecnologia tem sido utilizada, por exemplo, para caracterizar o perfil transcriptômico específico de vários tecidos de diversos lepidópteros (VOGEL et al., 2011).

Pauchet e colaboradores (2010) realizaram um estudo que forneceu informações sobre o perfil transcriptômico do intestino médio de lepidópteros com o intuito de avaliar as interações ecológicas-fisiológicas desempenhadas nesse tecido que apresenta um papel fundamental na digestão e desintoxicação de alimentos e de xenobióticos. Este estudo apresentou dados de sequências de enzimas envolvidas nesses processos, além de demostrar a presença de proteínas relacionadas à resposta imune, uma vez demonstrado que lagartas herbívoras, tais como *M. sexta*, necessitam ingerir grande quantidade de material vegetal durante o seu desenvolvimento e, durante a ingestão, ficam susceptíveis a uma variedade de microrganismos potencialmente nocivos, promovendo o risco de infecção através do intestino. Além da existência de evidências que sugerem que o intestino médio

participa ativamente no desencadeamento da resposta imune após a ingestão de alimentos contaminados por microrganismos e, pelo menos em *Drosophila*, isso envolve uma complexa resposta imune sistêmica capaz de distinguir entre a flora intestinal nativa e patógenos invasores (PAUCHET et al., 2010).

Vogel e colaboradores (2011) realizaram um estudo do perfil transcriptômico relacionado à imunidade da traça *Galleria mellonella,* que é um sistema modelo confiável e comprovado para estudos de imunidade, uma vez que a espécie é adequada para identificar características antigas da imunidade inata em lepidópteros, pois a *G. mellonella* pertence à família Pyralidae, que se encontra em posição filogenética basal dentro da ordem Lepidoptera. As sequências geradas foram analisadas através da combinação de duas plataformas de sequenciadores de última geração e resultou na identificação de novas sequências e de variações de proteínas e peptídeos anteriormente identificados em outras espécies.

Embora a ordem Lepidoptera apresente mais de 160.000 espécies distribuídas em 124 famílias, um levantamento de dados demonstrou que pouco mais de 50 espécies apresentam sequências gênicas depositadas nos bancos de dados do *NCBI*. De acordo com as informações disponíveis no *Genbank*, dentre as espécies e famílias de Lepidoptera que apresentam sequências gênicas depositadas em seus bancos de dados, não foi encontrada nenhuma sequência originada de espécies da família Megalopygidae, nem ao menos de um único membro da superfamília Zygaenoidea (NCBI, 2016).

Atualmente, o sequenciamento genômico completo de 23 espécies está disponível ao público e os dados de suas sequências e da anotação, facilitam a análise comparativa de outras espécies com informação genômica menos completa (NCBI, 2016; PAPANICOLAOU et al., 2008). Essas 23 espécies são todas representantes de apenas 15 famílias: Arctiidae, Bombycidae, Crambidae, Geometridae, Hesperiidae, Nymphalidae, Noctuidae, Papilionidae, Pieridae, Plutellidae, Pyralidae, Saturniidae, Sphingidae, Tineidae, Tortricidae (NCBI, 2016).

Como resultado, há uma riqueza de conhecimento ecológico e genético para Lepidoptera, contudo, muitas espécies e até mesmo famílias da ordem não apresentam estudos e ainda encontram-se praticamente desconhecidas, com ausência de dados até mesmo sobre sua biologia e habitat, como é o caso da lagarta *Podalia orsilochus*, que mesmo tendo sido descrita há mais de três séculos, somente é encontrada na literatura em

relatos médicos, relacionados a acidentes com os seres humanos em casos de envenenamento (LEPESQUEUR, 2012).

2.7 Uma visão global da bioinformática

A bioinformática é considerada uma linha de pesquisa que envolve aspectos multidisciplinares e a união de diversas linhas de conhecimento, como a ciência da computação, a engenharia de softwares, a matemática, a estatística e a biologia molecular. Ela pode ser definida como uma modalidade que abrange todos os aspectos de aquisição, processamento, armazenamento, distribuição, análise e interpretação da informação biológica (BORÉM; SANTOS, 2001; POP; SALZBERG, 2008).

A bioinformática surgiu no momento em que se iniciou a utilização de ferramentas computacionais para a análise de dados genéticos, bioquímicos e de biologia molecular e tem como finalidade principal desvendar a grande quantidade de dados que vem sendo obtidos através de sequências de DNA e proteínas. Para a análise de genomas completos, a informática é imprescindível e a biologia molecular moderna não estaria tão avançada hoje, não fossem os recursos computacionais existentes. A bioinformática, apesar de ser uma ciência nova e em desenvolvimento é hoje uma necessidade para a análise de dados em biologia molecular, devido ao seu crescimento acelerado (POP; SALZBERG, 2008).

2.7.1 Agrupamento de sequências

O agrupamento de sequências (*sequence assembly*) é o procedimento que permite a montagem do genoma e é importante porque ainda não existe nenhuma técnica que permita o sequenciamento de moléculas de DNA de mais de mil pares de bases. Várias sequências representando pedaços de genes são geradas e observa-se a posição onde elas se sobrepõem. Realizando a sobreposição de vários trechos de sequência é possível montar todo o genoma, entretanto, não é possível realizar esta montagem à mão e para isso são utilizados os algoritmos de montagem de genoma (PROSDOCIMI et al., 2002).

Diversos programas utilizam algoritimos de montagem, que comparam as sequências entre si, de forma a encontrar quais delas são idênticas ou contêm regiões parecidas o suficiente para que sejam reunidas em uma só, assim como na análise genômica. Assim,

esses programas apresentam uma saída contendo as sequências que foram agrupadas, chamadas de consensos ou *contigs*, e as sequências que não foram reunidas (por não apresentarem similaridade suficiente com nenhuma outra), chamadas de *singlets* ou *singletons* (Figura 11) (MUDADO, 2007; PROSDOCIMI et al., 2002).

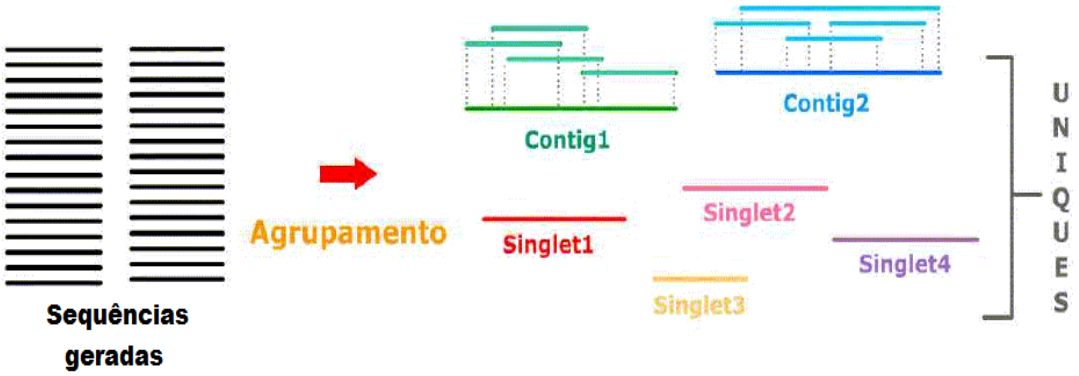


Figura 11 - Agrupamento de sequências.
Legenda: O agrupamento produz as sequências não-redundantes chamadas de *uniques*. As *uniques* são o conjunto das sequências consenso e das *singlets*.
Fonte: Prosdocimi, F. et al. (2002).

Cada uma das sequências resultantes do agrupamento (seja ela uma *singlet* ou um *contig*) é chamada de *unique*, que consiste no conjunto não redundante de moléculas. Considerando uma análise ideal, cada uma das sequências unique deve representar um gene distinto, porém na prática, a presença de famílias gênicas (apresentando regiões de similaridade dentro dos genes) e de genes duplicados dificulta a obtenção desse resultado ideal e, muitas vezes, a sequência unique pode representar mais de um gene (AMARAL; REIS; SILVA, 2007; PROSDOCIMI et al., 2002).

O agrupamento das sequências é importante devido, principalmente, aos seguintes fatores: diminuição da redundância das sequências, aumento do tamanho das sequências facilitando a anotação por similaridade (OLIVEIRA; JOHNSTON 2001) e o aumento do nível de confiabilidade de cada sequência (MILLER et al., 1999).

2.7.2 Bancos de dados em biologia molecular

As bases de dados em biologia molecular são importantes principalmente para proporcionar à comunidade científica uma forma de tornar acessíveis os dados biológicos

produzidos em todo mundo, de forma fácil, rápida e inteligente. O *NCBI*, ou Centro Nacional para Informação Biotecnológica dos EUA, foi criado pelo *NIH* (*National Institutes of Health*) em 1988 para abrigar esse tipo de informação (AMARAL; REIS; SILVA, 2007; WELLER et al., 2002).

O *NCBI* é considerado o banco de dados central sobre informações genômicas. Vários outros bancos de dados similares estão distribuídos por países da Europa, como o *EMBL* (*European Molecular Biology Laboratory* ou Laboratório Europeu de Biologia Molecular) e Japão, como o *DDBJ* (*DNA Data Bank of Japan* ou Banco de dados de DNA do Japão) (TATENO et al., 2002), mas todos trocam dados em um intervalo de 24 horas com o *NCBI* (AMARAL; REIS; SILVA, 2007).

Existem basicamente dois tipos de bancos de dados disponíveis para utilização e pesquisa de genes e proteínas, denominados de bancos primários e secundários. Os bancos de dados primários apresentam resultados de dados experimentais que são publicados com alguma interpretação, onde não há uma análise cuidadosa desses dados com relação aos outros publicados anteriormente, que é o caso, por exemplo, do *GenBank*, *EMBL* (STOESSER et al., 2002) e *PDB* (WESTBROOK et al., 2002). O *GenBank* é o principal banco de dados do NCBI e armazena e disponibiliza publicamente todas sequências de DNA (de sequências pequenas a genomas inteiros), RNA e proteínas depositadas (STOESSER et al., 2002).

Os bancos de dados secundários são aqueles onde há uma compilação e interpretação dos dados de entrada de forma que podem ser obtidos dados mais representativos e interessantes. Esses são os bancos de dados curados, como o *SWISS-PROT* e o *TrEMBL* (AMARAL; REIS; SILVA, 2007; BAIROCH; APWEILER, 2002).

Para cada diferente propósito, são construídas novas ferramentas de bioinformática e muitas delas são disponibilizadas via servidores na Internet e, através da criação de bancos de dados com as informações já processadas, acelera a investigação em diversas áreas como a medicina, a biotecnologia, a agronomia, etc (BORÉM; SANTOS, 2001).

2.7.3 Alinhamento de sequências

O alinhamento de sequências consiste no processo de comparar duas sequências (de nucleotídeos ou proteínas) de forma a observar seu nível de identidade. Essa técnica de

comparação de sequências é um dos pilares de toda a bioinformática (POP; SALZBERG, 2008; PROSDOCIMI et al., 2002).

Existem centenas de aplicações do alinhamento de sequências, como na identificação de genes e proteínas desconhecidas, na comparação da ordem de genes em genomas de organismos proximamente relacionados (sintenia), no mapeamento de sequências expressas para identificação de genes dentro de um genoma, na montagem de genomas e em diversas outras aplicações. Por exemplo, podemos alinhar duas sequências para descobrirmos o grau de similaridade entre elas, de tal forma que possamos ou não, inferir a uma delas, alguma propriedade já conhecida da outra (PROSDOCIMI et al., 2002). O alinhamento entre duas sequências pode ser feito de forma global ou local (Figura 12).

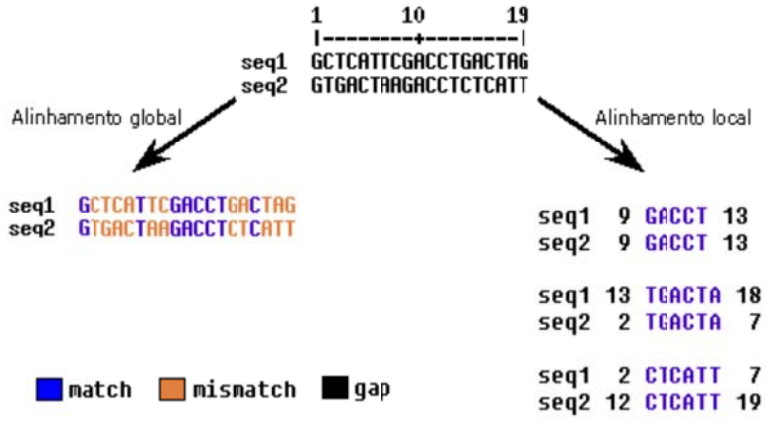


Figura 12 - Alinhamento global e local entre diferentes sequências.
Legenda: À esquerda, um exemplo de um alinhamento global das sequências e, à direita, um exemplo de um alinhamento local.
Fonte: Prosdocimi et al., 2002.

O alinhamento global é feito quando comparamos uma sequência de aminoácidos ou nucleotídeos com outra, ao longo de toda sua extensão. O algoritmo *Needleman-Wunsch* é o mais conhecido para realizar esse tipo de alinhamento, embora outros programas também o façam (CORPET, 1988). Nesse caso são dados valores em uma matriz de comparação para as similaridades (*matches*), divergências (*mismatches*) e falhas (*gaps*) encontradas durante o alinhamento das sequências. A soma dos valores do alinhamento, de acordo com essa matriz de comparação, resulta num valor, que é um escore de similaridade entre as sequências (MUDADO, 2007; PROSDOCIMI et al., 2002).

O alinhamento local acontece quando a comparação entre duas sequências não é feita ao longo de toda sua extensão, mas sim através de pequenas regiões destas. O principal programa utilizado para o alinhamento local de sequências de DNA é o *BLAST* (*Basic Local Alignment Search Tool* ou Ferramenta Básica de Procura por Alinhamento Local) que compreende um conjunto de algoritmos que possibilita a comparação de uma sequência de DNA ou proteína (Query) qualquer, com todas as sequências contidas em bases de dados genômicos de domínio público (AMARAL; REIS; SILVA, 2007; PROSDOCIMI et al., 2002).

Através do uso do alinhamento local, o *BLAST* identifica no banco de dados, a presença de uma sequência suficientemente parecida com a pesquisada, descartando, assim, rapidamente, os resultados não produtivos e estende a vizinhança da região de homologia detectada até não mais conseguir. Os programas *BLAST* foram desenvolvidos de modo a aumentar ao máximo a velocidade da busca por similaridade, mesmo correndo o risco de perder um pouco na sensibilidade do resultado (ALTSCHUL et al., 1997; POP; SALZBERG, 2008).

A rapidez da busca deve-se ao fato de que o programa utiliza uma heurística que quebra as sequências de entrada e das bases de dados e procura, inicialmente, similaridades entre elas. Outra vantagem de se utilizar o alinhamento local feito pelo *BLAST* é que, é possível identificar relações entre sequências que apresentam apenas regiões isoladas de similaridade (ALTSCHUL et al., 1997; MUDADO, 2007).

As sequências resultantes da comparação pelo *BLAST* são apresentadas de acordo com dois parâmetros: o valor do escore (*Score bits*) e o valor “E” (*e-value*). O valor de escore depende do tamanho do alinhamento, do número de matches/mismatches/gaps e da matriz de comparação de sequências utilizada. Já o valor “E”, corresponde à probabilidade de se obter, com outra sequência aleatória de mesmo tamanho e composição de letras, outro alinhamento com escore igual ou superior (MUDADO, 2007; PROSDOCIMI et al., 2002).

É interessante verificar que se utilizássemos um nucleotídio, "A" por exemplo, para pesquisar sequências humanas, a chance de encontrarmos uma região homóloga seria igual a 1 (100%). Se a nossa sequência pesquisada fosse mais complexa, 144 bases, por exemplo, a chance de encontrarmos uma sequência perfeitamente idêntica seria pequena. O valor de "E", expressa essa dificuldade e, assim, quanto menor o valor E, melhor o alinhamento e mais confiável será a consulta, de forma que (num banco de dados de grandes proporções) um valor de “E” igual a zero significa que não há chance de que um alinhamento entre as

duas sequências tenha ocorrido por mero acaso (AMARAL; REIS; SILVA, 2007; PROSDOCIMI et al., 2002).

O *BLAST* apresenta diferentes subprogramas que devem ser utilizados de acordo com o tipo de sequência de entrada e os bancos de dados que se deseja pesquisar. A tabela 3 apresenta as possibilidades de entrada, bancos de dados e programa a ser utilizado (MUDADO, 2007; PROSDOCIMI et al., 2002).

Tabela 3 - Subprogramas BLAST utilizados de acordo com o formato de entrada de sequência e banco de dados desejados.

Formato da sequência de entrada	Banco de dados	Formato da sequência que é comparado	Programa BLAST adequado
Nucleotídeos	Nucleotídeos	Nucleotídeos	BLASTn
Proteínas	Proteínas	Proteínas	BLASTp
Nucleotídeos	Proteínas	Proteínas	BLASTx
Proteínas	Nucleotídeos	Proteínas	TBLASTn
Nucleotídeos	Nucleotídeos	Proteínas	TBLASTtx

Fonte: Adaptação do NCBI < http://www.ncbi.nlm.nih.gov>
Acesso em: 15/12/2015.

2.7.4 Anotação de genomas

As sequências genômicas são fontes ricas de informações sobre a biologia dos organismos, mas devem ser traduzidas através de análises computacionais e de interpretação biológica para que possamos extrair delas a maior quantidade possível de dados úteis (LEWIS et al., 2004). A anotação genômica consiste num processo de vários passos e Stein (2001) divide-a, em três categorias básicas: a anotação de nucleotídeos, de proteínas e de processos.

No mapa físico de sequências, as diferentes regiões devem ser interpretadas com relação à sua função, através de um processo denominado anotação genômica. A similaridade existente entre genes presentes em diversos organismos é utilizada na anotação de função; assim um gene caracterizado numa levedura pode ajudar na identificação funcional do gene com a mesma função no homem, por exemplo. Vários algoritmos distintos foram desenvolvidos para facilitar o processo de anotação nas suas várias etapas (POP; SALZBERG, 2008; STEIN, 2001).

O processo de anotação genômica envolve a atribuição de padrões de funções e identificação de genes na sequência linear do DNA obtida através do sequenciamento e, sua principal tarefa é encontrar os genes. A anotação de nucleotídeos começa com a montagem do genoma, seguido de buscas para encontrar as partes que correspondem aos genes expressos, quais partes correspondem a genes de tRNA, quais correspondem aos clusters de genes de rRNA e assim por diante (PROSDOCIMI et al., 2002).

Para realizar a predição de genes, vários parâmetros podem ser avaliados, tais como, a existência de sequências no DNA que possam funcionar como promotores seguidos por sequências que possam gerar uma proteína funcional, ou que tenham similaridade com genes conhecidos, etc. O conhecimento prévio da proteína e a sua função em qualquer outra espécie facilita bastante o processo de anotação de genes. No entanto, atualmente, grande parte dos genes são ainda hipotéticos, isto é, não se conhece a função biológica destas sequências (LANDER et al., 2001; VENTER et al., 2001).

Na etapa da anotação de proteínas procura-se montar um catálogo das proteínas e genes presentes nos organismos, nomeá-los e associá-los a prováveis funções através, principalmente, de buscas por similaridades. Várias técnicas têm sido desenvolvidas para identificar automaticamente as proteínas pertencentes a diferentes grupos isofuncionais, chamados de grupos de ortologia, entretanto muitas dessas técnicas podem gerar classificações ambíguas. Na prática, o que é normalmente feito é a classificação das proteínas preditas com base em domínios funcionais, configurações espaciais e presença de padrões conservados, além de pesquisa ampla de similaridade contra proteínas bem caracterizadas (AUBOURG; ROUZÉ 2001; MUDADO, 2007).

A parte mais interessante e desafiadora do processo de anotação gênica é relacionar, finalmente, os genes com os processos biológicos. Para isso foi criado um consórcio chamado *Gene Ontology*, que busca criar um vocabulário padrão para descrever a função dos genes eucarióticos através de um sistema de classificação de genes de acordo com suas características. O *Gene Ontology*, popularmente conhecido como *GO*, não é exatamente um banco de dados, é um esforço colaborativo em reunir descrições consistentes de produtos gênicos em diferentes bancos de dados (GENE ONTOLOGY CONSORTIUM, 2000).

Os participantes do projeto GO desenvolveram três estruturas de vocabulário controlado, chamadas de ontologias, que descrevem os produtos gênicos de acordo com a sua associação com processos biológicos, componentes celulares e função molecular de

maneira não relacionada com qualquer organismo em especial. Basicamente, essas três ontologias estão relacionadas às principais perguntas que os pesquisadores têm sobre um determinado gene: (1) em quais processos biológicos o gene está envolvido; (2) qual a sua localização dentro da célula e; (3) como, molecularmente, o gene realiza sua função. Um determinado gene, por exemplo, pode realizar a transdução de sinais em uma célula (processo biológico), estar preso à membrana celular (localização celular) e ter uma função de fosforilar outra proteína (função molecular), sendo classificados no GO de acordo com essas três características (GENE ONTOLOGY CONSORTIUM, 2000; LEWIS, 2004).

A utilização de termos precisos de vocabulário do *GO* ajuda muito na chamada genômica comparativa, pois dessa forma podemos saber se um determinado organismo, por exemplo, apresenta um maior número percentual de genes relacionados a um determinado processo biológico ou não. Da mesma forma, a utilização destas ontologias permite saber quais são todas as proteínas quinases de um determinado organismo, ou quais são todas as proteínas envolvidas com metabolismo de DNA ou, ainda, quais são todas as proteínas que ficam associadas ao retículo (LEWIS, 2004; STEIN, 2001).

Para a anotação de processos é necessário mais do que trabalho computacional e, cada novo experimento adiciona mais informação e permite um melhor entendimento do genoma. Técnicas biológicas em larga escala, como análise de expressão em microarrays e identificação de proteínas por espectrometria de massa, têm sido de fundamental importância para identificar o papel de genes e proteínas nos processos biológicos (MUDADO, 2007; STEIN, 2001).

3 OBJETIVO

3.1 Objetivo Geral

O principal objetivo deste projeto é estudar e descrever o perfil transcriptômico do tegumento de *Podalia orsilochus* (Cramer, 1775), durante o estágio larval.

3.2 Objetivos específicos

- Gerar um banco de dados anotado dos transcritos encontrados no tegumento;
- Analisar as sequências geradas;
- Realizar um levantamento das sequências que apresentam indicações de atuação na imunidade;
- Identificar os peptídeos e proteínas que apresentam atividade antimicrobiana.

4 MATERIAL E MÉTODOS

4.1 Obtenção e preparo do lisado do tegumento da lagarta *Podalia orsilochus*.

O tegumento da lagarta da espécie *Podalia orsilochus* (família Megalopygidae) foi extraído no 6º instar de desenvolvimento, ou seja, na última etapa de desenvolvimento da fase larval. A lagarta utilizada no presente estudo, foi cedida pelo colaborador MSc. Roberto Henrique Pinto Moraes, do Laboratório de Parasitologia e Entomologia do Instituto Butantan, cujo laboratório recebe anualmente vários exemplares de diferentes espécies de lagartas. A identificação da espécie foi realizada pelo pesquisador científico do Instituto Biológico do Estado de São Paulo, Dr. Francisco José Zorzenon.

Uma parcela representativa do tegumento foi extraído com o auxílio de uma pinça de ponta fina e uma tesoura cirúrgica e imediatamente colocado em criotubo contendo RNA*later*® (Tissue Collection: RNA Stabilization Solution - Ambion), uma solução aquosa que permite o armazenamento de amostras de forma estável para a utilização do RNA, sem que o mesmo seja degradado, uma vez que, o RNA é uma molécula sabidamente instável. Os criotubos foram mantidos durante 24 horas a 4° C, após esse período o excesso de RNA*later*® foi removido e os criotubos foram estocados à -80° C até o momento do seu uso. O tegumento anteriormente extraído e mantido no RNA*later*®, foi pesado e alíquotado. O peso do tegumento utilizado foi de 50mg e para essa quantidade de tecido foi utilizado o padrão *standard* que define os volumes dos reagentes do kit utilizado para o isolamento e purificação do mRNA. O tegumento pesado foi triturado com o auxílio de um pistilo, formando um pó do tecido inicial. Foi adicionado 1250µl de tampão de lise (100mM Tris-HCl, pH 7.5; 500mM LiCl; 10mM EDTA, pH 8.0; 1% LiDS e 5mM DTT) ao tecido triturado e homogeneizado por aproximadamente 2 minutos, com a finalidade de se obter a lise completa do tecido. Todo o procedimento foi realizado de acordo com as indicações do manual do fabricante.

4.2 Isolamento e purificação do RNA mensageiro

O isolamento e a purificação do RNA mensageiro (mRNA) foi procedido de acordo com os protocolos do Kit Dynabeads® mRNA DIRECT™ (Invitrogen), que tem como princípio

o pareamento entre os resíduos de bases da cauda poliA da extremidade 3' do mRNA com os resíduos de Oligo (dT)25 que estão covalentemente ligados a superfície das dynabeads. Outras moléculas, como o DNA e proteínas e outras espécies de RNA (tais como RNA de transferência, micro RNAs e RNAs nucleolar pequeno) que eventualmente podem estar presentes na amostra, por não possuírem a cauda poliA, não hibridizam com os Oligo (dTs)25 ligados às beads e são facilmente removidos através de lavagens. Os protocolos do kit estão descritos a seguir.

4.2.1 Preparação de Dynabeads Oligo $(dT)_{25}$

As Dynabeads Oligo $(dT)_{25}$ foram totalmente resuspendidas e transferidas antes do uso para um microtubo RNase-free de 1.5ml, em seguida, foram colocadas em uma estante concentradora de partículas magnéticas. As beads se ligam à parede da estante magnética, separando-as facilmente da solução. O sobrenadante foi removido após 30 segundos e o tubo foi então retirado da estante magnética. As beads foram lavadas com tampão de lise do kit (100mM Tris-HCl, pH 7.5; 500mM LiCl; 10mM EDTA, pH 8; 1% LiDS e 5mM DTT).

4.2.2 Isolamento do RNA mensageiro

O tampão de lise foi removido do microtubo contendo as Dynabeads Oligo $(dT)_{25}$ e as beads foram completamente resuspendidas com a amostra do tegumento lisado. A amostra foi incubada por 5 minutos à temperatura ambiente, permitindo a hibridização da cauda poliA com os Oligo $(dT)_{25}$ que são ligados às beads. O microtubo foi colocado novamente na estante magnética e o sobrenadante foi removido. O complexo beads/mRNA foi submetido a uma série de lavagens, primeiramente utilizando 2ml do tampão de lavagem A (10mM Tris-HCl, pH 7.5; 0.15M LiCl; imM EDTA e 0.1% LiDS), seguido de 1.5ml do tampão de lavagem B (10mM Tris-HCl, pH 7.5; 0.15M LiCl e 1mM EDTA), utilizando sempre a estante magnética para separar as beads das soluções. Ao final das lavagens o mRNA foi eluído com 25µl de tampão de eluição (10mM Tris-HCl, pH 7.5) em temperatura entre 65-80°C por 2 minutos e imediatamente colocados na estante magnética para separar as beads do mRNA. O

sobrenadante contendo o mRNA foi transferido para um novo microtubo RNase-free. As beads que ficaram no tubo anterior foram reservadas.

4.2.3 Eliminação da contaminação com RNA ribossomal

Em muitos casos, a amostra de mRNA contém traços de RNA ribossomal (rRNA), que devem ser removidos para evitar interferências nas análises e interpretações dos resultados obtidos após a construção da biblioteca de cDNA. O rRNA é melhor eliminado realizando uma re-extração de mRNA a partir do eluído final. Esta última etapa de re-extração foi feita reutilizando as mesmas Dynabeads Oligo $(dT)_{25}$ que foram reservadas na etapa anterior.

4.3 Avaliação da quantidade e qualidade do RNA extraído

A quantidade e a qualidade do RNA extraído foram avaliadas utilizando-se o equipamento Bioanalyser com o RNA 6000 Pico LabChip Kit (Agilent Technologies, Palo Alto, CA) que avalia a degradação do RNA total e automaticamente detecta contaminação RNA ribossomal em RNA mensageiro. O Bioanalyzer utiliza um chip para realizar uma eletroforese capilar através da utilização de um corante fluorescente que se liga ao RNA para determinar a concentração e integridade do RNA. A análise eletroforética do chip é baseada nos princípios tradicionais de eletroforese em gel que foram transferidos para um formato de chip, que reduz drasticamente o consumo de amostra e também o tempo de separação. O chip foi preparado de acordo com as instruções e soluções fornecidas pelo kit. Brevemente, o chip foi preenchido com uma mistura de gel e corante, seguido pela adição de solução de condicionamento e do marcador. A amostra de RNA foi adicionada ao poço designado e o chip foi corrido no Bioanalyser.

4.4 Pirosequenciamento

Após purificação, o material foi sequenciado por pirosequenciamento na plataforma 454 GS-Junior (Roche) do Laboratório Especial de Toxinologia Aplicada , parte do CETICS, do Instituto Butantan. Todo o processo foi realizado utilizando-se kits e as instruções do fabricante. Em linhas gerais, o procedimento ocorreu da seguinte forma: O mRNA foi

fragmentado em tamanhos compatíveis com a metodologia (300 a 800 pb), e transformado em fragmentos abruptos fosforilados. Os fragmentos passaram por uma retrotranscrição e foram, então, ligados à adaptadores de sequência de oligonucleotídeos específica. A seguir, os fragmentos de cDNA ligados aos adaptadores, foram incubados e hibridados com microesferas individuais, denominadas beads, que contêm sequências de DNA complementares aos oligonucleotídeos do adaptador. As beads foram, então, isoladamente compartimentadas em microvesículas através de uma emulsificação com óleo, e a amplificação clonal das moléculas únicas de DNA ligadas às beads foi realizada durante a PCR em emulsão. Após a amplificação, a emulsão foi quebrada, e as beads contendo o DNA molde amplificado por clonagem foram enriquecidas. Após uma nova separação por diluição, as beads foram depositadas em uma placa de picotitulação, que possui superfície de aproximadamente 3.4 x 10^6 poços de reação, na escala de picolitros e, diâmetro suficiente para alojar uma única microesfera. A placa foi combinada com os reagentes e as enzimas necessárias para o sequenciamento e acoplada ao equipamento da plataforma 454 GS-Junior, juntamente com os reagentes que percorrem a placa em lavagens sucessivas. Em seguida, o aparelho iniciou a passagem dos fluidos, e a reação do pirosequencimento foi iniciado através da adição sucessiva dos quatro dNTPs (dATP, dTTP, dCTP ou dGTP). Um evento de incorporação de nucleotídeo no molde amplificado por clonagem, produz a liberação de um pirofosfato inorgânico (PPi), que é utilizado pela enzima Sulfurilase ATP para produzir ATP, cuja presença é suficiente para que a enzima Luciferase emita a luminescência que é capturada por uma câmera CCD (charge-coupled device) acoplada ao sistema. A intensidade luminosa do sinal, é detectada e gravada de acordo com o fluxo da reação de cada dNTP, possibilitando a determinação do nucleotídeo que foi incorporado à cadeia de DNA e sua proporção (quantidade) na síntese do ácido nucléico. (HUSE et at., 2007). A atribuição de bases aos cromatogramas gerados é feita automaticamente pelo 454 GS-Junior, através de um programa nomeador de bases próprio.

4.5. Bioinformática

4.5.1 Remoção das sequências de RNA ribossomal

Os dados "brutos" que foram gerados pela máquina de sequenciamento foram importados para o software CLC-Genomics (CLCBio, Aarhus, Dinamarca), onde passaram por uma etapa de pré-processamento, que consiste na remoção de bases consideradas de baixa qualidade, seguidas pela remoção de sequências com similaridade a rRNA. Inicialmente as sequências passaram por um processo de filtragem para eliminar aquelas de baixa qualidade (menor que 0,05) e de pequeno tamanho (menores que 30 pb). Para a remoção das sequências de RNA ribossomal (rRNA), primeiramente criou-se um banco de dados composto de sequências de rRNAs de Lepidoptera, a partir de dados disponíveis no banco de dados GenBank. A seguir, o programa CLC-Genomics foi utilizado para remover as sequências de rRNAs contaminantes dos dados de sequenciamento, através de comparação com o banco de dados de sequências de rRNA de Lepidoptera criado.

4.5.2 Montagem de novo dos contigs

Após as etapas anteriores, a estratégia de montagem das sequências de leitura (*reads*) em sequências contínuas (*contigs*), também foi realizada através do programa CLC Genomics, por ser um programa otimizado para montagem de grandes volumes de dados de alta qualidade de uma maneira rápida e eficiente. O programa tem capacidade de unir os *reads* por afinidade (A-T e C-G), sempre deixando pequenos espaços, que são preenchidos posteriormente com outros *reads* que se alinham, assim os *contigs* vão sendo montados. A montagem *de novo* dos *contigs* (consenso de sequências contínuas) foi realizada utilizando-se as sequências filtradas e os parâmetros padrões de montagem do próprio programa.

4.5.3 Comparação dos contigs *através do BLAST*

Os *contigs* obtidos foram então analisados e comparados com os bancos de dados disponíveis, como nr (não redundante) do *GenBank* (NCBI) utilizando-se as ferramentas

informáticas de busca automatizada através do algoritmo *BLAST* (Basic Local Alignment Search Tool), que realiza a busca de regiões de similaridade entre sequências biológicas já descritas e depositadas em seu banco de dados com os *contigs* montados. A opção de *BLAST* utilizada para essa análise foi o *BLASTx*, com os parâmetros padrões do mesmo conforme Altschul et al. (1997). O *BLASTx* realiza a busca proteínas dentro do banco de dados utilizando como consulta uma sequência de nucleotídeos que são automaticamente traduzidos.

4.5.4 Anotação functional

As ferramentas *BLAST* permitiram a busca de sequências similares ao conjunto de sequências de interesse, através de uma análise de homologia com os bancos de dados públicos. Para a anotação funcional das sequências proteicas obtidas através dos alinhamentos pelo *BLASTx*, foi utilizado o software Blast2GO®, com os parâmetros padrões conforme Conesa e Götz (2008). O software Blast2GO disponibiliza diversas ferramentas desenvolvidas para realizar a anotação funcional de milhares de sequências em uma única sessão e analizar alguns dados dessa anotação através de sua interação com diversos recursos, que fornecem diferentes informações para a análise funcional das sequências de proteínas, como o *GO* (*Gene Ontology Consortium*), mapas *KEGG* (*Kyoto Encyclopedia of Genes and Genomes*), *InterPro* (*Protein Sequence Analysis & Classification*) e *Enzyme nomenclature database*. A anotação funcional automatizada das sequências obtidas foi constituída em três principais passos: busca por analogia (*BLAST*), mapeamento para coletar os termos *GO* associados com os *BLAST hits* e anotação propriamente dita, que atribui a informação às sequências alvos, conforme indicado por Conesa e Götz (2008). Para obtenção das sequências proteicas dos *contigs* foram utilizadas a ferramenta *Translator* do software fr33.net (http://www.fr33.net/translator.php) e a ferramenta do programa Expasy (http://web.expasy.org/translate), ambas disponíveis online. Buscas para informações adicionais sobre sequências proteicas foram realizadas através de análises nos bancos de dados *UniprotKB* (*UniProt Knowledgebase*), do *PDB* (*Research Collaboratory for Structural Bioinformatics – Protein Data Bank*), do *Pfam* (*Protein Family Database*), e para alguns *contigs* foram identificados o domínio proteíco conservado através do banco de dados *CDD* (*Conserved Domains Database*).

4.5.4.1 Classificação dos contigs em categorias funcionais

Como um dos processos de anotação, a partir dos dados gerados foi realizado a divisão e a classificação de todos os *contigs* em categorias funcionais. Essa classificação foi realizada de acordo com as categorias funcionais disponibilizadas pelo banco de dados *COG* (*Clusters of Ortologous Groups* – http://www.ncbi.nlm.nih.gov/COG/) do *NCBI* (TATUSOV et al., 2001). A classificação foi realizada manualmente utilizando, para isso, informações funcionais presentes em arquivos do *NCBI* (WELLER et al., 2002), em informações do *SWISS-PROT*, do *UniProt*, em programas de busca da *Internet* (*Google* – http://www.google.com) (KANEHISA et al., 2002) e em artigos de diversas áreas. As categorias originais presentes nos COGs podem ser observadas na figura 13 e a tabela 6 preenchida com os *contigs* de *Podalia orsilochus*, encontra-se na seção 5.7 Classificação dos *contigs* em categorias funionais.

CATEGORIAS FUNCIONAIS DOS COGS
PROCESSAMENTO E ARMAZENAMENTO DE INFORMAÇÕES
J Tradução, estrutura ribossomal e biogênese
K Transcrição
L Replicação, recombinação e reparo do DNA
PROCESSOS CELULARES
D Divisão celular e particionamento cromossômico
O Modificação pós-traducional, turnover de proteínas e chaperones
M Biogênese do envelope celular, membrana externa
N Mobilidade celular e secreção
P Metabolismo e transporte de íons inorgânicos
T Mecanismos de transdução de sinais
METABOLISMO
C Produção e conversão de energia
G Metabolismo e transporte de carboidratos
E Metabolismo e transporte de aminoácidos
F Metabolismo e transporte de nucleotídeos
H Metabolismo de coenzimas
I Metabolismo de lipídeos
Q Biossíntese, catabolismo e transporte de metabólitos secundários
POBREMENTE CARACTERIZADOS
R Somente predição de função geral
S Função desconhecida

Figura 13 - Categorias funcionais dos COGs.
Legenda: O projeto COG classifica os genes em 4 grupos funcionais principais, sendo cada um deles dividido em subgrupos, representados por letras maiúsculas.
Fonte: COG <http://www.ncbi.nlm.nih.gov/COG/>.
Acesso em: 15/01/2013.

4.5.4.2 Análise dos contigs da subcategoria funcional Imunidade

Os *contigs* que foram categorizados na subcategoria Imunidade foram traduzidos para a obtenção das sequências proteicas e, em seguida, foram analisados através da ferrramenta de busca *BLASTp*, que realiza buscas por analogias entre sequências de aminoácidos. A partir dos resultados dos *BLAST hits*, as sequências foram analisadas em diversos bancos de dados, descritos anteriormente na seção 4.5.1, para obtenção de informações mais apuradas sobre estruturas, famílias, domínios e funções das proteínas, na tentativa de elucidar o papel de cada sequência na resposta de defesa das larvas de *Podalia orsilochus*. Algumas sequências específicas foram selecionadas de acordo com sua função predita e foram analisadas através do banco de dados de peptídeos antimicrobianos *PDB* (*The Antimicrobial Peptide Database*), na tentativa de identificar o perfil de peptídeos antimicrobianos produzidos por esses insetos.

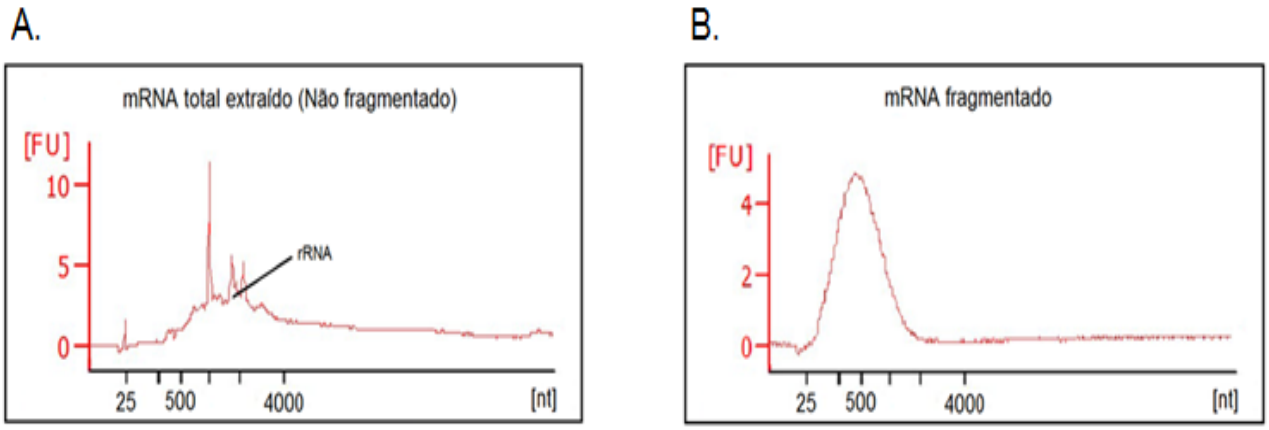
5 RESULTADOS

5.1 Obtenção do RNA mensageiro do tegumento da *Podalia orsilochus*

O tegumento de uma lagarta da espécie *Podalia orsilochus* foi devidamente dissecado e armazenado a -80 °C. Esta amostra de tegumento foi utilizada para a extração do RNA mensageiro e posterior sequenciamento de cDNA. O mRNA foi devidamente isolado e purificado diretamente do lisado proveniente de uma amostra de 50mg do tegumento extraído através do Kit Dynabeads mRNA DIRECT - Invitrogen.

5.1.1 Avaliação da qualidade do mRNA

Anteriormente ao início do pirosequenciamento foi realizada uma avaliação da qualidade do mRNA, através do Bioanalyzer, para verificar possíveis degradações e a presença excessiva de RNA ribossomal (rRNA). Para essa análise, foi utilizado 1µL da amostra. O perfil cromatográfico mostrado na figura 14 (A) indica que a quantidade visualmente perceptível de rRNA (pico por volta de 2000 nt) era tolerável para prosseguir com a experimentação. Após a análise de contaminação da amostra por rRNA, o mRNA foi fragmentado utilizando-se cloreto de zinco. A qualidade dessa amostra foi medida novamente utilizando o Bioanalyzer, e os fragmentos produzidos distribuíram-se conforme o esperado (pico entre 450 a 900 nt) (Figura 14 B). A figura 14 (C) mostra a representação virtual do gel da extração do mRNA, antes e depois da fragmentação.



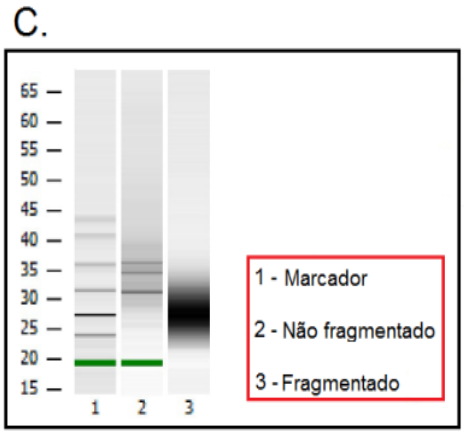


Figura 14 - Resultados das avaliações de tamanho e fragmentação da amostra de mRNA através do Bioanalyzer.
Legenda: (**A**) mRNA extraído do tegumento de *Podalia orsilochus*, apresentando pouca degradação e pouca contaminação por rRNA; (**B**) mRNA após a fragmentação; (**C**) Representação do gel virtual da extração do mRNA, antes e depois da fragmentação.
FU = unidades de fluorescência; nt = nucleotídeos.

5.2 Pirosequenciamento

Foi iniciado o pirosequenciamento e após outras etapas, que incluem reações de retrotranscrição, foi obtido o cDNA, que também foi avaliado através do Bioanalyzer. Novamente, o resultado indicou que a amostra estava dentro de um padrão esperado para a realização do pirosequenciamento, pois de acordo com as orientações, a biblioteca de cDNA deve produzir um pico entre 600 e 1200pb (Figura 15 A e B).

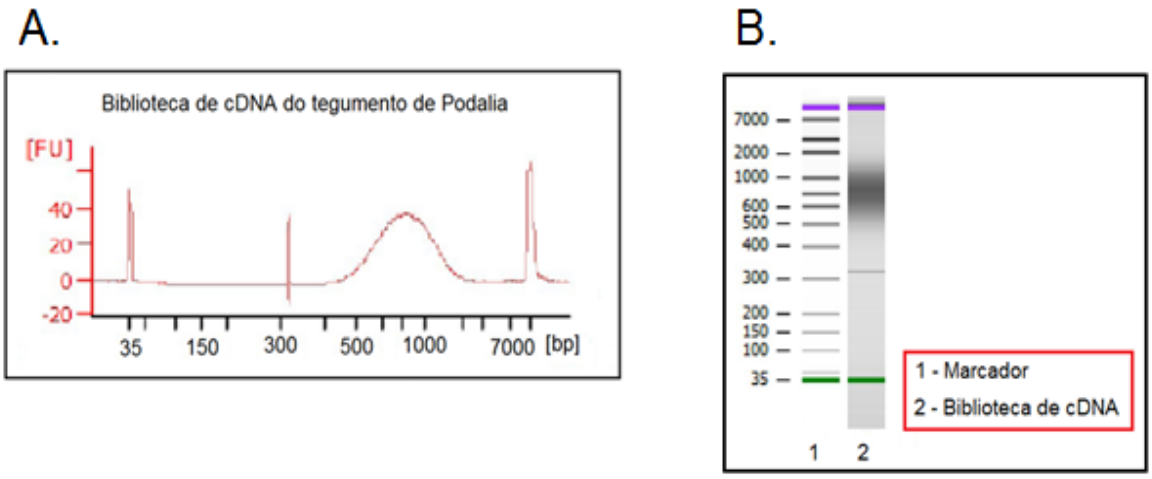


Figura 15 - Resultado da avaliação do cDNA através do Bioanalyzer.
Legenda: (**A**) O gráfico mostra o cDNA produzido a partir do mRNA fragmentado; (**B**) Representação do gel virtual da biblioteca de cDNA.
FU = unidades de fluorescência; bp = pares de bases.

5.2.1 PCR em emulsão, montagem da microplaca e sequenciamento

Após a obtenção do cDNA, foi realizada a PCR em emulsão seguida da montagem da microplaca. A microplaca preparada foi acoplada ao aparelho e pode-se verificar através do resultado obtido que a montagem foi eficiente, visto que a microplaca estava completamente preenchida de poços emitindo sinais luminosos positivos (Figura 16).

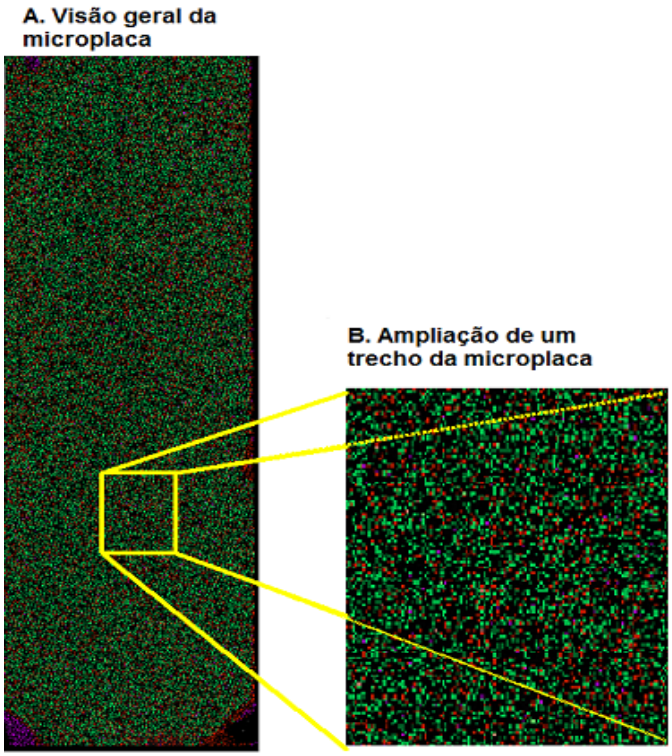


Figura 16 - Imagens captadas pelo 454 GS-Junior, mostrando a microplaca inteira e ampliada.

Legenda: (**A**) Microplaca inteira; (**B**) Microplaca ampliada. Os códigos de classificação de cada micropoço, determinados pela qualidade e tipo da informação obtida em cada micropoço é mostrada em pontos coloridos em (**A**) e (**B**). Pontos verdes = micropoços com sinais válidos; Pontos vermelhos = micropoços com sinais falhos; Pontos roxos = micropoços com verificação inválida; Pontos azuis = micropoços com DNA controle.

Um total de 272.404 micropoços foi identificado como válidos pelo leitor do aparelho (Tabela 4), destes, 263.648 passaram pela verificação de "chave" positiva (esta "chave" é composta por uma pequena sequência nucleotídica que faz parte dos adaptadores usados na construção da biblioteca). Dos micropoços que passaram na verificação de "chave", constatou-se que 42.151 deles possuíam sinal luminoso de baixa qualidade, de maneira que os 188.080 micropoços restantes foram considerados como válidos, e produziram sequências com uma média de tamanho de 425.51 pb e mediana de 471 pb (Figura 17).

Tabela 4 - Dados das contagens de poços obtidos no sequenciamento.

Categoria	Contagem
Micropoços totais identificados	272.404
Micropoços com verificação de “chave” positivos	263.648
Micropoços negativos (sem sequências)	20.338
Micropoços com sequências de baixa qualidade	42.151
Micropoços com mais de uma sequência	13.079
Micropoços com sequências que passaram no critério de qualidade	188.080
Porcentagem do total de sequências aceitas	71.3%

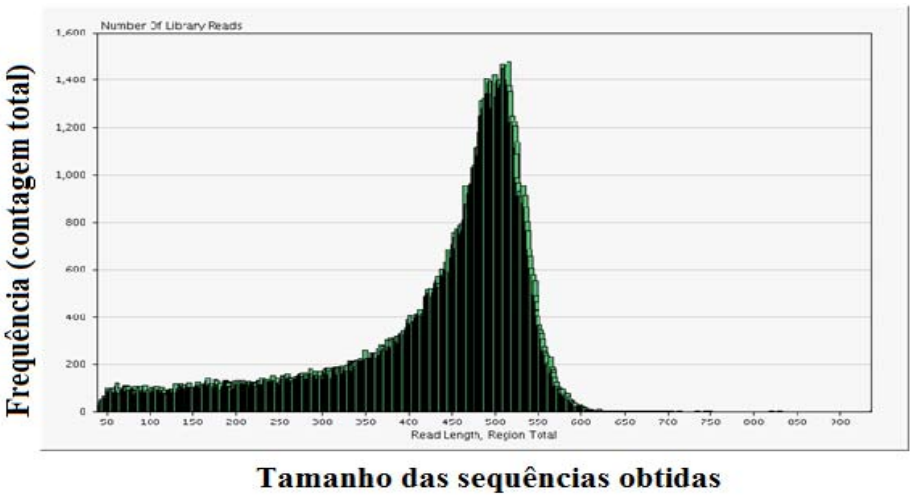


Figura 17 - Imagens captadas pelo 454 GS-Junior, mostrando a microplaca inteira e ampliada.

5.3 Tratamento das sequências e montagem dos *contigs*

As 180.080 sequências (*reads*) válidas foram obtidas a partir do pirosequenciamento do cDNA do tegumento de *P. orsilochus*. Após o sequenciamento, os *reads* foram filtrados através do programa CLC Genomics, para a remoção das sequências que correspondiam a cauda poli-A e rRNA. Esta remoção foi realizada através da comparação das sequências geradas com um banco de sequências ribossomais de Lepidoptera, criado a partir de sequências depositadas no banco de dados do Genbank. Após essa primeira filtragem, foi observado que aproximadamente 0,5% correspondiam a sequências de rRNA e, com isso foram obtidas 179.133 sequências (reads). A remoção das sequências de baixa qualidade e de pequeno tamanho resultou na obtenção de 179.084 sequências finais. Após passar por todas as etapas de tratamento, os *reads* finais foram utilizados para a criação de novas

sequências (*contigs*) que representam o consenso obtido a partir do agrupamento (a clusterização) dos *reads* semelhantes, utilizando-se o programa CLC-Genomics. Em seguida, foi realizada a montagem *de novo* dos *contigs*, que produziu 6.659 *contigs*, utilizando-se 125.805 das sequências iniciais (70,25% dos *reads* obtidos); as 53.279 sequências restantes foram consideradas como *singles* ou *singletons* (Tabela 5). O tamanho das 6.659 sequências *contigs* obtidas variou de 64 a 14474 pb, das quais apenas 7 sequências apresentaram menos de 100pb .

Tabela 5 - Análise das sequências e montagem dos contigs do tegumento de *Podalia orsilochus*.

Classes	**Quantidade**
Reads iniciais (sequências geradas)	180.080
1º triagem* – *reads* removidos	947
Resultado	179.133
2º triagem** – *reads* removidos	49
Resultado	179.084
Montagem	
Reads utilizados para a montagem	125.805
Singletons	53.279
Contigs Finais	6.659

*Remoção da cauda poliA e rRNA;
**Remoção das sequências de baixa qualidade e pequeno tamanho.

5.4 Anotação automatizada: função, ontologia gênica e sublocalização celular

5.4.1 Identificação dos transcritos e dados da anotação

Os 6.659 *contigs* da *P. orsilochus* foram analisados através do programa *Blast2GO* para o processo de anotação automática, para identificar os prováveis genes, as suas funções e possíveis papel biológico, determinando assim a ontologia gênica (GO) dos transcritos, através da análise por similaridade. Todas as sequências nucleotídicas, em formato *FASTA*, foram alinhadas contra o banco de dados *BLAST* para buscar regiões de similaridade entre sequências biológicas depositadas em seu banco de dados utilizando a opção *BLASTx*, implementada no *Blast2GO*.

Das 6.659 sequências analisadas, 4.382 (65%) apresentaram *hits* (similaridade) no *BLASTx*, enquanto que 2.228 (33%) não tiveram qualquer similaridade com sequências do banco de dados não-redundante. Também foram encontradas 49 sequências que não apresentaram resultados no *BLAST*. Das mesmas 6.659 sequências totais, 2.898 (43%) não apresentaram GO *hit* no mapeamento de termos funcionais. A partir desses números, podemos concluir que das 6.659 sequências totais analisadas, um total de 3761 sequências (56%) tiveram uma anotação completa, apresentando *hits* (similaridade) no *BLASTx* e *GO hit* no mapeamento de termos funcionais (Figura 18).

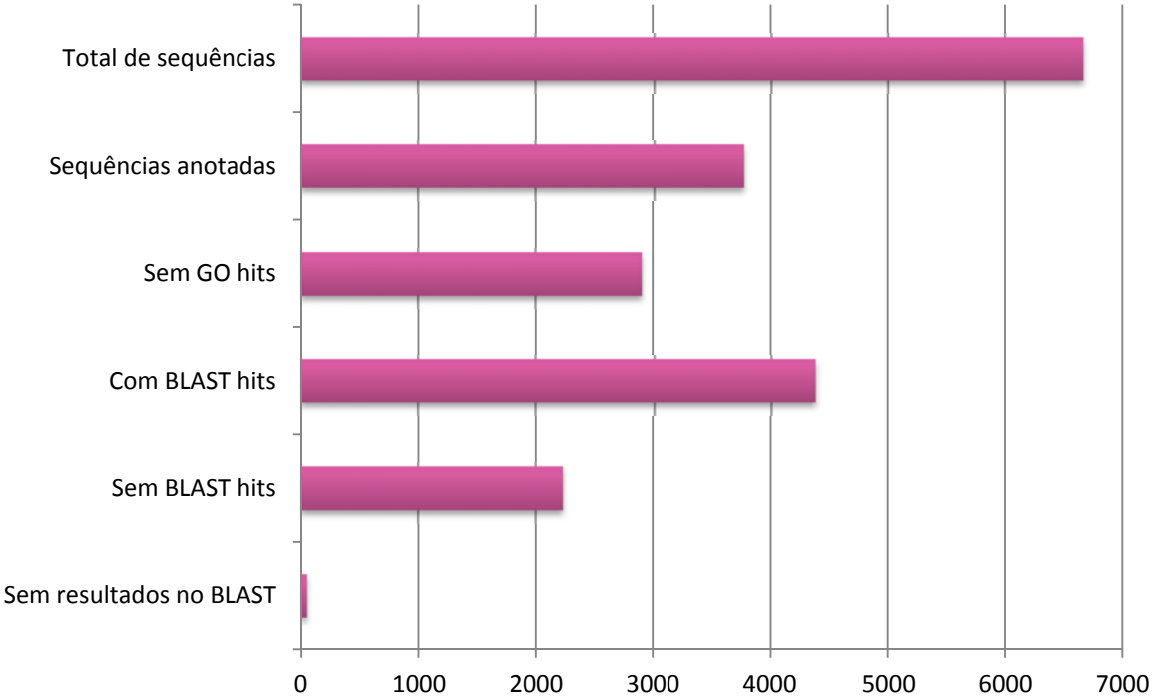


Figura 18 - Gráfico de barras representando o resultado do BLAST (BLASTx hits) e da anotação (GO hits), realizado pelo programa Blast2GO.

A figura 19 mostra o gráfico da distribuição de similaridades das sequências de *P. orsilochus* através do alinhamento das mesmas pelo *BLAST*. O resultado indica que a maior parte dos *contigs* apresentou similaridades entre 60 e 90%, levando em conta o comprimento do alinhamento e os aminoácidos similares entre as sequências analisadas.

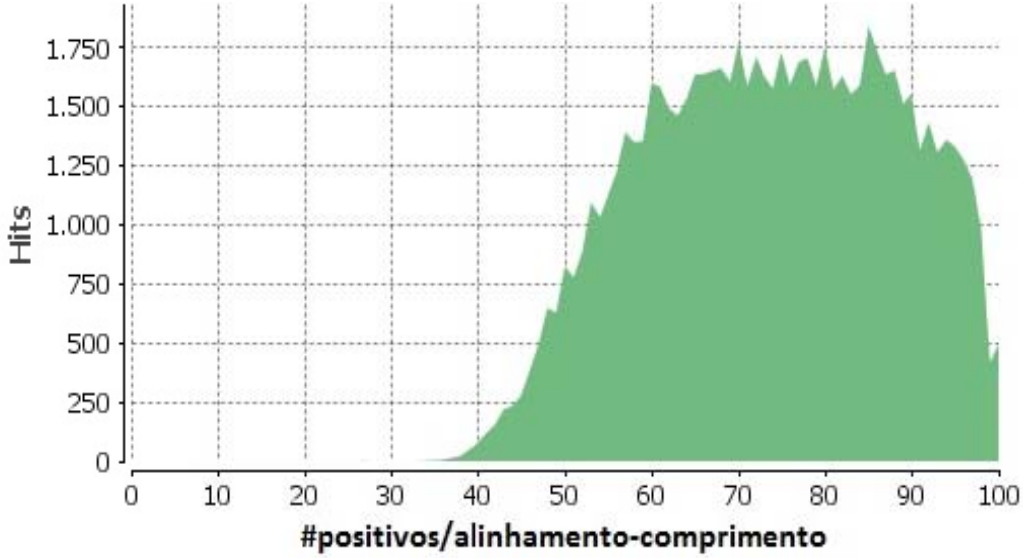


Figura 19 – Gráfico da distribuição de similaridades entre as sequências de *Podalia orsilochus* e as sequências alinhadas através do BLAST.

Durante o processo de anotação, os *contigs* apresentaram similaridade com sequências de diversas espécies. A figura 20 apresenta um gráfico que mostra as 12 espécies que mais vezes tiveram suas sequências comparadas com as sequências da *P. orsilochus*.

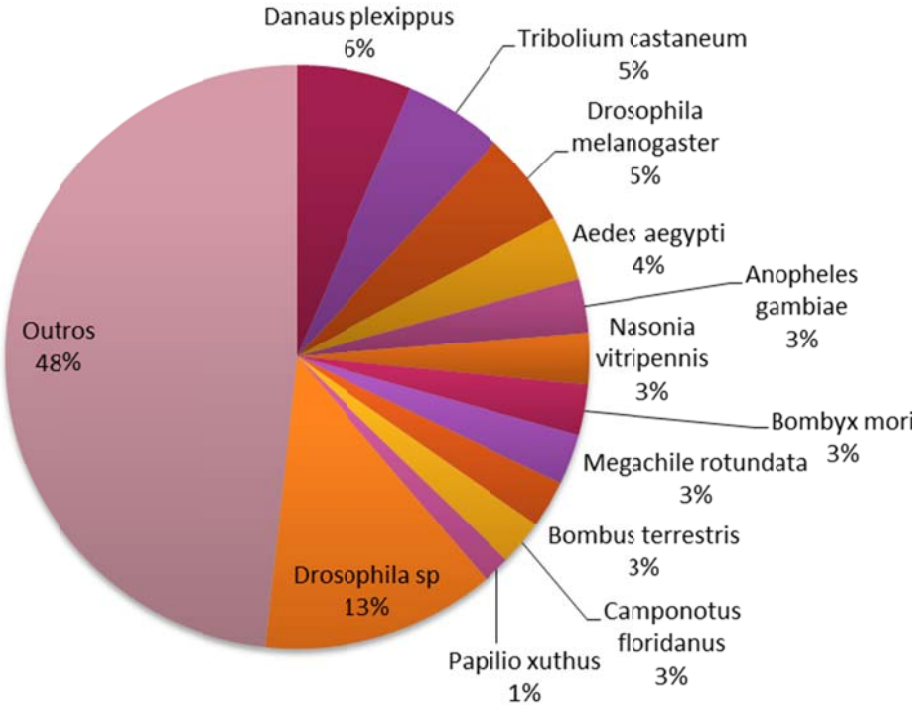


Figura 20 - Gráfico com a distribuição das 12 espécies que mais vezes apresentaram similaridade com as sequências de *Podalia orsilochus*.

Dentre as espécies que mais vezes tiveram suas sequências comparadas com as sequências da *P. orsilochus*, foi possível observar a presença de algumas espécies pertencentes à ordem Lepidoptera, principalmente a conhecida popularmente como borboleta-monarca *Danaus plexippus*, o bicho-da-seda *Bombyx mori* e a borboleta *Papilio xuthus.*

Das cinco espécies numericamente mais representativas durante a comparação de sequências através do BLAST, encontra-se em primeiro, a espécie *D. plexippus*, que obteve mais sequências comparadas com as sequências da *P. orsilochus*, chegando ao número total de 4.863 *hits*; seguida por *Tribolium castaneum*, com 4.159; *Drosophila melanogaster*, com 3.974; *Aedes aegypti*, com 2.742; e, *Anopheles gambiae*, com 2.271.

Outro dado importante do processo de anotação foi a distribuição das espécies com os melhores hits. Para obter esse resultado, o gráfico mostrado na figura 21, leva em consideração apenas a distribuição das espécies que tiveram o melhor alinhamento (menor E-value) dentre todos os possíveis resultados do BLAST para uma determinada sequência consultada. Nesse quesito, todos os primeiros cinco melhores *hits* ocorreram com espécies da ordem Lepidoptera, sendo novamente a espécie *D. plexippus* a de maior número, com 2.563 melhores *hits*, seguidos por *B. mori* com 373, duas espécies do gênero *Papilio*, a *Papilio xuthus* com 330 e *Papilio polytes* com 126 e *Manduca sexta* com 96.

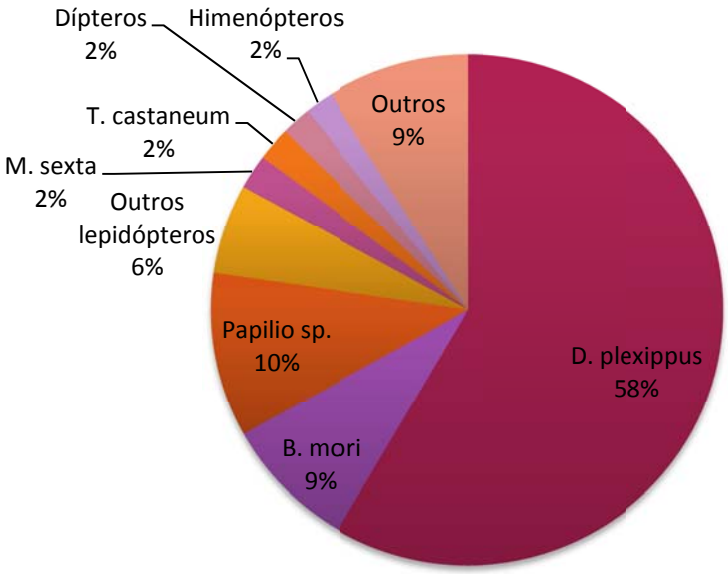


Figura 21 - Gráfico com a distribuição das espécies que apresentaram os melhores hits em comparação com as sequências de *Podalia orsilochus*.

Semelhante ao dado anterior, todos os melhores *hits* (*best hits*) foram obtidos com espécies da classe dos insetos, porém, com predominância de espécies da ordem Lepidoptera.

5.4.2 Ontologia gênica (GO)

O programa Blast2GO foi utilizado para a anotação de processos, através do recurso *Ontologia gênica* ou *GO*, que é uma iniciativa de bioinformática que visa padronizar a representação de genes e os produtos gênicos atribuídos entre as espécies e os bancos de dados, controlando o vocabulário dos termos para descrever as características dos produtos gênicos e criando ferramentas para acessar e processar esses dados. A ontologia abrange três domínios:

I - componente celular - as partes de uma célula ou do seu ambiente extracelular;

II - função molecular - as atividades elementares de um produto do gene, a nível molecular, tais como a ligação ou a catálise;

III - processo biológico - operações ou conjuntos de eventos moleculares com começo e fim definidos, pertinentes para o funcionamento de unidades vivas integradas: células, tecidos, órgãos e organismos.

Para cada *contig* foram obtidos os termos GO com os parâmetros padrões do programa. A partir dos dados gerados por esse recurso, foi possível inferir quantas sequências estão relacionadas com cada um dos domínios GO e quais as principais atividades dos seus produtos gênicos.

Foram encontradas 794 sequências com diferentes combinações dos termos GO para o domínio componente celular. No domínio componente celular, os termos GO mais encontrados foram os termos Membranas e Integral de membrana, que juntos apareceram em 201 sequências, seguidos pelos termos Citoplasma e Citosol, com 108 sequências e os termos Núcleo e Nucléolo, com 61 sequências. Esses termos GO citados e outros termos do domínio componente celular que foram encontrados nas sequências de *P. orsilochus*, e a quantidade de sequências que foram associadas a esses termos estão demonstrados na figura 22.

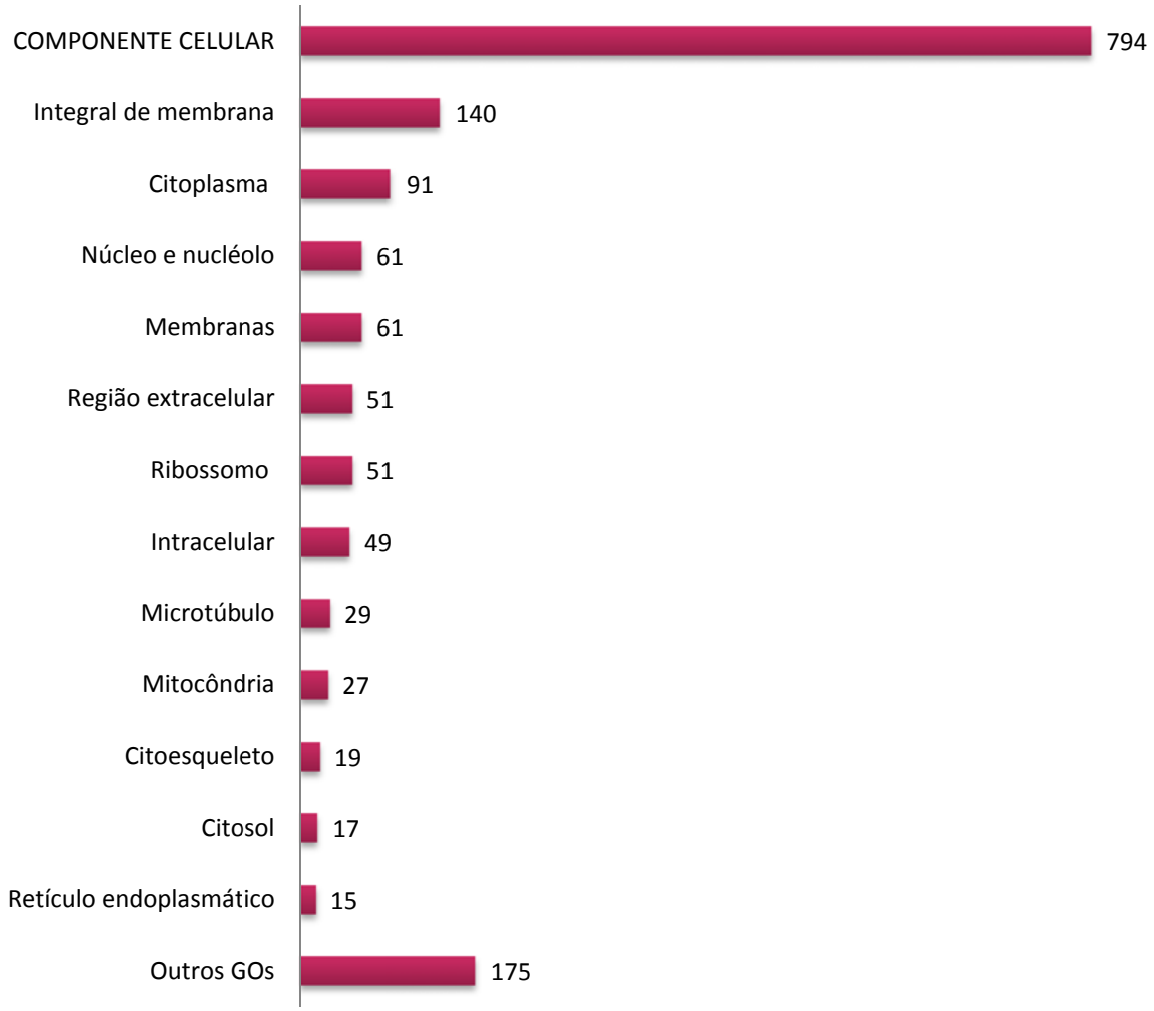


Figura 22 – Termos GO mais encontrados do domínio componente celular e a quantidade de sequências de Podalia orsilochus associadas a esses termos.

A figura 23 representa um gráfico combinado gerado pelo programa Blast2GO, das prováveis localizações celulares das diversas proteínas codificadas pelas sequências dos *contigs* da *P. orsilochus* que foram associadas com o domínio GO componentes celulares. Existem variações nos números das sequências observadas para cada termo GO, uma vez que, por exemplo, uma sequência que está associada com o termo GO Organela, também pode estar associada com os termos GO Parte celular ou membrana.

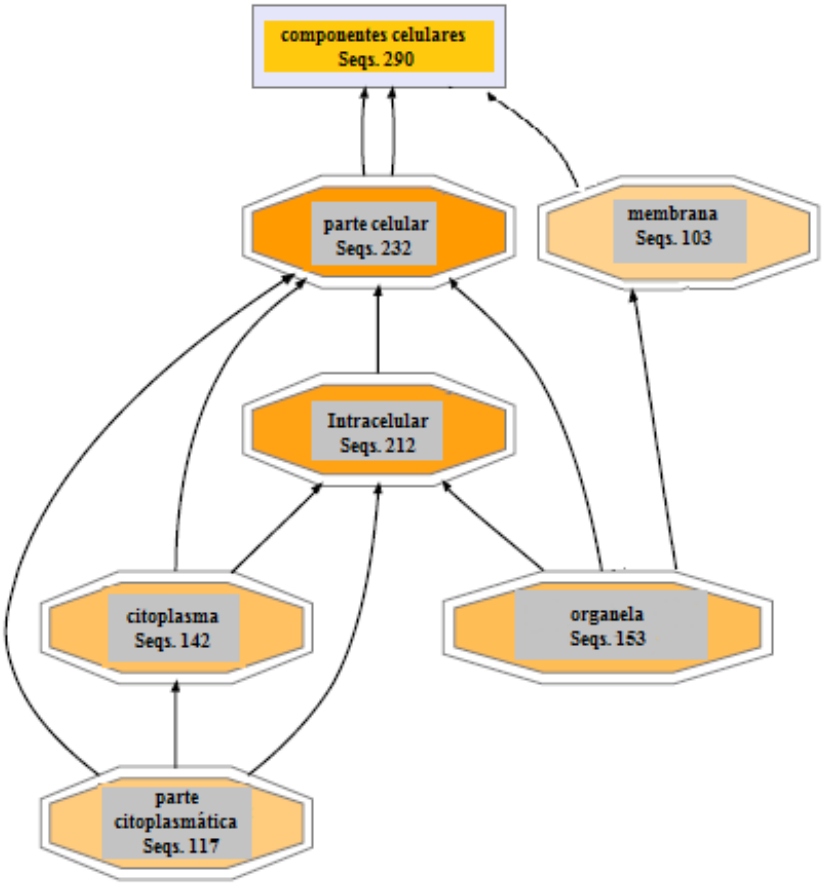


Figura 23 – Termos GO mais encontrados do domínio componente celular e a quantidade de sequências de Podalia orsilochus associadas a esses termos.

Das 6.659 sequências, 2.517 foram associadas ao domínio GO de função molecular. Através das análises realizadas pôde-se inferir que as proteínas com função de ligação (*binding proteins*) foram as numericamente mais encontradas, totalizando 1.686 sequências. Nessa categoria, as sequências mais abundantes foram descritas como ligantes de proteínas, de ácidos nucléicos, de cálcio e de pequenas moléculas. A segunda categoria de função mais encontrada foi a de atividade catalítica, com 1.471 sequências relacionadas e a maior abundância eram de hidrolases, transferases, oxidoredutases e ligases. Também foram encontradas 187 sequências com atividade de transporte, onde a maioria foi descrita como sendo de transporte transmembrana e/ou transporte transmembrana substrato-específico. As funções descritas menos encontradas no total das sequências foram as de atividade de molécula estrutural, com 151, as de atividade de receptor, com 113 sequências e as de atividade de enzima regulatória, com apenas 105 (Figura 24).

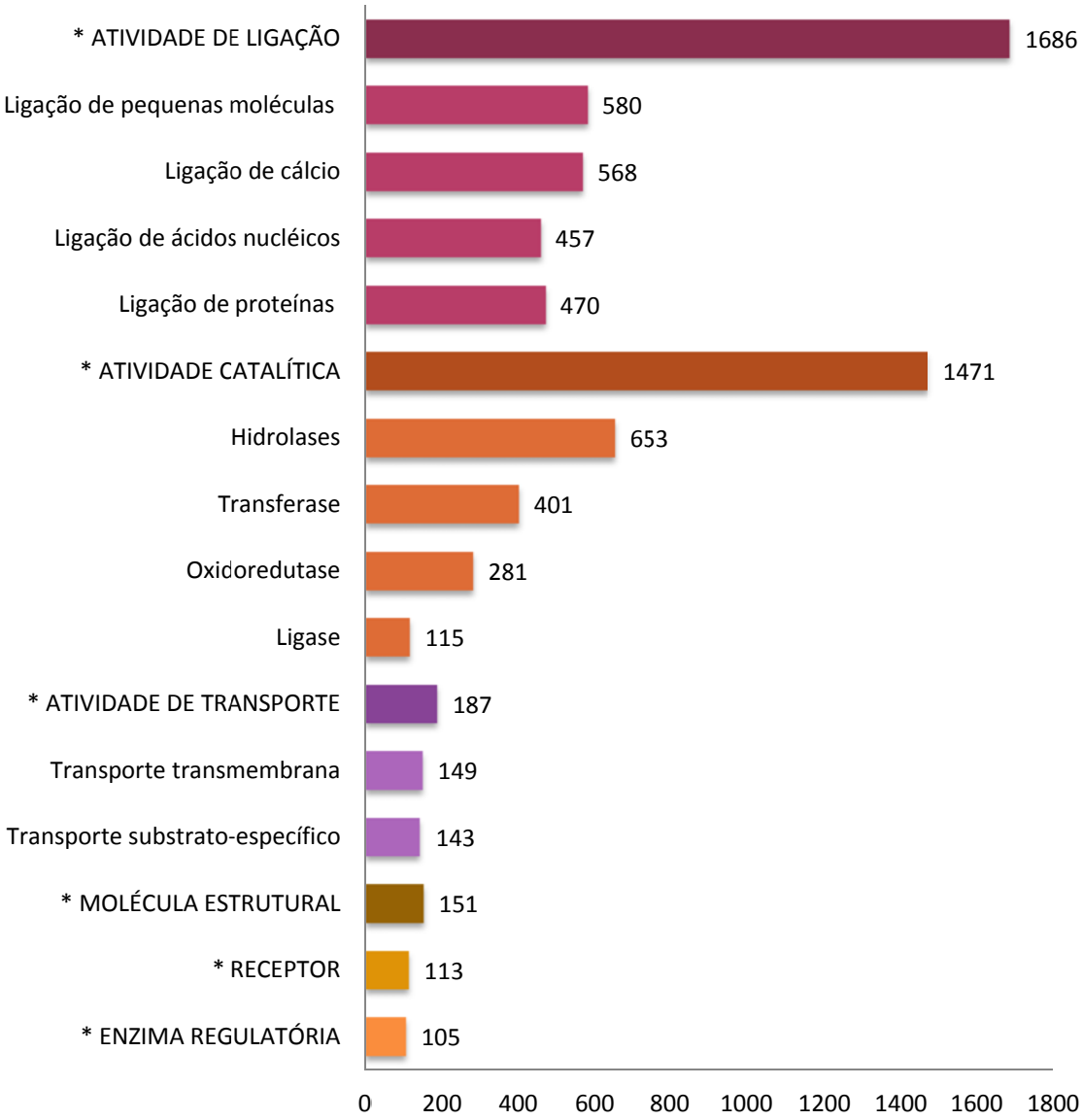


Figura 24 - Gráfico de distribuição das sequências associadas aos termos GO de funções moleculares.
Legenda: O gráfico mostra o número de sequências encontradas em cada categoria funcional principal e o número de sequências encontradas em cada categoria de atividade específica.

Para a última análise de ontologia gênica, um total de 1.204 sequências foram associadas com os termos GO do domínio Processos biológicos, não havendo termos com superioridade muito relevante. O processo biológico de proteólise foi o mais encontrado, apresentando 50 sequências associadas com esse termo, Seguidos pelo processo metabólico e processo de oxidação-redução, ambos com 32 sequências. O quarto termo GO de processo biológico mais encontrado foi o de *folding* de proteína, com 31 sequências. Outros termos GO encontrados nesse domínio se encontram na figura 25.

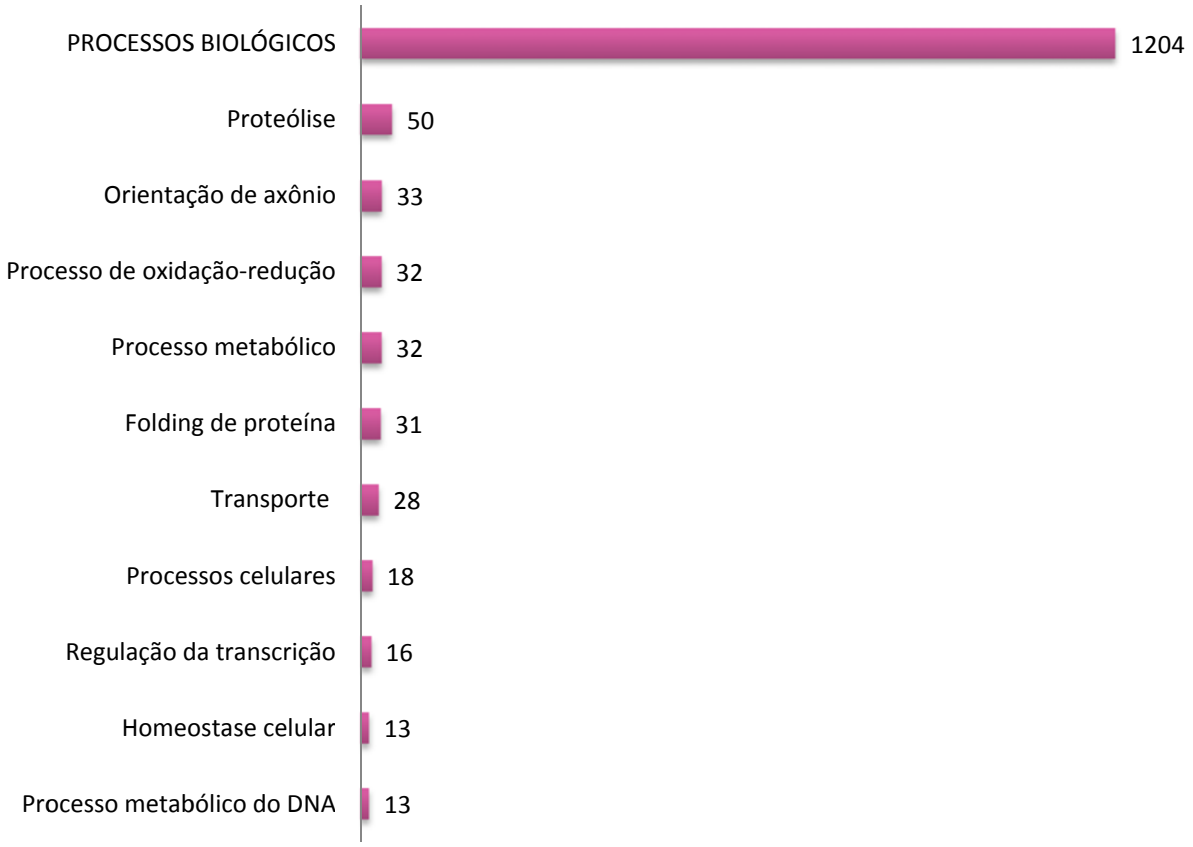


Figura 25 - Gráfico de distribuição das sequências associadas aos termos GO do domínio Processos biológicos.

5.5 Predição da presença de peptídeo sinal e de segmentos transmembrânicos

O programa *Blast2GO* possui a opção de interação com um consórcio de banco de dados denominado *InterProScan*, que integra as informações gerais de predição de função das proteínas que são utilizadas para classificar as proteínas em famílias e prever a presença de domínios e sítios importantes. Para essa classificação, são utilizados modelos preditivos, conhecidos como assinaturas, fornecidos por vários bancos de dados diferentes (chamados de bancos de dados membros) que compõem o consórcio *InterPro*. Nesse contexto, esse recurso foi utilizado para realizar as predições da presença de peptídeo sinal de secreção de proteína e da presença de sequências características de proteína transmembrana. Das 6.659 sequências analizadas, apenas 1333 tiveram resultados positivos apresentados pelo *InterPro* e, 5326 não apresentaram nenhum resultado (Figura 26).

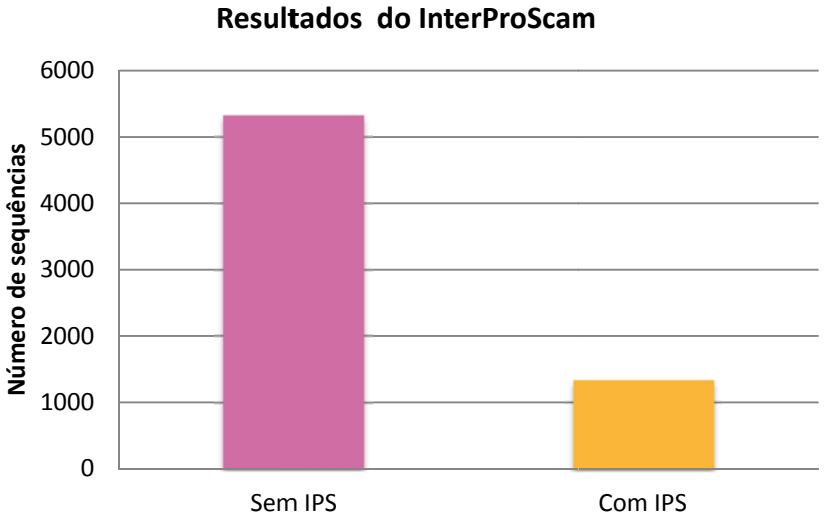


Figura 26 - Gráfico demonstrando a porcentagem de sequências que apresentaram predições da presença de peptídeo sinal de secreção de proteína e da presença de sequências características de proteína transmembrana através do recurso InterPro.

A porcentagem de predição para peptídeo sinal que caracteriza proteína secretada, encontrado no contigs de *P. orsilochus* foi de 22% (1.465 sequências) e de não secretada foi de 78% (5.194 sequências) (Figura 27).

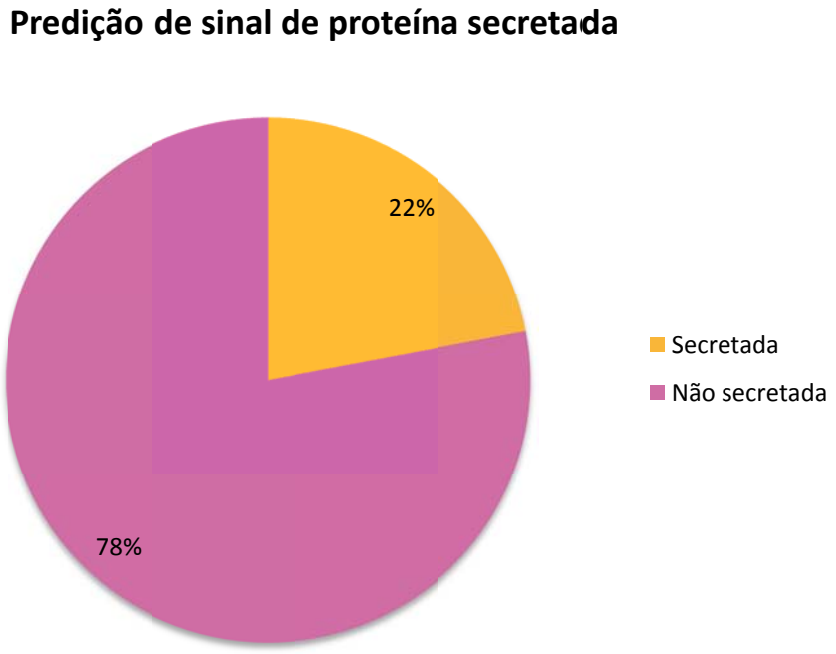


Figura 27 - Gráfico dos dados da porcentagem de predição de sinal de proteína secretada entre os 6.659 contigs analisados.

O cálculo de predição para proteína transmembrânica nos contigs de *P. orsilochus* foi de 16 % (1.065 sequências) e de não transmembrânica foi de 84 % (5.594 sequências) (Figura 28).

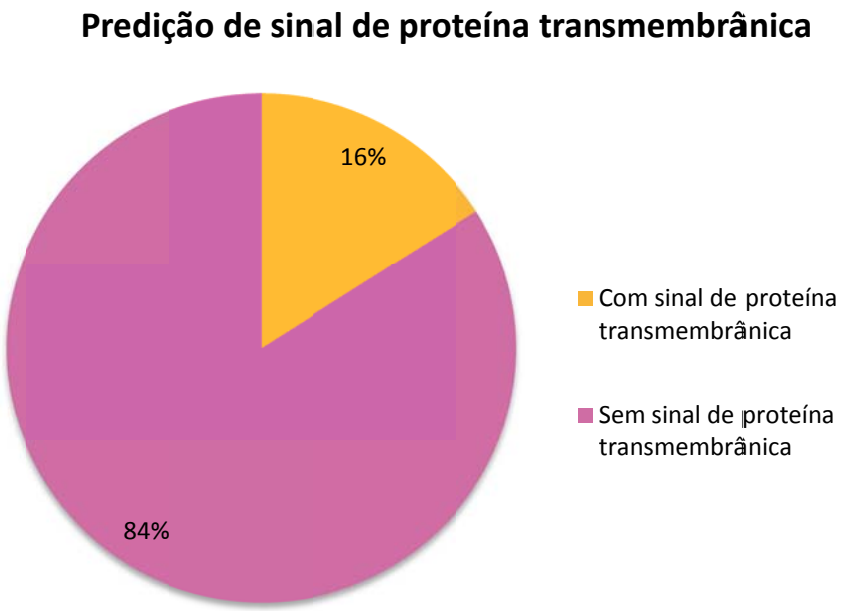


Figura 28 - Gráfico dos dados de porcentagem de predição de sinal de proteína transmembrânica entre os 6.659 contigs analisados.

5.6 Classificação enzimática

As sequências de *P. orsilochus* foram analisadas através da indicação dos termos de classificação enzimática, que utiliza códigos (EC codes) para identificar as principais classes de enzimas já descritas. Foram observadas diferentes quantidades de sequências relacionadas a membros de todas as classes de enzimas, apresentando um total de 859 sequências com predição de atividade enzimática.

Foram encontradas 163 sequências classificadas na classe das oxidoredutases (EC 1), que representa as enzimas que catalisam reações de oxidação-redução. As enzimas da classe das transferases (EC 2) foram encontradas em 251 sequências. Contudo, as enzimas da classe das hidrolases (EC 3), foram as mais representadas dentro das sequências da P. orsilochus, apresentando um total de 303 sequências. Em quantidades menores, foram encontradas enzimas das classes liases (EC 4), isomerases (EC 5) e ligases (EC 6) representadas por 30, 24 e 88 sequências respectivamente. O gráfico da figura 29 apresenta

a porcentagem de sequências da *Podalia orsilochus* em cada classe enzimática correspondente.

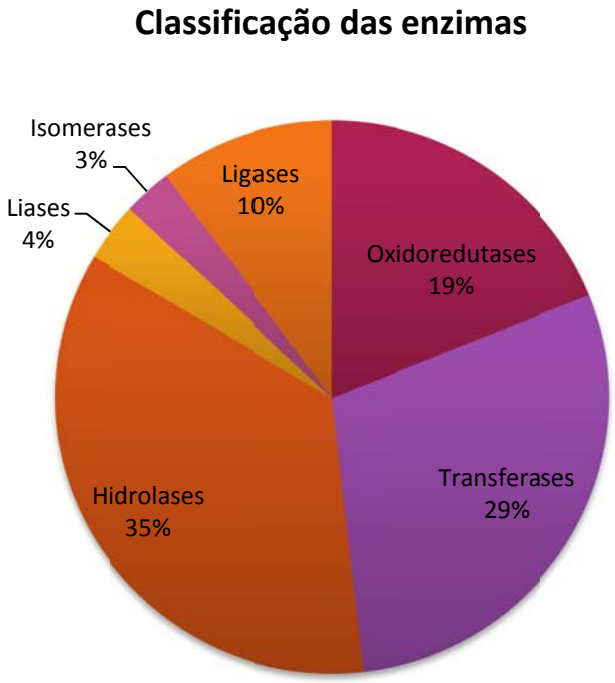


Figura 29 – Termos gerais dos códigos de classificação das enzimas (EC).
Legenda: (EC:1.x.x) oxidoredutases; (EC:2.x.x) transferases; (EC: 3.x.x) hidrolases; (EC:4.x.x) liases; (EC:5.x.x) isomerases; e (EC:6.x.x) ligases.

5.7 Classificação dos *contigs* em categorias funcionais

Após a geração de todos os dados dos processos de anotação, foi elaborada uma tabela que serviu de base para a classificação de cada *contig* em sua respectiva categoria funcional. Essa classificação foi realizada manualmente e individualmente, analizando todos os dados gerados para cada *contig* e comparando as informações obtidas com as informações disponíveis nos bancos de dados. Para essa categorização, o banco de dados mais utilizado foi o *UniProt* (*Resource Protein Universal*), que representa uma fonte abrangente para a pesquisa de sequências de proteína e dados de anotação. As bases de dados *UniProt* são, *UniProt Knowledgebase* (*UniProtKB*), o *UniProt Reference Clusters* (*UniRef*), e o *UniProt Archive* (*UniParc*).

O projeto *COG* classifica os genes em quatro grupos funcionais principais, sendo cada um deles dividido em subgrupos. Da tabela original, foi acrescido um grupo funcional, representado pela categoria “Morte celular e resposta imune”, que compreende duas

subcategorias: “Apoptose e morte celular” e “Imunidade”. Além da criação de duas subcategorias “Matriz extracelular e adesão” e “Molécula estrutural e de estoque”, que foram adicionadas a categoria “Processos celulares”. Essas quatro subcategorias foram criadas e incluídas na tabela original por se tratar de funções que não se enquadram em nenhuma outra subcategoria disponibilizada pelo *COG*. A tabela de categorias funcionais dos *COGs* foi utilizada para a classificação dos *contigs* e foi preenchida com a quantidade total de *contigs* enquadrados em cada categoria e subcategoria (Tabela 6). Por convenção da tabela original foi mantida a formatação de subcategorias representadas por letras.

Tabela 6 – Classificação dos contigs em categorias funcionais.

CATEGORIAS FUNCIONAIS

- **PROCESSAMENTO E ARMAZENAMENTO DE INFORMAÇÕES = 796 *contigs***
  J Tradução, estrutura ribossomal e biogênese **= 306 *contigs***
  K Transcrição **= 388 *contigs***
  L Replicação, recombinação e reparo do DNA **= 102 *contigs***

- **PROCESSOS CELULARES = 1955 *contigs***
  D Divisão celular e particionamento cromossômico = **22 *contigs***
  O Modificação pós-traducional, turnover de proteínas e chaperones **= 499 *contigs***
  M Biogênese do envelope celular, membrana externa **= 71 *contigs***
  N Mobilidade celular e secreção **= 154 *contigs***
  P Metabolismo e transporte de íons inorgânicos **= 324 *contigs***
  T Mecanismos de transdução de sinais **=748 *contigs***
  W Matriz extracelular e adesão* **= 54 *contigs***
  X Molécula estrutural e de estoque* **= 83 *contigs***

- **METABOLISMO = 951 *contigs***
  C Produção e conversão de energia **= 318 *contigs***
  G Metabolismo e transporte de carboidratos **= 53 *contigs***
  E Metabolismo e transporte de aminoácidos **= 24 *contigs***
  F Metabolismo e transporte de nucleotídeos **= 13 *contigs***
  I Metabolismo de lipídeos **= 66 *contigs***
  Q Biossíntese, catabolismo e transporte de metabólitos secundários **= 477 *contigs***

- **POBREMENTE CARACTERIZADOS = 2803 *contigs***
  R Somente predição de função geral **= 95 *contigs***
  S Função desconhecida = **2708 *contigs***

- **MORTE CELULAR E RESPOSTA IMUNE* = 154 *contigs***
  Y Apoptose e morte celular* **= 44 *contigs***
  Z Imunidade* **= 110 *contigs***

* Categorias novas não encontradas na tabela de categorias funcionais do COG original.

Após a classificação de cada *contig* na sua respectiva categoria principal, foi possível observar que 43% de todos os *contigs* obtidos do tegumento de *P. orsilochus*, correspondem a genes que são pobremente caracterizados. A segunda categoria com maior representação, foi a categoria de Processos celulares, com 29% de todos os *contigs* analizados, seguido por Metabolismo, com 14 %, Processamento e armazenamento de informações, com 12 % e 2 % corresponde a nova categoria Morte celular e resposta imune (Figura 30).

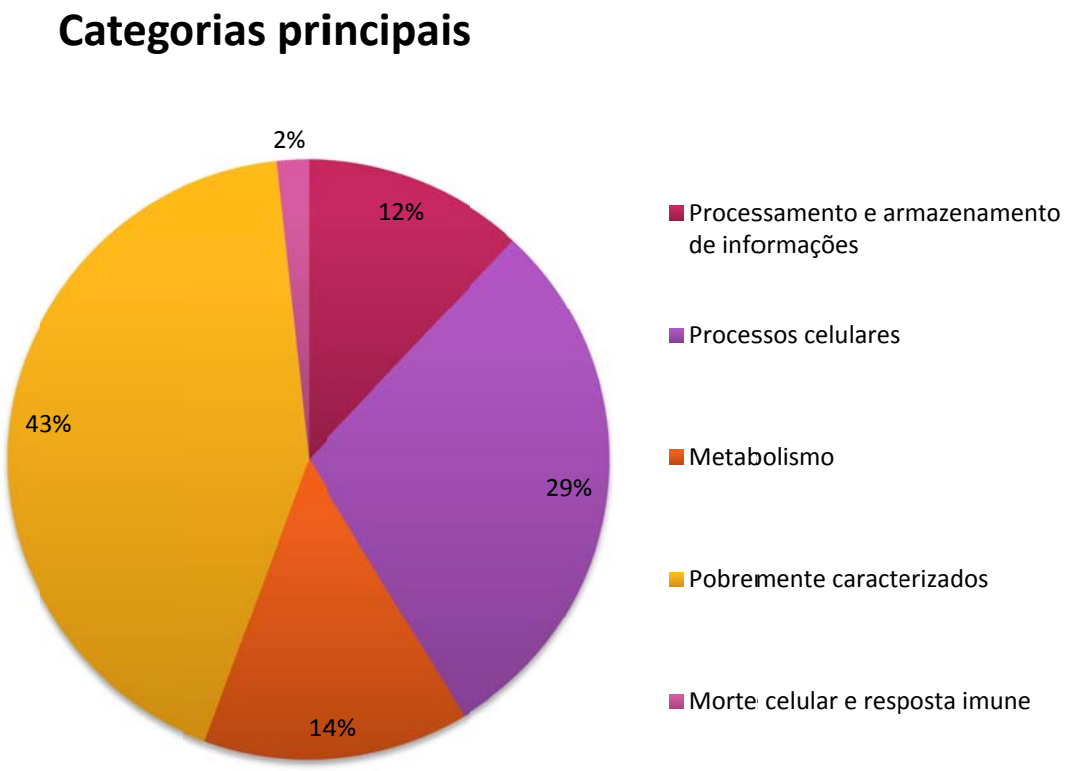


Figura 30 – Gráfico com a porcentagem de distribuição de contigs correspondente a cada categoria funcional.

A figura 31 mostra a porcentagem de *contigs* inclusos em cada subcategoria dentro de sua respectiva categoria principal. As letras representadas dentro de cada gráfico correspondem a letra designada para cada subcategoria de acordo com a tabela de categorias funcionais dos COGs apresentada na tabela 6.

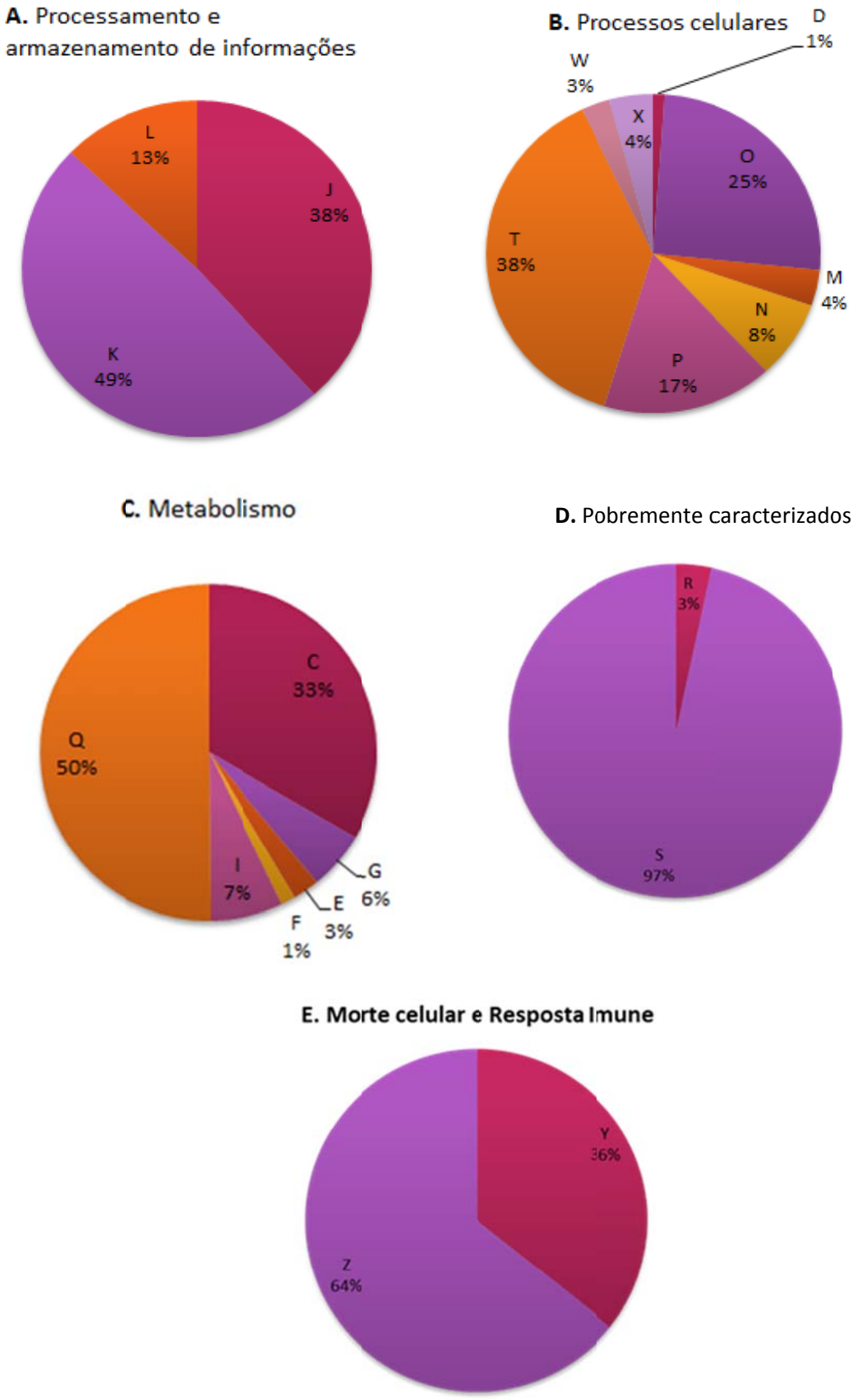


Figura 31 – Gráficos das porcentagens de contigs incluídos em cada subcategoria dentro de cada categoria geral.

Legenda: (**A**) Porcentagem das subcategorias de Processamento e armazenamento de informações: **J** Tradução, estrutura ribossomal e biogênese; **K** Transcrição; **L** Replicação, recombinação e reparo do DNA; (**B**) Porcentagem das subcategorias de Processos celulares; **C.** Porcentagem das subcategorias de Metabolismo; **D.** Porcentagem das subcategorias de Pobremente caracterizados e; **E.** Porcentagem das subcategorias de Morte celular e resposta imune. **D** Divisão celular e particionamento cromossômico; **O** Modificação pós-traducional, turnover de proteínas e chaperones; **M** Biogênese do envelope celular, membrana externa; **N** Mobilidade celular e secreção; **P** Metabolismo e transporte de íons inorgânicos; **T** Mecanismos de transdução de sinais; **W** Matriz extracelular e adesão; **X** Molécula estrutural e de estoque; **C** Produção e conversão de energia; **G** Metabolismo e transporte de carboidratos; **E** Metabolismo e transporte de aminoácidos; **F** Metabolismo e transporte de nucleotídeos; **I** Metabolismo de lipídeos; **Q** Biossíntese, catabolismo e transporte de metabólitos secundários; **R** Somente predição de função geral; **S** Função desconhecida; **Y** Apoptose e morte celular; **Z** Imunidade.

5.7.1 Análise contigs *com maior número de reads.*

Os *contigs* foram renomeados com a designação POT (uma sigla para Tegumento de *Podalia orsilochus*), seguida por uma identificação númerica baseada na quantidade de *reads* atribuído a cada *contig*. Este procedimento foi adotado para facilitar a análise, identificação e descrição dos transcritos. A Tabela 7 apresenta os 50 *contigs* de maior expressão.

Tabela 7 - Transcritos do tegumento de Podalia orsilochus que possuem altos níveis de expressão gênica (maior frequências em número de reads*), totalizando os contigs 50 mais expressos.

***CONTIGS* POTG**	**DESCRIÇÃO DO *CONTIG***	**CATEGORIA FUNCIONAL E SUBCATEGORIA (LETRA)**	**Nº DE *READS***	**TAMANHO (pb)**	**E-value**	**SIMILARIDADE**
POT0001	Heat shock protein 1	Metabolismo - Q	1914	1257	9,59E-43	47.05%
POT0002	Muscular protein 20	Processos celulares - N	1390	1588	6,38E-57	70.30%
POT0003	Long form-like	Processos celulares - N	1255	3257	0.0	85.20%
POT0004	Elongation factor 2	Processos celulares - O	1216	4887	0.0	93.90%
POT0005	Desconhecido	Pobremente caracterizado - S	1089	14474	_	_
POT0006	Actin	Processos celulares - N	1037	1433	0.0	79.00%
POT0007	Adenylate cyclase type 2	Processos celulares - T	1019	2247	1,17E-26	42.65%
POT0008	Elongation factor 1-alpha-like	Processos celulares - O	924	1667	0.0	77.50%
POT0009	Lethal essential for life l2efl	Metabolismo - Q	917	1492	1,58E-47	61.85%
POT0010	Desconhecido	Pobremente caracterizado - S	898	10116	-	-
POT0011	Actin	Processos celulares - N	895	1572	0.0	83.95%
POT0012	Myosin heavy muscle isoform 1	Processos celulares - X	811	6246	0.0	88.70%
POT0013	Desconhecido	Pobremente caracterizado - S	804	9565	-	-
POT0014	Desconhecido	Pobremente caracterizado - S	766	9252	-	-
POT0015	Myosin heavy nonmuscle or smooth muscle	Processos celulares - X	706	4844	0.0	87.80%
POT0016	Desconhecido	Pobremente caracterizado - S	690	5064	-	-
POT0017	Elongation factor 1-alpha-like	Processos celulares - O	659	1736	0.0	87.15%
POT0018	Isoform a	Processos celulares – W	607	1715	0.0	67.45%
POT0019	Desconhecido	Pobremente caracterizado - S	603	12995	-	-
POT0020	Cg41536 cg41536- partial	Pobremente caracterizado - R	586	3555	5,25E-70	65.75%
POT0021	Polyadenylate-binding protein 1-like isoform 1	Processamento e armazenamento de informações – K	580	2546	0.0	83.75%
POT0022	Serpin 1	Metabolismo – Q	566	1308	3,24E-83	66.00%
POT0023	Hypothetical protein kgm_05776	Pobremente caracterizado - R	543	2745	2,45E-95	78.00%
POT0024	Desconhecido	Pobremente caracterizado - S	529	4749	-	-

POT0025	Troponin c	Processos celulares – P	524	782	3,02E-64	78.70%
POT0026	Muscle-specific protein 300	Processos celulares – X	522	7432	0.0	57.85%
POT0027	Ribosomal protein l23a	Processamento e armazenamento de informações – J	517	1113	6,15E-89	92.15%
POT0028	Cytochrome p450	Metabolismo – C	512	2072	0.0	87.80%
POT0029	Tropomyosin isoform b	Processos celulares – X	506	1034	5,59E-14	83.30%
POT0030	Desconhecido	Pobremente caracterizado - S	491	9670	-	-
POT0031	Cytochrome p450	Metabolismo – C	478	1997	0.0	80.45%
POT0032	Tpa: cuticle protein	Processos celulares – X	464	681	9,25E-13	52.95%
POT0033	Myosin heavy muscle-like isoform 2	Processos celulares – X	459	3782	0.0	91.25%
POT0034	Retinol dehydrogenase 14	Metabolismo – C	452	3720	2,17E-74	79.05%
POT0035	Catalase	Metabolismo – C	451	1676	0.0	80.50%
POT0036	Muscle-specific protein isoform d	Processos celulares – X	413	7401	0.0	72.50%
POT0037	Titin-like protein	Metabolismo – I	411	3285	2,88E-75	63.20%
POT0038	Adenylate cyclase	Processos celulares – T	402	1917	1,05E-15	42.68%
POT0039	Titin-like protein	Processos celulares – P	400	3659	4,95E-43	56.90%
POT0040	Muscle protein 20-like protein	Processos celulares – X	399	2407	0.0	75.65%
POT0041	Vitellogenin	Processos celulares – X	396	5614	0.0	64.15%
POT0042	Serine protease	Metabolismo – Q	377	1008	5,75E-35	45.55%
POT0043	Cg41536 cg41536- partial	Pobremente caracterizado - R	373	2969	1,43E-77	76.20%
POT0044	Titin-like protein	Processos celulares – X	371	3387	1,61E-13	58.25%
POT0045	Translation elongation factor 2	Processamento e armazenamento de informações – K	367	2836	0.0	93.15%
POT0046	Serine protease	Metabolismo – Q	365	1105	6,85E-24	45.85%
POT0047	Fatty acid synthase	Metabolismo – I	364	7916	0.0	66.10%
POT0048	Hypothetical protein kgm_00217	Pobremente caracterizado - R	350	3828	6,91E-94	53.33%
POT0049	Juvenile hormone binding protein	Pobremente caracterizado - R	341	913	7,53E-77	52.95%
POT0050	Adp atp translocase	Processos celulares – P	336	1265	0.0	89.55%

*Dados obtidos através do processo de anotação pelo programa *Blast2GO*.

5.8 Análise dos contigs da subcategoria Imunidade

Dentre as categorias funcionais apresentadas, o foco foi concentrado na subcategoria Z – Imunidade para realizar uma análise mais aprofundada, com a intenção de disponibilizar novas informações sobre o perfil transcriptômico de genes ligados à resposta imune dos insetos e viabilizar a pesquisa de novos peptídeos ativos, que possam ter atividades biológicas de interesse farmacológico e biotecnológico. Foram incluídos 110 *contigs* na subcategoria Imunidade e foi atribuída a provável função imunológica para cada *contig* de *P. orsilochus* baseado na similaridade de sua sequência proteica com genes previamente caracterizados com função na resposta imune e em seus padrões de reconhecimento. O gráfico apresentado na figura 32 representa a abundância em número de *reads* dos *contigs* dos genes imunes encontrados em *P. orsilochus* que apresentaram mais de 50 *reads*.

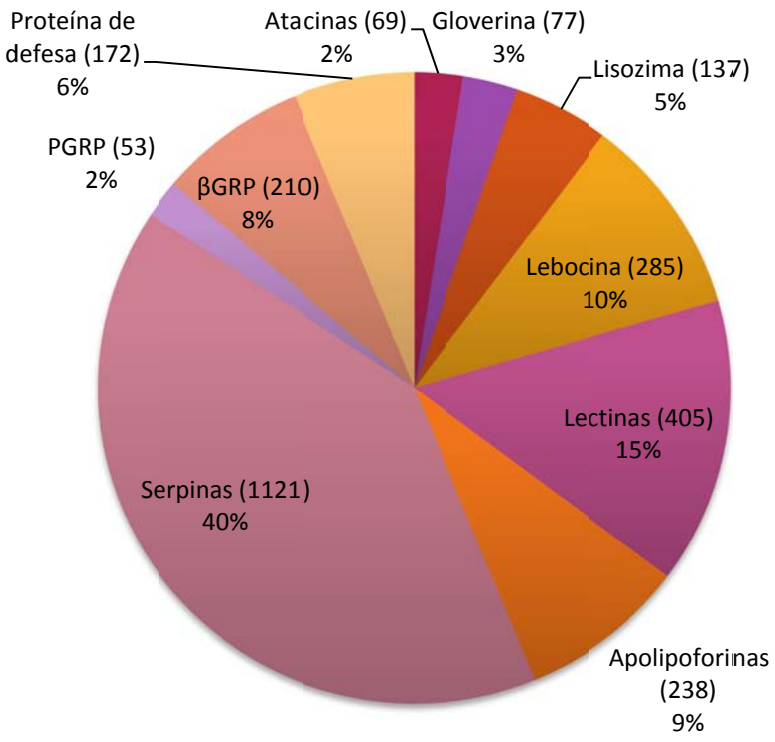


Figura 32 - Distribuição dos transcritos relacionados à imunidade em *Podalia orsilochus*.
Legenda: O número indicado em parênteses representa os reads encontrados em cada família de gene específico.

Um subconjunto de 39 desses prováveis genes imunes está representado na tabela 8 com os respectivos detalhes como a quantidade de *reads* de cada *contig*, os números de acesso *GenBank*, similaridade, provável função e E-value (*BLAST*). Desse subconjunto foram excluídos apenas os *contigs* que apresentaram mais de um homólogo.

Os *contigs* foram novamente renomeados, desta vez com a designação POTI (uma sigla para Tegumento de *Podalia orsilochus*, subcategoria Imunidade), seguida por uma identificação númerica baseada nos nomes das sequências em ordem alfabética. Este procedimento foi adotado para facilitar a análise, identificação e descrição dos transcritos durante a discussão dos resultados.

Tabela 8 – Dados do subconjunto de contigs da Imunidade (totalizando 39 contigs).

***CONTIG* POTI**	***READS***	**PROTEÍNA PREDITA**	**SIMILARIDADE**	**E-value**	**PREDIÇÃO DE FUNÇÃO GERAL**
POTI0001	3	18 wheeler precursor *B. mori* [GenBank: NP_001116821.1]	97%	1e-81	Transdução de sinal/ Receptor transmembrânico
POTI0002	19	6Tox *G. mellonella* [GenBank: AAN15783.1]	39%	0.29	Resposta de defesa antimicrobiana
POTI0003	4	akirin-2-like - PREDICTED *P. xylostella* [GenBank: XP_011559837.1]	92%	1e-32	Resposta de defesa contra bactérias Gram-negativas
POTI0004	24	Antimicrobial protein 6Tox precursor *B. mori* [GenBanK: NP_001037004.1]	41%	2e-22	Resposta de defesa antimicrobiana

POTI0005	28	Apolipophorin III *T. ni* [GenBank: ABV68867.1]	82%	3e-80	Resposta de defesa contra bactérias Gram-positiva
POTI0006	15	Argonaute 2 *B. mori* [GenBank: NP_001036995.2]	77%	8e-152	Regulação negativa da replicação do genoma viral
POTI0007	26	Attacin *P. xuthus* [GenBank: ACR82290.1]	68%	4e-32	Resposta de defesa contra bactérias
POTI0008	24	Beta-1,3-glucan recognition protein 1 *H. armigera* [GenBank: ACI32825.1]	85%	0.0	Reconhecimento de superfície celular bacteriana
POTI0009	4	Caspase 6 *G. mellonela* [GenBank: AEK20837.1]	70%	1e-60	Resposta contra bactéria Gram-negativa e vírus
POTI0010	15	Cecropin precursor *B. mori* [GenBank: NP_001037392.1]	90%	1e-05	Resposta de defesa contra bactérias
POTI0011	12	Chlorophyllide A binding protein, partial *O. brumata* [Genbank: KOB65254.1]	63%	0.005	Não estabelecida
POTI0012	116	Conserved hypothetical protein *C. quinquefasciatus* [GenBank: XP_001869170.1]	55%	2e-05	Resposta de defesa
POTI0013	26	Defensin 2 *M. sexta* [GenBank: ADQ00386.1]	94%	1e-34	Resposta de defesa contra fungos e bactérias Gram+
POTI0014	19	Defense protein 3-like (LOC106110585), mRNA-PREDICTED *P. polytes* [GenBank: XM_013292466.1]	72%	1e-34	Resposta de defesa contra bactérias
POTI0015	172	Defense protein Putative *A. selene* [GenBank: AJW81891.1]	55%	1e-33	Resposta contra protozoários
POTI0016	19	Endochitinase A-like - PREDICTED *B. mori* [GenBank: XP_004929147.1]	73%	1e-43	Resposta de defesa contra fungos
POTI0017	3	Gallerimycin *S. exigua* [GenBank: ADJ95798.1]	58%	2e-07	Resposta de defesa contra fungos
POTI0018	50	Gloverin *C. suppressalis* [GenBank: AHC94952.1]	58%	1e-29	Resposta de defesa contra bactérias
POTI0019	11	Heliomicin precursor *H. virescens* [GenBank: ACR78445.1]	89%	2e-33	Resposta de defesa contra fungos e bactérias Gram-positivas
POTI0020	31	Integumentary mucin C.1 PREDICTED *M. domestica* [GenBank: XP_011291734.1]	85%	1e-11	Resposta de defesa contra microrganismos
POTI0021	28	Hemicentin-2 *P. xuthus* [GenBank: KPI98301.1]	49%	1e-14	Resposta a estímulos de defesa
POTI0022	177	Immunlectin A *A. pernyi* [GenBank: AFC35299.1]	54%	2e-101	Reconhecimento de superfície celular bacteriana
POTI0023	169	Lebocin 1 *S. exigua* [GenBank: AKJ54499.1]	75%	2e-22	Resposta de defesa contra bactéria Gram-positiva e fungos
POTI0024	2	Lysosomal-trafficking regulator-like - PREDICTED *P. machaon* [GenBank: XP_014360803.1]	95%	1e-90	Resposta de defesa contra bactéria, protozoário e vírus
POTI0025	32	Lysozyme c-1 *P. xuthus* [GenBank: KPI97757.1]	89%	1e-75	Atividade bacteriolótica
POTI0026	3	MAM domain-containing glycosylphosphatidylinositol anchor protein 1 *P. xuthus* [GenBank: KPJ04907.1]	77%	1e-60	Não estabelecida
POTI0027	3	Methylthioribulose-1-phosphate dehydratase Putative *P. xuthus* [GenBank: KPI95852.1]	91%	6e-41	Controle de proliferação bacteriana
POTI0028	4	Peptidoglycan recognition protein *D. plexippus* [GenBank: EHJ74577.1]	81%	1e-77	Reconhecimento de superfície celular bacteriana
POTI0029	18	Prophenoloxidase-1 *B. mori* [GenBank: AAG09304.1]	82%	0.0	Integrante da cascata da Pro-fenoloxidase
POTI0030	3	Protein toll-like - PREDICTED *B. mori* [GenBank: XP_012546802.1]	46%	1e-11	Resposta de defesa contra bactéria Gram-positiva e fungos
POTI0031	3	Scavenger receptor - PREDICTED class B member 1 *P. polytes* [GenBank: XP_013146600.1]	95%	1e-105	Regulação da fagocitose e interação vírus – hospedeiro
POTI0032	59	Serine protease inhibitor 012 *C. suppressalis* [GenBank: AEW46891.2]	43%	1e-12	Não estabelecida
POTI0033	566	Serpin 1b *P. xylostella* [GenBank: BAF48334.1]	67%	1e-87	Inibidor de protease
POTI0034	14	Serine protease persephone, partial *P. xuthus* [GenBank: KPJ01245.1]	75%	1e-81	Resposta de defesa contra fungos
POTI0035	135	TNFSF13 *B. mori* [Genbank: NP_001243914.1]	92%	1e-76	Regula o crescimento e função de outras células
POTI0036	5	Toll-interacting protein *P. machaon* [GenBank: KPJ07013.1]	83%	1e-109	Resposta de defesa contra bactéria Gram-positiva e fungos
POTI0037	14	Uncharacterized protein LOC106117995 - PREDICTED *P. xuthus* [GenBank: XP_013167996.1]	81%	1e-94	Resposta de defesa
POTI0038	62	Zonadhesin-like – PREDICTED *A. transitella* [GenBank: XP_013184383.1]	66%	1e-45	Não estabelecida
POTI0039	4	WD repeat-containing protein 48 homolog *P. xylostella* [GenBank: XP_011560744.1]	99%	2e-90	Possível resposta de defesa contra vírus

* Dados obtidos através do processo de anotação do programa Blast2GO e através da ferramenta BLASTx.

5.8.1 Grupos funcionais dos contigs *da subcategoria Imunidade*

Para uma melhor compreensão dos genes imunes e seus papéis funcionais dentro dessas respostas, as prováveis proteínas imunes foram categorizadas em diferentes grupos funcionais, conforme Gandhe et al. (2006), sendo estes:

I- Peptídeo anticrobiano (AMP);

II – Receptor de reconhecimento de padrão (PRRs) e de vias de sinalização;

III - Protease e inibidor de protease;

IV - Provável proteína de defesa com função desconhecida.

Os homólogos de várias outras proteínas envolvidas na imunidade, tais como antioxidantes, ferritinas e componentes da via apoptótica também foram encontrados, mas não foram incluídos nesses grupos funcionais, devido a não apresentarem somente funções específicas à resposta imune.

5.8.1.1 Grupo I- Peptídeo antimicrobiano (AMP)

O repertório do grupo de peptídeos antimicrobianos encontrados nas sequências de *Podalia orsilochus* apresentou 9 diferentes AMPs, sendo estes: Atacina, Cecropina, Galerimicina, Galiomicina, Gloverina, Heliomicina, Lebocina, Lisozima e X-tox, que apresentaram uma variedade de frequências de expressão representadas pelos números de *reads* encontrados, além da presença de sequências homólogas em 5 dos 9 peptídeos encontrados.

Um total de 651 *reads* e 16 isoformas de sequências de AMPs foi encontrado. A tabela 9 apresenta os dados descritos acima, além de informações suplementares sobre esses AMPs, como seu espectro de ação antimicrobiana e a classe estrutural a qual pertence.

Tabela 9 - Abundância em *reads** dos prováveis contigs de peptídeos antimicrobianos e os números de isômeros encontrados em *Podalia orsilochus*.

AMPs de *P. orsilochus*	Nº de *reads*	Nº de homólogos	Atividade antimicrobiana	Classe estrutural dos AMPs
Atacina	69 *reads*	3	Bactérias Gram + e - / Alguns fungos	AMPs rico em glicina
Cecropina	15 *reads*	-	Bactérias Gram + e - / fungos / parasitas e vírus	AMPs α-hélice
Galerimicina	3 *reads*	-	Fungos	AMPs cíclico
Galiomicina	26 *reads*	-	Fungos	AMPs cíclico
Gloverina	77 *reads*	2	Bactérias Gram + e - / Fungos / vírus	AMPs rico em glicina

Heliomicina	11 *reads*	-	Fungos	AMPs cíclico
Lebocina	285 *reads*	2	Bactérias Gram + e - / Fungos	AMPs rico em prolina
Lisozima	137 *reads*	7	Bactérias Gram +	-
X-tox	28 *reads*	2	Provavelmente apenas fungos	Provável AMPs cíclico
Total de AMPs	651 *reads*	16		

*O número de *reads* indica a abundância de expressão de cada sequência.

O provável peptídeo lebocina foi o *contig* expresso em maior abundância em termos de *reads* dentro da subcategoria Imunidade e apresentou apenas dois homólogos diferentes de lebocina, representando 44% (285 *reads*) do repertório de AMPs encontrados em *Podalia orsilochus* (Figura 33). Lebocinas, assim como as gloverinas, que apresentaram o terceiro *contig* mais expresso são peptídeos antimicrobianos presentes apenas em lepidópteros. Gloverinas apresentaram 12% (77 *reads*) dos genes de AMPS e duas sequências homólogas. O homólogo de lisozimas foi a segunda proteína mais expressa, com 137 *reads* representando 21% dos AMPs e sequências que mais apresentou isoformas, com um total de 7 variantes. Atacina foi o quarto mais abundante em número de *reads*, representando 11% (69 *reads*) do total de sequências de AMPs encontrados. Três peptídeos derivados de defensinas foram encontrados nos transcritos de Podalia, sendo estes galiomicina 26 *reads*, heliomicina com 11 *reads* e galerimicina, com 3 *reads*. A nova classe de AMP derivada de defensinas, X-tox apresentou 28 reads e duas isoformas e cecropina com 15 *reads* e apenas um tipo de sequência.

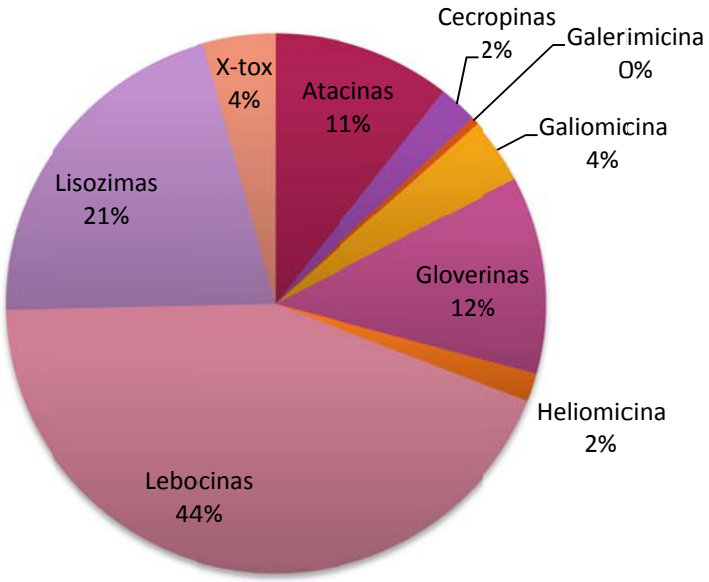


Figura 33 - Porcentagem dos peptídeos antimicrobianos encontrados nas sequências de *Podalia orsilochus.*

5.8.1.2 Grupo II - Receptores de reconhecimento de padrões (PRRs) e de vias de sinalização

Em insetos os PRRs detectam e ligam-se a padrões moleculares associados à patógenos (PAMPs) como lipopolissacarídeo, peptideoglucano, β1-3 glucano, ácido lipoteicóico, entre outros. Os prováveis PRRs identificados no perfil transcriptômico imune de *P. orsilochus* foram as proteínas de reconhecimento β-glucano (βGRPs), proteínas de reconhecimento peptideoglicano (PGRPs), apolipoforinas, lectinas e mucinas.

Dentro desse grupo, as proteínas apoliforinas e as proteínas de reconhecimento β-glucano (βGRPs), foram as mais representativas, tanto em número de homólogos, quanto em número de *reads*, onde apolipoforina apresentou 11 sequências homólogas e um total de 238 *reads* e as βGRPs apresentaram 7 tipos distintos de homólogos, representados por 210 *reads*.

Quatro homólogos da proteína PGRP, com um total de 53 *reads* foram encontrados no repertório de genes imunes de *P. orsilochus.* Para lectinas foram encontrados 17 homólogos e um total de 405 *reads*. Para mucina foi encontrado apenas um tipo, representado por 31 *reads*.

Em insetos, a sinalização celular contra fungos e bactérias patógenas ocorre através das vias Toll, IMD e Jak-STAT, que induzem a expressão de peptídeos antimicrobianos e outras moléculas através da interação com os fatores de NFkB. As principais vias de sinalização são as vias Toll e IMD, que são representadas por receptores centrais, como Toll, toll-like e 18 wheeler. Foram identificadas 4 proteínas Toll diferentes no conjunto de contigs de *P. orsilochus* e uma sequência muito semelhante ao precursor da proteína 18 wheeler encontrado em *B. mori* (GenBank NP_001116821.1).

5.8.1.3 Grupo III - Proteases e inibidores de proteases

Várias proteases e inibidoras de proteases regulam diversos mecanismos da imunidade, como a melanização, fagocitose e indução de peptídeos antimicrobianos. Nos transcritos de *P. orsilochus* foram observadas a presença de 4 homólogos de proteinases que ativam a cascata da pro-fenoloxidase, possuindo 53 *reads* no total. Serpinas foram as proteínas mais bem representativas no repertório de genes imunes de *P. orsilochus*, sendo representadas por 1121 reads e apresentando 12 isoformas da proteína.

5.8.1.4 Grupo IV – Prováveis proteínas de defesa com função desconhecida

Muitos potenciais novos membros de famílias de proteínas conhecidas foram identificados no presente estudo, porém, todos apresentam pouca ou nenhuma informação sobre sua atuação em respostas imunológicas. Entre os novos membros, um era uma proteína lisozima-like putativa descrita na seção anterior. Duas moléculas semelhantes a imunoglobulinas (Ig) foram identificadas pela presença de domínios conservados e representavam 36 reads. Uma variedade de inibidores de proteases e proteases putativas também foi encontrada. Assim como a proteína denominada "proteína de defesa", que foi uma das mais bem representadas em número de *reads* dentro dos genes imunes, sendo representada por 172 *reads* em apenas uma única sequência.

6 DISCUSSÃO

Através dos resultados observados foi possível verificar que a extração, a purificação e a avaliação do mRNA obtidos do tegumento da lagarta *Podalia orsilochus* foram realizados com sucesso, e o pirosequenciamento foi considerado satisfatório, fato evidenciado pela a grande quantidade de sequências aceitas pelos critérios de qualidade exigidos pelo programa do equipamento utilizado e através das análises posteriores.

O sequenciamento realizado através do aparelho 454 GS-Junior foi fundamental devido a sua característica de geração de *reads* de maior comprimento, que facilitou a montagem das sequências, por se tratar de um sequenciamento *de novo* (sequenciamento de genomas desconhecidos) (MARGULIES et al., 2005; SCHUSTER, 2008; VOELKERDING; DAMES; DURTSCHI, 2009).

Alguns fatos apresentados contribuíram para evidenciar o sucesso obtido na montagem das sequências de *P. orsilochus*. O primeiro fato pode ser observado na figura 20 e está relacionado às altas similaridades encontradas com sequências de diversas espécies de insetos, sendo observada a presença de algumas espécies pertencentes à ordem Lepidoptera, principalmente a conhecida popularmente como borboleta-monarca *Danaus plexippus*, que obteve a maior quantidade de *hits* alinhados com as sequências analisadas, o bicho-da-seda *Bombyx mori* e a borboleta *Papilio xuthus.*

Grande parte da similaridade encontrada com espécies de outras ordens de insetos, que não a Lepidoptera, seja provavelmente, devido ao fato de que essas espécies sejam muito mais estudas dentre todas as espécies pertencentes à classe dos insetos e, por isso, tenham uma maior quantidade de sequências depositadas no banco de dados do BLAST. Vale ressaltar que dentre as 160.000 (valor aproximado) espécies de lepidópteros descritos até o momento, apenas 23 delas possuem seu genoma completo depositado nos bancos de dados do Genbank e nenhuma dessas espécies é do gênero *Podalia* ou de alguma espécie que faça parte da família Megalopygidae.

O segundo fato que corrobora o sucesso observado na montagem dos *contigs* é mostrado na figura 21, que indica que os melhores alinhamentos (*best hits*) realizados entre as sequências da *P. orsilochus* e todas as outras sequências depositadas no GenBank foram obtidos com sequências provenientes de insetos, mas com uma predominância muito maior com sequências de lepidópteros. Dentre as 30 espécies que apresentaram os melhores

alinhamentos, 17 são representantes da ordem Lepidoptera (*D. plexippus, B. mori, P. xuthus, P. polytes, M. sexta, H. armigera, S. frugiperda, C. suppressalis, S. exígua, H. melpomene, S. litura, H. virescens, B. betularia, L. obliqua, P. rapae, S. littoralis e, E. aurinia*). Os melhores alinhamentos (best *hits*) levam em consideração os alinhamentos com a maior área de cobertura de similaridades, o menor valor E-value e a porcentagem de identidade entre as sequências analisadas. Esses dados são de grande importância, uma vez que, todas as espécies elencadas nos gráficos das figuras 20 e 21, são espécies representantes da classe dos insetos, sugerindo que a montagem dos *contigs* da *P. orsilochus* foi realizada com êxito.

Os *contigs* provenientes de *P. orsilochus* foram submetidos a um processo de anotação, no qual todas as sequências correspondentes aos mRNAs produzidos pela lagarta no momento da extração da amostra são utilizados como entrada em um banco de dados que realiza uma busca de similaridade com outras sequências de DNA ou proteína que foram encontradas de diversos organismos e depositadas para consultas. No momento em que essas sequências encontram similaridades com outras sequências conhecidas, pode-se inferir diversas características a essa sequência desconhecida, tais como as prováveis estruturas, funções e processos que essa proteína possa estar envolvida.

Os processos através dos quais as proteínas são encaminhadas para o seu destino final adequado dentro de uma célula, é um aspecto fundamental da vida celular. Em muitos casos, a triagem depende de “sinais" que podem ser identificados examinando a estrutura primária de uma proteína. Assim, o direcionamento de uma determinada proteína para a via secretora ou para organelas celulares normalmente depende de uma pré-sequência N-terminal do peptídeo alvo que são reconhecidos por receptores na superfície da organela apropriada (NIELSEN; BRUNAK; VON HEIJNE, 1999).

A triagem da previsão de sinais da proteína a partir da sequência de aminoácidos é de grande importância no campo da proteômica. Recentemente, o crescimento de bases de dados de proteínas, em combinação com abordagens de bioinformática, tais como redes neurais e modelos como “*hidden Markov model*” (TMHMM), tornou possível alcançar um nível de confiabilidade no caso da utilização prática de base de dados de anotações automáticos. Por definição, a célula é capaz de reconhecer todos os tipos de sinais de triagem de proteínas com quase 100 % de seletividade e especificidade. Dado que os sinais de triagem acima mencionados parecem ser definidos por um estiramento N-terminal linear dos polipeptídeos, essas ferramentas de bioinformática geralmente podem reconhecer estes

sinais com uma eficiência que se aproxima a da própria célula (NIELSEN; BRUNAK; VON HEIJNE, 1999).O peptídeo sinal para secreção de proteína trata-se, tipicamente, de uma sequência peptídica N-terminal formada por 15 a 40 aminoácidos composto por três regiões: uma região “n” carregada positivamente, uma região composta de aminoácidos hidrofóbicos denominada região “h” e uma região “c” terminal formada por pelo menos 6 resíduos, até o sítio de clivagem do peptídeo sinal (EMANUELSSON et al., 2000; NIELSEN et al., 1997; NIELSEN, KROG, 1998).

A topologia de uma proteína de membrana, isto é, a especificação dos segmentos transmembrânicos e sua orientação geral na membrana, é uma característica estrutural fundamental. Estudos experimentais e estatísticos da chamada hélice de empacotamento de proteínas de muitos diferentes sistemas de membrana têm apontado os resíduos com carga que flanqueiam os segmentos transmembranares hidrofóbicos como topologia predominante determinante (VON HEIJNE, 1997), embora outras propriedades, tais como o comprimento de um segmento hidrofóbico e a cinética de dobragem de domínios extramembranosos também contribua (DENZER et al., 1995; WAHLBERG; SPIESS, 1997).

O modelo “*hidden Markov model*” (TMHMM), foi utilizado para a predição da topologia transmembrânica (KROGH et al., 2001). Esse modelo, como uma estrutura probabilística, têm sido amplamente aplicado em biologia computacional com notável sucesso (DURBIN et al., 1998). Baseado nessas predições é estimado que cerca de 20 a 30 % de todos os genes na maioria dos genomas codifiquem proteínas de membrana (KROGH et al., 2001).

As sequências do transcriptoma de *P. orsilochus* foram avaliadas quanto à presença ou não de características como o peptídeo sinal e a topologia transmembrânica e foi observada a presença de peptídeo sinal em 22 % e 16 % para a topologia transmembrânica.

As enzimas possuem papel fundamental no metabolismo, apresentando atividade tanto intra quanto extracelulares, atuando como catalisadores em todas as vias metabólicas que ocorrem nos seres vivos. As enzimas são divididas em classes e uma variedade de subclasses distintas, de acordo com o tipo de reações químicas que estão envolvidas e são classificadas em 6 principais classes: oxidoredutases (EC1); transferases (EC2); hidrolases (EC3); Liases (EC4); Isomerases (EC5); e ligases (EC6). O código enzimático para as sequências de *P. orsilochus* foi obtido através do recurso *Enzyme code* do programa Blast2GO, o qual

verifica a presença de domínios que são associados com as funções catalíticas das enzimas e atribui o código para as sequências que apresentam esses domínios. Os domínios podem ser considerados unidades funcionais e / ou estruturais distintas de uma proteína que, independente da dobragem de uma cadeia polipeptídica, realiza uma função específica. Domínios conservados são unidades recorrentes em evolução molecular, cuja extensão em uma cadeia polipeptídica contém padrões de sequências conservados ou motivos, que permitem a sua detecção por análise de sequência e estrutura. Em evolução molecular tais domínios podem ter sido utilizados como blocos de construção, e podem ter sido recombinados nos diferentes arranjos para modular a função da proteína (MARCHLER-BAUER et al., 2015).

Através desse resultado podemos observar quais enzimas e a quantidade respectiva que estão sendo produzidas no momento da extração da amostra do tegumento da lagarta. Notamos uma maior quantidade de *reads* que codificam enzimas da classe das hidrolases, sendo um total de 303 *reads*. As transferases (251 *reads*) e oxidoredutases (163 *reads*) tiveram uma quantidade de *reads* superior as liases, isomerases e ligases, que respectivamente apresentaram 30, 24 e 88 *reads* cada.

A partir da obtenção das informações fornecidas pelo processo de anotação e dos dados adicionais, como a caracterização de uma enzima ou da predição de peptídeo sinal, podemos agrupar e classificar essas sequências de acordo com suas características em comum. Um padrão de classificação eficiente, simples e de qualidade é difícil de ser criado, pois os genes e proteínas não exibem padrões fixos que seriam necessários para o melhor entendimento do funcionamento dos genomas. Dessa forma, foi optado por utilizar o padrão de classificação de função de proteínas adotado pelo banco de dados *COG* (http://www.ncbi.nlm.nih.gov/COG/), que é amplamente conhecido na comunidade científica e que foi criado e é utilizado num dos maiores centros de bioinformática do mundo, o *NCBI*. É importante notar que, devido à artificialidade dos grupos de classificação, alguns genes foram incluídos em mais de uma categoria funcional. Além disso, devido ao fato do padrão de classificação criado pelo *COG* ter sido desenvolvido para lidar com organismos procariotos (apesar de conter também classificações de eucariotos), foi necessária a criação de novas subcategorias e de adaptar certas famílias de proteínas em categorias já existentes.

As 6.659 sequências provenientes da *Podalia* foram analisadas e após a classificação de cada *contig* na sua respectiva categoria principal, foi possível observar que 43% de todos os *contigs* obtidos do tegumento de *P. orsilochus*, correspondem a genes que são pobremente caracterizados. Esse resultado já era esperado, uma vez que o objeto desse estudo representa uma espécie pouco conhecida e até o momento nunca estudada com tal abrangência. A segunda categoria com maior representação, foi a categoria de Processos celulares, com 29% de todos os *contigs* analizados, seguido por Metabolismo, com 14%, Processamento e armazenamento de informações, com 12 % e 2 % corresponde a nova categoria Morte celular e resposta imune. Estes dados estão apresentados na figura 30.

Foi realizada uma breve verificação dos primeiros cinco *contigs* com maior frequência de ocorrência no transcriptoma do tegumento de *P. orsilochus*, que é indicado pelo número de *reads* que foram utilizados para a montagem de um único *contig*. Os cinco *contigs* mais representativos foram analizados manualmente no *BLAST* e avaliados em bancos de dados de informações de proteínas, como o UNIPROT, na tentativa de identificar os genes mais importantes para a sobrevivência da espécie nessa etapa de transição de larva para pupa, e verificar através de similaridade, se as proteínas que estavam sendo transcritas em maior número de vezes no momento da extração do mRNA do tegumento da *P. orsilochus* são proteínas geralmente transcritas nesse tipo de tecido e de organismo.

Os dados indicam que os *reads* encontrados em maior frequência no transcriptoma são característicos de proteínas presentes no tegumento de lagartas, destacando-se a presença de proteínas estruturais musculares, além de proteínas que são expressas durante o estágio larval. Abaixo estão descritas de forma resumida as possíveis funções inferidas para os cinco *reads* mais encontrados.

O *contig* de maior expressão foi o POT0001 (“*heat shock protein 1”l*), que possui 1257bp de comprimento e foi inserido na categoria Metabolismo, subcategoria Q - Biossíntese, catabolismo e transporte de metabólitos secundários, totalizando 1914 *reads*. Proteínas de choque térmico (*heat shock proteins - HSPs*) são chaperonas importantes, que estão envolvidas em várias vias de sinalização e regulam muitos processos fisiológicos. As primeiras pesquisas sugerem que algumas *HSPs* estão envolvidas no desenvolvimento dos insetos e, um estudo recente, forneceu informações importantes sobre a atuação dessas proteínas durante a metamorfose ocorrida entre larva e pupa, demonstrando que essas proteínas são, provavelmente, as principais intervenientes na metamorfose do intestino

médio em *Spodoptera litura*. O produto desta sequência gênica possui um papel vital no desenvolvimento embrionário e é altamente expresso durante todo o desenvolvimento larval (GU et al., 2012). Os resultados da análise desse *contig* no *BLAST*, demonstrou em seus dez melhores hits, grande similaridade com sequências encontradas em insetos da ordem Lepidoptera. Toda informação encontrada sobre essa proteína vai de encontro com o estágio de desenvolvimento em que se encontrava a lagarta *Podalia* no momento da extração do seu tegumento, que estava próximo ao seu último instar larval.

O *contig* POT0002 foi o segundo *contig* com maior frequência de *reads*, com ocorrência de 1390 *reads* e mostrou alta similaridade com uma sequência protéica descrita como "*muscular protein 20".* Essa proteína foi primeiramente isolada em *Drosophila melanogaster*, porém, é largamente encontrada em diversas espécies de inseto, incluindo os da ordem Lepidoptera e codifica uma proteína de 20-KD específica de músculo (AYME-SOUTHGATE, 1989). Esse *contig* também apresentou homologia parcial, com uma proteína de mesmo nome, encontrada em um catálogo de genes obtido através de uma construção de biblioteca de ESTs da epiderme de larvas de *Papilio xunthus* e *P. polytes*, que contêm respectivamente, 20.736 e 5.376 clones e, representa atualmente uma das maiores bibliotecas de genes de borboletas disponíveis (FUTAHASHI et al., 2012). A comparação da sequência nucleotídica do *contig* da lagarta *Podalia* com a sequência disponibilizada nesse catálogo, teve uma cobertura de 61%, com 77% de identidade e obteve um E-value de 4e-121.

O *contig* POT0003 apresentou a terceira maior ocorrência de *reads*, com 1255 *reads*, onde, no processo de anotação realizado pelo programa *Blast2GO*, foi descrita como "*long form-like"* e, não foi inferida nenhuma função, localização ou processo no qual essa proteína poderia estar envolvida nos bancos de dados consultados através do programa. Porém, uma análise mais aprofundada dessa sequência, mostrou uma alta similaridade com uma sequência gênica descrita como *"paramyosin"*, que corresponde a uma sequência encontrada em diversos insetos da ordem Lepidoptera. A proteína *"paramyosin long form"* foi isolada e caracterizada em *Drosophila melanogaster* e, corresponde a uma proteína isolada a partir de músculos de invertebrados, que corresponde a principal componente estrutural dos filamentos grossos musculares. A *"paramyosin long form"* é expressa em todos os tecidos musculares de larvas e adultos e sua expressão é cinco vezes maior nos músculos tubulares do que nos músculos fibrilares (VINÓS et al., 1991).

O *contig* POT0004 possuiu a quarta maior frequência de *reads*, com ocorrência de 1216 *reads*. Esse *contig* apresentou alta similaridade com uma sequência gênica descrita como "*elongation factor 2*". Uma análise funcional da proteína "*elongation factor 2*" realizada através do banco de dados UniProt/SWISS-PROT, a descreve como uma catalisadora da translocação ribossomal dependente de GTP durante a etapa de alongamento das cadeias polipeptídicas durante a tradução. Durante esta etapa, o ribossomo altera entre o seu estado conformacional de pré-traducional para pós-traducional. A proteína em questão, é responsável por catalisar o movimento coordenado de duas moléculas de mRNA e as mudanças conformacionais no ribossomo. Por similaridade, esta proteína "*Elongation factor2"* pertence a família "*GTP-binding elongation factor"* e a subfamília "*EF-G/EF-2"*. Uma análise da sequência do *contig* 6569 no banco de dados Interpro, apresentou um domínio protéico denominado "*Translation elongation factor EFG/EF2"*, que é o responsável pelos dois processos que ocorrem durante a síntese de proteínas no ribossomo que estão descritos acima, sendo responsáveis pela precisão da tradução e tanto *EFG e EF2* são extremamente conservados ao longo da evolução (WINTERMEYER et al., 2001). A análise dessa proteína no banco de dados KEGG, mostrou que essa proteína está ligada a apenas uma via de sinalização, que é a via de sinalização AMPK. A indicação é de que quando essa proteína é ativada por AMPK, ela inibe a síntese protéica.

O *contig* POT0005 foi o quinto de maior frequência em *reads*, com ocorrência de 1037 *reads*. A similaridade ocorreu com a sequência descrita como "*Actina*", responsável por codificar a proteína actina. A actina é uma proteína estrutural encontrada em todas as células eucarióticas, é a maior componente dos filamentos de células musculares e do citoesqueleto presente em todas as células. Em conjunto com a proteína miosina e moléculas de ATP, gera movimentos celulares e musculares.

Após uma avaliação geral dos transcritos produzidos pela *P. orsilochus*, foi realizada uma análise do perfil dos genes relacionados às respostas imunológicas nesse inseto. Essa análise foi realizada através dos 110 *contigs* que foram incluídos na subcategoria Imunidade e, podemos considerar que a quantidade de genes que apresentaram indicações de participação nesse processo foi satisfatória, levando em consideração sua especificidade em comparação com as outras subcategorias mais abrangentes, ressaltando ainda que, dentre os 6.659 *contigs* totais, 42% representam *contigs* com função desconhecida, além dos que

apresentaram apenas predição de função geral, podendo existir uma relativa parcela desses *contigs* com funções ligadas à resposta imune da espécie *P. orsilochus* que ainda são desconhecidos.

Informações gerais dos principais *contigs* que apresentaram atividade bem definida, com suas funções específicas ligadas à resposta imune estão brevemente introduzidas a seguir.

O contig POTI0003 apresentou similaridade com uma proteína chamada *akirin protein.* Inicialmente descoberta em Drosófila, akirina mostrou ser altamente conservada em diversas espécies, inclusive foi encontrada no genoma humano contendo dois homólogos. Em Drosófila, duas vias de sinalização distintas, a denominada deficiência imune (*immune deficiency* - Imd) e a via *Toll* controlam a transcrição de genes de peptídeos antimicrobianos. Infecções causadas por fungos ou bactérias Gram-positivas ativam a via *Toll*. Akirina possui localização nuclear e têm uma função essencial à jusante do receptor do tipo *Toll*, fator de necrose tumoral e interleucina 1-β (IL-1β), as vias de sinalização que conduzem à produção de IL-6. Assim, akirina é um fator nuclear conservado necessário para resposta imune inata (GOTO et al., 2008).

O *contig* POTI0014 apresentou similaridade com sequências descritas como *defense protein* em diversas espécies de insetos, com predominância nos insetos da ordem Lepidoptera. Na literatura disponível sobre essa proteína consta um estudo realizado em *B. mori*, que sugere que a *defense protein* está envolvida na ativação da cascata da pro-fenoloxidase. Os resultados fornecem novas evidências no mecanismo da cascata de melanização da imunidade de insetos (BAO et al., 2011).

Sete *contigs* representados pelo POTI0008 apresentaram similaridade com uma proteína denominada β-*glucan recognition protein*, também conhecida como proteína de ligação à bactérias Gram-negativas (βGNBP), pois possui uma forte afinidade com a parede celular de bactérias Gram-negativas. A proteína de 50-kDa foi purificada da hemolinfa de *B. mori* e a comparação da sua sequência de aminoácidos com sequências conhecidas, revelaram que GNBP contém uma região que exibe uma significativa homologia com uma região catalítica bacteriana de um grupo de beta-1,3 glucanases e beta-1,3-1,4glucanases (LEE et al., 1996).

Foram encontrados quatro *contigs* representados pelo POTI0028 que apresentaram similaridade com a proteína de reconhecimento ao peptideoglicano (PGRP), que é

considerada uma das principais proteínas de reconhecimento de padrões na imunidade inata dos insetos, ligando-se especificamente ao peptideoglicano. O PGRP é um componente essencial no desencadeamento da cascata da pro-fenoloxidase, que é um importante mecanismo de defesa do inseto. O gene PGRP é expresso constitutivamente no corpo gorduroso, células epiteliais e hemócitos de insetos, mas sua expressão é intensificada frente a uma infecção. Uma pesquisa de homologia revelou que homólogos do PGRP de insetos estão presentes em camundongos, nematóides e bacteriófagos, sugerindo que o reconhecimento de peptideoglicano estrangeiro, tanto em vertebrados como em invertebrados, é efetuado por PGRP homólogos com uma origem evolucionária comum (OCHIAI; ASHIDA, 1999).

Em insetos os PRRs detectam e ligam-se à padrões moleculares associados à patógenos (PAMPs) como lipopolissacarídeo, peptideoglucano, β1-3 glucano, ácido lipoteicóico, entre outros. Os prováveis PRRs identificados no perfil transcriptômico imune de *P. orsilochus* foram as proteínas de reconhecimento β-glucano, proteínas de reconhecimento peptideoglicano (PGRPs), lectina e mucina.

Dentro desse grupo, as proteínas de reconhecimento β-glucano foram as mais representativas, tanto em número de homólogos, quanto em número de *reads*, onde apresentou 7 tipos distintos de homólogos, representados por 210 *reads*.

Quatro homólogos da proteína PGRP, com um total de 53 *reads* foram encontrados no repertório de genes imunes de *P. orsilochus*, em comparação com *A. mylitta* que foram encontrados três tipos e em *D. melanogaster* que são conhecidos 13 PGRPs envolvidos na ativação de várias vias efetoras da imunidade.

O repertório do grupo de peptídeos antimicrobianos encontrados nas sequências de *Podalia orsilochus* foi bem representativo, contendo a presença de pelo menos um AMP representante de cada classe estrutural definida para essas moléculas envolvidas em respostas imunes.

Os prováveis peptídeos relacionados à atacina apresentaram três tipos diferentes de homólogos em *P. orsilochus*, em comparação com outras espécies como *A. mylitta* e em pupas de *H. cecropia* previamente imunizadas, demonstrou a expressão de seis tipos diferentes de atacinas (GANDHE et al., 2006). Em um estudo anterior, foi demonstrada a presença de quatro tipos diferentes de atacinas em *D. melanogaster*. Com relação às lisozimas encontradas, foi observada a presença de três tipos distintos homólogos as

lisozimas bacteriolíticas de outros lepidópteros, que são conhecidos por causar a lise de paredes celulares bacterianas [21] e dois tipos semelhantes à lisozima, que carece de um ou ambos os resíduos catalíticos essenciais para a atividade de muramidase. Esta proteína já foi relatada na espécie *A. mylitta* e é homóloga a lisozimas de dípteros, mas apresentam função desconhecida.

Os homólogos dos peptídeos atacina e lisozima foram os terceiros em maior abundância de *reads*, ambos representando 12,43% (69 *reads*) do total de sequências de AMPs encontrados. Dois homólogos de defensinas foram encontrados com uma frequência de 37 *reads*, representando 6,66% dos AMPs. Um deles apresentou grande simularidade com a sequência de um tipo defensina de *Heliothis virescens* denominada heliomicina, já o outro apresentou similaridade com um tipo de defensina de *Galleria mellonella*, chamado galiomicina.

Os últimos da lista de abundância de expressão de AMPs em *P. orsilochus* foram cecropina, 2,7% (15 *reads*) e apenas um tipo de sequência e uma sequência do peptídeo antifúngico galerimicina, com 0,54% (3 *reads*) de representatividade. Três tipos homólogos de cecropina foram encontrados nos transcriptomas de *A. mylitta* e *H. cecropia*, e o peptídeo foi o segundo mais expresso em ambas as espécies.

Galerimicina, até o momento, só foi encontrada em espécies de lepidópteros (*D. plexippus; A. pernyi; H. armigera; S. frugiperda; G. mellonella; S. litura; S. exígua; S. ricini; T. ni; L. obliqua*). Galerimicina é encontrada na literatura como um peptídeo antifúngico semelhante à defensinas que, inicialmente, foi clonado a partir de uma biblioteca de cDNA de hemócitos de *Galleria mellonella*. A sequência também mostra semelhanças com os peptídeos antifúngicos drosomicina de *D. melanogaster* e heliomicina de *Heliothis virescens* (SCHUHMANN et al., 2003).

Os dados gerados através do transciptoma de *P. orsilochus* estão em conformidade com a tendência da presença de genes antimicrobianos representantes de várias famílias de AMPs, como observado em muitas outras espécies de insetos e destacou a essencialidade destes genes no organismo.

Estudos com lagartas de *Antheraea mylitta* infectadas 24 horas antes com bactérias *E. coli* (GANDHE et al., 2006) demonstraram que a resposta imune em insetos é dinâmica e diferentes genes efetores são provavelmente expressos em diferentes pontos temporais

durante a infecção o que contribui para a capacidade dos insetos evitarem infecções, apesar da ausência da imunidade adaptativa.

Contudo, a lagarta *P. orsilochus*, aparentemente produz uma grande quantidade desses genes imunes constitutivamente, uma vez que a lagarta não fora infectada com nenhum tipo de microrganismo anterior a extração do tegumento para análise. Ressaltando que a maioria dos peptídeos antimicrobianos específicos, como foi o caso do peptídeo lebocina, a expressão em *P. orsilochus* superou a expressão de lebocina em *A. mylitta* previamente infectada com *E. coli*. Embora não seja possível afirmar que a lagarta *P. orsilochus* não estivesse sendo alvo de uma infecção natural, por não ter sido realizado nenhum tipo de teste para avaliar essa questão, os componentes da via IMD que estão implicados na ativação de várias vias efetoras sobre infecção bacteriana em *Drosophila* não estavam presentes no transcriptoma.

Outro dado interessante a destacar foi os diferentes tecidos utilizados para os transcriptomas. O transcriptoma de *A. mylitta* foi realizado com o corpo gorduroso da lagarta, que é conhecido como o principal tecido produtor de proteínas e peptídeos relacionados às respostas imunológicas dos insetos e, esperava-se que a expressão de genes imunes nesse tecido, superasse em muito a expressão dos genes imunes no tegumento que fora utilizado para a realização do transcriptoma de *P. orsilochus*.

7 CONCLUSÕES

Pelos dados obtidos e descritos acima, podemos concluir que, a extração, a purificação e a avaliação da qualidade do mRNA obtido do tegumento da lagarta *Podalia orsilochus* foram realizados com sucesso.

O resultado do pirosequenciamento feito a partir da amostra do mRNA extraído foi satisfatório, visto a grande quantidade de sequências aceitas pelos critérios de qualidade exigidos pelo programa do equipamento utilizado.

A montagem dos *contigs* foi correta, uma vez que uma grande quantidade dos mesmos apresentou alta porcentagem de similaridade com diversas sequências já depositadas e bem caracterizadas em insetos de diversas ordens, inclusive, apresentando os melhores alinhamentos com espécies da ordem Lepidoptera.

Foram apontadas várias características importantes para uma melhor identificação proteica, tanto funcional quanto estrutural, dos *contigs* representantes do transcriptoma do tegumento de *Podalia orsilochus*, como a presença de peptídeo sinal, o sinal de proteína transmembrânica e as funções enzimáticas, que foram classificadas de acordo com a nomenclatura de enzimas dadas pelo EC (*Enzyme code*), onde foram atribuídas suas principais classes.

Foi obtido um resultado esperado na classificação dos *contigs* em categorias funcionais, onde dos 6.659 *contigs* totais, 2.803 foram classificados na categoria Pobremente caracterizados, sendo que 95 *contigs* apresentaram apenas predição de função geral e, 2.708 foram incluídos na subcategoria de função desconhecida, indicando que realmente estamos lidando com uma espécie muito pouco estudada e com grandes possibilidades de descobertas de novas proteínas que possam ter funções de interesse biotecnológico e farmacológico.

Foram classificados noventa e nove *contigs* na subcategoria Imunidade, onde algumas sequências correspondem a proteínas com funções regulatórias do sistema imune, atuando tanto na regulação da expressão de genes imunes, quanto na ativação de diversas vias e da cascata da pro-fenoloxidase dos insetos e outras moléculas com atividade

específica efetora da imunidade inata, mais conhecidas como peptídeos antimicrobianos, que possuem ação direta aos organismos patogênicos invasores.

Foram encontrados três *reads* do *contig* POTI0017, que corresponde a um peptídeo antimicrobiano chamado galerimicina, isolado inicialmente na espécie de mariposa *Galleria mellonella* e que até o atual momento, só foi identificado em espécies da ordem Lepidoptera.

O perfil transcriptômico apresentado neste trabalho representa genes que provavelmente são expressos constitutivamente no tegumento de *P. orsilochus*, uma vez que a lagarta utilizada para o transcriptoma encontrava-se em seu estado natural no momento da extração do mRNA.

O presente estudo contribui para o aumento do repertório de prováveis genes específicos em insetos da ordem Lepidoptera, e descreve os genes expressos em uma lagarta de uma família (Megalopygidae) dessa ordem que nunca havia sido estudada nesse nível de abordagem, embora seja uma família que apresente espécies de importância médica e de considerável ocorrência no Brasil, como é o caso da *Podalia orsilochus*, dessa forma aumentando o valor de sua contribuição para o início de novos trabalhos nessa área de pesquisa.

As informações coletadas através da descrição dos genes envolvidos na resposta imunológica de *Podalia orsilochus*, faz desse trabalho um recurso valioso para estudos imunológicos específicos para lepidópteros, em particular, para os estudos das respostas de defesa dos insetos em geral.

REFERÊNCIAS*

ABDALLAH, N. A.; SHAH, D.; ABBAS, D.; MADKOUR, M. Stable integration and expression of a plant defensin in tomato confers resistance to fusarium wilt. **GM Crops,** v. 1, p. 344–350, 2010.

AERTS, A. M.; FRANÇOIS, I. E.; CAMMUE, B. P.; THEVISSEN, K. The mode of antifungal action of plant, insect and human defensins. **Cell Mol. Life Sci**., v. 65, p. 2069–2079, 2008.

ALEXANDER, J. O. **Arthropods and Human Skin**. Berlin: Springer Verlag-Heidelberg, 1984. 292 p.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. **Nucl. Acids Res.**, v. 25, p. 3389-3402, 1997.

ALVARENGA, Z. A taturana. VII Congresso Brasileiro de Medicina Cirúrgica. **Anais Belo Horizonte II**, v. 2, p. 132-135, 1912.

AMARAL, A. M.; REIS, M. S.; SILVA, F. R. **O programa BLAST:** guia prático de utilização. Brasília, DF: EMBRAPA Recursos Genéticos e Biotecnologia, 2007. 24 p.

ARDAO, M. I.; SOSA PERDOMO, C.; PELLATON, M. G. Venom of the *Megalopyge urens* (Berg) caterpillar. **Nature**, v. 209, p. 1139–1140, 1966.

AROCHA-PIÑANGO, C. L. Fibrinolisis producida por contacto con orugas: comunicación preliminar**. Acta Cientifica Venezolana**, v. 18, p. 136-139, 1967.

ARTUSO, A. Bioprospecting, benefit sharing, and biotechnological capacity building. **World Development**, v. 30, n. 8, p. 1355-1368, 2002.

AUBOURG, S.; ROUZÉ, P. Genome annotation. **Plant Physiology and Biochemistry**, v. 39, n. 3-4, p. 181-193, 2001.

AYME-SOUTHGATE, A.; LASKO, P.; FRENCH, C.; PARDUE, M.; L. Characterization of the gene for mp20: A *Drosophila* muscle protein that is not found in asynchronous oscillatory flight muscle. **The Journal of Cell Biology.** v. 108, p. 521-531, 1989.

AXEN, A.; CARLSSON, A.; ENGSTRÖM, A.; BENNICH, H. Gloverin, an antibacterial protein from the immune hemolymph of *Hyalophora pupae*. **Eur. J. Biochem.**, v. 247, n. 2, p. 614-619, 1997.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acid. Res.**, v. 28, n. 1, p. 45-48, 2000.

BAO, Y. Y.; XUE, J.; WU, W. J.; WANG, Y.; LV Z. Y.; ZHANG, C. X. An immune-induced reeler protein is involved in the *Bombyx mori* melanization cascade. **Insect Biochem. Mol. Biol.**, v. 41, n. 9, p. 696-706, 2011.

---

*De acordo com:
ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

BARTH, R. Estudos histológicos nas células glandulares dos insetos peçonhentos: II. Os órgãos urtificantes da largata de *Megalopyge albicollis* superba Edwards (Lepidoptera, Megalopygidae). **Memórias do Instituto Oswaldo Cruz**, v. 52, p. 125-147, 1954.

BELL, A. Antimalarial peptides: the long and the short of it. **Curr. Pharm. Des**., v. 17, p. 2719–2731, 2011.

BORÉM, A.; SANTOS, F. R. **Biotecnologia simplificada.** Minas Gerais: Editora Suprema, 2001.

BOURQUIN, F. Metamorfosis de *Megalopyge undulata vulpina* Berg, 1882 (Lep. Megalopygidae). **Revista de la Sociedad Entomológica Argentina**, v. 12, p. 387-392, 1945.

BROWN JR, K. S.; FREITAS, A. V. L. Lepidoptera. In: BRANDÃO, C. R. F.; CANCELLO, E. M. Biodiversidade do Estado de São Paulo, Brasil: síntese do conhecimento ao final do século XX. **Invertebrados Terrestres**. São Paulo: FAPESP, 1999. p. 225-243.

BROWN, S. E.; HOWARD, A.; KASPRZAK, A. B.; GORDON, K. H.; EAST, P. D. The discovery and analysis of a diverged family of novel antifungal moricin-like peptides in the wax moth *Galleria mellonella*. **Insect Biochem. Mol. Biol**., v. 38, p. 201‑212, 2008.

BULET, P.; HETRU, C.; DIMARCQ, J.; HOFFMANN, D. Antimicrobial peptides in insects; structure and function. **Developmental and Comparative Immunology**, v. 23, p. 329-344, 1999.

BULET, P.; STOCKLIN, R. Insect antimicrobial peptides: structures, properties and gene regulation. **Prot. Pep. Letters**, v. 12, n. 1, p. 3-11, 2005.

BURDMANN, E. A.; ANTUNES, I.; SALDANHA, L. B.; ABDULKADER, R. C. Severe acute renal failure induced by the venom of Lonomia caterpillars. **Clinical Nephrology**, v. 46, p. 337-339, 1996.

CALIXTO, J. B. **Medicamento fitoterápico no Brasil: produção indiscriminada sem o cumprimento da lei**. Disponível em: <http://www.conteudoexpresso.com.br/Mostra Conteúdo.asp>. Acesso em: 20 mar. 2014.

CARDOSO, J. L. C. **Animais peçonhentos no Brasil:** biologia, clínica e terapêutica dos acidentes. 2. ed. São Paulo: Sarvier, 2009. 540 p.

CARDOSO, A. E. C.; HADDAD JR, V. Acidentes por Lepidópteros (larvas e adultos de mariposas): estudo dos aspectos epidemiológicos, clínicos e terapêuticos**. An. Bras. Dermatol.,** v. 6, n. 80, p. 571-578, 2005.

CARDOSO, J. L. C.; FRANÇA, F. O. S.; WEN, F. H.; MALAQUE, C. M. S.; HADDAD JR, V. Animais peçonhentos no Brasil: biologia, clínica e terapêutica dos acidentes. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 45, n. 6, p. 338, 2003.

CARLSSON, A.; NYSTROMT, D. E.; COCK, H.; BENNICH, H. Attacin - an insect immune protein-binds LPS and triggers the specific inhibition of bacterial outer-membrane protein synthesis. **Microbiology,** v. 8, n. 144, p. 2179‑2188, 1998.

CARMO, A. C. V.; GIOVANNI, D. N. S.; CORREA, T. P.; MARTINS, L. M.; STOCCO, R. C.; SUAZO, C. A. T.; MORAES, R, H, P.; VEIGA, A. B. V.; MENDONÇA, R. Z. Expression of an antiviral protein from *Lonomia obliqua* hemolymph in baculovirus/insect cell system. **Antiviral Res**. v. 94, p. 126-130, 2012.

CARVALHO ADE, O.; GOMES, V. M. Plant defensins and defensin-like peptides - biological activities and biotechnological applications. **Curr. Pharm. Des**., v. 17, p. 4270‑4293, 2011.

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, n. 3, p.735-744, 2010.

CHAPMAN, A. D. **Numbers of Living Species in Australia and the World**. Canberra: Australian Biological Resources Study, 2009. 80 p.

CHERNYSH, S.; KIM, S. I.; BEKKER, G.; PLESKACH, V. A.; FILATOVA, N. A.; ANIKIN, V. B.; PLATINO, V. G.; BULLET, P. Antiviral and antitumor peptides from insects. **Proc. Natl. Acad. Sci.,** v. 99, p. 12628-12632, 2002.

COCIANCICH, S.; DUPONT, A.; HEGY, G.; LANOT, R.; HOLDER, F.; HETRU, C.; HOFFMANN, J. A.; BULET, P. Novel inducible antibacterial peptides from a hemipteran insect, the sap-sucking bug *Pyrrhocoris apterus*. **Biochem. J**., v. 300, p. 567–575, 1994.

COCK, M. J. W. On the Number of Moths (Lepidoptera) that Occur in Trinidad and Tobago. **Living World, J. Trinidad and Tobago Field Naturalists' Club**, p. 49–58, 2003.

CONESA, A.; GÖTZ, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. **International Journal of Plant Genomics**, v. 2008, 2008.

CORNET, B.; BONMATIN, J. M.; HETRU, C.; HOFFMANN, J. A.; PTAK, M.; VOVELLE, F. Refined three-dimensional solution structure of insect defensin A. **Structure,** v. 3, p. 435–448, 1995.

CORPET, F. Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Res**., v. 16, p. 10881-10890, 1988.

COSTA LIMA, A. M. Lepidópteros. In:_____(Ed.). **Insetos do Brasil**. Rio de Janeiro: Escola Nacional de Agronomia, 1945. Série Didática, tomo 5, nº 7, 376 p.

CREMA, A.; CASTELO BRANCO, M. Impact of temperature and photoperiod on ovarian development and oviposition of the Diamondback Moth. **Horticultura Brasileira**, v. 22, p. 305-308, 2004.

D'ALENÇON, E.; BIERNE, N.; GIRARD, P. A.; MAGDELENAT, G.; GIMENEZ, S.; SENINET, I.; ESCOUBAS, J. M. Evolutionary history of x-tox genes in three lepidopteran species: origin, evolution of primary and secondary structure and alternative splicing, generating a repertoire of immunerelated proteins. **Insect Biochem. Mol. Biol**., v. 43, p. 54–64, 2013.

DAVIES, K. **Decifrando o genoma:** a corrida para desvendar o DNA humano. São Paulo: Companhia das Letras, 2001.

DE ROODT, A. R.; SALONOM, O. D.; ORDURA, T. A. Accidents due to lepidoptera with special reference to *Lonomia sp*. **Medicina**, v. 60, p. 964-972, 2000.

DELCROIX, M.; RILEY, L. W. Cell-penetrating peptides for antiviral drug development. **Pharmaceuticals**, v. 3, n. 3, p. 448-470, 2011.

DENZER, A. J.; NABHOLZ, C. E.; SPIESS, M. Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. **EMBO J**. v. 14, p. 6311-6317, 1995.

DEPARTMENT OF ENERGY - US (2006). **Genomics GTL Roadmap**: systems biology for energy and environment. Washington: p. 76. Disponível em: <http://genomicsgtl.energy.gov/roadmap/pdf/GenomicsGTL_Roadmap_lowres.pdf>. Acesso em: 02 Jul. 2014.

DIMARCQ, J. L.; ZACHARY, D.; HOFFMANN, J. A.; HOFFMANN, D.; REICHHART, J.M. Insect immunity: expression of the two major inducible antibacterial peptides, defensin and diptericin, in *Phormia terranovae*. **EMBO J**., v. 9, p. 2507–2515, 1990.

DONG, Y.; TAYLOR, H.E.; DIMOPOULOS, G. AgDscam, a Hypervariable Immunoglobulin Domain-Containing Receptor of theAnopheles gambiae Innate Immune System. **PLoS Biol.**, v. 4, n. 7, p. 1137-1146, 2006.

DUARTE, M.; MARCONATO, G.; SPECHT, A.; CASAGRANDE, M. M. Lepidoptera. In: RAFAEL, J. A.; MELO, G. A. R.; CARVALHO, C. J. B.; CASARI, S. A.; CONSTANTINO, R. **Insetos do Brasil**. Diversidade e taxonomia. Holos, Ribeirão Preto, SP, 2012. p. 625-682.

DURBIN, R.; EDDY, S.; KROGH, A.; MITCHISON, G. Biological sequence analysis. Cambridge: UK Cambridge University Press, 1998.

DYAR, H.G. Notes on Megalopygidae. **Proceedings of the Entomological Society**, v. 12, p. 161-176, 1910.

EAGLEMAN, D. M. Envenomation by the asp caterpillar (*Megalopyge opercularis*). **Clinical Toxicology**, v. 46, p. 201-215, 2008.

EL-MALLAKH, R. S.; BAUMGARTNER, D. L.; FARES, N. "Sting" of the puss caterpillar *Megalopyge opercularis* (Lepidoptera:Megalopygidae): first report of cases from Florida and review of literature. **Journal of the Florida Medical Association**, v. 73, p. 521-525, 1986.

EMANUELSSON, O.; NIELSEN, H.; BRUNAK, S.; VON HEIJNE, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. **J. Mol. Biol.**, v. 300, n. 4, p. 1005-1016, 2000.

ENGSTROM, P.; CARLSSON, A.; ENGSTROM, A.; TAO, Z. J.; BENNICH, H. The antibacterial effect of attacins from the silk moth *Hyalophora cecropia* is directed against the outer membrane of *Escherichia coli*. **EMBO J**., v. 3 , p. 3347–3351, 1984.

EPSTEIN, M. E. Revision and phylogeny of the limacodid-group families, with evolutionary studies on slug caterpillars (Lepidoptera: Zygaenoidea). **Smithsonian Contributions to Zoology**, v. 582, p. 1-102, 1996.

EPSTEIN, M. E.; GEERTSMA, H.; NAUMANN, C. M.; TARMANN, G. M. 10. The Zygaenoidea. *Handbook of Zoology***. Arthropoda Insecta,** v. 4, p. 159-180, 1999.

ETEBARI, K.; PALFREYMAN, R. W.; SCHLIPALIUS, D.; NIELSEN, L. K.; GLATZ, R. V.; ASGARI, S. Deep sequencing-based transcriptome analysis of *Plutella xylostella* larvae parasitized by Diadegma semiclausum. **BMC Genomics**. v. 12, p. 446, 2011.

FAYE, I.; PYE. A.; RASMUSON, T.; BOMAN, H. G.; BOMAN, I. A. Insect immunity. 11. Simultaneous induction of antibacterial activity and selection synthesis of some hemolymph proteins in diapausing pupae of *Hyalophora cecropia* and *Samia cynthia*. **Infect Immun**., v.12, p. 1426–1438, 1975.

FIECK, A.; HURWITZ, I.; KANG, A. S.; DURVASULA, R. *Trypanosoma cruzi*: synergistic cytotoxicity of multiple amphipathic anti-microbial peptides to *T. cruzi* and potential bacterial hosts. **Exp. Parasitol**., v. 125, p. 342–347, 2010.

FIORE, A. E.; FRY, A.; SHAY, D.; GUBAREVA, L.; BRESEE, J. S.; UYEKI, T. M.; Antiviral Agents for the Treatment and Chemoprophylaxis of Influenza. **MMWR Recomm. Rep.**, v. 60, n. 1, 2011.

FONSECA F. Acidentes por lagartas. In: _____ **Animais peçonhentos**. São Paulo: Instituto Butantan, 1950. p. 284-296.

FRAIHA, H.; BALLARINI, A. J.; LEÃO, R. N. Q.; COSTA JR, D.; DIAS, L. B. Síndrome hemorrágica por contato com larvas de mariposa (Lepidoptera, Saturniidae). In: **Instituto Evandro Chagas - 50 anos de contribuição às ciências biológicas e à medicina tropical.** Belém: Fundação Serviços de Saúde Pública, 1986. v. 2, p. 811-818.

FUTAHASHI, R.; SHIRATAKI, H.; NARITA, T.; MITA, K.; FUJIWARA, H. Comprehensive microarray-based analysis for stage-specific larval camouflage pattern-associated genes in the swallowtail butterfly, Papilio xuthus. **BMC Biol**., v. 31, p. 10-46, 2012.

GANDHE, A. S.; ARUNKUMAR, K. P.; JOHN, S. H.; NAGARAJU, J. Analysis of bacteria-challenged wild silkmoth, *Antheraea mylitta* (lepidoptera) transcriptome reveals potential immune genes. **BMC Genomics**, n. 7, v. 184, p. 1-10, 2006.

GANZ, T.; LEHRER, R. I. Defensins. **Curr. Opin. Immunol**., v. 6, p. 584–589, 1994.

GENE ONTOLOGY CONSORTIUM. The Gene Ontology (GO) database and informatics resource. **Nucleic Acids Research**, v. 32, 2004.

GHAG, S. B.; SHEKHAWAT, U. K.; GANAPATHI, T. R. Petunia floral defensins with unique prodomains as novel candidates for development of fusarium wilt resistance in transgenic banana plants. **PLoS One,** v. 7, p. 53-57, 2012.

GOTO, A.; MATSUSHITA, K.; GESELLCHEN, V.; CHAMY, L.; KUTTENKEULER, D.; TAKEUCHI, O.; HOFFMANN, J. A.; AKIRA, S.; BOUTROS, M.; REICHHART, J. Akirins, highly conserved nuclear proteins, required for NF-κB dependent gene expression in *Drosophila* and mice. **Nat. Immunol.**, v. 9, n. 1, p. 97–104, 2008.

GRECO, K. N.; MENDONÇA, R. M.; MORAES, R. H.; MANCINI, D. A.; MENDONÇA, R. Z. Antiviral activity of the hemolymph of *Lonomia obliqua* (Lepidoptera: Saturniidae). **Antiviral Res.**, v. 84, n. 1, p. 84-90, 2009.

GRIMALDI, D.; ENGEL, M. S. **Evolution of the insects.** Cambridge University Press, 2005. 755p.

GU, J.; HUANG, L. X.; SHEN, Y.; HUANG, L. H.; FENG, Q. L. Hsp70 and small Hsps are the major heat shock protein members involved in midgut metamorphosis in the common cutworm, *Spodoptera litura*. **Insect Mol. Biol.**, v. 21, n. 5, p. 535-543, 2012.

GUSMÃO, H. H.; FORATINI, O. P.; ROTBERG, A. Dermatite provocada por lepidóptero do gênero *Hylesia*. São Paulo: **Revista do Instituto de Medicina Tropical**, v. 3, p. 114-120, 1961.

GUŽVIĆ, M. The History of DNA Sequencing/ISTORIJAT SEKVENCIRANJA DNK. **Journal of Medical Biochemistry**, v. 32, p. 301-312, 2013.

HADDAD JR, V.; CARDOSO J. L. C. Erucismo e Lepidopterismo. In: CARDOSO J. L. C.; WEN, F. H.; FRANÇA, F. O. S.; MALAQUE, C. M. S.; HADDAD JR, V. **Animais peçonhentos no Brasil: biologia, clínica e terapêutica dos acidentes**. São Paulo: Sarvier, 2003. p. 236-239.

HAMBLETON, E. J.; FORBES, W. T. M. Uma lista de Lepidoptera (Heterocera) do Estado de Minas Gerais. **Archivos do Instituto Biológico**, v. 6, p. 213-256, 1935.

HARA, S.; YAMAKAWA, M. A novel antibacterial peptide family isolated from the silkworm, *Bombyx mori.* **Biochem. J.**, v. 310, p. 651-656, 1995.

HEMMI, H.; ISHIBASHI, J.; HARA, S.; YAMAKAWA, M. Solution structure of moricin, an antibacterial peptide, isolated from the silkworm *Bombyx mori*. **FEBS Lett**., v. 518, p. 33–38, 2002.

HEPPNER, J. B. Faunal regions and the diversity of Lepidoptera. **Tropical Lepidoptera**, v. 2, p. 1-85, 1991.

HIRAKI, A.; YUKAWA, M.; KIM, J.; UEDA, S. Antiviral substance from silkworm faeces: purification and its chemical characterization. **Biol. Pharm. Bull**., v. 20, n. 5, p. 547-555, 1997.

HIRUMA, K.; RIDDIFORD, L. M. The molecular mechanisms of cuticular melanization: the ecdysone cascade leading to dopa decarboxylase expression in *Manduca sexta*. **Insect Biochem. Mol. Biol.,** v. 39, p. 245–253, 2009.

HOLLAND, D. L.; ADAMS, D. P. "Puss caterpillar" envenomation: a report from North Carolina. **Wilderness & Environmental Medicine**, v. 9, p. 213-216, 1998.

HOPKINS, T. L.; KRCHMA, L. J.; AHMAD, S. A.; KRAMER, K. J. Pupal cuticle proteins of *Manduca sexta*: characterization and profiles during sclerotization. **Insect Biochem. Mol. Biol.,** v. 30, p. 19–27, 2000.

HOPP, W. 1934. **Megalopygidae.** In: SEITZ, A. The macrolepidoptera of the world. Stuttgart: Alfred Keman. [Plates published in 1935]. p. 1071-1101.

HUBERMAN, A.; AGUILAR, M. B.; NAVARRO-QUIROGA, I.; RAMOS, L.; FERNANDEZ, I.; WHITE, F. M.; HUNT, D. F.; SHABANOWITZ, J. A hyperglycemic peptide hormone from the Caribbean shrimp Penaeus (litopenaeus) schmitti. **Peptides**, v. 21, p. 331-338, 2000.

HULTMARK, D.; STEINER, H.; RASMUSON, T.; BOMAN H. G. Insect immunity. Purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of *Hyalophora cecropia*. **Eur. J. Biochem**., v. 106, p. 7–16, 1980.

HUSE, S. M.; HUBER, J. A.; MORRISON, H. G.; SOGIN, M. L.; WELCH, D. M. Accuracy and quality of massively parallel DNA pyrosequencing. **Genome Biol**., v. 8, 2007.

HWANG, B.; HWANG, J. S.; LEE, J.; KIM, J. K.; KIM, S. R.; KIM, Y.; LEE, D. G. Induction of yeast apoptosis by an antimicrobial peptide, Papiliocin. **Biochem. Biophys. Res. Commun**., v. 408, p. 89–93, 2011.

HWANG, J.; KIM, Y. RNA Interference of an antimicrobial peptide, gloverin, of the beet armyworm, *Spodoptera exigua*, enhances susceptibility to Bacillus thuringiensis. **J. Invertebr. Pathol**., v. 108, p. 194–200, 2011.

ISERHARD, C. A. A.; ROMANOWSKI, H. P. Lista de espécies de borboletas (Lepidoptera, Papilionoidea e Hesperioidea) da região do rio Maquiné, Rio Grande do Sul, Brasil. **Revista Brasileira de Zoologia**, v. 21, p. 649–662, 2004.

JAN, P. S.; HUANG, H. Y; CHEN, H. M. Expression of a synthesized gene encoding cationic peptide cecropin B in transgenic tomato plants protects against bacterial diseases. **Appl. Environ. Microbiol**., v. 76, p. 769–775, 2010.

JAYNES, J. M.; BURTON, C. A.; BARR, S. B.; JEFFERS, G. W.; JULIAN, G. R.; WHITE, K. L.; ENRIGHT, F. M.; KLEI, T. R.; LAINE, R. A. In vitro cytocidal effect of novel lytic peptides on *Plasmodium falciparum* and *Trypanosoma cruzi*. **FASEB J**., v. 2, p. 2878–2883, 1988.

JHA, S.; CHATTOO, B. B. Expression of a plant defensin in rice confers resistance to fungal phytopathogens. **Transgenic. Res**., v. 19, p. 373–384, 2010.

KALYUZHNAYA, M. G.; LAPIDUS, A.; IVANOVA, N.; COPELAND, A. C.; MCHARDY, A. C.; SZETO, E.; SALAMOV, A.; GRIGORIEV, I. V.; SUCIU, D.; LEVINE, S. R.; MARKOWITZ, V. M.; RIGOUTSOS, I.; TRINGE, S. G.; BRUCE, D. C.; RICHARDSON, P. M.; LIDSTROM, M. E.; CHISTOSERDOVA, L. High-resolution metagenomics targets specific functional types in complex microbial communities. **Nature Biotechnology**, v. 26, p. 1029 – 1034, 2008.

KANEHISA, M.; GOTO, S.; KAWASHIMA, S.; NAKAYA, A. The KEGG databases at GenomeNet. **Nucleic Acids Res.**, v. 30, p. 42–46, 2002.

KANEKO, Y.; TANAKA, H.; ISHIBASHI, J.; IWASAKI, T.; YAMAKAWA, M. Gene expression of a novel defensin antimicrobial peptide in the silkworm, *Bombyx mori*. **Biosci. Biotechnol. Biochem**., v. 72, p. 2353–2361, 2008.

KANOST, M. R.; JIANG, H.; YU, X. Q. Innate imune responses of a lepidopteran insect, *Manduca sexta*. **Immunol. Rev**., v. 198, p. 97–105, 2004.

KAUR, J.; THOKALA, M.; ROBERT-SEILANIANTZ, A.; ZHAO, P.; PEYRET, H.; BERG, H.; PANDEY, S.; JONES, J.; SHAH, D. Subcellular targeting of an evolutionarily conserved plant defensin MtDef4.2 determines the outcome of plant-pathogen interaction in transgenic *Arabidopsis*. **Mol Plant Pathol**. v. 13, p. 1032–1046, 2012.

KAWAMOTO, F.; KUMADA, N. Biology and venoms of Lepidoptera. **Handbook of Natural Toxins**, New York, v. 2, p. 291-330, 1984.

KAWAOKA, S.; KATSUMA, S.; DAIMON, T.; ISONO, R.; OMURO, N.; MITA, K.; SHIMADA, T. Functional analysis of four Gloverin-like genes in the silkworm, *Bombyx mori*. **Arch. Insect Biochem. Physiol**., v. 67, p. 87–96, 2008.

KIM. S. R.; HONG. M. Y.; PARK, S. W.; CHOI, K. H.; YUN, E. Y.; GOO, T. W.; KANG, S. W.; SUH, H. J.; KIM, I.; HWANG, J. S. Characterization and cDNA cloning of a cecropin-like antimicrobial peptide, papiliocin, from the swallowtail butterfly, *Papilio xuthus*. **Molecules and Cells,** v. 29, p. 419–423, 2010.

KITCHING, R. L.; ORR, A. G.; THALIB, L.; MITCHELL, H.; HOPKINS, S.; GRAHAM, A. W. Moth assemblages as indicator of environmental quality in remants of upland Australian rain forest. **Jornal of Applied Ecology,** v. 37, p. 284-297, 2000.

KOCKUM, K.; FAYE, I.; HOFSTEN, P. V.; LEE, J. Y.; XANTHOPOULOS, K. G.; BOMAN, H. G. Insect immunity. Isolation and sequence of two cDNA clones corresponding to acidic and basic attacins from *Hyalophora cecropia*. **EMBO J**., v. 3, p. 2071–2075, 1984.

KOKOZA, V.; AHMED, A.; WOON-SHIN, S.; OKAFOR, N.; ZOU, Z.; RAIKHEL, A. S. Blocking of *Plasmodium* transmission by cooperative action of Cecropin A and Defensin A in transgenic *Aedes aegypti* mosquitoes. **Proc. Natl. Acad. Sci. U S A**, v. 107, p. 8111–8116, 2010.

KRISTENSEN, N. P.; SCOBLE, M. J.; KARSHOLT, O. K. Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. **Zootaxa,** v. 1668, p. 699-747, 2007.

KRISTENSEN, N. P.; SKALSKI, A. W. **Phylogeny and Palaeontology**. Pp. 7-25. In: KRISTENSEN, N. P. (Ed.). Handbook of zoology, vol. 4, Arthropoda: Insecta - Lepidoptera, moths and butterflies, vol. 1: Evolution, Systematics, and Biogeography, part 35. Walter de Gruyter, Berlin, NY., 1999. 487p.

KRISTENSEN, N. P. **Lepidoptera:** moths and Butterflies 1. Evolution, Systematics and Biogeography. *Handbuch der Zoologie/Handbook of Zoology*. Walter de Gruyter, Berlin & New York. 1998, 491 p.

KROGH, A.; LARSSON, B.; VON HEIJNE, G.; SONNHAMMER, E. L. Prediction transmembrane protein topology with a hidden markov model: application to complete genomes. **J. Mol. Biol.,** n. 305, v. 3. p. 567–580, 2001.

KWON, Y. M.; KIM, H. J.; KIM, Y. I.; KANG, Y. J.; LEE, I. H.; JIN, B. R.; HAN, Y. S.; CHEON, H. M.; HA, N. G.; SEO, S. J. Comparative analysis of two attacin genes from *Hyphantria cunea*. **Comp. Biochem. Physiol. B Biochem. Mol. Biol**., v. 151, p. 213–220, 2008.

LAMBERT, J.; KEPPI, E.; DIMARCQ, J. L.; WICKER, C.; REICHHART, J. M.; DUNBAR. B.; LEPAGE, P.; VAN DORSSELAER, A.; HOFFMANN, J.; FOTHERGILL, J.; HOFFMANN, D. Insect immunity: isolation from immune blood of the dipteran *Phormia terranovae* of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. **Proc. Natl. Acad. Sci. U S A,** v. 86, p. 262–266, 1989.

LAMY, M.; PASTUREAUD, M. H.; NOVAK, F.; DUCOMBS, G.; VINCEDEAU, P.; MALEVILLE, J.; TEXIER, L. Thaumetopoein: an urticating protein from the hairs and integument of the pine processionary caterpillar (Thaumetopoeapityocampa Schiff., Lepidoptera, Thaumetopoeidae). **Toxicon**, v. 24, p. 347-356, 1986.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C.; BALDWIN, J.; et al. Initial seqüencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, 2001.

LANDON, C.; MEUDAL, H.; BOULANGER, N.; BULET, P.; VOVELLE, F. Solution structures of stomoxyn and spinigerin, two insect antimicrobial peptides with an alpha-helical conformation. **Biopolymers,** v. 81, p. 92–103, 2006.

LEE, W. J.; LEE, J. D.; KRAVCHENKO, V. V.; ULEVITCH, R. J.; BREY, P. T. Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. **Proc. Natl. Acad. Sci. U S A**, v. 93, n. 15, p. 7888-7893, 1996.

LEMAITRE, B.; HOFFMANN, J. The host defense of *Drosophila melanogaster*. **Annu. Rev. Immunol.**, v. 25, p. 697–743, 2007.

LEPESQUEUR, C. **Megalopygidae (Lepidoptera, Zygaenoidea): biologia, diversidade e biogeografia.** 2012. 105 f. Tese (Doutorado em Ecologia) - Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, 2012.

LEWIS, S. E. Gene ontology: looking backwards and forwards. **Genome Biology**, v. 6, n. 1, p. 1-4, 2004.

LI, Z. Q.; MERRIFIELD, R. B.; BOMAN, I. A.; BOMAN, H. G. Effects on electrophoretic mobility and antibacterial spectrum of removal of two residues from synthetic sarcotoxin IA and addition of the same residues to cecropin. **B. FEBS Lett**., v. 231, p. 299–302, 1988.

LI, Z.; ZHOU, M.; ZHANG, Z.; REN, L.; DU, L.; ZHANG, B.; XU, H.; XIN, Z. Expression of a radish defensin in transgenic wheat confers increased resistance to *Fusarium graminearum* and *Rhizoctonia cerealis*. **Funct. Integr. Genom**., v. 11, p. 63–70, 2011.

LORDELLO, L. G. E. Notas sobre a evolução da Sassurana-*Megalopyge lanata* (Stoll, 1780). (Lep., Heterocera, Megalopygidae). **Anais da Escola Superior de Agricultura Luiz de Queiroz**, v. 8, p. 205-218, 1951.

LOWENBERGER, C. Innate immune response of *Aedes aegypti.* **Insect Bioch. Mol. Biol.**, v. 31, p. 219–229, 2001.

LOWENBERGER, C.; CHARLET, M.; VIZIOLI, J.; KAMAL, S.; RICHMAN, A.; CHRISTENSEN, B. M.; BULET, P. Antimicrobial activity spectrum, cDNA cloning, and mRNA expression of a newly isolated member of the cecropin family from the mosquito vector *Aedes aegypti.* **J. Biol. Chem**., v. 274, p. 20092-20097, 1999.

LUNDSTROM, A.; LIU, G.; KANG, D.; BERZINS, K.; STEINER, H. *Trichoplusia ni* gloverin, an inducible immune gene encoding an antibacterial insect protein. **Insect Biochem. Mol. Biol**., v. 32, p. 795–801, 2002.

MACKINTOSH, J. A.; GOOLEY, A. A.; KARUSO, P. H.; BEATTIE, A. J.; JARDINE, D. R.; VEAL, D. A. A gloverin-like antibacterial protein is synthesized in *Helicoverpa armigera* following bacterial challenge. **Dev. Comp. Immunol**., v. 22, p. 387–399, 1998.

MALLET, J. Subspecies, semispecies, superspecies. In: LEVIN, S. **Encyclopedia of biodiversity.** Oxford: Elsevier, 2007. v. 5, p. 523-526.

MARCHLER-BAUER, A.; DERBYSHIRE, M. K.; GONZALES, N. R.; LU, S.; CHITSAZ, F.; GEER, L. Y.; GEER, R. C.; HE, J.; GWADZ, M.; HURWITZ, D. I.; LANCZYCKI, C. J.; LU, F.; MARCHLER, G. H.; SONG, J. S.; THANKI, N.; WANG, Z.; YAMASHITA, R. A.; ZHANG, D.; ZHENG, C.; BRYANT, S. H. CDD: NCBI's conserved domain database. **Nucleic Acids Research**, v. 43, p. 222–226, 2015.

MARGULIES, M.; EGHOLM, M.; ALTMAN, W. E.; ATTIYA, S.; BADER, J. S.; BEMBEN, L. A. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437, p. 376–380, 2005.

MATSUYAMA, K.; NATORI, S. Molecular cloning of cDNA for sapecin and unique expression of the sapecin gene during the development of *Sarcophaga peregrina*. **J. Biol. Chem**., v. 263, p. 17117–17121, 1988.

McGOVERN, J.P.; BARKIN, G.D.; MCELHENNEY, T.R.; WENDE, R. *Megalopyge opercularis*: observations of its life history, natural history of its sting in man, and report of an epidemic. **The Journal of the American Medical Association**, v. 175, p. 1155-1158, 1961.

MILLER, R. T.; CHRISTOFFELS, A. G.; GOPALAKRISHNAN, C.; BURKE, J.; PTITSYN, A. A.; BROVEAK, T. R.; HIDE, W. A. A Compreensive approach to clustering of expressed h gene sequence: The sequence tag alignment and consensus knowledge base. **Genome Research**, v. 9, p. 1143-1155, 1999.

MITSUHARA, I.; MATSUFURU, H.; OHSHIMA, M.; KAKU, H.; NAKAJIMA, Y.; MURAI, N.; NATORI, S.; OHASHI, Y. Induced expression of sarcotoxin IA enhanced host resistance against both bacterial and fungal pathogens in transgenic tobacco. **MPMI**, v. 13, p. 860–868, 2000.

MOELLERING, R. C. JR. Vancomycin-resistant enterococci. **Clin. Infect Dis**., v. 26, 1196-1199, 1998.

MONTAÑO, A. M.; TSUJINO, F.; TAKAHATA, N.; SATTA, Y. Evolutionary origin of peptidoglycan recognition proteins in vertebrate innate immune system. **BMC Evolutionary Biology**, v. 11, n. 79, p. 1-10, 2011.

MOORE, A. J.; BEAZLEY, W. D.; BIBBY, M. C.; DEVINE, D. A. Antimicrobial activity of cecropins. **J. Antimicrob. Chemother**., v. 37, p. 1077–1089, 1996.

MORAES, R. H. P. Lepidópteros de importância médica. In: CARDOSO, J. L. C.; FRANÇA, F. O. S.; WEN, F. H.; MÁLAQUE, C. M. S.; HADDAD JR, V. **Animais peçonhentos no Brasil:** Biologia, clínica e terapêutica dos acidentes. São Paulo, SP: Sarvier, 2003. p. 211-219.

MORALES, M. M. Desafio em relação a terapias avançadas: terapia utilizando células-tronco, terapias gênicas e nanotecnologia aplicada à saúde. **Anais da 58ª Reunião Anual da SBPC**. Florianópolis, 2006.

MUDADO, M. A. **Uso da Base de Dados Secundária KOG como Ferramenta para Caracterização de Expressão Gênica e Mineração de Dados em Projetos Transcriptoma.** 2007. 134 f. Tese (Doutorado em Bioinformática) – Instituto de Ciências Biológicas, Universidade de São Paulo, São Paulo, 2007.

NACIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **BLAST**. Disponível em: <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/query_tutorial.html>. Acesso em: 14 jun. 2014.

NANBU, R.; NAKAJIMA, Y.; ANDO, K.; NATORI, S. Novel feature of expression of the sarcotoxin IA gene in development of *Sarcophaga peregrina*. **Biochem. Biophys. Res. Commun**., v. 150, p. 540–544, 1988.

NATION, J. L. **Insect physiology and biochemistry**. Boca Raton: CRC Press, 2002. 553 p.

NATURE <http://www.nature.com/scitable/knowledge/library/factors-affecting-global-climate-17079163.> Acesso em 13 nov. 2015.

NETO, E. M. C. **Etnoentomologia no povoado de pedra branca, município de Santa Terezinha, Bahia. Um estudo de caso das interações seres humanos/insetos.** 2003. 251 f. Tese (Doutorado em Ciências) – Centro de Ciências Biológicas e da Saúde, São Carlos, São Paulo, 2003.

NEUSTATER, B. R.; STOLLMAN, N.H.; MANTEN, H. D. Sting of the puss caterpillar: an unusual cause of acute abdominal pain**. Southern Medical Journal**, v. 89, p. 826-827, 1996.

NIELSEN, H.; BRUNAK, S.; VON HEIJNE, G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. **Protein Eng.**, v. 1, n. 12, p. 3-9, 1999.

NIELSEN, H.; ENGELBRECHT, J.; BRUNAK, S.; VON HEIJNE, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. **Protein Eng.**, v. 10, n. 1, p. 1-6, 1997.

NIELSEN, H.; KROGH, A. Prediction of signal peptides and signal anchors by a hidden Markov model. In: Proceedings of the sixth international conference on intelligent systems for molecular biology. California: **AAAI Press**, p. 122-130, 1998.

NTUI, V. O.; THIRUKKUMARAN, G.; AZADI, P.;, KHAN, R. S.; NAKAMURA, I.; MII, M. Stable integration and expression of wasabi defensin gene in "Egusi" melon (*Colocynthis citrullus L.*) confers resistance to Fusarium wilt and Alternaria leaf spot. **Plant Cell Rep**., v. 29, p. 943–954, 2010.

NYREN, P.; PETTERSSON, B.; UHLEN, M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. **Anal. Biochem**., v. 208, p. 171–175, 1993.

OCHIAI, M.; ASHIDA, M. A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, *Bombyx mori*. **J. Biol. Chem.,** v. 274, n. 17, p. 1854-1858, 1999.

OIZUMI, Y.; HEMMI, H.; MINAMI, M.; ASAOKA, A.; YAMAKAWA, M. Isolation, gene expression and solution structure of a novel moricin analogue, antibacterial peptide from a lepidopteran insect, *Spodoptera litura*. **Biochim. Biophys. Acta**, v. 1752, p. 83–92, 2005.

OKADA, M.; NATORI, S. Primary structure of sarcotoxin I, an antibacterial protein induced in the hemolymph of *Sarcophaga peregrina* (flesh fly) larvae. **J. Biol. Chem**., v. 260, p. 7174–7177, 1985.

OLIVEIRA, G.; JOHNSTON, D. A. Mining the schistosome DNA sequence database. **Trends in Parasitology**, v. 17, p. 501-503, 2001.

OTVOS, L. J. R. Antibacterial peptides isolated from insects. **J. Pept. Sci**., v. 6, p. 497–511, 2000.

OURTH, D. D. Antiviral activity against human immunodeficiency virus-1 in vitro by myristoylated-peptide from *Heliothis virescens*. **Biochem. Biophys. Res. Commun.**, v. 320, n. 1, p. 190-196, 2004.

PAPANICOLAOU, A.; GEBAUER-JUNG, S.; BLAXTER, M. L.; MCMILLAN, W. O.; JIGGINS, C. D. ButterflyBase: a platform for lepidopteran genomics. **Nucleic Acids Res.,** v. 36, p 582-587, 2008.

PAUCHET, Y.; WILKINSON, P.; VOGEL, H.; NELSON, D. R.; REYNOLDS, S. E.; HECKEL, D. G.; FRENCH-CONSTANT, R. H. Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxificationand defence. **Insect Molecular Biology,** n. 19, v. 1, p. 61–75, 2010.

PEREIRA, P. R. V. S.; HALFELD-VIEIRA, B. A.; NECHET, K. L. *Norape* sp. (Lepidoptera: Megalopygidae): lagarta desfolhadora em plantios comerciais de *Acacia mangium*. **Boa Vista: Embrapa Roraima**, Rr. p, 6. (Comunicado Técnico, 20), 2003.

PEREIRA, R. A. S.; PRADO, A. P.; KJELLBERG, F. Gynandromorphism in pollinating fig wasps (Hymenoptera: Agaonidae). **Entomological News**, v. 114, p. 152-155, 2003.

PHOENIX, D. A.; DENNISON, S. R.; HARRIS, F. **Antimicrobial Peptides**. In: Antimicrobial peptides: their history, evolution, and functional promiscuity. Wiley-VCH Verlag GmbH & Co. KGaA, 2013, p. 1–37.

POIVET, E.; GALLOT, A.; MONTAGNE, N.; GLASER, N.; LEGEAI, F.; JACQUIN-JOLY, E. A Comparison of the olfactory gene repertoires of adults and larvae in the Noctuid Moth *Spodoptera littoralis*. **PLoS One**, n. 8, v. 4, p. 1-15, 2013.

POP, M.; SALZBERG, S. L. Bioinformatics challenges of new sequencing technology. **Trends in Genetics**, v. 3, p. 142-149, 2008.

POPHAM, H. J.; SHELBY, K. S.; BRANDT, S. L.; COUDRON, T. A. Potent virucidal activity in larval *Heliothis virescens* plasma against *Helicoverpa zea* single capsid nucleopolyhedrovirus. **J. Gen. Virol.**, v. 85, n. 8, p. 2255-2561, 2004.

PRETZEL, J.; MOHRING, F.; RAHLFS, S.; BECKER, K. Antiparasitic peptides. **Adv. Biochem. Eng. Biotechnol**., v. 135, p. 157–192, 2013.

PROSDOCIMI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A. F.; REIS, A. N.; JUNQUEIRA, A. C. M.; SANTOS, A. C. F.; NHANI-JÚNIOR, A.; WUST, C. I.; CAMARGO-FILHO, F.; KESSEDJIAN, J. L.; PETRETSKI, J. H.; CAMARGO, L. P.; FERREIRA, R. G. M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. S.; FOLGUERASFLATSCHART, A. V. Bioinformática: manual do usuário. **Biotec. Ci. Des.**, v. 29, p. 18-31, 2002.

RAVI, C.; JEYASHREE, A.; RENUKA DEVI, K. Review article antimicrobial peptides from insects: an overview. **Research in Biotechnology**, v. 2, n. 5, p. 1-7, 2011.

RAHNAMAEIAN, M.; VILCINSKAS, A. Defense gene expression is potentiated in transgenic barley expressing antifungal peptide Metchnikowin throughout powdery mildew challenge. **J. Plant Res**., v. 125, p. 115–124, 2012.

RAO, X. J.; XU, X. X.; YU, X. Q. Functional analysis of two lebocin related proteins from *Manduca sexta*. **Insect Biochem. Mol. Biol**., v. 42, p. 231–239, 2012.

RIDDIFORD, L. M.; HIRUMA, K.; ZHOU, X.; NELSON, C. A. Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster*. **Insect Biochem. Mol. Biol.,** v. 33, p. 1327–1338, 2003.

ROBERTSON, M.; POSTLETHWAIT, J. H. The humoral antibacterial response of *Drosophila* adults. **Dev. Comp. Immunol**., v. 10, p. 167–179, 1986.

ROBERTSON, H. M.; MARTOS, R.; SEARS, C. R.; TODRES, E. Z.; WALDEN, K. K.; NARDI, J. B. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. **Insect Mol. Biol.,** v. 8, p. 501–518, 1999.

ROMANELLI, D.; CASATI, B.; FRANZETTI, E.; TETTAMANTI, G. Review Article. A Molecular View of Autophagy in Lepidoptera. **BioMed Res. Int.,** v. 2014, p. 1-11, 2014.

RONAGHI, M.; KARAMOHAMED, S.; PETTERSSON, B.; UHLEN, M.; NYREN, P. Real-time DNA sequencing using detection of pyrophosphate release. **Anal. Biochem.**, v. 242, p. 84–89, 1996.

RONAGHI, M.; UHLEN, M.; NYREN, P. A sequencing method based on real-time pyrophosphate. **Science**, v. 281, p. 363–365, 1998.

ROSEN, T. Caterpillar Dermatitis. **Dermatol. Clin**., v. 8, p. 245-252, 1990.

RUBIO, G. B. G. Vigilância epiemiológica de distribuição da lagarta *Lonomia obliqua* Walke, 1855, no Estado do Paraná, Brasil**.** Rio de Janeiro. **Caderno de Saúde Pública**, v. 17, p. 1036, 2001.

SACCARO JR, N. L. **Desafios da bioprospecção no Brasil.** Brasília: Ipea, 2011.

SADD, B. M.; SCHMID-HEMPEL, P. Insect immunity shows specificity in protection upon secondary pathogen exposure. **Curr. Biol**., v. 16, p. 1206–1210, 2006.

SAMAKOVLIS, C.; KIMBRELL, D. A.; KYLSTEN, P.; ENGSTROM, A.; HULTMARK, D. The immune response in *Drosophila*: pattern of cecropin expression and biological activity. **EMBO J**., v. 9, p. 2969–2976, 1991.

SCHERRER, S.; DINIZ, I. R.; MORAIS, H. C. Climate and host plant characteristics effects on lepidopteran caterpillar abundance on *Miconia ferruginata* DC. and *Miconia pohliana* Cogn (Melastomataceae). **Brazilian Journal of Biology**, v. 70, p. 103-109, 2010.

SCHUHMANN, B.; SEITZ, V.; VILCINSKAS, A.; PODSIADLOWSKI, L. Cloning and expression of gallerimycin, an antifungal peptide expressed in immune response of greater wax moth larvae, *Galleria mellonella*. **Arch Insect Biochem Physiol.**, v. 53, n.3, p. 125-133, 2003.

SCHUSTER, S. C. Next-generation sequencing transforms today's biology. **Nature Methods**, v. 5, p. 16-18, 2008.

SCOBLE, M. J. The Lepidoptera form, function and diversity. New York, NY: **Oxford University Press**, 1995. p. 404.

SCOCCHI, M.; TOSSI, A.; GENNARO, R. Proline-rich antimicrobial peptides: converging to a non-lytic mechanism of action. **Cell Mol. Life Sci**., v. 68, p. 2317‑2330, 2011.

SEITZ, V.; CLERMONT, A.; WEDDE, M.; HUMMEL, M.; VILCINSKAS, A.; SCHLATTERER, K.; PODSIADLOWSKI, L. Identification of immunorelevant genes from greater wax moth (*Galleria mellonella*) by a subtractive hybridization approach. **Dev. Comp Immunol**., v. 27, p. 207‑215, 2003.

SEUFI, A. M.; HAFEZ, E. E.; GALAL, F. H. Identification, phylogenetic analysis and expression profile of an anionic insect defensin gene, with antibacterial activity, from bacterial-challenged cotton leafworm, *Spodoptera littoralis*. **BMC Mol. Biol**., v. 12, p. 47, 2011.

SHIELDS, V. D.; HILDEBRAND, J. G. Recent advances in insect olfaction, specifically regarding the morphology and sensory physiology of antennal sensilla of the female sphinx moth *Manduca sexta*. **Microsc Res Tech**., v. 55, p. 307–329, 2001.

SILVA, C. C. A. Aspectos do sistema imunológico dos insetos. **Biotecnologia Ciência & Desenvolvimento**, n. 24, p. 68-72, 2002.

SILVA, J. L.; BARBOSA, J. F.; BRAVO, J. P.; SOUZA, E. M.; HUERGO, L. F.; PEDROSA, F. O.; ESTEVES, E.; DAFFRE, S.; FERNANDEZ, M. A. Induction of a gloverin-like antimicrobial polypeptide in the sugarcane borer *Diatraea saccharalis* challenged by septic injury. **Braz. J. Med. Biol. Res**., v. 43, p. 431–436, 2010.

SILVA, N. A. P.; DUARTE, M.; DINIZ, I. R.; MORAIS, H. C. Host plants of Lycaenidae on inflorescences in the central Brazilian cerrado. **The Journal of Research on the Lepidoptera,** v. 44, p. 95-105, 2011.

SOARES, J. W.; MELLO, C. M. Antimicrobial Peptides: A Review of How Peptide Structure Impacts Antimicrobial Activity. **Proceedings of Spie,** v. 5271, p. 20 – 27, 2004.

SPECHT, A.; CORSEUIL, E.; FORMENTINI, A.C.; PRESTES, A.S. Lepidópteros de importância médica ocorrentes no Rio Grande do Sul. I. Megalopygidae. **Biociências,** v. 12, p. 173-179, 2004.

SPECHT, A.; CORSEUIL, E. Megalopygidae. In: SPECHT, A.; CORSEUIL, E.; ABELLA, H. B. (Org.). **Lepidópteros de importância médica**: principais espécies no Rio Grande do Sul. Pelotas, RS: Useb, 2008. p. 11-41.

STEIN, L. Genome annotation: from sequence to biology. **Nat. Ver. Genet**., v. 2, n. 7, p. 493-503, 2001.

STEINER, H.; HULTMARK, D.; ENGSTROÈM, A.; BENNICH, H.; BOMAN, H. G. Sequence and specicity of two antibacterial proteins involved in insect immunity. **Nature,** v. 292, p. 246-248, 1981.

STIPETIC, M. E.; ROSEN, P. B.; BORYS, D. J.A retrospective analysis of 96" asp. *Megalopyge opercularis*. **Clinical Toxicology**, v. 37, p. 457-462, 1996.

STOESSER, G.; BAKER, W.; VAN DEN BROEK, A.; CAMON, E.; GARCIA-PASTOR, M.; KANZ, C.; KULIKOVA, T.; LEINONEN, R.; LIN, Q.; LOMBARD, V.; LOPEZ, R.; REDASCHI, N.; STOEHR, P.; TULI, M. A.; TZOUVARA, K.; VAUGHAN, R. The EMBL nucleotide sequence database. **Nucleic Acid. Res.**, v. 30, n. 1, p. 21-26, 2002.

SUN, S. C.; LINDSTROM, I.; LEE, J. Y.; FAYE, I. Structure and expression of the attacin genes in *Hyalophora cecropia*. **Eur. J. Biochem**., v. 196, p. 247–254, 1991.

SWATHI ANURADHA, T.; DIVYA, K.; JAMI, S. K.; KIRTI, P. B. Transgenic tobacco and peanut plants expressing a mustard defensin show resistance to fungal pathogens. **Plant Cell Rep**., v. 27, p. 1777–1786, 2008.

TANIAI, K.; FURUKAWA, S.; SHONO, T.; YAMAKAWA, M. Elicitors triggering the simultaneous gene expression of antibacterial proteins of the silkworm, *Bombyx mori*. **Biochem. Biophys. Res. Commun**., v. 226, p. 783–790, 1996.

TAPADIA, M. G.; VERMA, P. Immune response and antimicrobial peptides expression in malpighian tubules of Drosophila melanogaster is under developmental regulation, **PLOS One**, v. 7, n. 7, p, 1-12, 2012.

TATENO, Y.; IMANISHI, T.; MIYAZAKI, S.; FUKAMI-KOBAYASHI, K.; SAITOU, N.; SUGAWARA, H.; GOJOBORI, T. The DNA Data Bank of Japan (DDBJ) for genome scale research in life sciences. **Nucleic Acid Res.**, v. 30, n. 1, p. 27-30, 2002.

THE ANTIMICROBIAL PEPTIDE DATABASE. Disponível em: <http://aps.unmc.edu/AP/main.php> Acesso em: 14 jan. 2016.

TIAN, C.; GAO, B.; FANG, Q.; YE, G.; ZHU, S. Antimicrobial peptide-like genes in *Nasonia vitripennis*: a genomic perspective. **BMC Genomics**, v. 11, n. 187, p. 1-19, 2010.

VALLE, J. R.; PICARELLI, Z. P.; PRADO, J. L. Histamine Contend and pharmacological properties of crude extrats from setae of urticating caterpillar. **Arch. Int. Pharmacody Terer**, v. 98, p. 324-344, 1954.

VAN NIEUKERKEN, E. J.; KAILA, L.; KITCHING, I. J.; KRISTENSEN, N. P.; LEES, D. J.; MINET, J.; ZOLOTUHIN, V. V. Order Lepidoptera Linnaeus, 1758. **Zootaxa**, v. 3148, p. 212-221, 2011.

VENTER, J. C.; ADAMS, M. D.; MYERS, E. W.; LI, P. W.; MURAL, R. J.; SUTTON, G. G.; SMITH, H. O.; YANDELL, M. et al. The sequence of the human genome. **Science**, v. 291, n. 5507, p. 1304-1351, 2001.

VILLAS-BOAS, I. M.; GONCALVES-DE-ANDRADE, R. M.; PIDDE-QUEIROZ, G.; ASSAF, S. L. M. R.; PORTARO, F. C.; SANT'ANNA, O. A.; TAMBOURGI, D. V. Premolis semirufa (Walker, 1856) envenomation, disease affecting rubber tappers of the Amazon: searching for caterpillar bristles toxic components. **PLoS Neglected Tropical Diseases**, v. 6, p. 1531, 2012.

VINÓS, J.; DOMINGO, A.; MARCO, R.; CERVERA, M. Identification and characterization of *Drosophila melanogaster* paramyosin. **J. Mol. Biol.**, v. 220, n. 3, p. 687-700, 1991.

VITAL BRAZIL, O. Farmacologia do canal do sódio: modo de ação de toxinas**. Ciência e Cultura**, v. 38, p. 324-328, 1986.

VIZIOLI, J.; BULET, P.; CHARLET, M.; LOWENBERGER, C.; BLASS, C.; MULLER, H. M.; DIMOPOULOS, G.; HOFFMANN, J.; KAFATOS, F. C.; RICHMAN, A. Cloning and analysis of a cecropin gene from the malaria vector mosquito, *Anopheles gambiae*. **Insect Mol. Biol**., v. 9, p. 75–84, 2000.

VOELKERDING, K. V.; DAMES, S. A.; DURTSCHI, J. D. Next-generation sequencing: from basic research to diagnostics. **Clin Chem**, v. 55, n. 4, p. 641-658, 2009.

VOGEL, H.; ALTINCICEK, B.; GLÖCKNER, G.; VILCINSKAS, A. A comprehensive transcriptome and immunegene repertoire of the lepidopteran model host *Galleria mellonella*. **BMC Genomics,** n. 12, v. 308, p. 1-19, 2011.

VON HEIJNE, G. Principles of membrane protein assembly and structure. **Prog. Biophys. Mol. Biol**., v. 66, p. 113-139, 1997.

WACHINGER, M.; KLEINSCHMIDT, A.; WINDER, D.; VON PECHMANN, N.; LUDVIGSEN, A.; NEUMANN, M.; HOLLE, R.; SALMONS, B.; ERFLE, V.; BRACK-WERNER, R. Antimicrobial peptides melittin and cecropin inhibit replication of human immunodeficiency virus 1 by suppressing viral gene expression. **J. Gen. Virol**., n. 79, v. 4, p. 731–740, 1998.

WAHLBERG, J. M.; SPIESS, M. Multiple determinants direct the orientallon of signal-anchor proteins: The topogenic role of the hydrophobic signal domain. **J. Cell Biol.,** v. 137, p. 555-562, 1997.

WANG, Z.; WANG, G. APD: the Antimicrobial Peptide Database. **Nucleic Acids Research,** v. 32, p. 590-592, 2004.

WANG, G.; LI, X.; WANG, Z. APD3: the antimicrobial peptide database as a tool for research and education. **Nucleic Acids Research,** p. 44, 2016.

WANG, Q.; LIU, Y.; HE, H. J.; ZHAO, X. F.; WANG, J. X. Immune responses of *Helicoverpa armigera* to different kinds of pathogens. **BMC Immunol**., v. 11, p. 9, 2010.

WANG, Y. P.; LAI, R. Insect antimicrobial peptides: structures, properties and gene regulation. **Dongwuxue Yanjiu**, v. 31, v. 1, p. 27-34, 2010.

WARD, K. E.; LANDOLF, P. J. Influence of multiple matings on fecundity and longevity of female cabbage looper moths (Lepidoptera: Noctuidae). **Annals of the Entomological Society of America**, v. 88, p. 768-772, 1995.

WELLER, D. L.; CHURCH, D. M.; LASH, A. E.; LEIPE, D. D.; MADDEN, T. L.; PONTIUS, J. U.; SCHULER, G. D.; SCHRIML, L. M.; TATUSOVA, T. A.; WAGNER, L.; RAPP, B. A. Database resources of the National Center for Biotechnology information: 2002 update. **Nucleic Acid Res.**, v. 30, n. 1, p. 13-16, 2002.

WEN, F. H. **Ensaio clínico para avaliação da eficácia e segurança de um antiveneno específico no tratamento da síndrome hemorrágica causada por lagartas do Gênero Lonomia.** 2002. Tese (Doutorado em Ciências médicas) - Universidade Estadual de Campinas, São Paulo, 2002.

WESTBROOK, J.; FENG, Z.; JAIN, S.; BHAT, T. N.; THANKI, N.; RAVICHANDRAN, V.; GILLILAND, G. L.; BLUHM, W.; WEISSIG, H.; GREER, D. S.; BOURNE, P. E.; BERMAN, H. M. The Protein Data Bank: unifying the archive. **Nucleic. Acid. Res.**, v. 30, n. 1, p. 245-248, 2002.

WICKER. C.; REICHHART, J. M.; HOFFMANN, D.; HULTMARK, D.; SAMAKOVLIS, C.; HOFFMANN, J. A. Insect immunity. Characterization of a Drosophila cDNA encoding a novel member of the diptericin family of immune peptides. **J. Biol. Chem**., v. 265, p. 22493–22498, 1990.

WILMES, M.; CAMMUE, B. P.; SAHL, H. G.; THEVISSEN, K. Antibiotic activities of host defense peptides: more to it than lipid bilayer perturbation. **Nat. Prod. Rep**., v. 28, p. 1350–1358, 2011.

WINTERMEYER, W.; SAVELSBERGH, A.; SEMENKOV, Y. P.; KATUNIN, V. I.; RODNINA, M. V. Mechanism of elongation factor G function in tRNA translocation on the ribosome. **Cold Spring Harb Symp Quant Biol.**, v. 66, p. 449-458, 2001.

XU, X. X.; ZHONG, X.; YI, H. Y.; YU, X. Q. *Manduca sexta* gloverin binds microbial components and is active against bacteria and fungi. **Dev. Comp. Immunol**., v. 38, p. 275–284, 2012.

YEVTUSHENKO, D. P.; ROMERO, R.; FORWARD, B. S.; HANCOCK, R. E.; KAY, W. W.; MISRA, S. Pathogen-induced expression of a cecropin A-melittin antimicrobial peptide gene confers antifungal resistance in transgenic tobacco. **J. Exp. Bot**., v. 56, p. 1685–1695, 2005.

YI, H.Y.; CHOWDHURY, M.; HUANG, Y. D.; YU, X. Q. Insect antimicrobial peptides and their applications. **Appl. Microbiol. Biotechnol**. n. 98, v. 13, p. 5807-5822, 2014.

YOE, S. M.; KANG, C. S.; HAN, S. S, BANG, I. S. Characterization and cDNA cloning of hinnavin II, a cecropin family antibacterial peptide from the cabbage butterfly, *Artogeia rapae*. **Comp. Biochem. Physiol. B Biochem. Mol. Biol**., v. 144, p. 199–205, 2006.

ZANNIN, M.; LOURENÇO, D. M.; MOTTA, G.; DALLA-COSTA, L. R.; GRANDO, M.; GAMBORGI, G. P.; CHUDZINSKI-TAVASSI, A. M. Blood coagulation and fibrinolytic factors in 105 patients with hemorrhagic syndrome caused by accidental contact with Lonomia obliqua caterpillar in Santa Catarina, southern Brazil. **Thromb Haemost**, v. 89, p. 355-364, 2003.

ZANUNCIO, J. C.; ALVES, J. B.; SANTOS, G. P.; CAMPOS, W. O. Levantamento e flutuação populacional de lepidópteros associados à eucaliptocultura. **Pesquisa Agropecuária Brasileira**, v. 28, p. 1121-1127, 1993.

ZHAO, L.; LU, W. Defensins in innate immunity. **Curr. Opin. Hematol**., v. 21, p. 37–42, 2014.

ZHU, S.; LI, W.; JIANG, D.; ZENG, X. Evidence for the existence of insect defensin-like peptide in scorpion venom. **IUBMB Life**, v. 50, p. 57-61, 2000.

ZHU, Y.; JOHNSON, T. J.; MYERS, A. A.; KANOST, M. R. Identification by subtractive suppression hybridization of bacteria-induced genes expressed in *Manduca sexta* fat body. **Insect Biochem. Mol. Biol.,** v. 33, p. 541–559, 2003

ZORZENON, F. J. Principais pragas das palmeiras. In: ALEXANDRE, M. A. V; DUARTE, L. M. L.; CAMPOS-FARINHA, A. E. C. **Plantas ornamentais**: doenças e pragas. 2008. Cap. 10 p. 207-247.