

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE  
DARCY RIBEIRO**

**LARISSA LUANA DE OLIVEIRA MANHÃES**

**ANÁLISE COMPARATIVA DE TRANSCRIPTOMAS EM OVOS DO  
MOSQUITO *Anopheles gambiae* (Diptera: Culicidae) E DO BESOURO  
*Tribolium castaneum* (Coleoptera: Tenebrionidae): BUSCA DE GENES  
ANCESTRAIS RELACIONADOS COM A FORMAÇÃO DA CUTÍCULA  
Serosa**

**CAMPOS DOS GOYTACAZES - RJ  
SETEMBRO - 2023**

**LARISSA LUANA DE OLIVEIRA MANHÃES**

**ANÁLISE COMPARATIVA DE TRANSCRIPTOMAS EM OVOS  
DO MOSQUITO *Anopheles gambiae* (Diptera: Culicidae) E DO  
BESOURO *Tribolium castaneum* (Coleoptera: Tenebrionidae):  
BUSCA DE GENES ANCESTRAIS RELACIONADOS COM A  
FORMAÇÃO DA CUTÍCULA SEROSA**

Dissertação apresentada ao Centro de  
Biotecnologia e Biociências da Universidade  
Estadual do Norte Fluminense Darcy  
Ribeiro, como parte das exigências para  
obtenção do título de Mestre em Biociências  
e Biotecnologia."

**Prof. Dr. Gustavo Lazzaro Rezende**

**CAMPOS DOS GOYTACAZES - RJ  
SETEMBRO - 2023**

**FICHA CATALOGRÁFICA**

UENF - Bibliotecas

Elaborada com os dados fornecidos pela autora.

M277 Manhães, Larissa Luana de Oliveira.

"ANÁLISE COMPARATIVA DE TRANSCRIPTOMAS EM OVOS DO MOSQUITO *Anopheles gambiae* (Diptera : *Culicidae*) E DO BESOURO *Tribolium castaneum* (Coleoptera : *Tenebrionidae*) : BUSCA DE GENES ANCESTRAIS RELACIONADOS COM A FORMAÇÃO DA CUTÍCULA SEROSA". / Larissa Luana de Oliveira Manhães. - Campos dos Goytacazes, RJ, 2023.

125 f. : il.

Inclui bibliografia.

Dissertação (Mestrado em Biociências e Biotecnologia) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, 2023.

Orientador: Gustavo Lazzaro Rezende.

1. cutícula. 2. serosa. 3. evolução. 4. homologos. 5. ovos de insetos. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 570

**ANÁLISE COMPARATIVA DE TRANSCRIPTOMAS EM OVOS DO  
MOSQUITO *Anopheles gambiae* (Diptera: Culicidae) E DO BESOURO  
*Tribolium castaneum* (Coleoptera: Tenebrionidae): BUSCA DE GENES  
ANCESTRAIS RELACIONADOS COM A FORMAÇÃO DA CUTÍCULA  
SEROSA**

Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Biociências e Biotecnologia."

**Prof. Dr. Gustavo Lazzaro Rezende**

\_\_\_\_\_ em \_\_\_\_/\_\_\_\_/\_\_\_\_

**Banca examinadora:**

\_\_\_\_\_  
Dr. Gustavo Lazzaro Rezende - UENF

*Ana Eliza Zeraik*

\_\_\_\_\_  
Dra. Ana Eliza Zeraik - UENF

\_\_\_\_\_  
Dr. José Roberto Da Silva - UFRJ

*Kátia Valevski Sales Fernandes*

\_\_\_\_\_  
Dra. Kátia Valevski Sales Fernandes - UENF

**CAMPOS DOS GOYTACAZES - RJ  
SETEMBRO - 2023**



Governo do Estado do Rio de Janeiro  
Universidade Estadual do Norte Fluminense Darcy Ribeiro  
Diretoria do Centro de Biociências e Biotecnologia

## DECLARAÇÃO

A **versão final** da dissertação de Mestrado intitulada "Análise comparativa de transcriptomas em ovos do mosquito *Anopheles gambiae* (Diptera: Culicidae) e do besouro *Tribolium castaneum* (Coleoptera: Tenebrionidae): busca de genes ancestrais relacionados com a formação da cutícula serosa", defendida pela aluna Larissa Luana de Oliveira Manhães (Matrícula 202112120003) do Programa de Biociências e Biotecnologia da UENF no dia 01 de setembro de 2023, foi **aprovada** pelos membros da banca: Gustavo Lazzaro Rezende (Presidente e membro interno) e José Roberto Silva (membro externo), impossibilitados de assinar presencialmente o documento.

Campos dos Goytacazes, 17 novembro de 2023

Gustavo Lazzaro Rezende  
Professor Associado, LQFPP, CBB, UENF  
ID. Funcional 4391413-6



Documento assinado eletronicamente por **Gustavo Lazzaro Rezende, Professor**, em 21/11/2023, às 15:06, conforme horário oficial de Brasília, com fundamento nos art. 28º e 29º do [Decreto nº 48.209, de 19 de setembro de 2022](#).



A autenticidade deste documento pode ser conferida no site [http://sei.rj.gov.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=6](http://sei.rj.gov.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=6), informando o código verificador **63595505** e o código CRC **99593091**.

Referência: Processo nº SEI-260009/004364/2023

SEI nº 63595505

Avenida Alberto Lamego, 2000, - Bairro Pq. Califórnia, Campos dos Goytacazes/RJ, CEP 28013-602  
Telefone: - www.uenf.br

Dedico à memória da minha avó Vilma de Oliveira Santos, que sempre foi um exemplo de garra, fé e coragem, e faleceu em setembro de 2022.

Ninguém ignora tudo. Ninguém sabe tudo. Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa. Por isso aprendemos sempre."

Paulo Freire

## AGRADECIMENTOS

Agradeço a Deus por ter me sustentado até aqui, por me proporcionar a conclusão de mais uma etapa em minha vida, em meio aos obstáculos que enfrentei pelo caminho. Agradeço a minha família, em especial a minha mãe, que sempre esteve ao meu lado desde a graduação, dando todo o apoio necessário para que eu pudesse alcançar os meus objetivos e que nunca me deixou desistir. Um exemplo de mulher, a qual eu desejo ser motivo de muito orgulho para sempre. Agradeço também ao meu pai, por me entender e me amparar em momentos difíceis. Também gostaria de agradecer a minha irmã querida, a Karol, por toda paciência, apoio e carinho durante esses anos, principalmente antes de eu me casar, quando nós dividíamos o quarto e eu ficava por horas estudando e você sempre presente acompanhando tudo de perto e apoiando do seu jeitinho. E ao meu esposo, por me apoiar a buscar sempre o melhor para a minha vida acadêmica e profissional, desde o namoro. Além de todos os outros membros, os quais apresento total gratidão. Obrigada por entenderem os momentos de exaustão e principalmente, por acreditarem no meu potencial, isso é um dos motivos pelo qual me empenho a ser melhor a cada dia.

Gostaria de agradecer aos meus amigos, Aline da Gama Melila Licurgo, Sandralice Marins da Silva Dias, Tuane Cristina da Silva e Wanderson Júnior dos Santos Chagas que estão em minha vida há muitos anos e que se fizeram presentes em todas as etapas, desde a seleção do mestrado a esse momento. E todos os demais amigos que se fizeram presente e não foram mencionados.

Gostaria de agradecer às minhas colegas de laboratório por todo suporte fornecido desde quando iniciei na equipe como bolsista TCT até o momento, como mestranda, Ulli Barros de Oliveira, Mariana Mendonça, Josiana Gomes e Cinthia Mello. Vocês tiveram um papel fundamental em minha formação, contribuindo com todo apoio e amizade e sendo uma equipe maravilhosa, tornando nosso ambiente de pesquisa agradável e acolhedor.

Quero também agradecer ao cientista Samuel Hund por ter me ajudado no início da pesquisa, fornecendo os dados de meu interesse atualizados do banco de dados Vector Base.

Agradeço também ao Francisnei Pedrosa, que me ensinou a utilizar linhas de comando necessárias para realizar meu trabalho.

Também sou grata a CAPES por ter me concedido a Bolsa de Mestrado e tornado o sonho desse título possível, visto que eu precisava de uma renda para poder me dedicar exclusivamente a pesquisa.

Por fim, gostaria de agradecer ao meu orientador Dr. Gustavo Lazzaro Rezende por ter me ajudado durante esse processo de pesquisa, aprendizagem e escrita, mesmo com todos os seus compromissos pessoais e profissionais, me auxiliou no desenvolvimento de todo trabalho, sanando todas as minhas dúvidas, sempre me aconselhando, acreditando no meu potencial e contribuindo para o meu progresso. Gustavo, muito obrigada por tudo!

## RESUMO

A embriogênese da maioria dos insetos se inicia no momento da oviposição, quando o ovo passa a ficar exposto a estresses bióticos e abióticos. Durante a embriogênese inicial as células da serosa, um tecido extraembrionário, expressam vários genes que serão responsáveis por produzir e secretar uma matriz extracelular denominada de cutícula serosa. Essa cutícula se associa internamente à casca do ovo aumentando a proteção do ovo contra a perda de água e contra predadores. Até o presente momento apenas dois transcriptomas de ovos, relacionados com a serosa e produção da cutícula serosa, foram publicados: em 2009, feito a partir de microarranjo do mosquito africano *Anopheles gambiae* (Giles, 1902), principal vetor da malária no mundo, e em 2020, feito a partir de RNAseq do besouro da farinha *Tribolium castaneum* (Herbst, 1897), que é uma praga de grãos armazenados. Esses dois transcriptomas ainda não foram analisados e comparados de forma sistemática e este é o objetivo do presente trabalho. A associação de genes anotados no genoma com as sondas de microarranjo foi atualizada e, em ambos os transcriptomas foi preciso: identificar, descrever e realizar a anotação funcional dos genes supostamente expressos na serosa, identificar genes homólogos e selecionar genes a serem investigados experimentalmente no futuro. Foram utilizados os bancos de dados *non redundant* do *GenBank*, *Ensembl Metazoa*, *IBeetle-Base*, *VectorBase*, *FlyBase*, *Interpro*, *PFAM* e ferramentas como *BLAST* e *GeneTree*. Obtivemos como resultados: i) a diminuição de sondas órfãs (sondas que não eram associadas a genes) do microarranjo de 165 para 30 e um total de 472 genes diferencialmente expressos no mosquito; ii) a descrição e anotação de função de 371 e 242 genes dos transcriptomas de *A. gambiae* e *T. castaneum*, dos totais de 472 e 318 genes, respectivamente; iii) 101 e 76 genes desconhecidos no mosquito e no besouro, o que equivale a 21 e 24% dos transcriptomas; iv) ortologia de 30 genes de *A. gambiae* com 31 genes de *T. castaneum* e paralogia entre espécies de 37 genes de mosquito com 21 genes de besouro, o que equivale a 14 e 16%, respectivamente, de genes homólogos em ambas as espécies; v) compartilhamento de diversas funções gênicas em ambos os transcriptomas: metabolismo de lipídeos, proteases, transportadores transmembranares, proteína cuticular, relacionado ao citoesqueleto, Proteína G, citocromos P450 (CYP450), fator de transcrição, metabolismo de quitina, sinalização de cálcio, proteína quimiossensorial e relacionado ao relógio circadiano. A presença de genes homólogos e as mesmas funções em ambas as espécies indica que estes são essenciais para a formação da cutícula serosa, visto que estão conservados a pelo menos 330 milhões de anos. Tais resultados auxiliam no entendimento da evolução dos ovos de insetos e dos genes cuticulares e podem ajudar no estabelecimento de novas estratégias de controle de vetores e pragas.

Palavras-chave: cutícula, serosa, evolução, homólogos, ovos de insetos,

## ABSTRACT

The embryogenesis of most insects begins at the time of oviposition, when the egg becomes exposed to biotic and abiotic stresses. During early embryogenesis, serosa cells, an extraembryonic tissue, express several genes that will be responsible for producing and secreting an extracellular matrix called serosal cuticle. This cuticle is associated internally with the eggshell, increasing egg protection against water loss and predators. So far, only two egg transcriptomes, related to the serosa and the serosal cuticle production, have been published: in 2009, made with microarray of the African mosquito *Anopheles gambiae* (Giles, 1902), the main vector of malaria in the world, and in 2020, made with RNAseq from the flour beetle *Tribolium castaneum* (Herbst, 1897), a pest of stored grains. These two transcriptomes have not yet been systematically analyzed and compared and this is the objective of the present work. The association of genes annotated in the genome with the microarray probes was updated and, in both transcriptomes, it was necessary to: identify, describe and perform the functional annotation of genes supposedly expressed in the serosa, identify homologous genes and select genes to be experimentally investigated in the future. The following databases were employed: non redundant from GenBank, Ensembl Metazoa, IBeetle-Base, VectorBase, FlyBase, Interpro, PFAM as well as tools such as BLAST and GeneTree. The following results were obtained: i) microarray orphan probes decreased (probes that were not associated with genes) from 165 to 30 and a total of 472 differentially expressed genes in the mosquito; ii) the description and function annotation of 371 and 242 genes from the transcriptomes of *A. gambiae* and *T. castaneum*, out of a total of 472 and 318 genes, respectively; iii) 101 and 76 unknown genes in the mosquito and in the beetle, which accounts to 21 and 24% of the transcriptomes; iv) orthology of 30 *A. gambiae* genes with 31 *T. castaneum* genes and paralogy among species of 37 mosquito genes with 21 beetle genes, which is equivalent to 14 and 16%, respectively, of homologous genes in both species; v) sharing of diverse gene functions in both transcriptomes: lipid metabolism, proteases, transmembrane transporters, cuticular protein, cytoskeleton-related, G protein, cytochrome P450 (CYP450), transcription factor, chitin metabolism, calcium signaling, chemosensory protein and related to the circadian clock. The presence of homologous genes and the same functions in both species indicates that they are essential for the formation of the serosal cuticle, since they have been conserved for at least 330 million years. Such results assist in understanding the evolution of insect eggs and cuticular genes and may help in establishing new vector and pest control strategies.

Keywords: cuticle, serosa, evolution, homologues, insects eggs.

## LISTA DE FIGURAS

<b>Figura 1.</b> Estágios de vida do mosquito <i>A. gambiae</i> e processo de repasto sanguíneo .....	19
<b>Figura 2.</b> Estágios de vida do besouro <i>T. castaneum</i> : Ovo, instares larvais, pupa e adulto ...	20
<b>Figura 3.</b> Estágios iniciais da embriogênese de insetos. Esquema de um ovo cortado em seção cruzada.....	21
<b>Figura 4.</b> A serosa é uma novidade evolutiva dos insetos .....	22
<b>Figura 5.</b> Esquema da dinâmica de formação e estrutura da cutícula serosa de insetos .....	23
<b>Figura 6.</b> Isolamento de células da serosa aderidas a cascas de ovos para análise de microarranjo .....	24
<b>Figura 7.</b> Perfil de expressão de genes alongase ao longo da embriogênese de <i>D. melanogaster</i> e <i>A. gambiae</i> .....	25
<b>Figura 8.</b> Importância do fator de transcrição Zen1 para a formação da serosa .....	26
<b>Figura 9.</b> Diferentes relações de homologia entre genes exemplificado em <i>Homo sapiens</i> e <i>Mus musculus</i> .....	30
<b>Figura 10.</b> Fluxograma das etapas metodológicas empregadas .....	33
<b>Figura 11.</b> Análise das sondas órfãs .....	42
<b>Figura 12.</b> Resultados do BLASTx da região da sonda Ag.UNKN.1126.0_CDS_s_at	43
<b>Figura 13.</b> Gráfico das funções de todos os genes de <i>A. gambiae</i> (parte superior da figura) e de <i>T. castaneum</i> (parte inferior da figura) .....	59
<b>Figura 14.</b> Gráfico das funções dos genes de <i>A. gambiae</i> (parte superior da figura) e de <i>T. castaneum</i> (parte inferior da figura) contendo somente as funções conhecidas e sem a "cauda" (funções representadas por apenas um gene).....	60
<b>Figura 15.</b> Perfis de expressão gênica dos 30 ortólogos ao longo da embriogênese de <i>A. gambiae</i> .....	67

## LISTA DE TABELAS

<b>Tabela 1.</b> Lista das 41 sondas de <i>A. gambiae</i> que deixaram de ser órfãs.....	44
<b>Tabela 2.</b> Análise dos 20 genes desconhecidos mais expressos na serosa de <i>A. gambiae</i> .....	46
<b>Tabela 3.</b> Identificação dos genes expressos na serosa do mosquito <i>Anopheles gambiae</i> agrupados pelas funções atribuídas em ordem alfabética .....	48
<b>Tabela 4.</b> Disposição do “top 10” das funções conhecidas mais presentes na serosa de <i>A. gambiae</i> .....	52
<b>Tabela 5.</b> Identificação dos genes que possivelmente estão sendo expressos na serosa do besouro <i>T. castaneum</i> agrupados pelas funções atribuídas em ordem alfabética .....	54
<b>Tabela 6.</b> Disposição do “top 10” das funções conhecidas dos genes provavelmente presentes na serosa de <i>T. castaneum</i> .....	56
<b>Tabela 7.</b> Comparação da anotação funcional dos genes de <i>A. gambiae</i> e <i>T. castaneum</i> .....	58
<b>Tabela 8.</b> Trinta genes em <i>A. gambiae</i> são ortólogos de 31 genes de <i>T. castaneum</i> .....	62
<b>Tabela 9.</b> Trinta e sete genes de <i>A. gambiae</i> possuem parálogos entre espécies de 21 genes de <i>T. castaneum</i> .....	68
<b>Tabela 10.</b> Genes selecionados a serem investigados experimentalmente.....	71

## LISTA DE ANEXOS

<b>ANEXO I.</b> Passo a passo FINDSTR: Busca de sequências automatizadas .....	78
<b>ANEXO II.</b> Identificação dos genes associados a sondas de <i>A. gambiae</i> (atualização dos dados originais de 2009).....	79
<b>ANEXO III.</b> Passo a passo - GREP: Para obter as sequências proteicas através do proteoma OGS3 .....	92
<b>ANEXO IV.</b> Passo a passo – AWK: Para obter as sequências em uma única linha.....	93
<b>ANEXO V.</b> Tabela com 30 sondas órfãs de <i>A. gambiae</i> com suas respectivas informações para serem analisadas e relacionadas a genes no futuro .....	94
<b>ANEXO VI.</b> Tabela com descrição e função dos 472 genes expressos na serosa do <i>A. gambiae</i> .....	95
<b>ANEXO VII.</b> Tabela com descrição e função dos 318 genes possivelmente expressos na serosa do <i>T. castaneum</i> .....	113

## GLOSSÁRIO

**BLAST: Basic Local Alignment Search Tool.** Ferramenta de Busca de Alinhamento Local Básico. É uma ferramenta de comparação que acha regiões de similaridades entre sequências. A ferramenta compara sequências de nucleotídeos ou de proteínas e calcula a significância estatística dos pareamentos (*matches*). O BLAST pode ser utilizado para inferir relações funcionais ou evolutivas entre sequências assim como ajudar a identificar membros de famílias gênicas.

**BLASTn: Nucleotide BLAST.** Algoritmo que realiza a comparação da sequência nucleotídeo-nucleotídeo. É usado quando temos uma sequência de DNA de busca (*query*) e queremos encontrar sequências de DNA similares (*subject*) dentro do banco de dados.

**BLASTx: Translated Nucleotide to Protein BLAST.** Algoritmo que realiza a comparação da sequência nucleotídeo-proteína. As sequências nucleotídicas do *query* são traduzidas em seis quadros de leitura diferentes e essas seis traduções proteicas são comparadas contra um banco de dados de proteínas, buscando encontrar proteínas similares (*subject*).

**BLASTp: Protein BLAST.** Algoritmo que realiza a comparação da sequência proteína-proteína. A sequência da proteína em questão (*query*) é comparada com sequências de proteínas depositadas no banco de dados, buscando-se as sequências de proteínas mais similares (*subjects*).

**ESTs: Expressed Sequence Tags.** Dados de expressão de RNA mensageiro provenientes de pequenas sequências de cDNA. Esses fragmentos podem servir como evidências de dados experimentais indicando que determinada região do genoma possui expressão. Também é útil no processo de descoberta de novos genes.

**Fator de transcrição:** Proteína reguladora necessária para a transcrição genética. A transcrição é o primeiro passo da expressão genética e envolve a transferência das informações contidas no DNA para uma molécula de RNA que é posteriormente processada para dar origem aos produtos dos genes.

**Formato FASTA:** Formato de arquivo de computador baseado em texto para representar tanto sequências de nucleotídeos quanto sequências de proteínas, em que os resíduos de nucleotídeos ou de aminoácidos são representados usando códigos de uma única letra.

**Microarranjo:** Microarranjo de DNA ou Chip de DNA é uma coleção de pequenos fragmentos de moléculas de DNA que correspondem a sondas para regiões do genoma produzindo resultados quantitativos de expressão gênica a partir de sinais de fluorescência.

**PFAM:** Banco de dados de famílias e domínio de proteínas, incluindo suas anotações e alinhamentos de sequências utilizando Modelos Ocultos de Markov (*Hidden Markov Models*).

**qPCR: quantitative Polymerase Chain Reaction.** OqPCR é uma reação em cadeia da polimerase (PCR) em que os dados de amplificação são medidos em tempo real. É uma técnica de Biologia Molecular que permite a replicação in vitro de um fragmento de DNA de interesse de forma extremamente rápida e quantitativa.

**RNAseq: RNA sequencing:** Técnica de sequenciamento de próxima geração (NGS) capaz de analisar padrões de expressão de uma amostra biológica, em uma determinada condição.

**Similaridade entre proteínas:** Percentual obtido no alinhamento entre duas proteínas que indica a soma do percentual de resíduos de aminoácidos idênticos em uma mesma posição mais o percentual de resíduos de aminoácidos que possuem características semelhantes em uma mesma posição.

**Sistemática:** Área da Biologia dedicada a inventariar e descrever a biodiversidade e compreender as relações filogenéticas entre os organismos. Inclui a taxonomia e também a filogenia, compreendendo assim, a classificação dos diversos organismos vivos.

**Sonda de microarranjo do tipo “CDS”:** Sonda que se associa a região codante do transcrito (CDS: *coding sequence*), que vai constituir algum exon do gene em questão. Exemplo: Ag.3L.479.0\_CDS\_at, leia-se: sonda 479.0\_CDS\_at do cromossomo 3, braço L de *A. gambiae*.

**Sonda de microarranjo do tipo “UTR”:** Sonda que se associa a uma região não traduzida

do transcrito (UTR: *untranslated region*). Exemplo: Ag.2R.461.0\_UTR\_a\_at, leia-se: sonda 461.0\_UTR\_a\_at do cromossomo 2, braço R de *A. gambiae*.

**TMHMM:** Programa que prevê a probabilidades de regiões transmembranares em uma sequência de proteína.

## SUMÁRIO

1	INTRODUÇÃO.....	17
1.1	Os insetos.....	17
1.2	Ovos de insetos, serosa e cutícula serosa.....	20
1.2.1	Estrutura da cutícula serosa.....	22
1.2.2	Composição bioquímica da cutícula serosa.....	23
1.2.3	A serosa e cutícula serosa de <i>A. gambiae</i> .....	23
1.2.4	A serosa e cutícula serosa de <i>T. castaneum</i> .....	26
1.3	O uso da Transcriptômica e técnicas utilizadas para obtenção dos genes relacionados a serosa do mosquito <i>A. gambiae</i> e do besouro <i>T. castaneum</i> .....	27
1.4	Relações de homologia gênica entre espécies .....	28
2	OBJETIVOS.....	32
2.1	Objetivo geral .....	32
2.2	Objetivos específicos .....	32
3	METODOLOGIA.....	33
3.1	Obtenção dos dados e identificação de genes de <i>Anopheles gambiae</i> diferencialmente expressos na serosa, associados às sondas órfãs.....	34
3.1.1	Obtenção dos dados .....	34
3.1.2	Identificação dos genes de <i>A. gambiae</i> .....	35
3.2	Descrição dos genes expressos na serosa de <i>Anopheles gambiae</i> e investigação dos genes sem descrição.....	36
3.3	Atribuição funcional dos genes expressos na serosa do <i>Anopheles gambiae</i> .....	36
3.4	Descrição dos genes possivelmente relacionados com a serosa do <i>Tribolium castaneum</i> .....	37
3.5	Atribuição funcional de genes possivelmente expressos na serosa do <i>Tribolium castaneum</i> .....	38
3.6	Análise de genes considerados homólogos.....	38
3.6.1	BLASTp do <i>Tribolium castaneum</i> para o <i>Anopheles gambiae</i> .....	38
3.6.2	BLASTp recíproco do <i>Anopheles gambiae</i> para o <i>Tribolium castaneum</i> .....	39
3.7	Análise de genes considerados parálogos entre espécies.....	39
3.8	Designação de genes a serem estudados experimentalmente no futuro .....	39
4	RESULTADOS E DISCUSSÃO.....	41
4.1	Identificação de genes de <i>Anopheles gambiae</i> associados a sondas consideradas órfãs .....	41
4.2	Descrição e anotação funcional dos genes expressos na serosa de <i>A. gambiae</i> .....	45
4.3	Descrição e anotação funcional dos genes possivelmente expressos na serosa de <i>T. castaneum</i> .....	52
4.4	Comparação da anotação funcional entre ambas as espécies .....	56
4.5	Análise dos genes homólogos presentes em ambos os transcriptomas .....	61
4.6	Genes selecionados para serem estudados experimentalmente no futuro .....	61
5	CONCLUSÕES.....	72
6	REFERÊNCIAS BIBLIOGRÁFICAS .....	73

# 1 INTRODUÇÃO

## 1.1 Os insetos

Os insetos são os animais invertebrados do filo Arthropoda e subfilo Hexapoda, formados por pernas articuladas e corpo segmentado. O subfilo dos hexápodos (do Grego que significa "seis pernas") compreende a maior parte das espécies de artrópodes e inclui a classe Insecta, além de três pequenos grupos de seres ápteros: Collembola, Protura e Diplura (Grimaldi e Engel, 2005; Gullan, 2017). Os insetos possuem uma estrutura corpórea composta por cabeça tórax e abdome, tendo três segmentos torácicos com um par de pernas em cada segmento torácico. O segundo e terceiro segmentos torácicos podem apresentar asas, dependendo da espécie. O exoesqueleto dos insetos é composto por duas camadas cuticulares secretadas pela epiderme: a camada mais externa é a epicutícula, rica em lipídeos e a camada mais interna é a procutícula, constituída por proteínas e quitina, um polissacarídeo estrutural. A cutícula possui algumas funções cruciais como a proteção contra predadores e agentes patológicos e proteção contra a perda de água. Desta forma, a cutícula é uma das principais características contribuintes para o sucesso da classe Insecta (Gullan, 2017).

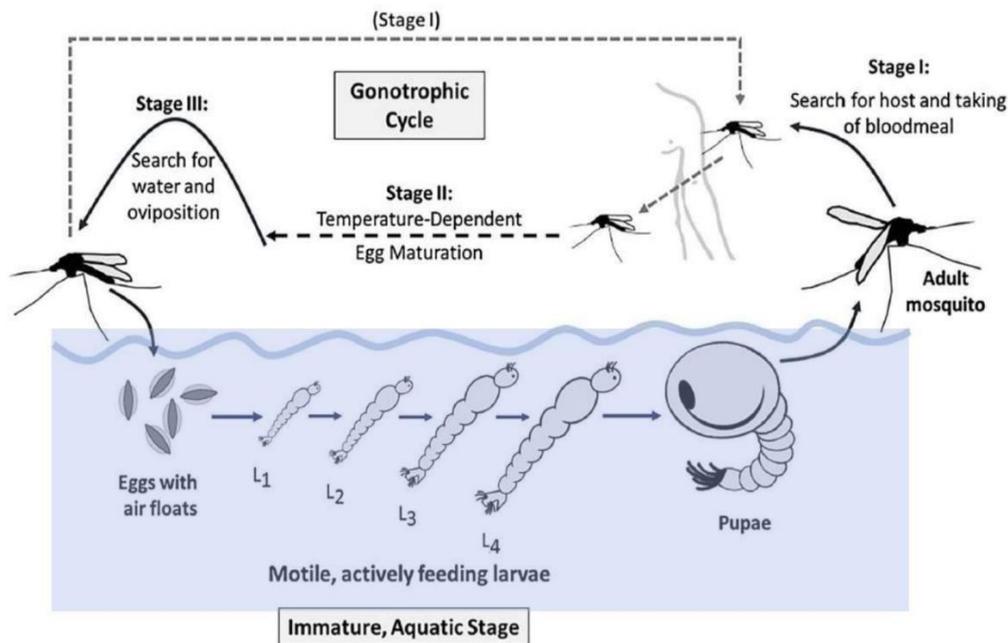
Os insetos são o grupo animal com maior número de espécies, sendo importantes para algumas culturas humanas como polinizadores e fontes de alimentos, dentre outros papéis, mas também sendo prejudiciais como pragas agrícolas e vetores de agentes etiológicos de doenças (Grimaldi e Engel, 2005). Mais de um milhão de espécies de insetos já foram descritas em publicações de taxonomia (Gullan, 2017). Eles fornecem benefícios econômicos e ambientais e as características de alguns os fazem modelos úteis para entender processos biológicos gerais, como por exemplo a mosca *Drosophila melanogaster*, que é um excelente modelo de estudo devido a facilidade de criação em laboratório, tempo de geração, alta fecundidade e facilidade de manipulação (Gullan, 2017)

Cinco ordens de insetos são consideradas as principais, pois destacam-se por sua riqueza de espécies: os besouros (Coleoptera), moscas e mosquitos (Diptera), vespas, abelhas e formigas (Hymenoptera), borboletas e mariposas (Lepidoptera), e percevejos (Hemiptera) (Gullan, 2017). As espécies de interesse para a presente pesquisa são o mosquito *Anopheles gambiae* (Diptera: Culicidae) (Giles, 1902) e o besouro *Tribolium castaneum* (Coleoptera: Tenebrionidae) (Herbst, 1897), que possuem histórias de vida e ecologias muito distintas e cujo último ancestral comum existiu a 330 milhões de anos atrás (Timetree, 2023).

Mosquitos do gênero *Anopheles* são vetores da malária em mamíferos, sendo a espécie

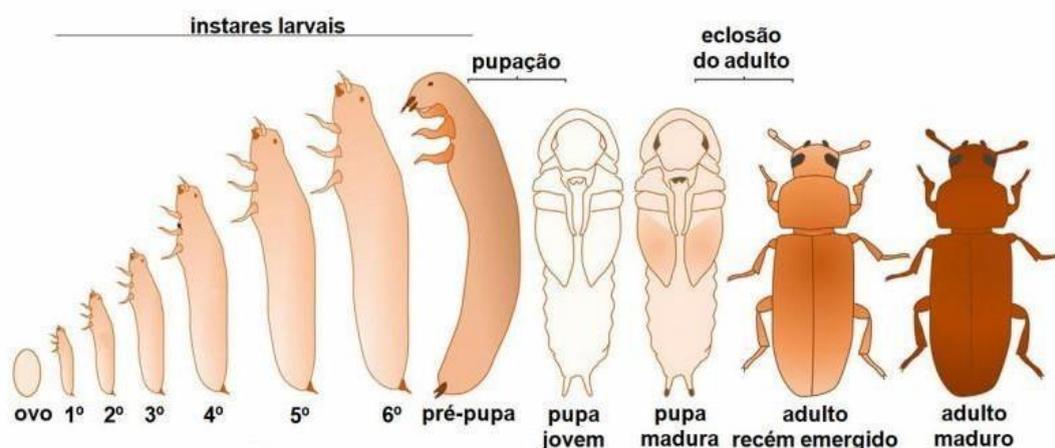
*A. gambiae* considerada o principal vetor do agente etiológico da malária na África (Gullan, 2017). A Organização Mundial de Saúde registrou 247 milhões de casos de malária em 84 países endêmicos no ano de 2021, ocorrendo em grande parte em países da região Africana. Por outro lado, houve uma diminuição na quantidade de óbitos por malária de 625.000 em 2020 para 619.000 em 2021 (Ministério da Saúde, [s.d.]). No Brasil, a maioria dos casos de malária se concentram na região amazônica (Ministério da Saúde, [s.d.]). Os parasitas que provocam a malária são os protozoários do gênero *Plasmodium*. Cinco espécies são responsáveis pela malária nos humanos: *P. falciparum*, *P. ovale*, *P. knowlesi*, *P. malariae* e *P. vivax* (Gullan, 2017).

O *A. gambiae* é holometábolo, ou seja, durante o desenvolvimento os indivíduos passam por uma metamorfose completa, baseada em quatro fases de vida: ovo, larva, pupa e adulto (Clements, 1992) (Figura 1). A fase de vida larval é subdividida em 4 estágios, de L1 a L4. Machos e fêmeas se alimentam de néctar de plantas e as fêmeas se alimentam também do sangue de vertebrados, pois é através dele que obtém os nutrientes necessários para a produção de seus ovos (Clements, 1992; Okuneye *et al.*, 2019). É no processo de repasto sanguíneo que as fêmeas podem se infectar com *Plasmodium* de um vertebrado infectado, ou transmitir o *Plasmodium* para um indivíduo saudável (Foster e Walker 2009). Adultos de *A. gambiae*, quando privados de água, exibem uma média de vida útil maior em alta umidade relativa (Wang *et al.*, 2011). As fêmeas muitas vezes depositam os seus ovos em poças que precisam de reabastecimento e, conseqüentemente, o aumento na taxa de chuva favorece a reprodução do mosquito (Gullan, 2017). Assim, as populações de mosquito aumentam de maneira significativa e rápida após o início da estação chuvosa (Jawara *et al.*, 2008), mostrando a susceptibilidade desta espécie para variações de umidade relativa. Essa mesma susceptibilidade ocorre no estágio de vida do ovo, que é posto na água (Goltsev *et al.*, 2009, Farnesi *et al.*, 2017).



**Figura 1.** Estágios de vida do mosquito *A. gambiae* e processo de repasto sanguíneo. No estágio 1 (Stage I) o mosquito adulto fêmea busca pelo hospedeiro para se alimentar de sangue e ocorre o repasto sanguíneo. No estágio 2 (Stage II) o sangue é digerido e a maturação dos ovários depende da temperatura. No estágio 3 (Stage III) a fêmea busca um ambiente aquático para ovipor. A esquerda a fêmea está colocando seus ovos em ambiente aquoso. Os estágios imaturos são aquáticos: ovos, L1, L2, L3 e L4 (larvas de primeiro ao quarto instar, respectivamente) e pupa. Apenas o adulto vive fora da água e é o estágio de vida responsável por transmitir o *Plasmodium*. Figura adaptada de Okuneye *et al.*, 2019.

O besouro *T. castaneum*, também conhecido como besouro da farinha, é uma espécie cosmopolita, sendo uma praga agrícola secundária e depende do ataque de pragas primárias para se instalar nos grãos armazenados. Ataca diversos cereais como arroz, trigo e milho, e também cereais moídos como farinha de trigo e fubá, causando prejuízos elevados pela sua presença e atividade biológica levando a deterioração dos grãos ou farinhas (Lorini, 2002). Ocorrem com maior frequência em ambientes quentes e úmidos, mas vivem bem em ambiente seco e as fêmeas ovipõem no seco (Duarte *et al.*, 2021). Assim como o mosquito, é holometábolo, apresentando seu ciclo de vida também nos 4 estágios de ovo, larva, pupa e adulto (Figura 2). As condições ótimas (*i.e* maiores taxas de viabilidade) para oviposição é temperatura em torno de 30 °C. Nessa temperatura o período de embriogênese é de 3,6 dias e as larvas passam por 6 a 9 instares de desenvolvimento. O ciclo de vida desde a postura do ovo até a eclosão do adulto dura em torno de 27 dias (Bucher, 2006). *Tribolium castaneum* é considerado um popular modelo emergente para estudos de desenvolvimento de insetos holometábolos, incluindo estudos que utilizam a indução da perda de função por interferência de RNA (RNAi) (Gullan, 2017; Gurska *et al.*, 2020).



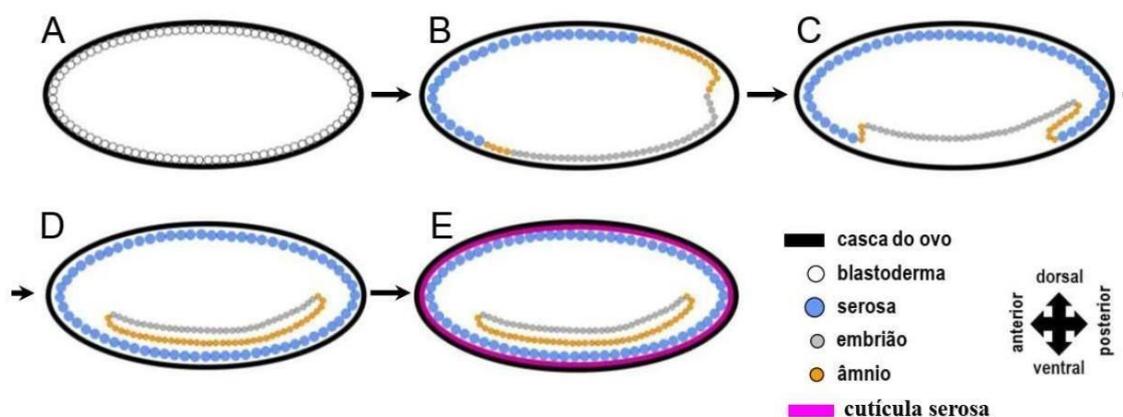
**Figura 2.** Estágios de vida do besouro *T. castaneum*: Ovo, instares larvais, pupa e adulto. Na parte superior o esquema mostra o ovo e os estágios larvais (que podem ser de 6 a 9 instares) seguido dos estágios de pupa (jovem e madura) e adulto, sendo o recém emergido menos melanizado que o adulto maduro. A parte inferior são imagens do ovo (**A**), larva (**B**), pupa (**C**) e adulto (**D**). As imagens não estão na mesma escala. (Figura adaptada de Vargas, 2015 e 2019).

## 1.2. Ovos de insetos, serosa e cutícula serosa

O ovo é o estágio de vida menos conhecido dos insetos (Beament, 1989), são imóveis e susceptíveis a estresses bióticos e abióticos, como predadores e a perda de água. Essa perda de água ocorre no processo de trocas gasosas durante o desenvolvimento embrionário e ambientes com baixa umidade relativa levam a maior dessecação dos ovos, tanto em insetos que depositam seus ovos em ambientes terrestres (caso do *T. castaneum*) quanto em insetos que depositam os ovos em ambientes aquáticos (caso do *A. gambiae*) (Hinton, 1981; Zeh *et al.*, 1989). Um fator importante para proteger os ovos contra a dessecação e garantir a sua sobrevivência é a cutícula serosa, uma matriz extracelular formada no início da embriogênese pelas células da serosa (Rezende *et al.*, 2008, Goltsev *et al.*, 2009; Jacobs *et al.*, 2013).

A serosa é um dos tecidos extraembrionários (junto com o âmnio) considerados uma novidade evolutiva dos insetos. No início da embriogênese (Figura 3) o blastoderma celular dará origem às células da serosa, do âmnio e do embrião (Figura 3A, B). As células da serosa

envolvem o âmnio e o embrião enquanto o âmnio envolve o embrião apenas em sua parte ventral (Figura 3C, D). Após envolver completamente o âmnio e embrião a serosa sintetiza a cutícula serosa (Figura 3E), uma matriz extracelular que protege o ovo contra a perda de água; assim como existe a cutícula (exoesqueleto) do inseto adulto, protegendo-o contra essa perda. (Rezende *et al.*, 2016; Goltsev *et al.*, 2009; Jacobs *et al.*, 2013; Gullan 2017).

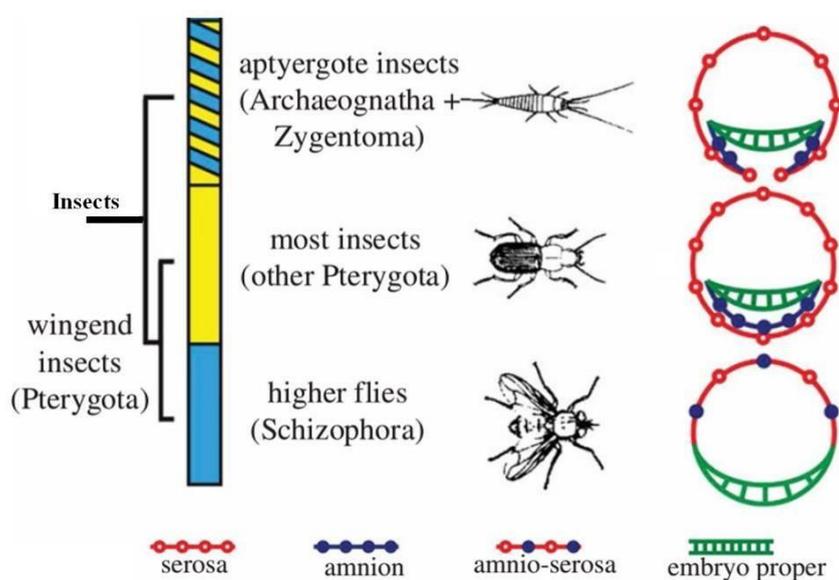


**Figura 3.** Estágios iniciais da embriogênese de insetos. Esquema de um ovo cortado em seção cruzada. (A) Estágio de blastoderma celular não diferenciado. (B) Estágio de blastoderma celular diferenciado com as células do embrião e dos dois tecidos extraembrionários - serosa e âmnio - já definidas. (C, D) As células extraembrionárias se expandem, a serosa recobre embrião e âmnio e o âmnio recobre apenas o embrião, ventralmente. (E) Em seguida, as células da serosa secretam a sua cutícula, que aumenta a proteção do ovo contra a dessecação. Figura criada por Gustavo Rezende a partir de Vargas *et al.*, 2021.

A cutícula serosa é uma estrutura ancestral em insetos sendo também uma novidade evolutiva entre eles, estando ausente em outros artrópodes como os colêmbolas, crustáceos, miriápodes e aracnídeos (Jacobs *et al.*, 2013; Rezende *et al.*, 2016). É entendido que a existência da serosa e de sua cutícula facilitou o processo de terrestrialização dos insetos que ocorreu a cerca de 460 milhões de anos atrás. Os ancestrais dos insetos, que já possuíam uma cutícula protetora, viviam no mar, e o processo de terrestrialização ocorreu de forma gradual, provavelmente ao longo de alguns milhões de anos. A existência de uma cutícula protetora facilitou na transição que ocorreu dos insetos do ambiente aquático para o terrestre, o que foi um ganho evolutivo importante. (Jacobs *et al.*, 2013; Rota-Stabelli *et al.*, 2013; Vargas *et al.*, 2021; Zeh *et al.*, 1989).

A presença da cutícula serosa já foi relatada em espécies das seguintes ordens de insetos: *Archaeognatha*, *Zygentoma*, *Ephemeroptera*, *Zoraptera*, *Dermaptera*, *Plecoptera*, *Orthoptera*, *Grylloblattodea*, *Embioptera*, *Phasmatodea*, *Isoptera*, *Thysanoptera*, *Hemiptera*,

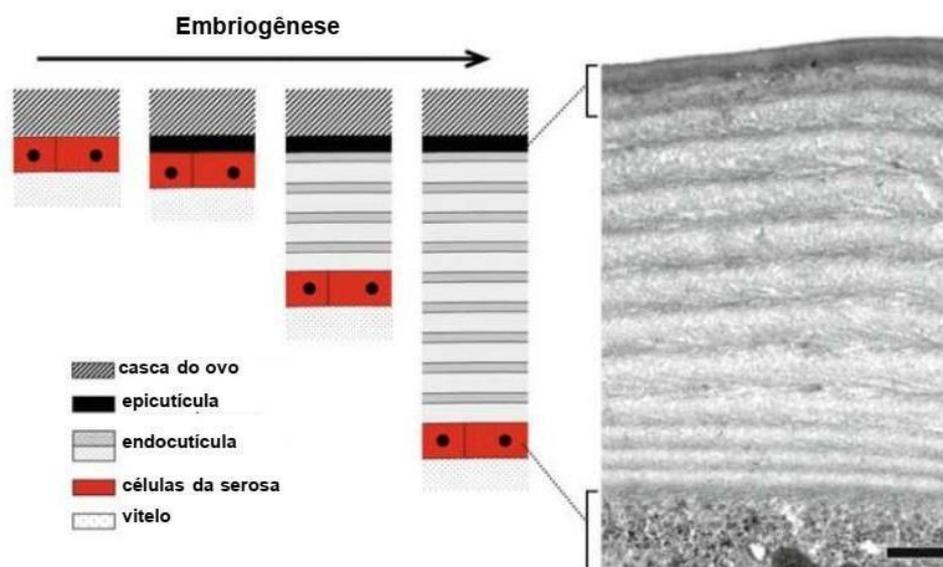
*Psocodea*, *Coleoptera*, *Lepidoptera* e *Diptera* (Rezende *et al.*, 2016). Alguns insetos perderam a cutícula serosa apesar de ainda possuírem a serosa, como por exemplo o percevejo *Oncopeltus fasciatus* (Hemiptera: Lygaeidae). Outros insetos sofreram alterações ainda mais profundas, como por exemplo a mosca *Drosophila Melanogaster* (Diptera: Drosophilidae) que não possui cutícula serosa, serosa e nem o âmnio, passando a ter apenas um tecido extraembrionário denominado amnioserosa (Panflio, 2008, Jacobs *et al.*, 2013) (Figura 4).



**Figura 4.** A serosa é uma novidade evolutiva dos insetos. A faixa azul indica espécies com ambientes de oviposição restritos a locais aquáticos ou úmidos, a faixa amarela indica as espécies que colocam os seus ovos em ambientes terrestres e as faixas diagonais amarela com azul indicam as espécies que vivem em ambientes muito diferentes, quanto à umidade. Nos insetos ápteros (Archaeognatha e Zyngetoma) a serosa não se funde sob o embrião e um plug cuticular é produzido (Rezende *et al.*, 2016). Em *Drosophila melanogaster* ocorreu uma redução secundária, dando origem a um tecido denominado de “amnioserosa” (Figura adaptada de Jacobs *et al.*, 2013).

### 1.2.1 Estrutura da cutícula serosa

A cutícula serosa é composta por duas camadas: a endocutícula, interna e mais espessa e a epicutícula, externa e mais fina e a primeira a ser sintetizada (Figura 5). A endocutícula e a epicutícula da serosa também são conhecidas, respectivamente, como cutícula branca e cutícula amarela. A endocutícula é uma camada dura e fibrosa que apresenta resistência a impactos mecânicos (Rezende, 2008; Rezende *et al.*, 2016; Slifer, 1937).



**Figura 5.** Esquema da dinâmica de formação e estrutura da cutícula serosa de insetos. Os esquemas à esquerda representam a ordem de secreção das camadas da cutícula serosa. No início da embriogênese tem-se o vitelo e as células da serosa (em vermelho) e a casca do ovo. Em seguida uma fina epicutícula é produzida abaixo da casca do ovo e, posteriormente, uma endocutícula espessa composta de várias lamelas é inserida gradualmente. Do lado direito encontra-se uma micrografia eletrônica de transmissão da cutícula serosa de *T. castaneum*. A parte de cima em destaque é a epicutícula e a parte de baixo é a serosa. Adaptado de Rezende *et al.*, 2016.

### 1.2.2 Composição bioquímica da cutícula serosa

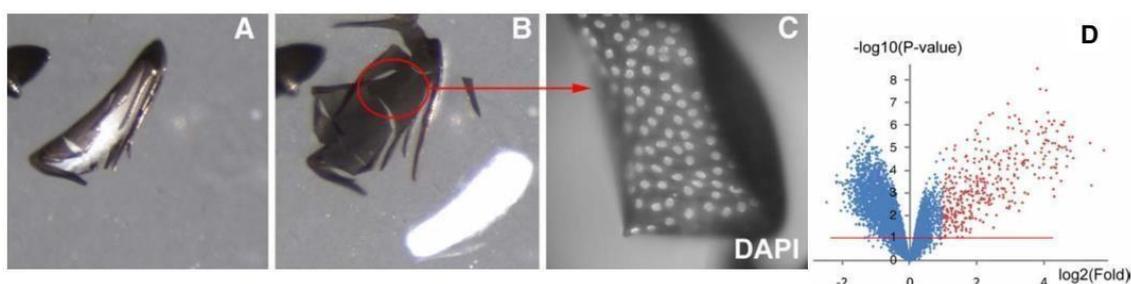
A endocutícula é feita de quitina e proteínas e a presença de quitina na endocutícula foi confirmada em *T. castaneum* através do silenciamento genético (Jacobs *et al.*, 2013). Já a epicutícula não contém quitina e acredita-se que seja composta por proteínas, lipídeos, ceras e/ou hidrocarbonos (Vargas, 2019). Também contém aminoácidos, como fenilalanina e tirosina relacionados a melanização e esclerotização (Rezende *et al.*, 2016).

Já foram encontradas na serosa e cutícula serosa de alguns ortópteros, moléculas de dopa, dopamina e NADA, e uma enzima que pode melanizar as estruturas e estão relacionados com a melanização e vias de esclerotização (Rezende *et al.*, 2016). A maior parte dos dados de genes relacionados com a formação da cutícula serosa são provenientes de trabalhos realizados com *A. gambiae* e *T. castaneum*, que serão detalhados a seguir.

### 1.2.3 A serosa e cutícula serosa de *A. gambiae*

Os trabalhos de Rezende (2008) e Goltsev *et al.* (2009) descrevem em *A. gambiae* o processo de formação da cutícula serosa, bem como investiga genes expressos na serosa no período em que esta sintetiza sua cutícula. Nesse mosquito a cutícula serosa é formada entre 7 e 11 horas após a oviposição, com aproximadamente 20% de embriogênese (Rezende, 2008). Já para os genes expressos na serosa, Goltsev e colaboradores (2009) separaram fisicamente os conteúdos dos ovos fixados de suas cascas (Figura 6) e observaram que as células da serosa

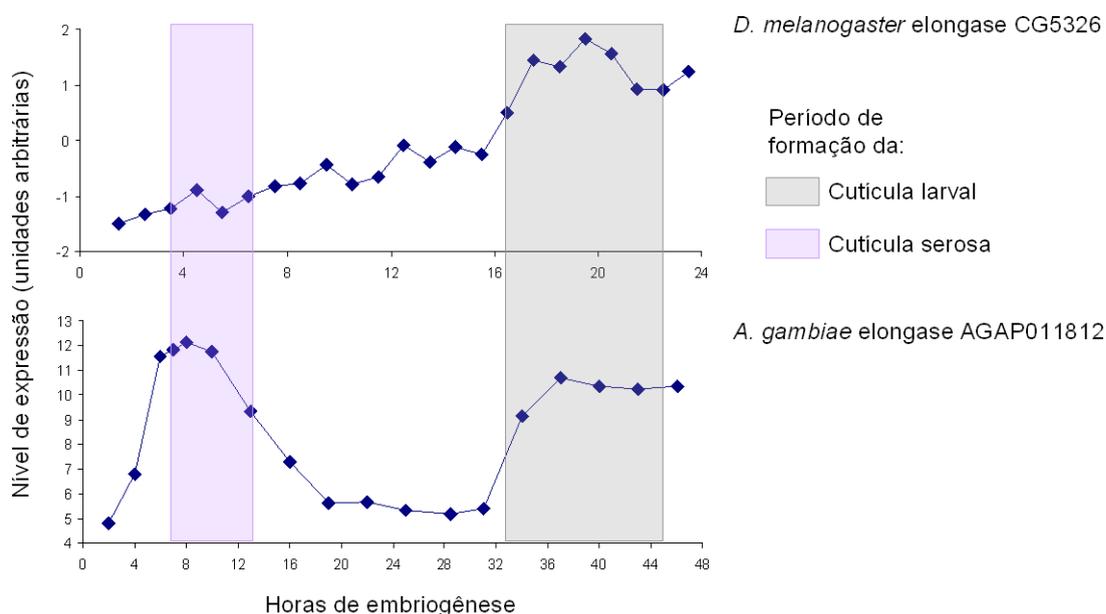
fixadas estavam presentes na casca (Figura 6A-C). Com esses dois materiais isolados (de um lado as células do embrião com o âmnio, o vitelo e os vitelófagos e do outro lado as células da serosa) foi realizada uma análise comparativa em chips de microarranjo e foi possível detectar quais genes foram diferencialmente expressos na serosa desse mosquito (Figura 6D). Na ocasião (ano de 2009) e empregando o genoma de referência do *A. gambiae* de então, os autores encontraram 372 sondas de microarranjo que foram associadas a 359 genes diferencialmente expressos (*i.e.* mais presentes na serosa), de um total de 537 sondas. Ou seja, ainda restaram 165 sondas de microarranjo que apresentaram mais sinal na serosa, mas que eram órfãs, posto que não possuíam correspondência com os genes anotados no genoma na época.



**Figura 6. Isolamento de células da serosa aderidas a cascas de ovos para análise de microarranjo.** (A) Conteúdo do ovo com 10,5 horas de embriogênese fixado dentro da casca. (B) Mesmo campo que em A, mas após a separação da casca de todo o resto. O embrião, junto com âmnio e vitelo, branco, é visto no canto inferior direito da imagem. (C) Uma casca de ovo preparada de forma semelhante a B foi marcada com DAPI e observada em fluorescência, revelando a monocamada de células da serosa fixadas aderidas à superfície interna da casca. (D) Plotagem de vulcão. Os valores de p moderados transformados por log10 negativos estão plotados contra  $\log_2\text{Fold}$  (Nível de Enriquecimento na Serosa) para todos os genes do mosquito. A codificação de cores é usada para destacar as sondas de correspondência de genes com alteração de  $\log_2\text{Fold}$  acima de 1 (ou seja, fora da escala log, genes que são no mínimo duas vezes mais expresso na serosa do que no embrião). A linha vermelha indica o limiar de valor p moderado usado para selecionar genes estatisticamente significativos mais expressos na serosa (pontos vermelhos) (Adaptado de Goltsev *et al.*, 2009).

Além disso, também foi realizada uma série temporal de análises com chips de microarranjo com ovos intactos, frescos (não fixados). Foram utilizados 18 pontos temporais, de 2h às 46h durante o desenvolvimento do ovo, perfazendo toda a embriogênese do *A. gambiae*. Esse experimento de expressão temporal dos genes ao longo da embriogênese mostrou que diversos genes, dentro dos genes diferencialmente mais expressos na serosa (Figura 6D, pontos vermelhos), possuem dois períodos de elevada expressão (Figura 7), correspondendo à síntese da cutícula serosa, no início da embriogênese e a síntese da cutícula da larva, ao final da embriogênese. Os autores propuseram que existe uma "bateria de genes cuticulares" que seriam empregados todas as vezes que uma cutícula tenha que ser produzida,

quer seja uma cutícula da serosa, da larva, da pupa ou do adulto. Corroborando essa hipótese, foi observado que, enquanto esses genes possuem expressão bifásica durante a embriogênese de *A. gambiae*, os ortólogos desses mesmos genes possuem apenas um período de elevada expressão em *D. melanogaster*, correspondendo ao final da embriogênese, quando a cutícula larval dessa mosca é produzida (Figura 7) (Rezende 2008, Goltsev *et al.*, 2009).



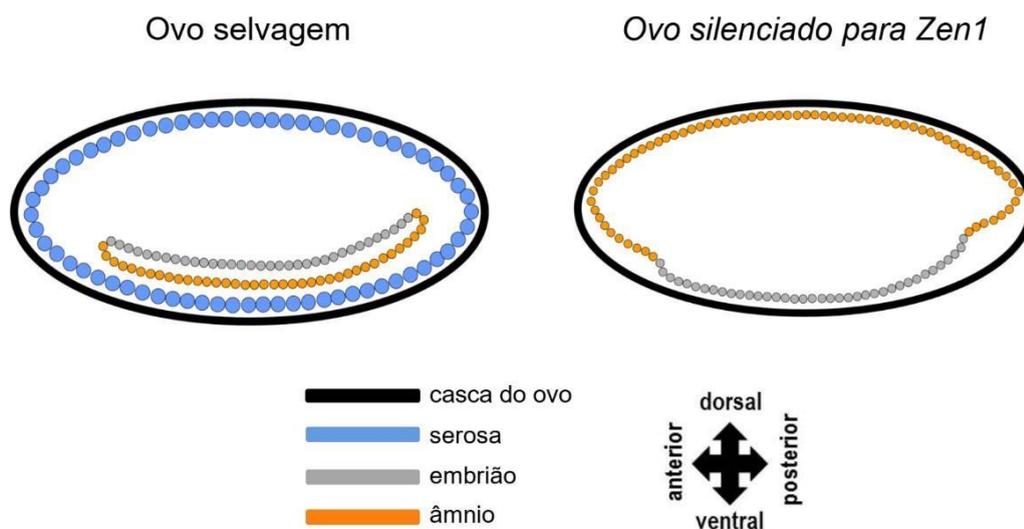
**Figura 7. Perfil de expressão de genes elongase ao longo da embriogênese de *D. melanogaster* e *A. gambiae*.** O tempo total de embriogênese de *D. melanogaster* e *A. gambiae* é de 24 e 48 horas, respectivamente. A elongase é um gene relacionado com a produção do exoesqueleto de insetos adultos. Na mosca só ocorre um período de elevada expressão ao final da embriogênese, relacionado com a síntese da cutícula da larva. No mosquito, além dessa expressão tardia, também ocorre um período de elevada expressão no início da embriogênese, relacionado com a síntese da cutícula serosa. Figura criada por Gustavo Rezende a partir de dados de expressão obtidos de Hooper *et al.*, 2007 e Goltsev *et al.*, 2009.

Também foram realizadas técnicas independentes para comprovar os achados descritos acima: hibridizações *in situ* foram realizadas para confirmar que um grupo seletivo de genes é de fato expresso na serosa e experimentos de PCR quantitativo em tempo real (qPCR) foram realizados para comprovar o perfil de expressão de alguns genes ao longo da embriogênese, confirmando os dados obtidos com as análises temporais de microarranjo. Especificamente, confirmou-se a expressão na serosa do gene de quitina sintase 1 (que produz quitina), serpentine (relacionado com a organização das microfibras de quitina), dopa descarboxilase e tirosina hidroxilase (relacionados com vias de melanização e esclerotização) e elongase (envolvido na síntese de ácidos graxos de cadeia muito longa). Outros genes também encontrados expressos na serosa estão envolvidos no processo de apoptose, e também ocorrem serinas peptidases, inibidores de serina protease, fatores de transcrição e citocromos

(Goltsev *et al.*, 2009). Apesar dessas informações estarem disponíveis, esse transcriptoma publicado em 2009 nunca foi analisado de forma sistemática ou comparativa: não existe outro trabalho parecido e nem que faça um contraponto a esse artigo.

#### 1.2.4 A serosa e a cutícula serosa de *T. castaneum*

Nesse besouro a formação da cutícula serosa acontece entre 8 e 14 horas após a oviposição, com aproximadamente 14% de embriogênese (Vargas, 2019). Ovos de *T. castaneum* silenciados para o fator de transcrição *Zen1* através de RNAi não possuem serosa pois *Zen1* é um fator de transcrição responsável pela especificação destas células. Na ausência de *Zen1*, posto que o ovo não desenvolve serosa o âmnio toma o seu lugar (van der Zee *et al.*, 2005) (Figura 8). Esses ovos sem serosa também não possuem cutícula serosa, são viáveis em condições de alta umidade relativa, mas deixam de sobreviver em baixa umidade relativa. Essa é uma demonstração direta do papel da cutícula serosa em proteger o ovo contra a perda de água (Jacobs *et al.*, 2013).



**Figura 8. Importância do fator de transcrição *Zen1* para a formação da serosa.** Ambos os ovos estão no estágio de início da extensão da banda germinal (vide Figura 3). Enquanto o ovo selvagem possui serosa o ovo silenciado para *Zen1* não possui serosa e o âmnio ocupa o lugar desta (Figura cedida por Gustavo Rezende, adaptada de van der Zee *et al.*, 2005 e Jacobs *et al.*, 2013).

Experimentos de silenciamento gênico através de RNAi foram realizados com os genes *Chitin synthase1* (*Chs1*), *Knickkopf1* (*knk1*), *Retroactive* (*rtv*) e *Laccase 2* (*Lac2*), sendo os três primeiros importantes para a organização das lamelas de quitina e o último para melanização, esclerotização e formação do córion (Moussian *et al.*, 2005; Chaudhari *et al.*,

2013; Rezende *et al.*, 2016). Na ausência de *Chs1*, a cutícula serosa se mantém, mas é carente de quitina e a endocutícula serosa torna-se uma camada amorfa, possivelmente composta somente por proteínas estruturais (Jacobs *et al.*, 2013). O silenciamento de *knk1* e *rtv* afetam severamente a estrutura laminar da endocutícula serosa que se torna amorfa, com lamelas desorganizadas, enquanto o silenciamento de *Lac2* aparentemente não afeta a estrutura da cutícula serosa (Jacobs *et al.*, 2015; Chaudari *et al.*, 2015). O silenciamento dos três genes relacionados com o metabolismo de quitina e, curiosamente, de *Lac2* também, acarretam um aumento da mortalidade dos ovos em baixas umidades, ainda que com níveis distintos (Jacobs *et al.*, 2013; Jacobs *et al.*, 2015), o que pode estar associado ao papel de cada gene.

No trabalho de Gurska *et al.* (2020) foi realizada uma análise transcriptômica para se investigar genes possivelmente expressos na serosa que sejam importantes para a produção da cutícula serosa. Os autores utilizaram ovos de *T. castaneum* selvagens e silenciados para *Zen 1*, que se desenvolvem normalmente em altas umidades, mas que não possuem serosa (Figura 8). Foi realizado um sequenciamento de RNA (RNAseq) em ovos de ambas as condições na etapa em que a serosa está expressando os genes necessários para a produção da cutícula serosa. Gurska e colaboradores observaram que 338 genes de *T. castaneum* são diferencialmente expressos quando *Zen1* é silenciado, sendo 331 genes menos expressos e 7 genes mais expressos nesse evento. Desta forma, se infere que muitos desses 331 genes serão expressos na serosa, visto que não há serosa em ovos silenciados para *Zen1*.

Foram realizados experimentos de hibridização *in situ* como uma técnica independente para confirmar que alguns dos genes menos expressos quando *Zen1* foi silenciado de fato são expressos apenas na serosa de *T. castaneum*. Ao todo 20 genes, dos 331, foram comprovadamente localizados na serosa (Vargas 2015, Gurska 2017, Rezende, dados não publicados).

### 1.3 O uso da Transcriptômica e técnicas utilizadas para obtenção dos genes relacionados a serosa do mosquito *A. gambiae* e do besouro *T. castaneum*

A Transcriptômica é o estudo do transcriptoma, do conjunto completo de todas as moléculas de ácido ribonucleico (RNA) expressas em uma célula, tecido ou organismo em um determinado momento e condição. Ou seja, a transcriptômica identifica e quantifica os RNAs expressos nos diferentes tecidos em distintas condições fisiológicas. A Reação em Cadeia da Polimerase quantitativa (qPCR, do Inglês *quantitative Polymeras Chain Reaction*), microarranjos e Sequenciamento de RNA (RNA-Seq.) são as técnicas mais usadas na Transcriptômica (Pacheco *et al.*, 2019). As técnicas de microarranjo e sequenciamento de RNA foram utilizadas em *A. gambiae* (Goltsev *et al.*, 2009) e *T. castaneum* (Gurska *et al.* 2020), respectivamente.

A técnica de microarranjo de DNA complementar (complementary DNA microarray) baseia-se na hibridização por complementaridade das moléculas de ácido nucleico, que ocorre entre a sonda depositada na lâmina de vidro e o seu RNAm correspondente, transformado em cDNA, extraído das amostras a serem analisadas e comparadas. Nessa técnica, duas amostras de cDNA (obtidas a partir de RNA), marcadas com fluorescências distintas, são avaliadas concomitantemente (teste vs. controle). De acordo com a fluorescência detectada, é possível aferir as concentrações relativas de transcritos nas amostras, sendo um sistema capaz de detectar a expressão de uma grande quantidade de genes em paralelo. Milhares de sondas para os genes de interesse são aderidas a pontos específicos na lâmina (um suporte sólido). Os microarranjos, ou chips de DNA, em alusão ao componente eletrônico que carrega milhões de transistores, são lâminas sólidas, nas quais fragmentos de DNA fita simples, denominados de sondas, são depositados e imobilizados de forma ordenada e em áreas específicas, chamadas de spots. Na lâmina, cada spot contém milhões de cópias de um único e determinado transcrito que posteriormente pode ser identificado. Com o uso de microarranjos, é possível avaliar os padrões de expressão gênica complexos e desenvolver sensores para o uso em diagnósticos clínicos (Pacheco *et al.*, 2019; Giachetto, 2010).

O sequenciamento de RNA (RNA-Seq) quantifica todos os transcritos expressos e é uma técnica moderna de biologia molecular que usa o sequenciamento profundo (*deep sequencing*) de DNA complementar (produzido a partir do RNA) para a quantificação da expressão gênica diferencial. Após o sequenciamento, as sequências elucidadas são mapeadas usando o genoma de referência, a avaliação da presença e da quantidade de cada RNA pode ser calculada e comparada às quantidades em outra amostra sequenciada. Com a utilização de RNA-Seq, é possível aferir a presença e a prevalência de transcritos conhecidos mas também o de transcritos previamente desconhecidos (Pacheco *et al.*, 2019). Essa é uma grande vantagem em relação à técnica de microarranjo, que utiliza sondas pré-fabricadas, ou seja, que implicam em um conhecimento prévio dos genes que são expressos. Essas técnicas (microarranjo e RNA-Seq) apresentam os resultados em fold-change (indicando quantas vezes a concentração do RNA aumentou ou diminuiu), e seus dados podem ser usados em estudos comparativos, por exemplo, analisando as mudanças na expressão gênica (Pacheco *et al.*, 2019).

#### 1.4 Relações de homologia gênica entre espécies

Para se realizar uma comparação apropriada entre o perfil de genes expressos em um mesmo tecido (serosa) em dois organismos diferentes (mosquito e besouro), é importante saber quais desses genes possuem algum tipo de homologia entre si. A relação de homologia

entre genes pode ser definida como dois genes distintos que possuem uma ancestralidade compartilhada. Essa ancestralidade compartilhada pode ser em relação a duas espécies distintas, ou em relação a um gene que se duplica dentro de uma mesma espécie. Em outras palavras, existem dois subtipos de homologia gênica: genes ortólogos e genes parálogos. A literatura costuma utilizar o valor mínimo de 30% de similaridade entre a sequência de duas proteínas distintas para considerá-las homólogas (Graur e Li, 2000, Kuzniar *et al.* 2008).

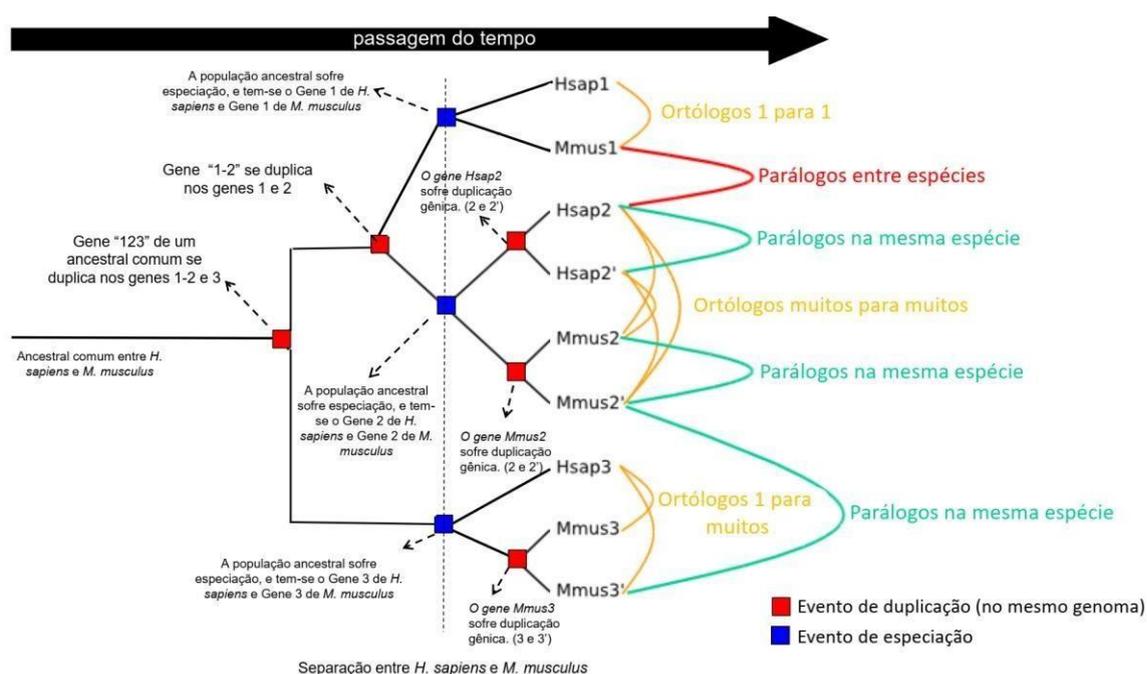
Genes ortólogos são definidos como genes homólogos derivados de um processo de especiação, a partir de um único gene presente no genoma do último ancestral comum das espécies que estão sendo comparadas (Graur e Li, 2000). Os genes ortólogos normalmente executam funções equivalentes em espécies próximas, porém a função desses genes também pode evoluir, ou seja, tornar-se diferente de sua função ancestral (Kuzniar *et al.*, 2008). A ortologia de genes entre duas espécies pode ser de três categorias diferentes: 1) Ortólogos 1 para 1 (*1-to-1*): quando apenas uma cópia do gene é encontrada em cada espécie; 2) Ortólogos 1 para muitos (*1-to-many*): quando um gene em uma espécie é ortólogo de vários genes em outra espécie; 3) Ortólogos muitos para muitos (*many-to-many*): quando vários ortólogos são encontrados em ambas as espécies (Figura 9) (Ensembl, 2023, Kuzniar *et al.*, 2008).

Genes parálogos são definidos como genes homólogos derivados de um evento de duplicação gênica a partir de um único gene. As relações de paralogia ocorrem tanto intra-espécies quanto inter-espécies. Os parálogos dentro de uma mesma espécie são definidos como genes homólogos que divergiram por um evento de duplicação gênica dentro de um mesmo genoma, quando um gene ancestral é duplicado no genoma de um organismo, resultando em duas ou mais cópias do mesmo gene. Dentro de uma espécie, um dos parálogos pode acumular mutações e adquirir novas funções ao longo da evolução, sendo uma das principais fontes de diversidade funcional. Ainda assim, algumas vezes essas novas funções podem ser biologicamente relacionadas à função ancestral (Gabaldón e Koonin, 2013, Graur e Li, 2000, Hurles 2004, Ensembl, 2023). Os parálogos entre espécies, ocorrem quando há uma relação entre genes de espécies diferentes em que o nó ancestral foi rotulado como um nó de duplicação, como exemplificado na Figura 9 (relação entre Mmus1 e Hsap2) (Graur e Li, 2000, Ensembl, 2023).

A existência de famílias gênicas é uma evidência de que a duplicação desempenha um papel importante na evolução de novas funções gênicas (Hurles M, 2004). Sobre o conceito de família gênica, e de acordo com o critério de Dayhoff (1978), considera-se que proteínas que apresentam pelo menos 50% de similaridade entre si, em termos de resíduos de aminoácidos, devem ser consideradas membros de uma família, enquanto proteínas

homólogas com menos de 50% de similaridade devem ser consideradas membros de uma super família (Graur e Li, 2000).

Existem diversos procedimentos para avaliar relações de ortologia e paralogia gênica entre espécies. Alguns destes procedimentos empregam comparação de sequências em pares (*pairwise*), como o BLAST, enquanto outros procedimentos empregam inferências feitas a partir de alinhamento de sequências múltiplas e construção de árvores filogenéticas. Outros procedimentos empregam uma abordagem híbrida (usando tanto a informação vinda da comparação de sequências em pares quanto vinda da construção de árvores filogenéticas), como é o caso da ferramenta Pan Taxonomic Compara, do Ensembl (Kuzniar *et al.* 2008, Gabaldón e Koonin, 2013).



**Figura 9. Diferentes relações de homologia entre genes exemplificado em *Homo sapiens* e *Mus musculus*.** Em um ancestral comum um gene chamado "123" sofre duplicação gênica gerando o gene "1-2" e o gene "3". Um segundo evento de duplicação faz com que o gene "1-2" dê origem aos genes "1" e ao gene "2". Em seguida ocorreu um evento de especiação na população desse ancestral comum, gerando as espécies *H. sapiens* e *M. musculus* e seus respectivos genes: Hsap1, Hsap2, Hsap3 e Mmus 1, Mmus 2, Mmus 3. Depois, os genes Hsap2, Mmus2 e Mmus3 sofrem duplicação gênica, originando respectivamente Hsap2', Mmus2' e Mmus3'. Após três eventos de duplicação e um evento de especiação, diversas relações de homologia podem ser observadas: Ortólogos 1 para 1 (1:1), Ortólogos muitos para muitos e Ortólogos 1 para muitos, além de Parálogos entre espécies e Parálogos na mesma espécie. Imagem adaptada da primeira figura presente na página [https://www.ensembl.org/info/genome/compara/homology\\_types.html](https://www.ensembl.org/info/genome/compara/homology_types.html).

Nesse sentido, o presente trabalho investiga quais genes estão presentes na serosa, durante a embriogênese inicial, tanto no mosquito *A. gambiae* quanto no besouro *T.*

*castaneum*. A hipótese de trabalho é a de que existem genes essenciais para a formação da cutícula serosa que são compartilhados entre ambas as espécies, ou seja, genes que estão conservados a centenas de milhões de anos. Os resultados poderão ajudar no entendimento da evolução de processos que ocorrem em ovos de insetos que possuem ecologias muito distintas e futuramente poderão auxiliar no estabelecimento de novas estratégias de controle de insetos vetores e insetos pragas.

## 2 OBJETIVOS

### 2.1 Objetivo geral

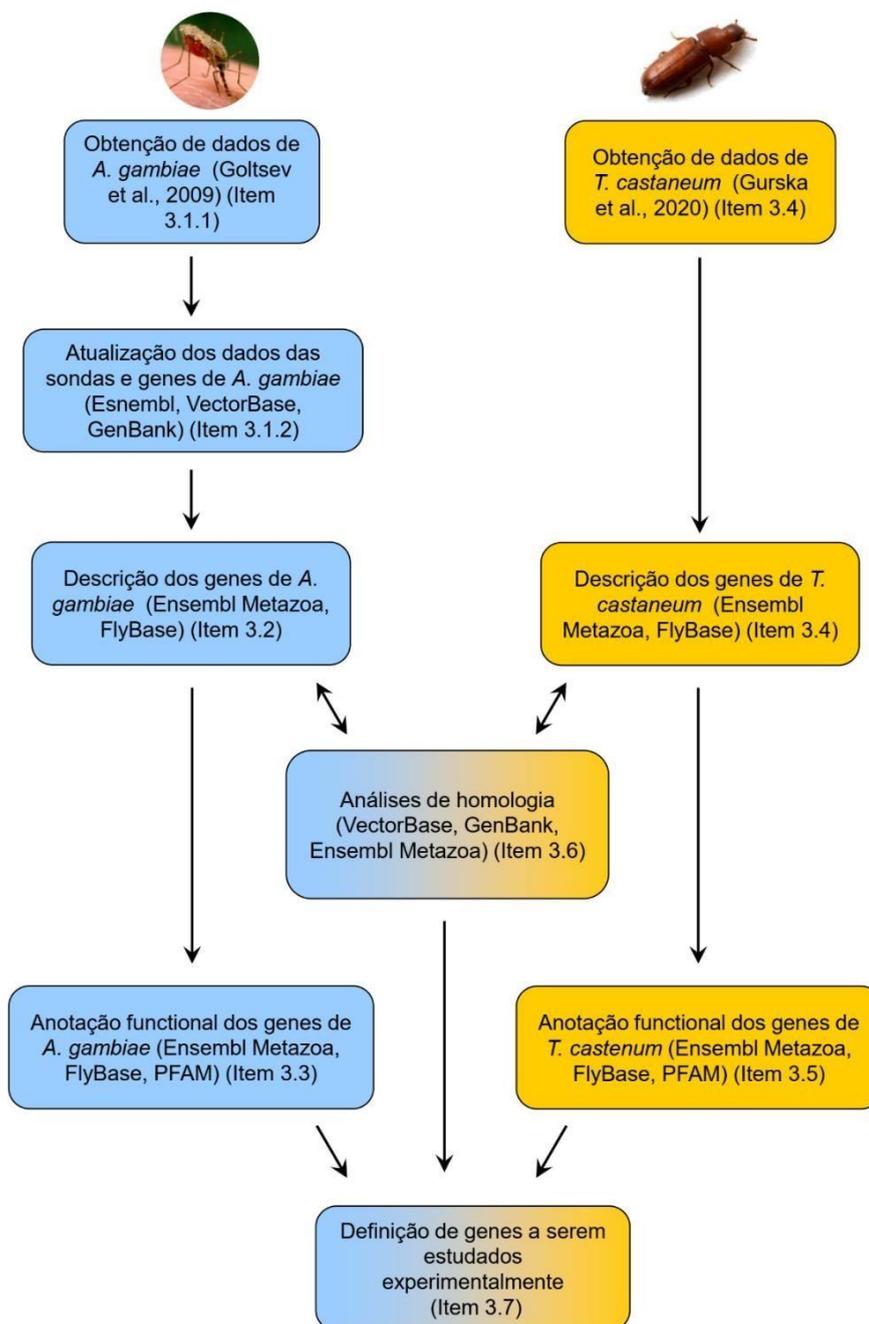
Analisar e comparar transcriptomas relacionados com as células da serosa do mosquito *A. gambiae* e do besouro *T. castaneum*, disponíveis em bancos públicos. Identificar genes homólogos relacionados com a produção da cutícula serosa.

### 2.2 Objetivos específicos

- Atualizar a lista de genes associados as sondas de microarranjo encontradas no transcriptoma da serosa de *A. gambiae*;
- Descrever os genes expressos na serosa de *A. gambiae*;
- Realizar a anotação funcional para os genes expressos na serosa de *A. gambiae*;
- Descrever os genes relacionados com a serosa de *T. castaneum*;
- Realizar a anotação funcional dos genes relacionados com a serosa de *T. castaneum*.
- Identificar quais homólogos são expressos em ambos os insetos.
- Selecionar genes candidatos a serem investigados experimentalmente.

### 3 METODOLOGIA

A Figura 10 apresenta um esquema resumindo toda a metodologia utilizada para realização desse trabalho, que é detalhada a seguir.



**Figura 10. Fluxograma das etapas metodológicas empregadas.** A esquerda, de azul, estão as etapas realizadas com *A. gambiae* e a direita, em laranja, as etapas realizadas com *T. castaneum*. Em cada caixa de texto está indicado o item correspondente da Metodologia.

### 3.1 Obtenção dos dados e identificação de genes de *Anopheles gambiae* diferencialmente expressos na serosa associados às sondas órfãs

#### 3.1.1 Obtenção dos dados:

As sequências de todas as sondas relacionadas ao genoma de *A. gambiae* presentes no microchip "GeneChip™ Plasmodium/Anopheles Genome Array" (Número de catálogo 900511) da Thermo Fisher/Affymetrix foram adquiridas na página do produto (<https://www.thermofisher.com/order/catalog/product/900511>), na parte de "Software & Data Analysis" => "Support Files" => "Sequence Files: Plasmodium\_Anopheles Probe Sequences, Tabular" que leva para o link [https://www.thermofisher.com/order/catalog/product/sec/assets?url=TFS-Assets/LSG/Support-Files/Plasmodium\\_Anopheles-probe\\_tab.zip](https://www.thermofisher.com/order/catalog/product/sec/assets?url=TFS-Assets/LSG/Support-Files/Plasmodium_Anopheles-probe_tab.zip). O arquivo baixado (Plasmodium\_Anopheles-probe\_tab.zip) contém a informação de todas as sondas, em formato tabular exportável para o programa Excel. Cada sonda deste microchip é constituída por 11 fragmentos, em que cada fragmento é composto por 25 nucleotídeos. Esses 11 fragmentos são descontínuos, apesar de estarem fisicamente próximos entre si. Em paralelo, baixou-se duas tabelas suplementares disponibilizadas no sítio da internet do artigo de Goltsev e colaboradores (2009) (<https://www.sciencedirect.com/science/article/pii/S0012160609001584>), a saber:

*Supplementary Table 2* (<https://ars.els-cdn.com/content/image/1-s2.0-S0012160609001584-mmc4.txt>) e *Supplementary Table 4* (<https://ars.els-cdn.com/content/image/1-s2.0-S0012160609001584-mmc6.txt>). A *Supplementary Table 2* contém a indicação das sondas que foram diferencialmente mais expressas na serosa, considerando o parâmetro determinado pelos autores, de valores de *Average Log2 Fold*  $\geq 1$  de expressão diferencial da serosa em relação ao resto do ovo (vide Introdução e pontos vermelhos na Figura 6D). Nessa mesma tabela encontram-se, quando possível, os dados do gene associado à sonda, do InterPro desse gene, bem como o nome do homólogo mais próximo em *D. melanogaster* desse gene. Já a *Supplementary Table 4* contém os dados das sondas/genes diferencialmente mais expressos na serosa em seu perfil de expressão temporal, com RNAs obtidos de ovos inteiros, ao longo dos 18 pontos temporais da embriogênese (vide Introdução e Figura 7).

Foram obtidas informações de outros artigos que citam as sondas consideradas órfãs (i.e. sem genes associados na *Supplementary Table 2*) em que nesses artigos essas sondas apresentam expressão diferencial com valor de *P-value* significativo. Essas informações foram obtidas utilizando a versão do "legacy Vector Base P4.12" que seria desativada em

pouco tempo. Essa etapa foi importante uma vez que não havia certeza de que essas informações seguiriam disponíveis dentro do VectorBase em sua nova versão repaginada.

Em seguida, entramos em contato com o Dr. Samuel Rund, pesquisador do VectorBase que nos forneceu dados atualizados do genoma de *A. gambiae*, utilizando a versão P4.12 *Release 47*. Essa nova versão, atualizada após 12 anos da publicação de Goltsev e colaboradores em 2009, mostrou que do total de 537 sondas, 466 sondas possuíam genes relacionados a elas e 71 sondas ainda eram órfãs. Além disso, foi observado que o número dos genes presentes no genoma versão P4.12 *Release 47* é de 13.057, sendo 11.213 genes associados a alguma sonda do microarranjo. O número de genes sem sondas associadas é de 1844, ou seja, 14,1% dos genes anotados no genoma versão P4.12 *Release 47* não fizeram parte da nossa análise, visto que não estão contemplados no chip de microarranjo empregado. Todas essas etapas acima foram realizadas em março de 2021.

### 3.1.2 Identificação dos genes de *A. gambiae*:

As sequências nucleotídicas das 71 sondas órfãs do arquivo "Plasmodium\_Anopheles.probe\_tab" foram extraídas utilizando o comando "findtsr" na linha de comando do *prompt* do DOS dentro do Sistema Operacional do Windows (Anexo I). A sequência de cada uma das sondas órfãs foi utilizada como *query* (consulta) para a realização de BLASTn contra o genoma do *A. gambiae* dentro da base de dados do *Ensembl Metazoa* (link de acesso: [https://metazoa.ensembl.org/Anopheles\\_gambiae/Info/Index](https://metazoa.ensembl.org/Anopheles_gambiae/Info/Index)). A página de resultados do BLASTn foi configurada para mostrar os genes anotados, a posição do alinhamento dos fragmentos da sonda e também as sequências de ESTs disponíveis. Dessa forma, foi possível observar graficamente a região do genoma associada a cada sonda orfã. Nos casos em que a região da sonda correspondeu a ESTs que eram contíguas a um gene conhecido, a sonda foi atribuída a esse gene. Nos casos em que a região da sonda não pode ser atribuída a algum gene, empregamos uma segunda abordagem: foi utilizado os dados de sequenciamento de RNA (RNAseq) presentes no VectorBase e a mesma lógica foi empregada: quando havia expressão de RNAseq na região da sonda que era contígua a algum gene, a sonda foi atribuída a esse gene. Nos casos em que sondas não foram atribuídas a nenhum gene nem com dados de EST e nem de RNAseq, foi utilizada uma terceira abordagem: a realização de BLASTx com a região da sonda empregada como *query* contra o banco de dados do GenBank (NCBI). Ao realizar o BLASTx, as sequências de nucleotídeos foram traduzidas nos 6 quadros de leitura possíveis para buscar e verificar similaridade com alguma proteína descrita.

Também realizou-se a observação da análise gráfica da expressão gênica associada as sondas consideradas órfãs a partir dos dados de microarranjo obtidos ao longo da embriogênese (Goltsev *et al.* 2009), vide Introdução e Figura 7, sendo necessário verificar a quantidade de genes envolvidos na análise dos dados obtidos na *Supplementary Table 2*, visto que a relação entre sondas e genes não era de 1:1 pois haviam genes que eram representados por mais de uma sonda. Para obter o número total de genes únicos, sem redundância, foi preciso retirar todos os genes que apareciam mais de uma vez na lista, ou seja, que se associavam a mais de uma sonda. As redundâncias foram removidas organizado o nome dos genes em ordem crescente e utilizando a fórmula de "Verdadeiro ou Falso" do Microsoft Excel. Foram mantidas as entradas que possuíam maior valor de expressão na serosa (AvgLog2Fold). Por exemplo, tanto a sonda "Ag.2R.322.0\_CDS\_at" (AvLog2Fold de 2,35) quanto a sonda "Ag.2R.322.0\_UTR\_at" (AvLog2Fold de 1,91) se referem ao mesmo gene: AGAP002685. Para esse gene AGAP002685 foi mantida apenas a representação da sonda "Ag.2R.322.0\_CDS\_at" já que esta apresentou o maior nível de expressão diferencial. Essas etapas foram realizadas entre abril e maio de 2021.

### **3.2 Descrição dos genes expressos na serosa de *Anopheles gambiae* e investigação dos genes sem descrição:**

A descrição dos genes expressos na serosa de *A. gambiae* foi realizada através do banco de dados do *Ensembl Metazoa*, buscando cada gene individualmente, em julho de 2021.

Em seguida, foi feita uma análise com os 20 genes mais expressos com descrição n/a (*not available*), ou seja, sem descrição e com função desconhecida no *Ensembl Metazoa*. Para verificar a presença de regiões transmembranares foi utilizado o programa TMHMM versão 2.0 (Krogh *et al.*, 2001) (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>). para investigar se as proteínas codificadas por estes 20 genes eram solúveis ou transmembranares. Foi verificado com auxílio da ferramenta BLASTp a homologia de cada uma das proteínas codificadas por estes 20 genes com sequências já anotadas em outros organismos que tenham descrição e/ou função conhecidas. Este BLASTp foi realizado no GenBank do NCBI, utilizando o banco de dados *non redundant* (nr) e os parâmetros padrões do algoritmo. Também fiz uma busca no banco de dados do PFAM.

### **3.3. Atribuição funcional dos genes expressos na serosa do *Anopheles gambiae*:**

A atribuição funcional dos genes foi realizada através dos dados originais das sondas usadas em Goltsev *et al.* (2009), já atualizados com a versão P4.12 *Release 47* do genoma e

identificação dos genes associados a sondas que eram previamente consideradas órfãs (ANEXO II), conforme descrito acima (Itens 3.1.1 e 3.1.2). Para atribuir funções aos genes expressos na serosa de *A. gambiae*, observou-se a descrição do gene no Ensembl, e utilizou-se o FlyBase para verificar a presença de ortólogos 1:1 em *D. melanogaster* e sua função, quando foi o caso. Tendo em vista que a *D. melanogaster* é o inseto mais conhecido geneticamente, a informação do ortólogo direto é útil. Quando não era possível obter o ortólogo direto, observou-se os genes de *D. melanogaster* homólogos ao gene de *A. gambiae*, anotando a principal função. Quando o FlyBase não forneceu nenhuma informação foi feita uma busca no PFAM, empregando a sequência da proteína predita em questão, para verificar a existência de algum domínio conhecido, o que também contribuiu na atribuição da função. Em paralelo, também foram vistos artigos obtidos no PubMed ou Google para se ter uma perspectiva da função do gene em questão.

#### **3.4. Descrição dos genes possivelmente relacionados com a serosa do *Tribolium castaneum*:**

Dados obtidos no trabalho de Gurska *et al.* (2020), mostram 311 genes diferencialmente menos expressos, quando silenciados para *Tc-Zen1*. Além destes, foi observado que mais sete genes também eram diferencialmente menos expressos, mas foram desconsiderados no artigo supracitado devido ao baixo suporte estatístico (Kristen Panfilio, comunicação pessoal). Optamos por incluir esses sete genes nas análises do presente trabalho, fornecendo assim um total de 318 genes diferencialmente menos expressos ao silenciar *Zen1*. Baixou-se a tabela "42003\_2020\_1250\_MOESM3\_ESM.xlsx" disponível em [https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-020-01250-3/MediaObjects/42003\\_2020\\_1250\\_MOESM3\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-020-01250-3/MediaObjects/42003_2020_1250_MOESM3_ESM.xlsx), nomeada como *Supplementary Data 1-4* em setembro de 2021. Foram utilizados os dados da aba "1A" da tabela para realização das análises, considerando os 331 transcritos com Log2Fold change de -4,814917736 a -1,002119087. Desses 331 transcritos houve 20 genes que se repetiram pois forneceram mais de um transcrito (por exemplo o gene TC002459 apareceu duas vezes, com o transcrito TC002459 -RA e com o transcrito TC002459-RB). Essas redundâncias foram removidas. A partir daí foi realizada a descrição desses genes utilizando os banco de dados *Ensembl Metazoa* ([https://metazoa.ensembl.org/Tribolium\\_castaneum/Info/Index](https://metazoa.ensembl.org/Tribolium_castaneum/Info/Index)) em outubro de 2021.

### 3.5 Atribuição funcional de genes possivelmente expressos na serosa do *Tribolium castaneum*:

A atribuição funcional dos genes de *T. castaneum* foi realizada no *iBeetleBase*(<https://ibeetle-base.uni-goettingen.de/search>) entre março e abril de 2023. O ortólogo em *D. melanogaster* presente no *iBeetleBase* foi indicado somente quando a relação era de 1:1. Quando não era possível obter o ortólogo direto, era anotado a principal função dos genes de *D. melanogaster* homólogos ao gene de *T. castaneum*. Quando ainda não era possível obter nenhuma informação realizamos uma busca no PFAM, empregando a sequência da proteína predita em questão, para anotação de algum domínio conhecido, se houvesse. Também foi observada as informações no Interpro (link de acesso: <https://www.ebi.ac.uk/interpro/entry/InterPro/>) e Uniprot (link de acesso: <https://www.uniprot.org/>), e feita uma busca com artigos obtidos no PubMed ou Google, para ser ter uma ideia da função do gene em questão.

### 3.6. Análise de genes considerados homólogos:

#### 3.6.1. BLASTp do *Tribolium castaneum* para o *Anopheles gambiae*:

A versão mais atual do proteoma de *T. castaneum* é a associada à nova versão da montagem do genoma, de nome Tcas5.2, e ao *Official Gene Set 3* (OGS3). Este proteoma foi obtido no formato FASTA, no arquivo “OGS3\_proteins.fasta” disponível no link [https://ibeetle-base.uni-goettingen.de/downloads/OGS3\\_proteins.fasta](https://ibeetle-base.uni-goettingen.de/downloads/OGS3_proteins.fasta). Foram extraídas as 338 sequências das proteínas preditas de interesse com o comando "grep" no bash do Sistema Operacional Linux (ANEXO III). Estas 338 sequências proteicas foram utilizadas como *query* em um BLASTp feito no VectorBase com as quais verificamos a existência de homólogos em *A. gambiae*. Foi necessário utilizar o banco de dados do VectorBase pois a identificação das proteínas preditas no genoma de *A. gambiae* no banco de dados do GenBank do NCBI estava desatualizada. Os parâmetros utilizados para a realização do BLASTp foram: E-value: 10, Max target sequences: 100, “no filter”. Das 338 proteínas analisadas, foram obtidos *best hits* em *A. gambiae*. E a partir de cada *best hit* obtido, foram utilizados, para a consideração de genes homólogos entre as duas espécies, os seguintes parâmetros: cobertura (*coverage* ou tamanho) das proteínas  $\geq 50\%$ , porcentagem de alinhamento  $\geq 40\%$  e porcentagem de similaridade entre proteínas  $\geq 30\%$ . Com esses parâmetros foram considerados que 243 proteínas de *T. castaneum*, das 338 proteínas obtidas como diferencialmente expressas quando *Zen1* foi silenciado (Gurska *et al.*, 2020), possuíam homólogos no genoma de *A. gambiae*.

Essas etapas foram realizadas em setembro e outubro de 2021.

### 3.6.2. BLASTp recíproco do *Anopheles gambiae* para o *Tribolium castaneum*:

Com as proteínas de *A. gambiae* obtidas como *best hit* para cada uma das 243 proteínas de *T. castaneum*, foi realizado um BLASTp recíproco (que é um procedimento padrão para verificar a existência de ortólogos entre espécies). Para isso obtivemos o proteoma do *A. gambiae* em sua versão 47 do VectorBase no formato FASTA, arquivo "AgamProteom47.fasta", disponível no link [https://vectorbase.org/common/downloads/release-47/AgambiaePEST/fasta/data/VectorBase-47\\_AgambiaePEST\\_AnnotatedProteins.fasta](https://vectorbase.org/common/downloads/release-47/AgambiaePEST/fasta/data/VectorBase-47_AgambiaePEST_AnnotatedProteins.fasta). Em seguida as sequências das proteínas obtidas foram dispostas em apenas uma linha através do comando "awk" (Anexo III) seguido do comando "grep" (Anexo III), para extrair as 243 sequências de interesse, ambos na linha de comando do Sistema Operacional Linux. O BLASTp recíproco (de *A. gambiae* para *T. castaneum*), empregando cada uma das 243 proteínas como *query*, foi realizado no GenBank do NCBI e foram empregados os mesmos parâmetros do primeiro BLASTp (de *T. castaneum* para *A. gambiae*). Essas etapas foram realizadas entre março e junho de 2022.

A ortologia dos genes foi confirmada no *Ensembl Metazoa - Release 52*, de agosto a dezembro de 2022. Buscou-se um determinado gene de *A. gambiae* e foi verificado os ortólogos desse gene dentro da ferramenta *Pan Taxonomic Compara* do *Ensembl Metazoa*.

### 3.7. Análise de genes considerados parálogos entre espécies:

Quando o BLASTp recíproco forneceu como resultado (*subject*) de *best hit* uma proteína de *T. castaneum* diferente da proteína de busca empregada como *query* no primeiro BLASTp (de *T. castaneum* para *A. gambiae*), considerou-se que nesses casos podem existir famílias gênicas e, portanto, parálogos entre espécies. A existência de parálogos entre espécies foi confirmada com a ferramenta *Gene Tree* dentro do *Pan Taxonomic Compara* do *Ensembl Metazoa*. Foram obtidos os dados do *Gene Tree* para cada um dos genes codificantes das proteínas de *A. gambiae* que não apresentaram ortólogos em *T. castaneum* e buscou-se a presença, dentro dos genes homólogos, dos genes de *T. castaneum* empregados como *query* no primeiro BLASTp (de *T. castaneum* para *A. gambiae*).

### 3.8. Designação de genes a serem estudados experimentalmente no futuro:

Nesta etapa foi gerado uma lista de genes candidatos a serem estudados em *T. castaneum*. Os parâmetros utilizados para criar a lista foram:

- Genes ortólogos entre as duas espécies com perfil de expressão bifásica no microarranjo temporal de *A. gambiae* ou com uma fase de maior expressão no período de formação da cutícula serosa no mosquito.
- Genes parálogos entre espécies com perfil de expressão bifásico ou com uma fase de maior expressão no período de formação da cutícula serosa em *A. gambiae*.
- Genes desconhecidos com perfil de expressão bifásica ou com uma fase de maior expressão no período de formação da cutícula serosa em *A. gambiae*.
- Genes conhecidos que em *A. gambiae* possuem expressão bifásica ou uma fase de maior expressão no período de formação da cutícula serosa.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Identificação de genes de *Anopheles gambiae* associados a sondas consideradas órfãs:

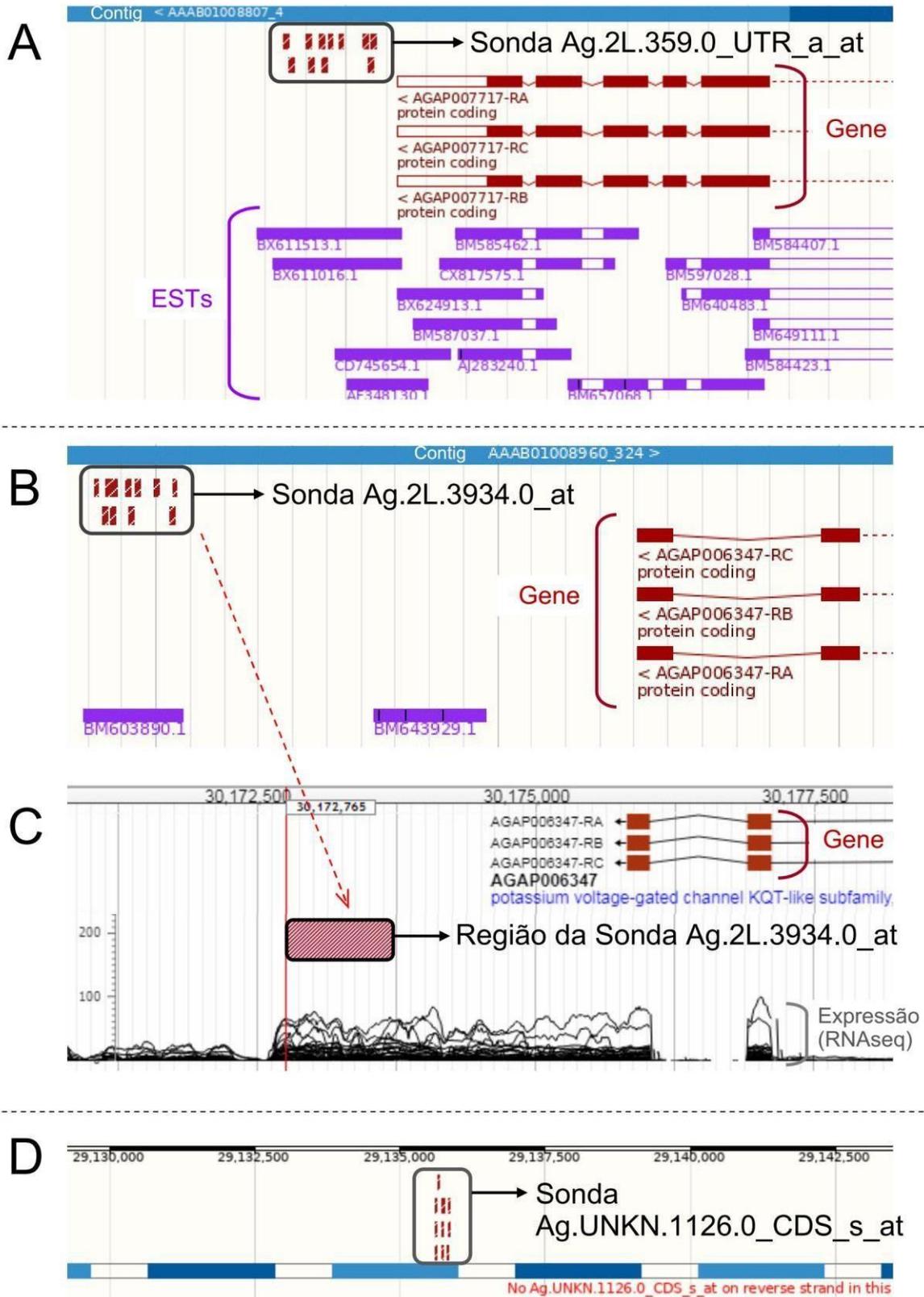
As análises foram iniciadas com os dados obtidos por microarranjo e disponíveis publicamente (Goltsev *et al.* 2009). Nesse trabalho, havia então um total de 537 sondas diferencialmente expressas na serosa, sendo 372 sondas atribuídas a 359 genes e 165 sem gene associado sendo consideradas sondas órfãs, ou seja, 30% das sondas eram órfãs. A relação entre sondas e genes não é de 1:1 devido a duas situações encontradas em alguns casos: i) uma sonda se associando a mais de um gene e ii) um gene que é detectado por mais de uma sonda. Quando os dados das 537 sondas foram atualizados em 2021, utilizando a versão P4.12 "Release 47" do genoma do *A. gambiae* no VectorBase, 466 sondas foram atribuídas a 486 genes e o número de sondas órfãs diminuiu para 71 (13% do total de sondas).

Foram realizadas análises com essas 71 sondas órfãs buscando a localização das mesmas no genoma através de BLASTn e em seguida utilizando dados de ESTs do Ensembl Metazoa ou dados de RNAseq do VectorBase e em alguns casos, através de BLASTx (Figura

11). Por exemplo, os nucleotídeos correspondentes a sonda Ag.2L.359.0\_UTR\_a\_at se localizam na região de ESTs que são contínuos ao gene AGAP007717 (Figura 11A). Em outro exemplo, o local da sonda Ag.2L.3934.0\_at não apresentou nenhum EST (Figura 11B),

mas no VectorBase essa região possui expressão de RNAseq que é contígua ao gene AGAP006347 (Figura 11C). Em outro caso, a sonda Ag.UNKN.1126.0\_CDS\_s\_at se localiza em uma região do genoma que não possui genes descritos (Figura 11D). Foi realizado um BLASTx dessa região "deserta" que demonstrou que a região da sonda Ag.UNKN.1126.0\_CDS\_s\_at codifica uma proteína *rolling stone-like*, descrita no genoma de outros anofelinos (Figura 12). Com isso, entendemos que esse gene está presente no genoma do *A. gambiae*, mas que ainda não foi anotado como tal.

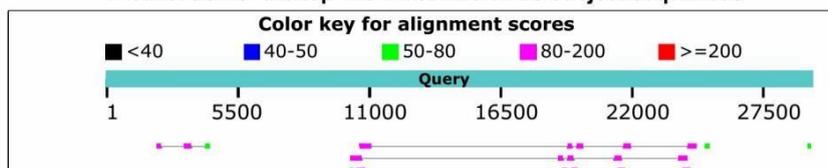
Através dessa abordagem foi possível identificar os genes associados a 41 sondas que até então eram órfãs (Tabela 1): Das 41 sondas, 22, 16 e 3 atribuições de genes às sondas foram feitas a partir de dados de EST, RNAseq e BLASTx, respectivamente. Assim, das 537 sondas originais, apenas 30 (5,6%) não conseguiram ser atribuídas a algum gene. Foram obtidos alguns resultados prévios das regiões genômicas correspondentes a essas 30 sondas que ainda precisam ser investigados com mais detalhe (Anexo V).



**Figura 11. Análise das sondas órfãs.** A) Localização da sonda Ag.2L.359.0\_UTR\_a\_at que foi atribuída ao gene AGAP007717 a partir dos dados de ESTs presentes no *Ensembl Metazoa*. B) A localização da sonda Ag.2L.3934.0\_at não corresponde a ESTs que sejam contíguos a algum gene descrito. C) No VectorBase, a região da sonda Ag.2L.3934.0\_at foi atribuída ao gene AGAP006347 graças aos dados de RNAseq. D) Localização da sonda Ag.UNKN.1126.0\_CDS\_s\_at, onde não há gene descrito.

## Graphic Summary

Distribution of the top 126 Blast Hits on 33 subject sequences



## Descriptions

Sequences producing significant alignments:

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
protein rolling stone-like [Anopheles arabiensis]	192	801	4%	2e-52	97.85%	<a href="#">XP_040167250.1</a>
protein rolling stone-like [Anopheles merus]	191	799	4%	5e-52	96.77%	<a href="#">XP_041778347.1</a>

## Alignments

protein rolling stone-like [Anopheles arabiensis]

Sequence ID: **XP\_040167250.1** Length: 297 Number of Matches: 6

▶ See 6 more title(s)

Range 1: 205 to 297

Score	Expect	Method	Identities	Positives	Gaps	Frame
192 bits(488)	2e-52()	Compositional matrix adjust.	91/93(98%)	92/93(98%)	0/93(0%)	+2

Features:

Query	20777	SHRFGRPYIYNVLDWNEPGKAMVTVVGTTILLAMIVHMLMFAIYKLRVKIYLRYPNHKPII	20956
		+ RFGRPYIYNVLDWNEPGKAMVTVVGTTILLAMIVHMLMFAIYKLRVKIYLRYPNHKPII	
Sbjct	205	TDRFGRPYIYNVLDWNEPGKAMVTVVGTTILLAMIVHMLMFAIYKLRVKIYLRYPNHKPII	264
Query	20957	PTNSTLQLQTGKCDGISMVYGS DNKAFSISERY	21055
		PTNSTLQLQTGKCDGISMVYGS DNKAFSISERY	
Sbjct	265	PTNSTLQLQTGKCDGISMVYGS DNKAFSISERY	297

Range 2: 205 to 297

Score	Expect	Method	Identities	Positives	Gaps	Frame
192 bits(487)	2e-52()	Compositional matrix adjust.	91/93(98%)	92/93(98%)	0/93(0%)	+1

Features:

Query	23326	SHRFGRPYIYNVLDWNEPGKAMVTVVGTTILLAMIVHMLMFAIYKLRVKIYLRYPNHKPII	23505
		+ RFGRPYIYNVLDWNEPGKAMVTVVGTTILLAMIVHMLMFAIYKLRVKIYLRYPNHKPII	
Sbjct	205	TDRFGRPYIYNVLDWNEPGKAMVTVVGTTILLAMIVHMLMFAIYKLRVKIYLRYPNHKPII	264
Query	23506	PTNSTLQLQTGKCDGISMVYGS DNKAFSISERY	23604
		PTNSTLQLQTGKCDGISMVYGS DNKAFSISERY	
Sbjct	265	PTNSTLQLQTGKCDGISMVYGS DNKAFSISERY	297

**Figura 12. Resultados do BLASTx da região da sonda Ag.UNKN.1126.0\_CDS\_s\_at.** Os nucleotídeos da região associada à sonda Ag.UNKN.1126.0\_CDS\_s\_at foram utilizados como *query* para um BLASTx no banco de dados "nr" do *GenBank* do NCBI. O resultado apresenta algumas das espécies de anofelinos que possuem proteínas preditas que produzem alinhamentos significativos

**Tabela 1. Lista das 41 sondas que deixaram de ser órfãs.** AvLog2fold: nível de expressão dessas sondas em escala Log na serosa, em relação ao embrião. Probe Set ID: código de cada sonda. ID do gene: código de cada gene, quando possível.

AvLog2Fold	Probe Set ID	ID do gene	Evidência
4,317	Ag.2R.3767.0_at	AGAP004373	EST
4,246	Ag.2R.33.1_at	AGAP029769	EST
4,154	Ag.2L.3934.0_at	AGAP006347	RNAseq
4,142	Ag.2L.359.0_UTR_a_at	AGAP007717	EST
3,990	Ag.UNKN.1126.0_CDS_s_at	Protein Rolling-stone like	BLASTx
3,614	Ag.2L.391.0_UTR_at	AGAP005170	EST
3,484	Ag.3R.194.0_UTR_at	AGAP007781	EST
3,210	Ag.2R.47.0_UTR_at	AGAP001979	EST
3,033	Ag.X.153.0_UTR_s_at	AGAP000011	EST
2,987	Ag.2L.2940.0_at	AGAP007674	EST
2,980	Ag.2R.550.0_UTR_a_at	AGAP003141	EST
2,859	Ag.2R.274.0_UTR_a_at	AGAP002358	EST
2,718	Ag.2R.463.1_CDS_at	AGAP004631	EST
2,493	Ag.2L.402.0_UTR_a_at	AGAP005091	EST
2,396	Ag.2R.3734.0_at	AGAP004071	RNAseq
2,360	Ag.3R.403.0_UTR_at	AGAP008821	EST
2,225	Ag.2L.4032.0_at	AGAP007065	RNAseq
2,195	Ag.2R.289.15_UTR_at	AGAP001797	EST
2,087	Ag.2L.355.0_UTR_at	AGAP006901	RNAseq
2,032	Ag.X.11.0_UTR_at	AGAP000573	RNAseq
2,000	Ag.UNKN.1520.0_CDS_at	CLIP domain-containing serine protease 14D-like (AGAP003249-like)	BLASTx
1,948	Ag.2L.3261.0_at	AGAP005091	RNAseq
1,904	Ag.2R.464.0_UTR_at	AGAP001313	EST
1,817	Ag.3L.124.0_UTR_at	AGAP010957	RNAseq
1,711	Ag.2R.463.0_CDS_a_at	AGAP004631	RNAseq
1,706	Ag.2L.2212.0_CDS_at	AGAP007327	RNAseq
1,687	Ag.3L.245.0_UTR_a_at	AGAP010640	RNAseq
1,494	Ag.3R.567.0_UTR_a_at	AGAP007970	EST
1,488	Ag.2R.994.1_UTR_a_at	AGAP003492	EST
1,486	Ag.2R.749.0_UTR_a_at	AGAP001411	RNAseq
1,373	Ag.2R.352.0_UTR_a_at	AGAP003153	RNAseq
1,315	Ag.3R.3531.0_at	AGAP008636	EST
1,270	Ag.2R.4371.0_at	AGAP002644	EST
1,216	Ag.2L.303.0_UTR_a_at	AGAP005400	EST
1,136	Ag.UNKN.177.1_a_at	solute carrier family 15 member-like	BLASTx
1,136	Ag.2R.360.0_UTR_s_at	AGAP001548	RNAseq
1,108	Ag.2R.706.0_UTR_a_at	AGAP001520	EST
1,104	Ag.2L.2880.0_a_at	AGAP005091	RNAseq
1,082	Ag.3L.2205.0_at	AGAP011771	RNAseq
1,051	Ag.2R.5024.0_at	AGAP002898	RNAseq
1,024	Ag.2R.580.0_UTR_a_at	AGAP002428	EST

Holt et al. (2002) realizaram o primeiro sequenciamento do genoma de *A. gambiae* há mais de 20 anos, porém, o mesmo vem sendo atualizado até os dias atuais, tendo em vista que diversos genes ainda não foram anotados.

Ainda não se sabe qual é a função do gene “Rolling stone-like” em *A. gambiae* (Figura 12), apesar da função do gene Rolling stone ser definida. Rolling stone é uma proteína transmembranar envolvida na fusão de mioblastos e seu gene é expresso em células ectodérmicas e células mesodérmicas (Paululat *et al.*, 1997). Proteínas transmembranares são aquelas que ficam inseridas na membrana plasmática da célula ou na membrana da organela. A função básica dos mioblastos é a restauração da função do músculo esquelético. O gene TC005180 em *T. castaneum* é homólogo de Rolling stone, e já está descrito como tal no Ensembl, mas não está entre os 338 genes de *T. castaneum* estudados.

Encontramos em *A. gambiae* outros dois casos de possíveis genes não anotados: “solute carrier family 15 member like” e “clip domain serine protease”. *Solute carrier family 15 member like* é um tipo de transportador transmembranar, que transporta moléculas através da membrana celular e pode estar envolvido no transporte de substâncias importantes para as células da serosa. Já o *Clip domain serine protease* é uma protease, uma enzima responsável pela clivagem de proteínas em fragmentos menores. A presença de muitas dessas proteases na serosa de *A. gambiae* sugere que elas desempenham papéis importantes nesse tecido. Isso pode ter várias funções, como a ativação de outras proteínas ou degradação de proteínas indesejadas. Estudos adicionais são necessários para entender melhor as funções específicas dessas proteases na serosa.

#### **4.2 Descrição e anotação funcional dos genes expressos na serosa de *A. gambiae*:**

A primeira abordagem utilizada para descrever os genes de *A. gambiae* foi registrar a descrição disponível no *Ensembl*. Essa etapa foi necessária pois a descrição realizada em 2009 (Goltsev *et al.* 2009) pode ter sido atualizada e novos genes foram incluídos conforme descrito acima (item 4.1). Dessa listagem inicial (dados não mostrados), os 20 genes mais expressos considerados desconhecidos pelo Ensembl foram selecionados para uma análise mais detalhada, buscando ortólogos em *D. melanogaster*, avaliando a existência de domínios transmembrana, realizando BLASTp no *GenBank* e buscando domínios conhecidos no PFAM (Tabela 2). Constatamos que esses genes, que são considerados desconhecidos pelo Ensembl, na verdade são conhecidos quando estes outros bancos de dados foram utilizados. Por exemplo, AGAP001984 é uma proteína relacionada com hormônio juvenil, AGAP011936 é uma quitina deacetilase (Serpentine) e AGAP012946 é uma serina protease.

**Tabela 2. Análise dos 20 genes desconhecidos mais expressos na serosa de *A. gambiae*.** Informações obtidas através do FlyBase, TMHMM, BLASTp e PFAM.

AvLog2Fold Ser/Emb	ID do gene	Ortólogo em <i>Drosophila</i> 1:1	Descrição em <i>Drosophila melanogaster</i>	TMHMM	BlastP	Family id	Função
5,78	AGAP010799	not available	not available	solúvel	Proteína específica de anofelinos	Não encontrado	Desconhecida
5,41	AGAP001984	FBgn0038894	Juvenile hormone binding protein 16	solúvel	Presente em insetos, circadian clock-controlled protein daywake-like, related to Take Out	JHBP	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
5,07	AGAP011936	FBgn0260653	serpentine	solúvel	serpentine, chitin deacetylase 1	CBM_14	Metabolismo de quitina
4,87	AGAP002962	FBgn0030503	Transport and Golgi organization 2	solúvel	transport and Golgi organization protein 2 isoform X2	TANGO2	Organização de Golgi
4,85	AGAP000697	FBgn0052694	not available	solúvel	uncharacterized protein	Não encontrado	Desconhecida
4,67	AGAP006410	not available	not available	transmembranar	presente em insetos, one cut domain, keratin, type II cytoskeletal 2 epidermal isoform X2	Não encontrado	Desconhecida
4,63	AGAP001508	not available	not available	solúvel	hypothetical protein RP20, glycine-rich protein 5-like	Não encontrado	Desconhecida
4,60	AGAP000821	FBgn0052816	not available	solúvel	presente em insetos, uncharacterized protein	Não encontrado	Desconhecida
4,51	AGAP002085	FBgn0051997	not available	solúvel	presente em insetos, uncharacterized protein	SVWC	Miscelânea
4,44	AGAP029052	FBgn0261015	PFTAIRE-interacting factor 1A	solúvel	Putative leucine rich repeat containing protein DDB, restin homolog isoform	JAKMIP_CC3	Relacionado ao citoesqueleto
4,42	AGAP011778	not available	not available	solúvel	histone-lysine N-methyltransferase	Não encontrado	Desconhecida
4,39	AGAP012946	not available	not available	solúvel	serine protease snake like,	Trypsin	Desconhecida
4,38	AGAP009635	not available	not available	solúvel	Proteína específica de anofelinos, glycine rich RNA binding protein like	Não encontrado	Desconhecida
4,33	AGAP013005	not available	not available	solúvel	Presente em anofelinos e culex, glycine rich wall structural protein	Não encontrado	Desconhecida
4,29	AGAP010385	not available	not available	solúvel	probable serine/ threonine- proteinkinase	Não encontrado	Desconhecida
4,24	AGAP010441	not available	not available	solúvel	Uncharacterized mosquito protein	Não encontrado	Desconhecida
4,24	AGAP001718	not available	not available	solúvel	Uncharacterized mosquito protein	DUF4789	Desconhecida
4,21	AGAP006151	not available	not available	transmembranar	hypothetical protein ZHAS	Não encontrado	Desconhecida
4,15	AGAP003515	not available	not available	solúvel	omega amidase NIT2	CN_hydrolase	Metabolismo de Aminoácidos
4,14	AGAP007717	not available	not available	solúvel	uncharacterized protein LOC118513600, CAP isoform R	SH3_9	Desconhecida

Vale destacar que o gene mais expresso na serosa, AGAP010799, é um gene específico de anofelinos, totalmente desconhecido, o segundo mais expresso é um gene que codifica uma proteína de ligação ao Hormônio Juvenil e relacionada ao relógio circadiano (que o Ensembl todavia classifica como desconhecida), seguido de uma elongase e serpentine, genes conhecidos por serem importantes para os metabolismos de lipídeos e de quitina, respectivamente (Rezende *et al.* 2016, Vargas, 2019).

A partir desse resultado, foi feita uma análise detalhada, mas sem a avaliação do TMHMM, com todos os 472 genes associados às 507 sondas mais expressas na serosa. A descrição destes genes, bem como a anotação funcional de cada um deles está descrito no Anexo VI (páginas 95 a 112). Destes, 3 genes foram descritos, mas não estão anotados no genoma (vide item 4.1. e Tabela 1), 368 genes possuem descrição e função e 101 genes são desconhecidos. Destes 368 genes, é importante mencionar que 73 deles não possuem descrição no Ensembl, mas conseguimos atribuir funções para eles a partir de outros bancos de dados. Os genes foram agrupados de acordo com a função atribuída a cada um na Tabela 3.

**Tabela 3. Identificação dos genes expressos na serosa do mosquito *Anopheles gambiae* agrupados pelas funções atribuídas em ordem alfabética. "Miscelânea" são os genes para os quais não conseguimos atribuir funções.**

Função	Quantidade	ID dos genes
Adesão celular	1	AGAP009522, AGAP002428
Aldose redutase	1	AGAP003408
Aminotransferase, Transaminase	1	AGAP012404
Anidrase carbônica	1	AGAP002359
Apoptose	3	AGAP001411, AGAP011832 e AGAP009224
Associado ao Olfato	1	AGAP007751
Associado ao Retículo Endoplasmático	3	AGAP008615, AGAP001069, AGAP001548
Ativação da proliferação de hemócitos	1	AGAP000297
Ativação de Gamma-secretase	1	AGAP008636
Atividade de acyltransferase	4	AGAP010111, AGAP010118, AGAP010125, AGAP012821
Atividade de ligação de caderina e tubulina	1	AGAP003238
Atividade de Ligação de íons de zinco	1	AGAP009051
Atividade de peroxidase	1	AGAP008350
Atividade inibidora de peptidase	1	AGAP011460
Atividade inibidora de serina/treonina fosfatase	1	AGAP007886
Biossíntese de Isoprenóides, e Ubiquinona	2	AGAP000591, AGAP006015
Canal de potássio dependente de voltagem	1	AGAP006347
Canal de próton	1	AGAP013356
Carboxylesterase	1	AGAP001101
Catabolismo do RNAm	1	AGAP008181
Ciclo da Uréia	1	AGAP008141
Clivagem/Inativação de peptídeos	1	AGAP009791
Crescimento e proliferação celular	1	AGAP010131
CYP450	8	AGAP001864, AGAP002417, AGAP002419, AGAP012292, AGAP012295, AGAP012291, AGAP000500, AGAP006082.
Degradação de epóxidos	1	AGAP011972
Desconhecida	101	AGAP010799, AGAP000697, AGAP006410, AGAP001508, AGAP000821, AGAP011778, AGAP012946, AGAP009635, AGAP013005, AGAP01035, AGAP010441, AGAP001718, AGAP006151, AGAP007717, AGAP000222, AGAP000726, AGAP001206, AGAP001346, AGAP001510, AGAP001616, AGAP001719, AGAP001889, AGAP003513, AGAP003620, AGAP003746, AGAP003942, AGAP003960, AGAP004144, AGAP004814, AGAP004815, AGAP005093, AGAP005242, AGAP0055612, AGAP005888, AGAP005982, AGAP006063, AGAP006193, AGAP006398, AGAP006967, AGAP006972, AGAP006973, AGAP007077, AGAP007314, AGAP007372, AGAP007534, AGAP007564, AGAP007645, AGAP007754, AGAP008014, AGAP008403, AGAP008923, AGAP009106, AGAP009251, AGAP009252, AGAP009634, AGAP010066, AGAP010640, AGAP010805, AGAP011070, AGAP011119, AGAP011431, AGAP011432, AGAP011605, AGAP011641, AGAP011771, AGAP011958, AGAP012181, AGAP012270, AGAP012530, AGAP012531, AGAP012571, AGAP012613, AGAP012740, AGAP012926, AGAP012956, AGAP013060, AGAP013134, AGAP013365, AGAP013476, AGAP013529, AGAP028038, AGAP028049, AGAP028067, AGAP028095, AGAP028186, AGAP028227, AGAP028402, AGAP029106, AGAP029117, AGAP029133, AGAP029284, AGAP029292, AGAP029365, AGAP029378, AGAP029419, AGAP029560, AGAP029607, AGAP029611, AGAP029673, AGAP029747, AGAP029750
Desenvolvimento do sistema nervoso	1	AGAP003141

Desenvolvimento muscular	3	AGAP000769, AGAP006901, AGAP010784
Domínio PDZ	1	AGAP005235
Exocitose	1	AGAP007721
Fator de transcrição	8	AGAP000237, AGAP005137, AGAP007327, AGAP011655, AGAP007044, AGAP006386, AGAP007065, AGAP002036
Formação de axônios	1	AGAP011334
Gene responsivo a Ecdisona	1	AGAP001751
Glucuronidação, transferência de ácido glucurônico	1	AGAP009403
Inibidor de Serina Protease	7	AGAP001377, AGAP009213, AGAP001376, AGAP009670, AGAP007693, Regulador negativo da melanização: AGAP006911 e AGAP003194
Ligação da cromatina	1	AGAP000707
Ligação de quitina	1	AGAP001203
Metabolismo de acetil-CoA	1	AGAP005847
Metabolismo de adenosina	1	AGAP000380
Metabolismo de Aminoácidos	15	AGAP000011, AGAP002425, AGAP013231, AGAP003515, AGAP007730, AGAP012662, AGAP010928, AGAP006023, De asparagina, de glutamina, de oxaloacetato: AGAP003516 e AGAP012801, Estoque de aminoácidos: AGAP001659, Processo catabólico de L-fenilalanina: AGAP009609 e AGAP000327, Glutaciona: AGAP011172, Cisteína: AGAP005153
Metabolismo de Carboidratos	13	AGAP004397, AGAP006430, AGAP007412, AGAP008844, AGAP010769, AGAP007589, AGAP000241, AGAP001449, AGAP013329, AGAP005552, AGAP004451, Açúcar: AGAP007920, Interconversão de malato e oxaloacetato: AGAP009510
Metabolismo de compostos endógenos e xenobióticos.	2	AGAP007121, AGAP028055
Metabolismo de Ferro	1	AGAP007258
Metabolismo de Glicogênio	1	AGAP001520
Metabolismo de Glutamina	1	AGAP008988
Metabolismo de glutaciona	10	AGAP004164, AGAP004383, AGAP009194, AGAP009197, AGAP010404, AGAP000947, AGAP002898, AGAP000165, AGAP000163, AGAP029495.
Metabolismo de Lipídeos	52	AGAP005175, AGAP008544, AGAP007358, AGAP000519, AGAP000275, AGAP012953, AGAP009282 AGAP005556, AGAP003514, AGAP004450, AGAP009104, AGAP029668, AGAP002644, AGAP012746, AGAP005091, AGAP011810, AGAP007881, AGAP010714, AGAP004532, AGAP000546, AGAP029758, AGAP004372, AGAP001763, AGAP005515, AGAP002279, AGAP004784, AGAP007593, AGAP003687, AGAP007354, AGAP006395, AGAP000180, AGAP007921, AGAP003814, AGAP011357, AGAP005553, AGAP011812, AGAP003600, AGAP004373, AGAP003606, AGAP011992, AGAP029445, Ácidos graxos: AGAP013071, Degradação de ácidos graxos: AGAP008667, e Metabolismo energético: AGAP007636, Ativação de esfingolipídios: AGAP001082, Biossíntese de Esteróis: AGAP002113, ligação lipídica: AGAP009281, Proteína de ligação ao Oxisterol: AGAP010893, transporte de colesterol: AGAP008137 e AGAP002857, Vitamina E: AGAP003733, Reação de Oxirredução: AGAP008570
Metabolismo de Pirimidina	2	AGAP003968, Degradação: AGAP001021
Metabolismo de Purina	3	AGAP000577, AGAP009648, AGAP006906
Metabolismo de Quitina	8	AGAP004500, AGAP011936, AGAP002457, AGAP000359, AGAP013260; Desenvolvimento de cutícula: AGAP005634, AGAP006898 e AGAP011937
Metabolismo de Ribossomo, putativo	1	AGAP029442
Metabolismo de tirosina	3	AGAP004802, AGAP005865, AGAP002438
Metabolismo energético	1	AGAP029610

Migração das células germinativas	1	AGAP007104
Miscelânea	18	AGAP009651, AGAP002738, AGAP001730, AGAP006615, AGAP005400, AGAP003487, AGAP002085, AGAP003738, AGAP005170, AGAP007384, AGAP028649, AGAP007140, AGAP012696, AGAP029783, AGAP007209, AGAP004861, AGAP000311, AGAP010733, AGAP000009
Organização de Golgi	1	AGAP002962
Organização do Retículo Endoplasmático	1	AGAP007940
Oxidação de aldeídos	2	AGAP005124, AGAP003578
Palmitoilação	2	AGAP008321, AGAP010164
Produção de cAMP	1	AGAP010436
Produção de peróxido de hidrogênio, formação da casca do ovo	1	AGAP009978
Protease/Proteólise	25	AGAP000573, AGAP002270, AGAP003057, AGAP003251, AGAP003252, AGAP008835, AGAP009214, AGAP011780, AGAP011781, AGAP011789, AGAP011792, AGAP029769, AGAP011788, AGAP004860, AGAP004859, AGAP004567, AGAP000315, AGAP029770, CLIP domain-containing serine protease 14D-like-AGAP003249-like - sem ID- obtido por BLASTx, Envolvido na ativação da cascata de melanização:AGAP000572; Metaloprotease: AGAP007982; Miscelânea: AGAP001979; tripsina: AGAP003627; via de Toll: AGAP003250 e AGAP004148
Proteína cuticular	18	AGAP000344, AGAP006867, AGAP000820, AGAP000987, AGAP000988, AGAP000989, AGAP003308, AGAP009790, AGAP028178, AGAP028156, AGAP028137, AGAP028160, AGAP012676, AGAP012356, AGAP007980, AGAP006007, AGAP010369, AGAP003334.
Proteína G	11	AGAP009953, AGAP006521, AGAP007452, AGAP005976, AGAP006056, AGAP010189, AGAP008419, AGAP001617, AGAP004769, AGAP000243, AGAP009739.
Proteína quinase	3	AGAP029459; Receptor de membrana: AGAP008813; Relacionado ao Citoesqueleto: AGAP007587.
Proteína quimiossensorial	7	AGAP005716, AGAP008059, AGAP001189, AGAP007281, AGAP000278, AGAP008051, AGAP008054
Proteína Rolling-stone like, função indefinida	1	Protein Rolling-stone like, a confirmar experimentalmente – sem ID- obtido por BLASTx, AGAP009583.
Reação de Oxirredução	4	AGAP002551, AGAP002552, AGAP003784, AGAP010696
Regulação da contração muscular	1	AGAP001797
Regulação da expressão gênica	1	AGAP008794
Regulação da tradução	1	AGAP007781
Regulação da transcrição	5	AGAP029650, AGAP000410, AGAP005311, AGAP010623, Via de Notch e Ras: AGAP004795.
Relacionado a Autofagia	2	AGAP002685, AGAP007970
Relacionado a oviposição	1	AGAP029130
Relacionado ao Citoesqueleto	13	AGAP007674, AGAP003790, AGAP003263, AGAP004071, AGAP000035, AGAP010445, AGAP004211, AGAP029052, AGAP007006, AGAP011118, AGAP009210, AGAP012008, AGAP029606
Relacionado ao relógio circadiano	7	AGAP009755, AGAP009756, AGAP004261, AGAP001982, AGAP005350, AGAP004263, Ligação de Hormônio Juvenil: AGAP001984
Sinalização celular, putativo	1	AGAP008983
Sinalização de cálcio	8	AGAP007666, AGAP010957, AGAP012844, AGAP029559, AGAP004631, AGAP001501, AGAP012406, Domínio calpaína: AGAP010395
Sistema Imune, proteção contra bactéria Gram-negativa	1	AGAP009619
Superoxide dismutase, balanço Redox	1	AGAP005234
Thiol Protease	1	AGAP005509

Thiol Reductase	1	AGAP002989
Transdução de sinal	2	AGAP008821, Fagocitose: AGAP006921
Transporte de vesícula, complexo de Golgi	1	AGAP002357
Transporte Transmembranar	23	AGAP008931, AGAP012097, AGAP010370, solute carrier family 15 member-like; De açúcar:AGAP007340, AGAP003492, AGAP009745, AGAP011493; De Aminoácidos: AGAP009896, AGAP010857; De cobre: AGAP002109; De íons: AGAP009616; De zinco: AGAP011388, zinco ou outros metais: AGAP012772; ADP:ATP antiporter: AGAP002358; ATPase Na+ K+: AGAP007790, AGAP002858; íons sódio: AGAP00783; Transportador ABC: AGAP009466, AGAP002638, AGAP000440(de ácidos graxos), AGAP003680(desenvolvimento de cutícula); transporte de desoxinucleotídeos :AGAP009333;
Ubiquitinação	6	AGAP002691, AGAP005324, AGAP029666, AGAP029707, Proteólise: AGAP003839; Regulador negativo- Via de Notch: AGAP012139.
V- proton ATPase	3	AGAP009486, AGAP028154, AGAP003153
Via da Gliconeogênese	1	AGAP003350
Via de BMP	2	AGAP008640, AGAP002032
Via de JNK	1	AGAP007651
Via de Melanização/Esclerotização	3	AGAP006441, AGAP004549, AGAP010280
Via de Notch	2	AGAP008822, AGAP003070
Via de sinalização mediada por hormônio juvenil	1	AGAP001313
Via de Toll	1	AGAP008360
Via de Wnt	1	AGAP012717

As anotações funcionais dos genes predominantes presentes na serosa de *A. gambiae* foram dispostas com os seus respectivos percentuais na Tabela 4. Nesta Tabela não foram incluídos os 101 genes com funções desconhecidas, que representam 21,4% do total dos 472 genes.

**Tabela 4.** Disposição do “top 10” das funções conhecidas mais presentes na serosa de *A. gambiae*. Sempre que o percentual com uma casa decimal é o mesmo entre duas funções, ambas assumem o mesmo nível de ranking. Por isso, esse "top 10" possui 17 funções. "Miscelânea" são os genes para os quais não conseguimos atribuir funções.

Ranking	Função	Porcentagem (%)	Número de genes
1	Metabolismo de Lipídeos	14,0	52
2	Protease/Proteólise	6,7	25
3	Transporte transmembranar	6,2	23
4	Miscelânea	4,9	18
4	Proteína cuticular	4,9	18
5	Metabolismo de Aminoácidos	4,0	15
6	Metabolismo de Carboidratos	3,5	13
6	Relacionado ao citoesqueleto	3,5	13
7	Proteína G	3,0	11
8	Metabolismo de Glutamina	2,7	10
9	CYP450	2,2	8
9	Fator de transcrição	2,2	8
9	Metabolismo de Quitina	2,2	8
9	Sinalização de cálcio	2,2	8
10	Inibidor de Serina Protease	1,9	7
10	Proteína quimiossensorial	1,9	7
10	Relacionado ao relógio circadiano	1,9	7

#### **4.3 Descrição e anotação funcional dos genes possivelmente expressos na serosa de *T. castaneum*:**

De forma similar, foi feita uma análise detalhada com todos os 318 genes possivelmente expressos na serosa do *T. castaneum*. A descrição destes genes, bem como a anotação funcional de cada um deles está descrito no Anexo VII (páginas 113 a 125). Do total de 318 genes, 242 possuem descrição e função e 76 genes são desconhecidos. Destes 242, 69 não possuíam descrição no Ensembl, mas conseguimos atribuir funções para eles a partir de outros bancos de dados. Vale destacar que o gene mais diferencialmente expresso na serosa

(TC001141) é um gene específico de artrópodes, totalmente desconhecido. O segundo e terceiro genes mais diferencialmente expressos codificam respectivamente uma Fatty acyl-CoA reductase e uma elongase (Elongation of very long chain fatty acids protein), ambos relacionados com metabolismo de lipídeos. Os genes com as mesmas funções foram agrupados juntos na Tabela 5, pela ordem alfabética das funções.

**Tabela 5. Identificação dos genes que possivelmente estão sendo expressos na serosa do besouro *Tribolium castaneum* agrupados pelas funções atribuídas em ordem alfabética.** "Miscelânea" são os genes para os quais não conseguimos atribuir funções. Os genes de fato presentes na serosa estão destacados em azul e verde, sendo que os destacados em verde são aqueles cujo RNAi mostrou efeito na serosa. Os genes que não foram detectados na serosa estão destacados em amarelo. A comprovação experimental da presença dos genes na serosa foi efetuada através de hibridizações in situ (Vargas 2015, Gurská 2017, Rezende, dados não publicados).

Função	Quantidade	ID dos genes
Adesão celular	5	TC001129, TC013296, TC014942, TC007861 e TC030754
Amidase	1	TC002790
Aminotransferase, Transaminase	1	TC034300
Anidrase carbônica	1	TC003286
Atividade de peroxidase	1	TC005493
Biossíntese de Purina	3	TC034834, TC005249, TC005250
Canal de Potássio, dependente de voltagem	1	TC031062
Catabolismo de Peróxido de Hidrogênio	1	TC004948
Chaperona	1	TC013191
Ciclo Celular	1	TC010675
Ciclo do Centrossomo	1	TC010590
CYP450	7	TC009877, TC012495, TC000856, TC034858, TC000411, TC034363, TC014346
Desconhecida	76	TC001141, TC003884, TC031852, TC007258, TC015555, TC000089, TC006727, TC033541, TC030984, TC031851, TC034460, TC008024, TC009442, TC032920, TC012208, TC032290, TC032848, TC008235, TC009377, TC005104, TC012183, TC015592, TC033081, TC033893, TC008048, TC012970, TC010817, TC001860, TC033203, TC016330, TC005998, TC011099, TC014831, TC014688, TC032847, TC031037, TC032087, TC008047, TC009269, TC032476, TC033185, TC000428, TC032514, TC007806, TC033623, TC014345, TC034461, TC003982, TC010815, TC012828, TC034364, TC034862, TC032983, TC030992, TC034950, TC010023, TC001176, TC031591, TC033874, TC033330, TC008005, TC014943, TC033543, TC034269, TC007091, TC014638, TC007679, TC033916, TC005048, TC032831, TC034792, TC032641, TC002586, TC031487, TC034273, TC012929
Desenvolvimento de Cutícula	3	Impermeabilização: TC015564, Regulação de Endocitose: TC031827, TC031922
Diferenciação Epitelial	2	TC010099, TC009191
Divisão celular	1	TC001320
Endonuclease	1	TC005292
Fator de Transcrição	13	TC014708, TC033158, TC009560, TC032304, TC003707, TC012346, TC033029, TC031386, TC003355, TC002315, TC014225, TC000921(define a serosa), TC000922(função na serosa)
Guanilato Kinase, Desenvolvimento Traqueal	1	TC033082
Ligação a açúcares	3	Lectina (TC003708, TC006978), TC008247
Ligação de Diacilglicerol (DAG)	1	TC001994
Metabolismo de açúcar	4	TC011889, TC013648, TC032113, TC031239
Metabolismo de Aminoácidos	1	TC014177
Metabolismo de Glutathiona	3	TC006284, TC000067, TC004115
Metabolismo de Lipídeos	34	TC000107, TC016278, TC015108, TC033309, TC013465, TC015379, TC011937, TC010855, TC015188, TC032736, TC015111, TC010864, TC034275, TC011283, TC006124, TC015392, TC012027, TC015380, TC016280, TC015554, TC034274, TC013464, TC015161, TC014329, TC015158, TC000803, TC003327, TC014245, TC032874, TC000236, TC005503, TC013311, TC031518, TC000611
Metabolismo de Lisina, Degradação	1	TC033769

Metabolismo de Quitina	11	TC014100, TC014101, TC006846, TC015481, TC009178, TC010653, TC001120, TC012819, TC012734, TC031511, Chitina Sintase 1 (TC014634)
Metabolismo de Serina	1	TC000542
Miscelânea	9	TC032543, TC011863, TC031718, TC014954, TC033624, TC003085, TC000957, TC003371, TC008734
Nuclease	1	TC002459
Nucleotidase	1	TC002083
Organização da membrana plasmática	1	TC006832
Palmiltoilação	1	TC008472
Polimerização de Actina	1	TC002504
Protease/proteólise	11	09 Proteases ou enzimas proteolíticas (TC030609, TC000635, TC031198, TC009092 TC001036, TC013415, TC004635, TC005908, TC006033), 01 Inibidor de Protease (TC032053), 01 Via de Protease/proteólise (TC012124)
Proteína Cuticular	9	TC008400, TC011141, TC011140, TC011101, TC031150, TC033228, TC009894, TC015720, TC013128
Proteína quinase	5	TC002107, TC034998, TC014143, TC002609, TC007584
Proteína quimiossensorial	4	TC030104, TC031823, TC032844, TC014534
Reação de Oxirredução	5	TC033746, TC015721, TC015717, TC015712, TC015715
Regulação da epiderme	1	TC002669
Regulação de Receptor Tirosina kinase	1	TC010825
Regulador da expressão gênica, Via de BMP	1	TC000827
Relacionado ao citoesqueleto	11	Organização do Citoesqueleto (TC014517, TC033997, TC010445, TC032708, TC031338, TC031666, TC013974), Associado ao Citoesqueleto (TC009525, TC009549, TC015671, TC014285)
Relacionado ao relógio circadiano	10	TC007661, Ligação de hormônio juvenil:TC013560, TC033106, TC013404, TC015517, TC014397, TC034730, TC013559, TC032992, TC013403
Sinalização de Cálcio	7	TC005571, TC011791, TC008169, TC000431, TC030730, TC032788, TC000753
Thiol reductase	1	TC000174
Tirosina Fosfatase	2	TC033738, TC033739
Transcrição	1	TC001319
Transcriptase reversa	2	TC030923, TC004170
Transdução de sinal	1	TC006771
Transferência de Açúcar	2	TC007632, TC034813
Transporte Transmembranar	19	TC001278, TC015598, TC034705, TC010516, TC013544, TC006631, TC007583, TC008454, De Açúcar: TC013484, De Aminoácido: TC003203, TC002616; De Neurotransmissor: TC033778, TC013646; Sódio e Bicarbonato: TC033053; Sulfato: TC015236; Transportador ABC: TC034861, TC014775; Transportador ABC, Desenvolvimento de cutícula: TC012512, TC034444
Via de BMP	3	TC012674, Antagonista: TC007044, Regulador da Via de BMP (TC010525)
Via de Ecdisona	1	TC008101
Proteína G	11	TC032258, TC033522, TC008856, TC033520, TC033244, TC007083,TC006723, TC010654, TC031358, TC003668, Receptor de Octopamina: TC011639
Via de JNK	1	TC013579
Via de Melanização/Esclerotização	4	TC006229, TC008204, DDC: TC013480, Só esclerotização: TC010001
Via de Notch	1	TC031361
Via de Pigmentação	1	TC010593
Via de Toll	8	TC013304,TC004438, TC000520, TC011529, TC006726, TC012787, TC032151, Desenvolvimento do tegumento:TC007252

As funções predominantes presentes na serosa de *T. castaneum* foram dispostas com os seus respectivos percentuais na Tabela 6. Nesta Tabela não foram incluídos os 76 genes com funções desconhecidas, que representam 23,9% do total dos 318 genes.

**Tabela 6.** Disposição do “top 10” das funções conhecidas dos genes provavelmente presentes na serosa de *T. castaneum*. Sempre que o percentual com uma casa decimal é o mesmo entre duas funções, ambas assumem o mesmo nível de ranking. Por isso, esse “top 10” possui 19 funções. “Miscelânea” são os genes para os quais não conseguimos atribuir funções.

Ranking	Função	Porcentagem (%)	Número de genes
1	Metabolismo de Lipídeos	14	34
2	Transporte Transmembranar	7,9	19
3	Fator de Transcrição	5,4	13
4	Via de GTP	4,5	11
4	Metabolismo de Quitina	4,5	11
4	Protease/proteólise	4,5	11
4	Relacionado ao citoesqueleto	4,5	11
5	Relacionado ao relógio circadiano	4,1	10
6	Miscelânea	3,7	9
6	Proteína Cuticular	3,7	9
7	Via de Toll	3,3	8
8	CYP450	2,9	7
8	Sinalização de Cálcio	2,9	7
9	Adesão celular	2,1	5
9	Proteína kinase	2,1	5
9	Reação de Oxirredução	2,1	5
10	Metabolismo de açúcar	1,7	4
10	Proteína quimiossensorial	1,7	4
10	Via de Melanização/Esclerotização	1,7	4

#### 4.4 Comparação da anotação funcional entre ambas as espécies:

Muitas funções estão presentes em ambas as listas “top 10” (lembrando que esse “top 10” possui 17 funções em *A. gambiae* e 19 funções em *T. castaneum*). O metabolismo de lipídeos é a primeira função em ambas as espécies, com o mesmo percentual (14%), o que aponta para uma grande importância, provavelmente ancestral, dos lipídeos para a serosa e para a cutícula serosa, pois a epicutícula é possivelmente formada por lipídeos e eles podem contribuir na proteção contra a perda de água, visto que são hidrofóbicos (Vargas, 2019). Existem outras 11 funções em comum no “top 10” das duas espécies: proteases,

transportadores transmembranares, proteína cuticular, relacionado ao citoesqueleto, Proteína G, citocromos P450 (CYP450), fator de transcrição, metabolismo de quitina, sinalização de cálcio, proteína quimiossensorial, relacionado ao relógio circadiano.

No "top 10" das funções mais presentes existem 3 que só existem em *A. gambiae*: inibidor de serina protease, metabolismo de aminoácido e metabolismo de glutatona. A função de metabolismo de aminoácido e metabolismo de glutatona está dentre os genes do *T. castaneum*, mas não estão no "top 10" deles. Já os genes com a função de inibir serina protease não estão no transcriptoma da serosa do besouro. Por outro lado, existem 5 funções que só existem no "top 10" do besouro: via de Toll, via de melanização/esclerotização, adesão celular, proteína kinase e reação de oxirredução. Todas essas estão presentes em *A. gambiae* porém em percentual inferior quando comparado ao besouro. Genes relacionados com a via de Toll podem participar do sistema imune e/ou da padronização antero posterior (Jacobs et al., 2014). A via de melanização/esclerotização deve ser importante para a rigidez da cutícula, e é curioso que no mosquito essa função não esteja no "top 10", pois a casca do ovo de *A. gambiae* é melanizada, diferentemente da casca do *T. castaneum*. O fato de sete genes com função de inibir serina protease só estarem presentes em *A. gambiae* nos faz refletir se essa função é algo específico de mosquito, visto que as outras funções podem não pertencer ao top 10 (funções mais presentes) de ambos, mas existem em ambos, em percentuais menores.

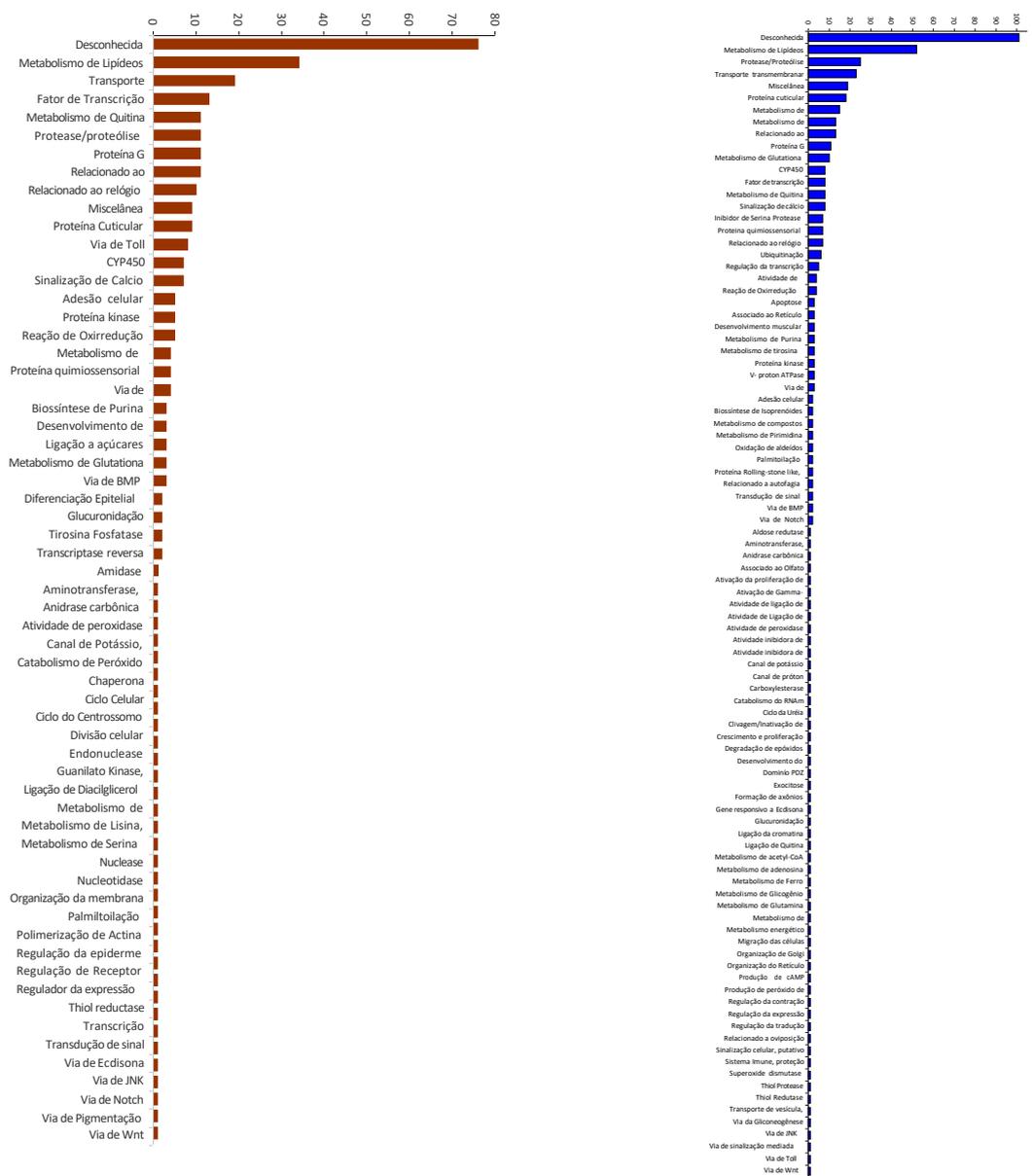
A Tabela 7 e as Figura 13 e 14 mostram uma análise global da comparação da anotação funcional dos transcriptomas de *A. gambiae* e *T. castaneum*. Do total de 17 funções mais presentes em *A. gambiae* e 19 em *T. castaneum*, existem 12 funções em comum no "top 10" das duas espécies. O percentual de genes com função desconhecidas são muito próximos: 21 e 24% em *A. gambiae* e *T. castaneum*, respectivamente. E ambas as espécies os genes desconhecidos apresentam o maior percentual, seguindo de metabolismo de lipídeos. Também ocorrem muitas funções representadas por apenas um gene, sendo considerados como uma "cauda" (funções que só tem 1 gene) (Figura 13). Em relação ao número total de funções conhecidas, *Anopheles gambiae* e *Tribolium castaneum* possuem, respectivamente, 93 e 59 funções. Essa diferença pode ocorrer pelo fato de existir mais genes na serosa do *A. gambiae* ou por serem genes específicos de mosquito. É possível visualizar essa diferença no número de funções e genes na Figura 13. A Figura 14 apresenta a mesma comparação, porém sem os dados dos desconhecidos e da cauda, somente com o "top 10" e as funções restantes de cada espécie (Figura 14).

**Tabela 7. Comparação da anotação funcional dos genes de *A. gambiae* e *T. castaneum*.** A distribuição percentual indica os genes presentes nas funções do "top 10", os de função desconhecida e a cauda (cada função relacionada com apenas um gene). Não estão representados outros genes, que correspondem a 13,8 e 7,2% em *A. gambiae* e *T. castaneum*, respectivamente.

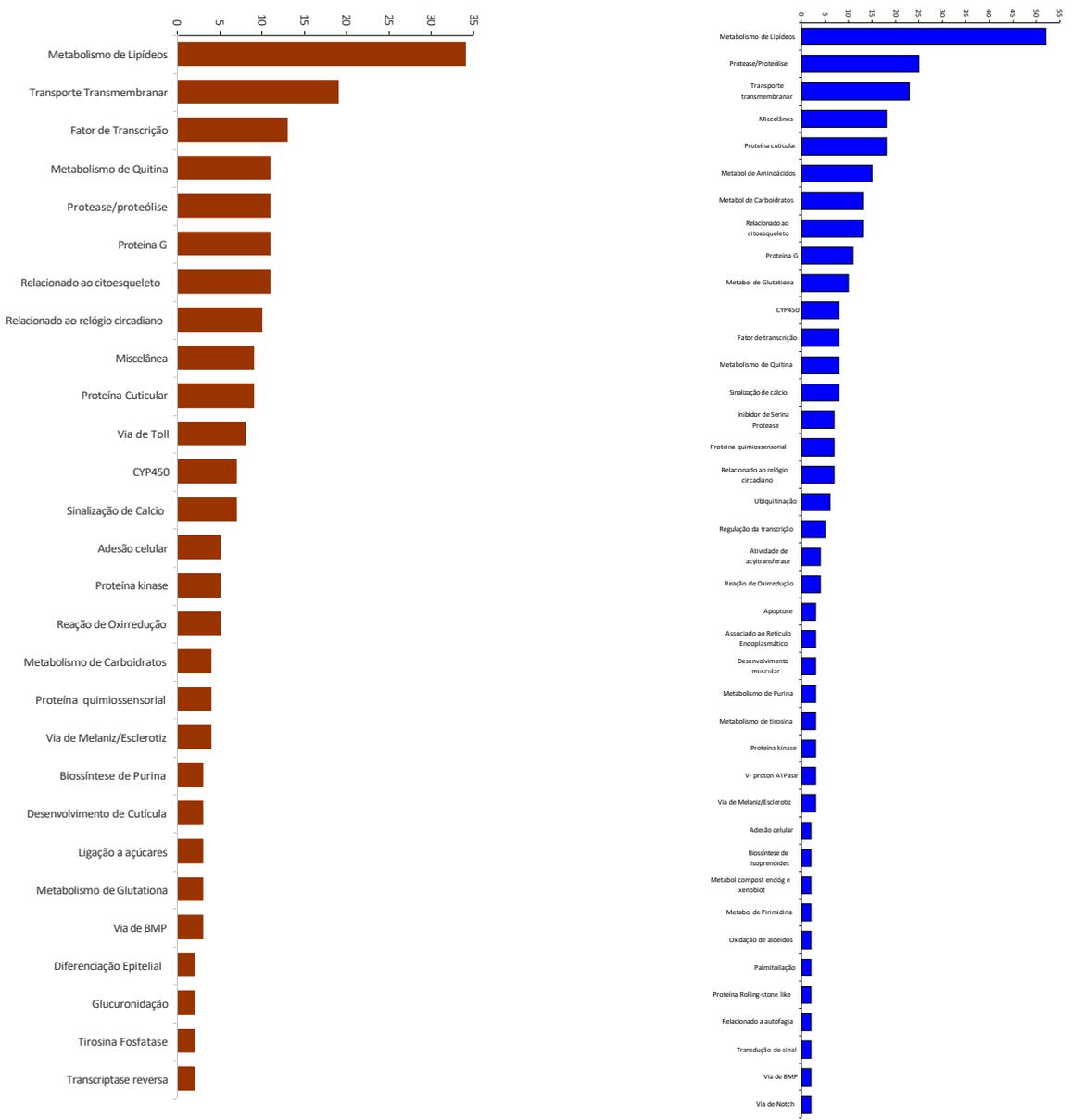
Espécie	Total de genes	Distribuição percentual			Funções no "top10"	Funções "top" 10 em comum
		"genes no "top 10"	Genes desconhecidos	cauda		
<i>A. gambiae</i>	472	53,2%	21,4%	11,7%	17	12
<i>T. castaneum</i>	318	58,8%	23,9%	10,1%	19	12

A função de proteínas cuticulares apresenta uma grande diferença entre *A. gambiae* e *T. castaneum*. O *A. gambiae* e o *T. castaneum* apresentam, respectivamente, 18 e 9 genes que codificam proteínas cuticulares. O *A. gambiae* pode ter mais proteínas cuticulares devido a sua maior susceptibilidade à perda de água (Goltsev *et al.*, 2009, Farnesi *et al.* 2017): o ambiente úmido do ovo do mosquito pode secar e com isso precisará se proteger mais contra a perda de água do que os ovos de *T. castaneum* que já são postos em ambiente seco.

Figura 13. Gráfico das funções de todos os genes de *A. gambiæ* (parte superior da figura) e de *T. castaneum* (parte inferior da figura). Estão presentes os genes com função desconhecida e as funções representadas por apenas um gene (“cauda”).



**Figura 14. Gráfico das funções dos genes de *A. gambiæ* (parte superior da figura) e de *T. castaneum* (parte inferior da figura) contendo somente as funções conhecidas e sem a "cauda" (funções representadas por apenas um gene).**



A presença de 12 funções biológicas presentes em ambos os transcriptomas sugere que essas funções são essenciais para a serosa e a cutícula serosa e estão conservadas há mais de 330 milhões de anos (TimeTree, 2023). Essa semelhança torna-se ainda mais relevante quando se considera que ambos os insetos possuem história de vida, ecologia e hábitos de oviposição muito distintos, e que ambos os transcriptomas foram obtidos com técnicas muito diferentes (microarranjo e RNAseq). O estudo detalhado de alguns desses genes pode abrir caminho para novas descobertas e aplicações biotecnológicas relacionadas à resistência e adaptabilidade dessas espécies, bem como abrir caminho para novos métodos de controle de insetos nocivos, em particular quando leva-se em conta que esses genes cuticulares podem também ser importantes para a cutícula de outros estágios de vida.

#### **4.5 Análise dos genes homólogos presentes em ambos os transcriptomas.**

As análises de homologia mostram que 68 genes de *A. gambiae* são homólogos a 51 genes de *T. castaneum*; em percentual dos genes totais (472 e 318) esses valores equivalem a 14,4 e 16,0%, respectivamente. Dentro dos subgrupos de homólogos, *A. gambiae* e *T. castaneum* possuem 30 (6,6%) e 31 (9,4%), respectivamente, de genes ortólogos (Tabela 8) e 37 (7,8%) e 21 (6,6%), respectivamente, de genes parálogos entre espécies (Tabela 9).

**Tabela 8. Trinta genes em *A. gambiae* são ortólogos de 31 genes de *T. castaneum*. O gene AGAP000989 é ortólogo de dois genes: TC011140 e TC011141.**

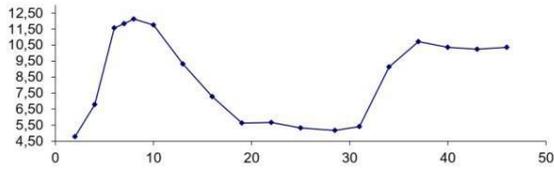
Av log 2 fold <i>T. castaneum</i>	ID <i>Tribolium castaneum</i>	Av log 2 fold <i>A. gambiae</i>	ID <i>Anopheles gambiae</i>	Confirmação Ensembl Pan-Taxonomic	Função
-2,334	TC011140	4,51	AGAP000989	1 to many	Proteína cuticular
-2,767	TC011141	4,51	AGAP000989	1 to many	Proteína cuticular
-2,938	TC011937	3,76	AGAP003600	1 to many	Metabolismo de Lipídeos
-1,234	TC010653	2,02	AGAP004500	1 to many	Metabolismo de quitina
-2,975	TC015379	1,21	AGAP005515	1 to many	Metabolismo de Lipídeos
-1,22	TC014245	2,08	AGAP029758	1 to many	Metabolismo de Lipídeos
-2,165	TC033522	3,1	AGAP000243	1-to-1	Proteína G
-1,28	TC002083	1,45	AGAP000380	1-to-1	Nuclease/Metabolismo de adenosina em <i>A. gambiae</i>
-1,232	TC034861	1,55	AGAP000440	1-to-1	Transporte transmembranar, transportador ABC
-2,372	TC013404	2,89	AGAP001982	1-to-1	Relacionado ao relógio circadiano
-2,389	TC033106	5,41	AGAP001984	1-to-1	Relacionado ao relógio circadiano
-1,281	TC014177	1,77	AGAP002425	1-to-1	Metabolismo de Aminoácido
-3,443	TC012512	1,2	AGAP002638	1-to-1	Transporte transmembranar, transportador ABC
-1,706	TC010675	1,01	AGAP003513	1-to-1	ciclo celular/ desconhecida em <i>A. gambiae</i>
-1,473	TC034444	2,01	AGAP003680	1-to-1	Transporte transmembranar, transportador ABC
-1,772	TC011099	2,95	AGAP003746	1-to-1	Desconhecida
-1,451	TC015715	1,11	AGAP003784	1-to-1	Reação de Oxirredução
-1,374	TC006229	1,27	AGAP004549	1-to-1	Via de Melanização/Esclerotização
-2,838	TC003708	3,85	AGAP006430	1-to-1	Ligação a açúcares/ Metabolismo de carboidratos
-1,569	TC015481	1,11	AGAP006898	1-to-1	Metabolismo de Quitina
-1,688	TC031823	1,97	AGAP007564	1-to-1	Proteína Quimiossensorial
-1,579	TC005493	1,08	AGAP008350	1-to-1	Atividade de peroxidase
-1,616	TC007584	1,54	AGAP008813	1-to-1	Proteína kinase
-1,594	TC002616	4,21	AGAP009896	1-to-1	Transporte transmembranar
-1,005	TC030730	2,34	AGAP010395	1-to-1	Sinalização de cálcio
-1,862	TC012027	4,19	AGAP011357	1-to-1	Metabolismo de Lipídeos
-4,504	TC016278	5,37	AGAP011812	1-to-1	Metabolismo de Lipídeos

-2,803	TC014100	5,07	AGAP011936	1-to-1	Metabolismo de Quitina
-2,782	TC014101	2,11	AGAP011937	1-to-1	Metabolismo de Quitina
-1,546	TC006978	1,04	AGAP029559	*sem confirmação	Ligação a açúcares/ Sinalização de cálcio
-2,935	TC010855	4,11	AGAP000275	Many to Many	Metabolismo de Lipídeos

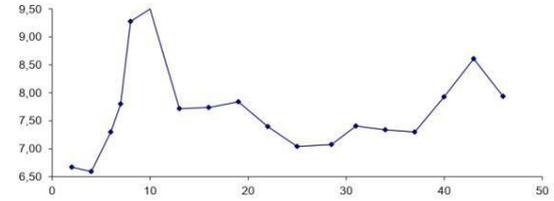
. \*sem confirmação: esse foi o único caso em que o GeneTree não confirmou os dados dos BLASTp recíprocos. Entendemos que nesse caso houve uma falha do GeneTree (conforme visto para outro gene - dados não mostrados) pois em ambos os BLASTPs que confirmaram a ortologia entre TC006978 e AGAP029559 os valores de E-value foram muito baixos ( $2e-93$  e  $9e-94$ ).

Os perfis de expressão temporal, ao longo de toda a embriogênese de *A. gambiae*, dos 30 genes ortólogos da Tabela 8 estão dispostos na Figura 15. Muitos desses genes possuem expressão bifásica, indicando um papel tanto na formação da cutícula serosa quanto na da cutícula larval, ao final da embriogênese. Um desses é o gene AGAP001984, relacionado ao relógio circadiano e com ligação ao hormônio juvenil. Antes do presente trabalho, tanto o AGAP001984 quanto o seu ortólogo TC033106 eram considerados genes com descrição e função desconhecidas.

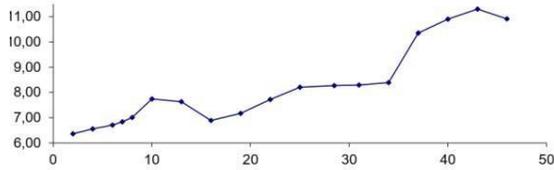
AGAP011812, Metabolismo de lipídeos



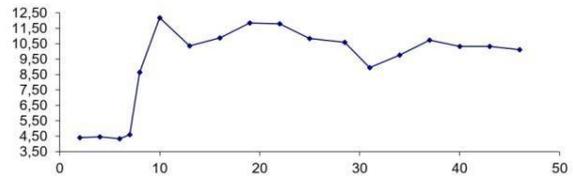
AGAP002638, Transporte transmembranar



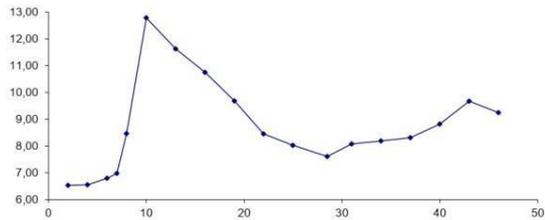
AGAP005515, Metabolismo de lipídeos



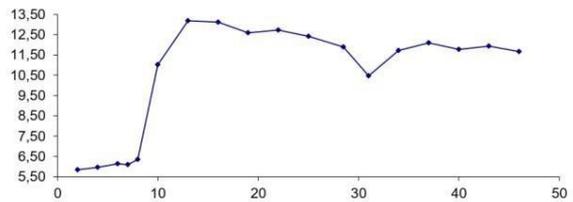
AGAP003600, Metabolismo de lipídeos



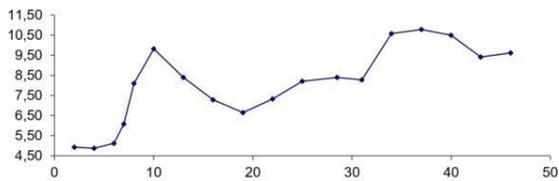
AGAP000275, Metabolismo de lipídeos



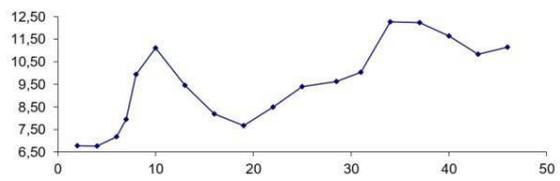
AGAP006430, Metabolismo de carboidratos



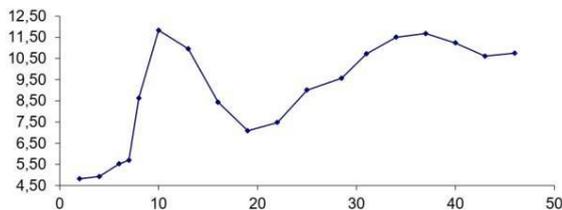
AGAP011936, Metabolismo de quitina



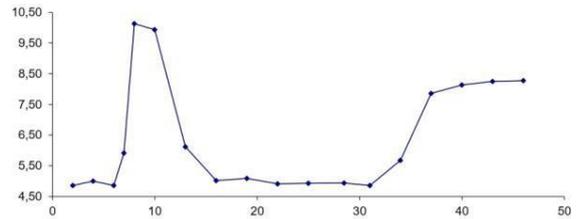
AGAP011937, Metabolismo de quitina



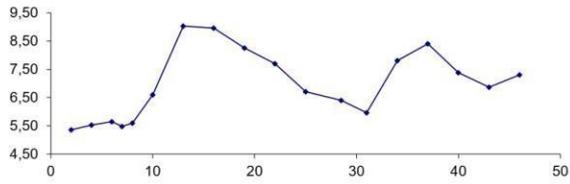
AGAP000989, Proteína cuticular



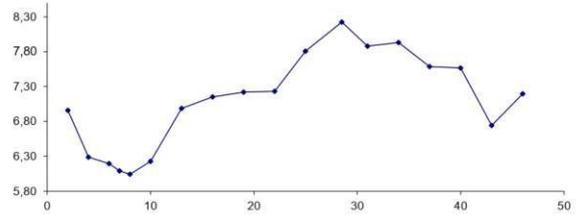
AGAP001984, Relacionado ao relógio circadiano



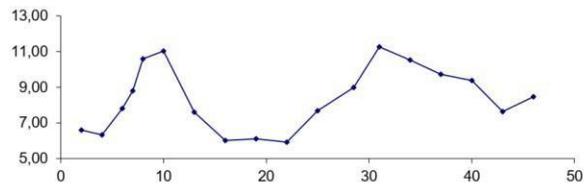
AGAP001982, Relacionado ao relógio circadiano



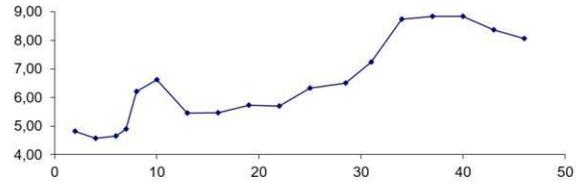
AGAP000243, Proteína G



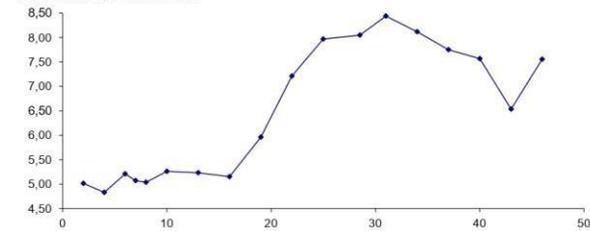
AGAP011357, Metabolismo de lípidos



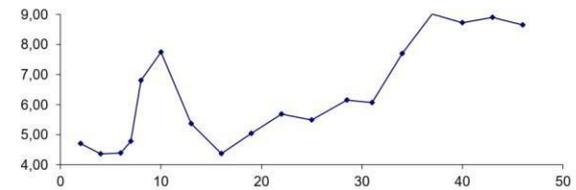
AGAP003746, Desconhecida



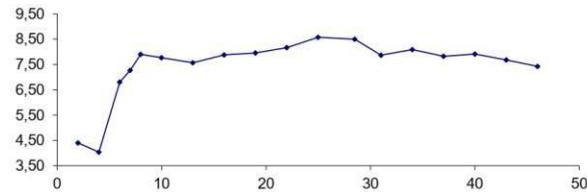
AGAP003513, Desconhecida



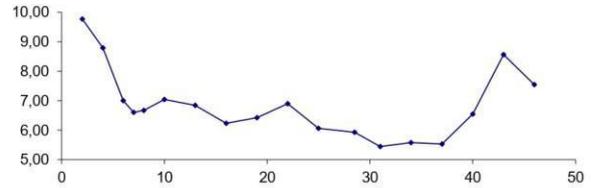
AGAP007564, Proteína quimiossensorial



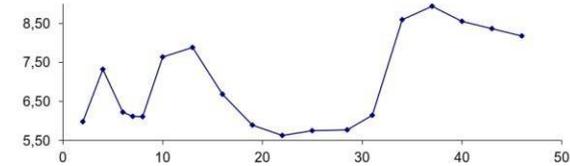
AGAP008813, Proteína quinase



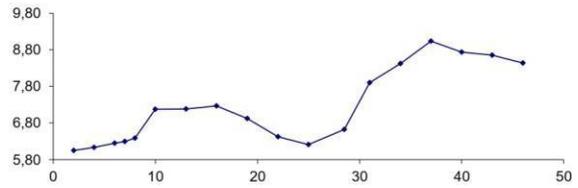
AGAP009896, Transporte transmembranar

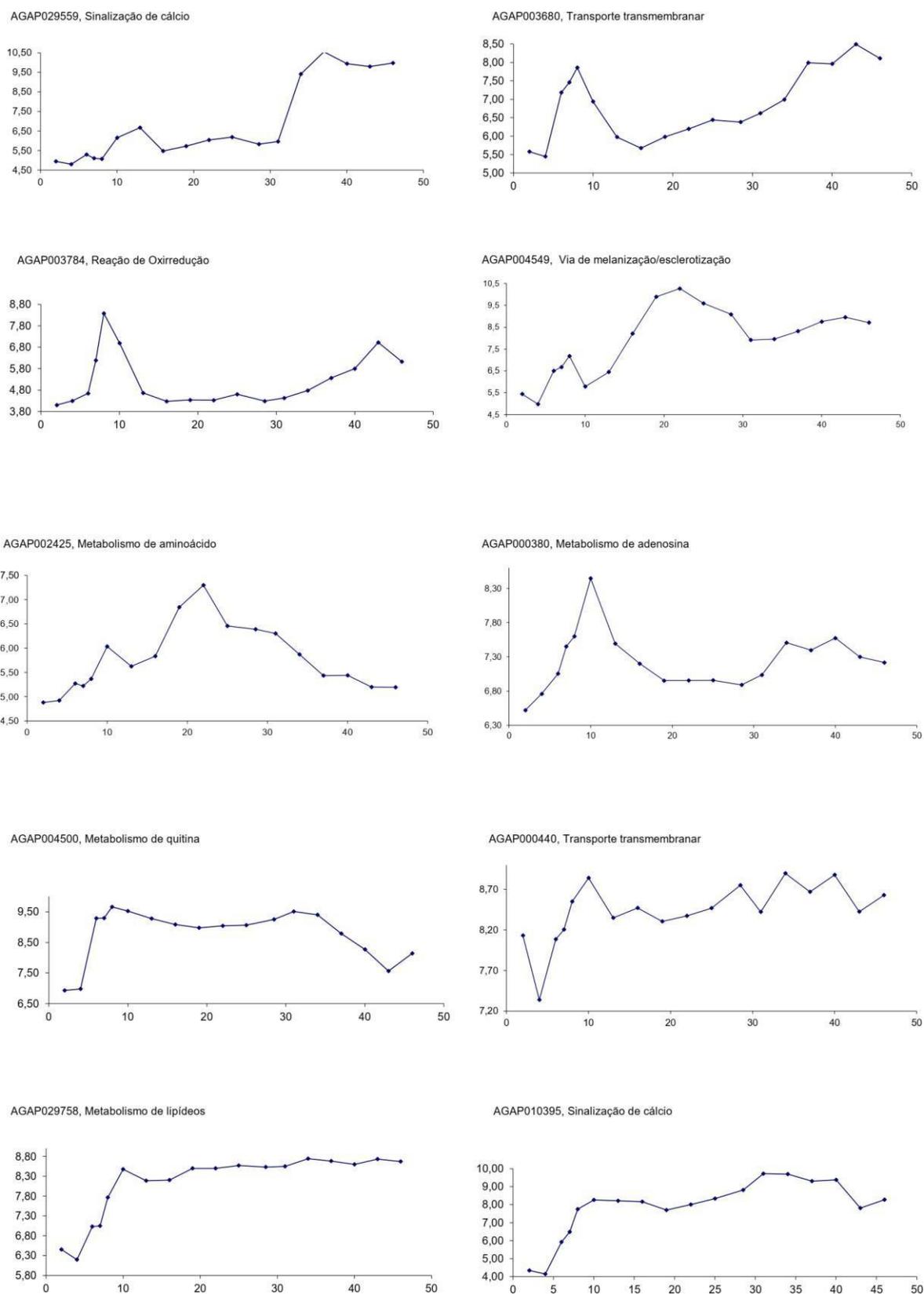


AGAP008350, Atividade de peroxidase



AGAP006898, Metabolismo de quitina





**Figura 15. Perfis de expressão gênica dos 30 ortólogos ao longo da embriogênese de *A. gambiae*.** Cada gráfico mostra o perfil de expressão da sonda de microarranjo associada ao gene (AGAP) indicado no canto superior direito de cada painel. Figura criada a partir de dados obtidos através do microarranjo temporal de Goltsev e colaboradores (2009). No eixo x está o tempo de embriogênese em horas e no eixo y o nível de expressão em unidades arbitrárias

**Tabela 9.** Trinta e sete genes de *A. gambiae* possuem parálogos entre espécies de 21 genes de *T. castaneum*.

<i>T. castaneum</i>		<i>A. gambiae</i>		Função
Gene homólogo	AvgLog2 Fold	Gene homólogo	AvgLog2 Fold	
TC014397	-1,54	AGAP001984	5,41	Relacionado ao relógio circadiano
TC003085	-1,16	AGAP008059	4,78	Miscelânea/Proteína quimiossensorial
TC014534	-1,04	AGAP008059	4,78	Proteína quimiossensorial
TC011791	-1,35	AGAP003790	4,41	Sinalização de cálcio/ Relacionado ao citoesqueleto
TC013415	-1,43	AGAP012946	4,39	Protease-proteólise/ Desconhecida
TC014397	-1,54	AGAP004263	4,32	Relacionado ao relógio circadiano
TC015161	-1,48	AGAP003606	4,15	Metabolismo de Lipídeos
TC032874	-1,20	AGAP002738	4,05	Metabolismo de Lipídeos/ Miscelânea
TC003286	-2,80	AGAP002359	3,72	Anidrase carbônica
TC013415	-1,43	AGAP029292	3,26	Protease-proteólise/ Desconhecida
TC013415	-1,43	AGAP029106	3,19	Protease-proteólise/ Desconhecida
TC000236	-3,17	AGAP009104	2,93	Metabolismo de Lipídeos
TC014397	-1,54	AGAP001982	2,89	Relacionado ao relógio circadiano
TC015161	-1,48	AGAP002279	2,72	Metabolismo de Lipídeos
TC013415	-1,43	AGAP008835	2,64	Protease-proteólise
TC034813	-1,09	AGAP007589	2,29	Transferência de açúcar/ Metabolismo de carboidratos
TC013415	-1,43	AGAP029117	2,25	Protease-proteólise/ Desconhecida
TC013415	-1,43	AGAP029747	2,25	Protease-proteólise/ Desconhecida
TC013415	-1,43	AGAP029750	2,25	Protease-proteólise/ Desconhecida
TC013415	-1,43	AGAP011431	2,25	Protease-proteólise/ Desconhecida
TC000174	-1,49	AGAP002989	2,23	Thiol reductase
TC000520	-2,092	AGAP008360	2,19	Via de Toll
TC032151	-1,03	AGAP008360	2,19	Via de Toll
TC032874	-1,20	AGAP005716	2,09	Metabolismo de Lipídeos/ Proteína quimiossensorial
TC011283	-2,10	AGAP004784	2,07	Metabolismo de Lipídeos
TC013415	-1,43	AGAP000573	2,03	Protease-proteólise
TC013415	-1,43	AGAP000572	1,92	Protease-proteólise
TC003085	-1,16	AGAP008054	1,88	Miscelânea/Proteína quimiossensorial
TC033738	-1,63	AGAP002438	1,82	Tirosina Fosfatase/ Metabolismo de tirosina
TC000067	-1,37	AGAP010404	1,73	Metabolismo de glutatona
TC003085	-1,16	AGAP008051	1,71	Miscelânea/Proteína quimiossensorial
TC003355	-1,32	AGAP005311	1,69	Fator de Transcrição/ Regulação da transcrição
TC007632	-1,55	AGAP007920	1,68	Glucuronidação, transferência de ácido glucurônico/ Metabolismo de carboidratos
TC011863	-1,35	AGAP007209	1,57	Miscelânea
TC013415	-1,43	AGAP004859	1,49	Protease-proteólise
TC008400	-3,32	AGAP006867	1,39	Proteína Cuticular
TC013415	-1,43	AGAP003627	1,34	Protease-proteólise
TC011283	-2,10	AGAP005515	1,21	Metabolismo de Lipídeos
TC033746	-2,59	AGAP003784	1,11	Reação de Oxirredução

Houve situações em que o BLASTp recíproco realizado com uma proteína predita de *A. gambiae* não fornecia como primeiro resultado *best hit* a proteína predita de *T. castaneum* que foi usada como busca (*query*) do primeiro BLASTp. Estes poderiam ser casos de famílias gênicas em que existem parálogos entre espécies. A confirmação dessas paralogias foi feita através do *GeneTree* do *Pan-Taxonomic Compara* do *Ensembl Metazoa*. Em alguns casos não foi possível confirmar a paralogia entre genes através do *GeneTree* e foi preciso olhar o resultado do BLASTp recíproco novamente para confirmar se a proteína predita de busca do primeiro BLASTp fazia parte da família gênica, ou seja, se eram parálogos entre espécies.

#### 4.6 Genes selecionados para serem estudados experimentalmente no futuro

Os genes selecionados a serem estudados futuramente em *T. castaneum* (Tabela 10), são genes que apresentam uma fase de elevada expressão no período de formação da cutícula serosa em *A. gambiae*. E a maioria das funções atribuídas a esses genes são consideradas funções muito presentes em ambas as serosas de acordo com os resultados das análises.

As informações aqui obtidas com os transcriptomas de *Anopheles gambiae* e *Tribolium castaneum* podem ser importantes para todas as cutículas: serosa, larval, da pupa e do adulto. Isto é provável já que alguns genes são essenciais para a cutícula em geral, como por exemplo os necessários para a síntese de quitina, que está presente em todas essas cutículas.

Observando o gráfico da expressão gênica do microarranjo temporal (Figura 15) é possível ver que alguns genes apresentam elevada expressão em horários próximos, e quedas de expressão mais ou menos abruptas. Os genes AGAP011812 e AGAP001984 exemplificam os dois processos: seus pontos iniciais de alta expressão são com 6 e 8 horas de embriogênese; o pico de expressão de AGAP011812 é mais largo, das 6 as 13h enquanto o pico de expressão de AGAP001984 é mais abrupto, das 8 as 10h. Isso ocorre pois o processo de síntese de uma matriz extracelular complexa como a cutícula serosa (vide Figura 5) é subdividida em várias etapas, o que faz com que alguns genes tenham que ser expressos antes de outros (Goltsev *et al.*, 2009).

Em relação ao transcriptoma de *T. castaneum* feito com RNAseq (Gurska *et al.*, 2020) dados prévios corroboram os achados aqui encontrados: o silenciamento para os genes *Chs1* e *knk1*, ambos essenciais para o metabolismo de quitina, afeta a formação da cutícula serosa e diminui a proteção que essa confere contra a perda do ovo (Jacobs *et al.*, 2013, 2015 e vide Introdução).

Os dados de transcriptomas são cada vez mais utilizados para estudos da filogenia dentro e entre ordens de insetos (Gullan, 2017). As semelhanças entre transcriptomas referem-se às similaridades nas expressões gênicas e os transcriptomas de diferentes espécies podem compartilhar uma série de genes comuns. Eles podem compartilhar vias metabólicas comuns, indicando uma função celular semelhante, como por exemplo os ortólogos AGAP011812 e TC016278 que estão envolvidos na síntese de ácidos graxos de cadeia muito longa (Rezende, comunicação pessoal). E podem ter padrões semelhantes de regulação da expressão gênica, indicando a ativação ou desativação de vias específicas em resposta a uma determinada condição, como por exemplo os genes ortólogos TC007584 e AGAP008813 que codificam uma proteína kinase, com possível papel regulatório. E essa semelhança entre os transcriptomas podem ocorrer em diferentes estágios do desenvolvimento, o que indica a expressão de genes específicos necessários para cada estágio, como por exemplo, nesse trabalho, que foi possível observar a expressão dos genes no início da embriogênese para a formação da cutícula serosa pelas células da serosa.

Ao analisarmos os genes homólogos e conseqüentemente, genes ortólogos e parálogos presentes nas serosas de ambas as espécies é possível observar os genes que são conservados, ou seja, genes essenciais a esse tecido e a formação da cutícula serosa.

A lista de genes criada, para se estudar experimentalmente no futuro, servirá para confirmar a presença desses genes na serosa do besouro através de hibridização *in situ*. Após confirmação da presença do gene na serosa, poderá ser feito silenciamento via RNAi para analisar o efeito no ovo com o gene escolhido silenciado, e assim observar a importância desse gene na serosa e na formação da cutícula serosa.

**Tabela 10. Genes selecionados a serem investigados experimentalmente.**

<b>Gene de <i>A. gambiae</i></b>	<b>Homologia</b>	<b>picos de expressão</b>	<b>Gene de <i>T. castaneum</i></b>	<b>Função</b>
AGAP004263	parálogos	expressão só no período da CS (1 pico só)	TC014397	Relacionado ao relógio circadiano
AGAP002279	parálogos	expressão principal no período da CS (1 pico só)	TC015161	Metabolismo de Lipídeos
AGAP000572	parálogos	expressão principal no período da CS, pouca expressão na CL	TC013415	Protease-proteólise
AGAP002738	parálogos	expressão só no período da CS (1 pico só)	TC032874	Metabolismo de Lipídeos/ Miscelânea
AGAP007589	parálogos	2 picos de expressão (CS e CL)	TC034813	Metabolismo de carboidratos
AGAP002638	ortólogos	2 picos de expressão (CS e CL)	TC012512	Transporte transmembranar
AGAP000275	ortólogos	expressão só no período da CS (1 pico só)	TC010855	Metabolismo de lipídeos
AGAP001984	ortólogos	2 picos de expressão (CS e CL)	TC033106	Relacionado ao relógio circadiano
AGAP003784	ortólogos	expressão principal no período da CS, pouca expressão na CL	TC015715	Reação de Oxirredução
AGAP000380	ortólogos	expressão principal no período da CS, pouca expressão na CL	TC002083	Metabolismo de adenosina
AGAP000989	ortólogos	2 picos de expressão (CS e CL)	TC011140,TC011141	Proteína cuticular
AGAP003746	ortólogos	2 picos de expressão (CS e CL)	TC011099	Desconhecida

## 5 CONCLUSÕES

- O processo de identificação de genes associados a sondas órfãs contribui para a anotação do genoma de *Anopheles gambiae*, visto que foram descobertos 3 novos genes ainda não descritos;
- A existência de genes de funções comuns na serosa de ambas as espécies sugere a importância destas funções para a serosa e talvez para a síntese da cutícula serosa;
- Algumas funções são específicas para cada espécie, o que pode ter relação com características biológicas únicas, não compartilhadas, de cada organismo;
- A presença de genes homólogos indica a importância destes para a serosa e talvez para a cutícula serosa.
- A ancestralidade observada para algumas funções e para os genes homólogos apontam que estes estão conservados a pelo menos 330 milhões de anos, apesar das grandes diferenças biológicas existentes entre o mosquito *A. gambiae* e o besouro *T. castaneum*.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

Beament, J.W. (1989). John Hull Grundy lecture. Eggs--the neglected insects. *Journal of the Royal Army Medical Corps*, 135 2, 49-56 .

Bucher G. The Beetle book. 2006.

Chaudhari SS, Noh MY, Moussian B, Specht CA, Kramer KJ, Beeman RW, Arakane Y, Muthukrishnan S. Knickkopf and retroactive proteins are required for formation of laminar serosal procuticle during embryonic development of *Tribolium castaneum*. *Insect Biochem Mol Biol*. 2015 May;60:1-6. doi: 10.1016/j.ibmb.2015.02.013. Epub 2015 Mar 4. PMID: 25747009.

Clements, A. N., 1992. The biology of mosquitoes. Development, nutrition and reproduction. Chapman and Hall, London.

Dayhoff, M. O., 1978. Atlas of Protein Sequence and Structure, Vol. 5, Supplement 3. National Biomedical Research Foundation, Silver Spring, MD.

Duarte, S., Limão, J., Barros, G., Bandarra, N. M., Roseiro, L. C., Gonçalves, H., ... Carvalho, M. O. (2021). Nutritional and chemical composition of different life stages of *Tribolium castaneum* (Herbst). *Journal of Stored Products Research*, 93, 101826. doi:10.1016/j.jspr.2021.101826

Farnesi LC, Vargas HCM, Valle D, Rezende GL (2017) Darker eggs of mosquitoes resist more to dry conditions: Melanin enhances serosal cuticle contribution in egg resistance to desiccation in *Aedes*, *Anopheles* and *Culex* vectors. *PLoS Negl Trop Dis* 11(10): e0006063. <https://doi.org/10.1371/journal.pntd.0006063>

Foster WA, Walker ED. 2009. Mosquitoes (Culicidae). In Mullen, G, Durden, L. (Eds.) *Medical and veterinary entomology*. 2nd Ed. (pp. 207-259). Academic Press, Burlington, MA. 637 pp17

Gabaldón, T., Koonin, E. Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics* 14 , 360–366 (2013). <https://doi.org/10.1038/nrg3456>.

Giachetto, Poliana Fernanda. A tecnologia de microarranjos na identificação de genes de interesse na bovinocultura - Campinas: Embrapa Informática Agropecuária, 2010

Graur D, Li WH. - Fundamentals of Molecular Evolution, 2000.

Goltsev Y, Rezende GL, Vranizan K, Lanzaro G, Valle D, Levine M. Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo. *Dev Biol.* 2009 Jun 15;330(2):462-70. doi: 10.1016/j.ydbio.2009.02.038. Epub 2009 Mar 17. PMID: 19298808; PMID:PMC4084816.

Grimaldi D, Engel M. 2005. Evolution of the insects. Cambridge University Press, 755 p.

Gurska D. The changing roles of Hox3 genes in insect evolution: characterizing the zen paralogues in the Beetle *Tribolium castaneum*.2017.

Gurska D, Vargas Jentsch IM, Panfilio KA. 2020. Unexpected mutual regulation underlies paralogue functional diversification and promotes epithelial tissue maturation in *Tribolium*. *Commun Biol.* 2020 Oct5;3(1):552. doi: 10.1038/s42003-020-01250-3.

Hurles M. Gene duplication: the genomic trade in spare parts. *PLoS Biol.* 2004 Jul;2(7):E206. doi: 10.1371/journal.pbio.0020206. Epub 2004 Jul 13. PMID: 15252449; PMID:PMC449868.

Hinton H. E., 1981. Biology of insect eggs. Pergamon Press 1: 473p.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin

E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*. 2002 Oct 4;298(5591):129-49. doi: 10.1126/science.1076181. PMID: 12364791.

Hooper, Sean D; Boué, Stephanie; Krause, Roland; Jensen, Lars J; Mason, Christopher E; Ghanim, Murad; White, Kevin P; Furlong, Eileen EM; Bork, Peer (2007). *Identification of tightly regulated groups of genes during Drosophila melanogaster embryogenesis*. *Molecular Systems Biology*, 3(), -. doi:10.1038/msb4100112

Jacobs CG, Rezende GL, Lamers GE, van der Zee M. 2013. The extraembryonic serosa protects the insect egg against desiccation. *Proc Biol Sci*. Jun 19;280(1764):20131082. doi: 10.1098/rspb.2013.1082. Print 2013 Aug 7.

Jacobs CG, Spaink HP, van der Zee M. The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response. *Elife*. 2014 Dec 9;3:e04111. doi: 10.7554/eLife.04111. PMID: 25487990; PMCID: PMC4358341.

Jacobs CG, Braak N, Lamers GE, van der Zee M. Elucidation of the serosal cuticle machinery in the beetle *Tribolium* by RNA sequencing and functional analysis of *Knickkopf1*, *Retroactive* and *Laccase2*. *Insect Biochem Mol Biol*. 2015 May;60:7-12. doi: 10.1016/j.ibmb.2015.02.014. Epub 2015 Mar 4. PMID: 25747006.

Jawara M, Pinder M, Drakeley CJ, Nwakanma DC, Jallow E, Bogh C, Lindsay SW, Conway DJ. Dry season ecology of *Anopheles gambiae* complex mosquitoes in The Gambia. *Malar J*. 2008 Aug 18;7:156. doi: 10.1186/1475-2875-7-156. PMID: 18710559; PMCID: PMC2533673.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001 Jan 19;305(3):567-80. doi: 10.1006/jmbi.2000.4315. PMID: 11152613.

Kuzniar A, van Ham RC, Pongor S, Leunissen JA. The quest for orthologs: finding the corresponding gene across genomes. *Trends Genet.* 2008 Nov;24(11):539-51. doi: 10.1016/j.tig.2008.08.009. Epub 2008 Sep 24. PMID: 18819722.

Lorini, I. Descrição, biologia e danos das principais pragas de grãos armazenados. In: Lorini, I., Miike, L.H., Scussel, V.M. (eds.). *Armazenagem de grãos*. Campinas: Instituto Bio Geneziz, p. 378-397, 2002.

Maurijn van der Zee; Nicola Berns; Siegfried Roth (2005). Distinct Functions of *the Tribolium zerku'llt* Genes in Serosa Specification and Dorsal Closure. , 15(7), 624–636. doi:10.1016/j.cub.2005.02.057

Ministério da Saúde - Malária - Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/m/malaria> {s.d.] acessado em: julho, 2023.

Okuneye, K., Eikenberry, S. E., & Gumel, A. B. (2019). Weather-driven malaria transmission model with gonotrophic and sporogonic cycles. *Journal of Biological Dynamics*, 1–37. doi:10.1080/17513758.2019.1570363

Pacheco, C., Ceccatto, V.M., Maia, C.M., Rosa, S.D., & Leite, C.R. (2019). Pesquisa translacional na era pós-genômica: avanços na área da transcriptômica. *Saúde em Debate*.

Panfilio KA. Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev Biol.* 2008 Jan 15;313(2):471-91. doi: 10.1016/j.ydbio.2007.11.004. Epub 2007 Nov 17. PMID: 18082679.

P.J. Gullan e P.S. Cranston, Eduardo da Silva Alves dos Santos. *Insetos Fundamentos da Entomologia 5ª Edição*, ROCA , 2017.

REZENDE, G.L. (2008) Investigações sobre a aquisição de resistência à dessecação durante a embriogênese de *Aedes aegypti* e *Anopheles gambiae*.

Rezende GL, Martins AJ, Gentile C, Farnesi LC, Pelajo-Machado M, Peixoto AA, Valle D. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. *BMC Dev Biol.* 2008 Sep 13;8:82. doi: 10.1186/1471-213X-8-82. PMID: 18789161; PMCID: PMC2561029.

Rezende, G.L., Vargas, H.C., Moussian, B., & Cohen, E. (2016). Composite Eggshell Matrices: Chorionic Layers and Sub-chorionic Cuticular Envelopes.

Rota-Stabelli O, Daley AC & Pisani D 2013. Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Current Biology* 23 (5): 392-398.

Slifer E H. The Origin and fate of the Membranes surrounding the Grasshopper Egg; together with some Experiments on the Source of the Hatching Enzyme. 1937.

VARGAS, H. C. M. (2015). CARACTERIZAÇÃO DA EXPRESSÃO DE GENES *elovl* NA SEROSA DE *Tribolium castaneum* (COLEOPTERA: TENEBRIONIDAE)

VARGAS, H. C. M. (2019) Relações hídricas em ovos de Hexapoda: A proteção contra a perda de água conferida pelo gene *elovl-4* ao besouro *Tribolium castaneum* e pela cutícula blastodermal nos colêmbolos *Orchesella cincta* e *Folsomia candida*.

Vargas HCM, Panfilio KA, Roelofs D, Rezende GL. 2020. Increase in egg resistance to desiccation in springtails correlates with blastodermal cuticle formation: Eco-evolutionary implications for insect terrestrialization. *J Exp Zool B Mol Dev Evol.* 2020 Jul 10. doi: 10.1002/jez.b.22979.

Wang MH, Marinotti O, Vardo-Zalik A, Boparai R, Yan G. Genome-wide transcriptional analysis of genes associated with acute desiccation stress in *Anopheles gambiae*. *PLoS One.* 2011;6(10):e26011. doi: 10.1371/journal.pone.0026011. Epub 2011 Oct 4. PMID: 21991392; PMCID:PMC3186805

Zeh DW, Zeh JA, Smith RL. 1989. Ovipositors, amnions and eggshell architecture in the diversification of terrestrial arthropods. *Quarterly Review of Biology* 64, 147-168

## ANEXO I

Passo a passo FINDSTR: Busca de sequências automatizadas

Em "Local Disk (c): Criar uma pasta e copiar o arquivo de interesse, no nosso caso, o arquivo "Plasmodium Anopheles. probe.tab" renomeado, baixado do Affymetrix, link: "[Sequence Files: Plasmodium\\_Anopheles Probe Sequences, Tabular](#)", e colar nessa pasta no "LocalDisk". Em seguida, alterar o nome para um menor, ex: "anop.tab". Ir em "Comand Prompt", e voltar para a raiz "Cd.." seguido de "Cd..", pois desta forma chegará em "c".

A partir daí você entrará no diretório da pasta criada, através do comando:

```
C:\Cd nomedapasta
```

E conseguirá obter a sequência desejada através do comando:

```
C: nomedapasta > findstr - espaço - nome da sonda - espaço - anop.tab - espaço- chave(>) -  
espaço - nome da sonda.xls
```

## ANEXO II

Identificação dos genes associados a sondas de *A. gambiae* (atualização dos dados originais de 2009).

AvLog2Fold Ser/Emb	Probe Set ID 2009	Ensemble ID 2009	ID 2021
5,78	Ag.3L.920.0_CDS_s_at	AGAP010799-11:11	AGAP010799
5,41	Ag.2R.1851.0_CDS_at	AGAP001984-10:10	AGAP001984
5,37	Ag.3L.479.0_CDS_at	AGAP011812-7:9	AGAP011812
5,07	Ag.3L.324.0_UTR_a_at		AGAP011936
4,87	Ag.2R.1713.0_CDS_at	AGAP002962-10:10	AGAP002962
4,85	Ag.X.399.1_at	AGAP000697-6:6	AGAP000697
4,78	Ag.3R.570.1_CDS_a_at	AGAP008059-10:10	AGAP008059
4,77	Ag.3R.408.0_CDS_at	AGAP009790-10:10	AGAP009790
4,75	Ag.3L.1081.0_CDS_at	AGAP010370-11:11	AGAP010370
4,69	Ag.2R.16.0_CDS_at	AGAP003250-11:11	AGAP003250
4,67	Ag.2R.442.1_CDS_s_at	AGAP003308-10:11	AGAP003308
4,67	Ag.2L.259.1_CDS_a_at	AGAP006410-16:9	AGAP006410
4,63	Ag.2R.677.0_CDS_at	AGAP001508-10:10	AGAP001508
4,60	Ag.X.1529.0_a_at	AGAP000821-11:11	AGAP000821
4,56	Ag.2L.115.0_CDS_at	AGAP007412-11:11	AGAP007412
4,54	Ag.3L.798.0_CDS_at	AGAP010928-11:11	AGAP010928
4,51	Ag.X.460.0_CDS_at	AGAP000989-11:11	AGAP000989
4,51	Ag.2R.528.1_CDS_a_at	AGAP002085-9:9	AGAP002085
4,46	Ag.3L.1610.0_CDS_at	AGAP011992-11:11	AGAP011992
4,44	Ag.2R.3686.0_a_at	AGAP003752-12:10	AGAP029052
4,42	Ag.3L.2571.0_at		AGAP011778
4,41	Ag.UNKN.255.0_CDS_s_at	AGAP012930-11:11 AGAP003790-11:11	AGAP003790
4,39	Ag.2R.537.0_CDS_at	AGAP003627-11:11	AGAP012946
4,38	Ag.3R.390.0_CDS_at	AGAP009635-11:11	AGAP009635
4,37	Ag.2L.550.0_UTR_at	AGAP005553-1:1	AGAP005552
4,36	Ag.3R.556.0_UTR_at	AGAP009745-3:3	AGAP009745
4,35	Ag.2R.2084.0_CDS_at	AGAP001101-11:11	AGAP001101
4,33	Ag.2R.450.1_at		AGAP013005
4,32	Ag.2R.17.0_CDS_at	AGAP004263-10:11	AGAP004263
4,32	Ag.2R.3767.0_at		AGAP004373
4,30	Ag.3L.16.0_CDS_a_at	AGAP011780-11:11	AGAP011780
4,30	Ag.2L.1513.0_CDS_at	AGAP007636-9:10	AGAP007636
4,29	Ag.3L.1299.0_CDS_at	AGAP010385-9:9	AGAP010385

4,28	Ag.2L.165.0_CDS_at	AGAP004802-11:11	AGAP004802
4,24	Ag.3L.688.0_CDS_at	AGAP010441-9:10	AGAP010441
4,24	Ag.2R.410.0_CDS_a_at	AGAP001718-18:9	AGAP001718
4,21	Ag.3R.557.0_UTR_at	AGAP009896-6:6	AGAP009896
4,21	Ag.3R.224.0_UTR_a_at		AGAP008988
4,21	Ag.2L.2030.1_at	AGAP006151-10:10	AGAP006151
4,19	Ag.3L.818.1_CDS_at		AGAP011357
4,15	Ag.2L.3934.0_at		AGAP006347
4,15	Ag.2R.1442.0_CDS_at	AGAP003606-5:10	AGAP003606
4,15	Ag.2R.551.0_CDS_at	AGAP003515-11:11	AGAP003515
4,14	Ag.2L.359.0_UTR_a_at		AGAP007717
4,13	Ag.2R.710.0_CDS_at	AGAP001719-10:10	AGAP001719
4,12	Ag.X.1102.0_CDS_at		AGAP029673
4,11	Ag.X.354.0_CDS_at	AGAP000275-10:10	AGAP000275
4,10	Ag.2R.450.4_at	AGAP001510-11:11	AGAP001510
4,07	Ag.X.262.0_CDS_at	AGAP000988-9:9	AGAP000988
4,06	Ag.2R.3698.0_s_at	AGAP003620-11:11	AGAP003620
4,05	Ag.2R.92.0_CDS_at	AGAP002738-8:9	AGAP002738
4,04	Ag.UNKN.307.0_CDS_s_at	AGAP003733-9:9	AGAP003733
4,03	Ag.2R.76.0_CDS_at	AGAP003194-11:11	AGAP003194
4,01	Ag.2R.33.0_CDS_a_at	AGAP003058-9:9	AGAP029770
3,99	Ag.2R.856.0_CDS_at	AGAP004148-11:11	AGAP004148
3,99	Ag.UNKN.1126.0_CDS_s_at		Protein Rolling-stone like, a confirmar experimentalmente
3,99	Ag.2L.1068.0_CDS_at	AGAP005612-9:9	AGAP005612
3,98	Ag.2L.580.0_CDS_at	AGAP005509-11:11	AGAP005509
3,98	Ag.2L.1491.0_CDS_at	AGAP007354-10:10	AGAP007354
3,96	Ag.X.239.0_CDS_a_at	AGAP000344-10:11	AGAP000344
3,93	Ag.3R.26.0_CDS_at	AGAP009194-11:11	AGAP009194
3,92	Ag.3L.1517.1_a_at	AGAP011460-9:9	AGAP011460
3,92	Ag.X.295.0_CDS_at	AGAP000359-10:11	AGAP000359
3,91	Ag.3R.265.0_UTR_at		AGAP008640
3,89	Ag.UNKN.1342.0_CDS_at		AGAP029133
3,88	Ag.2L.1312.0_CDS_at	AGAP004814-10:10	AGAP004814
3,86	Ag.3R.180.0_CDS_at	AGAP007980-9:10	AGAP007980
3,85	Ag.2L.39.0_CDS_at	AGAP006430-16:8	AGAP006430

3,83	Ag.3R.41.0_CDS_at	AGAP009214-10:10 AGAP009215-2:2	AGAP009214
3,80	Ag.X.279.0_CDS_a_at	AGAP000591-11:11	AGAP000591
3,79	Ag.X.193.1_a_at		AGAP000241
3,78	Ag.3L.18.0_CDS_at	AGAP011789-11:11	AGAP011789
3,76	Ag.2R.461.0_UTR_a_at		AGAP003600
3,73	Ag.2L.397.0_UTR_at	AGAP006441-14:7	AGAP006441
3,72	Ag.2R.1097.0_CDS_a_at	AGAP002359-6:11	AGAP002359
3,72	Ag.2R.3343.0_CDS_at		AGAP013071
3,68	Ag.2L.1407.0_CDS_at		AGAP006056
3,66	Ag.2L.705.0_CDS_a_at	AGAP006193-11:11	AGAP006193
3,61	Ag.2L.391.0_UTR_at		AGAP005170
3,54	Ag.2L.1770.0_CDS_at	AGAP007645-11:11	AGAP007645
3,52	Ag.2R.655.0_CDS_s_at	AGAP001501-11:11	AGAP001501
3,50	Ag.3L.364.0_CDS_s_at	AGAP011788-3:3	AGAP011788
3,49	Ag.2L.604.0_CDS_at	AGAP004861-9:10	AGAP004861
3,48	Ag.3R.194.0_UTR_at		AGAP007781
3,47	Ag.2R.1819.0_CDS_at	AGAP001449-10:11	AGAP001449
3,47	Ag.3R.565.0_CDS_a_at	AGAP007843-9:10	AGAP029445
3,44	Ag.2L.606.1_UTR_s_at	AGAP005124-20:11	AGAP005124
3,44	Ag.X.1396.0_at		AGAP000311
3,43	Ag.3R.1176.0_CDS_at	AGAP007920-10:11	AGAP028055
3,38	AF203339.1_s_at	AGAP006911-11:11	AGAP006911
3,38	Ag.X.226.0_CDS_at	AGAP000987-10:11	AGAP000987
3,38	Ag.2R.424.0_CDS_a_at	AGAP003350-11:11	AGAP003350
3,36	Ag.2R.953.0_CDS_at	AGAP001376-10:10	AGAP001376
3,34	Ag.2R.24.0_CDS_at	AGAP003057-10:10	AGAP003057
3,33	Ag.2L.293.0_CDS_at	AGAP007666-11:11	AGAP007666
3,32	Ag.X.1152.0_s_at		AGAP012956
3,31	Ag.2R.1248.2_CDS_a_at		AGAP013329
3,26	Ag.3L.981.0_CDS_at	AGAP011432-11:11	AGAP011432;AGAP029292
3,26	Ag.3L.1002.0_CDS_at	AGAP011606-11:11	AGAP011605
3,22	Ag.3R.1928.1_CDS_a_at	AGAP009583-16:9	AGAP009583
3,22	Ag.3R.762.2_at		AGAP007836
3,21	Ag.2R.47.0_UTR_at		AGAP001979
3,19	Ag.2L.1544.0_CDS_at	AGAP004857-10:11	AGAP029106

3,15	Ag.2L.191.2_s_at	AGAP006398-11:11	AGAP006398
3,14	Ag.3L.355.0_CDS_a_at	AGAP012404-10:11	AGAP012404
3,14	Ag.2R.666.0_CDS_at	AGAP003814-9:10	AGAP003814
3,14	Ag.2R.192.0_CDS_at	AGAP004567-11:11	AGAP004567
3,13	Ag.2R.865.0_UTR_a_at	AGAP001751-9:9	AGAP001751
3,12	Ag.2R.2454.0_CDS_at	AGAP001659-11:11	AGAP001659
3,11	Ag.3R.415.0_CDS_at	AGAP008570-8:9	AGAP008570
3,11	Ag.2L.941.0_UTR_at	AGAP007730-11:11	AGAP007730
3,10	Ag.X.932.0_CDS_a_at	AGAP000243-9:10	AGAP000243
3,07	Ag.2R.33.1_a_at	AGAP003058-10:11	AGAP029769
3,04	Ag.2R.29.0_CDS_at	AGAP003251-11:11	AGAP003251
3,03	Ag.X.153.0_UTR_s_at		AGAP000011
3,03	Ag.3R.587.0_CDS_at	AGAP009106-7:7	AGAP009106
3,03	Ag.2R.20.0_CDS_at	AGAP001377-10:10	AGAP001377
2,99	Ag.X.80.0_UTR_a_at		AGAP000820
2,99	Ag.2L.2940.0_at		AGAP007674
2,98	Ag.2R.550.0_UTR_a_at		AGAP003141
2,98	Ag.3R.593.0_CDS_a_at	AGAP009634-9:10	AGAP009634
2,98	Ag.3R.2923.0_at		AGAP029459
2,97	Ag.2R.3257.0_CDS_at	AGAP002551-11:11	AGAP002551
2,97	Ag.2R.3178.2_at		AGAP002109
2,96	Ag.3L.2098.0_CDS_s_at	AGAP012355-1:26	AGAP028137;AGAP028156;AGAP028160;AGAP028178
2,96	Ag.2R.1206.0_CDS_at	AGAP003578-10:10	AGAP003578
2,95	Ag.2R.2009.0_CDS_at	AGAP003746-8:8	AGAP003746
2,93	Ag.3L.1250.0_CDS_at	AGAP012172-11:11	AGAP029607
2,93	Ag.3L.2137.0_s_at	AGAP012292-11:11	AGAP012292
2,93	Ag.3R.553.0_CDS_a_at	AGAP009104-9:11	AGAP009104
2,92	Ag.3R.754.0_CDS_a_at	AGAP010066-10:10	AGAP010066;AGAP029419
2,91	Ag.2L.242.0_UTR_at		AGAP029365
2,91	Ag.2R.721.0_CDS_s_at	AGAP012571-9:9	AGAP012571;AGAP013060
2,91	Ag.3R.2644.0_at		AGAP010164
2,89	Ag.3L.1393.0_CDS_at	AGAP010857-11:11	AGAP010857
2,89	Ag.2R.3153.0_CDS_at	AGAP001982-9:9	AGAP001982
2,86	Ag.3R.877.0_CDS_a_at	AGAP008615-11:11	AGAP008615
2,86	Ag.2R.274.0_UTR_a_at		AGAP002358

2,83	Ag.3L.1661.0_CDS_s_at	AGAP012356-9:9	AGAP012356;AGAP028227
2,80	Ag.2R.1242.0_CDS_at	AGAP001346-9:9	AGAP001346
2,79	Ag.3L.252.0_CDS_at	AGAP010769-10:10	AGAP010769
2,79	Ag.3R.631.0_CDS_at	AGAP009281-10:10	AGAP009281
2,79	Ag.3R.745.0_CDS_a_at	AGAP009609-10:10	AGAP009609
2,77	Ag.UNKN.2307.0_at		AGAP012008
2,74	Ag.2R.1082.0_CDS_at	AGAP001763-11:11	AGAP001763
2,72	Ag.3R.673.1_at		AGAP008822
2,72	Ag.2R.1869.0_CDS_at	AGAP002279-9:9	AGAP002279
2,72	Ag.2R.463.1_CDS_at		AGAP004631
2,68	Ag.3R.647.0_UTR_at	AGAP009978-7:7	AGAP009978
2,67	Ag.X.88.0_UTR_at	AGAP000327-9:9	AGAP000327
2,67	Ag.2R.420.1_CDS_a_at	AGAP002857-10:10	AGAP002857
2,66	Ag.2R.509.0_CDS_at	AGAP004451-11:11	AGAP004451
2,64	Ag.3R.76.0_CDS_at	AGAP008835-11:11	AGAP008835
2,62	Ag.3L.529.0_UTR_at	AGAP011118-2:2	AGAP011118
2,62	Ag.2L.1535.0_CDS_s_at	AGAP012530-5:5 AGAP012531-9:10 AGAP004815-9:10	AGAP004815;AGAP012530;AGAP012531
2,62	Ag.3R.868.0_UTR_at		AGAP008321
2,61	Ag.2L.1267.0_CDS_a_at	AGAP006967-11:11	AGAP006967
2,61	Ag.3L.326.0_UTR_a_at		AGAP029666
2,61	Ag.3R.2238.0_CDS_a_at	AGAP008544-11:15	AGAP008544
2,58	Ag.2L.1694.0_CDS_at	AGAP006521-10:11	AGAP006521
2,57	Ag.3L.886.0_CDS_a_at	AGAP010436-11:11	AGAP010436
2,56	Ag.2R.63.0_CDS_at	AGAP002270-11:11	AGAP002270
2,52	Ag.3R.1289.0_CDS_s_at	AGAP009252-8:11	AGAP009251;AGAP009252
2,50	Ag.3R.23.0_CDS_a_at	AGAP009670-11:11	AGAP009670
2,49	Ag.2L.402.0_UTR_a_at		AGAP005091
2,49	Ag.2L.440.0_CDS_a_at	AGAP005153-22:11	AGAP005153
2,48	Ag.X.29.0_CDS_at	AGAP000163-9:10	AGAP000163
2,45	Ag.3R.321.0_UTR_a_at		AGAP009953
2,44	Ag.UNKN.314.0_CDS_at	AGAP004261-11:11	AGAP004261
2,43	Ag.3L.2099.0_CDS_s_at	AGAP012676-8:8 AGAP012355-3:18	AGAP012676;AGAP028067
2,43	Ag.X.4.0_CDS_at	AGAP000278-11:11	AGAP000278
2,40	Ag.3R.286.0_CDS_at	AGAP007921-11:11	AGAP007921
2,40	Ag.2R.3734.0_at		AGAP004071
2,39	Ag.2L.148.0_CDS_at	AGAP007693-10:10	AGAP007693

2,38	Ag.2L.206.0_CDS_a_at	AGAP005234-19:11	AGAP005234
2,38	Ag.2L.496.0_CDS_at	AGAP006082-10:11	AGAP006082
2,37	Ag.3R.704.0_CDS_a_at	AGAP007886-8:8	AGAP007886
2,37	Ag.2L.1121.0_CDS_at	AGAP006906-11:11	AGAP006906
2,37	Ag.3R.633.0_CDS_a_at	AGAP009333-10:10	AGAP009333
2,36	Ag.3R.403.0_UTR_at	AGAP008821-4:4	AGAP008821
2,36	Ag.X.307.0_CDS_a_at	AGAP001021-9:9	AGAP001021
2,35	Ag.2R.322.0_CDS_at	AGAP002685-9:9	AGAP002685
2,34	Ag.3L.893.0_CDS_at	AGAP010395-11:11	AGAP010395
2,33	Ag.2L.3844.0_at		AGAP005847
2,32	Ag.X.354.1_at	AGAP000275-4:9	AGAP012953
2,31	Ag.2L.2221.0_CDS_at	AGAP007372-9:9	AGAP007372
2,31	Ag.X.1445.0_a_at		AGAP000519
2,29	Ag.2L.1026.0_CDS_at	AGAP007589-11:11	AGAP007589
2,29	Ag.2L.2444.0_CDS_a_at	AGAP005634-11:11	AGAP005634
2,26	Ag.3L.160.0_CDS_at	AGAP011172-9:9	AGAP011172
2,25	Ag.2L.274.0_CDS_at	AGAP007104-10:10	AGAP007104
2,25	Ag.3L.1510.0_CDS_s_at	AGAP011430-11:11 AGAP011431-11:11	AGAP011431;AGAP029117;AGAP029747;AGAP029750
2,24	Ag.2L.254.0_CDS_at	AGAP005865-11:11	AGAP005865
2,23	Ag.2R.3332.0_CDS_a_at	AGAP002989-10:10	AGAP002989
2,23	Ag.3L.1005.0_CDS_at	AGAP011641-10:10	AGAP011641
2,22	Ag.2L.4032.0_at		AGAP007065
2,21	Ag.2R.922.0_CDS_at	AGAP002417-10:10	AGAP002417
2,21	Ag.2L.583.0_CDS_at	AGAP005982-10:10	AGAP005982
2,20	Ag.3L.2469.0_at		AGAP029284
2,19	Ag.2R.289.15_UTR_at		AGAP001797
2,19	Ag.3R.85.0_CDS_at	AGAP008360-10:10	AGAP008360
2,19	Ag.2L.610.1_CDS_at	AGAP005556-10:10	AGAP005556
2,19	Ag.3R.213.0_UTR_s_at	AGAP008667-2:2	AGAP008667
2,18	Ag.2L.3095.0_at	AGAP006921-22:11	AGAP006921
2,17	Ag.2L.2014.1_at	AGAP006063-9:11	AGAP006063
2,16	Ag.3R.653.0_UTR_a_at	AGAP007751-11:11	AGAP007751
2,16	Ag.X.142.0_UTR_a_at		AGAP000009
2,16	Ag.3R.2409.0_CDS_a_at	AGAP009213-11:11	AGAP009213
2,15	Ag.3L.1552.0_CDS_at	AGAP011655-11:11	AGAP011655
2,14	Ag.3L.456.0_CDS_a_at	AGAP011810-11:11	AGAP011810

2,13	Ag.3L.26.0_CDS_at	AGAP012295-11:11	AGAP012295
2,12	Ag.3R.738.1_at		AGAP009210
2,12	Ag.2L.2714.0_CDS_at	AGAP007044-11:11	AGAP007044
2,11	Ag.3R.3557.0_at		AGAP008794
2,11	Ag.3R.1101.0_CDS_a_at	AGAP008983-9:10	AGAP008983
2,11	Ag.3L.221.0_CDS_a_at	AGAP011937-11:11	AGAP011937
2,10	Ag.3R.783.0_CDS_at	AGAP008403-10:11	AGAP008403
2,10	Ag.3L.313.2_CDS_a_at	AGAP011725-10:10	AGAP029606
2,09	Ag.2L.69.0_CDS_at	AGAP005716-9:9	AGAP005716
2,09	Ag.3R.1042.0_UTR_at		AGAP007754
2,09	Ag.2L.3034.0_at		AGAP005242
2,09	Ag.2L.355.0_UTR_at		AGAP006901
2,08	Ag.2R.445.0_UTR_a_at		AGAP003687
2,08	Ag.3L.1605.0_CDS_a_at	AGAP011972-10:11	AGAP011972
2,08	Ag.3L.455.2_CDS_a_at	AGAP011798-8:9	AGAP029758
2,07	Ag.2L.1033.0_CDS_at	AGAP004784-9:10	AGAP004784
2,06	Ag.3L.1802.0_CDS_at	AGAP012696-11:11	AGAP012696;AGAP029783
2,03	Ag.X.11.0_UTR_at	AGAP000573-3:3	AGAP000573
2,02	Ag.2R.1394.1_CDS_at	AGAP002457-8:11	AGAP002457
2,02	Ag.2R.1350.0_CDS_at	AGAP004500-11:11	AGAP004500
2,01	Ag.2R.1127.0_CDS_a_at	AGAP003680-10:10	AGAP003680
2,01	Ag.3L.197.0_CDS_at	AGAP011781-10:10	AGAP011781
2,00	Ag.UNKN.1520.0_CDS_at	AGAP012614-11:11	CLIP domain-containing serine protease 14D-like, AGAP003249-like
1,98	Ag.3R.374.1_UTR_a_at	AGAP009976-10:11	AGAP028649
1,98	Ag.3L.1301.0_CDS_at	AGAP010369-4:7	AGAP010369
1,97	Ag.2L.1146.0_CDS_at	AGAP007564-10:10	AGAP007564
1,97	Ag.2L.1210.0_CDS_at	AGAP006015-11:11	AGAP006015
1,94	Ag.3L.2293.0_at		AGAP029560
1,93	Ag.2R.2381.0_CDS_at	AGAP001203-11:11	AGAP001203
1,92	Ag.X.766.0_CDS_at	AGAP000572-7:7	AGAP000572
1,90	Ag.2R.464.0_UTR_at		AGAP001313
1,90	Ag.3L.462.2_CDS_s_at		AGAP010445
1,89	Ag.3L.1500.0_CDS_s_at	AGAP012772-9:10 AGAP011388-9:10	AGAP011388;AGAP012772
1,88	Ag.3R.282.1_a_at		AGAP008054
1,88	Ag.3R.1582.0_CDS_a_at	AGAP010280-11:11	AGAP010280

1,88	Ag.2R.379.0_CDS_a_at		AGAP013365
1,87	Ag.3R.1753.1_at		AGAP029610
1,85	Ag.3L.919.0_CDS_at	AGAP010784-11:11	AGAP010784
1,85	Ag.2L.784.0_CDS_at	AGAP007140-10:10	AGAP007140
1,85	Ag.2R.638.0_CDS_at	AGAP002457-11:11	AGAP013260
1,85	Ag.3R.210.0_CDS_at	AGAP009651-9:10	AGAP009651
1,84	Ag.2L.843.2_CDS_s_at	AGAP007534-30:10	AGAP007534
1,82	Ag.2L.238.0_UTR_at	AGAP005175-22:11	AGAP005175
1,82	Ag.2R.4942.0_at		AGAP002438
1,82	Ag.2R.2486.0_CDS_at	AGAP001864-1:12	AGAP001864
1,82	Ag.3R.2662.0_s_at	AGAP007790-6:6	AGAP007790
1,82	Ag.3R.2431.0_CDS_at	AGAP009282-9:10	AGAP009282
1,82	Ag.3L.124.0_UTR_at		AGAP010957
1,81	Ag.2L.176.0_CDS_a_at	AGAP007593-18:9	AGAP007593
1,80	Ag.3R.1937.0_CDS_at	AGAP009616-10:11	AGAP009616
1,80	Ag.2L.744.0_UTR_a_at	AGAP005324-22:11	AGAP005324
1,80	Ag.2R.1625.0_CDS_at	AGAP004450-11:11	AGAP004450
1,80	Ag.2R.98.0_CDS_at	AGAP004383-10:10	AGAP004383
1,79	Ag.2R.2608.0_CDS_at	AGAP002552-10:11	AGAP002552
1,78	Ag.2R.2884.0_CDS_at		AGAP013134
1,77	Ag.3R.334.0_CDS_a_at	AGAP008931-10:10	AGAP008931
1,77	Ag.2R.530.0_CDS_at	AGAP002425-11:11	AGAP002425
1,77	Ag.UNKN.415.0_CDS_at	AGAP012821-5:5 AGAP007075-8:8	AGAP012821
1,74	Ag.3R.3668.0_at		AGAP009510
1,74	Ag.X.320.0_UTR_at		AGAP000410
1,73	Ag.3L.12.0_UTR_s_at	AGAP010404-10:10	AGAP010404
1,73	Ag.2R.506.0_CDS_at	AGAP003408-9:9	AGAP003408
1,71	Ag.3R.7.2_a_at	AGAP008051-11:11	AGAP008051
1,71	Ag.2R.749.0_CDS_a_at	AGAP001411-9:9	AGAP001411
1,71	Ag.2L.2212.0_CDS_at		AGAP007327
1,69	Ag.2R.566.0_CDS_at	AGAP003960-6:11	AGAP003960
1,69	Ag.2L.483.0_UTR_at	AGAP006615-11:11	AGAP006615
1,69	Ag.2L.1589.0_CDS_at	AGAP005311-10:10	AGAP005311
1,69	Ag.2R.3349.0_CDS_at	AGAP003072-10:11	AGAP003070
1,69	Ag.2R.1537.0_CDS_at	AGAP002419-11:11	AGAP002419
1,69	Ag.3L.245.0_UTR_a_at		AGAP010640

1,68	Ag.3R.1104.1_a_at		AGAP028095
1,68	Ag.3R.655.0_CDS_at	AGAP007920-7:10	AGAP007920
1,67	Ag.X.1500.0_at		AGAP000707
1,67	Ag.2R.992.0_UTR_at	AGAP003263-11:11	AGAP003263
1,67	Ag.2L.1771.0_CDS_at	AGAP007651-11:11	AGAP007651
1,67	Ag.2L.2003.0_CDS_at	AGAP006007-10:11	AGAP006007
1,65	Ag.2L.1754.0_CDS_at	AGAP007384-10:10	AGAP007384
1,64	Ag.2R.4959.0_at		AGAP029378
1,64	Ag.3L.378.0_CDS_at	AGAP011119-10:10	AGAP011119
1,62	Ag.2R.571.0_UTR_s_at	AGAP001616-6:6	AGAP001616
1,61	Ag.3L.595.0_CDS_a_at	AGAP010714-9:10	AGAP010714
1,61	Ag.2L.469.0_CDS_at	AGAP007258-11:11	AGAP007258
1,61	Ag.2R.3721.0_at	AGAP001889-2:3	AGAP001889
1,60	Ag.2L.1467.0_CDS_at	AGAP006972-9:10 AGAP006973-9:10	AGAP006972;AGAP006973
1,57	Ag.3R.399.0_UTR_at		AGAP029707
1,57	Ag.3L.2023.0_CDS_at	AGAP011958-10:10	AGAP011958
1,57	Ag.2R.95.0_CDS_at	AGAP003738-10:10	AGAP003738
1,57	Ag.2L.1150.0_CDS_at	AGAP007587-18:10	AGAP007587
1,57	Ag.2L.183.0_CDS_a_at	AGAP007209-50:10	AGAP007209
1,56	Ag.2R.3304.0_CDS_at	AGAP002855-3:3	AGAP013529
1,55	Ag.2L.2168.0_CDS_at	AGAP007077-10:11	AGAP007077
1,55	Ag.3R.258.0_UTR_a_at	AGAP009224-6:6	AGAP009224
1,55	Ag.2R.1977.0_CDS_at	AGAP003487-10:10	AGAP003487
1,55	Ag.X.322.0_UTR_at		AGAP000440
1,55	Ag.2R.298.13_UTR_a_at	AGAP002858-7:7	AGAP002858
1,55	Ag.3R.2629.0_at		AGAP007881
1,55	Ag.X.16.0_UTR_a_at		AGAP000500
1,55	Ag.2R.722.0_CDS_at	AGAP003252-10:10	AGAP003252
1,54	Ag.2L.870.0_CDS_at	AGAP005350-10:11	AGAP005350
1,54	Ag.X.198.0_UTR_at		AGAP000769
1,54	Ag.3R.882.0_CDS_at	AGAP008813-10:10	AGAP008813
1,54	Ag.X.1147.0_at		AGAP000726
1,53	Ag.2R.71.2_CDS_at	AGAP004164-11:11	AGAP004164
1,53	Ag.2R.327.0_CDS_a_at	AGAP001730-10:11	AGAP001730
1,51	Ag.3L.286.0_CDS_at	AGAP010733-10:10	AGAP010733
1,50	Ag.3R.3257.0_at		AGAP009466

1,50	Ag.2L.2345.0_CDS_at	AGAP005137-11:11	AGAP005137
1,49	Ag.3R.567.0_UTR_a_at		AGAP007970
1,49	Ag.2L.544.0_CDS_at	AGAP004859-10:10	AGAP004859
1,49	Ag.2L.2956.0_at		AGAP005976
1,49	Ag.2R.994.1_UTR_a_at		AGAP003492
1,47	Ag.2L.226.0_CDS_at	AGAP005888-10:11	AGAP005888
1,45	Ag.2L.377.1_CDS_a_at	AGAP007006-18:9	AGAP007006
1,45	Ag.X.486.0_CDS_at		AGAP013476
1,45	Ag.2L.820.0_UTR_at	AGAP006043-9:9	AGAP029130
1,45	Ag.3L.2135.1_a_at	AGAP011832-3:3	AGAP011832
1,45	Ag.X.1179.0_at		AGAP000380
1,44	Ag.2R.2203.0_CDS_at	AGAP002691-7:8	AGAP002691
1,43	Ag.UNKN.53.0_UTR_a_at	AGAP010623-1:1	AGAP010623
1,43	Ag.UNKN.2292.0_s_at	AGAP012613-10:10 AGAP012740-10:10 AGAP012926-10:10	AGAP012613;AGAP012740;AGAP012926
1,42	Ag.2L.2761.0_CDS_at	AGAP007281-8:8	AGAP007281
1,42	Ag.UNKN.128.0_CDS_at	AGAP012717-10:11	AGAP012717
1,40	Ag.2L.17.0_UTR_at	AGAP007121-11:11	AGAP007121
1,40	Ag.3R.36.0_CDS_s_at	AGAP007940-10:11	AGAP007940
1,39	Ag.3R.1005.0_CDS_a_at	AGAP009648-11:11	AGAP009648
1,39	Ag.2L.1457.0_CDS_at	AGAP006867-11:11	AGAP006867
1,39	Ag.X.313.0_UTR_at		AGAP000222
1,38	Ag.2L.669.0_CDS_at	AGAP007314-9:9	AGAP007314
1,38	Ag.3L.425.0_UTR_at		AGAP010696
1,37	Ag.2R.352.0_UTR_a_at		AGAP003153
1,37	Ag.3R.39.0_CDS_a_at	AGAP009197-10:11	AGAP009197
1,37	Ag.2L.1020.1_CDS_a_at	AGAP007452-20:10	AGAP007452
1,37	Ag.2R.539.0_UTR_at	AGAP003968-5:5	AGAP003968
1,37	Ag.2L.299.0_CDS_at	AGAP004860-10:10	AGAP004860
1,37	Ag.2R.2828.0_CDS_s_at	AGAP012801-11:11 AGAP003516-11:11	AGAP003516;AGAP012801
1,37	Ag.3L.124.0_CDS_s_at	AGAP012844-11:11 AGAP010957-10:11	AGAP010957;AGAP012844
1,37	Ag.3L.273.0_CDS_at		AGAP028038
1,36	Ag.3L.1231.2_s_at		AGAP029495
1,36	Ag.2R.108.0_CDS_a_at	AGAP001189-11:11	AGAP001189
1,36	Ag.3L.680.2_at		AGAP012139
1,36	Ag.3R.2058.0_CDS_s_at	AGAP010111-10:10 AGAP010118-11:11 AGAP010125-11:11	AGAP010111;AGAP010118;AGAP010125

1,35	Ag.2L.184.0_UTR_a_at		AGAP006023
1,35	Ag.X.71.2_s_at	AGAP000180-11:11	AGAP000180
1,34	Ag.2L.3463.0_at		AGAP006386
1,34	Ag.2R.2846.0_CDS_at	AGAP003627-11:11	AGAP003627
1,34	Ag.X.31.0_CDS_at	AGAP000571-10:10	AGAP000315
1,33	Ag.2R.2345.0_CDS_a_at	AGAP004397-10:10	AGAP004397
1,33	Ag.X.1646.0_at		AGAP000237
1,32	Ag.2L.1493.0_CDS_a_at	AGAP007358-22:11	AGAP007358
1,31	Ag.3R.3531.0_at		AGAP008636
1,31	Ag.2R.830.0_CDS_at	AGAP002032-10:10	AGAP002032
1,30	Ag.X.14.1_CDS_a_at	AGAP000165-11:11	AGAP000165
1,29	Ag.2R.1456.0_CDS_at	AGAP004144-10:10	AGAP004144
1,28	Ag.3L.512.0_UTR_a_at	AGAP012406-10:10	AGAP012406
1,28	Ag.2L.727.1_a_at	AGAP004769-9:10	AGAP004769
1,28	Ag.2R.848.0_CDS_s_at	AGAP003238-10:10	AGAP003238
1,27	Ag.3R.2681.0_at		AGAP028049
1,27	Ag.2R.3758.0_at		AGAP003839
1,27	Ag.2R.4371.0_at		AGAP002644
1,27	Ag.2R.1147.0_CDS_a_at	AGAP004549-11:11	AGAP004549
1,25	Ag.2R.2029.0_UTR_at	AGAP003942-11:11	AGAP003942
1,25	Ag.3R.3250.0_at		AGAP009403
1,24	Ag.3R.197.1_CDS_s_at	AGAP010131-10:10	AGAP010131
1,24	Ag.2R.740.0_UTR_a_at		AGAP004372
1,23	Ag.X.9.0_CDS_a_at	AGAP000947-11:11	AGAP000947
1,22	Ag.3L.1820.1_at		AGAP011070
1,22	Ag.3R.1631.0_CDS_at	AGAP007982-11:11	AGAP007982
1,22	Ag.3L.2065.0_CDS_at	AGAP012181-9:9	AGAP012181
1,22	Ag.2L.303.0_UTR_a_at		AGAP005400
1,21	Ag.2L.1187.0_CDS_at	AGAP005515-11:11	AGAP005515
1,21	Ag.3R.1011.0_CDS_a_at	AGAP009791-9:9	AGAP009791
1,21	Ag.3L.538.0_CDS_at	AGAP011493-11:11	AGAP011493
1,20	Ag.2L.550.0_CDS_at	AGAP005553-10:10	AGAP005553
1,20	Ag.2R.1105.0_CDS_at	AGAP002638-10:10	AGAP002638
1,19	Ag.3R.63.0_CDS_at	AGAP008844-11:11	AGAP008844
1,18	Ag.3L.2426.0_at		AGAP010805
1,17	Ag.3R.1274.0_CDS_a_at	AGAP009051-10:10	AGAP009051

1,16	Ag.X.479.1_a_at	AGAP000297-8:10	AGAP000297
1,15	Ag.UNKN.1138.0_CDS_s_at	AGAP012746-9:10 AGAP003514-9:10	AGAP003514;AGAP012746
1,15	Ag.3L.13.0_CDS_a_at	AGAP011792-10:10	AGAP011792
1,15	Ag.3R.3455.0_at		AGAP008141
1,15	Ag.2R.2522.0_CDS_a_at	AGAP002113-22:11	AGAP002113
1,14	Ag.3R.612.2_a_at	AGAP008014-10:10	AGAP008014
1,14	Ag.UNKN.177.1_a_at	AGAP012154-9:9	solute carrier family 15 member-like
1,14	Ag.2R.360.0_UTR_s_at		AGAP001548
1,13	Ag.2R.3186.0_CDS_at		AGAP028402
1,13	Ag.2R.3519.0_CDS_at	AGAP003969-11:11	AGAP029442
1,12	Ag.X.122.0_CDS_a_at	AGAP000577-11:11	AGAP000577
1,12	Ag.2R.571.0_CDS_s_at	AGAP001617-13:9	AGAP001617
1,11	Ag.UNKN.1847.0_CDS_at	AGAP003784-10:10	AGAP003784
1,11	Ag.2L.929.0_CDS_a_at	AGAP007340-20:11	AGAP007340
1,11	Ag.2L.1333.0_CDS_at	AGAP005093-11:11	AGAP005093
1,11	Ag.2R.706.0_UTR_a_at		AGAP001520
1,11	Ag.X.658.0_CDS_at		AGAP013356
1,11	Ag.2L.1716.0_CDS_at	AGAP006898-11:11	AGAP006898
1,10	Ag.3R.1661.0_CDS_at	AGAP008181-11:11	AGAP008181
1,10	Ag.3L.861.0_UTR_at	AGAP012060-8:8	AGAP029650
1,10	Ag.UNKN.50.2_CDS_a_at	AGAP001082-11:11	AGAP001082
1,09	Ag.2R.2090.0_CDS_at		AGAP001206
1,09	Ag.3R.3047.0_at		AGAP008137
1,08	Ag.2L.3119.0_at	AGAP004796-9:9	AGAP004795
1,08	Ag.3R.1992.0_CDS_at	AGAP009754-3:3 AGAP009755-9:9 AGAP009756-11:11	AGAP009755;AGAP009756
1,08	Ag.3R.664.0_CDS_at	AGAP008350-10:11	AGAP008350
1,08	Ag.3L.2205.0_at		AGAP011771
1,08	Ag.2R.1263.0_UTR_at	AGAP002036-1:1	AGAP002036
1,08	Ag.3R.310.0_UTR_a_at	AGAP008419-11:11	AGAP008419
1,07	Ag.3R.649.0_UTR_at		AGAP010189
1,07	Ag.3R.274.0_UTR_a_at		AGAP009486
1,07	Ag.3L.2086.0_CDS_at	AGAP012270-11:11	AGAP012270
1,07	Ag.3L.212.0_UTR_a_at		AGAP011334
1,07	Ag.2R.714.0_CDS_a_at	AGAP002357-10:11	AGAP002357
1,07	Ag.2R.390.0_CDS_at	AGAP003334-5:7	AGAP003334

1,06	Ag.X.412.0_CDS_a_at	AGAP001069-11:11	AGAP001069
1,06	Ag.3L.2136.1_a_at		AGAP028154
1,06	Ag.2R.813.0_CDS_s_at	AGAP012662-9:10	AGAP012662;AGAP013231
1,05	Ag.2R.5024.0_at		AGAP002898
1,05	Ag.3L.2635.0_at		AGAP012291
1,05	Ag.2L.601.1_a_at	AGAP007721-30:10	AGAP007721
1,05	Ag.2L.421.0_UTR_a_at		AGAP029668
1,04	Ag.2L.1684.0_CDS_a_at	AGAP006395-11:11	AGAP006395
1,04	Ag.2L.53.0_CDS_at	AGAP005332-26:10	AGAP029559
1,04	Ag.3L.929.0_CDS_at	AGAP010853-10:10	AGAP029611
1,04	Ag.3R.2331.1_a_at	AGAP008923-20:10	AGAP008923;AGAP028186
1,04	Ag.2L.2372.0_CDS_at	AGAP005235-9:10	AGAP005235
1,03	Ag.3R.2690.0_a_at		AGAP009619
1,03	Ag.X.345.0_UTR_at	AGAP000034-2:2	AGAP000035
1,03	Ag.X.1155.0_at	AGAP000546-11:11	AGAP000546
1,02	Ag.2R.819.0_CDS_a_at	AGAP004532-11:11	AGAP004532
1,02	Ag.2R.580.0_UTR_a_at		AGAP002428
1,02	Ag.3R.3267.0_at		AGAP009522
1,02	Ag.3L.262.0_UTR_a_at	AGAP010893-8:8	AGAP010893
1,01	Ag.2R.4634.0_at		AGAP004211
1,01	Ag.2R.1747.0_CDS_a_at	AGAP003513-9:11	AGAP003513
1,00	Ag.3L.552.0_UTR_a_at		AGAP012097
1,00	Ag.3R.1133.0_CDS_at	AGAP009739-11:11	AGAP009739

### ANEXO III

Passo a passo - GREP: Para obter as sequências protéicas através do proteoma OGS3

Comando:

```
grep -A 1 -f test.txt Tcas5_2_prot.txt > output.xls
```

Onde:

*-A 1 significa que vai pegar uma linha após (After) a linha do string/query/busca*

*-f significa que o grep vai pegar o string/query/busca de dentro de um arquivo*

*test.txt é o arquivo que contém o string/query/busca*

*Tcas5\_2\_prot.txt é o arquivo com o proteoma inteiro de onde quero extrair as sequências que nos interessa significa que o resultado de saída será "impresso" em um arquivo a ser definido*

*output.xls significa o arquivo de saída; pode ser também output.txt.*

## ANEXO IV

Passo a passo – AWK: Para obter as sequências em uma única linha.

```
awk '/^>/ {printf("\n%s\n",$0);next; } { printf("%s",$0);} END {printf("\n");}' <  
AgamProteom47.fasta | tail -n +2 > AgamProteom47_one_line.fasta
```

Esse comando vai gerar um arquivo contendo a sequência de aminoácidos em apenas uma linha (além do cabeçalho)

## ANEXO V

**30 sondas órfãs com suas respectivas informações para serem analisadas e relacionadas a genes no futuro:**

<b>Avlog2fol d</b>	<b>Sonda</b>	<b>ID do gene - Descrição</b>
3,888	Ag.X.1629.0_at	AGAP000198, voltar depois no VectorBase/NCBI - Chitinase, talvez
3,631	Ag.UNKN.2704.0_at	checar RNASeq VectorBase/NCBI, fazer BlastX
3,060	Ag.3L.2464.0_at	checar RNASeq VectorBase/NCBI, fazer BlastX
2,962	Ag.UNKN.725.0_CDS_at	Trabalhar com região AAAB, fazer BlastX
2,877	Ag.2R.4812.0_at	fazer BlastX
2,770	Ag.3R.2743.0_at	AGAP007970, voltar depois no VectorBase/NCBI - autophagy related gene
2,705	Ag.2R.442.0_UTR_at	AGAP003308, olhar BlastN de novo - cuticular protein
2,688	Ag.2L.2815.0_at	AGAP005710? olhar no VectorBase/NCBI, fazer BlastsX - prominin-like protein
2,333	Ag.X.1156.0_at	Olhar VectorBase/NCBI, fazer BlastX
2,192	Ag.UNKN.378.0_CDS_at	fazer BlastX- Cuticular protein family? Specific to <i>Anopheles gambiae</i> species complex?
1,974	Ag.2R.2733.0_CDS_s_at	fazer BlastX
1,679	Ag.3R.3573.0_at	São muitas regiões do genoma! Transposon?
1,649	Ag.3L.156.0_UTR_s_at	fazer BlastX
1,611	Ag.2R.4494.0_at	voltar, olhar SNAP Prediction
1,532	Ag.2R.4570.0_at	fazer BlastX
1,473	Ag.X.1616.0_a_at	fazer BlastX
1,426	Ag.2R.5141.0_at	fazer BlastX
1,332	Ag.3R.3626.0_at	Voltar depois no VectorBase/NCBI, Fazer BlastX
1,252	Ag.2L.3355.0_at	fazer BlastX
1,239	Ag.UNKN.854.0_CDS_at	Se associa a cinco regiões, fazer BlastX
1,225	Ag.2L.2832.0_at	fazer BlastX
1,221	Ag.UNKN.587.0_CDS_s_at	Se associa a 6 regiões, várias predições corretas do SNAP, fazer BlastX
1,176	Ag.2L.3383.0_at	São muitas regiões do genoma! Transposon?
1,167	Ag.2R.4736.0_at	fazer BlastX
1,152	Ag.3R.3428.0_at	?AGAP007971? voltar depois, olhar SNAP Predictions, VectorBase, NCBI
1,137	Ag.X.1133.0_at	AGAP000198? voltar depois no VectorBase, NCBI, fazer BlastX - Chitinase, talvez.
1,094	Ag.UNKN.2279.0_at	AGAP001616, fazer BlastX na região AABB
1,071	Ag.3L.2287.0_at	voltar depois, VectorBase, NCBI, fazer BlastX, entender SNAP
1,069	Ag.2L.3037.0_at	AGAP005245? voltar depois, VectorBase, NCBI, fazer BlastX
1,008	Ag.2R.3814.0_at	voltar depois, VectorBase, NCBI, fazer Blast X

## ANEXO VI

Descrição e função dos 472 genes expressos na serosa do *A. gambiae*. Para a descrição e definição de função foram utilizadas as informações presentes no Ensembl. Quando se fez necessário, outros bancos de dados foram empregados (FlyBase - para dados de *D. melanogaster* - e PFAM).

AvgLog 2Fold	ID gene	Descrição Ensembl	Ortólogo 1:1 em <i>Drosophila</i> , ID	Ortólogo 1:1 em <i>Drosophila</i> , nome completo	Descrição PFAM	Função
5,78	AGAP010799	not available	n/a	n/a		Desconhecida
5,41	AGAP001984	not available	FBgn0038894	Juvenile hormone binding protein 16		Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
5,37	AGAP011812	elongation of very long chain fatty acids	FBgn0038983	n/a	-	Metabolismo de Lipídeos
5,07	AGAP011936	not available	FBgn0260653	serpentine		Metabolismo de quitina
4,87	AGAP002962	not available	FBgn0030503	Transport and Golgi organization 2		Organização de Golgi
4,85	AGAP000697	not available	FBgn0052694	n/a		Desconhecida
4,78	AGAP008059	chemosensory protein 1	n/a	n/a	-	Proteína quimiossensorial
4,77	AGAP009790	cuticular protein	FBgn0027600	obstructor-B	-	Proteína cuticular
4,75	AGAP010370	solute carrier family 17 (anion/sugar transporter), member 5	n/a	n/a		Transporte transmembranar
4,69	AGAP003250	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise, via de Toll
4,67	AGAP003308	cuticular protein	FBgn0026077	Gasp	-	Proteína cuticular
4,67	AGAP006410	not available	n/a	n/a		Desconhecida
4,63	AGAP001508	not available	n/a	n/a		Desconhecida
4,60	AGAP000821	not available	FBgn0052816	n/a		Desconhecida
4,56	AGAP007412	C-type lectin (CTL) - mannose binding	n/a	n/a		Metabolismo de Carboidratos
4,54	AGAP010928	tetratricopeptide repeat protein 36 homolog	FBgn0036359	n/a	-	Metabolismo de aminoácidos
4,51	AGAP000989	cuticular protein	FBgn0031097	obstructor-A	-	Proteína cuticular
4,51	AGAP002085	not available	FBgn0051997	n/a	SVWC	Miscelânea
4,46	AGAP011992	lysosomal acid lipase	n/a	n/a	-	Metabolismo de Lipídeos
4,44	AGAP029052	not available	FBgn0261015	PFTAIRE-interacting factor 1A		Relacionado ao citoesqueleto
4,42	AGAP011778	not available	n/a	n/a		Desconhecida
4,41	AGAP003790	annexin B9	FBgn0000083	Annexin B9	-	Relacionado ao Citoesqueleto
4,39	AGAP012946	not available	n/a	n/a		Desconhecida
4,38	AGAP009635	not available	n/a	n/a		Desconhecida

4,37	AGAP005552	peptidoglycan recognition protein (long)	FBgn0260458	Peptidoglycan recognition protein LD		Metabolismo de Carboidratos
4,36	AGAP009745	Sugar transporter ERD6-like 4	n/a	n/a		Transporte transmembranar de açúcar
4,35	AGAP001101	carboxylesterase	FBgn0029690	n/a	-	Carboxylesterase
4,33	AGAP013005	not available	n/a	n/a		Desconhecida
4,32	AGAP004263	takeout 1	n/a	n/a		Relacionado ao relógio circadiano
4,32	AGAP004373	elongation of very long chain fatty acids protein 1	FBgn0051522	n/a	-	Metabolismo de Lipídeos
4,30	AGAP011780	CLIP-domain serine protease	FBgn0058160	n/a	-	Protease/Proteólise
4,30	AGAP007636	phosphatidate phosphatase LPIN	FBgn0263593	Lipin	-	Metabolismo de Lipídeos e Metabolismo energético
4,29	AGAP010385	not available	n/a	n/a		Desconhecida
4,28	AGAP004802	4-hydroxyphenylpyruvate dioxygenase	FBgn0036992	4-hydroxyphenylpyruvate dioxygenase		Metabolismo de tirosina
4,24	AGAP010441	not available	n/a	n/a		Desconhecida
4,24	AGAP001718	not available	n/a	n/a		Desconhecida
4,21	AGAP009896	proton-coupled amino acid transporter	FBgn0036116	n/a		Transporte transmembranar de Aminoácidos
4,21	AGAP008988	Glutamine synthetase	FBgn0001142	Glutamine synthetase 1		Metabolismo de Glutamina
4,21	AGAP006151	not available	n/a	n/a		Desconhecida
4,19	AGAP011357	All-trans/9-cis/11-cis	FBgn0038610	n/a	-	Metabolismo de Lipídeos
4,15	AGAP006347	potassium voltage-gated channel KQT-like subfamily, invertebrate	FBgn0033494	KCNQ potassium channel		Canal de potássio dependente de voltagem
4,15	AGAP003606	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
4,15	AGAP003515	not available	n/a	n/a		Metabolismo de Aminoácidos
4,14	AGAP007717	not available	n/a	n/a		Desconhecida
4,13	AGAP001719	not available	n/a	n/a		Desconhecida
4,12	AGAP029673	not available	n/a	n/a		Desconhecida
4,11	AGAP000275	epidermal retinal dehydrogenase	n/a	n/a	-	Metabolismo de Lipídeos
4,10	AGAP001510	not available	n/a	n/a		Desconhecida
4,07	AGAP000988	cuticular protein	n/a	n/a	-	Proteína cuticular
4,06	AGAP003620	not available	n/a	n/a		Desconhecida
4,05	AGAP002738	Class B Scavenger Receptor (CD36 domain).	FBgn0058006	n/a		Miscelânea
4,04	AGAP003733	alpha-tocopherol transfer protein-like protein	n/a	n/a		Metabolismo de lipídeos, Vitamina E

4,03	AGAP003194	serine protease inhibitor (serpin) 8	FBgn0262057	Serpin 77Ba		Inibidor de Serina Protease, regulador negativo da melanização
4,01	AGAP029770	not available	n/a	n/a		Protease/Proteólise
3,99	AGAP004148	CLIP-domain serine protease	FBgn0000533	easter		Protease/Proteólise, via de Toll
3,99	Protein Rolling-stone like, a confirmar experimentalmente					Proteína Rolling-stone like, função indefinida
3,99	AGAP005612	not available	n/a	n/a		Desconhecida
3,98	AGAP005509	Ufm1-specific protease 1	FBgn0050157	UFM1 specific peptidase 1	-	Thiol Protease
3,98	AGAP007354	phosphatidylinositol transfer protein SEC14	n/a	n/a	-	Metabolismo de Lipídeos
3,96	AGAP000344	cuticular protein RR-1 family 127	FBgn0039481	Cuticular protein 97Eb	-	Proteína cuticular
3,93	AGAP009194	glutathione S-transferase epsilon class 2	n/a	n/a	-	Metabolismo de glutationa
3,92	AGAP011460	salivary cysteine-rich protein	FBgn0050197	n/a		Atividade inibidora de peptidase
3,92	AGAP000359	venom allergen	FBgn0052499	Chitin deacetylase-like 4		Metabolismo de Quitina
3,91	AGAP008640	not available	n/a	n/a		Via de BMP
3,89	AGAP029133	not available	n/a	n/a		Desconhecida
3,88	AGAP004814	not available	n/a	n/a		Desconhecida
3,86	AGAP007980	cuticular protein CPCFC family (CPCFC1)	FBgn0033308	n/a	-	Proteína cuticular
3,85	AGAP006430	C-type lectin (CTL) - galactose binding	FBgn0038017	n/a		Metabolismo de Carboidratos
3,83	AGAP009214	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
3,80	AGAP000591	decaprenyl-diphosphate synthase subunit 1	n/a	n/a		Biossíntese de Isoprenóides, e Ubiquinona
3,79	AGAP000241	N-acetylglucosamine kinase	FBgn0038321	N-acetylglucosamine kinase		Metabolismo de Carboidratos
3,78	AGAP011789	CLIP-domain serine protease	n/a	n/a		Protease/Proteólise
3,76	AGAP003600	Elongation of very long chain fatty acids protein	FBgn0037534	ELOVL fatty acid elongase 7	-	Metabolismo de Lipídeos
3,73	AGAP006441	GTP cyclohydrolase I	FBgn0003162	Punch		Via de Melanização/Esclerotização
3,72	AGAP002359	carbonic anhydrase	n/a	n/a	-	Anidrase carbônica
3,72	AGAP013071	Stearoyl-CoA desaturase 5	n/a	n/a	-	Metabolismo de Lipídeos / ácidos graxos

3,68	AGAP006056	not available	FBgn0028380	faltan		Proteína G
3,66	AGAP006193	not available	n/a	n/a		Desconhecida
3,61	AGAP005170	not available	FBgn0034067	n/a	Reeler, DOMON, Cytochrom_B 561	Miscelânea
3,54	AGAP007645	not available	n/a	n/a		Desconhecida
3,52	AGAP001501	Regucalcin protein	n/a	n/a	-	Sinalização de cálcio
3,50	AGAP011788	CLIP-domain serine protease	n/a	n/a		Protease/Proteólise
3,49	AGAP004861	Tetraspanin	FBgn0035936	Tetraspanin 66E		Miscelânea
3,48	AGAP007781	eukaryotic translation initiation factor 4E-binding protein 1	FBgn0261560	Thor		Regulação da tradução
3,47	AGAP001449	not available	n/a	n/a		Metabolismo de Carboidratos
3,47	AGAP029445	O-acyltransferase	FBgn0004797	midway	-	Metabolismo de Lipídeos
3,44	AGAP005124	aldehyde dehydrogenase	n/a	n/a		Oxidação de aldeídos
3,44	AGAP000311	UNC93-like protein MFSD11	FBgn0038053	n/a		Miscelânea
3,43	AGAP028055	UDP-glucuronosyltransferase	FBgn0015663	UDP- glycosyltransferase family 36 member A1		Metabolismo de compostos endógenos e xenobióticos.
3,38	AGAP006911	serine protease inhibitor (serpin) 2	n/a	n/a		Inibidor de Serina Protease, regulador negativo da melanização
3,38	AGAP000987	cuticular protein	n/a	n/a	-	Proteína cuticular
3,38	AGAP003350	phosphoenolpyruvate carboxykinase (GTP)	FBgn0003067	Phosphoenolpyruvate carboxykinase 1		Via da Gliconeogênese
3,36	AGAP001376	serine protease inhibitor (serpin) 17	n/a	n/a		Inibidor de Serina Protease
3,34	AGAP003057	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
3,33	AGAP007666	calcyphosin-like protein	n/a	n/a	-	Sinalização de cálcio
3,32	AGAP012956	not available	n/a	n/a		Desconhecida
3,31	AGAP013329	not available	n/a	n/a		Metabolismo de Carboidratos
3,26	AGAP011432	not available	n/a	n/a		Desconhecida
3,26	AGAP029292	not available	n/a	n/a		Desconhecida
3,26	AGAP011605	not available	n/a	n/a		Desconhecida
3,22	AGAP009583	not available	FBgn0032084	n/a		Proteína Rolling-stone like, função indefinida
3,22	AGAP007836	sodium-coupled monocarboxylate transporter 2	n/a	n/a		Transporte transmembranar, íons sódio
3,21	AGAP001979	Class A Scavenger Receptor (SRCR domain) with Serine Protease domain	n/a	n/a		Protease/Proteólise, Miscelânea

3,19	AGAP029106	not available	n/a	n/a		Desconhecida
3,15	AGAP006398	not available	FBgn0265186	n/a		Desconhecida
3,14	AGAP012404	alanine--glyoxylate aminotransferase 2, mitochondrial	FBgn0037186	n/a		Aminotransferase, Transaminase
3,14	AGAP003814	very-long-chain (3R)-3-hydroxyacyl-(acyl-carrier protein) dehydratase	n/a	n/a	-	Metabolismo de Lipídeos
3,14	AGAP004567	Transmembrane protease serine 11B	n/a	n/a	-	Protease/Proteólise
3,13	AGAP001751	Oxidation resistance protein	FBgn0013576	mustard		Gene responsivo a Ecdisona
3,12	AGAP001659	hexamerin	n/a	n/a		Metabolismo de Aminoácidos, estoque de aminoácidos
3,11	AGAP008570	Steroid dehydrogenase	n/a	n/a	-	Metabolismo de Lipídeos, Reação de Oxirredução
3,11	AGAP007730	not available	FBgn0033538	n/a		Metabolismo de Aminoácidos
3,10	AGAP000243	Rhopilin-2	FBgn0026374	Rhopilin		Proteína G
3,07	AGAP029769	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
3,04	AGAP003251	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
3,03	AGAP000011	branched-chain-amino-acid aminotransferase 2	FBgn0030482	Branched chain amino acid transaminase	-	Metabolismo de aminoácidos
3,03	AGAP009106	not available	n/a	n/a		Desconhecida
3,03	AGAP001377	serine protease inhibitor (serpin) 11	n/a	n/a		Inibidor de Serina Protease
2,99	AGAP000820	cuticular protein RR-2 family 125	FBgn0039805	Cuticular protein 100A	-	Proteína cuticular
2,99	AGAP007674	GIPC PDZ domain containing family, member 2	FBgn0010504	kermit		Relacionado ao Citoesqueleto
2,98	AGAP003141	Insulin-related peptide binding protein	FBgn0001257	Ecdysone-inducible gene L2		Desenvolvimento do sistema nervoso
2,98	AGAP009634	not available	n/a	n/a		Desconhecida
2,98	AGAP029459	not available	FBgn0000721	foraging		Proteína kinase
2,97	AGAP002551	glucose dehydrogenase (acceptor)	n/a	n/a		Reação de Oxirredução
2,97	AGAP002109	solute carrier family 31 (copper transporter), member 1	FBgn0062413	Copper transporter 1A		Transporte transmembranar de cobre
2,96	AGAP028137	cuticular protein (putative) CPLCP17	n/a	n/a	-	Proteína cuticular
2,96	AGAP028156	cuticular protein (putative) CPLCP14	n/a	n/a	-	Proteína cuticular
2,96	AGAP028160	cuticular protein (putative) CPLCP20	n/a	n/a	-	Proteína cuticular
2,96	AGAP028178	cuticular protein (putative) CPLCP13	n/a	n/a	-	Proteína cuticular
2,96	AGAP003578	aldehyde dehydrogenase (NAD+)	n/a	n/a		Oxidação de aldeídos
2,95	AGAP003746	not available	FBgn0034253	n/a	SEA	Desconhecida
2,93	AGAP029607	not available	n/a	n/a		Desconhecida

2,93	AGAP012292	cytochrome P450	n/a	n/a	-	CYP450
2,93	AGAP009104	not available	FBgn0032029	n/a		Metabolismo de Lipídeos
2,92	AGAP010066	not available	n/a	n/a	Mpv17_PMP 22	Desconhecida
2,92	AGAP029419	not available	n/a	n/a	Mpv17_PMP 22	Desconhecida
2,91	AGAP029365	not available	n/a	n/a		Desconhecida
2,91	AGAP012571	not available	n/a	n/a		Desconhecida
2,91	AGAP013060	not available	n/a	n/a		Desconhecida
2,91	AGAP010164	protein withered, carnitine O-palmitoyltransferase	FBgn0261862	withered	-	Palmitoilação, palmitoiltransferase
2,89	AGAP010857	sodium-dependent nutrient amino acid transporter 5	n/a	n/a	-	Transporte transmembranar de Aminoácidos
2,89	AGAP001982	not available	FBgn0038394	Juvenile hormone binding protein 4		Relacionado ao relógio circadiano
2,86	AGAP008615	Canopy homolog 2	FBgn0263260	seele		Associado ao Retículo Endoplasmático
2,86	AGAP002358	ADP,ATP carrier protein 2	FBgn0003360	stress-sensitive B	-	Transporte transmembranar, ADP:ATP antiporter
2,83	AGAP012356	cuticular protein (putative) CPLCP27	n/a	n/a	-	Proteína cuticular
2,83	AGAP028227	not available	n/a	n/a		Desconhecida
2,80	AGAP001346	not available	FBgn0030040	n/a		Desconhecida
2,79	AGAP010769	glucosamine-phosphate N-acetyltransferase	FBgn0039690	Glucosamine-phosphate N-acetyltransferase		Metabolismo de Carboidratos
2,79	AGAP009281	Apolipoprotein D precursor	FBgn0053126	Neural Lazarillo		Metabolismo de Lipídeos, ligação lipídica
2,79	AGAP009609	homogentisate 1,2-dioxygenase	FBgn0040211	homogentisate 1,2-dioxygenase		Metabolismo de Aminoácidos, processo catabólico de L-fenilalanina
2,77	AGAP012008	not available	FBgn0010620	n/a		Relacionado ao citoesqueleto
2,74	AGAP001763	fatty acid transporter protein 1	n/a	n/a		Metabolismo de Lipídeos
2,72	AGAP008822	FK506-binding protein 14	FBgn0010470	FK506-binding protein 14		Via de Notch
2,72	AGAP002279	Fatty acyl-CoA reductase	FBgn0039131	n/a	-	Metabolismo de Lipídeos
2,72	AGAP004631	not available	FBgn0038829	n/a		Sinalização de cálcio
2,68	AGAP009978	DUOX: dual oxidase: peroxidase and NADPH-oxidase domains	FBgn0283531	Dual oxidase		Produção de peróxido de hidrogênio, formação da casca do ovo
2,67	AGAP000327	tyrosine aminotransferase	FBgn0030558	Tyrosine aminotransferase		Metabolismo de Aminoácidos, processo catabólico de L-fenilalanina e

						tirosina
2,67	AGAP002857	not available	FBgn0038198	Niemann-Pick type C-2b		Metabolismo de Lipídeos, transporte de colesterol
2,66	AGAP004451	UDPglucose--hexose-1-phosphate uridylyltransferase	FBgn0263200	Galactose-1-phosphate uridylyltransferase		Metabolismo de Carboidratos
2,64	AGAP008835	CLIP-domain serine protease	n/a	n/a		Protease/Proteólise
2,62	AGAP011118	focal adhesion kinase 1	FBgn0020440	Focal adhesion kinase		Relacionado ao citoesqueleto
2,62	AGAP004815	not available	n/a	n/a		Desconhecida
2,62	AGAP012530	not available	n/a	n/a		Desconhecida
2,62	AGAP012531	not available	n/a	n/a		Desconhecida
2,62	AGAP008321	membrane protein, palmitoylated	n/a	n/a		Palmitoilação
2,61	AGAP006967	not available	n/a	n/a		Desconhecida
2,61	AGAP029666	not available	FBgn0004919	goliath		Ubiquitinação
2,61	AGAP008544	AMP dependent ligase	n/a	n/a		Metabolismo de Lipídeos
2,58	AGAP006521	adenylate cyclase-stimulating G alpha protein	FBgn0010223	G protein $\alpha$ f subunit	-	Proteína G
2,57	AGAP010436	adenylate cyclase 2	FBgn0004852	Adenylyl cyclase 76E	-	Produção de cAMP
2,56	AGAP002270	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
2,52	AGAP009251	not available	n/a	n/a		Desconhecida
2,52	AGAP009252	not available	n/a	n/a		Desconhecida
2,50	AGAP009670	serine protease inhibitor (serpin) 4	n/a	n/a		Inibidor de Serina Protease
2,49	AGAP005091	protein GDAP2 homolog	FBgn0042135	Gdap2		Metabolismo de Lipídeos
2,49	AGAP005153	cysteine dioxygenase	FBgn0034364	n/a	-	Metabolismo de aminoácidos/ metabolismo de cisteína
2,48	AGAP000163	glutathione transferase microsomal 2	n/a	n/a	-	Metabolismo de glutatona
2,45	AGAP009953	guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-13	n/a	n/a		Proteína G
2,44	AGAP004261	cryptochrome 2	n/a	n/a		Relacionado ao relógio circadiano
2,43	AGAP012676	cuticular protein (putative) CPLCP26	n/a	n/a	-	Proteína cuticular
2,43	AGAP028067	not available	n/a	n/a		Desconhecida
2,43	AGAP000278	odorant-binding protein 9	n/a	n/a		Proteína quimiossensorial
2,40	AGAP007921	sterol carrier protein 2	FBgn0015808	Sterol carrier protein X-related thiolase		Metabolismo de Lipídeos
2,40	AGAP004071	Fimbrin	FBgn0024238	Fimbrin		Relacionado ao Citoesqueleto
2,39	AGAP007693	serine protease inhibitor (serpin) 7	n/a	n/a		Inibidor de Serina Protease
2,38	AGAP005234	copper-zinc superoxide dismutase 2	FBgn0033631	Superoxide dismutase 3		Superoxide dismutase, balanço Redox
2,38	AGAP006082	not available	FBgn0033753	Cytochrome P450		CYP450

				301a1		
2,37	AGAP007886	not available	FBgn0032297	n/a		Atividade inibidora de serina/treonina fosfatase
2,37	AGAP006906	Cat eye syndrome critical region protein 1	n/a	n/a		Metabolismo de Purina
2,37	AGAP009333	mitochondrial carrier protein (Rim2)	FBgn0031359	Replication in mitochondria 2		Transporte transmembranar, transporte de desoxinucleotídeos
2,36	AGAP008821	N-myc downstream regulated	FBgn0043070	Misexpression suppressor of KSR 2		Transdução de sinal
2,36	AGAP001021	dihydropyrimidine dehydrogenase (NADP+)	FBgn0086450	suppressor of rudimentary		Metabolismo de Pirimidina, Degradação
2,35	AGAP002685	autophagy related gene	FBgn0052672	Autophagy-related 8a		Relacionado a Autofagia
2,34	AGAP010395	calpain-C	FBgn0260450	Calpain C	-	Sinalização de cálcio, Domínio calpaína
2,33	AGAP005847	hydroxymethylglutaryl-CoA synthase	FBgn0010611	HMG Coenzyme A synthase	-	Metabolismo de acetil-CoA
2,32	AGAP012953	epidermal retinal dehydrogenase	n/a	n/a	adh_short	Metabolismo de Lipídeos
2,31	AGAP007372	not available	n/a	n/a		Desconhecida
2,31	AGAP000519	diacylglycerol kinase (ATP dependent)	FBgn0261549	retinal degeneration A	-	Metabolismo de Lipídeos
2,29	AGAP007589	glucosyl/glucuronosyl transferases	n/a	n/a		Metabolismo de Carboidratos
2,29	AGAP005634	chitinase	FBgn0022702	Chitinase 2		Metabolismo de quitina, desenvolvimento de cutícula
2,26	AGAP011172	cystathionine gamma-lyase	FBgn0000566	Cystathionine gamma-lyase	-	Metabolismo de aminoácidos/ glutationa
2,25	AGAP007104	farnesyl diphosphate synthase	FBgn0025373	Farnesyl pyrophosphate synthase		Migração das células germinativas
2,25	AGAP011431	not available	n/a	n/a		Desconhecida
2,25	AGAP029117	not available	n/a	n/a		Desconhecida
2,25	AGAP029747	not available	n/a	n/a		Desconhecida
2,25	AGAP029750	not available	n/a	n/a		Desconhecida
2,24	AGAP005865	fumarylacetoacetase	n/a	n/a		Metabolismo de tirosina
2,23	AGAP002989	gamma-interferon-inducible lysosomal thiol reductase	FBgn0037721	n/a	-	Thiol Reductase
2,23	AGAP011641	not available	n/a	n/a		Desconhecida
2,22	AGAP007065	transcription factor CP2 and related proteins	FBgn0050011	gemini		fator de transcrição
2,21	AGAP002417	cytochrome P450	n/a	n/a	-	CYP450
2,21	AGAP005982	not available	FBgn0015801	Rhythmically expressed	n/a	Desconhecida

				gene 5		
2,20	AGAP029284	not available	n/a	n/a		Desconhecida
2,19	AGAP001797	Tropomyosin 1	FBgn0003721	Tropomyosin 1		Regulação da contração muscular
2,19	AGAP008360	spaetzle-like cytokine 3	FBgn0031959	spatzle 3		Via de Toll
2,19	AGAP005556	Motile sperm domain-containing protein 2	FBgn0053523	n/a		Metabolismo de Lipídeos
2,19	AGAP008667	Hydroxysteroid (17-beta) dehydrogenase 4	FBgn0030731	peroxisomal Multifunctional enzyme type 2	-	Metabolismo de Lipídeos / degradação de ácidos graxos
2,18	AGAP006921	PID domain-containing protein	FBgn0029092	ced-6		Transdução de sinal, Fagocitose
2,17	AGAP006063	not available	FBgn0015766	Msr-110	n/a	Desconhecida
2,16	AGAP007751	not available	FBgn0016926	Pinocchio		Associado ao Olfato
2,16	AGAP000009	syntaxin-binding protein	FBgn0030412	Tomosyn		Miscelânea
2,16	AGAP009213	serine protease inhibitor (serpin) 16	n/a	n/a		Inibidor de Serina Protease
2,15	AGAP011655	not available	FBgn0001325	Kruppel		Fator de transcrição
2,14	AGAP011810	Retinol dehydrogenase 12 (all-trans/9-cis/11-cis)	FBgn0034500	n/a		Metabolismo de Lipídeos
2,13	AGAP012295	cytochrome P450	n/a	n/a	-	CYP450
2,12	AGAP009210	not available	FBgn0020503	Cytoplasmic linker protein 190		Relacionado ao citoesqueleto
2,12	AGAP007044	potassium channel tetramerisation domain containing 1	FBgn0034636	tiwaz		Fator de transcrição
2,11	AGAP008794	not available	FBgn0053526	Phosphatase 1 nuclear targeting subunit		Regulação da expressão gênica
2,11	AGAP008983	not available	FBgn0259978	vulcan	GKAP	Sinalização celular, putativo
2,11	AGAP011937	not available	FBgn0261341	vermiform		Metabolismo de quitina, desenvolvimento de cutícula
2,10	AGAP008403	not available	n/a	n/a		Desconhecida
2,10	AGAP029606	not available	FBgn0010246	Myosin 61F		Relacionado ao citoesqueleto
2,09	AGAP005716	Class B Scavenger Receptor (CD36 domain)	FBgn0035815	Sensory neuron membrane protein 2		Proteína quimiossensorial
2,09	AGAP007754	not available	n/a	n/a		Desconhecida
2,09	AGAP005242	not available	FBgn0034021	n/a		Desconhecida
2,09	AGAP006901	not available	FBgn0265991	Z band alternatively spliced PDZ-motif protein 52		Desenvolvimento muscular
2,08	AGAP003687	glycerol-3-phosphate O-acyltransferase 1/2	FBgn0027579	minotaur	-	Metabolismo de Lipídeos

2,08	AGAP011972	Epoxide hydrolase	n/a	n/a		Degradação de epóxidos
2,08	AGAP029758	Acyl-coenzyme A oxidase	n/a	n/a	-	Metabolismo de Lipídeos
2,07	AGAP004784	fatty acyl-CoA reductase 2	n/a	n/a	-	Metabolismo de Lipídeos
2,06	AGAP012696	Sulfotransferase (Sult)	n/a	n/a	-	Miscelânea
2,06	AGAP029783	Sulfotransferase (Sult)	n/a	n/a	-	Miscelânea
2,03	AGAP000573	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
2,02	AGAP002457	not available	FBgn0038180	Chitinase 5		Metabolismo de quitina
2,02	AGAP004500	Knickkopf	n/a	n/a		Metabolismo de Quitina
2,01	AGAP003680	ATP-binding cassette transporter (ABC transporter) family H member 2	FBgn0039594	snustorr	-	Transporte Transmembranar, Transportador ABC, desenvolvimento de cutícula
2,01	AGAP011781	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
2,00	CLIP domain-containing serine protease 14D-like, AGAP003249-like					Protease/Proteólise
1,98	AGAP028649	not available	FBgn0086711	moladietz	DuoxA	Miscelânea
1,98	AGAP010369	cuticular protein RR-2 family 112	n/a	n/a	-	Proteína cuticular
1,97	AGAP007564	not available	FBgn0036780	n/a		Desconhecida
1,97	AGAP006015	decaprenyl-diphosphate synthase subunit 2	n/a	n/a		Biossíntese de Isoprenóides, e Ubiquinona
1,94	AGAP029560	not available	n/a	n/a		Desconhecida
1,93	AGAP001203	not available	FBgn0052036	n/a		Ligação de quitina
1,92	AGAP000572	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise, envolvida na ativação da cascata de melanização
1,90	AGAP001313	Muscular protein 20	FBgn0035499	Chd64		Via de sinalização mediada por hormônio juvenil
1,90	AGAP010445	Rab11 family-interacting protein 3/4	FBgn0013718	nuclear fallout		Relacionado ao Citoesqueleto
1,89	AGAP011388	solute carrier family 39 (zinc transporter), member 5	n/a	n/a		Transporte transmembranar de zinco
1,89	AGAP012772	not available	n/a	n/a	Zip	Transporte Transmembranar de zinco ou outros metais, putativo
1,88	AGAP008054	sensory appendage protein 3	n/a	n/a		Proteína quimiossensorial: proteína específica de bulbo ejaculatório 3
1,88	AGAP010280	yellow protein	FBgn0041713	yellow-c		Via de Melanização/Esclerotização
1,88	AGAP013365	not available	n/a	n/a		Desconhecida
1,87	AGAP029610	not available	FBgn0036746	CREB-regulated transcription		Metabolismo energético

				coactivator		
1,85	AGAP010784	dolichyl-phosphate-mannose-protein mannosyltransferase	FBgn0003292	rotated abdomen		Desenvolvimento Muscular, Miscelânea
1,85	AGAP007140	Ser/Thr protein phosphatase/nucleotidase	n/a	n/a	-	Miscelânea
1,85	AGAP013260	chitinase	n/a	n/a		Metabolismo de quitina
1,85	AGAP009651	CD63 antigen	FBgn0032074	Tetraspanin 29Fa		Miscelânea
1,84	AGAP007534	not available	n/a	n/a		Desconhecida
1,82	AGAP005175	acetyl-CoA carboxylase / biotin carboxylase	FBgn0033246	Acetyl-CoA carboxylase	-	Metabolismo de Lipídeos
1,82	AGAP002438	tyrosine-protein phosphatase non-receptor type 11	FBgn0000382	corkscrew		Metabolismo de tirosina
1,82	AGAP001864	cytochrome P450	n/a	n/a	-	CYP450
1,82	AGAP007790	Sodium/potassium-transporting ATPase subunit beta	FBgn0015776	nervana 1	-	Transporte Transmembranar, ATPase Na+ K+
1,82	AGAP009282	Hyphantrin	n/a	n/a	Lipocalin	Metabolismo de Lipídeos
1,82	AGAP010957	Calmodulin	FBgn0000253	Calmodulin	-	Sinalização de cálcio
1,81	AGAP007593	glycerol-3-phosphate dehydrogenase (NAD+)	FBgn0001128	Glycerol-3-phosphate dehydrogenase 1	-	Metabolismo de Lipídeos
1,80	AGAP009616	not available	FBgn0029733	n/a	Neur_chan_L BD, Neur_chan_m emb	Transporte transmembranar de íons
1,80	AGAP005324	UBIQUITIN_CONJUGAT_2 domain-containing protein	FBgn0033544	n/a		Ubiquitinação
1,80	AGAP004450	not available	FBgn0261446	n/a	adh_short	Metabolismo de Lipídeos
1,80	AGAP004383	glutathione S-transferase delta class 10	n/a	n/a	-	Metabolismo de glutatona
1,79	AGAP002552	glucose dehydrogenase (acceptor)	n/a	n/a		Reação de Oxirredução
1,78	AGAP013134	not available	n/a	n/a		Desconhecida
1,77	AGAP008931	inorganic phosphate cotransporter	n/a	n/a		Transporte Transmembranar
1,77	AGAP002425	cysteine sulfinic acid decarboxylase	FBgn0036975	n/a		Metabolismo de Aminoácidos
1,77	AGAP012821	not available	n/a	n/a		Atividade de acyltransferase
1,74	AGAP009510	malate dehydrogenase	FBgn0262782	Malate dehydrogenase 1		Metabolismo de Carboidratos, interconversão de malato e oxaloacetato
1,74	AGAP000410	histone deacetylase 4/5	FBgn0041210	Histone deacetylase 4		Regulação da transcrição
1,73	AGAP010404	glutathione S-transferase sigma class 1	FBgn0010226	n/a	-	Metabolismo de glutatona
1,73	AGAP003408	aldose reductase	n/a	n/a		Aldose redutase

1,71	AGAP008051	sensory appendage protein 1	n/a	n/a		Proteína quimiossensorial: proteína específica de bulbo ejacatório 3
1,71	AGAP001411	BCL2/adenovirus E1B 19 kDa protein-interacting protein 3	FBgn0037007	n/a		Apoptose
1,71	AGAP007327	not available	FBgn0000567	Ecdysone-induced protein 74EF		Fator de transcrição
1,69	AGAP003960	not available	n/a	n/a		Desconhecida
1,69	AGAP006615	FK506-binding protein 8	FBgn0034368	zonda	-	Miscelânea
1,69	AGAP005311	not available	FBgn0008636	homeobrain		Regulação da transcrição
1,69	AGAP003070	not available	FBgn0039688	Kuzbanian-like		Via de Notch
1,69	AGAP002419	cytochrome P450	n/a	n/a	-	CYP450
1,69	AGAP010640	not available	n/a	n/a		Desconhecida
1,68	AGAP028095	not available	n/a	n/a		Desconhecida
1,68	AGAP007920	glucuronosyltransferase	n/a	n/a		Metabolismo de Carboidratos, açúcar
1,67	AGAP000707	not available	FBgn0016754	six-banded		Ligação da cromatina
1,67	AGAP003263	CDC42 small effector protein-like protein	FBgn0044823	Spec2		Relacionado ao Citoesqueleto
1,67	AGAP007651	growth arrest and DNA-damage-inducible protein	n/a	n/a		Via de JNK
1,67	AGAP006007	cuticular protein RR-1 family 28	n/a	n/a	-	Proteína cuticular
1,65	AGAP007384	not available	FBgn0034540	Leucine-rich tendon-specific protein	LRR_8	Miscelânea
1,64	AGAP029378	not available	n/a	n/a		Desconhecida
1,64	AGAP011119	not available	n/a	n/a		Desconhecida
1,62	AGAP001616	not available	n/a	n/a		Desconhecida
1,61	AGAP010714	very-long-chain enoyl-CoA reductase	FBgn0035471	Sc2	-	Metabolismo de Lipídeos
1,61	AGAP007258	aconitate hydratase 1 / homoaconitase	n/a	n/a		Metabolismo de Ferro
1,61	AGAP001889	not available	n/a	n/a		Desconhecida
1,60	AGAP006972	not available	n/a	n/a		Desconhecida
1,60	AGAP006973	not available	n/a	n/a		Desconhecida
1,57	AGAP029707	not available	n/a	n/a	PB1, ZZ, UBA_5	Ubiquitinação
1,57	AGAP011958	not available	n/a	n/a		Desconhecida
1,57	AGAP003738	not available	FBgn0032116	Multicopper oxidase 1		Miscelânea
1,57	AGAP007587	not available	FBgn0020621	Protein kinase N		Proteína Kinase, Relacionado ao Citoesqueleto
1,57	AGAP007209	Tetraspanin	n/a	n/a		Miscelânea

1,56	AGAP013529	not available	n/a	n/a		Desconhecida
1,55	AGAP007077	not available	n/a	n/a		Desconhecida
1,55	AGAP009224	not available	FBgn0033188	Death resistor Adh domain containing target		Apoptose
1,55	AGAP003487	Mesenchymal stem cell protein DSCD75	FBgn0029838	n/a	4HBT_2, HotDog_dom_sf	Miscelânea
1,55	AGAP000440	ATP-binding cassette transporter (ABC transporter) family D member 2	FBgn0031069	ATP binding cassette subfamily D member 3	-	Transporte Transmembranar, Transportador ABC, de ácidos graxos
1,55	AGAP002858	Sodium/potassium-transporting ATPase subunit alpha	FBgn0002921	Na pump $\alpha$ subunit	-	Transporte Transmembranar, ATPase Na <sup>+</sup> K <sup>+</sup>
1,55	AGAP007881	Steroid dehydrogenase	n/a	n/a	adh_short	Metabolismo de Lipídeos
1,55	AGAP000500	NADPH cytochrome P450 reductase	FBgn0015623	Cytochrome P450 reductase	-	CYP450
1,55	AGAP003252	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
1,54	AGAP005350	not available	FBgn0028622	quasimodo		Relacionado ao relógio circadiano
1,54	AGAP000769	not available	FBgn0030850	Stasimon		Desenvolvimento muscular
1,54	AGAP008813	FMS-like tyrosine kinase 1	FBgn0032006	PDGF- and VEGF-receptor related		Proteína kinase, receptor de membrana
1,54	AGAP000726	not available	n/a	n/a		Desconhecida
1,53	AGAP004164	glutathione S-transferase delta class 1	n/a	n/a	-	Metabolismo de glutationa
1,53	AGAP001730	cyclin G2	FBgn0039858	Cyclin G		Miscelânea
1,51	AGAP010733	Vanin-like protein 1	n/a	n/a		Miscelânea
1,50	AGAP009466	ATP-binding cassette transporter (ABC transporter) family G member 8	FBgn0031516	n/a	-	Transporte Transmembranar, Transportador ABC
1,50	AGAP005137	not available	FBgn0052105	LIM homeobox transcription factor 1 alpha		Fator de transcrição
1,49	AGAP007970	autophagy related gene	FBgn0032935	Autophagy-related 18b		Relacionado a Autofagia
1,49	AGAP004859	serine protease snake	n/a	n/a	-	Protease/Proteólise
1,49	AGAP005976	not available	FBgn0035533	Cdc42-interacting protein 4		Proteína G
1,49	AGAP003492	Facilitated glucose transporter, solute carrier family 2	n/a	n/a		Transporte transmembranar de açúcar
1,47	AGAP005888	not available	n/a	n/a		Desconhecida
1,45	AGAP007006	filamin	FBgn0028371	jitterbug		Relacionado ao citoesqueleto
1,45	AGAP013476	not available	FBgn0039804	n/a		Desconhecida

1,45	AGAP029130	not available	FBgn0061492	logjam		Relacionado a oviposição
1,45	AGAP011832	Death-associated protein 1	FBgn0033624	n/a		Apoptose
1,45	AGAP000380	5'-nucleotidase	FBgn0052549	5' nucleotidase B	-	Metabolismo de adenosina
1,44	AGAP002691	arrestin domain containing 2	n/a	n/a		Ubiquitinação
1,43	AGAP010623	X box-binding protein 1	FBgn0021872	X box binding protein-1		Regulação da transcrição
1,43	AGAP012613	not available	n/a	n/a		Desconhecida
1,43	AGAP012740	not available	n/a	n/a		Desconhecida
1,43	AGAP012926	not available	n/a	n/a		Desconhecida
1,42	AGAP007281	odorant-binding protein 63	n/a	n/a		Proteína quimiossensorial
1,42	AGAP012717	vesicle transport protein SEC22	FBgn0260855	Secretory 22		Via de Wnt
1,40	AGAP007121	cytochrome b5 protein	FBgn0264294	Cytochrome b5	-	Metabolismo de compostos endógenos e xenobióticos.
1,40	AGAP007940	Reticulon-like protein	n/a	n/a	Reticulon	Organização do Retículo Endoplasmático
1,39	AGAP009648	2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase	n/a	n/a		Metabolismo de Purina
1,39	AGAP006867	cuticular protein RR-1 family 141	FBgn0036880	Cuticular protein 76Bc	-	Proteína cuticular
1,39	AGAP000222	not available	n/a	n/a		Desconhecida
1,38	AGAP007314	not available	n/a	n/a		Desconhecida
1,38	AGAP010696	Uncharacterized oxidoreductase yqjQ	FBgn0031360	n/a		Reação de Oxirredução
1,37	AGAP003153	VATA, V-Type proton ATPase catalytic subunit A	n/a	n/a	-	V- proton ATPase
1,37	AGAP009197	glutathione S-transferase epsilon class 3	n/a	n/a	-	Metabolismo de glutatona
1,37	AGAP007452	neuronal guanine nucleotide exchange factor	n/a	n/a		Proteína G
1,37	AGAP003968	UMP-CMP kinase	FBgn0028833	Cytidine/uridine monophosphate kinase 1		Metabolismo de Pirimidina
1,37	AGAP004860	protease m1 zinc metalloprotease	n/a	n/a	-	Protease/Proteólise
1,37	AGAP003516	not available	n/a	n/a		Metabolismo de Aminoácidos, de asparagina, de glutamina, de oxaloacetato
1,37	AGAP012801	not available	n/a	n/a		Metabolismo de Aminoácidos, de asparagina, de glutamina, de oxaloacetato
1,37	AGAP012844	Calmodulin	n/a	n/a	-	Sinalização de cálcio

1,37	AGAP028038	not available	n/a	n/a		Desconhecida
1,36	AGAP029495	not available	FBgn0046114	Glutamate-cysteine ligase modifier subunit		Metabolismo de glutationa
1,36	AGAP001189	odorant-binding protein 10	n/a	n/a	-	Proteína quimiossensorial
1,36	AGAP012139	ubiquitin carboxyl-terminal hydrolase 12/46	FBgn0039025	Ubiquitin-specific protease 12/46		Ubiquitinação, regulador negativo, Via de Notch
1,36	AGAP010111	not available	n/a	n/a		Atividade de acyltransferase
1,36	AGAP010118	not available	n/a	n/a		Atividade de acyltransferase
1,36	AGAP010125	not available	n/a	n/a		Atividade de acyltransferase
1,35	AGAP006023	tyrosine 3-monooxygenase	FBgn0005626	pale		Metabolismo de aminoácidos
1,35	AGAP000180	phosphoribosylaminoimidazole carboxylase	FBgn0020513	PAICS bifunctional enzyme	-	Metabolismo de Lipídeos
1,34	AGAP006386	transcription factor AP-1	FBgn0001291	Jun-related antigen		fator de transcrição
1,34	AGAP003627	not available	n/a	n/a	Trypsin	Protease/Proteólise, tripsina
1,34	AGAP000315	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
1,33	AGAP004397	alpha-1,6-mannosyl-glycoprotein beta- 1,2-N-acetylglucosaminyltransferase	n/a	n/a		Metabolismo de Carboidratos
1,33	AGAP000237	cAMP response element-binding protein (CREB)	FBgn0265784	Cyclic-AMP response element binding protein B		fator de transcrição
1,32	AGAP007358	CRAL-TRIO domain-containing protein	n/a	n/a		Metabolismo de Lipídeos
1,31	AGAP008636	Pigeon protein	FBgn0010309	pigeon		Ativação de Gamma-secretase
1,31	AGAP002032	not available	FBgn0265048	crossveinless d		Via de BPM
1,30	AGAP000165	glutathione transferase microsomal 1	n/a	n/a	-	Metabolismo de glutationa
1,29	AGAP004144	not available	n/a	n/a		Desconhecida
1,28	AGAP012406	uncharacterized protein slr0305	FBgn0037185	n/a		Sinalização de cálcio
1,28	AGAP004769	Ras-related protein Rap-1b precursor	n/a	n/a		Proteína G
1,28	AGAP003238	N-myc downstream regulated	FBgn0027608	n/a		Atividade de ligação de caderina e tubulina
1,27	AGAP028049	not available	n/a	n/a		Desconhecida
1,27	AGAP003839	Fem-1 homolog c	n/a	n/a		Ubiquitinação, proteólise
1,27	AGAP002644	phospholipid-translocating ATPase	FBgn0037989	ATPase 8B		Metabolismo de Lipídeos
1,27	AGAP004549	Protein yellow	n/a	n/a		Via de Melanização/Esclerotização
1,25	AGAP003942	not available	n/a	n/a	zf-AD, zf- C2H2	Desconhecida
1,25	AGAP009403	not available	FBgn0031738	n/a		Glucuronidação, transferência de ácido

						glucurônico
1,24	AGAP010131	ornithine decarboxylase antizyme 1	n/a	n/a		Crrescimento e proliferação celular
1,24	AGAP004372	elongation of very long chain fatty acids protein 1	FBgn0051523	n/a	-	Metabolismo de Lipídeos
1,23	AGAP000947	glutathione S-transferase unclassified 1	FBgn0250732	GST-containing FLYWCH zinc-finger protein	-	Metabolismo de glutationa
1,22	AGAP011070	not available	FBgn0261553	n/a	n/a	Desconhecida
1,22	AGAP007982	angiotensin-converting enzyme 8	FBgn0032535	Ance-2		Protease/Proteólise, metaloprotease
1,22	AGAP012181	not available	FBgn0027585	n/a		Desconhecida
1,22	AGAP005400	four and a half LIM domains	n/a	n/a	LIM	Miscelânea
1,21	AGAP005515	Fatty acyl-CoA reductase	FBgn0033464	n/a	-	Metabolismo de Lipídeos
1,21	AGAP009791	neprilysin, neutral endopeptidase 1	FBgn0029843	Neprilysin 1		Clivagem/Inativação de peptídeos
1,21	AGAP011493	UDP-xylose and UDP-N-acetylglucosamine transporter-like	n/a	n/a		Transporte Transmembrar de açúcar
1,20	AGAP005553	cellular retinaldehyde binding protein	FBgn0035636	Cellular retinaldehyde binding protein	-	Metabolismo de Lipídeos
1,20	AGAP002638	ATP-binding cassette transporter (ABC transporter) family H member 1	FBgn0053970	oskyddad	-	Transportador ABC
1,19	AGAP008844	galectin 1	n/a	n/a		Metabolismo de Carboidratos
1,18	AGAP010805	not available	FBgn0050015	n/a		Desconhecida
1,17	AGAP009051	lipopolysaccharide-induced TNF-alpha transcription factor (LITAF-like 1)	n/a	n/a		Atividade de Ligação de íons de zinco
1,16	AGAP000297	not available	FBgn0030309	n/a		Ativação da proliferação de hemócitos
1,15	AGAP003514	not available	n/a	n/a	PhyH	Metabolismo de Lipídeos
1,15	AGAP012746	Phyhd1 protein	n/a	n/a	PhyH	Metabolismo de Lipídeos
1,15	AGAP011792	CLIP-domain serine protease	n/a	n/a	-	Protease/Proteólise
1,15	AGAP008141	argininosuccinate lyase	FBgn0032076	Argininosuccinate lyase	-	Ciclo da Uréia
1,15	AGAP002113	cytochrome b5	FBgn0029854	n/a	-	Metabolismo de Lipídeos, Biossíntese de Esteróis
1,14	AGAP008014	not available	n/a	n/a		Desconhecida
1,14	solute carrier family 15 member-like		n/a	n/a		Transporte Transmembrar
1,14	AGAP001548	Stress-associated endoplasmic reticulum protein 2	n/a	n/a	RAMP4	Associado ao Retículo Endoplasmático
1,13	AGAP028402	not available	n/a	n/a		Desconhecida
1,13	AGAP029442	not available	FBgn0039339	n/a	GTP-bdg_N,	Metabolismo de Ribossomo, putativo

					GTP-bdg_M, MMR_HSR1	
1,12	AGAP000577	AMP deaminase	FBgn0052626	AMP deaminase	-	Metabolismo de Purina
1,12	AGAP001617	Ras-related protein Rab-7A	FBgn0015795	Rab7		Proteína G
1,11	AGAP003784	glucose dehydrogenase (acceptor)	FBgn0030592	n/a		Reação de Oxirredução
1,11	AGAP007340	facilitated glucose transporter (solute carrier family 2)	n/a	n/a		Transporte transmembranar de açúcar
1,11	AGAP005093	not available	n/a	n/a		Desconhecida
1,11	AGAP001520	protein phosphatase 1, regulatory (inhibitor) subunit 3	FBgn0036428	Glycogen binding subunit 70E		Metabolismo de Glicogênio
1,11	AGAP013356	otopetrin	n/a	n/a		Canal de próton
1,11	AGAP006898	chitinase	FBgn0035398	Chitinase 7		Metabolismo de quitina, Desenvolvimento de cutícula
1,10	AGAP008181	protein SMG5	FBgn0019890	Smg5 nonsense mediated mRNA decay factor		Catabolismo do RNAm
1,10	AGAP029650	not available	FBgn0033166	ELL-associated factor		Regulação da transcrição
1,10	AGAP001082	saposin	FBgn0000416	Saposin-related		Metabolismo de Lipídeos, Ativação de esfingolipídeos
1,09	AGAP001206	not available	FBgn0052037	n/a	n/a	Desconhecida
1,09	AGAP008137	Niemann-Pick C1 protein	FBgn0024320	Niemann-Pick type C-1a		Metabolismo de Lipídeos, transporte de colesterol
1,08	AGAP004795	RNA polymerase II elongation factor ELL	FBgn0014037	Suppressor of Triplolethal		Regulação da transcrição, Via de Notch e Ras
1,08	AGAP009755	angiotensin-converting enzyme 5	n/a	n/a		Relacionado ao relógio circadiano
1,08	AGAP009756	angiotensin-converting enzyme 6	n/a	n/a		Relacionado ao relógio circadiano
1,08	AGAP008350	Double Oxidase: Two Peroxidase domains.	FBgn0032685	n/a		Atividade de peroxidase
1,08	AGAP011771	not available	n/a	n/a		Desconhecida
1,08	AGAP002036	zinc finger, FYVE domain containing 1	n/a	n/a		Fator de transcrição
1,08	AGAP008419	Ras-related protein Rab-14	FBgn0015791	Rab14		Proteína G
1,07	AGAP010189	not available	FBgn0020304	drongo		Proteína G
1,07	AGAP009486	V-type H <sup>+</sup> -transporting ATPase 54 kD subunit	FBgn0027779	Vacuolar H <sup>+</sup> -ATPase SFD subunit		V- proton ATPase
1,07	AGAP012270	not available	n/a	n/a		Desconhecida
1,07	AGAP011334	Failed axon connections protein	FBgn0014163	failed axon connections		Formação de axônios
1,07	AGAP002357	vesicle transport protein GOT1B	FBgn0052576	n/a		Transporte de vesícula, complexo de

						Golgi
1,07	AGAP003334	cuticular protein unclassified	n/a	n/a	-	Proteína cuticular
1,06	AGAP001069	not available	FBgn0019925	Surfeit 4	SURF4	Associado ao Retículo Endoplasmático
1,06	AGAP028154	V-type proton ATPase proteolipid subunit	FBgn0262736	Vacuolar H <sup>+</sup> ATPase 16kD subunit 1	-	V- proton ATPase
1,06	AGAP012662	omega-amidase	n/a	n/a		Metabolismo de aminoácidos
1,06	AGAP013231	Nitrilase homolog 2	n/a	n/a		Metabolismo de aminoácidos
1,05	AGAP002898	glutathione S-transferase zeta class 1	FBgn0037697	Glutathione S transferase Z2		Metabolismo de glutatona
1,05	AGAP012291	Cytochrome P450	n/a	n/a	-	CYP450
1,05	AGAP007721	synaptobrevin	n/a	n/a	Synaptobrevin, VAMP	Exocitose
1,05	AGAP029668	not available	FBgn0035087	n/a		Metabolismo de Lipídeos
1,04	AGAP006395	phospholipase b, plb1	n/a	n/a	-	Metabolismo de Lipídeos
1,04	AGAP029559	C-type lectin (CTL) - mannose binding	FBgn0053978	n/a		Sinalização de cálcio
1,04	AGAP008923	not available	n/a	n/a		Desconhecida
1,04	AGAP029611	not available	n/a	n/a		Desconhecida
1,04	AGAP028186	not available	n/a	n/a		Desconhecida
1,04	AGAP005235	not available	FBgn0082585	sprite	n/a	Domínio PDZ
1,03	AGAP009619	S-adenosylmethionine decarboxylase	FBgn0019932	S-adenosylmethionine decarboxylase		Sistema Imune, proteção contra bactéria Gram-negativa
1,03	AGAP000035	not available	FBgn0029688	lava lamp		Relacionado ao Citoesqueleto
1,03	AGAP000546	acyl-CoA-binding protein	n/a	n/a	-	Metabolismo de Lipídeos
1,02	AGAP004532	3-ketoacyl-CoA reductase	FBgn0029975	spidey	-	Metabolismo de Lipídeos
1,02	AGAP002428	not available	FBgn0036433	n/a	n/a	Adesão celular, Septate Junction
1,02	AGAP009522	not available	FBgn0259735	miles to go	n/a	Adesão celular
1,02	AGAP010893	oxysterol-binding protein 1	FBgn0020626	Oxysterol binding protein	-	Metabolismo de lipídeos, Proteína de ligação ao Oxisterol
1,01	AGAP004211	not available	FBgn0027598	CIN85 and CD2AP related		Relacionado ao citoesqueleto
1,01	AGAP003513	not available	n/a	n/a		Desconhecida
1,00	AGAP012097	mitochondrial 2-oxoglutarate/malate transporter (solute carrier family 25)	n/a	n/a		Transporte transmembranar
1,00	AGAP009739	vacuolar protein sorting 52	FBgn0031710	Vacuolar protein sorting 52		Proteína G

## ANEXO VII

Descrição e função dos 318 genes possivelmente expressos na serosa do *T. castaneum*. Diferente de *A. gambiae*, a lista de *T. castaneum* tem AvgLog2fold negativo, visto que são os genes menos expressos quando *Zen1* é silenciado (Gurska *et al.*, 2020). Para a descrição e definição de função foram utilizadas as informações presentes no Ensembl. Quando se fez necessário, outros bancos de dados foram empregados (iBeetleBase, FlyBase - para dados de *D. melanogaster* - e PFAM).

AvLog2 Fold	ID gene	Descrição Ensembl	Ortólogo 1:1 em Drosophila, ID	Ortólogo 1:1 em Drosophila, Nome Completo	Descrição PFAM	Função
-4,815	TC001141	n/a	FBgn0039161	n/a	n/a	Desconhecida
-4,714	TC000107	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-4,504	TC016278	Elongation of very long chain fatty acids protein	n/a	n/a	-	Metabolismo de Lipídeos
-4,379	TC003884	n/a	n/a	n/a	n/a	Desconhecida
-4,236	TC031852	n/a	n/a	n/a	n/a	Desconhecida
-4,235	TC007258	n/a	FBgn0034860	n/a	n/a	Desconhecida
-4,150	TC015555	n/a	n/a	n/a	n/a	Desconhecida
-4,022	TC015108	Acyl-CoA Delta(11) desaturase-like Protein	n/a	n/a	-	Metabolismo de Lipídeos
-4,010	TC000089	n/a	FBgn0024366	n/a	-	Desconhecida
-3,850	TC006727	n/a	n/a	n/a	Domínios PFAM: 5x DUF4758, EGF_CA, SEA	Desconhecida
-3,778	TC033541	n/a	n/a	n/a	n/a	Desconhecida
-3,753	TC013480	Aromatic-L-amino-acid decarboxylase-like Protein	n/a	n/a	-	Via de Melanização/Esclerotização, DDC
-3,564	TC030984	n/a	n/a	n/a	n/a	Desconhecida
-3,469	TC033309	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-3,451	TC013465	Alpha-tocopherol transfer protein-like Protein	n/a	n/a	-	Metabolismo de lipídeos
-3,443	TC012512	ABC transporter G family member 23-like Protein	n/a	n/a	-	Transporte Transmembranar, Transportador ABC, Desenvolvimento de cutícula
-3,352	TC031851	n/a	n/a	n/a	n/a	Desconhecida
-3,352	TC034460	n/a	FBgn0001253	Ecdysone-inducible gene E1	-	Desconhecida
-3,321	TC008400	Adult-specific rigid cuticular protein 12.4-like Protein	FBgn0036619	Cuticular protein 72Ec	-	Proteína Cuticular
-3,256	TC008204	Dopamine N-acetyltransferase-like	FBgn0287831	Arylalkylamine N-	-	Via de Esclerotização e Melanização

		Protein		acetyltransferase 1		
-3,255	TC010590	n/a	FBgn0034530	Reduction in Cnn dots 6	-	Ciclo do Centrossomo
-3,222	TC030609	Serine Proteínase stubble-like Protein	FBgn0033362	n/a	-	Protease/proteólise
-3,173	TC000236	Vitellogenin-3-like Protein	n/a	n/a	-	Metabolismo de Lipídeos, homolog fb
-3,139	TC031518	n/a	FBgn0032136	Apolipoprotein lipid transfer particle	-	Metabolismo de Lipídeos, transporte
-3,072	TC033082	Peripheral plasma membrane protein CASK-like Protein	FBgn0250785	varicose	-	Guanilato Kinase, Desenvolvimento Traqueal
-3,050	TC008024	Icarapin-like	FBgn0043841	virus-induced RNA 1	-	Desconhecida
-3,040	TC000635	Serine protease P16	FBgn0038447	n/a	-	Protease/proteólise
-2,992	TC009877	Cytochrome P450 49A1	n/a	n/a	-	CYP450
-2,975	TC015379	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-2,938	TC011937	Elongation of very long chain fatty acids protein	n/a	n/a	-	Metabolismo de Lipídeos
-2,935	TC010855	Peroxisomal multifunctional enzyme type 2-like Protein	n/a	n/a	-	Metabolismo de lipídeos
-2,911	TC002107	Ovarian-specific serine/threonine-protein kinase Lok-like Protein	FBgn0040056	n/a	-	Proteína kinase
-2,890	TC014708	Nuclear factor of activated T-cells 5-like Protein	FBgn0030505	NFAT nuclear factor	-	Fator de Transcrição
-2,871	TC009442	n/a	FBgn0034563	n/a	n/a	Desconhecida
-2,871	TC032920	n/a	n/a	n/a	n/a	Desconhecida
-2,865	TC002504	Protein enabled-like Protein	FBgn0000578	enabled	-	Polimerização de Actina
-2,853	TC000611	Alpha-tocopherol transfer protein-like	n/a	n/a	-	Metabolismo de lipídeos, Vitamina E
-2,847	TC005503	Vitellogenin-2-like Protein	FBgn0034166	n/a	-	Metabolismo de Lipídeos, homolog fb
-2,838	TC003708	n/a	n/a	n/a	Lectin_C	Ligação a açúcares
-2,803	TC014100	Chitin deacetylase 1	n/a	n/a	-	Metabolismo de quitina
-2,796	TC003286	Carbonic anhydrase 1-like Protein	FBgn0027844	Carbonic anhydrase 1	-	Anidrase carbônica
-2,782	TC014101	n/a	n/a	n/a	CBM_14, Ldl_recept_a	Metabolismo de quitina
-2,767	TC011141	Cuticular protein analogous to peritrophins 3-A2	n/a	n/a	-	Proteína Cuticular
-2,740	TC015188	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-2,698	TC012208	n/a	n/a	n/a	n/a	Desconhecida
-2,661	TC014517	Tubulin polyglutamylase TTLL4-like Protein	FBgn0037838	Tubulin tyrosine ligase-like 15	-	Relacionado ao Citoesqueleto
-2,619	TC013304	n/a	FBgn0035379	spatzle 5	-	Via de Toll

-2,614	TC032736	Putative methylmalonate-semialdehyde dehydrogenase	FBgn0023537	n/a	-	Metabolismo de Lipídeos
-2,597	TC013560	n/a	n/a	n/a	JHBP	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-2,597	TC031198	Transmembrane protease serine 9-like Protein	n/a	n/a	-	Protease/proteólise
-2,592	TC015111	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-2,591	TC033746	Glucose dehydrogenase [FAD, quinone]-like Protein	n/a	n/a	-	Reação de Oxirredução
-2,560	TC005571	NVP-like protein	n/a	n/a	-	Sinalização de Cálcio
-2,545	TC001278	n/a	FBgn0035173	n/a	MFS_1 - Major facilitator superfamily	Transporte Transmembranar
-2,544	TC032290	n/a	FBgn0262508	n/a	n/a	Desconhecida
-2,536	TC032258	Phosphoinositide phospholipase C	n/a	n/a	-	Proteína G
-2,504	TC031511	n/a	FBgn0260386	mind the gap	-	Metabolismo de quitina, formação de traquéia
-2,482	TC001186	Protein notum homolog-like Protein	FBgn0044028	Notum	-	Via de Wnt
-2,471	TC032848	n/a	FBgn0030161	n/a	n/a	Desconhecida
-2,389	TC033106	n/a	FBgn0038894	n/a	JHBP	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-2,384	TC008235	n/a	n/a	n/a	n/a	Desconhecida
-2,372	TC013404	Protein takeout-like Protein	FBgn0038394	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-2,362	TC015564	n/a	FBgn0031646	snustorr snarlik	-	Desenvolvimento de Cutícula, impermeabilização
-2,334	TC011140	Cuticular protein analogous to peritrophins 3-A1	n/a	n/a	-	Proteína Cuticular
-2,316	TC033158	Zinc finger protein 395-like Protein	FBgn0267336	Glucose transporter 4 enhancer factor	-	Fator de Transcrição
-2,297	TC013484	n/a	FBgn0051100	n/a	3x Sugar_tr	Transporte Transmembranar de Açúcar
-2,291	TC010864	n/a	FBgn0031538	n/a	JHBP, Grp7_allergen	Metabolismo de lipídeos
-2,289	TC003203	B(0,+)-type amino acid transporter 1-like Protein	n/a	n/a	-	Transporte Transmembranar de Aminoácido
-2,281	TC000921	Zen	n/a	n/a	-	Fator de Transcrição, Define a serosa
-2,280	TC009377	n/a	FBgn0083951	n/a	n/a	Desconhecida

-2,275	TC004948	Peroxiredoxin 1-like Protein	n/a	n/a	-	Catabolismo de Peróxido de Hidrogênio
-2,218	TC034275	Elongation of very long chain fatty acids protein	FBgn0053110	n/a	-	Metabolismo de Lipídeos
-2,169	TC005104	n/a	FBgn0034501	n/a	n/a	Desconhecida
-2,166	TC006846	n/a	FBgn0051973	Chitin deacetylase-like 5	-	Metabolismo de quitina
-2,165	TC033522	Rhophilin-2-like Protein	FBgn0026374	Rhophilin	-	Proteína G
-2,163	TC004438	Toll-like protein	n/a	n/a	-	Via de Toll
-2,144	TC006771	n/a	n/a	n/a	SH2	Transdução de sinal
-2,128	TC012495	Cytochrome P450-like protein	n/a	n/a	-	CYP450
-2,110	TC008856	n/a	FBgn0264753	Rad, Gem/Kir family member 1	-	Proteína G
-2,109	TC012183	n/a	n/a	n/a	n/a	Desconhecida
-2,104	TC011283	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-2,101	TC015592	n/a	n/a	n/a	n/a	Desconhecida
-2,092	TC000520	Spaetzle	FBgn0003495	spatzle	-	Via de Toll
-2,049	TC034998	n/a	FBgn0040056	n/a	n/a	Proteína kinase
-2,048	TC033081	n/a	n/a	n/a	n/a	Desconhecida
-2,041	TC006124	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-2,019	TC001994	n/a	n/a	n/a	C1_1 - C1 domain protein	Ligação de Diacilglicerol (DAG)
-1,993	TC015392	Alpha-tocopherol transfer protein-like	n/a	n/a	-	Metabolismo de Lipídeos
-1,993	TC015598	n/a	FBgn0029898	n/a	MFS_1	Transporte Transmembranar
-1,983	TC033893	n/a	n/a	n/a	zf-CCHC	Desconhecida
-1,973	TC033520	G-protein coupled receptor Mth-like Protein	FBgn0034219	methuselah-like 4	-	Proteína G
-1,951	TC008048	n/a	FBgn0265714	n/a	n/a	Desconhecida
-1,949	TC015721	Glucose dehydrogenase [FAD, quinone]-like Protein	n/a	n/a	-	Reação de Oxirredução
-1,930	TC009560	n/a	FBgn0003053	pebbled/Hindsight	-	Fator de Transcrição
-1,924	TC012970	n/a	n/a	n/a	DUF4371, Dimer_Tnp_hAT	Desconhecida
-1,919	TC010817	n/a	FBgn0035094	n/a	n/a	Desconhecida
-1,911	TC032113	n/a	n/a	n/a	4x Alpha-amylase, 4x Alpha-amylase_C	Metabolismo de Carboidratos, Degradação

-1,908	TC033778	Transporter	n/a	n/a	SNF	Transporte Transmembranar de Neurotransmissor
-1,906	TC033053	Anion exchange protein	FBgn0259111	Na <sup>+</sup> -driven anion exchanger 1	-	Transporte transmembranar, Sódio e Bicarbonato
-1,901	TC001860	n/a	n/a	n/a	n/a	Desconhecida
-1,897	TC001129	Cadherin 23	FBgn0038247	Cadherin 88C	-	Adesão celular
-1,872	TC033203	n/a	n/a	n/a	n/a	Desconhecida
-1,862	TC012027	WW domain-containing oxidoreductase-like Protein	n/a	n/a	-	Metabolismo de lipídeos
-1,848	TC033244	Putative G-protein coupled receptor Mth-like 3	n/a	n/a	-	Proteína G
-1,830	TC034705	Monocarboxylate transporter 5-like protein	FBgn0023549	Monocarboxylate transporter 1	-	Transporte Transmembranar
-1,822	TC000922	Zen2	n/a	n/a	-	Fator de Transcrição, Função na serosa
-1,791	TC015380	Fatty acyl-CoA reductase	n/a	n/a	-	Metabolismo de Lipídeos
-1,779	TC016330	n/a	n/a	n/a	n/a	Desconhecida
-1,777	TC032304	Protein atonal-like Protein	FBgn0024249	cousin of atonal	-	Fator de Transcrição
-1,777	TC005998	n/a	n/a	n/a	n/a	Desconhecida
-1,772	TC011099	n/a	FBgn0034253	n/a	SEA	Desconhecida
-1,769	TC011101	Cuticular protein analogous to peritrophins 1-J	n/a	n/a	-	Proteína Cuticular
-1,764	TC003707	n/a	n/a	n/a	zf-H2C2_2, ichor - Zinc finger C2H2 transcription factors,	Fator de Transcrição
-1,764	TC008247	n/a	n/a	n/a	CBM39	Ligação a açúcares
-1,761	TC014831	n/a	n/a	n/a	n/a	Desconhecida
-1,744	TC034834	Trifunctional purine biosynthetic protein adenosine-3	FBgn0000053	GART trifunctional enzyme	-	Biossíntese de Purina
-1,743	TC014285	Protein PIH1D3-like Protein	FBgn0036437	n/a	-	Relacionado ao Citoesqueleto, Dynein
-1,735	TC016280	Elongation of very long chain fatty acids protein	n/a	n/a	-	Metabolismo de Lipídeos
-1,731	TC009092	Serine protease P93	n/a	n/a	-	Protease/proteólise
-1,729	TC009525	n/a	n/a	n/a	TPH	Relacionado ao Citoesqueleto
-1,725	TC006832	n/a	n/a	n/a	2x Ank_2	Organização da membrana plasmática
-1,724	TC015554	Dehydrogenase/reductase SDR family protein 7-like	n/a	n/a	-	Metabolismo de Lipídeos
-1,721	TC000431	Calponin	FBgn0038774	n/a	-	Sinalização de Cálcio

-1,720	TC007083	n/a	n/a	n/a	n/a	Proteína G
-1,706	TC010675	n/a	FBgn0262717	Skeletor	-	Ciclo Celular
-1,693	TC030104	Gustatory receptor	FBgn0035468	Gustatory receptor 63a	-	Proteína quimiossensorial
-1,692	TC014688	n/a	n/a	n/a	n/a	Desconhecida
-1,690	TC005249	Phosphoribosylformylglycinamide synthase-like Protein	FBgn0000052	Phosphoribosylformylglycinamide synthase	-	Biossíntese de purina
-1,688	TC031823	n/a	n/a	n/a	n/a	Proteína quimiossensorial
-1,687	TC034274	Elongation of very long chain fatty acids protein	n/a	n/a	-	Metabolismo de Lipídeos
-1,680	TC015717	Glucose dehydrogenase	n/a	n/a	-	Reação de Oxirredução
-1,677	TC014143	Protein kinase DC2-like Protein	n/a	n/a	-	Proteína Kinase
-1,675	TC006284	Gamma-glutamyltranspeptidase 1-like Protein	n/a	n/a	-	Metabolismo de Glutathione
-1,671	TC012346	Dorsocross	n/a	n/a	-	Fator de Transcrição
-1,662	TC013579	n/a	FBgn0033483	eiger	-	Via de JNK
-1,660	TC013464	Protein real-time-like Protein	n/a	n/a	-	Metabolismo de lipídeos
-1,659	TC032847	n/a	n/a	n/a	n/a	Desconhecida
-1,648	TC010001	n/a	FBgn0034797	nahoda	-	Via de Esclerotização
-1,648	TC011529	Gram-negative bacteria binding protein 2	n/a	n/a	-	Via de Toll
-1,637	TC031037	n/a	n/a	n/a	n/a	Desconhecida
-1,631	TC010516	UNC93-like protein	n/a	n/a	-	Transporte transmembranar
-1,631	TC033738	n/a	n/a	n/a	2x fn3, 2x Y_phosphatase	Tirosina Fosfatase
-1,628	TC000827	Held out wings	FBgn0264491	held out wings	-	Regulador da expressão gênica, Via de BMP
-1,621	TC031827	n/a	FBgn0261260	Megalin	-	Regulação de Endocitose, Desenvolvimento de Cutícula
-1,620	TC032087	n/a	n/a	n/a	n/a	Desconhecida
-1,616	TC007584	Receptor protein-tyrosine kinase	FBgn0032006	PDGF- and VEGF-receptor related	-	Proteína kinase, receptor de membrana
-1,614	TC006726	Spatzle 4	FBgn0032362	spatzle 4	-	Via de Toll
-1,613	TC013646	Transporter	n/a	n/a	-	Transporte Transmembranar de Neurotransmissor
-1,594	TC002616	Proton-coupled amino acid transporter 4-like Protein	n/a	n/a	-	Transporte Transmembranar de Aminoácido
-1,592	TC013544	MFS-type transporter SLC18B1-like	n/a	n/a	-	Transporte Transmembranar

		Protein				
-1,580	TC004115	Gamma-glutamylcyclotransferase	FBgn0039109	n/a	-	Metabolismo de Glutaciona, Via de Notch
-1,579	TC005493	n/a	FBgn0032685	n/a	An_peroxidase	Atividade de Peroxidase
-1,575	TC010593	C-1-tetrahydrofolate synthase, cytoplasmic-like Protein	FBgn0020385	pugilist	-	Via de Pigmentação
-1,569	TC015481	Putative chitinase 3-like Protein	FBgn0035398	Chitinase 7	-	Metabolismo de quitina
-1,568	TC015517	Protein takeout-like Protein	FBgn0039203	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-1,565	TC005250	n/a	FBgn0000052	Phosphoribosylformylglucosylaminidase	-	Biossíntese de purina
-1,561	TC033769	n/a	FBgn0086909	n/a	2x APH	Metabolismo de Lisina, Degradação
-1,561	TC015712	Glucose dehydrogenase [FAD, quinone]-like Protein	n/a	n/a	-	Reação de Oxirredução
-1,558	TC033739	n/a	n/a	n/a	2x fn3, 4x Y_phosphatase	Tirosina Fosfatase
-1,551	TC007632	UDP-glucuronosyltransferase 2B14-like Protein	FBgn0027070	UDP-glycosyltransferase family 36 member E1	-	Glucuronidação, transferência de ácido glucurônico
-1,550	TC033997	n/a	FBgn0050069	n/a	n/a	Relacionado ao Citoesqueleto
-1,547	TC008047	n/a	n/a	n/a	n/a	Desconhecida
-1,546	TC006978	Brevican core protein-like Protein	FBgn0035199	triforce	-	Ligação a açúcares
-1,540	TC009269	n/a	n/a	n/a	n/a	Desconhecida
-1,540	TC014397	Circadian clock-controlled protein-like Protein	n/a	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-1,535	TC032476	n/a	n/a	n/a	Myb_DNA-bind_5	Desconhecida
-1,524	TC009549	Zinc finger MYND domain-containing protein 10 homolog-like Protein	FBgn0266709	Zinc finger MYND-type containing 10	-	Relacionado ao Citoesqueleto
-1,521	TC015236	Sodium-independent sulfate anion transporter-like Protein	n/a	n/a	-	Transporte Transmembranar, sulfato
-1,517	TC007044	Gremlin 1	n/a	n/a	-	Via de BMP, antagonista
-1,516	TC034300	Alanine--glyoxylate aminotransferase 2-like	n/a	n/a	-	Aminotransferase, Transaminase
-1,512	TC001036	Fizzy-related protein homolog-like Protein	n/a	n/a	Anaphase-promoting complex subunit 4 WD40 domain	Protease/proteólise

-1,511	TC031150	n/a	FBgn0002440	lethal (3) malignant blood neoplasm	-	Proteína Cuticular
-1,507	TC033185	n/a	n/a	n/a	n/a	Desconhecida
-1,493	TC000174	Gamma-interferon-inducible lysosomal thiol reductase-like Protein	n/a	n/a	-	Thiol reductase
-1,491	TC006631	Putative inorganic phosphate cotransporter-like Protein	n/a	n/a	-	Transporte Transmembranar
-1,490	TC000428	n/a	n/a	n/a	-	Desconhecida
-1,488	TC009178	Putative chitinase 3-like Protein	n/a	n/a	-	Metabolismo de quitina
-1,488	TC033029	Protein tramtrack, alpha isoform-like Protein	n/a	n/a	-	Fator de Transcrição
-1,481	TC007583	Synaptotagmin-6-like Protein	FBgn0261090	Synaptotagmin $\beta$	-	Transporte transmembranar
-1,478	TC015161	n/a	n/a	n/a	adh_short	Metabolismo de Lipídeos
-1,477	TC031361	Tropomodulin-like Protein	FBgn0082582	tropomodulin	-	Via de Notch
-1,476	TC010825	n/a	FBgn0033668	expansion	MH2	Regulação de Receptor Tirosina kinase
-1,473	TC034444	ABC transporter G family member 23-like Protein	n/a	n/a	-	Transporte Transmembranar, Transportador ABC, Desenvolvimento de cutícula
-1,454	TC032514	n/a	n/a	n/a	n/a	Desconhecida
-1,451	TC015715	Glucose dehydrogenase [FAD, quinone]-like Protein	n/a	n/a	-	Reação de Oxirredução
-1,441	TC005292	n/a	n/a	n/a	Endonuclease_7	Endonuclease
-1,439	TC007806	n/a	n/a	n/a	n/a	Desconhecida
-1,427	TC006723	n/a	n/a	n/a	FOG_N, related with G-protein signalling	Proteína G
-1,427	TC014634	n/a	FBgn0001311	krotzkopf verkehrt	-	Chitina Sintase 1
-1,427	TC010525	n/a	FBgn0262169	magu	-	Regulador da Via de BMP
-1,425	TC013415	Serine protease P141	n/a	n/a	-	Protease/proteólise
-1,420	TC034730	Protein takeout-like protein	n/a	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-1,416	TC010445	n/a	FBgn0264502	n/a	n/a	Relacionado ao Citoesqueleto
-1,413	TC010099	n/a	FBgn0002306	stranded at second	-	Diferenciação Epitelial
-1,412	TC033623	n/a	n/a	n/a	n/a	Desconhecida
-1,411	TC032708	n/a	FBgn0263873	sickie	n/a	Relacionado ao Citoesqueleto
-1,410	TC014345	n/a	FBgn0024989	n/a	n/a	Desconhecida

-1,404	TC031386	cyclic AMP-responsive element-binding protein 3-like protein 2	FBgn0004396	Cyclic-AMP response element binding protein A	-	Fator de Transcrição
-1,403	TC012124	Cullin homolog 1-like Protein	n/a	n/a	-	Via de Protease/proteólise
-1,400	TC013559	Protein takeout-like Protein	n/a	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-1,398	TC002459	Three prime repair exonuclease 2-like Protein	FBgn0031484	n/a	-	Nuclease
-1,394	TC032788	Synaptotagmin 1-like Protein	FBgn0261085	Synaptotagmin 12	-	Sinalização de Cálcio, Excitose
-1,380	TC000856	Cytochrome P450 6B51	n/a	n/a	-	CYP450
-1,379	TC030923	Putative RNA-directed DNA polymerase from transposon BS-like Protein	n/a	n/a	-	Transcriptase reversa
-1,374	TC006229	Protein yellow-like Protein	n/a	n/a	-	Via de Melanização/Esclerotização
-1,374	TC034461	n/a	n/a	n/a	DUF229	Desconhecida
-1,370	TC000067	Glutathione S-transferase S1-like Protein	FBgn0010226	Glutathione S transferase S1	-	Metabolismo de Glutathione
-1,361	TC032543	n/a	FBgn0037941	n/a	TSP_1, Laminin_G_2	Miscelânea
-1,359	TC003982	n/a	n/a	n/a	n/a	Desconhecida
-1,352	TC011863	Tetraspanin	n/a	n/a	-	Miscelânea
-1,347	TC011791	Annexin	n/a	n/a	-	Sinalização de Cálcio
-1,343	TC008472	Palmitoyltransferase	FBgn0029137	Patsas	-	Palmitoilação
-1,340	TC007661	Myoinhibiting peptide	FBgn0036713	Myoinhibiting peptide precursor	-	Relacionado ao relógio circadiano
-1,336	TC010815	n/a	n/a	n/a	n/a	Desconhecida
-1,333	TC012828	n/a	n/a	n/a	n/a	Desconhecida
-1,320	TC031338	Protein sickie-like Protein	FBgn0263873	Sickie	-	Relacionado ao Citoesqueleto
-1,319	TC003355	Orthodenticle-2	n/a	n/a	-	Fator de Transcrição
-1,317	TC034364	n/a	FBgn0029836	n/a	n/a	Desconhecida
-1,297	TC000753	Calexcitin-1-like Protein	FBgn0020907	Sarcoplasmic calcium-binding protein 2	-	Sinalização de Cálcio, Ligação de GTP
-1,293	TC032992	Protein takeout-like Protein	n/a	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-1,286	TC013648	Glucose-6-phosphate 1-dehydrogenase	FBgn0004057	Zwischenferment	-	Metabolismo de Carboidratos, Via das Pentoses Fosfato
-1,281	TC014177	Glutamate decarboxylase-like Protein	FBgn0036975	n/a	-	Metabolismo de Aminoácidos

-1,280	TC002083	Cytosolic purine 5'-nucleotidase-like Protein	FBgn0052549	n/a	-	Nucleotidase
-1,276	TC031718	n/a	n/a	n/a	Acyl_transf_3, SBP_bac_3	Miscelânea
-1,273	TC013403	Protein takeout-like Protein	FBgn0038395	n/a	-	Relacionado ao relógio circadiano, Ligação de Hormônio Juvenil
-1,267	TC034862	n/a	n/a	n/a	n/a	Desconhecida
-1,265	TC012674	Crossveinless 2	FBgn0000395	crossveinless 2	-	Via de BMP
-1,258	TC010654	Putative G-protein coupled receptor Mth-like 1	n/a	n/a	-	Proteína G
-1,251	TC032983	n/a	n/a	n/a	n/a	Desconhecida
-1,246	TC032844	Sensory neuron membrane protein 1-like Protein	FBgn0260004	Sensory neuron membrane protein 1	-	Proteína quimiossensorial
-1,244	TC004170	n/a	n/a	n/a	RVT_1, Reverse Transcriptase	Transcritpase Reversa
-1,235	TC013296	n/a	n/a	n/a	n/a	Adesão celular
-1,234	TC010653	Knickkopf	FBgn0001321	knickkopf	-	Metabolismo de quitina
-1,232	TC034861	Multidrug resistance protein homolog 65-like protein	FBgn0031069	Peroxisomal Membrane Protein 70 kDa	-	Transporte Transmembranar, Transportador ABC
-1,228	TC034858	Cytochrome P450 18a1-like protein	FBgn0010383	Cytochrome P450-18a1	-	CYP450
-1,224	TC030992	n/a	n/a	n/a	n/a	Desconhecida
-1,221	TC015671	Actin-related protein 2/3 complex subunit 4	n/a	n/a	-	Relacionado ao Citoesqueleto
-1,220	TC014954	Protein croquemort-like Protein	FBgn0010435	epithelial membrane protein	-	Miscelânea
-1,220	TC014245	Acyl-coenzyme A oxidase	n/a	n/a	-	Metabolismo de lipídeos, Degradação
-1,219	TC002790	Peptidoglycan-recognition protein LE-like Protein	n/a	n/a	N-acetylmuramoyl-L-alanine amidase	Amidase
-1,215	TC034950	Low-density lipoprotein receptor-related protein 2-like protein	n/a	n/a	-	Desconhecida
-1,212	TC010023	n/a	n/a	n/a	DUF745, IFT57	Desconhecida
-1,211	TC031062	n/a	FBgn0262593	Shaker cognate b	-	Canal de Potássio, dependente de voltagem
-1,209	TC033228	n/a	n/a	n/a	2x Chitin_bind_4	Proteína Cuticular
-1,202	TC032874	n/a	n/a	n/a	2x Abhydro_lipase	Metabolismo de lipídeos, Degradação
-1,194	TC009894	Cuticular protein analogous to peritrophins 1-H	FBgn0035844	n/a	-	Proteína Cuticular

-1,186	TC001176	n/a	n/a	n/a	n/a	Desconhecida
-1,179	TC031591	n/a	n/a	n/a	n/a	Desconhecida
-1,177	TC031239	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1-like Protein	FBgn0027621	6-phosphofructo-2-kinase	-	Metabolismo de Carboidratos, regulador da glicólise
-1,177	TC033624	n/a	FBgn0053207	Pxb	-	Miscelânea
-1,170	TC002315	n/a	FBgn0001138	Grain	GATA zinc finger	Fator de Transcrição
-1,168	TC000411	Cytochrome P450 304E1	FBgn0038095	Cyp304a1	-	CYP450
-1,168	TC033874	n/a	n/a	n/a	n/a	Desconhecida
-1,165	TC015720	n/a	FBgn0033942	Cuticular protein 51A	-	Proteína Cuticular
-1,164	TC000542	Phosphoserine aminotransferase	FBgn0014427	n/a	-	Metabolismo de Serina
-1,160	TC003085	Chemosensory protein 2	n/a	n/a	-	Miscelânea
-1,159	TC001120	n/a	n/a	n/a	Chitin_bind_4	Metabolismo de quitina
-1,155	TC013191	DnaJ protein homolog 1-like Protein	n/a	n/a	-	Chaperona
-1,147	TC033330	n/a	n/a	n/a	n/a	Desconhecida
-1,146	TC008005	n/a	n/a	n/a	n/a	Desconhecida
-1,145	TC014943	n/a	n/a	n/a	n/a	Desconhecida
-1,144	TC011639	Putative G-protein coupled receptor No9-like Protein	FBgn0024944	Octopamine receptor in mushroom bodies	-	Proteína G, Receptor de Octopamina
-1,134	TC033543	n/a	n/a	n/a	n/a	Desconhecida
-1,134	TC004635	Serine protease P53	n/a	n/a	-	Protease/proteólise
-1,132	TC014775	Putative multidrug resistance-associated protein lethal(2)03659-like Protein	n/a	n/a	-	Transporte Transmembranar, Transportador ABC
-1,132	TC031922	Low-density lipoprotein receptor-related protein 2-like Protein	n/a	n/a	-	Regulação de Endocitose, Desenvolvimento de Cutícula
-1,130	TC031666	n/a	FBgn0263873	Sickie	-	Relacionado ao Citoesqueleto
-1,130	TC002609	Mitogen-activated protein kinase kinase kinase	FBgn0263395	happyhour	-	Proteína kinase
-1,127	TC012819	n/a	FBgn0037427	Osiris 17	-	Metabolismo de quitina
-1,126	TC000957	WD repeat and FYVE domain-containing protein 3-like Protein	FBgn0043362	blue cheese	-	Miscelânea
-1,124	TC005908	Serine protease H64	n/a	n/a	-	Protease/proteólise
-1,116	TC014942	Fasciclin-3-like Protein	FBgn0000636	Fasciclin 3	-	Adesão celular
-1,113	TC034269	n/a	n/a	n/a	n/a	Desconhecida
-1,109	TC012734	Putative chitinase 3-like Protein	FBgn0250907	Chitinase 10	-	Metabolismo de quitina
-1,107	TC007091	n/a	n/a	n/a	n/a	Desconhecida
-1,104	TC007861	n/a	FBgn0031571	bark beetle	-	Adesão celular

-1,100	TC014638	n/a	FBgn0083978	n/a	n/a	Desconhecida
-1,099	TC007679	n/a	FBgn0030716	n/a	WW	Desconhecida
-1,098	TC014329	Alpha-tocopherol transfer protein-like	n/a	n/a	-	Metabolismo de lipídeos
-1,087	TC034813	UDP-glucuronosyltransferase 2B19-like protein	n/a	n/a	-	Glucuronidação, transferência de ácido glucurônico
-1,087	TC012787	Heparan sulfate 2-O-sulfotransferase pipe-like Protein	n/a	n/a	-	Via de Toll
-1,084	TC009191	Protein transport protein Sec23A-like Protein	FBgn0262125	Secretory 23	-	Diferenciação Epitelial
-1,083	TC033916	n/a	n/a	n/a	n/a	Desconhecida
-1,080	TC005048	n/a	n/a	n/a	n/a	Desconhecida
-1,077	TC031358	Tachykinin-like peptides receptor 99D	FBgn0035610	Leucokinin receptor	-	Proteína G
-1,075	TC013311	Choline-phosphate cytidyltransferase B-like Protein	n/a	n/a	-	Metabolismo de lipídeos, Síntese
-1,075	TC030754	n/a	FBgn0001083	furrowed	-	Adesão celular
-1,075	TC015158	Dehydrogenase/reductase SDR family protein 7-like	n/a	n/a	-	Metabolismo de Lipídeos
-1,065	TC014225	Transcription factor E3-like Protein	FBgn0263112	Mitf	-	Fator de Transcrição
-1,065	TC034363	Cytochrome P450 4c3-like Protein	n/a	n/a	-	CYP450
-1,064	TC011889	FGGY carbohydrate kinase domain-containing protein-like Protein	FBgn0035484	n/a	-	Metabolismo de Carboidratos
-1,063	TC003668	Ras-like protein 3	FBgn0050158	n/a	-	Proteína G
-1,061	TC003371	Lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog-like Protein	n/a	n/a	-	Miscelânea
-1,059	TC014346	Cytochrome P450 346A1	n/a	n/a	-	CYP450
-1,052	TC032831	n/a	n/a	n/a	n/a	Desconhecida
-1,052	TC006033	Serine protease P68	n/a	n/a	-	Protease/proteólise
-1,050	TC008101	n/a	FBgn0030060	n/a	EcKinase	Via de Ecdisona
-1,049	TC034792	n/a	n/a	n/a	n/a	Desconhecida
-1,040	TC032641	n/a	n/a	n/a	n/a	Desconhecida
-1,039	TC014534	Chemosensory protein 12	n/a	n/a	-	Proteína quimiosensorial
-1,038	TC002586	n/a	n/a	n/a	n/a	Desconhecida
-1,037	TC001320	Protein inscuteable homolog-like Protein	FBgn0011674	inscuteable	-	Divisão celular
-1,033	TC000803	Carboxylic ester hydrolase	n/a	n/a	-	Metabolismo de Lipídeos
-1,031	TC031487	n/a	n/a	n/a	n/a	Desconhecida

-1,030	TC013974	n/a	FBgn0263973	Javelin	-	Relacionado ao Citoesqueleto
-1,027	TC032151	n/a	FBgn0003495	Spatzle	-	Via de Toll
-1,024	TC002669	n/a	FBgn0004511	dusky		Regulação da epiderme
-1,015	TC032053	n/a	n/a	n/a	2x Kazal_1	Inibidor de Protease
-1,011	TC001319	n/a	n/a	n/a	Rho_N - transcription termination	Transcrição
-1,010	TC013128	Pupal cuticle protein-like Protein	n/a	n/a	-	Proteína Cuticular
-1,010	TC034273	n/a	n/a	n/a	n/a	Desconhecida
-1,009	TC008169	n/a	n/a	n/a	n/a	Sinalização de Cálcio
-1,009	TC012929	n/a	n/a	n/a	DUF4371	Desconhecida
-1,009	TC008734	n/a	FBgn0039527	n/a	Domínios: 4x Antistasin, Kunitz_BPTI, 3x Lustrin_cystein, 6x Thyroglobulin_1, 2x WAP	Miscelânea
-1,007	TC003327	1-acylglycerophosphocholine O- acyltransferase 1-like Protein	FBgn0052699	Lysophosphatidylcholin e acyltransferase	-	Metabolismo de Lipídeos
-1,005	TC030730	Calpain-C-like Protein	FBgn0260450	Calpain C	-	Sinalização de Cálcio
-1,002	TC007252	MD-2-related lipid-recognition protein-like Protein	n/a	n/a	-	Via de Toll, Desenvolvimento do tegumento
-1,001	TC008454	Protein white-like Protein	FBgn0032167	n/a	-	Transporte transmembranar

