HIGO NASSER SANT'ANNA MOREIRA

ANOTAÇÃO E MONTAGEM DE TRANSCRIPTOMAS DE INTESTINO MÉDIO E OVÁRIOS DO CARRAPATO *AMBLYOMMA SCULPTUM*, ANTES E APÓS A INFECÇÃO POR *RICKETTSIA AMBLYOMMII*

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2016

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
M838a
2016

Moreira, Higo Nasser Santanna, 1987-
Anotação e montagem de transcriptomas de intestino médio e ovários do carrapato *Amblyomma Sculptum*, antes e após a infecção por *Rickettsia Amblyommii* / Higo Nasser Santanna Moreira. – Viçosa, MG, 2016.
xviii, 156f. : il. (algumas color.) ; 29 cm.

Inclui anexos.
Orientador: Cláudio Lisias Mafra de Siqueira.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Carrapato - genes. 2. Expressão gênica - avaliação. 3. *Amblyomma Sculptum. 4. Rickettsia Amblyommii*.
I. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de Pós-graduação em Bioquímica Aplicada. II. Título.

CDD 22. ed. 595.42

HIGO NASSER SANT'ANNA MOREIRA

ANOTAÇÃO E MONTAGEM DE TRANSCRIPTOMAS DE INTESTINO MÉDIO E OVÁRIOS DO CARRAPATO *AMBLYOMMA SCULPTUM*, ANTES E APÓS A INFECÇÃO POR *RICKETTSIA AMBLYOMMII*

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Doctor Scientiae*.

APROVADA: 29 de Fevereiro de 2016.

Otávio J. B. Brustollini

Jorge Abdalla Dergam

Donald H. Bouyer

Gustavo Ferreira Martins

Cláudio Lísias Mafra de Siqueira
(Orientador)

“Grandes e maravilhosas são as tuas obras, Senhor Deus Todo-Poderoso! Justos e verdadeiros são os Teus Caminhos, ó Rei das Nações.

Quem Te não temerá, ó Senhor, e não magnificará O Teu Nome? Porque só Tu és Santo; por isso todas as nações virão, e se prostrarão diante de Ti, porque os Teus Juízos são manifestos.”

Apocalipse 15:3,4

“A Glória do Senhor é ocultar certas coisas. Tentar descobrí-las é a motivação dos homens.”

Provérbios 25:2

“E no princípio criou Deus os céus e a terra. E a terra era sem forma e vazia; e havia trevas sobre a face do abismo; e O Espírito de Deus se movia sobre a face das águas...
...E disse Deus: Produza a terra alma vivente conforme a sua espécie; gado e répteis, e bestas-feras da terra conforme a sua espécie. E assim foi.
E fez Deus as bestas-feras da terra conforme a sua espécie, e o gado conforme a sua espécie, e todo réptil da terra conforme a sua espécie. E viu Deus que tudo isso era bom.
E disse Deus: **Façamos o homem à nossa imagem e semelhança, conforme à nossa semelhança; e domine sobre os peixes do mar, e sobre as aves do céus, e sobre todo o gado, e sobre toda à terra, e sobre todo o réptil que se move sobre a terra.**
E criou Deus o homem à sua imagem e semelhança: à sua própria imagem o criou; macho e fêmea os criou.
...E assim os céus, e a terra e todo o seu exército foram acabados. E havendo Deus acabado no sexto dia a sua obra, que tinha feito, descansou no sétimo dia de toda sua obra, que tinha feito.
E abençoou Deus o dia sétimo, e o santificou; porque nele descansou de toda sua obra, que criara e fizera.”

Gênesis 1:1, 25-27 e 2:1-3

“Não sabes, não ouviste que O Eterno Deus , O Senhor, O Criador dos fins da terra, nem se cansa nem se fatiga? Não há esquadrinhação do Seu Entendimento.
Ele dá esforço ao cansado, e multiplica as forças ao que não tem nenhum vigor. Os jovens se cansarão e se fatigarão, e os mancebos certamente cairão. Mas os que esperam no Senhor renovarão as suas forças, subirão com asas como águias: correrão, e não se cansarão; caminharão, e não se fatigarão.”

Isaías 40: 28-31

“Agora cinge os teus lombos como, como homem; e perguntar-te-ei, e tu responde-me.
Onde estavas tu, ó homem, quando eu fundava a terra? Faze-mo saber, se tens inteligência. Quem lhe pôs as medidas, se tu o sabes? Onde está o caminho da morada da luz? E, quanto às trevas, onde está o seu lugar?”

Jó 38: 3-4; 19

“Ele é o que está assentado sobre o **globo** terra, cujos moradores são para Ele como gafanhotos..” (texto escrito pelo profeta Isaías, mais de 2 mil anos antes da descoberta de que a terra realmente é redonda).

Isaías 40:22

“Ora ninguém subiu ao céu, senão O que desceu do céu, O Filho do Homem, que está no céu.

Porque Deus amou o mundo de tal maneira que deu O Seu Filho Unigênito, para que todo aquele que Nele crê, não pereça, mas tenha A Vida Eterna. Porque Deus enviou Seu Filho mundo, não para que condenasse o mundo, mas para que o mundo fosse salvo por Ele."

João 3:15-17

" E vi um novo céu e uma nova Terra. Porque já o primeiro céu e a primeira terra passaram. E Deus limpará de seus olhos toda a lagrima, e não haverá mais a morte, nem pranto, nem clamor, nem dor, porque já as primeiras coisas serão passadas. E O que estava assentado No Trono disse: Eis que faço novas todas as coisas. E disse-me: Escreve, porque estas palavras são verdadeiras e fiéis. E disse ainda: Eu sou o Alfa e o Ômega, o Princípio e o Fim. A quem quer que tiver sede, de graça lhe darei da fonte da água da Vida.

Apocalipse 21: 1-6

À Deus, por Ele ser O Provedor e O Mantenedor de todas as coisas

AGRADECIMENTOS

Pela realização deste trabalho, gostaria de agradecer:

✓ À Deus, pela amizade, por ser a A Minha Rocha, Confiança e Provisão nos momentos difíceis. Pelo constante cuidado e direção em todos os momentos, inclusive durante as etapas de elaboração, desenvolvimento e conclusão deste trabalho, que se iniciou no mestrado. Especificamente por ter respondido uma oração que fiz no dia 28/11/09 me dando a idéia deste projeto (mestrado e doutorado), e já na manhã do dia 29/11/09, ter provido as parceirias necessárias (Projeto GENOPROT/CNPq) para que o mesmo fosse iniciado no mestrado.

✓ À minha Mãe, Lucila, por sua amizade e exemplo de sabedoria, força e determinação, que me servem de inspiração. Pelas sábias palavras de apoio e pelas constantes orações. Ao meu Pai, Nasser, pela amizade e constante apoio, mas principalmente por ter me ensinado deste cedo a lutar pelos objetivos e valores que realmente valem à pena na vida. Abaixo de Deus, são os meus heróis, referenciais e Aqueles à quem mais admiro e me espelho;

✓ À Polina, por ser minha amiga e companheira nos momentos de bonança e dificuldade, pelas incessantes orações por mim e por sonhar comigo os sonhos de Deus;

✓ Ao Prof. Cláudio Mafra, por ter acreditado neste projeto, pela Fé, em 29/11/09, quando ainda não tínhamos as condições técnicas e financeiras para a realização do mesmo. Pela amizade e por ter aberto as portas de seu laboratório para mim, permitindo-me grande crescimento profissional e pessoal. Pela importante orientação e *insights* no tocante à aspectos técnicos deste trabalho, desde a concepção do desenho experimental, desenvolvimento e até a análise crítica dos dados obtidos;

✓ Ao Prof. Mathias Szabó e à Maria Marlene Olegário (UFU) por gentilmente terem cedidos carrapatos *A. sculptum* para a realização deste trabalho; das bibliotecas de cDNA, essencial para realização deste trabalho;

✓ Ao Prof. Marcelo Bahia Labruna (FMVZ-USP) por gentilmente ter fornecido o cultvo de *Rickettsia amblyommii* para as etapas de infecção experimental dos carrapatos;

✓ Aos Drs. Otavio Brustollini e Talles Maciel, pelo aprendizado e valiosa cooperação durante a análise in silico dos dados gerados, prestando valiosos insights no tocante ao processamento e interpretação dos dados de RNASeq;

✓ Aos Professores, Dr. Jorge A. Dergam, cooperação e pelas valiosas sugestões na minha banca de qualificação e pela importante disponibilidade de contribuir para o enriquecimento da discussão dos dados, através desta banca de doutorado.

✓ Dr. Donald H. Bouyer e Dr. Gustavo Ferreira Martins pela valiosa cooperação em participarem desta banca de doutorado, prestando valiosa contribuição através da análise crítica dos dados, sugestões e enriquecimento no tocante à discussão das informações geradas à partir desta tese;

✓ Á todos os amigos e companheiros de trabalho do Laboratório de Epidemiologia e Parasitologia molecular, pelo companheirismo, amizade e valiosa ajuda em todas as etapas deste trabalho, em especial Adriano Soares Moreno, Paulo Lima (Paulão) e Juliana Areal Carrizo (Juju). Verdadeiros irmãos na minha caminhada no doutorado. Ao Hugo Genier (Departamento de Microbiologia) pela amizade e valiosa ajuda no tocante à minha estadia em Viçosa.

✓ Aos amigos do Departamento de Bioquímica e demais amigos de Viçosa, pela valiosa amizade;

✓ Aos amigos da Igreja Adventista, que são a minha família em Viçosa;

✓ Á FAPEMIG e ao CNPq, pela financiamento dos projetos necessários para a execuação deste doutorado. Á CAPES, pela concessão da bolsa de estudos no doutorado;

✓ Enfim, agradeço à todas aquelas pessoas que, de uma forma ou de outra, estiveram do meu lado durante este grande desafio, contribuindo para realização deste importante trabalho;

BIOGRAFIA

Higo Nasser Sant'anna Moreira, nascido em 6 de fevereiro de 1987, na cidade de Vitória, Espírito Santo, Brasil, filho de Nasser Fransciso Moreira e Lucila Sant'anna Moreira.

No ano de 2005, ingressou no Bacharelado em Bioquímica pela Universidade Federal de Viçosa, obtendo o título de Bacharel em Bioquímica em 22 de janeiro de 2010. Durante a graduação, exerceu diversas atividades como monitor e tutor efetivo em matérias das áreas de Bioquímica e Química Orgânica. Também foi bolsista de iniciação científica pela FAPEMIG, no Laboratório de Parasitologia e Epidemiologia Molecular (DBB/UFV) em projeto intitulado: Marcadores moleculares para identificação e diferenciação de carrapatos vetores de agentes riquetsiais, sob orientação do Prof. Dr. Cláudio Mafra.

Em fevereiro de 2012, defendeu o Mestrado pelo Programa de Pós-graduação em Bioquímica Agrícola, com a Dissertação intitulada: Genoma funcional de glândulas salivares de fêmeas de *A. cajennense* em condições de infecção e não infecção com *Rickettsia* spp.

Como continuidade da linha de pesquisa, iniciou, em março de 2012, o Doutorado pelo Programa de Pós-graduação em Bioquímica Aplicada na UFV, dando prosseguimento aos estudos relacionados à transcriptomas do carrapato *A. cajennense* (*A. sculptum*). Sob orientação do Prof. Dr. Cláudio Mafra, elaborou e aprovou o projeto de pesquisa referente à esta tese (Obtenção e estudo de transcriptomas de intestino e ovários do carrapato *Amblyomma cajennense* - *A. sculptum* - em condições de infecção e não infecção com Rickettsia amblyommii) no Edital FAPEMIG 01/2011, Demanda Universal. Em julho de 2014, já com resultados preliminares da tese de doutorado, foi aprovado em processo de qualificação pelo Programa de Pós-graduação em Bioquímica Aplicada - UFV

Em janeiro de 2016, concluiu as análises *in silico* referentes ao Genoma funcional dos órgãos internos do carrapato *Amblyomma sculptum* (*Amblyomma cajennense*), finalizando assim, a parte experimental desta tese de doutorado.

APRESENTAÇÃO DA TESE

Esta tese está apresentada em formato de capítulos, redigidos como artigos científicos, e organizada da seguinte forma:

Resumo, Abstract, Introdução e Objetivos. Em seguida, três artigos científicos

Capítulo 1 – Obtenção do transcriptoma total de referência e construção de um catálogo de genes dos órgãos internos (intestino, ovário e glândulas salivares) do carrapato *Amblyomma* sculptum, apresentado e discutido no formato de artigo científico, intitulado**:** "*A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of A. sculptum".*

Capítulo 2 – Genômica funcional e análise comparativa (semi-quantitativa) à partir de dados RNAseq de intestinos de fêmeas do *Amblyomma sculptum* sob condições de infecção e não infecção por *Rickettsia amblyommii*. Apresentado e discutido no formato de de artigo científico, intitulado: "*RNASeq data analysis reveals deep changes in gene expression patterns of Amblyomma sculptum tick midguts in response to rickettsial infection*".

Capítulo 3 – Genômica funcional e análise comparativa (semi-quantitativa) à partir de dados RNAseq de ovários de fêmeas do carrapato *Amblyomma sculptum* sob condições de infecção e não infecção por *Rickettsia amblyommii*. Apresentado e discutido no formato de artigo científico, intitulado: "*The first functional genomics of Amblyomma sculptum tick ovaries under conditions of infection and non-infection with Rickettsia amblyommii*."

SUMÁRIO

LISTA DE TABELAS ..... **xiii**

LISTA DE FIGURAS ..... **xiv**

LISTA DE ABREVIATURAS, SIGLAS E SIMBOLOS ..... **xv**

RESUMO ..... **xvi**

ABSTRACT ..... **xviii**

1. INTRODUÇÃO E JUSTIFICATIVA **1**

2. OBJETIVOS **3**

2.1. Objetivo Geral **3**

2.2. Objetivos Específicos **3**

CAPÍTULO 1: A DEEP INSIGHT INTO THE WHOLE TRANSCRIPTOME OF MIDGUTS, OVARIES AND SALIVARY GLANDS OF *A. SCULPTUM* **4**

ABSTRACT **5**

2. MATERIALS AND METHODS **7**

2.1. Ticks feeding experiments, internal organs dissections and total RNA isolation **7**

2.2. cDNA libraries, NGS sequencing and datasets availability **8**

2.3. *De novo* assembly and gene prediction by Augustus **9**

2.4. Functional annotation **9**

3. RESULTS AND DISCUSSION **10**

3.1. General features of NGS data, *de novo* assembly and gene prediction of the *A. sculptum* transcriptomes **10**

3.2. Functional annotation of the whole transcriptome from *A. sculptum* internal organs **12**

3.2.1. Overview on whole transcriptome of the A. sculptum midguts (mialome) 12

3.2.2. Overview on the main ovaries transcripts 14

3.2.3. Overview on the previously described putative salivary glands transcripts 14

3.2.4. Comparative analysis of the profiles of gene expression from A. sculptum internal organs 14

3.2.5. Comparative analysis of the expression of housekeeping genes among the internal organs 18

3.2.5.1. Nuclear regulation and transcription factors families *18*

3.2.5.2. Protein synthesis and modification machinery *19*

3.2.5.3. Protein export machinery *20*

3.2.5.4. Proteasome machinery *21*

3.2.5.5. Cytoskeletal machinery *22*

3.2.5.6 . Digestive enzymes *22*

3.2.5.7. Nutrients transporters *23*
3.2.5.8. Carbohydrates/Energy metabolism *24*
3.2.5.9. Lipids Metabolism *25*
3.2.5.10. Amino acids and nucleotide metabolism *25*
3.2.5.11. Detoxification *25*
3.2.5.12. Signal transduction *26*
3.2.5.13. Immunity related products *26*
3.2.6. Putative secreted proteins 27
3.2.7. Transposable elements families 29
3.2.8. Unknown fragments and viral related proteins 29
4. CONCLUSION **29**
5. ACKOWNLEDGEMENTS **30**
6. REFERENCES **30**
CAPÍTULO 2: RNASEQ DATA ANALYSIS REVEALS DEEP CHANGES IN GENE EXPRESSION PATTERNS OF *AMBLYOMMA SCULPTUM* TICK MIDGUT IN RESPONSE TO RICKETTSIAL INFECTION **37**
ABSTRACT **38**
1. INTRODUCTION **39**
2. METHODS **40**
2.1. Ticks, experimental infection, midguts isolation and ethics **40**
2.2. Total RNA extraction, cDNA libraries and NGS sequencing **40**
2.3. Additional data, de novo assembly and CDS prediction **41**
2.4. Functional annotation and gene expression inferences **42**
3. RESULTS AND DISCUSSION **43**
3.1. Characteristics of the *A. sculptum* mialomes 43
3.2. Overview on Housekeeping genes (H) from *A. sculptum* mialomes 44
3.2.1. Transcripts putatively related with SFG rickettsiae infection mechanisms at A. sculptum midgut tract 46
3.2.2. More abundance of CDS related with gene expression control in A. sculptum infected midguts 48
3.2.2.1 *Nuclear regulation machinery* 48
3.2.2.2. *Transcription factors and other transcription related CDS* 50
3.2.2.3. *Protein synthesis and modification machinery* 50
3.2.3. Endocytosis, phagocytosis and protein export machinery genes during rickettsial infection in A. sculptum midguts 51
3.2.4. Signal transduction 53
3.2.5. Overexpression of genes related with nutrient transporters and catabolic

pathways in response to rickettsial infection 55
3.2.5.1. *Protein digestion* 55
3.2.5.2. *Carbohydrate digestion* 55
3.2.5.3. *Lipid digestion and metabolism* 55
3.2.5.4. *Nutrient transporters* 56
3.2.5.5. *Carbohydrate, amino acid and nucleotide metabolism* 56
3.2.6. Detoxification proteins 57
3.2.7. Imune related products 58
3.2.8. Proteasome machinery 60
3.3. Protein containing signal peptides 60
3.4. Transposable elements and unknown products 60
4. CONCLUSIONS 61
5. ACKNOWLEDGEMENTS 61
6. SUPPLEMENTARY DATA (ANEXOS) 61
7. REFERENCES 61
CAPÍTULO 3: THE FIRST FUNCTIONAL GENOMICS OF *AMBLYOMMA SCULPTUM* TICK OVARIES UNDER CONDITIONS OF INFECTION AND NON-INFECTION WITH *RICKETTSIA AMBLYOMMII* 67
1. INTRODUCTION 68
2. METHODS 69
2.1. Obtainment of the ticks and treatment imposition 69
2.2. Extraction of total RNA, cDNA libraries and NGS sequencing 70
2.3. *De Novo* assembly and gene prediction 70
2.4. Functional annotation and gene expression analysis 71
3. RESULTS AND DISCUSSION 72
3.1. Overall features of the ovaries transcriptomes of *A. sculptum* 72
3.2. Functional annotation and effect of rickettsial infection on gene expression of *A. sculptum* ovaries 72
3.3. Housekeeping genes 73
3.3.1. Extracellular matrix and cytoskeletal machinery 73
3.3.1.1. Extracellular matrix 73
3.3.1.2. Cytoskeletal machinery 74
3.3.2. Control machinery of the gene expression 77
3.3.2.1. Nuclear regulation 77
3.3.2.2. Transcription factors 78
3.3.3. Signal transduction 80

3.3.3.1. mTOR signaling pathway *80*
3.3.3.2. Hormones related *80*
3.3.4. Reproductive processes, nutrients assimilation and metabolism *81*
3.3.4.1. Storage/vitellogenesis *81*
3.3.4.2. Nutrient transporters *81*
3.3.4.3. Digestive proteases and amino acid metabolism *82*
3.3.4.4. Carbohydrate metabolism and digestive enzymes 82
3.3.5. *Proteins turnover, detoxification and immunity related processes* 83
3.3.5.1. Proteasome machinery *83*
3.3.5.2. Detoxification related processes *84*
3.3.5.3. Immunity related 84
3.4. Putative secreted proteins from *A. sculptum* ovaries 85
3.5. Transposable elements, viral related transcripts and unknown products 85
4. CONCLUSION 86
5. ACKNOWLEDGEMENTS 86
6. SUPLEMENTARY DATA 87
7. REFERENCES 88
ANEXOS 93
ANEXOS CAPÍTULO 1: A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of *A. sculptum* 93
Spreadsheet S1. Functional classification and distribution of CDS expressed among the *A. scultpum* internal organs 93
Spreadsheet S2. Functional classification and distribution of CDS at least 5-Fold more abundant among the *A. scultpum* internal organs 107
ANEXOS CAPÍTULO 2: RNASeq data analysis reveals deep changes in gene expression patterns of *Amblyomma sculptum* tick midguts in response to rickettsial infection 123
Spreadsheet S1. Functional classification of transcripts from the whole *A. sculptum* mialome (MDG) 123
ANEXOS CAPÍTULO 3: The first functional genomics of *Amblyomma sculptum* tick ovaries, under conditions of infection and non-infection with *Rickettsia amblyommii* 138
Spreadsheet S1. Functional classification of transcripts from the whole transcriptome of *A. sculptum* ovaries (OVA) 138

LISTA DE TABELAS

CAPÍTULO 1: A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of *Amblyomma sculptum*

Table 1. Functional classification and tissue expression profiles of the transcriptomes from *A. sculptum* internal organs **13**

Table 2. Functional classification and distribution of the CDS at least 5 fold-more abundant among *A. sculptum* internal organs **17**

CAPÍTULO 2: RNASeq data analysis reveals deep changes in gene expression patterns of *Amblyomma sculptum* tick midgut in response to rickettsial infection

Table 1. Functional classification of the 25,569 all expressed CDS from *A. sculptum* mialomes **44**

Table 2. Classification of Housekeeping genes in the *A. sculptum* mialomes (MI and MNI) **45**

Table 3. More abundance of *A. sculptum* midgut proteins putatively involved in the rickettsial infection mechanism **47**

CAPÍTULO 3: The first functional genomics of *Amblyomma sculptum* tick ovaries, under conditions of infection and non-infection with *Rickettsia amblyommii*

Table 1. General distribution of reads and coding sequences from the *A. sculptum* ovaries transcriptomes. **76**

ANEXOS CAPÍTULO 1: A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of *A. sculptum*

Spreadsheet S1. Functional classification and distribution of CDS expressed among the *A. scultpum* internal organs **93**

Spreadsheet S2. Functional classification and distribution of CDS at least 5-Fold more abundant among the *A. scultpum* internal organs **107**

ANEXOS CAPÍTULO 2: RNASeq data analysis reveals deep changes in gene expression patterns of *Amblyomma sculptum* tick midguts in response to rickettsial infection

Spreadsheet S1. Functional classification of transcripts from the whole *A. sculptum* mialome (MDG) **123**

ANEXOS CAPÍTULO 3: The first functional genomics of *Amblyomma sculptum* tick ovaries covered by Illumina HiSeq analysis data

Spreadsheet S1. Functional classification of transcripts from the whole transcriptome of *A. sculptum* ovaries (OVA) **138**

LISTA DE FIGURAS

CAPÍTULO 1: A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of *Amblyomma sculptum*

Figure 1. Functional distribution of the all 27,308 CDS from the whole transcriptome of *A. sculptum* internal organs. **11**

Figure 2. Distribution of the all 27,308 CDS from the whole transcriptome of *A. sculptum* internal organs among the three internal organs evaluated in this study **15**

Figure 3. Functional classification of the 2,177 coding sequences related to secreted proteins, in the context of the whole transcriptome of *A. sculptum* internal organs. **28**

CAPÍTULO 2: RNASeq data analysis reveals deep changes in gene expression patterns of *Amblyomma sculptum* tick midgut in response to rickettsial infection

Figure 1. Differential expression of CDS from *A. sculptum* mialomes homologues to proteins involved in the regulation of actin cytoskeleton during baterial invasion of epitheliel cells (KEGG consensus pathway). **49**

Figure 2. More abundance of CDS related to protein synthesis and modification machinery in response to rickettsial infection (MI versus MNI). **51**

Figure 3. More abundance of CDS related to vesicle transport and endocytosis in response to rickettsial infection (MI versus MNI). **52**

Figure 4. Differential expression of CDS from *A. sculptum* mialomes related to Rap1 signaling pathway in response to *R. amblyommii* infection (MI *versus* MNI). **54**

Figure 5. Differential expression of CDS from *A. sculptum* mialomes related to Glycolysis and Gluconeogenesis pathways in response to *R. amblyommii* infection (MI *versus* MNI). **59**

CAPÍTULO 3: The first functional genomics of *Amblyomma sculptum* tick ovaries, under conditions of infection and non-infection with *Rickettsia amblyommii*

Figure 1. Functional annotation of the 21.230 CDS derived from whole *A. sculptum* ovarian transcriptome (OVA) **75**

LISTA DE ABREVIATURAS, SIGLAS E SIMBOLOS

Acari	Banco de dados para sequências gênicas de ácaros
BGI	Beijing Genome Institute
BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide Basic Local Alignment Search Tool
BLASTx	Nucleotide 6-frame translation-protein Basic Local Alignment Search Tool
BLASTp	Protein Basic Local Alignment Search Tool
RPSBLAST	Reversed Position Specific Basic Local Alignment Search Tool
CDD	Conserved Domains Database
CDS	Coding sequnces
COG	Cluster of Orthologous Groups
GO	Gene Ontology
GRPs	Glycine-rich proteins
H	Housekeeping genes
KOG	Eukaryotic Orthologous Groups of Proteins
LAPEM	Laboratório de Parasitologia e Epidemiologia Molecular
MDG	Whole Transcriptome of Amblyomma sculptum midguts or midguts
Mialome	Midgut transcriptome
MI	Transcriptome of midguts from infected *Amblyomma sculptum* ticks
MNI	Transcriptome of midguts from non-infected *Amblyomma sculptum* ticks
NCBI	National Center for Biotechnology Information
nt	Nucleotides
NR	Non-Redundant database
OVA	Whole Transcriptome of Amblyomma sculptum ovaries or ovaries
OI	Transcriptome of ovaries from infected *Amblyomma sculptum* ticks
ONI	Transcriptome of ovaries from non-infected *Amblyomma sculptum* ticks
bp	Base pairs
PFAM	The Protein Family Database
SMART	Simple Modular Architecture Research Tool domain family database
S	Secreted proteins
SG	Whole salivary glands transcriptome
Sialome	Salivary glands transcriptome
SwissProt	Curated protein database
SFG	Universidade Federal de Viçosa
SPF	Specific Pathogen Free
TE	Transposable elements
V	Viral
UFV	Universidade Federal de Viçosa
U	Unknown proteins
UNIPROT	Universal Protein Resource database

RESUMO

MOREIRA, Higo Nasser Sant'anna, D.Sc., Universidade Federal de Viçosa, fevereiro de 2016. **Anotação e montagem de transcriptomas de intestino médio e ovários do carrapato *Amblyomma sculptum*, antes e após a infecção por *Rickettsia amblyommii***. Orientador: Cláudio Lísias Mafra de Siqueira.

O objetivo deste trabalho foi construir um catálogo de genes para o genoma funcional dos órgãos internos do carrapato *A. sculptum*, bem como, avaliar os padrões de expressão gênica no intestino médio (MDG) e ovários (OVA) frente à condições de não-infecção (MNI e ONI) e infecção por *Rickettsia amblyommii* (MI e OI). Como primeiro passo, foi realizada a construção de um transcriptome de referência à partir da montagem de 9 transcriptomas (200 milhões de reads) derivados do intestino médio (3), dos ovários (2) e das glândulas salivares (4). O De novo assembly gerou um total de 460,445 contigs, a partir dos quais foram preditas 27.308 CDS expressas em no transcriptome total do *A. sculptum*. O mapeamento das reads derivadas de cada órgão contra o transcriptoma de referencia permitiu a identificação de 25,569 CDS expressas nos intestinos (MDG), 21,230 CDS expressas em ovários (OVA), enquanto que 10.697 CDS foram derivados dos transcriptomas de glândula salivar (SG). A anotação do transcriptoma de referência (27.308 CDS) permitiu a identificação de 23,228 CDS relacionados a processos housekeeping, enquanto que 2.177 CDS apresentaram sequência de peptídeos (SignalP) e foram anotadas como proteínas secretads.Um total de 261 CDS foram anotadas como elementos transponíveis, 20 CDS foram relacionadas à transcritos virais, enquanto que 1,622 permaneceram sem anotação, sendo anotadas com Unknown. Entre as 29 classes de genes housekeeping, destacam-se aqueles relacionados com Sinal de transdução (3.158 CDS), seguido de 1887 CDS anotadas com os transportadores e 1.523 CDS relacionados maquinaria de transcrição. A avaliação individualizada dos transcriptomas MDG e OVA no tocante a infecção e não infecção por *R. amblyommii* (infectados vs não infectado) revelou um padrão oposto entre os dois órgãos, com quase todos os processos celulares am MDG superexpressos em resposta à infecção por riquétsia, enquanto a maioria dos processos celular em ovário (OVA) foram down regulados em resposta à infecção por *R. amblyommii*. Essa down regulação de processes em ovários de carrapatos infectados por rickettsiae corrobora estudos in vitro com outras especíceis de carraátos infectados com *Rickettsia* spp. o que concorda com outros estudos in vivo com foco ovipostue de carrapatos infectados. com queda das funções fisiológicas da femeas ingurgitada e queda no rendimento da oviposição. Também foi observado up-regulação de proteínas no intestino, em resposta à infecção, sendo essas relacionadas com processes de infecção rickettsial em ´celulas de mamíferos bem como outros invertebrados, tais como actina, complexo Arp 2/3, domínio WASP, proteína tirosina quinase, clatrinas, intergrinas e ontre outras. Estes resultados sugerem que, de forma semelhante à verificada em células de mamíferos e em outras

espécies de carrapatos, estas proteínas estar envolvidas durante o processo de infecção do trato intestinal de A. sculptum. No entanto, a down regulação das principais vias metabólicas, maquinaria de replicação e de síntese proteica em ovários em resposta à infecção por *Rickettsia* (OI contra ONI) indica que esta batéria se constiui uma carga metabólica e fisiológica para os ovários. Alpem de sugerir possível alvos a sere avaliados na interação rickettsia-carrapato, estes resultados também dão suporte à outros estudos e estratégicas para uma melhor compreensão da biologia deste vetor, bem como para o desenvolvimento de estratégias de combate e controle biológico, como vacinas e acaricidas.

ABSTRACT

MOREIRA, Higo Nasser Sant'anna, D.Sc., Universidade Federal de Viçosa, february, 2016. **Annotation and de novo assembly of transcriptomes from midguts and ovaries of *Amblyomma sculptum* ticks, before and after infection with *Rickettsia amblyommii*.** Adviser: Cláudio Lísias Mafra de Siqueira

The aim of this work was to construct a gene catalogue for the functional genome of internal organs from A. sculptum, as well as, to evaluate the patterns of gene expression from midguts and ovaries against non-infection and *Rickettsia amblyommii* infection condition. First, we construct a reference transcriptome for *A. sculptum* internal organs assembling 9 transcriptomes (200 million reads) derived from midguts (3), ovaries (2) and salivary glands (4). The assemble generated 460,445 contigs, of which we identified 27,308 CDS expressed at whole A. sculptum transcriptome. The mapping of the reads derived from each organ against this CDS reference allowed the identification of 25,569 CDS expressed at midguts (MDG), 21,230 CDS expressed at ovaries (OVA) while 10,697 CDS were derived from sialomes (SG). The annotation of the reference transcriptome (27,308 CDS) allowed the identification of 23,228 CDS related to housekeeping processes, 2,177 CDS related presented signal peptide sequence by SignalP analysis, been annotated as Secreted group; 261 CDS were annotated as transposable elements, 20 CDS were reçated to viral transcripts and 1,622 remained without annotation, been classified as unknown sequences. Among the 29 classes related to housekeeping processes, stand out those related to sinal transduction 3,158 CDS, followed by 1,887 related to transporters and 1,523 CDS related to transcription machinery. The individualized evaluation of the MDG and OVA dataset regarding rickettsial infection (infected vs non-infected) reveals an opposite pattern, with almost all midgut processes more abundant in response to rickettsial infection, while most of all ovarian processes were down regulated in response to rickettsial infection, which agrees to other in vivo studies focusing ovipostue of infected ticks. Regarding MDG datasets, we have observed more abundance of host protein related to rickettsial infection processes at mammals and tick cells, such as actin, Arp 2/3 complex, WASP domain, protein tyrosine kinase, clathrins, intergrins and other. These results suggest that, similarly to verified at mammals cells and other tick specie, these proteins could be play a role during the mechanism of infection of *A. sculptum* midguts by *R. amblyommii.* However, the down regulation of metabolic pathways, replication and protein synthesis machinery at ovaries in response to rickettsial infection (OI versus ONI) indicates that *R. amblyommii* play role as a metabolic and physiological burden for ovaries. These results suggest putative candidates proteins for a rickettsial infection mechanism of *A. sculptum* midguts and give support for further studies and strategies focused on biology control of the tick, as well, a better understanding about the interface *A. sculptum* – SFG rickettsiae.

1. INTRODUÇÃO E JUSTIFICATIVA

O carrapato *Amblyomma sculptum* destaca-se como uma das espécies de maior importância epidemiológica e zoonótica em uma extensa região do continente americano, quem vai deste o Sul dos Estados Unidos até o Norte da Argentina (Billings et al., 1998; Beck et al., 2011; Mastropaolo et al., 2011; Nava et al., 2014; Ramos et al., 2014). No Brasil, o *Amblyomma sculptum* é o carrapato que apresenta maior incidência de parasitismo em humanos, sendo o principal vetor da Febre Maculosa Brasileira, agravo de notificação compulsória que figura entre os de maior letalidade neste País (Lemos et al., 2002; Lamas et al., 2008; Nunes et al., 2015;).

O sucesso deste carrapato em espoliar uma grande variedade de hospedeiros vertebrados e, por conseguinte, transmitir zoonoses, esta intimamente relacionado à sua morfologia, bem como a fisiologia dos órgãos internos, sobretudo glândulas salivares, intestino e ovários. As glândulas salivares atuam produzindo um complexo coquetel de proteínas anti-inflamatórias, anti-hemostáricas, vasodilatadoras, que medeiam o respato sanguíneo (Francischetti et al., 2009) bem como proteínas ricas em glicina (componentes do cone de cemento) que auxiliam na fixação do hipostômio do ectoparasita ao sitio de alimentação na pele do hospedeiro vertebrado (Maruyama et al., 2010), além de atuarem como principal via de transmissão de riquétsias.

Além das glândulas salivares, o trato intestinal do carrapato é um importante órgão no processo de infecção do carrapato com a *Rickettsia* spp., visto realizar a maior parte da interface do carrapato com o meio externo, inclusive com o sangue contaminado obtido quando do repasto sanguíneo, podendo ser porta de entrada de patógenos. De fato, em outras espécies de carrapatos ixodídeos, como o Ixodes scapulares, antígenos intestinais do carrapato, tais como a Proteina TROSPA foram descritos como sendo a principal via de infecção do carrapato vetor pelo agente causador da Doença de Lyme nos EUA, a Borrelia burgdorferi (Narasimhan et al., 2014, Pal et al., 2004). Além disso, o conhecimento sobre antígenos intestinais destes ectoparasitas tem dado suporte à estratégicas promissoras de controle biológico e combate à estes artrópodes, como desenvolvimeno de vacinas e acaricidas (Galay et al., 2014; Jiang et al., 2014). No entanto, embora a ecologia e a taxonomia do *A. sculptum* tenha sido bem compreendida nos últimos anos, pouco se conhece sobre o repertório de genes ativados ou reprimidos na maior em seu trato intestinal, sobretudo quando dos eventos moleculares que regem a sua interação com riquétsias do grupo da frebre maculosa (GFM).

Outra questão chave na ecologia do *A. sculptum*, bem como em seu papel como agente de zoonoses, esta na sua grande capacidade de proliferação, podendo uma fêmea ovipor de 5 a 10 mil ovos (Barros-Battesti et al., 2006). Como agravante associado a este fator, está o fenômeno transmissão transovariana de riquétsias, através do qual ocorre a infecção dos ovos pela bactéria e

consequente infecçao da prole, o que resulta proliferação/manutenção de carrapatos infectados no ambiente (Labruna et al., 2011). Desta forma, um melhor entendimento dos processos moleculares envolvidos nestes órgãos constitui-se etapa fundamental para o melhor entendimento da biologia deste carrapato, bem como dos eventos moleculares que regem seu papel como agente vetor de riquétsias.

O advento das plataformas de sequenciamento de nova geração (NGS) tem proporcionado possibilidade sem precendentes de exploração do repertório/expressão de genes em várias espécies de carrapatos (Chmelar et al., 2015), inclusive com a identificação e bioprospecção de moléculas de interesse farmacológicos, como o agente antitumorais (Chudizinski-Tavassi et al., 2016). No entanto, embora estudos anteriores tenham descrito sialomas do *A. sculptum* (Batista et al., 2008; Maruyama et al., 2010, Garcia et al., 2014), o conhecimento do repertório de genes expressos no intestino e ovário deste importante carrapato vetor, como como estes orquestram processos de interação com riquétsias causadoras da febre maculosa permanece escasso. Tais conehcimentos seriam de grande valia, não somente para o melhor entendimento da interface riquétsia-carrapato, como tabém seria estratégia chave desenvolvimento de estratégias moleculares para o controle deste importante carrapato vetor.

Diante disso, o objetivo, o objetivo deste trabalho foi obter e construir um catálogo de genes dos principais órgãos internos do *A. sculptum* (glândulas salivares, trato intestinal e ovários) visando a melhor elucidação da interface carrapato-rickettsia ao nível do trato intestinal e dos ovários. Tal base de dados permitira também a identificação genes e moléculas promissoras como futuros alvos para estratégias de bioprospecção e controle biológico deste vetor.

2. OBJETIVOS

2.1. Objetivo Geral

Avaliar a expressão gênica dos ovários e intestinos do carrapato *A. sculptum*, agente transmissor da FMB, frente a situação de (i) Não-infecção e (ii) Infecção com *R.* a*mblyommii;*

2.2. Objetivos Específicos

2.2.1. Construção de um transcriptoma de referência e um catálogo dos genes expressos nos órgãos internos (glândulas salivares, intestino e ovários) de fêmeas do carrapato *A. sculptum.*

2.2.2. Obtenção de transcriptomas do trato intestinal (mialomas) de fêmeas semi-ingurgitadas (72-96 horas de alimentação) do carrapato *A. sculptum* nas condições: (i) não infecção e (ii) infecção artificial com *R. amblyommii.*

2.2.3. Obtenção de transcriptomes de ovários de fêmeas ingurgitadas (8-10 dias de alimentação) do carrapato *A. sculptum* nas condições de: (i) não-infecção e; (ii) infecção artificial com *R. amblyommii*.

2.2.4. Anotar e avaliar possíveis genes codificadores de proteínas-chave em mecanismos envolvidos no ciclo da *R. amblyommii* nos órgãos internos de seu carrapato transmissor.

CAPÍTULO 1

A DEEP INSIGHT INTO THE WHOLE TRANSCRIPTOME OF MIDGUTS, OVARIES AND SALIVARY GLANDS OF *A. SCULPTUM*

A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of *A. sculptum*

Higo Nasser Santanna MOREIRA[1], Talles MACIEL[1], Juliana CARRIZO[1], Rafael Mazioli BARCELOS[1], Carlos MONTANDON[1], Maria Marlene MARTINS[2], Paulo LIMA[2], Adriano Soares MORENO[2], Cláudio MAFRA[1¥]

[1] Biochemistry and Molecular Biology Department, Federal University of Vicosa, Brazil; [2] Veterinary Department, Federal University of Uberlandia, Brazil;

[¥] mafra@ufv.br

ABSTRACT

Ticks stand out as the second most important vector of pathogens to humans and its control is a key issue for public health strategies. In the Americas, *Amblyomma sculptum* stands out as one of the most important species as it is capable of infesting a wide range of vertebrate hosts, including humans. The success of this arthropod on its parasitism on vertebrate host, as well as, its importance in rickettsiae ecology are intimately related with the physiology of its main internal organs, such as midguts, ovaries and salivary glands, that participate as key sites for rickettsial entry into tick hemolymph, ovarian amplification and transmission to vertebrate hosts, respectively. Despite transcriptomic efforts have shed light on some aspects of *A. sculptum* sialomes, very few is known about the repertoire of genes that orchestrate the physiology of midguts and ovaries. In addition, as key sites for the development of tick control strategies, the knowledge about biological processes, antigens and molecular targets in *A. sculptum* midguts and ovaries remains uncovered. In this sense, the aim of this study was to construct a deep catalogue from the repertoire of genes and putative proteins expressed in the midguts, ovaries and salivary glands of *A. scultpum*, through the employment of RNAseq strategy. For this purpose, a total of 200 million reads derived from midguts, ovaries and salivary glands and sequenced by Illumina Hiseq, Ion Torrent PGM and 454 pyrosequencing, were assembled into contigs, of which were predicted 27,308 CDS. The mapping step allowed the identification of 25,569 CDS expressed in midguts, 21,230 at ovaries and 10,697 at salivary glands. Functional annotation allowed to identify 23,228 housekeeping genes expressed in all organs, while 2,177 CDS were annotated as secreted proteins. Midgut transcriptomes (mialomes) analysis reveals several enzymes related to digestion and metabolism enzymes of proteins, carbohydrates and lipids, as well as, transcripts related to detoxification processes and immunity processes. Ovaries transcriptomes analysis allows the identification of transcripts related to vitellogenin and vitellogenin receptors, which could be related with oogenesis/vitellogenesis processes. Classical proteins from sialomes, such as kunitz domain, serpins, lipocalins and microplusins were identified in salivary glands transcriptomes, but, also were very abundant at mialomes and ovarian transcriptomes, suggesting novel processes. Indeed, making the comparison among the three organs, we observed several CDS most expressed in mialomes and ovarian transcriptomes, in comparison to salivary glands, such as digestive and metabolic enzymes, transcription factors and protein synthesis, that suggests a higher metabolic activity in those organs. In opposite, we did not observe deep differences between the abundance of transcripts in midguts and ovarian transcriptomes, which could be related with growth conditions of these tissues during the tick engorgement. The annotation approach employed in this work allowed to discover a higher number of novel genes and processes at midguts and ovaries from *A. sculptum* and the overexpression of classical secreted proteins in these organs, such as lipocalins, serpins, kunitz, in comparison to salivary glands, could be related with novel physiological functions at the midguts and ovaries.

Key words

Amblyomma sculptum, high throughput sequencing, RNAseq, whole transcriptome, internal organs

1. INTRODUCTION

Ticks are obligate hematophagous arthropods (Acari Order, Class Arachnida) widely found in the most diverse ecosystems on the globe (Morshed et al., 2005; Dietrich et al., 2011; Beati et al., 2013). For being able to parasitize a wide range of vertebrate hosts, these arthropods figures at the second most important group of pathogen vectors agents to man, behind only the mosquitoes (Edlow et al., 2008; Parola & Raoult, 2001).

Among the main species of ticks of medical and veterinary importance, stand out *Amblyomma sculptum*, the main vector of spotted fever rickettsiosis in a large region of the Americas, since Southern United States to Northern Argentina (Billings et al., 1998; Beck et al., 2011; Mastropaolo et al., 2011; Alonso-Díaz et al., 2013; Nava et al., 2014; Ramos et al., 2014). This species presents an aggressive parasitism, being able to infest naturally a wide range of species, since wild and domestic vertebrates, to human (Lopes et al., 1998; Rojas et al., 1999; Labruna et al., 2002). Due morphological differences presented by tick populations from many regions of American continent, this specie was taxonomically renamed from *A. cajennense* to *A. sculptum* (Nava et al., 2014).

The success of this arthropod to infest a great variety of vertebrate hosts and disseminate Spotted Fever Group (SFG) rickettsiae depends on the physiology of its internal organs, especially midguts, ovaries and salivary glands. At the first time, ticks depend to the mouth parts adaptation and changes in the gene expression patterns of salivary glands during blood sucking and engorgement processes (Kovar, 2004; Anatriello et al., 2010; Francischetti et al., 2010; Maruyama et al., 2010). In turn, tick intestinal tract stands out as the largest interface with external environment, once metabolizes proteins and other nutrients from the ingested blood (Sojka et al., 2013). In addition, this organ plays role as the main entry path of pathogens, with epidemiological implications, such has been described for *Ixodes scapularis* midguts infection by *Borrelia burgdorferi* (Narasimhan et al., 2014, Pal et al., 2004). Regarding ovary, this organ stand out as the main tick biological structure for the SFG rickettsiae biology, once it play role in the transovarian transmission and amplification of rickettsiae in the environment (Soares et al., 2012). By other hand, the elucidation of molecular processes in this organ constitutes an essential step for the *A. sculptum* epidemiology, once could support further strategies for biological control of this important tick species.

In this context, in the last fourteen years, transcriptomics has emerged as a promising strategy to discover novel genes and peptides from tick tissues (Valenzuela et al., 2002; Anderson et al., 2008; Francischetti et al., 2009; Garcia et al., 2014; Chmelar et al., 2015) especially for the discovering of pharmacological peptides with antitumoral activities (Chudzinski-Tavassi et al.,

2016). Although there are three previous works have covering *A. sculptum* sialomes (Batista el al., 2008; Maruyama et al., 2010; Garcia et al., 2014), according our knowledge, there are no transcriptomic data about *A. scultpum* ovaries and midguts until this study. Thus, the knowledge about the transcriptomic repertoire that orchestrate biological and physiological processes at *A. sculptum* midguts and ovaries remains understudied.

Thus, aiming to shed light on molecular details about the biology of *A. sculptum* internal organs, our objective was to construct a deep catalogue from the repertoire of genes and putative proteins expressed in the midguts, ovaries and salivary glands of *A. scultpum*, the most important tick for rickettsiology in Central and South Americas.

2. MATERIALS AND METHODS

2.1. Ticks feeding experiments, internal organs dissections and total RNA isolation

Tick specimens of *A. sculptum* were maintained at BOD incubator, according protocol previously described (Kurti et al., 1996). For the midguts transcriptomes, two groups of 75 *A. sculptum* couples (one group previously infected with *R. amblyommii* and other non-infected) were fed on rabbits during 72-96 hours. The non-infected male ticks were utilized just to fertilize the *A. sculptum* females specimens for both treatments and ensure their feeding until semi-engorgement. The ovaries from *A. sculptum* were obtained from both the infected and non- infected tick groups. However, the females in both of these groups were allowed to feed until complete engorgement (8-10 days of feeding) in order to evaluate the oogenesis/vitellogenesis processes. Ticks from the infected groups were checked by the Real Time PCR in order to verify the presence of rickettsial DNA, according Labruna et al., 2007. In brief, a sample of ticks from infected treatments were tested by Real Time PCR assay with primers CS-5 (forward) and CS-6 (reverse), which are specific for a 147-bp fragment from citrate synthase gene (gltA) belonging for all genes *Rickettsia* species (Labruna et al., 2007). The reactions were performed for 25 µl per reaction, with 12.5 uL of reaction buffer (Maxima SYBR Green/ROX qPCR Master Mix, according manufactor recommendations) 1,25 µl of each primers C5 (forward) and C6 (reverse) at 15 µM, 1,5 µl of DNA from ticks for detection of DNA belonging to genus *Rickettsia*. The reactions were completed with 8,5 µl of molecular-grade water. These reactions were carried out at Applied Biosystems® StepOne™ Real-Time PCR System (Applied Biosystems, USA), according the following conditions: 1 cycle at 95°C for 5 min, followed by 50 cycles of 15 s at 95°C, 30 s at 53°C, and 30 s at 60°C. Molecular-grade water (1,5 uL) was used as negative control, while DNA of VERO cells infected with *R. amblyommii* was used as positive control.

The midguts and ovaries from each treatment were dissected according Maruyama et al. (2010) at RNAse free conditions, though treatment by RNAZap (Invitrogen, San Diego, CA, USA)

and the collected tissue were maintained at RNA Later (Ambion®, USA) at -80°C, to avoid RNA degradation. The total RNA extraction of the samples were carried out using the RNAqueous® RNA extraction kit (Ambion®, USA), according manufacturer's recommendations. The measurements of RNA quality and concentration were determined by NanoDrop 2000 (ThermoScientificTM, USA) and Bioanalyser Agilent 2100 (Agilent Technologies, Santa Clara, USA) at the Center of Biomolecules, Federal University of Vicosa, Brazil. All the procedures of this work were approved by Animal Experimentation Ethics Committee from the same University, under protocol number 07/2012, according to Ethical Principles on Animal Research established by the Brazilian College for Animal Experimentation (COBEA) and with actual Brazilian legislation.

2.2. cDNA libraries, NGS sequencing and datasets availability

The cDNA libraries representative of the two ovaries –samples and two midguts samples were constructed using Clontech SmartPCR cDNA kit (Clontech Laboratories, Inc, Mountain View, CA USA) through the temperature cycle of 6 min at 68°C, followed by 15 cycles of 10 min at 70 °C, according manufacturer protocols. The adapator sequences were removed using restriction digestion enzymes. After synthesis of double stranded cDNA, the conversion of cohesive cDNA ends was performed in blind ends (end repair) using T4 DNA polymerase, Klenow fragment of DNA polymerase I and T4 polynucleotide kinase. The subsequent purification was performed with aid of beads and magnetic hack from the kit. After this step, all the blunt ended cDNA samples were adenylated at the 3 'end. The sequencing adapters, containing the "T" ends, were ligated to the ends of cDNA containing the "A" ends with the aid of T4 DNA ligase. Then, cDNA sequences smaller than 200pb were selected with aid of electrophoresis in an agarose gel and then, amplified by PCR using specif primers targeting the adapters linked to double stranded cDNA. Finally, the amplified cDNA libraries were purified using QiaQuick PCR extraction kit (CA, USA) and eluted with elution buffer EB from the same kit. The evaluation of the quality, distribution and of libraries was performed with aid of Bioanalyser Agilent 2100 and the Illumina libraries for each treatment were prepared using NEBNext reagents (New England Biolabs, Ipswich, MA USA) according manufacturer recommendations. The quantity, quality and size distribution of these libraries were checked by Bioanalyser Agilent 2100. Finally, the sequencing of all these libraries were performed using Illumina Hiseq 2000, according standard protocol from the manufacturer. Archives contacting paired-ends sequences from each library were generated and their quality were checked through FASTQC algorithm (Babraham Institute, Cambridge, UK). An additional cDNA library constructed from the total RNA from *A. sculptum* females non-infected were sequenced by Ion Personal Genome Machine® (PGM™, USA) System, according manufacture protocols. The ovaries and midguts reads generated in this work were deposited at NCBI under the bioproject accession

number PRJNA309641. An additional dataset from constituted by four *A. sculptum* female sialomes (SG), previously published (Garcia et al., 2014), were downloaded from the Sequence Read Archive (SRA) under accession SRX498136.

2.3. *De novo* assembly and gene prediction by Augustus

The trimming step and *de novo* assembly of the nine transcriptomic datasets from midguts, ovaries and salivary glands into contigs were carried out by employment of CLC Genomics Workbench 8.0 package (CLC Bio, Denmark). The contigs dataset were searched for the presence of coding sequences (CDS) through the Augustus algorithm (http://augustus.gobics.de) that employs a generalized hidden Markov Model for identification of patterns associated with gene sequences. This particular approach is considered to be more accurate for gene prediction for eukaryotic sequences (Stanke et al., 2004; Stanke et al., 2005; Stanke et al., 2006; Hoff et al., 2015). The software was previously trained against the most current *Ixodes scapularis* structural genome deposited on GenBank (http://metazoa.ensembl.org/Ixodes_scapularis/Info/Index). The CDS dataset derived from this analysis were designated whole transcriptome from *Amblyomma sculptum* internal organs. For the identification of CDS expressed in each *A. sculptum* organ (MDG, OVA or SG), we performed the mapping of the reads from each organ against the CDS from the whole transcriptome of *A. sculptum* internal organs, through CLC Genomics Workbench 8.0 package (CLC Bio, Denmark). The CDS derived from midguts were designated whole transcriptome of *A. sculptum* midguts (MDG), while those related to ovaries and salivary glands were designated whole transcriptome of *A. sculptum* ovaries (OVA) and whole transcriptome of *A. sculptum* salivary glands (SG), respectively.

2.4. Functional annotation

The annotation step was carried out performing BLAST comparison of the 27,308 CDS related with the whole transcriptome of *A. sculptum* internal organs against several databases through the BLASTp, BLASTx, RPSBLAST and BLASTn algorithms (Altschul et al., 1990). We did adopt an E-value cut-off of 1E-04 for the BLAST result acceptance, as previously reported (Karim et al., 2011). For the BLAST searches of homology, we employed the following: (1) the non-redundant protein database (NR) of the National Center for Biology Information (NCBI); (2) Gene Ontology; (3) SwissProt; (4) UNIPROT; (5) NCBI-Acari; (6) CDD; (7) COG; (8) KOG; (9) PFAM; (10) SMART; (11) UNIPROT-Acari; (12) *Bombyx* mori; (13) *Chelicerata*; (14) rickettsiae; (15) *Amblyomma americanum*; (16) *A. cajennense* (*A. sculptum* complex); (17) *A. maculatum*; (18) *A. parvum*;(19) A. *triste*; (20) *Argas*; (21) *Dermacentor*; (22) *I. scapularis*; (23) *I. ricinus* ; (24) *Rhipicephalus* and (25) *Haemaphysalis*. In addition, for the functional annotation, we determined CDS coding signal peptides, transmembrane domains, addressing subcelular localization and related

metabolic pathways, through SignalP, TMHMM, TargetP and KEGG database, respectively. For the final annotation, consideration was given for the following parameters: (1) Minor E-value (<1E-04); (2) Higher score and (3) key-words from the main blast results (KOG, COG, GO, SwissProt, UniprotKb, Acari, etc). The annotation of all the CDS derived from whole transcriptome of *A. sculptum* internal organs was organized into an Excel spreadsheet that can be explored in a large scale, though Visual Basic programs (Guo et al., 2009). For the gene expression comparisons, normalization of the transcriptomes was performed using the RPKM calculus algorithm for each CDS, which considers the expression values in MDG, OVA and SG (Kotsyfakis et al. 2015). Venn diagram analysis was performed using R package (https://www.bioconductor.org/).

3. RESULTS AND DISCUSSION

3.1. General features of NGS data, *de novo* assembly and gene prediction of the *A. sculptum* transcriptomes

To construct the whole transcriptome database of *A. sculptum* internal organs, we performed the *de novo* assembly of 195,594,989 trimmed reads belonging to the nine *A. sculptum* datasets. The detailed distribution of the reads is as follows: (1) 45,417,679 reads from *A. sculptum* Ovaries Non-Infected (ONI); (2) 46,573,835 reads from *A. sculptum* Ovaries Infected with *R. amblyommii* (OI); (3) 49,209,818 reads from Midguts Infected with *R. amblyommii* (MI); 53.006.356 reads from Midguts Non-Infected (MNI). All these reads were obtained from Illumina HiSeq 2000 platform. An additional dataset from MNI were yielded by PGM Ion Torrent, with 844,529 reads, and included into the *de novo* assembly. We also assembled into this data 703,210 pyrosequencing reads belonging from four *A. sculptum* sialomes previously published (Garcia et al., 2014). These sialome datasets were included in our analysis in order to improve the *de novo* assemble analysis, as well as, to give representation for salivary glands genes in the whole transcriptome of *A. sculptum* during annotation and gene prediction analysis.

All these 195,494,989 trimmed reads were assembled into 460,445 contigs, with length ranging from 100 to 14,579pb and average size of 222pb. The contigs size distribution showed that 25,115 contigs (5,45%) were larger than 1,000pb, 8,372 contigs (1,82%) had a greater length than 2,000pb and 426,958 contigs (92,73%) were shorter than 1,000pb. From this contigs dataset, a total of 27,427 CDS were predicted by *Augustus* algorithm software, of which 27,308 CDS presented mapped reads and were selected to annotation step.

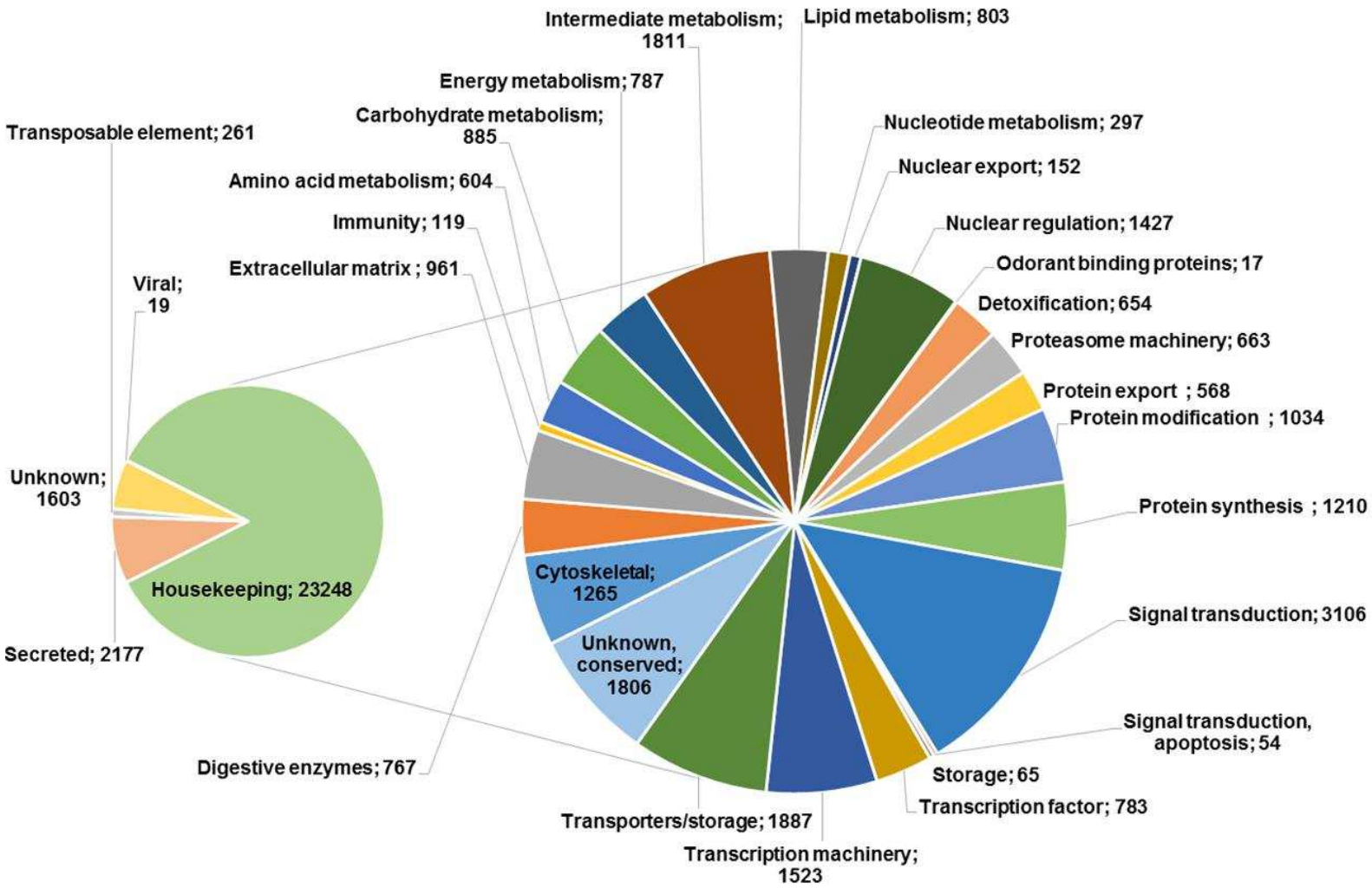

Figure 1. Functional classification of the whole transcriptome of internal organs from *A. sculptum* (27,308 CDS).

3.2. Functional annotation of the whole transcriptome from *A. sculptum* internal organs

All the 27,308 CDS identified from the whole transcriptome of *A. sculptum* internal organs (midgut tract, ovaries and salivary glands) were manually classified into 29 functional classes (Table 1 and Fig. 1), through annotation approach widely employed for other tick species (Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009). For details, see Methods. From these functional classes, twenty five were related to housekeeping processes (H), while the four other classes were related to CDS which presented signal peptide sequences at the SignalP analysis, classified as Secreted (S), Transposable elements (TE), Viral related products (V) and Unknown (U). The Unknown class contains CDS that did not present hits against databases or, whose results did not achieved the Evalue cut-off required for the annotation (See Methods). A deep insight into the 29 functional classes identified in the annotation process is provided by the Spreadsheet S1, which serves as a guide to the follow discussions. In order to give a brief insight on repertoire of genes expressed among the *A. sculptum* tissues, we performed the mapping analysis of the reads originated from each tissue against the 27,308 CDS from the whole *A. sculptum* transcriptome. The complete annotation of this CDS dataset, as well as, the complete profile of those expressed among the *A. sculptum* internal organs is discriminated at the Spreadsheet S3, which was constructed to serve as an informational source for the Spreadsheets S1 and S2.

3.2.1. Overview on whole transcriptome of the A. sculptum midguts (mialome)

There is a little information about midgut transcriptomes (mialomes) from the *A. sculptum*, been the present work one of the first attempts to give information toward this issue. Indeed, some review have discussed the advances on the NGS in the tick transcriptomes studies and recognized the minor information about tick mialomes, in comparison to the well described ticks sialomes (Anderson et al., 2008; Chmelar et al., 2015). Through the mapping of the midguts reads against the 27,308 CDS from the whole *A. sculptum* transcriptome, we have identified 25,569 CDS expressed in the *A. sculptum* mialome (MDG) tissue, (Table 1, Spreadsheet S1). We observed large number of CDS related to signal transduction processes (2,919), which was expected, once this organ orchestrates the largest interface of the tick with the external environment. In this sense, genes related to detoxification processes (614) and immunity related proteins (113), which play roles protecting the tick against vertebrate host immune response (Pichu et al., 2011; Duscher et al., 2014) and pathogens (Hajdusek et al., 2013; Takechi et al., 2016), has been annotated in MDG, as well. As previously observed for ixodidic tick *Ixodes ricinus* mialomes (Kotsyfakis et al., 2015), transcripts related to nutrient assimilation processes, such as transporters (1,755), digestive enzymes of sugars (142), proteins (417) and lipids (170), as well as, those related to several metabolic processes (4,947) were also abundant at the *A. sculptum* MDG.

Table 1. Functional classification and tissue expression profiles of the transcriptomes from *A. sculptum* internal organs

FUNCTIONAL CLASS	MDG	OVA	SG	M/O/S	M/O	M/S	S/O	MDG*	OVA *	SG*	TOTAL
HOUSEKEEPING (H)	**21,775**	**18,196**	**9,434**	**8,693**	**8,190**	**97**	**504**	**4,795**	**809**	**140**	**23,228**
Nuclear regulation	1,351	1,122	624	574	474	7	41	296	33	2	1,427
Transcription factor	735	744	382	355	346	4	18	30	25	5	783
Transcription machinery	1,438	1,175	743	685	423	12	28	318	39	18	1,523
Nuclear export	145	126	92	86	35	0	4	24	1	2	152
Protein synthesis machinery	1,158	895	533	509	341	4	13	304	32	7	1,210
Protein modification machinery	963	800	458	422	314	4	25	223	39	7	1,034
Protein export machinery	530	503	314	291	182	2	13	55	17	8	568
Proteasome machinery	610	638	405	382	207	4	15	17	34	4	663
Cytoskeletal	1,178	1,107	669	622	413	2	30	141	42	15	1,265
Extracellular matrix	890	759	394	352	343	6	30	189	34	6	960
Transporters	1,755	1,392	702	616	656	6	68	477	52	12	1,887
Lipid enzymes	170	171	56	52	112	1	3	5	4	0	177
Digestive proteases	417	431	182	175	230	0	7	12	19	0	443
Sugars enzymes	142	145	70	69	71	0	1	2	4	0	147
Storage	62	58	17	16	39	0	1	7	2	0	65
Carbohydrate metabolism	847	652	329	304	299	5	17	239	32	3	899
Lipid metabolism	751	636	272	250	339	3	14	159	33	5	803
Amino acid metabolism	580	388	200	190	174	2	8	214	16	0	604
Nucleotide metabolism	291	182	96	92	85	0	3	114	2	1	297
Energy metabolism	758	539	284	275	241	1	2	241	21	6	787
Detoxification	614	514	213	201	289	2	8	122	16	2	640
Intermediate metabolism	1,720	1,093	484	455	552	0	24	713	62	5	1,811
Signal transduction	2,919	2,725	1,477	1,355	1,151	25	77	388	142	20	3,158
Immunity	113	98	3	3	89	0	0	21	6	0	119
Unknown, conserved	1,638	1,303	435	362	785	7	54	484	102	12	1,806
SECRETED RELATED (S)	**2,033**	**1,937**	**963**	**879**	**935**	**4**	**59**	**215**	**64**	**21**	**2,177**
TRANSPOSABLE ELEMENTS (TE)	**256**	**194**	**101**	**96**	**94**	**3**	**1**	**63**	**3**	**1**	**261**
UNKNOWN (U)	**1,487**	**886**	**195**	**152**	**609**	**3**	**30**	**723**	**95**	**10**	**1,622**
VIRAL (V)	**18**	**17**	**4**	**4**	**11**	**0**	**0**	**3**	**2**	**0**	**20**
TOTAL	**25,569**	**21,230**	**10,697**	**9,824**	**9,839**	**107**	**594**	**5799**	**973**	**172**	**27,308**

Note: MDG: midgut transcripts; OVA: Ovarian transcripts; SG: Salivary glands transcripts; M/O: Midgut/Ovarian transcripts; M/O/S: Midgut/Ovarian/Salivary gland transcripts; M/S: Midguts/Salivary glands transcripts; S/O: Ovarian/Salivary glands transcripts. OVA*: transcripts exclusively expressed at ovaries; MDG*: Transcripts exclusively expressed at midguts; SG*: Transcripts exclusively expressed at salivary glands.

3.2.2. Overview on the main ovaries transcripts

The morphological features of the *A. sculptum* females ovaries have been described by Denardi et al. (2004). In this way, by the mapping of ovarian reads against the whole *A. scultpum* dataset, we have identified 21,230 coding sequences expressed at *A. sculptum* ovaries (OVA). Table 1 and Spreadsheet S1 show the functional classes as annotated from this dataset. Among the most abundant transcripts of note are those related to transcription factors (744), as well as, others related to gene expression control, such as transcription machinery (1,175), protein synthesis (895). It is worthy to note the presence of transcripts related to lipids metabolism and storage of nutrients, which could be involved in the increase of fat body cells, observed during the vitellogenesis/oogenesis processes (Dernardi et al., 2009). Once the OVA transcriptome is representative of engorged *A. sculptum* females, the abundance of these transcripts could be related to the deep changes that reproductive tissue undergoes during the oogenesis/vitellogenesis processes, with an organ size growth, as well as, increase in the number and size of lipids, proteins and sugar granules (Dernardi et al., 2004; Denardi et al., 2009).

3.2.3. Overview on the previously described putative salivary glands transcripts

The mapping analysis of the pyrosequencing reads dataset against the data base containing the entire transcriptome allowed the identification of 10,697 specific CDS from *A. sculptum* salivary glands (designated as SG). The observed number of annotated coding sequences was higher than others previously described for *A. sculptum* sialomes covered by Sanger sequencing (Batista et al., 2008) and 454 pyrosequencing (Garcia et al., 2014) with 1,754 ESTs and 4,604 CDS, respectively. These results could represent a gain of knownledge about genes and proteins from *A. sculptum* sialome, which corroborates the approach employed in the present work to discover novel genes and biological processes from this ixodidic tick.

It is worthy to note 963 coding sequences whose proteins presented signal peptide by SignalP analysis and, thus, annotated as Secreted proteins (Spreadsheet S1). Among the main secreted proteins families annotated, several previously described families stand out such as sulfotransferases (629 CDS), kunitz domains (28). lipocalins (59), serpins (51), secreted proteins related with host immunity modulation (107) and others (Valenzuela et al., 2002; Batista et al., 2008; Francischetti et al., 2010; Garcia et al., 2014; Kotsyfakis et al., 2015).

3.2.4. Comparative analysis of the profiles of gene expression from A. sculptum internal organs

In order to obtain a better understanding of the relationships of gene expression among the three *A. sculptum* tissues transcriptomes, we compared the expression of the CDS among the threes organs (Fig. 2). Our analysis identified (Table 1 and Spreadsheet S1) 9,824

CDS simultaneously expressed in the three organs (M/O/S) with the majority of the genes related to signal transduction pathways (1,355 CDS) followed by Secreted related proteins (879), Transcription machinery (685), Cytoskeletal (622) and Nutrient and Transporters (616). These results of this comparison indicate a high level of corroboration in the housekeeping genes and processes among the three internal organs.

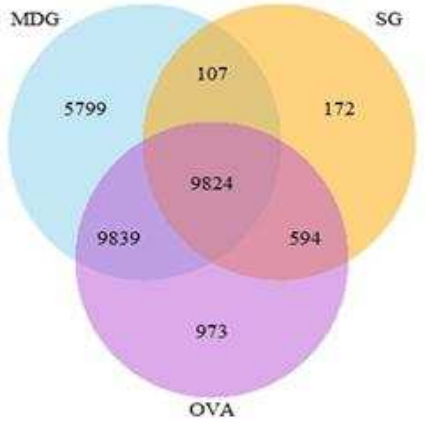

Figure 2. Distribution of the all 27,308 CDS from the whole transcriptome of *A. sculptum* internal organs among the three internal organs evaluated in this study. The functional classes belonging to each Venn diagram area are discriminated at the Table 1. MDG: Midgut; SG: Salivary glands; OVA: Ovarium.

The majority of the CDS were co-expressed between MDG and OVA (M/O) with 9,839 expressed CDS, which suggests a higher number of specific cellular processes concentrated between these organs. Indeed, these organs undergo a great development at the later stages of the blood feeding stages of *A. sculptum* females, which could be related with our results (Denardi et al., 2004; Caperucci et al., 2009; Caperucci et al., 2010). In contrast, we observed a lower number of co-expressed CDS between MDG and SG (M/S), with only 107 CDS. The most represented functional classes of this subset were those related to gene expression machinery, signal transduction pathways and nutrient transporters and protein synthesis and modification machinery (Table 1). A similar pattern was observed during our analysis of CDS co-expressed between S/O transcriptomes, with only 594 related CDS. It was our observation that transcripts related to cytoskeletal and extracellular matrix were more abundant in this subset in comparison to transcripts belonging to the same class in M/S.

In genes that exhibited exclusive expression in each tissue, we observed that MDG* expressed a higher number, with 5,799 CDS, followed by OVA* (973) and SG* (172), as shown in the Figure 1. This result suggests a greater number of specific cellular processes in the *A. sculptum* midgut in comparison to the other organs, which is consistent with its role as the major focus of the tick- vector interface with the external environment and nutrient assimilation as previously discussed. In other tissue-specific expression comparisons, transcripts related to intermediary metabolism represented the functional classes with higher number of genes expressed in MDG*.

These results could be explained by the need of the tick to metabolize a wide range of molecules from the blood of the vertebrate host (Sojka et al., 2013). There is also a possibility that the smaller numbers of CDS obtained from the SG dataset for this particular transcriptome is due to the pyrosequencing on the platform 454, which generates a smaller number of reads in comparison with Illumina HiSeq 2000 (Liu et al., 2012; Frey et al., 2014).

In order to obtain more details about the differentially expressed genes among *A. sculptum* internal organs, we performed a semi-quantitative comparison of the expression levels among the transcriptomes MDG, OVA and SG, utilizing the RPKM normalization method among the libraries (Dillies et al., 2012). For this purpose, the further discussion was carried regarding the results at the Spreadsheets S1, S2 and S3 (Supplemental information), which contains a deeper annotation about the functional classes listed at the Tables 1 and 2.

In our analysis, we verified a considerable abundance of overexpressed genes between the SG vs MDG comparisons, with 4,840 CDS at least 5-fold overexpressed at salivary glands in comparison to midguts. The overexpression of SG genes was also observed in our SG vs OVA comparison, with 4,318 overexpressed CDS in SG when compared to OVA. Interestingly, the SG CDS also produced notable differences with the down regulation of genes when compared to MDG and OVA. We observed deep differences regarding down regulated genes at SG in comparison to MDG (MDG vs SG) and to OVA (OVA vs SG) with respectively 4,333 and 4,438 CDS at least 5-fold underexpressed (Table 2). These results indicate that a more specific repertoire of genes is differentially expressed in the SG of *A. sculptum* which is understandable as the primary function of this organ is to mediate process the blood meal via the secretion of specific and diverse pool of antihemostatic and immunomodulators molecules, such as kunitz domain proteins, serpins, cystatins and others (Spreadsheet S1). This role has been previously described in several other tick sialomes (Kotál et al., 2015; Chmelar et al., 2012; Oliveira et al., 2011; Francischetti et al., 2009).

Regarding the analysis of differences between the gene expression profiles from midguts and ovaries the observed differences were not as drastic. When using the same markers of a 5 fold differences of expression, we observed that there was only a difference of 1,085 MDG CDS and 464 OVA CDS overexpressed in the MDG vs OVA comparison (Table 2). These results suggest that MDG and OVA transcriptome datasets may have similar closer physiological conditions to each other, than those for salivary glands. Indeed, the engorgement process of *A. sculptum* females post-fecundation is characterized by later apoptotic processes of salivary glands and the simultaneous higher development of midguts and ovarian tissues (Denardi et al., 2004; Nunes et al., 2008; Caperucci et al., 2009). A more detailed analysis of the biological processes at *A. sculptum* tissues (Table 2 and Spreadsheet 2), is discussed as follows.

Table 2. Functional classification and distribution of the CDS at least 5 fold-more abundant among *A. sculptum* internal organs.

FUNCTIONAL CLASS	MDG vs SG (5X)	SG vs MDG (5X)	MDG vs OVA (5X)	OVA vs MDG (5X)	OVA vs SG (5X)	SG vs OVA (5X)
HOUSEKEEPING	**3,422**	**4,356**	**793**	**261**	**3,604**	**3,880**
Cytoskeletal	144	315	24	7	152	290
Digestive enzymes	210	116	7	16	263	109
Extracellular matrix	147	178	22	47	150	158
Immunity	38	2	3	1	39	2
Amino acid metabolism	93	85	38	6	80	73
Carbohydrate metabolism	147	133	39	8	145	113
Energy metabolism	166	113	48	8	149	101
Intermediate metabolism	258	215	98	26	237	188
Lipid metabolism	137	127	25	4	141	116
Nucleotide metabolism	61	40	15	3	62	37
Nuclear export	14	52	7	1	13	47
Nuclear regulation	171	326	48	10	207	290
Detoxification	184	70	14	13	203	59
Proteasome machinery	99	188	6	6	135	169
Protein export machinery	91	154	11	5	97	139
Protein modification machinery	166	206	39	20	161	181
Protein synthesis machinery	258	173	72	19	258	119
Signal transduction	326	755	55	20	375	694
Storage	36	4	15	0	31	4
Transcription factor	79	195	10	3	111	172
Transcription machinery	186	398	43	12	205	329
Transporters	228	325	68	13	220	312
Unknown, conserved	183	186	86	13	170	178
SECRETED	**564**	**355**	**55**	**173**	**679**	**316**
TRANSPOSABLE ELEMENTS	**35**	**57**	**24**	**0**	**15**	**57**
UNKNOWN	**308**	**70**	**212**	**28**	**133**	**64**
VIRAL	**4**	**2**	**1**	**2**	**7**	**1**
TOTAL	**4,333**	**4,840**	**1,085**	**464**	**4,438**	**4,318**

3.2.5. Comparative analysis of the expression of housekeeping genes among the internal organs

3.2.5.1. Nuclear regulation and transcription factors families

This functional class represents 1,427 CDS annotated as nuclear regulation machinery (Spreadsheet S1). We have identified 36 functional families of proteins related to histones, histone modification enzymes, chromatin condensation control, replication machinery, among others, as annotated at the Spreadsheet S1. The most abundant proteins were those related to helicases (197), DNA polymerases subunits (188), Repair machinery (116), Topoisomerases (66) and genes related to chromatin structural proteins (66).

Making the comparison of the CDS RPKMs among the internal organs (Spreadsheet S2), we observed deeper differences regarding nuclear regulation machinery between MDG vs SG and OVA vs SG (Table 2). Regarding the MDG vs SG comparison, we identified 171 nuclear regulation CDS at least 5-fold overexpressed at MDG in comparison to SG, of which stand out several coding sequences related to DNA polymerases (27 CDS), DNA helicases (19 CDS), Chromatin structural proteins (11 CDS), DNA binding proteins and topoisomerases (8). We have observed this same pattern at the comparison between OVA vs SG, in which several coding related to DNA polymerases, DNA helicases and chromatin structural proteins were at least 5-fold overexpressed at OVA in comparison to SG. The complete list of the overexpressed nuclear regulation gene families is shown in Spreadsheet S2. Although, the MDG and OVA datasets represent organs collected at 72-96 hours and 8-10 days, respectively, and the SG dataset refers to tick fed during 72-96 hours, the overexpression of replication machinery at OVA and MDG transcriptomes could be related with later development of these organs observed during the engorgement and simultaneous salivary glands degradation, as previously described (Denardi et al., 2004; Nunes et al., 2008; Caperucci et al., 2009). In addition, the 5-fold overexpression of 17 CDS related to DNA repair machinery at MDG in comparison to SG is suggestive of the activation of protection mechanisms in *A. sculptum* midguts against the damage of oxidative molecules from the immune response present in the blood of the vertebrate host (Sojka et al., 2013; Chmelar et al., 2015).

There is also the overexpression of several transcripts at related to the control of histones, and their modification enzymes, such as histone acetylases, histones deacetylases, histones demethylases and histone methylases in the MDG and OVA transcriptomes when compared to the SG dataset. The higher expression of those coding sequences could be related with a higher activity of gene expression control at the level of transcription, in both the MDG and OVA datasets when compared to SG, which could be corroborated by the predominance of de-condensed chromatin previously observed in the midgut cells from *A. sculptum* females during the engorgement processes (Caperucci et al., 2010).

We have also annotated 783 coding sequences of transcription factosr from the total transcriptome of the three *A. sculptum* internal organs (Table 1), among of which were identified 59 different families, including Zinc finger, the C2H2 family, Bzip and Forkhead (Spreadsheet S1). We have identified 79 transcription factors that are 5 fold-overexpressed in the MDG CDS when compared to the SG data, of which there are 24 genes related to zinc finger family, 9 related to Forkhead family, 4 Leucine-rich repeat and 3 to bZIP (Spreadsheet S2).

Among those genes, it is worthy to note that transcription factors related to the Forkhead family have been previously described to drive gut development in *Drosophila* (Hoch & Pankratz, 1996) which corroborates with the expressive growth of the midgut tract of *A. sculptum* females during the engorgement processes (Caperucci et al., 2010). Thus, these results support the hypothesis that Forkhead transcription factors could play a role at midgut growth during the engorgement processes of *A. sculptum* females.

Regarding the comparison between OVA and SG dataset, we have identified 111 transcriptions factors at least 5-fold overexpressed at OVA (Table 2). Among these, it is worthy to note there was one annotated gene from the OVA dataset related to GATA binding protein family. This protein family has been previously described to orchestrate signaling pathways that activates the vitellogenesis processes in *Haemaphysalis longicornis*. Several other families, such as Forkhead (9), the C2H2 family, Homeobox, and TATA binding protein are also overexepressed in the OVA CDS, which could be related with midgut tissue development and differentiation during the engorgement (Spreadsheet S2). Finally, taking into consideration another overexpressed transcriptions factors in the SG vs MDG comparison, is NF KappaB with 2 CDS. This family is related to the activation of signaling pathways for immune innate response in *Drosophila* (Hetru & Hoffmann, 2009). It could tentatively be linked with the function of tick salivary glands as they encounter and engage a wide range of pathogens at the feeding site on vertebrate host (Francischetti et al., 2009). In addition, NFKappaB has also been implicated in the activation of apoptotic pathways in *Drosophila* (Primrose et al., 2007) with could be potentially involved with salivary glands degeneration at the later phase of engorgement (Nunes et al., 2008).

3.2.5.2. Protein synthesis and modification machinery

During our analysis of the entirety of the transcriptomes of the three *A. scultpum* internal organs, we identified total of 1,210 CDS related to protein synthesis machinery, of which the majority (1,158 CDS) were expressed in MDG. We also identified 895 and 533 CDS from the OVA and SG datasets, respectively (Table 1). These CDS were functionally distributed into 48 families, including ribosome structural proteins and aminoacyl-tRNA synthetases (Spreadsheet S1). Regarding the overexpressed genes from this class, we have identified 258 coding sequences that

were expressed in higher levels in MDG than in SG (Table 2), of which several were annotated as ribosome structural protein and aminoacyl-tRNA synthetases (Spreadsheet S2). This result is related with the great increase of the midguts during the later phase of the engorgement, with the simoutaneuous degradation of the salivary glands, as previously demonstrated (Caperucci et al., 2010).

In addition, we also identified and annotated 1,043 coding sequences as belonging to protein modification machinery, of which 963 were expressed in the MDG, 800 CDS in the OVA and 458 CDS at the SG (Table 1). This class covers gene families related to chaperones and chaperonins, as well as, those related to modification-proteins enzymes, such as signal peptidases, phosphatases, glucosilases, proteases and others (Spreadsheet S1).

In making the comparison analysis between the MDG and SG transcriptomes, we identified 166 coding sequences from MDG (Spreadsheet S2). A significant number of the CDS have been identified as modification enzymes, such as Proteases, with 30 overexpressed genes, followed by Peptidyl-prolyl cis-trans isomerases (9), ATPases (10), methyltransferases (5), alpha macroglobulin (2) and signal peptidases (1) (Spreadsheet S1). There are several chaperones, such as DnaJ (18), Chaperonins (12), ATPase (10), Hsp70 (4), Hsp90 (3), and DnaK (2) (Spreadsheet S2). The overexpression of these protein families could be related with post-translational processing and secretion of digestive enzymes during the digestion of host blood proteins at midguts (Sojka et al., 2014). Our comparison analysis also identified several protein modification CDS in the SG that were overexpressing in comparison with MDG and OVA. These SG specific overregulated genes include signal peptidases (2 CDS), DnaJ (20 CDS) Hsp70 (14) and 35 proteases (Spreadsheet S2). Besides their suggested role in the midguts, these proteins families could also be involved with post-translational processing of secreted proteins of tick saliva, as previously described in other studies (Schwarz et al., 2014; Chmelar et al., 2015). We also observed that 5 OVA overexpressed proteases during the OVA vs SG which is suggestive of their involvement in the proteolysis step of the vitellogenin maturation at *A. sculptum* ovaries, similarly as previously described for *Apis mielifera* (Havukainen et al., 2012).

3.2.5.3. Protein export machinery

We identified 569 coding sequences from the transcriptomes related to protein export machinery. These were further subcategorized into 24 functional families (Spreadsheet S1). Of these gene sequences there are 20 CDS related to N-ethylmalemine-sensitive factor attachment proteins (SNAPs) and their receptor proteins (SNARE). The SNAP and SNARE proteins have previously been described to drive the fusion of secretory vesicles with the membrane in *A. maculatum* and *A. americanum* salivary glands. Knockdown and RNAi approaches of the related

genes in those ixodidae ticks resulted in the decrease of the engorged tick weight and impaired oviposition along with a lack of production of eggs, suggesting that SNAREs play a key role at vitellogenesis in the ovaries (Villareal et al., 2013). During our MDG and SG transcriptome analysis, we also identified four coding sequences related to SNARE overexpressed in the MDG (Spreadsheet S2). This overexpression suggests that SNAREs could be specifically involved at secretion of digestive enzymes vesicles in the *A. sculptum* midguts and that SNAREs may be involved in the assimilation of nutrients in the midgut (Villareal et al., 2013). Additional over-regulated protein export machinery is shown in Spreadsheet S1. Of note is that several of the annotated genes are responsible for the orchestration of intracellular vesicle traffic including 22 types of sorting proteins, vacuolar ATPases (10), vesicle coat complex proteins (2), Golgi vesicle related proteins (5), Clathrins (7), Golgi vesicle related proteins (7) and SNARE proteins (4). Some of these proteins families, such as clathrins, have been described to play role on the endocytosis processes at ticks, including rickettsial infection mechanism, as well (Chang et al., 2009). These annotated proteins families could be related to the previous histological observations of lipids droplets and protein vesicles in *A. sculptum* females midguts (Caperucci et al., 2010).

Our analysis also identified during the OVA vs SG analysis, 11 overexpressed OVA proteins belonging to the SEC family (Spreadsheet S2). Proteins belonging to this family, in particular, SEC 5, have been described to play role on the trafficking of endocytic vesicles of oocytes from *Drosophila melanogaster* (Sommer et al., 2005). Thus, this overexpression could indicate a similar role during vitellogenesis processes *A. sculptum* ovaries.

3.2.5.4. Proteasome machinery

A total of 663 coding sequences from the whole *A. sculptum* transcriptome were related to proteasome machinery, including 302 CDS annotated as E2 ubiquitin ligases (Spreadsheet S1). Among a variety of functions, these enzymes play role as a marker of target proteins for degradation at the proteasome (d'Azzo et al., 2005). We observed overexpression of these enzymes in the MDG during the MDG vs SG with 43 coding sequences as well with 61 ubiquitin ligases as in the OVA dataset during the OVA vs SG library comparison. However, we verified a higher number of overexpressed E2 ubiquitin ligases from the SG dataset, with 90 CDS during the SG vs MDG and 81 SG derived CDS in the SG vs OVA comparison (Spreadsheet S2). This higher abundance of E2 ubiquitin ligases at salivary glands was expected as this organ begins to undergo degradation during the second half of the engorgement processes as the midgut and ovaries further develop (Denardi et al., 2004; Nunes et al., 2008; Caperucci et al., 2009). Thus, these results suggest a molecular explanation for the turnover of nutrients from salivary glands to the development of midguts and ovaries at the later phase of the engorgement processes.

3.2.5.5. Cytoskeletal machinery

We annotated 1,265 CDS related to cytoskeletal proteins, of which 1,178 CDS were expressed in MDG, 1,107 CDS in OVA and 669 CDS in the SG libraries (Table 1). This cytoskeletal related dataset was subcategorized into 42 functional families, as listed at the Spreadsheet S1. In our comparative analysis between MDG and SG libraries, we observed 144 coding sequences related to cytoskeletal machinery that were overexpressed in the MDG library in comparison to SG library (Spreadsheet S2). Sixteen CDS are related to cell cycle control which could be related with the great development observed in the midguts during the engorgement, once they could be related with mitotic processes (Caperucci et al., 2010). Analysis of the OVA and SG datasets allowed the identification of 152 cytoskeletal genes overexpressed in OVA in comparison to SG with 10% of the genes (15) related to cell cycle control. As the ovaries were collected from *A. sculptum* engorged females at 8-10 days of blood feeding, the abundance of transcripts related to cell cycle proteins could be justified by the ovarian development during the engorgement and vitellogenesis processes (Dernardi et al., 2004).

3.2.5.6 . Digestive enzymes

A total of 767 CDS were annotated as digestive enzymes (Table 1). This functional class were subdivided as 177 CDS found to be related to lipases and esterases; 443 CDS annotated as digestive proteases; and 147 CDS were related to digestive enzymes ot sugar (Spreadsheet S1). In general, the number of coding sequences related to these enzymes were slightly more abundant in the *A. sculptum* OVA transcriptome in comparison to MDG and SG (Table 1), which suggests their participation in the assimilation of nutrients during the vitellogenesis processes. Indeed, Dernardi et al (2004) though histological and histochemical approaches, identified the increased amounts of lipids, proteins and carbohydrates vesicles during the vitellogenesis processes at *A. sculptum* ovaries, which is supported by our results.

A total of 11 functional families were identified from the subset of lipid enzymes. The most significant families identified were those related to phosphodiesterases (37 CDS), carboxylesterases (19 CDS), lipases (19 CDS), esterases (13 CDS) and cholesterol esterases (8 CDS). The other families are listed at the Spreadsheet S1. Unfortunately, the lipid metabolism in ticks is not well understood at this moment. However, recent transcriptomes have reported and have annotated those enzymatic families (Kotsyfakis et al., 2015). There has been a previous study based on histochemichal approaches has described the ultrastructural features of *A. sculptum* midguts during the engorgement processes and characterized the presence of lipids droplets in midguts cells during all phases of the feeding process (Caperucci et al., 2009). Our observations of 39 CDS encoding lipid enzymes overexpressed in MDG during the MDG vs SG analytical comparison could suggest

a putative role for them in the assimilation of lipids in the *A. scultum* midgut.

As for the 443 CDS annotated as digestive proteases from the combined organ transcriptome, 417 CDS were derived from MDG, 431 CDS from OVA, and 182 CDS from the SG dataset (Table 1). These CDS were functionally annotated into 32 protease families as listed at the Spreadsheet S1. Further analysis of the genes during the MDG vs SG comparative analysis identified overexpression of the following families in the MDG database: leucine aminopeptidases (3 CDS), asparaginyl peptidases (10 CDS), aspartyl proteases (7 CDS), cathepsin B (1 CDS), cathepsin C (1 CDS), cathepsin D (1 CDS), legumains (2 CDS) and serine proteases (31 CDS). (Spreadsheet S2). These proteins have been described to play role on heme digestion processes in other ixodidic ticks midguts (Boldbaatar et al., 2006; Sojka et al., 2007; Sojka et al., 2013). Thus, these results suggest these proteins could also play role at digestive processes of blood proteins at *A. sculptum* midguts.

Another functional class related to digestive enzymes was that of carbohydrate digestion, with a total of 147 CDS annotated for the whole *A. sculptum* transcriptome (Spreadsheet S1). We annotated and identified several sugar enzymes families including glucosidases, amylases, fucosyltransferases, mannosidases, maltases, oligosaccharyltransferases, melibiases, and fucosidases (Spreadsheet S1). In the RPKMs values comparison between OVA and SG, theCDS sequences that could be assigned to one of the families. We identified fucosyltransferases (7 CDS), glucosidases (6 CDS), fucosidases (5 CDS), maltases (5 CDS), amylases (3 CDS) among others (Spreadsheet S2). The overexpression of these enzymes from ovaries from the engorged *A. sculptum* females could be related with vitellogenesis processes at this organ, especially when one considers the previous histological observations of carbohydrates vesicles in *A. sculptum* ovaries during the vitellogenesis (Denardi et al., 2004). We observed a similar pattern in the MDG vs SG comparison with MDG having 4 CDS from the fucosidases family over expressed (Spreadsheet S2). These enzymes have been previously characterized from *A. sculptum* midguts and its role in tick midgut is associated with fucose assimilation from glycoproteins and glycolipids from the blood, as well as a defense enzyme against fucosilated compound from parasites (Moreti et al., 2013).

3.2.5.7. Nutrients transporters

We annotated 1,887 coding sequences from the transcriptomes as nutrients transporters (Table 1). As expected, due the respective uptake function and vitellogenesis processes, the majority of the transcripts related to this class were derived from MDG (1,755 CDS), followed by OVA (1,392 CDS) and SG dataset (702 CDS). We identified 42 transporters families that were further classified as ion transporters (307 CDS), ABC superfamily (302 CDS), amino acid and peptide transporters (195 CDS), sodium symporters (193 CDS), metabolites transporters (192 CDS),

sugar/nucleotide transporters (108 CDS), and lipids transporters (65 CDS). These and other transporter families are listed at Spreadsheet S1.

In our comparative analysis between the MDG and SG datasets, we identified several transporter coding sequences overexpressed at MDG (Spreadsheet S2) including metabolites transporters (31 CDS), ion transporters (29 CDS), amino acid and peptides transporters (24 CDS), sugar/nucleotide transporters (15 CDS) and lipid transporters (10 CDS). In association with the previous results for digestive enzymes and protein associated with vesicles (Spreadsheet S2) this abundance of nutrient transporters families, especially those related with sugars, amino acids and lipids, corroborates the key role of *A. sculptum* midguts regarding nutrient uptake and these transporters could be explored as molecular aims for the development of A. sculptum specific acaricides. Indeed, other studies have discussed that tick midguts play role as storage and nutrients reserve through the intracellular vesicles of lipids and carbohydrates (Balashov, 1967; Diel et al., 1970). Until now those studies for *A. sculptum* were histology based approaches (Caperucci et al., 2009).

A similar pattern was observed during our OVA vs SG analysis, with metabolite transporters, amino acid and peptide transporters, sugar and nucleotide transporters and lipid transporters among those overexpressed in OVA CDS (Spreadsheet S2). The abundance of these transporters at *A. sculptum* ovaries suggests a role regarding the presence of carbohydrates, proteins and lipids vesicles observed during the vitellogenesis processes and yolk formation in this tick as previously reported (Denardi et al., 2004).

3.2.5.8. Carbohydrates/Energy metabolism

We have identified and annotated 899 CDS as belonging to carbohydrate metabolism, with several being identified as enzymes from catabolic pathways (Spreadsheet S1). As expected, we have observed several from these CDS overexpressed in the MDG dataset comparison to SG, especially glycolysis and gluconeogenesis related (25), tricarboxilic acid cycle (20), amino sugar metabolism (16), chitinase related (14), galactosidases (14), pentose phosphate pathway (10) and glucogen metabolism (4). These results confirms the increased abundance of transcripts related to sugar transporters in the MDG vs SG comparison (previously discussed) which suggest a higher activity of carbohydrates catabolism at MDG. Similarly, we observed 149 coding sequences from OVA overexpressed in comparison to the SG dataset, which highlights the higher nutritional demand of this organ during the oogenesis / vitellogenesis processes. Indeed, the previous observation of the sugar vesicles in midguts (Denardi et al., 2004) and ovaries of engorged females of *A. sculptum* (Caperucci et al., 2009; 2010) corroborates these results.

We also observed an overexpression of transcripts related with chitinases (53 CDS). These

are hydrolases that break glycosidic bonds in chitin, an important step in the assimilation/reabsorption processes in arthropods (Merzendorfer, 2013). Other relevant carbohydrate metabolism processes discovered by transcriptomic data analysis are those related to lactic fermentation (5 CDS), Xylose metabolism (6 CDS), Inositol metabolism (4 CDS) and fucosyltransferases (2 CDS), as shown at the Spreadsheet S1.

3.2.5.9. Lipids Metabolism

We have identified 803 CDS related with lipid metabolism, of which 751 were derived from *A. sculptum* midgut transcriptomes, while 636 CDS were derived from ovaries and 272 CDS from salivary glands (Table 1). Although lipid digestion by ticks is not well understood at this moment (Kotsyfakis et al., 2015), the higher expression of transcripts related to this class at MGD suggests the lipid assimilation could potentially be a nutrient source for *A. sculptum* during the blood meal.

3.2.5.10. Amino acids and nucleotide metabolism

We identified 604 CDS related to amino acids metabolism, of which 580 CDS derived from the *A. sculptum* mialome (MDG), 388 CDS from the ovaries transcriptome (OVA) and 200 CDS derived from salivary glands transcriptome (SG). Analysis of the nucleotide metabolism group indicated that a total of 297 CDS in the transcriptome of which 291 were derived from the *A. sculptum* midgut transcriptome (MDG), 182 CDS from ovaries (OVA) and 96 CDS derived from salivary gland transcriptome (Table 1 and Spreadsheet S1). These functional classes have described in previous tick transcriptomes (Kotsyfakis et al 2015, Chmelar et al., 2015, Garcia et al., 2014) and the higher number of transcripts related to these classes at MDG dataset could be related with the metabolism of proteins provided from the blood meal at tick midguts.

3.2.5.11. Detoxification

Enzymes related to detoxification processes have been described as essential in the tick defense and oxidative protection against inflammatory host response (Duscher et al., 2014; Pichu et al., 2011). Among this repertoire of detoxification enzymes, one of the key enzymes is glutathione S-transferase, which catalyzes the addition of a reductive glutathione group to toxic compounds for the tick, such as peroxide molecules yielded by intracellular pathogens and nitric oxide or peroxinitrite derived from vertebrate host inflammatory response (Duscher et al., 2014).

In our analysis of the entire transcriptome, we observed 640 CDS related to detoxification processes, of which 614 CDS were derived from MDG, 514 CDS from OVA and 213 CDS from SG dataset (Spreadsheet S1). Our analysis further classified into gene families including dehydrogenases (77 CDS), multidrug resistance related proteins (66 CDS), oxidases (50 CDS), peroxidases (41 CDS), superoxide dismutases (17 CDS), cytochrome oxidases (8 CDS), and 8

peroxidases (Spreadsheet S1). We identified 56 coding sequences related to Glutathione S-transferases in the MDG comparison to SG. These enzymes were found to have reductive activity on reactive molecules derived from intracellular pathogens or vertebrate host responses (Vuilleumier & Pagani, 2002). Another important enzyme overexpressed in MDG was the sulfotransferase family as represented by 153 coding sequences overexpressed (Spreadsheet S2). This enzyme catalyzes the transference of a sulfate group to hydroxyl or amine group, which inhibits toxic molecules. This sulfotransferase inhibitory activity has been described for *I. scapularis* enzymes from midguts, which can inhibit vertebrate host molecules, such dopamine and octopamine (Pichu et al., 2015).

3.2.5.12. Signal transduction

Signal transduction represented the functional class with the highest number of related CDS in the entire three organ transcriptome of *A. sculptum*, with 3,158 coding sequences annotated. A total of 2,919 signal transduction coding sequences were derived from the MDG transcriptome, while 2,525 were from the OVA and 1,477 CDS from the SG dataset (Table 1). Transcripts related to this class were functionally subcategorized into 49 families according the relevant signaling pathway (Spreadsheet S1 and Spreadsheet S3). Among these families one of the most important are the genes related to the MAP kinase signaling pathway. The MDG vs SG comparison identified 33 related coding sequences overexpressed in MDG. This pathway has been previously described in *Drosophila* as conducting the insulin-dependent glucose assimilation using the transcription factor ETS-1 and a mechanism in the process (Zhang et al., 2011). Analysis of our data has led us to propose a similar mechanism a similar mechanism for regulation of the glucose assimilation in *A. sculptum*.

Several additional other signal transduction pathways and CDS were identified in the comparison with Ras signaling pathway related (21 CDS), Calcium dependent signaling proteins (26), cAMP signaling related (9), Wnt signaling pathway (11), Hippo signaling pathway related (9) and JAK/STAT signaling pathway (11) being the most significant. Other CDS and signaling pathways are listed at Spreadsheet S2.

3.2.5.13. Immunity related products

During the blood meal and engorgement process, tick midguts have to deal with a wide range of stressors such as oxidative molecules from vertebrate host immune response, as well as external pathogens (Soneshine & Hynes, 2008; Narasimhan et al., 2013; Pichu et al., 2015). We were able to identify and annotate 119 CDS for immunity related products and their subfamilies (Table 1 and Spreadsheet S1). The transcripts to the immunological response were only expressed

in the *A. sculptum* midguts (MDG) and ovaries (OVA). From this dataset, we identified several functional families, such as those related to TNF receptors (51 CDS), penicillin binding proteins (6), immunoglobulin binding proteins (5 CDS), cytokine related proteins (2 CDS); interferon binding proteins (2 CDS). These and the other immunity related families are shown in Spreadsheet S1. It is worthy to note that we identified from this tick species Toll-like receptors (12 CDS), that can recognize pathogen associated molecular patterns or PAMPs (Hein & Ulmer, 2005; Kurata et al., 2006; Smith & Pau, 2014) The PAMPs as well as, those CDS related to immunoglobulin and interferon binding proteins play a role in neutralizing immunoglobulins and interferon from the blood of vertebrate host (Valenzuela et al., 2002; Gong et al., 2014).

3.2.6. Putative secreted proteins

Utilization of the SignalP software, allowed us to identify 2,177 coding sequences that contained signal peptides. These peptides were designated as secreted proteins (Table 1 and Spreadsheet S1). These were characterized into 23 different gene families, many of which have been previously reported by others in tick sialomes (Ribeiro et al., 2011; Garcia et al., 2014; Karim & Ribeiro, 2015). We also identified gene families such as serine proteinase inhibitors, immunomodulators, antimicrobial peptides, lipocalins, proteases in the *A. scultum* tick transcriptome (Figure 3). We also identified other secreted enzymes including those related to sulfotransferases (629), proteases, immune related proteins, lipocalins and kunitz domains. A total of 709 putative proteins (with signal peptide detected) did not present similarity to other known proteins from the databases evaluated in this study, and they were classified as specific *A. sculptum* secreted proteins (Spreadsheet S1 and S3).

Despite of these gene families having been previously identified in other tick mialomes, our comparative analysis data suggest different functions than what was reported. For example, we have identified several CDS that were overexpressed in MDG vs SG comparison including mucins (17), glycine-rich proteins, which have been described to play role on the cement formation during blood-feeding and perithrophic matrix formation, with 15 CDS (Maruyama et al., 2010), secreted lipases and esterases (10), lipocalins (18), peritrophins (8), serpins (11), perilipins (10), and Kunitz containning domains (16) among others (Spreadsheet S1). The overexpression of these “salivary” proteins families at midguts also has been reported for *Ixodes ricinus* mialomes and transcriptomes (Kotsyfakis et al., 2015) which suggests they could to play novel roles in the midgut physiology, which are ucovered at this moment. Extending this hypothesis, we also observed more abundance of these proteins in our OVA libraries when compared to SG datasets (Spreadsheet S1). In this case, the overexpression of lipases and esterase could be to be related to the increase of lipids vesicles during the yolk formation (Dernardi et al., 2004).

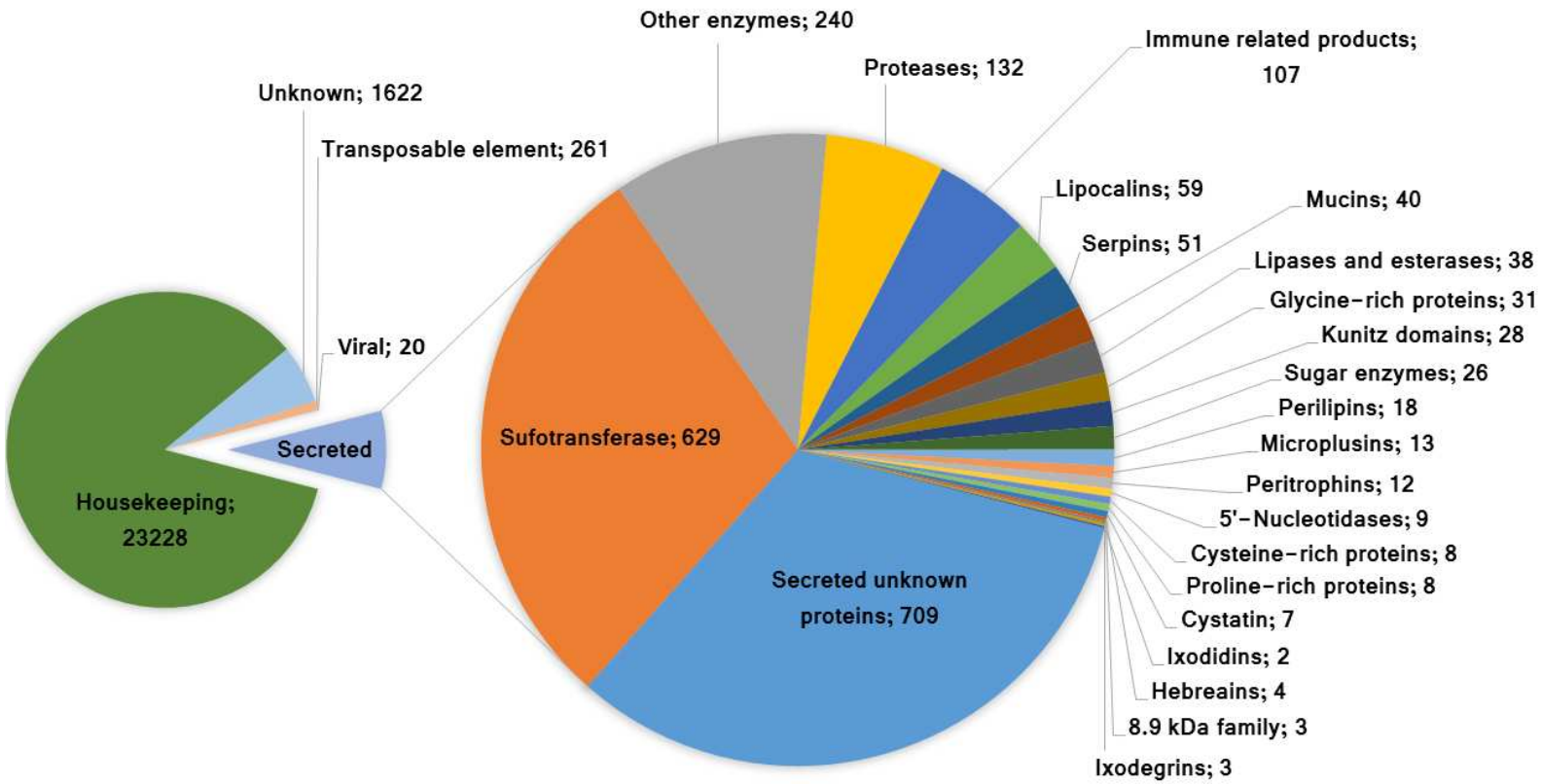

Figure 3. Functional classification of the 2,177 coding sequences related to secreted proteins, in the context of the whole transcriptome of *A. sculptum* internal organs (midguts, ovaries and salivary glands).

3.2.7. Transposable elements families

We have identified 261 coding sequences related to transposable elements, which indicates the strong possibility of transposition and recombination events, as observed with *Drosophila melanogaster* (Ecovoiu et al., 2016). Transcripts related to this functional class also have been reported from other tick transcriptomes projects (Kotsyfakis et al., 2015). Among the transposable elements annotated in our *A. sculptum* dataset, were those related to Maverick, Harbinger, Pogo and Pol polyprotein (Spreadsheet S1). In the SG comparison to both MDG and OVA, we noted an overexpression of transcripts related to Gag, Pogo, Pol and Tigger. We suggest that the presence of transposases can be associated with viral material into the genome (Anderson et al., 2016). The fact that there is greater abundance of these transcripts at SG could be explained due the increased exposure to viruses at feeding site on vertebrate hosts.

3.2.8. Unknown fragments and viral related proteins

A total of 1,603 CDS remains without homology to any deposited sequences, although they have been identified as coding sequences using the Augustus algorithm (Table 1), which suggests genes related to novel processes and could be an interesting source of information to be explored in further studies. We also have identified and annotated a smaller class of transcripts, with 20 CDS that presented homology to viral sequences. This class of viral transcripts also has been reported in several ixodidic ticks transcriptomes which suggests a putative viral insertion in tick genomes (Garcia et al., 2014; Kotsyfakis et al., 2015). However, this issue needs to be further studied.

4. CONCLUSION

In this study, we assembled 9 transcriptomes (200 million reads) derived from the main internal organs of *A. sculptum females* using with gene prediction based on the Augustus algorithm. This methodology allowed us to achieve a higher number of CDS for *A. sculptum*, than previously reported, which indicates the suitability of the assembly workflow application for transcriptome analysis. We identified and annotated a large number of transcripts from the MDG (25,569 CDS), OVA (21,230) and SG (10,697 CDS) transcriptome. Although we have identified classical proteins families from sialomes at the SG dataset (lipocalins, kunitz, serpins), our results indicate that those proteins families also play a role in the midguts and ovaries from *A. sculptum*. This has also been observed for the *I. ricinus* tick (Kotsyfakis et al., 2015). The *A. sculpum* MDG dataset revealed several CDS associated with digestive enzymes, especially proteases previously described to found to play a role in protein digestion with other ticks (leucine aminopeptidases, aspartyl proteases, cathepsin B, cathepsin L and others). SNARE proteins, which have been previously described in other *Amblyomma* ticks, are also expressed at the three organs of *A. sculptum*, indicating they play

role during housekeeping processes that are dependent on vesicle transport. We found that several cellular processes and overexpressed at least 5-fold over background in each of the studied transcriptomes. For example, the overexpression of detoxification processes, especially glutathione S- transferase in MDG may be related with the possible presence of pathogens as well as oxidative molecules from the host in the midgut of the tick. In general, we observed that the SG dataset contained the highest number of overexpressed CDS in comparison to both MDG and OVA which can be attributed to its specialized role of mediating the blood meal. The MDG and OVA datasets contained a lower number of overexpressed genes when compared against each other, which is due these organs represent the latter stage of the engorgement processes. The annotation approach employed in this work, along with our databases, allowed for the coverage of a higher number of genes and achieved a lower number of unknown sequences. We conclude that this approach was suitable to allow extended coverage a facilitated the identification and annotation of novel genes and processes from the *A. sculptum* internal organs.

5. ACKOWNLEDGEMENTS

We would like to thank the Foundation for Research Support of the State of Minas Gerais (FAPEMIG) for financial support; Dr. Marcelo B. Labruna, Faculty of Veterinary Medicine, University of São Paulo, for providing the inoculum of *Rickettsia amblyommii;* Dr. Pablo Juan Szabó Mathias, Laboratory of Ixodology Federal University of Uberlandia, for providing the specific pathogen free *Amblyomma sculptum* specimens. We would like also to thank Dr. Donald Bouyer, from University of Texas Medical Branch, for the editorial assistance for this paper.

6. REFERENCES

ALONSO-DÍAZ, M.A., FERNÁNDEZ-SALAS, A., MARTÍNEZ-IBÁÑEZ, F., OSORIO-MIRANDA, J. 2013. *Amblyomma cajennense* (Acari: Ixodidae) tick populations susceptible or resistant to acaricides in the Mexican Tropics. Vet. Parasitol. 197, 326–31.

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W., LIPMAN, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–10.

ANATRIELLO, E., RIBEIRO, J.M.C., DE MIRANDA-SANTOS, I.K.F., BRANDÃO, L.G., ANDERSON, J.M., VALENZUELA, J.G., MARUYAMA, S.R., SILVA, J.S., FERREIRA, B.R. 2010. An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. BMC Genomics 11, 450.

ANDERSON, J.M., SONENSHINE, D.E., VALENZUELA, J.G. 2008. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick, *Dermacentor variabilis* (Acari: Ixodidae). BMC Genomics 9, 552.

BALASHOV, Y.S., RAIKHEL, A.S., HOOGSTRAAL, H. 1983. An Atlas of Ixodid Tick Ultrastructure. College Park, Md.: Entomological Society of America.

BATISTA, I.F.C., CHUDZINSKI-TAVASSI, A.M., FARIA, F., SIMONS, S.M., BARROS-BATESTTI, D.M., LABRUNA, M.B., LEÃO, L.I., HO, P.L., JUNQUEIRA-DE-AZEVEDO,

I.L.M. 2008. Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acari: Ixodidae). Toxicon 51, 823–34.

BEATI, L., NAVA, S., BURKMAN, E.J., BARROS-BATTESTI, D.M., LABRUNA, M.B., GUGLIELMONE, A.A., CÁCERES, A.G., GUZMÁN-CORNEJO, C.M., LEÓN, R., DURDEN, L.A. FACCINI, J.L.H. 2013. *Amblyomma cajennense* (Fabricius, 1787) (Acari: Ixodidae), the Cayenne tick: phylogeography and evidence for allopatric speciation. BMC Evol. Biol. 13, 267.

BECK, D.L., ZAVALA, J., MONTALVO, E.O., QUINTANA, F.G. 2011. Meteorological indicators for *Amblyomma cajennense* and population dynamics in the Tamaulipan Biotic Province in Texas. J. Vector Ecol. 36, 135–46.

BILLINGS, A.N., RAWLINGS, J.A., WALKER, D.H. 1998a. Tick-borne diseases in Texas: a 10-year retrospective examination of cases. Tex. Med. 94, 66–76.

BILLINGS, A.N., YU, X.J., TEEL, P.D., WALKER, D.H. 1998b. Detection of a spotted fever group rickettsia in *Amblyomma cajennense* (Acari: Ixodidae) in south Texas. J. Med. Entomol. 35, 474–8.

BOLDBAATAR, D., SIKALIZYO SIKASUNGE, C., BATTSETSEG, B., XUAN, X., FUJISAKI, K. 2006. Molecular cloning and functional characterization of an aspartic protease from the hard tick *Haemaphysalis longicornis*. Insect Biochem. Mol. Biol. 36, 25–36.

CALVO, E., PHAM, V.M., MARINOTTI, O., ANDERSEN, J.F., RIBEIRO, J.M.C. 2009. The salivary gland transcriptome of the neotropical malaria vector *Anopheles darlingi* reveals accelerated evolution of genes relevant to hematophagy. BMC Genomics 10, 57.

CAPERUCCI, D., BECHARA, G.H., CAMARGO MATHIAS, M.I. 2010. Ultrastructure features of the midgut of the female adult *Amblyomma cajennense* ticks Fabricius, 1787 (Acari: Ixodidae) in several feeding stages and subjected to three infestations. Micron, 41(7), 710-21.

CAPERUCCI, D., CAMARGO MATHIAS, M.I., BECHARA, G.H. 2009. Histopathology and ultrastructure features of the midgut of adult females of the tick *Amblyomma cajennense* Fabricius, 1787 (Acari: Ixodidae) in various feeding stages and submitted to three infestations. Ultrastructural Pathology, 33(6), 249-59.

CHAN, Y.G.Y., CARDWELL, M.M., HERMANAS, T.M., UCHIYAMA, T., MARTINEZ, J.J. 2009. Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner. Cell. Microbiol. 11, 629–44.

CHMELAŘ, J., KOTÁL, J., KARIM, S., KOPACEK, P., FRANCISCHETTI, I.M., PEDRA, J.H., KOTSYFAKIS, M. 2015. Sialomes and Mialomes: A systems-biology view of tick tissues and tick-host interactions. Trends in Parasitology, 32(3), 242-54.

CHMELAR, J., CALVO, E., PEDRA, J.H., FRANCISCHETTI, I.M., KOTSYFAKIS, M., 2012. Tick salivary secretion as a source of antihemostatics. Journal of Proteomics, 75(13), 3842-54.

CHUDZINSKI-TAVASSI, A.M., MORAIS, K.L. PACHECO, M.T., PASQUALOTO, K.F., DE SOUZA, J.G. 2016. Tick salivary gland as potential natural source for the discovery of promising antitumor drug candidates. Biomedicine & Pharmacotherapy. 77, 14-9.

D'AZZO, A., BONGIOVANNI, A., NASTASI, T. 2005. E3 ubiquitin ligases as regulators of membrane protein trafficking and degradation. Traffic. 6(6), 429-41.

DENARDI, S.E., BECHARA, G.H., CAMARGO-MATHIAS, M.I. 2009. Fat body cells of *Amblyomma cajennense* partially engorged females (Acari: Ixodidae) and their role on vitellogenesis process. Experimental Parasitology. 121(3), 213-8.

DENARDI, S.E., BECHARA, G.H., OLIVEIRA, P.R., NUNES, E.T., SAITO, K.C., MATHIAS, M.I. 2004. Morphological characterization of the ovary and vitellogenesis dynamics in the tick *Amblyomma cajennense* (Acari: Ixodidae). Veterinary Parasitology. 125(3-4), 379-95.

DIETRICH, M., GÓMEZ-DÍAZ, E., MCCOY, K.D. 2011. Worldwide distribution and diversity of seabird ticks: implications for the ecology and epidemiology of tick-borne pathogens. Vector Borne Zoonotic Dis. 11, 453–70.

DILLIES, M.-A., RAU, A., AUBERT, J., HENNEQUET-ANTIER, C., JEANMOUGIN, M., SERVANT, N., KEIME, C., MAROT, G., CASTEL, D., ESTELLE, J., GUERNEC, G., JAGLA, B., JOUNEAU, L., LALOË, D., LE GALL, C., SCHAËFFER, B., LE CROM, S., GUEDJ, M., JAFFRÉZIC, F. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. 14, 671–83.

DUSCHER, G.G., GALINDO, R.C., TICHY, A., HUMMEL, K., KOCAN, K.M., DE LA FUENTE, J. 2014. Glutathione S-transferase affects permethrin detoxification in the brown dog tick, *Rhipicephalus sanguineus*. Ticks and Tick Borne Diseases. 5(3), 225-33.

ECOVOIU, A.A., GHIONOIU, I.C., CIUCA, A.M., RATIU, A.C. 2016. Genome ARTIST: a robust, high-accuracy aligner tool for mapping transposon insertions and self-insertions. Mob DNA. 7, 3.

EDLOW, J.A. 2008. Preface: tick-borne diseases, part II. Infectious Disease Clinics of North America, Sep,22(3):xiii-xv.

FERNANDES, F.F., FREITAS, E.P.S., COSTA, A.C., SILVA, I.G. 2015. Larvicidal potential of Sapindus saponaria to control the cattle tick Boophilus microplus. Pesquisa Agropecuária Brasileira, Dec:40(12) Brasília.

FRANCISCHETTI, I.M., SA-NUNES, A., MANS, B.J., SANTOS, I.M., RIBEIRO, J.M. 2009. The role of saliva in tick feeding. Front Biosci. 14, 2051-88.

FREY, K.G., HERRERA-GALEANO, J.E., REDDEN, C.L., LUU, T. V, SERVETAS, S.L., MATECZUN, A.J., MOKASHI, V.P., BISHOP-LILLY, K.A. 2014. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics. 15, 96.

GARCIA, G.R., GARDINASSI, L.G., RIBEIRO, J.M., ANATRIELLO, E., FERREIRA, B.R., MOREIRA, H.N., MAFRA, C., MARTINS, M.M., SZABÓ, M.P., DE MIRANDA-SANTOS, I.K., MARUYAMA, S.R. 2014. The sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. Parasites & Vectors. 7, 430.

GONG, H., QIN, S., WAN, X., ZHANG, H., ZHOU, Y., CAO, J., XUAN, X., SUZUKI, H., ZHOU, J. 2014. Immunoglobulin G binding protein (IGBP) from *Rhipicephalus haemaphysaloides*: identification, expression, and binding specificity. Parasitol. Res. 113, 4387–95.

GULIA-NUSS, M., NUSS, A.B., MEYER, J.M., SONENSHINE, D.E., ROE, R.M., WATERHOUSE, R.M., SATTELLE, D.B., DE LA FUENTE, J., RIBEIRO, J.M., MEGY, K., THIMMAPURAM, J., MILLER, J.R., WALENZ, B.P., KOREN, S., HOSTETLER, J.B., THIAGARAJAN, M., JOARDAR, V.S., HANNICK, L.I., BIDWELL, S., HAMMOND, M.P., YOUNG, S., ZENG, Q., ABRUDAN, J.L., ALMEIDA, F.C., AYLLÓN, N., BHIDE, K., BISSINGER, B.W., BONZON-KULICHENKO, E., BUCKINGHAM, S.D., CAFFREY, D.R., CAIMANO, M.J., CROSET, V., DRISCOLL, T., GILBERT, D., GILLESPIE, J.J., GIRALDO-CALDERÓN, G.I., GRABOWSKI, J.M., JIANG, D., KHALIL, S.M.S., KIM, D., KOCAN, K.M., KOČI, J., KUHN, R.J., KURTTI, T.J., LEES, K., LANG, E.G., KENNEDY, R.C., KWON, H., PERERA, R., QI, Y., RADOLF, J.D., SAKAMOTO, J.M., SÁNCHEZ-GRACIA, A., SEVERO, M.S., SILVERMAN, N., ŠIMO, L., TOJO, M., TORNADOR, C., VAN ZEE, J.P., VÁZQUEZ, J., VIEIRA, F.G., VILLAR, M., WESPISER, A.R., YANG, Y., ZHU, J., ARENSBURGER, P., PIETRANTONIO, P. V., BARKER, S.C.,

SHAO, R., ZDOBNOV, E.M., HAUSER, F., GRIMMELIKHUIJZEN, C.J.P., PARK, Y., ROZAS, J., BENTON, R., PEDRA, J.H.F., NELSON, D.R., UNGER, M.F., TUBIO, J.M.C., TU, Z., ROBERTSON, H.M., SHUMWAY, M., SUTTON, G., WORTMAN, J.R., LAWSON, D., WIKEL, S.K., NENE, V.M., FRASER, C.M., COLLINS, F.H., BIRREN, B., NELSON, K.E., CALER, E., HILL, C.A. 2016a. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. Nat. Commun. 7, 10507.

GUO, Y., RIBEIRO, J.M.C., ANDERSON, J.M., BOUR, S. 2009. dCAS: a desktop application for cDNA sequence annotation. Bioinformatics 25, 1195–6.

HATTA, T., TSUJI, N., MIYOSHI, T., ISLAM, M.K., ALIM, M.A., YAMAJI, K., ANISUZZAMAN, FUJISAKI, K. 2010. Leucine aminopeptidase, HlLAP, from the ixodid tick *Haemaphysalis longicornis*, plays vital roles in the development of oocytes. Parasitol. Int. 59, 286–9.

HAJDUŠEK, O., SÍMA, R., AYLLÓN, N., JALOVECKÁ, M., PERNER, J., DE LA FUENTE, J., KOPÁČEK, P. 2013. Interaction of the tick immune system with transmitted pathogens. Front Cell Infect Microbiol. 6, 3-26.

HATTA, T., TSUJI, N., MIYOSHI, T., ISLAM, M.K., ALIM, M.A., YAMAJI, K., ANISUZZAMAN, FUJISAKI, K. 2010. Leucine aminopeptidase, HlLAP, from the ixodid tick *Haemaphysalis longicornis*, plays vital roles in the development of oocytes. Parasitology International. 59(2), 286-9.

HAVUKAINEN, H., UNDERHAUG, J., WOLSCHIN, F., AMDAM, G., HALSKAU, Ø. 2012. A vitellogenin polyserine cleavage site: highly disordered conformation protected from proteolysis by phosphorylation. Journal of Experimental Biology. 215(Pt 11), 1837-46.

HETRU, C., HOFFMANN, J.A. 2009. NF-kappaB in the immune response of *Drosophila*. Cold Spring Harbor Perspectives in Biology. Dec,1(6), a000232.

HOCH, M., PANKRATZ, M.J. 1996. Control of gut development by fork head and cell signaling molecules in *Drosophila*. Mech. Dev. 58, 3–14.

KARIM, S., RIBEIRO, J.M. 2015. An Insight into the Sialome of the Lone Star Tick, *Amblyomma americanum*, with a Glimpse on Its Time Dependent Gene Expression. PLoS One. 10(7), e0131292.

KARIM, S., SINGH, P., RIBEIRO, J.M. 2011. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. PLoS One. 6(12), e28525.

KOTÁL, J., LANGHANSOVÁ, H., LIESKOVSKÁ, J., ANDERSEN, J.F., FRANCISCHETTI, I.M., CHAVAKIS, T., KOPECKÝ, J., PEDRA, J.H., KOTSYFAKIS, M., CHMELAŘ, J. 2015. Modulation of host immunity by tick saliva. Journal of Proteomics. 128, 58-68.

KOTSYFAKIS, M., SCHWARZ, A., ERHART, J., RIBEIRO, J.M.C. 2015. Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. Sci. Rep. 5, 9103.

KOVÁR, L. 2004. Tick saliva in anti-tick immunity and pathogen transmission. Folia Microbiol. (Praha). 49, 327–36.

KURATA, S., ARIKI, S., KAWABATA, S. 2006. Recognition of pathogens and activation of immune responses in *Drosophila* and horseshoe crab innate immunity. Immunobiology 211, 237–49.

KURTTI, T.J., MUNDERLOH, U.G., ANDREADIS, T.G., MAGNARELLI, L.A., MATHER, T.N. 1996. Tick cell culture isolation of an intracellular prokaryote from the tick *Ixodes scapularis*. J. Invertebr. Pathol. 67, 318–21.

LABRUNA, M.B., DE PAULA, C.D., LIMA, T.F., SANA, D.A., 2002a. Ticks (Acari: Ixodidae) on wild animals from the Porto-Primavera Hydroelectric power station area, Brazil. Memórias do Inst. Oswaldo Cruz 97, 1133–6.

LABRUNA, M.B., KASAI, N., FERREIRA, F., FACCINI, J.L.H., GENNARI, S.M., 2002b. Seasonal dynamics of ticks (Acari: Ixodidae) on horses in the state of São Paulo, Brazil. Vet. Parasitol. 105, 65–77.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L., LAW, M. 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 251364.

LOPES, C.M., LEITE, R.C., LABRUNA, M.B., DE OLIVEIRA, P.R., BORGES, L.M., RODRIGUES, Z.B., DE CARVALHO, H.A., DE FREITAS, C.M., VIEIRA JÚNIOR, C.R. 1998. Host specificity of *Amblyomma cajennense* (Fabricius, 1787) (Acari: Ixodidae) with comments on the drop-off rhythm. Memórias do Inst. Oswaldo Cruz 93, 347–51.

MARTINEZ, J.J., SEVEAU, S., VEIGA, E., MATSUYAMA, S., COSSART, P., 2005. Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for *Rickettsia conorii*. Cell 123, 1013–23.

MARUYAMA, S.R., ANATRIELLO, E., ANDERSON, J.M., RIBEIRO, J.M., BRANDÃO, L.G., VALENZUELA, J.G., FERREIRA, B.R., GARCIA, G.R., SZABÓ, M.P., PATEL, S., BISHOP, R., DE MIRANDA-SANTOS, I.K. 2010. The expression of genes coding for distinct types of glycine-rich proteins varies according to the biology of three metastriate ticks, *Rhipicephalus* (*Boophilus*) *microplus*, *Rhipicephalus sanguineus* and *Amblyomma cajennense*. BMC Genomics 11, 363.

MASTROPAOLO, M., NAVA, S., GUGLIELMONE, A.A., MANGOLD, A.J. 2011. Biological differences between two allopatric populations of *Amblyomma cajennense* (Acari: Ixodidae) in Argentina. Exp. Appl. Acarol. 53, 371–5.

MERZENDORFER, H. 2013. Chitin synthesis inhibitors: old molecules and new developments. Insect Sci. 20, 121–38.

MORETI, R., PERRELLA, N.N., LOPES, A.R. 2013. Carbohydrate digestion in ticks and a digestive α-L-fucosidase. J. Insect Physiol. 59, 1069–75.

MORSHED, M.G., SCOTT, J.D., FERNANDO, K., BEATI, L., MAZEROLLE, D.F., GEDDES, G., DURDEN, L.A. 2005. Migratory songbirds disperse ticks across Canada, and first isolation of the Lyme disease spirochete, *Borrelia burgdorferi*, from the avian tick, *Ixodes auritulus*. J. Parasitol. 91, 780–90.

NARASIMHAN, S., RAJEEVAN, N., LIU, L., ZHAO, Y.O., HEISIG, J., PAN, J., EPPLER-EPSTEIN, R., DEPONTE, K., FISH, D., FIKRIG, E. 2014. Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. Cell Host Microbe 15, 58–71.

NAVA, S., BEATI, L., LABRUNA, M.B., CÁCERES, A.G., MANGOLD, A.J., GUGLIELMONE, A.A. 2014. Reassessment of the taxonomic status of *Amblyomma cajennense* with the description of three new species, *Amblyomma tonelliae* n. p., *Amblyomma interandinum* n. sp. and *Amblyomma patinoi* n. sp., and reinstatement of *Amblyomma mixtum*, nd *Amblyomma sculptum* (Ixodida: Ixodidae). Ticks and Tick Borne Diseases. 5(3), 252-76.

NUNES, P.H., BECHARA, G.H., CAMARGO-MATHIAS, M.I. 2008. Morphological changes in the salivary glands of Amblyomma cajennense females (Acari: Ixodidae) in different feeding stages on rabbits at first infestation. Exp. Appl. Acarol. 45, 199–209.

OLIVEIRA, C.J.F., SÁ-NUNES, A., FRANCISCHETTI, I.M.B., CARREGARO, V., ANATRIELLO, E., SILVA, J.S., SANTOS, I.K.F. DE M., RIBEIRO, J.M.C., FERREIRA, B.R. 2011. Deconstructing tick saliva: non-protein molecules with potent immunomodulatory properties. J. Biol. Chem. 286, 10960–9.

PAL, U., LI, X., WANG, T., MONTGOMERY, R.R., RAMAMOORTHI, N., DESILVA, A.M., BAO, F., YANG, X., PYPAERT, M., PRADHAN, D., KANTOR, F.S., TELFORD, S., ANDER-

SON, J.F., FIKRIG, E., 2004. TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. Cell 119, 457–68.

PICHU, S., YALCIN, E.B., RIBEIRO, J.M., KING, R.S., MATHER, T.N. 2011. Molecular characterization of novel sulfotransferases from the tick, *Ixodes scapularis*. BMC Biochemistry. 12, 32.

PRIMROSE, D.A., CHAUDHRY, S., JOHNSON, A.G.D., HRDLICKA, A., SCHINDLER, A., TRAN, D., FOLEY, E., 2007. Interactions of DNR1 with the apoptotic machinery of *Drosophila melanogaster*. J. Cell Sci. 120, 1189–99.

RAMOS, V.N., OSAVA, C.F., PIOVEZAN, U., SZABÓ, M.P.J. 2014. Ticks on humans in the Pantanal wetlands, Brazil. Ticks Tick. Borne. Dis. 5, 497–9.

RIBEIRO, J.M., ANDERSON, J.M., MANOUKIS, N.C., MENG, Z., FRANCISCHETTI, I.M. 2011. A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. BMC Genomics 12, 136.

ROJAS, R., MARINI, M.A., ZANATTA COUTINHO, A.T. 1999. Wild birds as hosts of *Amblyomma cajennense* (Fabricius, 1787) (Acari: Ixodidae). Memórias do Inst. Oswaldo Cruz 94, 315–22.

SCHWARZ, A., TENZER, S., HACKENBERG, M., ERHART, J., GERHOLD-AY, A., MAZUR, J., KUHAREV, J., RIBEIRO, J.M.C., KOTSYFAKIS, M. 2014. A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. Mol. Cell. Proteomics 13, 2725–35.

SOARES, J.F., SOARES, H.S., BARBIERI, A.M., LABRUNA, M.B. 2012. Experimental infection of the tick *Amblyomma cajennense*, Cayenne tick, with *Rickettsia rickettsii*, the agent of Rocky Mountain spotted fever. Med. Vet. Entomol. 26, 139–51.

SOJKA, D., FRANTA, Z., HORN, M., CAFFREY, C.R., MAREŠ, M., KOPÁČEK, P. 2013. New insights into the machinery of blood digestion by ticks. Trends Parasitol. 29, 276–85.

SOJKA, D., FRANTA, Z., HORN, M., HAJDUSEK, O., CAFFREY, C.R., MARES, M., KOPÁCEK, P. 2008. Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases. Parasit. Vectors 1, 7.

SOMMER, B., OPRINS, A., RABOUILLE, C., MUNRO, S. 2005. The exocyst component Sec5 is present on endocytic vesicles in the oocyte of *Drosophila melanogaster*. J. Cell Biol. 169, 953–63.

SONENSHINE, D.E., HYNES, W.L. 2008. Molecular characterization and related aspects of the innate immune response in ticks. Front. Biosci. 13, 7046–63.

STANKE, M., KELLER, O., GUNDUZ, I., HAYES, A., WAACK, S., MORGENSTERN, B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research. 34, W435-9.

STANKE, M., MORGENSTERN, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research. 33, W465-7.

STANKE, M., STEINKAMP, R., WAACK, S., MORGENSTERN, B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Research. 32, W309-12.

TAKADA, S., COLLINS, E.R., KURAHASHI, K. 2015. The FHA domain determines Drosophila Chk2/Mnk localization to key mitotic structures and is essential for early embryonic DNA damage responses. Mol. Biol. Cell 26, 1811–28.

TAKECHI, R., GALAY, R.L., MATSUO, T., MAEDA, H., KUSAKISAKO, K., TALACTAC, M.R., MOCHIZUKI, M., FUJISAKI, K., TANAKA, T. 2016. Role of the tumor necrosis factor receptor-associated factor-type zinc finger domain containing protein 1 (TRAFD1) from the hard

tick *Haemaphysalis longicornis* in immunity against bacterial infection. Ticks Tick. Borne. Dis. 7, 36–45.

VALENZUELA, J.G., FRANCISCHETTI, I.M.B., PHAM, V.M., GARFIELD, M.K., MATHER, T.N., RIBEIRO, J.M.C. 2002. Exploring the sialome of the tick *Ixodes scapularis*. J. Exp. Biol. 205, 2843–64.

VILLARREAL, A.M., ADAMSON, S.W., BROWNING, R.E., BUDACHETRI, K., SAJID, M.S., KARIM, S. 2013. Molecular characterization and functional significance of the Vti family of SNARE proteins in tick salivary glands. Insect Biochem. Mol. Biol. 43, 483–93.

VUILLEUMIER, S., PAGNI, M. 2002. The elusive roles of bacterial glutathione S-transferases: new lessons from genomes. Appl. Microbiol. Biotechnol. 58, 138–46.

ZHANG, W., THOMPSON, B.J., HIETAKANGAS, V., COHEN, S.M. 2011. MAPK/ERK signaling regulates insulin sensitivity to control glucose metabolism in *Drosophila*. PLoS Genet. 7, e1002429. doi:10.1371/journal.pgen.1002429

CAPÍTULO 2

RNASEQ DATA ANALYSIS REVEALS DEEP CHANGES IN GENE EXPRESSION PATTERNS OF *AMBLYOMMA SCULPTUM* TICK MIDGUT IN RESPONSE TO RICKETTSIAL INFECTION

RNASeq data analysis reveals deep changes in gene expression patterns of *Amblyomma sculptum* tick midgut in response to rickettsial infection

Higo Nasser Santanna MOREIRA[1], Talles MACIEL[1], Juliana CARRIZO[1], Rafael Mazioli BARCELOS[1], Carlos MONTANDON[1], Maria Marlene MARTINS[2], Paulo LIMA[2], Adriano Soares MORENO[2], Cláudio MAFRA[1¥],

[1] Biochemistry and Molecular Biology Department, Federal University of Vicosa, Brazil; [2] Veterinary Department, Federal University of Uberlandia, Brazil;
[¥] mafra@ufv.br

ABSTRACT

Ticks are the second most important vectors group of diseases to humans, behind only the mosquitoes. Among the main species, stands out *Amblyomma sculptum*, a tick with great parasitic versatility and the main vector of a variety of *Spotted Fever Group* (SFG) *rickettsiae* in Americas. The midgut tract of ixodidic ticks play role as a primary interface with a wide range of pathogens species and constitutes the main entry path of rickettsiae to hemolymph and other internal organs. Amongst salivary glands and ovaries, midgut tract figures also as a key target for the vaccines and acaricides development. However, to spite its importance, information about the genes and proteins repertoire of the *A. sculptum* midguts remains scarce. In this context, Next Generation Sequencing (NGS) technologies have provided non-precedent possibilities to get high throughput information about genes and expanded the knowledge frontiers about tick genomics. In view of the necessity to better understand the biology of the *A. sculptum* and its interactions with SFG rickettsiae, the aim of this study was to elucidate gene repertoire of *A. sculptum* midguts, as well as, those differentially expressed front conditions of non-infection (MNI) and infection with *R. amblyommii* (MI), a SFG rickettsiae. Through Illumina Hiseq and PGM sequencing, we obtained two representative midguts transcriptomes (mialomes) for both conditions, with 49 million reads for MI and 53 million reads for MNI, with 20,013 CDS and 12,116 CDS generated from both dataset, respectively. After normalization by RPKM determination and gene expression analysis by GFOLD algorithm, we observed a great number genes more abundant in *A. sculptum* midguts in response to rickettsial infection, such as those related to gene expression machinery, protein synthesis, nutrient transporters, catabolism of carbohydrates, lipids and amino acids, digestive enzymes and others. We also observed more abundance of several proteins recognized to be involved in the infection process of rickettsiae GFM, such as actin filaments, Arp 2/3 complex, clathrins, N-WASP, a variety of kinases from signaling pathways, among others. We observed, as well, the more abundance of detoxification related proteins and antimicrobial peptides, that play as a defense response to intracellular pathogens. These are the first NGS transcriptomes obtained for *A. sculptum* midguts, main vector of FMB in Brazil. These results reinforce the hypothesis that rickettsiae GFM actively regulate the intestinal tract of its tick vector, modulating the expression of genes that facilitate the process of infection and establishment in its tick vector.

Key words: *Amblyomma sculptum*, midguts, RNASeq, mialomes, *Rickettsia amblyommii*

1. INTRODUCTION

Ticks are the second most important vectors group of diseases to humans, behind only the mosquitoes (Edlow et al., 2008; Parola & Raoult, 2001). Among the main species, stands out *Amblyomma sculptum*, a tick with great parasitic versatility and the main vector of a variety of Spotted Fever Group (SFG) *rickettsiae* in Americas (Nava et al., 2014). Among of them, stands out the Brazilian Spotted Fever (BSF) a vector-borne disease of high level of lethality (Nasser et al., 2015).

Much of the epidemiological importance of BSF lies in the biology of its tick vector, *A. sculptum*. This specie presents a wide geographical distribution, occurring in different ecosystems in a large region of the Americas, from the southern United States to northern Argentina (Billings et al., 1998; Beck et al., 2011; Mastropaolo et al., 2011; Alonso-Díaz et al., 2013; Nava et al., 2014;. Ramos et al., 2014). In Brazil, this vector occurs in almost all states, and its parasitism in humans has been reported in parks, peri-urban areas and in rural areas, usually for with local infested with larvae, contact with domestic sentinel animals or wild animals, especially capybaras (Toledo et al., 2011;. Santos et al., 2012; Krawczak et al., 2014). Because of its epidemiological importance and taxonomic issues, these species was recently renamed from *A. cajennense* to *A. sculptum* (Navas et al., 2014).

From the point of view of the internal morphology of ixodidic ticks, the midgut tract play role as the largest interface with the external environment, through which occurs the transportation of nutrients from host blood, as well as pathogens for hemolymph (Caperucci et al., 2010). Hemolymph functions as a circulatory system in ticks, communicating all other internal organs (salivary glands and the ovaries) and, thus, acting as a vehicle for establishment of SFG rickettsiae and other pathogens (Fogaça et al., 1999; Nakajima et al. 2003; Soneshine et al., 2005). In addition, tick midgut receptors have been described to develop a key role in the epidemiology of important tick-borne diseases, such as Lyme disease. In this case, midgut receptors TROSPA and Ixofin3D, from *Ixodes scapularis*, serves as an entry path for *Borrelia burgdorferi* into the hemolymph and facilitates its transmission (Narasimhan et al., 2014, Pal et al., 2004).

However, in spite of its importance, information about the genes and proteins repertoire of the *A. sculptum* midgut, especially those involved with rickettsial infection mechanisms, remains scarce. The understanding about the mechanism of rickettsial infection is a key tool to recognize molecular functions that could be explored to designing tick biological control strategies, as well as, to elucidate the dynamics of SFG rickettsiae on its environment in the vector and its implications on BSF epidemiology. In this context, Next Generation Sequencing (NGS) technologies have provided non-precedent possibilities to get high throughput information about genes and expanded the knowledge frontiers about tick genomics.

Thus, in this study we identify constitutive genes and processes, as well as, those differentially expressed in the midguts of *A. sculptum* tick under two treatments: Non-Infection (MNI) and artificial Midguts Infection (MI) with *R. amblyommii.*

2. METHODS

2.1. Ticks, experimental infection, midguts isolation and ethics

Adult female *A. sculptum* specific pathogen free were maintained according previously described (Kurtii et al.,1996). A total of 150 couples of *A. sculptum* adults were fed in rabbits by 72-96 hours and destinated to both treatments (75 Midguts Non-Infected – MNI and 75 Midguts Infected - MI). Specimens destinated to transcriptome MI were artificially infected with *R. amblyommii* activated in VERO cells, according previously described by Melles et al. (1999). Ticks were placed for 48 hours at BOD to enhance rickettsia proliferation and the checked by Real Time PCR in order to verify the presence of rickettsial DNA. In brief, a sample of ticks from infected treatments were tested by Real Time PCR assay with primers CS-5 (forward) and CS-6 (reverse), specific for a 147-bp fragment from citrate synthase gene (gltA) conserved into genus *Rickettsia* (Labruna et al., 2007). The reactions were performed for 25 µl per reaction, as follow: 12.5 uL of reaction buffer (Maxima SYBR Green/ROX qPCR Master Mix, according manufactor recommendations); 1,25 µl of each primers C5 (forward) and C6 (reverse) at 15 µM; 1,5 µl of sample DNA from ticks for detection of DNA belonging to genus *Rickettsia*. For the negative control, we used Molecular-grade water (1,5 uL) while for the positive control, we employed DNA from VERO cells infected with *R. amblyommii*. The reactions were completed with 8,5 µl of molecular-grade water and carried out at Applied Biosystems® StepOne™ Real-Time PCR System (Applied Biosystems, USA), according the following conditions: 1 cycle at 95°C for 5 min, followed by 50 cycles of 15 s at 95°C, 30 s at 53°C, and 30 s at 60°C.

Midguts from female specimens were dissected according Maruyama et al. (2010) at RNAse free conditions, through the aplication of RNAse inhibitor RNAZAP (Invitrogen, San Diego, CA, USA) and then, maintained into RNA Later (Ambion®, USA) at the bench, tips and samples. All the experiments were approved by Animal Experimentation Ethics Committee of the Federal University of Vicosa, under protocol number 07/2012, according to Ethical Principles on Animal Research established by the Brazilian College for Animal Experimentation (COBEA) and with actual Brazilian legislation. All the experimental steps were carried out at the Center of Biomolecules, Federal University of Vicosa, Brazil.

2.2. Total RNA extraction, cDNA libraries and NGS sequencing

The total RNA extraction of the samples (MI and MNI) were carried out using the

RNAqueous® RNA extraction kit (Ambion®, USA), according manufacturer's recommendations. The measurements of RNA quality and concentration were determined by NanoDrop 2000 (ThermoScientificTM, USA) and Bioanalyser Agilent 2100 (Agilent Technologies, Santa Clara, USA) following manufacturer procedures.

The construction of the cDNA libraries derived from the MI and MNI treatments were conducted using Clontech SmartPCR cDNA kit (Clontech Laboratories, Inc, Mountain View, CA USA) through a termal bath program of 6 min at 68°C, followed by 15 cycles of 10 min at 70 °C, with aid of termocycler (StepOne, Applied Biosystems, USA) according manufacturer protocols. After the remotion of the adaptors sequences from de cDNA through restriction enzymes, the cDNA double stranded ends were converted into blunt ends using T4 DNA polymerase, Klenow fragment of DNA polymerase I and T4 polynucleotide kinase and then, purified with aid of magnetic beads (Clontech Laboratories, Inc, Mountain View, CA USA). From these cDNA samples, only those smaller than 200pb were chosen through the electrophoresis in a agarose gel and following amplification by PCR with specific primers to the adaptors sequences. The purification of the double strand cDNA samples were performed using using QiaQuick PCR extraction kit (CA, USA) and eluted with elution buffer EB from the same kit. The confection of the Illumina libraries was carried out using NEBNext reagents (New England Biolabs, Ipswich, MA USA) following manufacturer guidelines and their quality and size distribution was evaluated through Bioanalyser Agilent 2100. The sequencing processes was carried out at Illumina Hiseq 2000, according instrumental recommendations and the quality of the reads was checked by FASTQC algorithm (Babraham Institute, Cambridge, UK). Once the transcriptomes were derived from midguts of non-infected *A. sculptum* (MNI) and infected with *R. amblyommii* (MI), we employed the term "mialome", that means midgut transcriptome, which were used first by Anderson et al. (2008).

2.3. Additional data, de novo assembly and CDS prediction

As additional libraries, we sequenced a total RNA samples from *A. sculptum* females non-infected by Ion Personal Genome Machine® (PGM™, USA) System, according manufacture protocols. Other 4 additional datasets (Garcia et al., 2014) constituted by *A. sculptum* female sialomes (SG), were downloaded from the NCBI under the accession under accession SRX498136. The NGS data generated from this work were deposited on NCBI under the Bioproject with the accession number PRJNA309641.

To construct the whole midgut transcriptome of *A. sculptum* (MDG), we performed the *de novo* assembly of dataset reads generated for MI (49.209.818 reads) and MNI (53.006.356 reads) together with other additional datasets of *A. sculptum*, as follow: (1) 45.417.679 reads from *A. sculptum* Ovaries Non-Infected (ONI); (2) 46.573.835 reads from *A. sculptum* Ovaries Infected with

R. amblyommii (OI); (3) an additional PGM Ion Torrent dataset of 844,529 reads belonging to MNI treatment; (5) a dataset of 703,210 pyrosenquencing reads from *A. sculptum* salivary glands transcriptomes previously described (Garcia et al., 2014). All this sequenced reads were *de novo* assembled into contigs using CLC Genomics Workbench 8.0 package (CLC Bio, Denmark). This assembly produced 460,445 contigs with a average size of 222pb and length ranging from 100 to 14,579pb.

In order to identify a greater number of novel genes for the *A. sculptum* mialome, we employed AUGUSTUS program (http://augustus.gobics.de) that employs the generalized hidden Markov Model, which allows the identification of gene sequences adopting patterns pre-established. This approach has been described to perform more accurately gene determination than other conventional gene predictors for eukaryotes (Stanke et al., 2004; Stanke et al., 2005; Stanke et al., 2006; Hoff et al., 2015). For the identification of the 27,308 CDS whole *A. sculptum* reference we trained the Augustus program using *Ixodes scapularis* structural genome that had been deposited in GenBank, at this moment: (http://metazoa.ensembl.org/Ixodes_scapularis/Info/Index). The mapping of the reads from MI and MNI transcriptomes against the 27,308 allowed the identification of 25,569 CDS expressed in both *A. sculptum* midguts dataset. This new dataset was denominated whole *A. sculptum* mialome (MDG).

2.4. Functional annotation and gene expression inferences

For the functional annotation, all the 25.569 CDS from whole *A. sculptum* mialome (MDG) were queried against several databases through the BLASTp, BLASTx, RPSBLAST and BLASTn algorithms (Altschul et al., 1990) adopting an E-value cut-off of 1E-04 for annotation validation, as made in previous works (Karim et al., 2011). In addition, as a help for the functional annotation, we determined CDS coding signal peptides, transmembrane domains, addressing subcelular localization and related metabolic pathways, through SignalP, TMHMM, TargetP and KEGG database, respectively.

For the BLAST searches of homology, we employed the following: (1) the non-redundant protein database (NR) of the National Center for Biology Information (NCBI); (2) Gene Ontology; (3) SwissProt; (4) UNIPROT; (5) NCBI-Acari; (6) CDD; (7) COG; (8) KOG; (9) PFAM; (10) SMART; (11) UNIPROT-Acari; (12) *Bombyx* mori; (13) *Chelicerata*; (14) rickettsiae; (15) *Amblyomma americanum*; (16) *A. cajennense* (*A. sculptum* complex); (17) *A. maculatum*; (18) *A. parvum*;(19) A. *triste*; (20) *Argas*; (21) *Dermacentor*; (22) *I. scapularis*; (23) *I. ricinus* ; (24) *Rhipicephalus* and (25) *Haemaphysalis*.

According approach adopted in many tick transcritomes works, we defined for the final annotation BLAST results with: (1) Minor E-value (<1E-04); (2) Higher score and (3) key-words

from the BLAST results (KOG, COG, GO, SwissProt, UniprotKb, Acari, Chelicerata, etc). According BLAST results, the 27,308 CDS dataset was functionally classified into 5 general functional classes: Housekeeping (H); Secreted (S); Transposable elements (TE); Unknown (U) and Viral related (V). Coding sequences that presented positive result in the SignalP analysis were annotated as Secreted. Coding sequences with no hits against databases, or BLAST hits with Evalue > 1E-04 were classified into Unknown products (U). Subcategorization of the general functional classes into functional classes was performed as approach extensively employed to tick transcriptomes (Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009) which considers keys words from the final annotation (best hits) for identification of gene families into functional classes previously annotated for tick transcriptomes.

The results were annotated in a excel spreadsheet, from which the data can be extracted, in a large scale, through Visual Basic programs designed by Guo et al. (2009). For the gene expression inference, the universe of reads belonging from a specific treatment was mapped of against the respective generated CDS dataset, followed by normalization through RPKM calculation. The determination of differentially expressed gene was performed by calculation of the ratio between RPKM values for MI and MNI CDS, followed by Log2 Fold-change determination, considering the pairwise comparison MI against MNI. The significative Log2FD values were determined by GFOLD program, using Blind mode (Feng et al., 2012).

For the prediction of metabolic pathways and gene expression inference, we employed the algorithm Pathway view from R package, using KEGG annotated list with their respective Log2FC values (MI versus MNI), as a input data. All the procedures to this analysis is at Pathway view document of R package (https://www.bioconductor.org/).

3. RESULTS AND DISCUSSION

3.1. Characteristics of the *A. sculptum* mialomes

A total of 25,569 CDS were obtained from both *A. sculptum* mialomes, of which 20,013 were yielded from midguts infected with *R. amblyommii* (MI) and 12,116 CDS from *A. sculptum* non-infected midguts (MNI). All the coding sequences were analyzed using BLASTn, BLASTp, BLASTx and RPSBLAST algorithms and compared against to public databases, as well as, other arthropod sequences datasets published at NCBI and UNIPROT, according described in Methods.

As a complementary annotation step, both mialomes were analyzed using the KEGG Pathview algorithm (Luo & Brouwer, 2013) whereby 5,168 CDS presented homology to conserved sequences from this database, with the identification 353 different putative metabolic pathways from the *A. sculptum* CDS (data not shown). The most important metabolic pathways identified in this analysis, related to glycolysis, citrate cycle and signaling pathways are discussed further in this

text. The predictions of secreted proteins and transmembrane domains were performed using *SignalP* and *TMHMM* algorithms, respectively (Emanuelsson et al., 2007; Sonnhammer et al., 1998; Krogh et al., 2001). The best match (annotation) among the BLAST results against the databases was manually chosen and curated according key-words of the main databases, as previously performed for others tick transcriptomes (Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009).

All 25,569 coding sequences from *A. sculptum* mialome were categorized into five general functional classes according to the best BLAST match and *SignalP* results, as summarized at the Table 1. A total of 21,177 CDS (85,17%) were annotated as Housekeeping (H) genes, while 2,049 (8,01%), that presented signal peptide at *SignalP* analysis were pre-classified as Secreted proteins (S). A total of 256 CDS (1%) were classified as Transposable elements (TE), while 19 CDS (0,074%) presented best hits to viral sequences (V). Finally, even with BLAST comparison against 30 databases, a subset of 1,468 predicted coding sequences remained without BLAST hits or did not present hits with E-value cut-off less than 1E-04 (see Methods). These classified into the Unknown (U) class. This unknown proteins dataset could represent novel proteins, therefore, could be an interesting *A. sculptum* gene dataset for future studies.

In order to go deeper into the *A. sculptum* mialomes information, this five general classes were further subcategorized into 29 functional groups (Table 2) by taking in consideration key-terms from the BLAST results and described gene families for ixodidic ticks, as previously employed for other Acari transcriptomes (Karim et al., 2011; Garcia et al., 2014; Kotsyfakis et al., 2014).

Table 1. Functional classification of the 25,569 all expressed CDS *A. sculptum* mialomes

CLASS	CDS	Reads	Reads/CDS	CDS*	Reads*
Housekeeping	21,777	26,305,341	1207.941	85.16954	65.28166
Secreted	2,049	9,040,561	4412.182	8.01361	22.43586
Transposble elements	256	41,237	161.082	1.001212	0.102337
Unknown	1,468	4,899,306	3337.402	5.741327	12.15855
Viral	19	8,700	457.8947	0.074309	0.021591
TOTAL	**25,569**	**40,295,145**	**1575.937**	**100**	**100**

*percentage of Reads and CDS

3.2. Overview on Housekeeping genes (H) from *A. sculptum* mialomes

All the 21,177 CDS classified as Housekeeping genes were subcategorized into 24 subclasses as shown at the Table 2 and each subclass is explored at the Spreadsheet S1. With a 2,870 CDS subset, the Signal transduction subclass arises as the most represented functional Housekeeping class in *A. sculptum* midguts, followed by Transporters coding sequences (1755 CDS). Cytoskeletal related transcripts (1196) and Unknown conserved sequences (1639).

Transcripts related with metabolic processes were categorized as Carbohydrate metabolism (835), Lipid metabolism (751), Amino acid metabolism (580), Nucleotide metabolism (291), Energy metabolism (758) and Intermediate metabolism (1720). This abundance of coding sequence related to metabolic processes was expected, regarding the importance of this organ to the assimilation of nutrients. Transcripts related with gene expression machinery were categorized as Nuclear regulation (1351), Transcription factors (735), Transcription machinery (1438), Nuclear export (145), Protein synthesis machinery (1158), Protein modification machinery (955), Protein export (528). Transcripts related with cellular turnover, detoxification and defense are represented by Proteasome machinery (610), Detoxification (628 CDS) and Immunity related products.

Table 2. Classification of Housekeeping genes in the *A. sculptum* mialomes (MI and MNI)**.**

Subclasses	MDG	MI	MNI	DEG	Up	Down
Signal transduction	2,870	2,513	1,123	1,647	1,238	409
Transporters	1,755	1,282	820	1,169	690	479
Intermediate metabolism	1,720	1,011	969	1,203	553	650
Unknown, conserved	1,639	1,160	702	874	414	460
Transcription machinery	1,438	1,146	742	1,000	717	283
Nuclear regulation	1,351	1,065	602	888	597	291
Cytoskeletal	1,196	1,085	510	735	583	152
Protein synthesis machinery	1,158	909	735	872	560	312
Protein modification machinery	955	735	480	652	431	221
Extracellular matrix/cell adhesion	871	687	402	573	376	197
Carbohydrate metabolism	835	616	447	586	367	219
Energy metabolism	758	519	457	619	382	237
Lipid metabolism	751	602	328	493	334	159
Digestive enzymes	737	718	287	531	492	39
Transcription factors	735	708	203	409	368	41
Detoxification	628	512	273	460	359	101
Proteasome machinery	610	599	242	399	368	31
Amino acid metabolism	580	367	334	423	218	205
Protein export machinery	528	478	226	372	312	60
Nucleotide metabolism	291	177	178	227	121	106
Nuclear export	145	121	66	86	62	24
Immunity	113	96	46	72	58	14
Storage	62	59	33	49	43	6
Signal transduction, apoptosis	51	50	20	34	31	3
TOTAL	**21,777**	**17,215**	**10,225**	**14,373**	**9,674**	**4,699**

MDG: Whole *A. sculptum* midgut transcriptome; MI: Transcriptome of *A. sculptum* midgut tract infected with *R. amblyommii*; MNI: Transcriptome of *A. sculptum* midgut tract non-infected; DEG: Differentially expressed CDS in MDG in response to infection (MI versus MNI); Up-regulated CDS in *A. sculptum* midguts in response to rickettsial infection (MI *versus* MNI). Down-regulated CDS in *A. sculptum* midguts in response to rickettsial infection (MI *versus* MNI).

In making the comparison between the *A. sculptum* mialomes (MI against MNI), we observed that 14,373 Housekeeping CDS were differentially expressed in response to rickettsial

infection, of which 9,674 were up-regulated in MI in comparison to MNI. These gene expression variations on *A. sculptum* midguts suggests tick defense responses against bacterial invasion or, perhaps gene expression manipulation of host cell by rickettsiae. This last issue has been described for several of intracellular pathogens, such as *Salmonella* (Hannemann et al., 2013; Verma et al., 2015), *Anaplasma phagocytophilum* (Garcia-Garcia et al., 2009), *Toxoplasma gondii* (Nelson et. al., 2007), *Coxiella burnetii* (Mahapatra et al., 2010) and *Legionella pneumophila* (Rolando et al., 2013).

The following text makes a relationship of the *A. sculptum* dataset annotated at the Spreadsheet S1 with possible host molecules involved in SFG rickettsiae infection mechanism in *A. sculpum* midguts.

3.2.1. Transcripts putatively related with SFG rickettsiae infection mechanisms at A. sculptum midgut tract

As a distinctive feature from the other rickettsial group, SFG rickettsiae depends on physical association with host actin filaments during its process of infection and establishment into mammalian (Heinzen et al., 1999;) and tick host cells (Petchampai et al., 2015).

We have identified 348 coding sequences putatively related with the rickettsial infection mechanism at *A. sculptum* midguts, of which 185 were more abundant at the infected condition (MI) in comparison to non-infected ticks (MNI) as shown at the Table 3. It is worthy to note the more abundance of Actin filaments (29 coding sequences more abundant) a key molecule for SFG rickettsiae infection (Martinez & Crossard, 2004) and tyrosine kinase proteins (20 CDS more abundant) previously described to be necessary during rickettsial invasion of mammal cells (Martinez & Cossart, 2004; Chan et al., 2010). We also observed more abundance of clathrins (19), Src kinase (17), Phosphatidylinositol 3 kinase (12), cdc42 (10), Rho GTPase Rac1 (10), Arp 2/3 complex (9), Dynamins (6), intergrins (4) and neural Wyskott-Aldrich syndrome protein, WASP (2) which is mimicked by rickettsial protein RickA (all these proteins are necessary for SPG rickettsiae infection mechanism (Martinez & Cossard, 2004; Chang et al., 2009; Petchampai et al., 2015).

Among the Cytoskeletal genes classified (Table 3 and Spreadsheet S1) 5 CDS related with Arp 2/3 complex were more abundant in midgut infected *A. sculptum* transcriptome (MI) in comparison to respective non-infected group (MNI). This result suggests the involvement of this cytoskeletal component on *R. amblyommii* infection mechanism of *A. sculptum* midguts, as reported for other tick species (Petchampai et al., 2015). Corroborating this evidence, Petchampai et al. (2013), showed by biochemical inhibition assays with *Dermacentor variabilis* embryonic cell lines, the involvement of the tick Arp 2/3 complex during *Rickettsia montanensis* infection. In addition, many studies about SFG rickettsiae infection on mammal cells have confirmed Arp2/3 as

key molecule for the rickettsial infection mechanism (Heinzen, 2003; Martinez & Crossard, 2004; Reed et al., 2012; Petchampai et al., 2015). These similarities suggest that the *A. sculptum*-SFG rickettsiae interface shares a similar infection mechanism.

In order to perform an *in silico* inference about the putative infection mechanism of *A. sculptum* midguts by *R. amblyommii*, we have identified a homologue KEGG pathway of epithelial cells when infected by *Listeria*, *Salmonela* and *Shigella*, deposited on KEGG database (Figure 1). Previous works have shown to similarities among the infection mechanisms employed by these intracellular pathogens and the rickettsial infection mechanisms (Ireton et al., 1996; Gouin et al., 1999). The more abundance of these proteins, in association with KEGG results, suggests that mechanism of rickettsial infection of *A. sculptum* midguts could be similar to those observed for others host cells, with the involvement of tick actin filaments, Arp 2/3 complex, WASP complex, phosphatydilinositol and tyrosine kinases. Other genes not covered by the KEGG annotation, such as fibronectins were identified and annotated through search homology against other databases, as shown at the Spreadsheet S1.

Table 3. Up regulation of *A. sculptum* midgut proteins (MI versus MNI) putatively involved in the rickettsial infection mechanism.

CATEGORY	CDS	Up	Annotation	Reference
Arp related	41	33	cs	Petchampai et al., 2015
Actin	44	29	cs	Heinzen et al., 1999
Tyrosine kinase	64	20	st	Petchampai et al., 2015
Clathrin	29	19	pe, extmat	Chang et al., 2009
Src kinase	28	17	st	Martinez & Cossart, 2004
PI3K	21	12	st	Martinez & Cossart, 2004
cdc42	32	10	st	Martinez & Cossart, 2004
Rho GTPase Rac1	29	15	st	Reed et al., 2012
Arp 2/3 complex	16	9	cs	Petchampai et al., 2015
Dynamin	8	6	cs, pe	Zhao et al., 2016
Focal adhesion kinases	12	5	st	Petchampai et al., 2015
Integrin	7	4	extmat	Hillman et al., 2013
Septin family	10	3	cs, extmat, nr	Haglung & Welch, 2011
WASP related proteins	5	2	cs	Petchampai et al., 2015
Cortactin	2	1	cs	Martinez & Cossart, 2004
TOTAL	**348**	**185**		

NOTE: cs= Cytoskeletal; extmat= Extracellular matrix; pe= Protein export machinery; nr= Nuclear regulation; CDS= expressed coding sequences at the whole midgut transcriptome; Up= More abundant CDS in *A. sculptum* infected midguts (MI) in comparison to non-infected midguts (MNI).

3.2.2. More abundance of CDS related with gene expression control in A. sculptum infected midguts

3.2.2.1 ***Nuclear regulation machinery***

A total of 1,359 CDS were annotated as Nuclear regulation related proteins, of which 603 were more abundant in *A. sculptum* infected midguts in response to rickettsial infection (MI *versus* MNI), as shown in the Table 1. This subset were subcategorized into 38 functional families, as shown at the Spreadsheet S1. Among the most more abundant gene families are those related to helicases, with 80 up-regulated genes, followed by DNA polymerases (75 CDS), Repair machinery related genes (53), Topoisomerases (19) and chromatin structural proteins, with 29 up-regulated genes. The other functional subcategories and their proportion of up-regulated genes in response to rickettsial infection are listed in the Spreadsheet S1 (Supplemental information).

Among these gene families, it is worthy to note those related to structural histones (27 up-regulated genes) of which, 7 CDS codify for H2B histones. Thepparit et al. (2010) observed that the physical association of OmpB antigen from *R. felis* with H2B histones is a necessary step to rickettsial infection of IS6 culture cells. These data suggest the hypothesis of modulation of the tick gene expression machinery by rickettsiae, similarly to what has been observed with other intracellular pathogens-host interactions (Nelson et al., 2008; Shaw et al., 2010). In confirmation of this point of view, we observed the more abundance of enzymes related to histones modification, such as histone acetylases (18) which weakens the histone association with DNA, resulting in higher gene expression and, histone deacetylases (11) that undoes the effects of histone acetylases, repressing the gene expression (Marmorstein & Trievel, 2009).

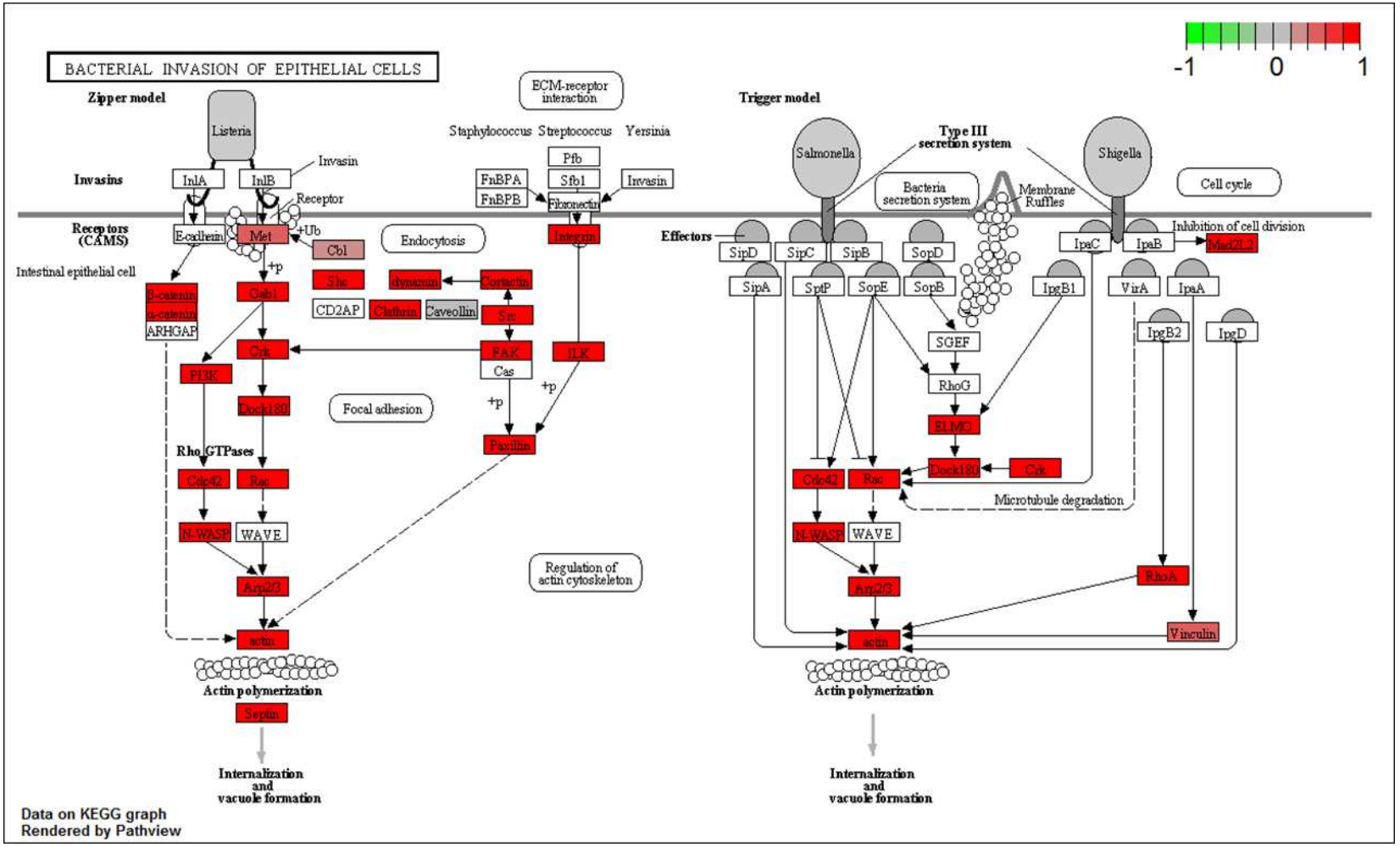

Figure 1. Differential expression of CDS from *A. sculptum* mialomes homologues to proteins involved in the regulation of actin cytoskeleton during baterial invasion of epithaliel cells (KEGG consensus pathway). In a scale from red to green, red boxes represent up-regulated genes at *A. sculptum* midguts in response to SFG rickettsial infection (MI versus MNI). Green boxes represent down-regulated CDS in response to infection (more expressed in MNI than MI). White boxes indicate genes not covered in the KEGG annotation of the whole *A. sculptum* mialome (MDG).

3.2.2.2. ***Transcription factors and other transcription related CDS***

We have identified 735 coding sequences related to Transcription factors in the whole *A. scultum* mialome (MDG), which were subcategorized into 55 families, as listed in the Spreadsheet S1. A total of 368 Transcription factors genes were more abundant in the tick midguts infected with *R. amblyommii* (MI) in comparison to non-infected condition (MNI). Among the up-regulated transcription factor genes, stand out those related to Zinc finger containning domains (133 CDS), followed by C2H2 family (32), Forkhead superfamily (27), Homeobox (16), bZIP family (12), Myb family (12), HOX family (8) and those related to TATA-binding proteins, with seven more abundant CDS in response to rickettsial infection. The great variety of transcription factor families more abundant in *A. sculptum* infected midguts reforces the hypothesis that rickettsiae bacteria, directly or indirectly, can induce changes in the host gene expression. These transcription factor families have been observed as well in *I. ricinus* mialomes (Kotsyfakis et al., 2015). Other transcription factor families identified in this study are annotated in the Spreadsheet S1.

3.2.2.3. ***Protein synthesis and modification machinery***

As previously shown (Table 1) 1,158 coding sequences were related to Protein synthesis machinery. This functional class were subcategorized into 58 functional families, as shown in Spreadsheet S1. A subset of 20 functional families containing 343 coding sequences were related to all 20 aminoacyl-tRNA synthetases, of which 142 CDS were up-regulated (MI versus MNI). In association to this result, from a subset of 202 CDS related with ribosomal structure proteins, 60 were up-regulated in *A. sculptum* midguts in response to rickettsial infection. These results suggest that rickettsial infection promotes an increase of protein synthesis activity in *A. sculptum* midguts, which is corroborated by KEGG Pathview analysis (Figure 2). Indeed, we have observed the more abundance (MI versus MNI) of several cellular processes that could possibly signalize a tick response against the infection, such as detoxification processes and immunity related processes (Spreadsheet S1).

Corroborating these results, we also verified a several more abundant enzymes (MI versus MNI) related to protein modification processes (Spreadsheet S1) such as Signal peptidases (5 up-regulated genes), Phosphatases (7), Acetyltransferases (2), Methyltransferases (9) and Serine proteinases belonging to KU family (7). It was also observed more abundance (MI vs MNI) of genes belonging to the chaperones families, such as DnaJ, with 45 genes more abundant, DnaK (4), Hsp 70 (20), chaperonins (20) and Hsp 90 (4), which suggest the *A. sculptum* midguts responses against cellular stress, in the case of rickettsial infection. This hypothesis is supported by the more abundance (MI versus MNI) of coding sequences related to antimicrobial peptides, such as

Microplusins (Spreadsheet s1)

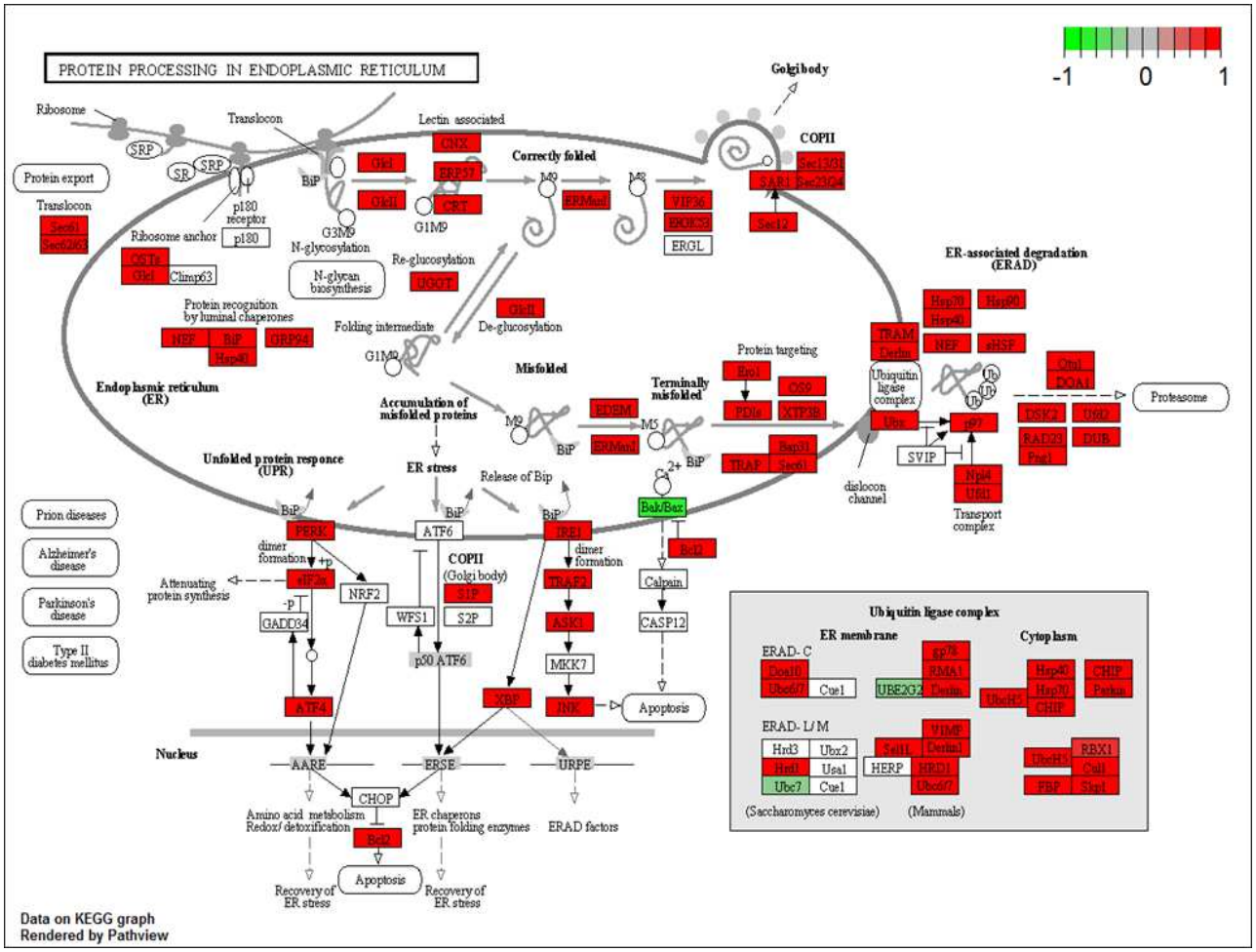

Figure 2. More abundance of CDS related to protein synthesis and modification machinery in response to rickettsial infection (MI versus MNI). In a scale from red to green, red boxes represent up-regulated genes at *A. sculptum* midguts in response to SFG rickettsial infection (MI versus MNI). Green boxes represent down-regulated CDS in response to infection (more expressed in MNI than MI). White boxes indicate genes not covered in the KEGG annotation of the whole *A. sculptum* mialome (MDG).

3.2.3. Endocytosis, phagocytosis and protein export machinery genes during rickettsial infection in A. sculptum midguts

There were 530 coding sequences related to Protein export machinery, of which 313 CDS were more abundant in *A. sculptum* female midguts in response to rickettsial infection (MI *versus* MNI) (Table 1). Spreadsheet S1 shows the distribution of those more abundant genes among 20 functional gene families. Among of then, it is worthy to note those related to 11 more abundant SNARE related proteins (MI *versus* MNI). These proteins were previously described to orchestrate the fusion of export vesicle proteins with cellular membrane in salivary glands, ovaries and midguts of *A. maculatum* and *A. americanum* (Villareal et al., 2013).

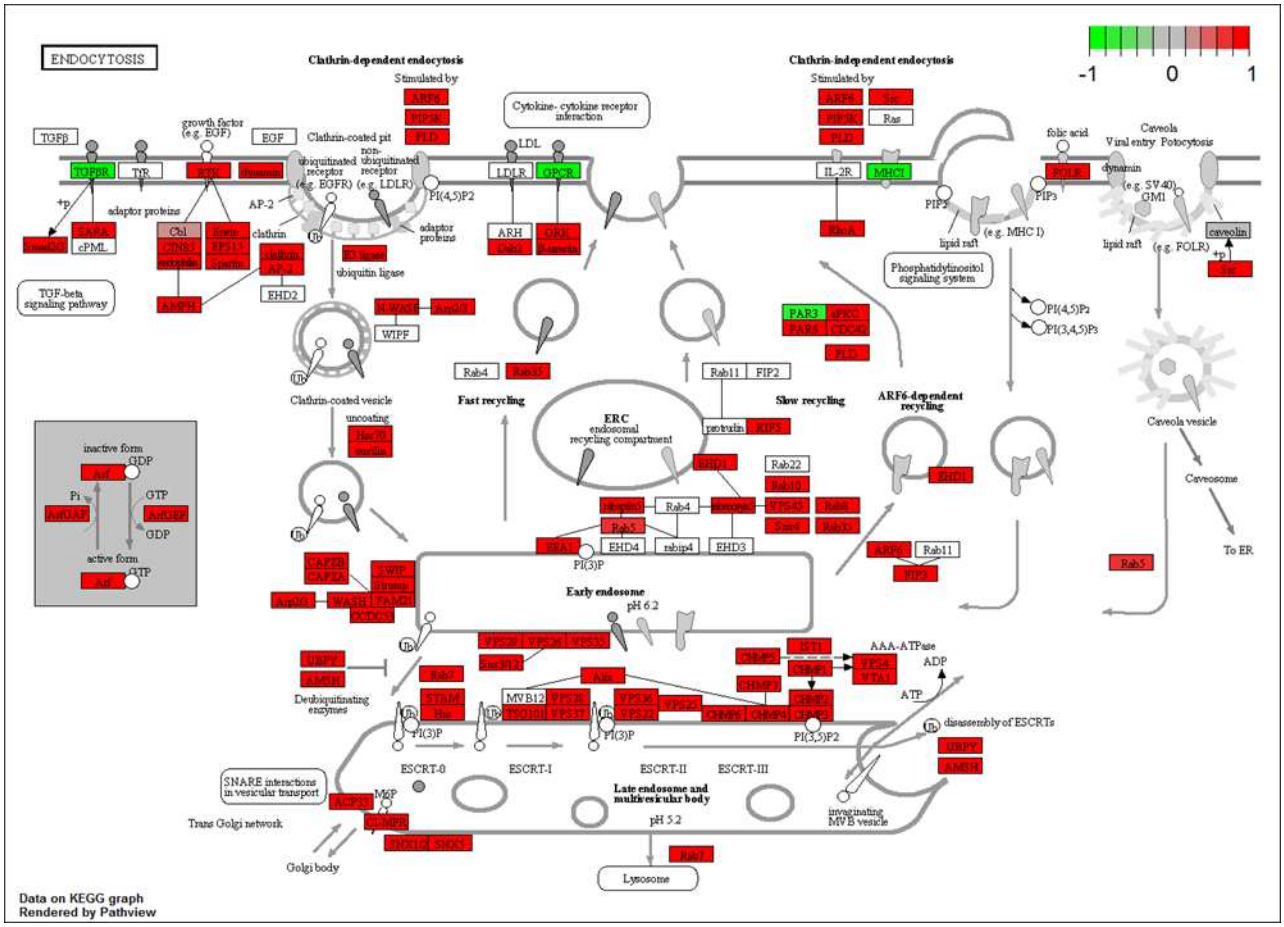

Figure 3. More abundance of CDS related to vesicle transport and endocytosis in response to rickettsial infection (MI versus MNI). More abundant proteins, such as clathrins, dynamins, PTK and Arp2/3 complex have previously described to participate of the rickettsial infection process in mammals and ticks (Chang et al., 2009; Petchampai et al., 2015). In a scale from red to green, red boxes represent up-regulated genes at *A. sculptum* midguts in response to SFG rickettsial infection (MI versus MNI). Green boxes represent down-regulated CDS in response to infection (more expressed in MNI than MI). White boxes indicate genes not covered in the KEGG annotation of the whole *A. sculptum* mialome (MDG).

The infection steps of host cells by SFG rickettsiae are: (1) invasion through endocytosis-mediated mechanism; (2) escape from the phagosome; (3) intracellular growth and (4) cell-to-cell spread (Martinez & Cossart, 2004). Through our analysis, we have identified several host proteins related with phagocytosis and endocytosis processes, such as clathrins and integrins (Spreadsheet S1) that were previously related to be essential for rickettsiae invasion (Petchampai et al., 2015; Martinez & Cossart, 2004). A total of 15 clathrins and 4 integrins were more abundant in *A. sculptum* midguts in response to rickettsial infection (Spreadsheet S1), which suggests that a similar mechanism of rickettsial infection might be happen during the infection and establishment of *R. amblyommii* in the midgut tract.

3.2.4. Signal transduction

We also observed more abundance (MI *versus* MNI) of several genes related to signaling pathways (1,238 CDS from 2,870), especially those that regulate cytoskeletal machinery and metabolic pathways activation (Table 2 and Spreadsheet S1). During our analysis of the more abundance (MI *versus* MNI) of cytoskeletal related genes, we verified the up-regulation of CDS related to intermediary components of the Rap1 signaling pathway, which has been described to regulates cytoskeletal machinery in *Dictyostelium discoideum*, an eukaryotic model (Lee & Jeon, 2012).

Through the Pathview analysis (see Methods) we have observed the more abundance of CDS related to RIAM, Profilin and F-actin (Spreadsheet S1 and Figure 2) which suggests an activation of cytoskeletal machinery recruitment mediated by Rap1. In addition, the more abundance of CDS related to Rap1, PI3K and Akt suggests the activation of PI3K-Akt signaling pathway, that regulates metabolic processes in eukaryotic cells (Ward & Thompson, 2012).Thus, the up-regulation of these pathways would contribute to a gain of performance of the tick vector. As SFG rickettsiae is dependent of the host metabolism, especially pyruvate and amino acids (Renesto et al., 2005; Fuxelius et al., 2007), the activation of those pathways would features a putative mutualistic relationship between *A. sculptum*-rickettsiae.

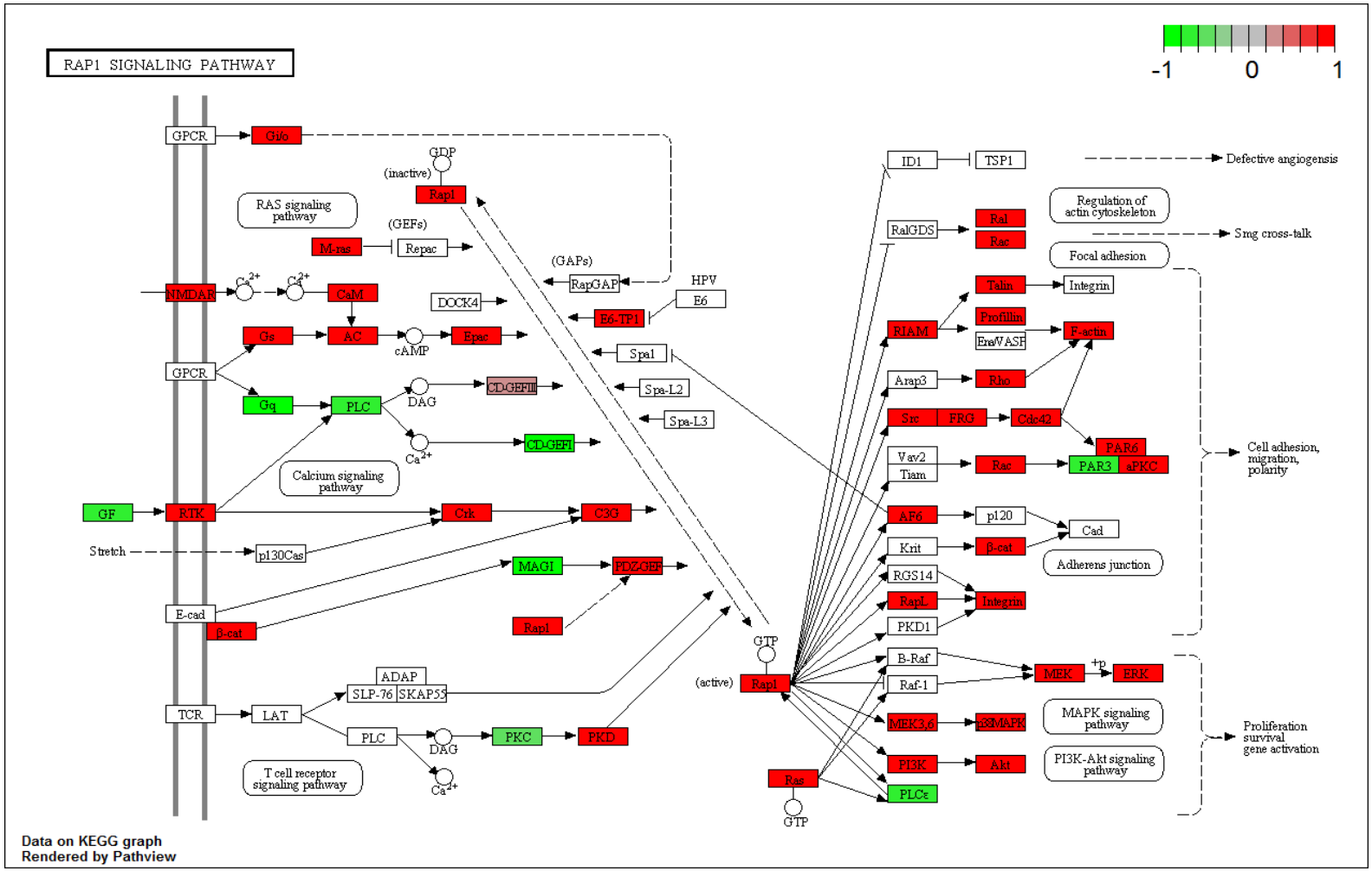

Figure 4. Differential expression of CDS from *A. sculptum* mialomes related to Rap1 signaling pathway in response to *R. amblyommii* infection (MI *versus* MNI). In a scale from red to green, red boxes represent up-regulated CDS in *A. sculptum* midguts in response to infection. Green boxes represent down-regulated CDS in response to infection (more expressed in MNI than MI). White boxes indicate genes not represented at the whole *A. sculptum* mialome (MDG).

3.2.5. Overexpression of genes related with nutrient transporters and catabolic pathways in response to rickettsial infection

3.2.5.1. ***Protein digestion***

Coding sequences related to this class are annotated into Digestive enzymes, in the Spreadsheet S1. A total of 737 CDS related to digestive proteases was observed in *A. sculptum* mialomes (MI and MNI), of which 492 were up-regulated in response to *R. amblyommii* infection (MI *versus* MNI). Among the more abundant proteins families, stand out those previously related to the metabolism of hemoglobin in other ixodidae ticks (Boldbaatar et al., 2006; Sojka et al., 2013) such as 1 leucine aminopeptidase, 32 serine proteases, 12 aspartyl proteases, 1 cathepsin B, 1 cathepsin C, 1 cathepsin D and 2 cathepsin L (Spreadsheet S1). This result suggests that rickettsial infection leads to an increase of the digestory activity in *A. sculptum* midguts, what could be a great advantage for the tick, which features a mutualism-like relationship, as observed for other rickettsialles species (Kremer et al., 2009). Other proteases families from *A. sculptum* midguts are annotated at the Spreadsheet S1 and serve as an interesting source of information to further development of inhibitors-based acaricides.

3.2.5.2. ***Carbohydrate digestion***

Moretti et al. (2013) recently characterized digestive carbohydratases from *A. sculptum* midguts such as, maltases, amylases, trehalases, chitinases, alpha-L-fucosidases and among others. We have annotated 142 coding sequences from *A. sculptum* mialome related to 17 carbohydrate enzymes families, as shown at the Spreadsheet S1. From this CDS dataset, 91 were up-regulated in MI in comparison to MNI, of which 10 were annotated as fucosyltransferases, 9 as glucosilases, 8 amylases, 8 lectins, 6 maltases, 6 fucosidases, 5 melibiases, 4 mannosyl transferases, 2 glucanases and others. The presence of these enzymes families in the *A. sculptum* mialome dataset corroborates the carbohydratases repertoire described in the literature (Moretti et al., 2013), which certifies the annotation approach employed in this work. Other less expressed carbohydratases in *A. sculptum* mialomes are annotated in the Spreadsheet S1.

3.2.5.3. ***Lipid digestion and metabolism***

Caperucci et al. (2010) observed an increase of lipids vesicles in *A. sculptum* midguts during the blood feeding. Transcripts related to lipids enzymes, such as lipases, phospholipases and sphyngomielinases have been annotated in *I. ricinus* mialomes (Kotsyfakis et al., 2014). However, there is still little information about lipid digestion in *A. sculptum*. From the present *A. sculptum* mialome dataset, we have annotated 205 CDS related to Lipid digestion, of which 130 were more abundant in response to rickettsial infection (MI *versus* MNI, as shown at the Spreadsheet S1).

Among them are lipases and esterases (25), Acetylcholinesterases and Butyrylcholinesterases (20), Phosphodiesterases (18), Thioesterase(14), Carboxyesterases (13), Isoamyl-acetate esterases (12), Triglyceride lipase-cholesterol esterases (4), Acetyl esterases (3) and Phosphoesterases (2). We observed the more abundance of several coding sequences related to Beta-oxidation of fatty acids and steroids metabolism, which suggest that rickettsial infection promotes an increase of lipid metabolism in the midgut.

3.2.5.4. ***Nutrient transporters***

We identified 1,755 coding sequences related to Transporters at the whole *A. sculptum* mialome (MDG), of which 690 CDS were up-regulated in response to rickettsial infection (Table 2). The most prevalent up-regulated transporters families are those annotated as Ion transporters (with 100 up-regulated genes), followed by Metabolites transporters (94), Sodium symporters (90), Amino acids and peptides transporters (68), sugar/nucleotide transporters (55) and Metal transporters (38). Other families are listed at the Spreadsheet S1. As with the more abundance of digestive enzymes, as discussed previously, the higher expression of nutrients transporters, especially those related with amino acids, sugars and other metabolites could be an advantage achieved by rickettsial infection condition, which suggests a gain of midgut capability to assimilate nutrients. This advantage could results in a higher blood-sucking activity by the tick vector and, as consequence, a higher probability to rickettsial transmission to the vertebrate host. Indeed, a similar nutritional mutualism has been verified between Rickettsiales *Wolbachia* and *Cimex lectularius*, where the bacteria increases the bedbug capability to assimilate B vitamin (Nikoh et al., 2014).

Another transporter family worthy to note is that annotated as ABC superfamily, represented by 293 CDS from MDG, of which 33 were up-regulated in response to rickettsial infection (Spreadsheet S1). Extensively studied in *Drosophila* (Huang et al., 2014; Prince et al., 2014), this transporter family employs ATP energy to develop a variety of functions, especially as xenobiotics transport (Dermauw & Leeuwen, 2014). The more abundance of this transporter family in *A. sculptum* infected midguts would confer resistance to acaricides.

3.2.5.5. ***Carbohydrate, amino acid and nucleotide metabolism***

A total of 852 coding sequences was annotated as enzymes of carbohydrate metabolism and subcategorized into 26 families (Spreadsheet S1). The most higher expressed families in the midgut in response to rickettsial infection (MI versus MNI) were those related with Glycolysis/gluconeogenesis pathways, with 58 more abundant CDS, followed by Tricarboxylic acid (63 more abundant CDS) and Pentose's phosphate pathway, with 20 CDS more abundant.

Analysis of rickettsial genomes has demonstrated that SFG rickettsiae lack several metabolic

pathways, such as glycolysis, pentose phosphate pathway, as well as, incomplete pathways, such as amino acids, nucleotide and lipid metabolism (Fuxelius et al., 2007). These features explain the metabolic dependence of the *Rickettsia* genus and their intracellular obligatory life cycle (Fuxelius et al., 2007).

This lack of rickettsial glycolytic enzymes CDS could explain the more abundance of homologues genes in the *A. sculptum*, as SFG rickettsiae are dependent on host cell pyruvate. The bacteria recruit this metabolite from the host cell through membrane transporters encoded by its genomes (Renesto et al., 2005; Fuxelius et al., 2007). Therefore, the more abundance of Glycolysis/gluconeogenesis coding sequences in the infected midguts could be a tick physiologic response to the lower levels of cytosolic pyruvate or a mechanism of manipulation directly promoted by bacteria, via the interaction with the gene expression machinery, such as histone H2B, as previously discussed (Thepparit et al., 2010). Indeed, the Pathview KEGG analysis corroborates the *A. sculptum* mialome analysis, with almost all enzymes related to Glycolysis/gluconeogenesis pathways more abundant (Figure 5).

We also observed the more abundance of CDS related to the tricarboxylic cycle (MI versus MNI) as shown in the Spreadsheet S1. Although this pathway is complete in the Rickettsiales (Renesto et al., 2005; Fuxelius et al., 2007), the Tricarboxylic cycle is a amphibolic pathway, serving both energy metabolism, as well as, anabolic pathways, such as amino acids biosynthesis. As *Rickettsia* are strictly dependent on host metabolites (Renesto et al., 2005; Fuxelius et al., 2007), the higher expression of tricarboxylic acid enzymes could be another positive point for *R. amblyommii* in the midguts. Indeed, from the 580 coding sequences annotated as amino acid metabolism enzymes, 218 were more abundant in *A. sculptum* midguts in response to rickettsial infection, with is in agreement with the observed more abundance of amino acids transporters and protein synthesis machinery (Spreadsheet S1).

In the analysis of the overexpression of genes annotated as replication machinery components we also observed the up-regulation of several genes related to the pentoses phosphate pathway and nucleotide metabolism, as shown at the Spreadsheet S1. The overexpression of metabolic pathways that give support to DNA and proteins synthesis, as previously discussed, suggests that rickettsial infection provides key conditions to a greater development of *A. sculptum* midguts.

3.2.6. Detoxification proteins

A total of 614 coding sequences were annotated as related to detoxification processes, of which 355 presented more abundance of transcripts in *A. sculptum* infected midguts (MI) in comparison to the non-infected *A. sculptum* mialome (MNI). The complete distribution of these

more abundant genes among the 15 detoxification gene families is annotated in Spreadsheet S1. We noted the more abundance of 118 Sulfotransferases and 83 glutathione S-transferases in MI in comparison to MNI (Spreadsheet S1). These proteins have been previously described to play role in the protection of ixodidic ticks against oxidative molecules from vertebrate host (Duscher et al., 2014; Pichu et al., 2011). These results suggests that *R. amblyommii* infection provides a gain of protection in *A. sculptum* midguts against vertebrate oxidative response, which could be a mutual advantage in the relationship between tick-rickettsiae.

3.2.7. Imune related products

From the subset of 113 putative genes annotated as immune related products, we identified 58 more abundant in *A. sculptum* midguts in response to *R. amblyommii* infection (Table 1). The functional distribution of these genes and the respective number of up-regulated genes for each immune related family is specified at the Spreadsheet S1. Among them, it is worthy to note those related to tumor necrosis factor receptors (42 overexpressed CDS in MI in comparison to MNI) which have been recently described to induce the expression of antimicrobial peptides in *Haemaphysalis longicornis* ticks (Takechi et al., 2015). Indeed, we also observed the more abundance of antimicrobial peptides, such as 7 microplusins, in *A. sculptum* midguts as a response to *R. amblyommii* infection (Spreadsheet S1).

In addition, one CDS related to immunoglobulin binding protein or IGBP was more abundant in *A. sculptum* midguts in response to rickettsial infection (MI RPKM 32,8081 vs MNI RPKM 5,16979). These proteins have been described as a tick inhibitory molecule against host vertebrate immunoglobulins (Valenzuela et al., 2002; Gong et al., 2014). This suggests that infected *A. sculptum* might employ a more well adapted immune response repertoire, when infected with *R. amblyommii* then uninfected ticks, which could be an issue to overcome in further vaccines strategies of vertebrate hosts against ticks.

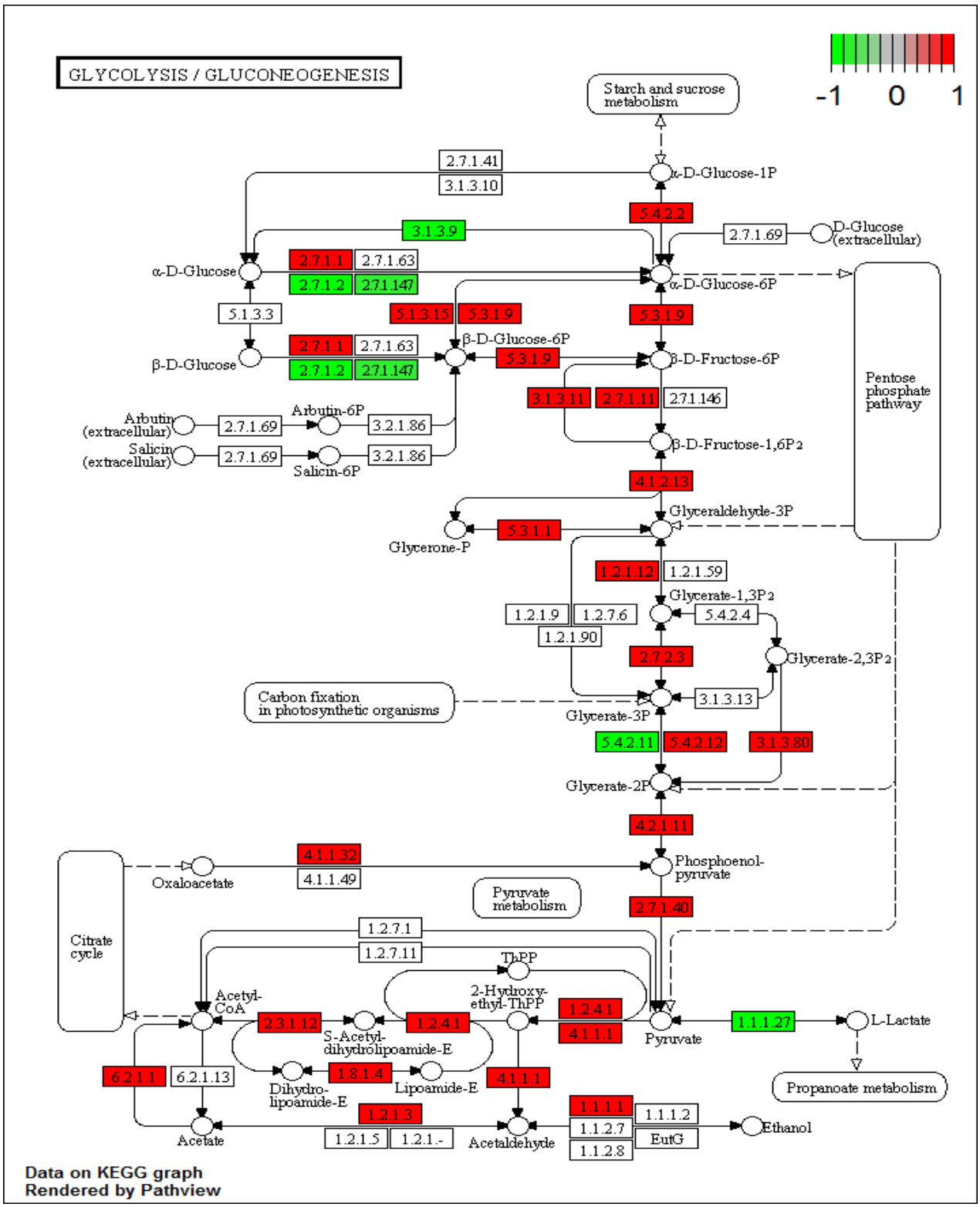

Figure 5. Differential expression of CDS from *A. sculptum* mialomes related to Glycolysis/Gluconeogenesis pathways in response to *R. amblyommii* infection (MI *versus* MNI). In a scale from red to green, red boxes represent up-regulated CDS in *A. sculptum* midguts in response to infection. Green boxes represent down-regulated CDS in response to infection (more expressed in MNI than MI). White boxes indicate genes not represented at the whole *A. sculptum* mialome (MDG).

3.2.8. Proteasome machinery

From the 610 putative genes related to Proteasome machinery, we identified 368 overexpressed genes from midguts of infected ticks in comparion to uninfected ticks (Table 1). This functional class was subcategorized into 15 gene families, as shown at the Spreadsheet S1 and the distribution of the up-regulated genes is as follows: Ubiquitin ligases (158/302); Other ubiquitin related (67/101); Proteasome regulatory machinery (56/71); Ubiquitin hydrolases (29/61); Other proteasome proteins related (26/34); ATPases (13/16); Neddylation related (13/16); Ubiquitin proteases (6/9).

About the more abundance of Ubiquitin ligases in *A. sculptum* mialomes (MI versus MNI). Previous studies evaluating the infection mechanisms of *Rickettsia conorii*, a SFG rickettsiae, on mammalian cells (Martinez et al., 2005) have reported that, the ubiquitination of the host receptor Ku, by an E3 Ubiquitin ligase, c-Cbl, is essential to rickettsial entry and infection (see Figure 3). In this way, the high number of more abundant coding sequences related to ubiquitin ligases at the rickettsial infected mialome (MI) suggests that *R. amblyommii* might to employ a similar mechanism during the infection of *A. sculptum* midguts.

3.3. Protein containing signal peptides

Similar observations by Kotsyfakis et al. (2014) in *I. ricinus* mialomes and sialomes, we have observed several genes previously associated with tick saliva more abundant in *A. sculptum* midguts in response to rickettsial infection (MI versus MNI). This suggests that these proteins may play physiological roles in this organ. Among of those proteins, there are several related to proteins containing kunitz domains. We have identified 29 of those genes at MDG, of which 24 were more abundant in infected *A. sculptum* midguts (MI) in comparison to non-infected dataset (MNI). Beyond its anticoagulant activity in tick saliva (Francischetti et al., 2009), these proteins also play bacteriostatic activity, inhibiting *R. montanensis* infection in *Dermacentor variabilis* (Ceraul et al., 2008; Ceraul et al., 2011). These results suggest that these proteins could play an additional role in the defense machinery of *A. sculptum* midguts against rickettsial infection.

3.4. Transposable elements and unknown products

We have identified 10 transposases families in MDG dataset, such as Gag (23 coding sequences), Harbinger (5) and Tigger transposases (16) , which might play role in horizontal transfer processes between *A. sculptum*-rickettsiae (Spreadsheet S1). In comparison to other ixodidic ticks transcriptomes, we have observed a lower number of unknown genes in our dataset. This observation corroborates the annotation strategy employed for this study, that combines CDS prediction by Augustus, through a *Ixodes scapularis* reference, in association to the annotation

procedures previously employed by several ticks transcriptomes (Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009). This unknown dataset could be an interesting source for further discovery of novel genes, peptides and biological processes in *A. sculptum*.

4. CONCLUSIONS

The rickettsial infection in the *A. sculptum* midguts results an impressive increase in a wide range of cellular processes, such as replication machinery, transcriptions factors, proteins synthesis and modification machinery, as well as, proteins related with vesicle traffic. The cytoskeletal proteins actin, dynamin, clathrin and proteins from Arp 2/3 complex undergo more abundance in *A. sculptum* when infected with *R. amblyommii*. Rickettsial infection promotes expressive more abundance in the repertoire of digestive enzymes of *A. sculptum* midguts, which could be a putative advantage for this vector. In addition, rickettsial infection in *A. sculptum* midguts leds to an overexpression of detoxification proteins and immunity proteins at this organ. All the proteins with recognized participation in rickettsial infection mechanisms in other species, are more abundant in *A. sculptum* midguts.

5. ACKNOWLEDGEMENTS

We would like to thank the Foundation for Research Support of the State of Minas Gerais (FAPEMIG) for financial support; Dr. Marcelo B. Labruna, Faculty of Veterinary Medicine, University of São Paulo, for providing the inoculum of *R. amblyommii;* Dr. Pablo Juan Szabó Mathias, Laboratory of Ixodology Federal University of Uberlandia, for providing the specific pathogen free *A. sculptum* specimens**.**

6. SUPPLEMENTARY DATA (ANEXOS)

Spreadsheet S1 (.DOC)

7. REFERENCES

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. Journal of Molecular Biology, 215: 403–410, 1990.

ANDERSON, J.M.; SONENSHINE, D.E.; VALENZUELA, J.G. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick, *Dermacentor variabilis* (Acari: Ixodidae). BMC Genomics, Nov 20;9:552, 2008.

BOLDBAATAR, D.; SIKALIZYO SIKASUNGE, C.; BATTSETSEG, B.; XUAN, X.; FUJISAKI, K. Molecular cloning and functional characterization of an aspartic protease from the hard tick *Haemaphysalis longicornis*. Insect Biochemistry and Molecular Biology, Jan;36(1):25-36, 2006.

CERAUL, S.M.; CHUNG, A.; SEARS, K.T.; POPOV, V.L.; BEIER-SEXTON, M.; RAHMAN, M.S.; AZAD, A.F. A Kunitz protease inhibitor from *Dermacentor variabilis*, a vector for spotted

fever group rickettsiae, limits *Rickettsia montanensis* invasion. Infection and Immunity, Jan;79(1):321-9, 2011.

CERAUL, S.M.; DREHER-LESNICK, S.M.; MULENGA, A.; RAHMAN, M.S.; AZAD, A.F. Functional characterization and novel rickettsiostatic effects of a Kunitz-type serine protease inhibitor from the tick *Dermacentor variabilis*. Infection and Immunity, Nov;76(11):5429-35, 2008.

CHANG, Y.G.; CARDWELL, M.M.; HERMANAS, T.M.; UCHIYAMA, T.; MARTINEZ, J.J. Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner. Cellular Microbiology, 11(4):629–644, 2009.

DERMAUW, W.; VAN LEEUWEN, T. 2014. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol,45:89-110.

DUSCHER, G.G.; GALINDO, R.C.; TICHY, A.; HUMMEL, K.; KOCAN, K.M.; DE LA FUENTE, J. Glutathione S-transferase affects permethrin detoxification in the brown dog tick, *Rhipicephalus sanguineus*. Ticks and Tick Borne Diseases, Apr;5(3):225-33, 2014.

EDLOW, J.A. Preface: tick-borne diseases, part II. Infectious Disease Clinics of North America, Sep;22(3):xiii-xv, 2008.

EMANUELSSON, O.; BRUNAK, S.; VON HEIJNE, G.; NIELSEN, H. 2007. Locating proteins in the cell using TargetP.; SignalP and related tools. Nat Protoc, 2(4):953-71.

FENG, J.; MEYER, C.A.; WANG, Q.; LIU, J.S.; SHIRLEY LIU X.; ZHANG, Y. 2012. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics. 2012 Nov 1;28(21):2782-8.

FRANCISCHETTI, I.M.; SA-NUNES, A.; MANS, B.J.; SANTOS, I.M.; RIBEIRO, J.M. The role of saliva in tick feeding. Frontiers in Bioscience, Jan 1;14:2051-88, 2009.

FUXELIUS, H.H.; DARBY, A.; MIN, C.K.; CHO, N.H.; ANDERSSON, S.G.The genomic and metabolic diversity of *Rickettsia*. Research in Microbiology, Dec;158(10):745-53, 2007.

GARCIA, G.R.; GARDINASSI, L.G.; RIBEIRO, J.M.; ANATRIELLO, E.; FERREIRA, B.R.; MOREIRA, H.N.; MAFRA, C.; MARTINS, M.M.; SZABÓ, M.P.; DE MIRANDA-SANTOS, I.K.; MARUYAMA, S.R. The sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. Parasites & Vectors, Sep 8;7:430, 2014.

GAVOTTE, L.; MERCER, D.; STOECKLE, J.; DOBSON, S. Costs and benefits of *Wolbachia* infection in immature *Aedes albopictus* depend upon sex and competition level. Journal of Invertebrate Pathology, Nov 105(3):341–346, 2010.

GONG, H.; QIN S.; WAN, X.; ZHANG, H.; ZHOU, Y.; CAO, J.; XUAN, X.; SUZUKI, H.; ZHOU, J. 2014. Immunoglobulin G binding protein (IGBP) from *Rhipicephalus haemaphysaloides*: identification, expression, and binding specificity. Parasitology Research, 113(12):4387-95.

GOUIN, E.; EGILE, C.; DEHOUX, P.; VILLIERS, V.; ADAMS, J.; GERTLER, F.; LI, R.; COSSART, P. The RickA protein of *Rickettsia* conorii activates the Arp2/3 complex. Nature, Jan 29;427(6973):457-61, 2004.

GUO, Y.; RIBEIRO, J.M.; ANDERSON, J.M.; BOUR, S. dCAS: a desktop application for cDNA

sequence annotation. Bioinformatics, May 1;25(9):1195-6, 2009.

HAGLUND, C.M.; WELCH, M.D. 2011. Pathogens and polymers: microbe-host interactions illuminate the cytoskeleton. Journal of the Cell Biology, 195(1):7-17.

HANNEMANN, S.; GAO, B.; GALÁN, J.E. 2013. *Salmonella* modulation of host cell gene expression promotes its intracellular growth. PLoS Pathog, 9(10):e1003668.

HEINZEN, R.A.; GRIESHABER, S.S.; VAN KIRK, L.S.; DEVIN, C.J. 1999. Dynamics of actin-based movement by *Rickettsia rickettsii* in vero cells. Infect Immun, 67(8):4201-7.

HOFF, K.J.; LANGE, S.; LOMSADZE, A.; BORODOVSKY, M.; STANKE, M. BRAKER. 2015. Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics, Nov 11. pii: btv661

HUANG, H.; LU-BO, Y.; HADDAD, G.G. A *Drosophila* ABC transporter regulates lifespan. PLoS Genetics, Dec 4;10(12):e1004844, 2014.

JENG, R.L.; GOLEY, E.D.; D'ALESSIO, J.A.; CHAGA, O.Y.; SVITKINA, T.M.; BORISY, G.G.; HEINZEN, R.A.; WELCH, M.D. A Rickettsia WASP-like protein activates the Arp2/3 complex and mediates actin-based motility. Cellular Microbiology, Aug;6(8):761-9, 2004.

KARIM, S.; RIBEIRO, J.M. An Insight into the Sialome of the Lone Star Tick, *Amblyomma americanum*, with a Glimpse on Its Time Dependent Gene Expression. PLoS One, Jul 1;10(7):e0131292, 2015.

KARIM, S.; SINGH, P.; RIBEIRO, J.M. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. PLoS One, 6(12):e28525, 2011.

KOTSYFAKIS, M.; KOPÁČEK, P.; FRANTA, Z.; PEDRA, J.H.; RIBEIRO, J.M. Deep Sequencing Analysis of the *Ixodes ricinus* Haemocytome. PLOS Neglected Tropical Diseases, May 13;9(5):e0003754, 2015.

KOTSYFAKIS, M.; SCHWARZ, A.; ERHART, J.; RIBEIRO, J.M. Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. Scientific Reports, Mar 13;5:9103, 2015.

KREMER, N.; VORONIN, D.; CHARIF, D.; MAVINGUI, P.; MOLLEREAU, B.; VAVRE, F. 2009. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. PLoS Pathog, Oct;5(10):e1000630.

KROGH, A.; LARSSON, B.; VON HEIJNE, G.; SONNHAMMER, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol, 305(3):567-80.

KURTTI, T.J.; MUNDERLOH, U.G.; ANDREADIS, T.G.; MAGNARELLI.; L.A.; MATHER.; T.N. 1996. Tick cell culture isolation of an intracellular prokaryote from the tick *Ixodes scapularis*. Journal Invertebrate Pathology, 67:318–21.

LEE, M.R.; JEON, T.J. 2012. Cell migration: regulation of cytoskeleton by Rap1 in *Dictyostelium discoideum*. Journal of Microbiology, 50(4):555-61.

LUO, W.; BROUWER, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics, Jul 15;29(14):1830-1, 2013.

MAHAPATRA, S.; AYOUBI, P.; SHAW, E.I. *Coxiella burnetii* Nine Mile II proteins modulate gene expression of monocytic host cells during infection. 2010. BMC Microbiology, 10:244.

MARTINEZ, J.J.; COSSART, P. 2004. Early signaling events involved in the entry of *Rickettsia conorii* into mammalian cells. Journal of Cell Science, 117(21):5097-106.

MARTINEZ, J.J.; COSSART, P. Early signaling events involved in the entry of *Rickettsia conorii* into mammalian cells. Journal fo Cell Science, 1;117(Pt 21):5097-106, 2004.

MARTINEZ, J.J.; SEVEAU, S.; VEIGA, E.; MATSUYAMA, S.; COSSART, P. Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for *Rickettsia conorii*. Cell, Dec 16; 123(6): 1013-1023, 2005.

MARUYAMA, S.R.; ANATRIELLO, E.; ANDERSON, J.M.; RIBEIRO, J.M.; BRANDÃO, L.; VALENZUELA, J.G.; FERREIRA, B.R.; GARCIA, G. R.; SZABÓ, M.P.J.; PATEL, P.; BISHOP, R.; MIRANDA-SANTOS, I.K.F. The expression of genes coding for distinct types of glycine-rick proteins varies according to the biology of three metastriate ticks, *Rhipicephalus* (*Boophilus*) *microplus*, *Rhipicephalus microplus* and *Amblyomma cajennense*. BMC Genomics, 11:363, 2010.

MORETI, R.; PERRELLA, N.N.; LOPES, A.R. 2013. Carbohydrate digestion in ticks and a digestive α-L-fucosidase. Journal of Insect Physiology, 59(10):1069-75.

NASSER, J.T.; LANA, R.C.; SILVA, C.M.; LOURENÇO, R.W.; DA CUNHA E SILVA, D.C.; DONALÍSIO, M.R. Urbanization of Brazilian spotted fever in a municipality of the southeastern region: epidemiology and spatial distribution. Revista Brasileira de Epidemiologia, Apr-Jun;18(2):299-312, 2015.

NASSER, J.T.; LANA, R.C.; SILVA, C.M.; LOURENÇO, R.W.; DA CUNHA E SILVA, D.C.; DONALÍSIO, M.R. Divergent environmental preferences and areas of sympatry of tick species in the *Amblyomma cajennense* complex (*Ixodidae*). International Journal for Parasitology, Dec;44(14):1081-9, 2014.

NAVA, S.; BEATI, L.; LABRUNA, M.B.; CÁCERES, A.G.; MANGOLD, A.J.; GUGLIELMONE, A.A. Reassessment of the taxonomic status of *Amblyomma cajennense* () with the description of three new species, *Amblyomma tonelliae* n. sp., *Amblyomma interandinum* n. sp. and *Amblyomma patinoi* n. sp., and reinstatement of *Amblyomma mixtum*, and *Amblyomma sculptum* (Ixodida: *Ixodidae*). Ticks and Tick Borne Diseases, Apr;5(3):252-76, 2014.

NELSON, M.M.; JONES, A.R.; CARMEN, J.C.; SINAI, A.P.; BURCHMORE, R.; WASTLING, J.M. 2008. Modulation of the host cell proteome by the intracellular apicomplexan parasite *Toxoplasma gondii*. Infect Immun. 76(2):828-44.

NIKOH, N.; HOSOKAWA, T.; MORIYAMA, M.; OSHIMA, K.; HATTORI, M.; FUKATSU, T. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc Natl Acad Sci U S A. 111(28):10257-62.

PETCHAMPAI, N.; SUNYAKUMTHORN, P.; BANAJEE, K.H.; VERHOEVE, V.I.; KEARNEY, M.T.; MACALUSO, K.R. 2015. Identification of host proteins involved in rickettsial invasion of tick cells. Infect Immun. 83(3):1048-55.

PETCHAMPAI, N.; SUNYAKUMTHORN, P.; GUILLOTTE, M.L.; THEPPARIT, C.; KEARNEY, M.T.; MULENGA, A.; AZAD, A.F.; MACALUSO, K.R. Molecular and functional characterization of vacuolar-ATPase from the American dog tick *Dermacentor variabilis*. Insect Molecular Biology,

Feb;23(1):42-51, 2014.

PETCHAMPAI, N.; SUNYAKUMTHORN, P.; GUILLOTTE, M.L.; VERHOEVE, V.I.; BANAJEE, K.H.; KEARNEY, M.T.; MACALUSO, K.R. 2014. Novel Identification of *Dermacentor variabilis* Arp2/3 complex and its role in rickettsial infection of the arthropod vector. PLoS One, 14;9(4), 2014.

PICHU S.; YALCIN, E.B.; RIBEIRO, J.M.; KING, R.S.; MATHER, T.N. 2011. Molecular characterization of novel sulfotransferases from the tick, *Ixodes scapularis*. BMC Biochemistry, 12:32.

PICHU, S.; YALCIN, E.B.; RIBEIRO, J.M.; KING, R.S.; MATHER, T.N. Molecular characterization of novel sulfotransferases from the tick, *Ixodes scapularis*. BMC Biochemistry, Jun 27;12:32, 2011.

PRINCE, L.; KORBAS, M.; DAVIDSON, P.; BROBERG, K.; RAND, M.D. 2014. Target organ specific activity of *Drosophila* MRP (ABCC1) moderates developmental toxicity of methylmercury. Toxicology Science, 140(2):425-35.

PRINCE, L.; KORBAS, M.; DAVIDSON, P.; BROBERG, K.; RAND, M.D. Target organ specific activity of drosophila MRP (ABCC1) moderates developmental toxicity of methylmercury. Toxicological Sciences, Aug 1;140(2):425-35, 2014.

REED, S.C.; SERIO, A.W.; WELCH, M.D. 2012. *Rickettsia parkeri* invasion of diverse host cells involves an Arp2/3 complex, WAVE complex and Rho-family GTPase-dependent pathway. Cell Microbiology, 14(4):529-45.

RENESTO, P.; OGATA, H.; AUDIC, S.; CLAVERIE, J.M.; RAOULT, D. Some lessons from *Rickettsia* genomics. FEMS Microbiology Reviews, Jan 29(1):99-117, 2005.

ROLANDO, M.; ESCOLL, P.; NORA, T.; BOTTI, J.; BOITEZ, V.; BEDIA, C.; DANIELS, C.; ABRAHAM, G.; STOGIOS, P.J.; SKARINA, T.; CHRISTOPHE, C.; DERVINS-RAVAULT, D.; CAZALET, C.; HILBI, H.; RUPASINGHE, T.W.; TULL, D.; MCCONVILLE, M.J.; ONG, S.Y.; HARTLAND, E.L.; CODOGNO, P.; LEVADE, T.; NADERER, T.; SAVCHENKO, A.; BUCHRIESER, C. 2016. *Legionella pneumophila* S1P-lyase targets host sphingolipid metabolism and restrains autophagy. Proc Natl Acad Sci USA, 113(7):1901-6.

SCHWARZ, A.; TENZER, S.; HACKENBERG, M.; ERHART, J.; GERHOLD-AY, A.; MAZUR, J.; KUHAREV, J.; RIBEIRO, J.M.; KOTSYFAKIS, M. 2014. A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinu*s midgut and salivary glands during early attachment and feeding. Molecular & Cellular Proteomics.; 13(10):2725-35.

SERBUS, L.R.; WHITE, P.M.; SILVA, J.P.; RABE, A.; TEIXEIRA, L.; ALBERTSON, R.; SULLIVAN, W. The impact of host diet on *Wolbachia* titer in *Drosophila*. PLoS Pathog, Mar 31;11(3):e1004777, 2015.

SOJKA, D.; FRANTA, Z.; HORN, M.; CAFFREY, CR.; MAREŠ, M.; KOPÁČEK, P. 2013. New insights into the machinery of blood digestion by ticks. Trends in Parasitology, 29(6):276-85.

SONNHAMMER, E.L.; VON HEIJNE, G.; KROGH, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol, 6:175-82.

STANKE, M.; KELLER, O.; GUNDUZ, I.; HAYES, A.; WAACK, S.; MORGENSTERN, B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research, Jul 1;34:W435-9, 2006.

STANKE, M.; MORGENSTERN, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research, Jul 1;33:W465-7, 2005.

STANKE, M.; STEINKAMP, R.; WAACK, S.; MORGENSTERN, B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Research, Jul 1;32:W309-12, 2004.

TAKECHI, R.; GALAY, R.L.; MATSUO, T.; MAEDA, H.; KUSAKISAKO, K.; TALACTAC, M.R.; MOCHIZUKI, M.; FUJISAKI, K.; TANAKA, T. 2016. Role of the tumor necrosis factor receptor-associated factor-type zinc finger domain containing protein 1 (TRAFD1) from the hard tick *Haemaphysalis longicornis* in immunity against bacterial infection. Ticks Tick Borne Dis. 7(1):36-45.

THEPPARIT, C.; BOURCHOOKARN, A.; PETCHAMPAI, N.; BARKER, S.A.; MACALUSO, K.R. Interaction of *Rickettsia felis* with histone H2B facilitates the infection of a tick cell line. Microbiology, Sep;156(Pt 9):2855-63, 2010.

TOZZI, M.G.; CAMICI, M.; MASCIA, L.; SGARRELLA, F.; IPATA, P.L. Pentose phosphates in nucleoside interconversion and catabolism. FEBS Journal, Mar;273(6):1089-101, 2006.

VALENZUELA, J.G.; FRANCISCHETTI, I.M.; PHAM, V.M.; GARFIELD, M.K.; MATHER, T.N.; RIBEIRO, J.M. Exploring the sialome of the tick *Ixodes scapularis*. Journal of Experimental Biology, Sep;205(Pt 18):2843-64, 2002.

VERMA, S.; MOHAPATRA, G.; AHMAD, S.M.; RANA, S.; JAIN, S.; KHALSA, J.K.; SRI-KANTH, C.V. 2015. Salmonella Engages Host MicroRNAs To Modulate SUMOylation: a New Arsenal for Intracellular Survival. Mol Cell Biol. 35(17):2932-46.

VILLARREAL, A.M.; ADAMSON, S.W.; BROWNING, R.E.; BUDACHETRI, K.; SAJID, M.S.; KARIM, S. Molecular characterization and functional significance of the Vti family of SNARE proteins in tick salivary glands. Insect Biochemistry and Molecular Biology, May;43(5):483-93, 2013.

WARD, P.S.; THOMPSON, C.B. 2012. Signaling in control of cell growth and metabolism. Cold Spring Harb Perspect Biol. 4(7):a006783.

WONG, Z.; BROWNLIE, J.C.; JOHNSON, K.N. Oxidative stress correlates with *Wolbachia*-mediated antiviral protection in *Wolbachia-Drosophila* associations. Applied and Environmental Microbiology, May 1;81(9):3001-5, 2015.

CAPÍTULO 3

THE FIRST FUNCTIONAL GENOMICS OF *AMBLYOMMA SCULPTUM* TICK OVARIES UNDER CONDITIONS OF INFECTION AND NON-INFECTION WITH *RICKETTSIA AMBLYOMMII*

The first functional genomics of *Amblyomma sculptum* tick ovaries, under conditions of infection and non-infection with *Rickettsia amblyommii*

Higo Nasser Santanna MOREIRA[1], Talles MACIEL[1], Juliana CARRIZO[1], Rafael Mazioli BARCELOS[1], Maria Marlene MARTINS[2], Paulo LIMA[2], Adriano Soares MORENO[2], Donald Hugh BOUYER[3], Cláudio MAFRA[1¥]

[1] Biochemistry and Molecular Biology Department, Federal University of Vicosa, Brazil; [2] Veterinary Department, Federal University of Uberlandia, Brazil; [3] Pathology Department, University of Texas Medical Branch, USA.

[¥] mafra@ufv.br

1. INTRODUCTION

Ticks are one of the most versatile blood-sucking arthropods group (Order: Acari, Class: Arachnida) being able to parasite a wide range of vertebrates hosts, including humans (Szabó et al., 2013). Because of its obligatory hematophagy and amazing capacity of reproduction in a variety of environments, the population control of these ectoparasites stands out as one of the main issues to be solved by public health agencies around the world (Kabowiak et al., 2015).

In the context of the American continent, stands out the *Amblyomma sculptum* tick, a species known for aggressive and promiscuous parasitism, being incriminated as one of the main vectors of Spotted Fever Group (SFG) rickettsiae in a large territory ranging from the Southern United States to Nothern Argentina. Among SFG rickettsiae, *Rickettsia rickettsii*, the causative agent of the Brazilian Spotted Fever (BSF), one of the most lethal illness in that country. Besides the capacity of play role as an efficient vector of SFG rickettsiae, *A. sculptum* ticks are also incriminated to play role on the amplification and maintenance of *Rickettsia* spp., at the environment, through the transovarial transmission processes, in which a female can transfer the bacteria to all her offspring (Labruna et al., 2011). Each female can oviposit from 5,000 to 10,000 eggs (Barros-Battesti et al., 2006), so the understanding of the molecular mechanism and genes repertoire in this processes rises as an issue of the first importance for the public health.

In the context of the recent advances of ticks transcriptomics, the emergence of next generation sequencing technologies, such as 454 pyrosequencing, Ion torrent PGM and Illumina Hiseq, has provided a lot of answers about issues around the ticks-host pathogens interface, especially regarding repertoire of genes from salivary glands and midguts (Chmelar et al., 2015). However, despite of the epidemiological relevance of *A. sculptum*, the molecular aspects that drive the interface between *A. sculptum* ovaries and rickettsiae remains not well understood until this moment. In addition, until this work, there is no transcriptomic data about the ovarian tissues of *A.*

sculptum, which limits the perspectives about the development of new strategies of biological control for this ectoparasite.

In order to shed light on molecular aspects of the biological processes that drive ovarian physiology of *A. sculptum* and its possible relationships with transovarian transmission of SFG rickettsiae, the purpose of this work was to construct a deep transcriptomic catalogue from *A. sculptum* ovaries to evaluate the standard conditions of non-infection and infection with *R. amblyommii*. For this purpose, the transcriptomes regarding these conditions were sequenced, the genes were predicted, functionally annotated, and the patterns of gene expression that drive these physiological aspects of *A. sculptum* ovarian transcriptomes analyzed.

2. METHODS

2.1. Obtainment of the ticks and treatment imposition

A. sculptum SPF specimens (Specific Pathogen Free) were maintained in a B.O.D. incubator according previously described by Fernandes et al., (2005). For the achievement of ovarian transcriptomes, a total of 150 females were divided into two cohorts groups: A subset of 75 females for the infection treatment (OI) and 75 females for non-infection treatment (ONI). Previously the feeding step, the OI group was artificially infected, through microinjection application with culture supernatant of VERO cells previously infected with *R. amblyommii*, according previous work (Melles et al., 1999). For control, the ONI group underwent the same procedure than OI group, but with only VERO cell culture supernatant. The ticks were maintained at B.O.D. incubator at 27°C for 48 hours, for rickettsial proliferation. After this period, in order to confirm the rickettsial infection, a sample of ticks for both were tested by Real Time PCR assay with primers CS-5 (forward) and CS-6 (reverse), which are specific for a 147-bp fragment from citrate synthase gene (gltA) belonging for all genes *Rickettsia* species (Labruna et al., 2007). The Real Time PCR reactions were performed for 25 µl per reaction, which contain 12.5 uL Maxima SYBR Green/ROX qPCR Master Mix, according manufactor recommendations, 1,25 µl of each primer at 15 µM, 1,5 µl of template DNA from each tick or VERO cell culture for detections of DNA belonging to any specie from genus *Rickettsia* and 8,5 µl of molecular-grade water. All the Real Time PCRs were conducted using Applied Biosystems® StepOne™ Real-Time PCR System according the following conditions: 1 cycle at 95°C for 5 min, followed by 50 cycles of 15 s at 95°C, 30 s at 53°C, and 30 s at 60°C. As negative control for the reactions, we used the same molecular-grade water from the reactions and, 1,5 uL of DNA of VERO cells infected with *R. amblyommii,* for the positive control.

After infection confirmation, the ticks were fed incubate at BOD incubator for 48 hours in order to provide rickettsial proliferation and then, fed on rabbits during 8-10 days, in the presence of male specimens, for fecundation and complete engorgement. After this period, the specimens were

dissected for recovering of internal organs, which were maintained into RNA later (Ambion®, USA) at -80°C for the next step of RNA extraction and cDNA synthesis.

2.2. Extraction of total RNA, cDNA libraries and NGS sequencing

The total RNA samples from the ovaries groups OI and ONI were extracted through the kit RNAqueous® RNA extraction kit (Ambion®, USA) kit per manufacter's guidelines. The quality of RNA samples from OI and ONI were measured through NanoDrop 2000 (ThermoScientificTM, USA) and Agilent Bioanalyser 2100 (Agilent Technologies, Santa Clara, USA) at the Center for Biomolecules at Federal University of Vicosa. The mRNA from total RNA was purified using Sera-mag Magnetic Oligo (dT) Beads (Thermo Fisher Scientific. USA). The cDNA libraries were synthesized through the Clontech SmartPCR cDNA kit (Clontech Laboratories, Inc, Mountain View, CA, USA), according manufacturer recommendations.The temperature cycle for the synthesis reaction was perfomed at the StepOne Termcycler (Applied Biosystems) according the follow: 6 min at 68°C, followed by 15 cycles of 10 min at 70 °C. The purification of the cDNA were performed with the aid of magnetic beads from the kit. Sequences of cDNA smaller than 200bp were selected through electrophoresis and amplified by PCR using primers specific to adaptor sequences from the kit. The pools of cDNA were purified wih aid of QiaQuick PCR extraction kit (CA, USA) and eluted with elution buffer EB from the same kit and subsequently its quality and purity were measured using Bionalayser Agilent 2100, according manufacturer protocols. After this step, the cDNA libraries from A. sculptum ovaries were sequenced using Illlumina HiSeq™ 2000 platform, according the Beijing Genome Institute recommendations.

2.3. *De Novo* assembly and gene prediction

We performed the construction of a reference transcriptome for *A. sculptum* internal organs assembling the both ovarian transcriptomes of this study together other seven dataset derived for internal organs of *A. sculptum*, as well, such as: (1) 53,850,855 reads from midguts non-infected (MNI) being 53,006,356 from Illumina HiSeq and 844,529 from PGM Ion Torrent. (2) 49,209,818 Illumina reads sequenced by derived from Midguts Infected (MI); (3) Four 454 pyrosequencing datasets from *A. sculptum* salivary glands obtained in a previous work (Garcia et al., 2014). These 9 transcriptomic datasets were assembled into contigs through CLC Genomics Workbench 8.0 package (CLC Bio, Denmark) which yields a total of 460,445 contigs, with length ranging from 100 to 14,579bp, with 222bp of average length. This contig dataset was submitted by gene prediction analysis through the Augustus algorithm (Stanke et al., 2004), which took into consideration the genome of *Ixodes scapularis* from GenBank as a template. A total of 27,308 CDS were identified and this dataset designated whole transcriptome of *A. sculptum* internal organs. After the mapping

step of OI and ONI Illumina reads against the 27,308 reference, we identified a total of 20,230 CDS expressed at least at one of the both ovarian transcriptomes of this study (OI and ONI), been this dataset designated whole transcriptome of *A. sculptum* ovaries (OVA).

2.4. Functional annotation and gene expression analysis

The 20,230 ovarian CDS (OVA) were compared against homologous sequences deposited in several databases, as follow: (1) the non-redundant protein database (NR) of the National Center for Biology Information (NCBI); (2) Gene Ontology; (3) SwissProt; (4) UNIPROT; (5) NCBI-Acari; (6) CDD; (7) COG; (8) KOG; (9) PFAM; (10) SMART; (11) UNIPROT-Acari; (12) *Bombyx* mori; (13) *Chelicerata*; (14) rickettsiae; (15) *Amblyomma americanum*; (16) *A. cajennense* (*A. sculptum* complex); (17) *A. maculatum*; (18) *A. parvum*; (19) A. *triste*; (20) *Argas*; (21) *Dermacentor*; (22) *I. scapularis*; (23) *I. ricinus*; (24) *Rhipicephalus* and (25) *Haemaphysalis*. BLAST included BLASTp, BLASTx, rpsBLAST and BLASTn algorithms (Altschul et al., 1990). For the annotation processes, we adopted an E-value cut-off of 1E-04, taking into consideration key words from the main databases employed, as approach previously employed for the annotation of a several blood-sucking arthropods transcriptomes (Ribeiro et al., 2015; Garcia et al., 2014). As a complementary approach for the annotation, we performed the determination of the presence of signal peptide and transmembrane domains at the 20,230 CDS through the employment of *SignalP* and TMHMM softwares, respectively. The best match for each CDS were annotated according key-words from the BLAST sorted into 5 general classes: Housekeeping (H); Secreted (S); Transposable elements (TE); Unknown (U) and Viral related (V). A deeper annotation was performed into each one of these classes, in order to get more specific information about the ovarian genes, in a manner similar to that done for other tick transcriptomes previously studied (Ribeiro et al., 2015, Karim et al., 2015; Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009). The comparison about gene expression between OI and ONI transcriptomes was performed by RPKM normalization and Log2 calculation, and significance differences determined through GFOLD algorithm (Feng et al., 2012).

3. RESULTS AND DISCUSSION

3.1. Overall features of the ovaries transcriptomes of *A. sculptum*

Two cDNA libraries derived from ovaries of female *A. sculptum* ticks fed for 7-10 days, under the conditions of non-infection (ONI) and infection with *R. amblyommii* (OI) were sequenced. A total of 45,417,679 reads were generated from ONI, while 46,573,835 reads were derived from OI sequencing, both datasets with 90nt of average size. This feeding period was chosen because represents the final stage of the blood meal period and the beginning of oogenesis, when the female detaches from the host for oviposition (Drummond & Whetstone, 1975; Lopes et al., 2008)

In order to construct a reference transcriptome for *A. sculptum* to predict a higher number of genes from ovaries and give support for further annotations, we performed the *de novo* assembly of OI and ONI datasets together with the other transcriptomics data generated from other *A. sculptum* organs, described as follows: (1) 49,209,818 reads from Midguts Infected with *R. amblyommii* (MI); (2) 53,006,356 reads from Midguts Non-Infected (MNI); (3) an additional MNI dataset of 844,529 reads generated by PGM Ion Torrent; (4) a dataset of 703,210 pyrosenquencing reads derived from four *A. sculptum* sialomes representing unfed and semi-engorged ticks, previously published (Garcia et al., 2014). This universe of 200 million reads were assembled into 460,445 contigs through CLC Genomics Workbench (http://www.clcbio.com), with length ranging from 100 to 14,579pb and average size of 222pb. This resultant contig dataset was submitted to gene prediction analysis by Augustus program (http://augustus.gobics.de), considering the most recent structural genome of *Ixodes scapularis* deposited on GenBank (http://metazoa.ensembl.org/Ixodes_scapularis/) and recently published (Gulia-Nuss et al., 2016) resulting in a total of 27,308 reference CDS dataset. The mapping of the two reads datasets from *A. sculptum* ovaries (OI and ONI) against the 27,308 reference CDS allowed the identification of 21,230 CDS expressed at *A. sculptum* ovaries. This subset of CDS related to the total number of CDS expressed at both ovaries treatments was designated “whole ovarian transcriptome of *A. sculptum*” (OVA). Individually, the mapping analysis of the reads allowed the identification of 19,944 CDS derived for the infected ovaries transcriptome (OI), while 19,935 CDS were expressed at non-infected ovaries transcriptome (ONI).

3.2. Functional annotation and effect of rickettsial infection on gene expression of *A. sculptum* ovaries

For the functional annotation, all the 21,230 CDS from OVA were compared against homologous sequences from several databases (see Methods), as well as, arthropods transcriptomic datasets deposited on Genbank and Uniprot (see Methods), through BLASTn, BLASTx, BLASTp

and rpsBLAST algorithms (Altschul et al., 1990). The final annotation were defined adopting an E-value cut-off of 1E-04, and key-words from the BLAST results of the main public databases, such as KOG, COG, GO, SwissProt and others. The prediction of signal peptides, transmembrane domains, addressing subcellular localization and related metabolic pathways were performed as an auxiliary approach for annotation, through the programs SignalP, TMHMM, TargetP and KEGG Pathview, respectively (Emanuelsson et al., 2007; Sonnhammer et al., 1998; Krogh et al., 2001; Luo & Brouwer, 2013). This approach has been employed for the annotation of a several ticks transcriptomes, as previously described (Ribeiro et al., 2015, Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009).

Using this approach, we have classified the OVA transcriptome CDS into 29 functional classes previously described for a several ixodidic ticks transcriptomes (Ribeiro et al., 2015, Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009), as shown at the Figure 1 and Table 1. A total of 25 functional classes were related with housekeeping genes (H), of which stand out those related to signal transduction, transporters and nuclear regulation as some of the those with higher numbers of coding sequences (Figure 1). A total of 1,953 coding sequences, that presented signal peptides detected by SignalP analysis, was classified as secreted proteins (S). The three remaining functional classes were 194 CDS related to transposable elements (TE), 14 CDS related to viral transcripts and 872 CDS classified as unknown, which represents those coding sequences that did not present homology with any sequences on the databases searched through BLAST algorithms. For the better understanding about the gene families expressed into each one of 29 function classes, we subcategorized each class into functional subcategories or gene families, as approach previously employed by tick transcriptomes annotation (Ribeiro et al., 2015, Karim et al., 2015; Schwarz et al., 2014, Garcia et al., 2014; Calvo et al., 2009; Guo et al., 2009). These results can be explored at the Spreadsheet S1.

3.3. Housekeeping genes

3.3.1. Extracellular matrix and cytoskeletal machinery

3.3.1.1. Extracellular matrix

We annotated 760 coding sequences related to extracellular matrix proteins from the whole ovarian transcriptome of *A. sculptum* females (OVA), of which 711 were expressed in infected ovaries (OI), whereas 723 presented expression at non-infected ovaries. Among the most represented families are the glycoproteins, with 61 genes, followed by cuticle related proteins (50), cell adhesion antigens (44), fibronectins (21) and mucins (21). All the functional families into extracellular matrix class is annotated at the Spreadsheet S1.

Regarding the effect of *R. amblyommi* infection on ovaries, we observed practically the same proportion between up and down regulated genes, with 131 and 128 genes respectively (Spreadsheet S1). Among the up-regulated genes, are those related to extracellular matrix kinases (4), catenins (3) and integrins (1). These host proteins have been previously reported to play role on the rickettsial infection mechanisms in both mammalian cells (Martinez & Crossard, 2004; Hillman et al., 2013) and invertebrate cells, in interactions with Arp 2/3 complex and tyrosine kinase (Petchampai et al., 2015). The overexpression of these proteins suggests a similar mechanism for *R. amblyommi* infection and establishment in the *A. sculptum* ovaries.

In contrast, we observed the downregulation of a several cuticle (16) and transmembrane proteins (51) in infected ovaries in comparison to the non-infected ovaries. Cuticle proteins have been annotated in a several ixodidic ticks transcriptomes and these proteins were described to play a variety of roles in ticks, such as giving waterproofing to the eggshell (Hackman, 1982). Once the less expression of these proteins could have an effect on eggshell structure and permeability , these results could be related with the high rates of transovarian infection and the larvae mortality from infected females of other ixodidic ticks, by increase of eggshell permeability to water, such as *D. andersoni* infected by *R. rickettsii* (Niebylski et al., 1999).

3.3.1.2. Cytoskeletal machinery

We have annotated 1,107 coding sequences from OVA, related to cytoskeletal related proteins, 1,039 were expressed in both OI and ONI. The most expressed cytoskeletal components in the OVA were those related with cell cycle processes (93), followed by tubulins (92), kinesin and actin binding proteins (55). The abundance of transcripts related to cell cycle processes could be related with the great increase of this organ at 7-10 day of blood-feeding, as related previously by Denardi et al. (2004). The complete list of annotated cytoskeletal proteins families is shown at the Spreadsheet S1.

In our analysis the differentially expressed genes in response to rickettsial infection (OI versus ONI), we observed the overexpression of 207 cytoskeletal genes, while 151 were down regulated (Spreadsheet S1). Among the overexpressed genes, it is worthy to note those related to Arp 2/3 complex, that has been described to play role during the infection mechanism of *Dermacentor variabilis* by *Rickettsia montanensis* (Petchampai et al., 2015) which suggests its possible involvement in the putative mechanism of infection and establishment of *A. sculptum* ovaries by *Rickettsia amblyommii*. Other host proteins previously described to play role at rickettsial infection on mammals cells, such as several genes for actin G (Martinez & Crossard, 2004) also have been overexpressed at *A. sculptum* ovaries in response to rickettsial infection (Spreadsheet S1), corroborating this hypothesis.

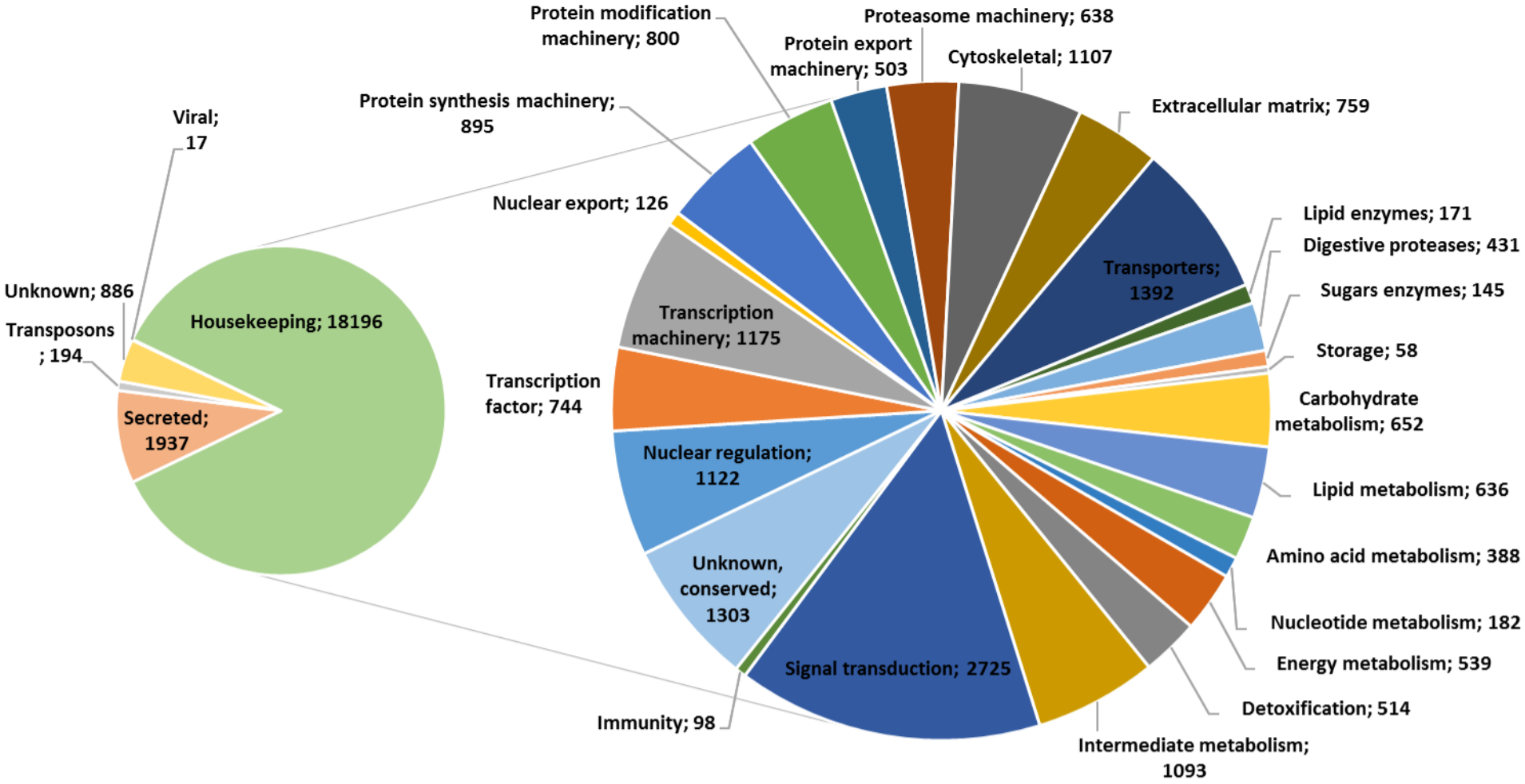

Figure 1. Functional annotation of the 21,230 CDS derived from whole *A. sculptum* ovaries transcriptomes (OVA).

Table 1. General distribution of reads in coding sequences from the *A. sculptum* ovaries transcriptomes

	OVARY INFECTED				OVARY NON-INFECTED			
FUNCTIONAL CLASS	**CDS**	**Reads**	**Reads/CDS**	**Reads***	**CDS**	**Reads**	**Reads/CDS**	**Reads***
HOUSEKEEPING	**17,142**	**16,218,534**	**946.1285**	**70.273**	**17,092**	**16,569,192**	**969.41**	**74.702**
Extracellular matrix	711	1,550,418	2180.62	6.72	723	1,352,004	1869.99	6.10
Cytoskeletal	1,039	924,756	890.04	4.01	1,039	908,295	874.20	4.10
Nuclear regulation	1,079	299,426	277.50	1,30	1,055	395,644	375.02	1.78
Transcription factor	707	281,565	398.25	1.22	701	301,980	430.78	1.36
Transcription machinery	1,123	1,429,080	1272.56	6.19	1,123	1,419,666	1264.17	6.40
Nuclear export	113	40,923	362.15	0.18	118	54,429	461.26	0.25
Protein synthesis machinery	852	3,965,210	4654.00	17.18	853	4,366,751	5119.29	19.69
Protein modification machinery	742	484,968	653.60	2.10	744	497,594	668.81	2.24
Protein export machinery	476	334,685	703.12	1.45	475	344,113	724.45	1.55
Signal transduction	2,558	1,028,357	402.02	4.46	2,493	1,049,839	421.11	4.73
Storage/Vitellogenesis	49	457,779	9342.43	1.98	55	433,286	7877.93	1.95
Transporters	1,297	338,970	261,35	1.47	1,310	363,243	277.28	1.64
Digestive enzymes	723	694,797	960.99	3.01	713	717,695	1006.58	3.24
Carbohydrate metabolism	604	644,829	1067.60	2.79	595	677,934	1139.38	3.06
Lipid metabolism	595	414,012	695.82	1.79	598	434,026	725.80	1.96
Amino acid metabolism	372	192,108	516.42	0.83	364	214,575	589.49	0.97
Nucleotide metabolism	173	205,844	1189.85	0.89	173	214,984	1242.68	0.97
Metabolism, energy	524	661,971	1263.30	2.87	525	709,432	1351.30	3.20
Intermediate metabolism	1,022	455,915	446.10	1.98	1,039	470,864	453.19	2.12
Proteasome machinery	611	496,764	813.03	2.15	602	543,089	902.14	2.45
Detoxification	504	1,027,444	2038.58	4.45	513	847,559	1652.16	3.82
Immunity	90	14,901	165.57	0.06	96	18,893	196.80	0.09
Unknown, conserved	1,178	273,812	232.44	1.19	1,185	233,297	196.88	1.05
SECRETED	**1,831**	**6,654,094**	**3634.13**	**28.83**	**1,890**	**5,437,826**	**2877.16**	**24.52**
TRANSPOSABLE ELEMENTS	**182**	**7,850**	**43.13**	**0.03**	**179**	**8,921**	**49.84**	**0.04**
UNKNOWN	**775**	**194,086**	**250.43**	**084**	**761**	**158,928**	**208.84**	**0.72**
VIRAL	**13**	**4,872**	**374.77**	**0.02**	**13**	**5,504**	**423.38**	**0.02**
TOTAL	**19,943**	**23,079,436**		**100**	**19,935**	**22,180,371**		**100**

Among the down regulated genes, stands out 17 coding sequences related to cell cycle proteins (Spreadsheet S1). The down regulation of cytoskeletal proteins that orchestrate cell division and, as consequence, tissue growth, suggest limitation of tissue growth due to the infection in OI, since this organ still increases at 7-10 day of blood feeding in non-infectious conditions (Denardi et al., 2004). Therefore, the down-regulation of these genes in response to infection (OI versus ONI) could be a negative effect on the ovarian tissue. Indeed, previous works have related that *R. rickettsii* infection affects negatively the oviposition processes at *A. aureolatum*, resulting in a lower egg mass (Labruna et al., 2011), what could explain these RNAseq results for *A. sculptum* ovaries infected with *R. amblyommii*, also a SFG rickettsiae (Labruna et al., 2007).

3.3.2. Control machinery of the gene expression

3.3.2.1. Nuclear regulation

We annotated a several functional classes related to control machinery of gene expression in the whole transcriptome of *A. sculptum* ovaries (OVA) as shown at Table 1 and Spreadsheet S1. A subset of 1,122 CDS were annotated as related to nuclear regulation processes, with 1,079 being expressed at OI, while 1,055 were expressed at non-infected conditions (ONI). Among the proteins families represented by this class at OVA, the most prominent are the DNA polymerases (140 CDS) and helicases (140), 98 coding sequences related to DNA repair, chromatin structural proteins (56); and other related to DNA and genome replication (Spreadsheet S1). This was expected, since the late development – and higher cell division activity - of *A. sculptum* ovaries was been previously reported (Denardi et al., 2004).

However, when considering the comparison between infected (OI) and non-infected (ONI) datasets, we have observed the down-regulation of 282 genes in response to rickettsial infection (OI versus ONI). In contrast, only 98 genes were overexpressed in OI in comparison to ONI, suggesting an overall reduction of the nuclear regulation activity in *A. sculptum* ovaries during infection with *R. amblyommii*. Indeed, among the gene families that presented a large number of down-regulated genes, are those related to DNA polymerases and helicases, both with 45 down-regulated genes, followed by machinery of DNA repair (26) and chromatin structural proteins. This suggests a lower activity of DNA replication, and consequently, ovarian tissue growth. We also noted the down-regulation of 7 genes related with nucleolar proteins (Spreadsheet S1). Previous works have observed the presence of two nucleoli at the epithelium cells that formed the ovary (Denardi et al., 2004). Therefore, the down regulation of these nucleoli-related genes (Spreadsheet S1) could suggest a lower development of these cells and, consequently, of the ovaries in infected *A. sculptum* females.

In addition, we have also observed the down-regulation of 7 histone acetylases (Spreadsheet

S1) related to transcription activators in *Drosophila* (Nightingale et al.,1998). The lower expression of these genes in OI in comparison to ONI (Spreadsheet S1) could be linked with the lower expression of nucleoli proteins, resulting in a lower overall gene expression activity in *A. sculptum* infected ovaries.

3.3.2.2. Transcription factors

The functional annotation of OVA transcriptome allowed for the identification of 744 genes encoding transcription factors, of which 707 were expressed in the infected ovaries; whereas were expressed at non-infected ovaries (ONI). We have annotated 59 transcription factors families from this dataset, such as 65 coding sequences for C2H2 family, leucine-rich repeat (21), Forkhead family (39), HMG family (18), bZIP family, among others (Spreadsheet S1). This large number of transcription factors families expressed in the OVA suggests a complex repertoire of genes and cellular processes regulation in the *A. sculptum* ovaries, which agrees of the late development and expressive growth observed for this organ (Denardi et al., 2004). This is especially noticeable with the transcription factors, it is worthy to note those belonging to C2H2 family, which has been described to regulates a several biological processes, such as signal transduction, cell growth, differentiation and development in *Drosophila* (Seetharam et al., 2010), which is in agreement with the morphological transformations previously observed for this organ (Denardi et al., 2004).

Regarding the differentially expressed genes, we have identified 96 overexpressed transcription factors at OI in comparison to ONI, with the most noted being the 23 coding sequences for zinc finger family, followed by C2H2 family (8), GA binding proteins (6), Forkhead (4) and bZIP (4) (Spreadsheet S1). The overexpression of Forkhead family, could indicate a decreasing of the ovarian cells in response to rickettsial infection, once transcription factors from this family, such as FOXO, have been described to activate the expression of apoptotic genes in *Drosophila*, leading to a reduction in organ size in the fly (Jünger et al., 2003). Indeed, physiological changes in the ovaries have been previously observed for *A. aureolatum* ovaries infected with *R. rickettsia* that resulted lower performance of egg oviposition (Labruna et al., 2011) and could explain these RNAseq results for infected ovaries from *A. sculptum*. However, further monohybrid studies will be necessary to gain more information about this *in silico* evidence. In addition, we have annotated 6 coding sequences for GATA binding protein (GABP) transcription factors that were up-regulated at OI in comparison to ONI (Spreadsheet S1). In opposition to what has been described for the Forkhead family, GABP transcription factors have been implicated in *Haemaphysalis longicornis* as an important activator for the expression of vitelligenin at fat bodies (Unemiya-Shirafuji et al., 2012). These results correlate the observations of by Labruna et al. (2011) for *A. aureolatum* ovaries infected with *R. rickettsia*, with the maintenance of the oviposition. Thus,

these expression profiles of GATA and Forkhead transcription factors suggest similar physiological features for *A. sculptum* ovaries infected with *R. amblyommii*.

3.3.2.3. Proteins synthesis, modification and export machineries

We have annotated 895 coding sequences related to protein synthesis machinery from OVA. Eight hundred and fifty two (852) were expressed in OI, while practically the same number were observed at ONI (Table 1). Among the most abundant families into the OVA dataset are those related to ribosomal structural proteins, as well as, all 20 aminoacyl-tRNA synthetases (Spreadsheet S1). Indeed, previous studies have observed the increase in the vitellogenin vesicles, a precursor of yolk, in *A. sculptum* ovaries during later development (Denardi et al., 2004). However, with the aminoacyl-tRNA synthetases, we observed that 84 coding sequences for this class were down regulated in ovaries in response to rickettsial infection (OI versus ONI), while only 22 were up regulated (Spreadsheet S1). This suggests a decrease in the protein synthesis activity in *A. sculptum* ovaries in response to rickettsial infection (OI versus ONI). Once vitellogenesis in *A. sculptum* ovaries is accompanied of protein synthesis activity (Denardi et al., 2004), the lower expression of aminoacyl-tRNA synthethase at OI could results in a deficient vitellogenesis processes and, consequently, lower rates of fertility and egg mass during the oviposition, as has been described during the infection of A. *aureolatum* with *R. rickettsii* (Labruna et al., 2011). Supporting this hypothesis, Fuxelius et al. (2007), though a phylogenomics approach, have described the incomplete presence of genes for amino acids metabolism in rickettsial genomes and, conversely, the abundance of amino acids transporters. This strongly indicates a rickettsial metabolic dependence on host cell metabolism. In this case, the lower expression of aminoacyl tRNA synthetases in OI could result from the lower concentration of amino acids, due the rickettsial infection.

In association with protein synthesis machinery, we have annotated 800 coding sequences related to protein modification machinery from OVA, of which 742 were expressed at OI, while 744 were expressed at ONI (Table 1). Among the main enzymes annotated from this functional class (Spreadsheet S1) are several that are related to signal peptidases (9) and phosphatases (24), which have been described to respective cleave and phosphorylate vitellogenin, leading the formation of phosphoproteins that participate of yolk composition in *Apis melifera* (Havukainen et al., 2012).

Among the differentially expressed genes related to protein modification during the OI *versus* ONI comparison, we identified 131 up regulated genes in response to rickettsial infection, and 172 were down regulated (Spreadsheet S1). Among the up-regulated genes, several were identified as chaperones, including DnaJ (12), chaperonins (9), heat shock protein 70 (7) and alpha crystalins (4). The up-regulation of those stress related chaperones could be an ovarian response to

the physiological burden caused by rickettsial infection (Henderson et al., 2006). Otherwise, coding sequences related to signal peptidases and phosphorylases were slightly up regulated. This indicates the maintenance of yolk synthesis and subsequent maintenance of a defective oviposition, as previously reported for *A. aureolatum* infected with *R. rickettsii* (Labruna et al., 2011). Regarding down regulated genes in response to rickettsial infection, a key family is the ubiquitin-like hydrolases (11 CDS), which have been described to promotes *Listeria* and *Salmonella* infection by a mechanism that also recruits actin cytoskeleton (Bassères et al., 2010). We propose that a similar mechanism employed by rickettsial infection (Martinez & Crossard, 2004; Petchampai et al., 2015). This lower expression of ubiquitin-like hydrolases at OI could be a defense mechanism of *A. sculptum* to protect ovaries against the rickettsial infection.

3.3.3. Signal transduction

3.3.3.1. mTOR signaling pathway

This functional class contains the largest number of coding sequences annotated in OVA dataset, with 2,725 coding sequences, of which 2,558 CDS were expressed in OI transcriptome, while 2,493 were derived from non-infected ovaries (Table 1). Among the proteins families related to this cellular function, it is worthy to note those that some target the rapamycin pathway, which has been described to control vitellogenesis in fat bodies of the parthenogenic tick, *H. longicornis*. The mechanism of action is through the activation of GABP transcription factor (Unemiya-Shirafuji et al., 2012). Once coding sequences for this transcription factor were up-regulated in response to rickettsial infection (OI versus ONI) and the presence of transcripts related to this pathway in the OVA dataset suggests a similar mechanism for vitellogenesis activation in *A. sculptum* ovaries.

3.3.3.2. Hormones related

We have identified 10 coding sequences related to putative hormones in ticks in the OVA dataset, such as juvenile hormones methyltransferases (JHAM) and 20-hydroxyecdysone receptors (Spreadsheet S1). Making the comparison between *A. scultpum* ovaries infected with *R. amblyommii* (OI) against non-infected ticks (ONI), we observed the up-regulation of 3 coding sequences for JHAM, while 2 CDS for the 20-hydroxyecdysone receptor were up-regulated as well. Previous studies with *H. longicornis* have described the up-regulation of vitellogenic receptors in response to 20-hydroxyecdysone and, in addition, the essential role of vitellogenin receptors for transovarian transmission of *Babesia gibsoni* (Boldbaatar et al., 2008). Thus, the up-regulation 20-hydroxyecdysone and vitellogenin receptors (further in the text) at infected ovaries of *A. sculptum* suggest similar aspects between the transovarian transmission pf *Babesia* at *H. longicornis* ovaries and *Rickettsia amblyommii* maintenance/transmission at *A. sculptum* ovaries. In addition, other

previous works have described the positive effect of juvenile hormone and JHAM on the expression of vitellogenin in *Tribolium castaneum* (Shang et al., 2011). As we found that the JHAM coding sequences were up-regulated in the infected ticks with our ovarian RNAseq data suggesting that rickettsial infection, even through negative regulation, could influence the regulation of genes responsible to maintain or increase the vitellogenesis activity in the infected ovaries. Indeed, this hypothesis could be supported by the observations of maintenance of oviposition, in *A. aureolatum* ovaries infected with *R. rickettsia*, even its negative effect on the mass and viability of the eggs in that tick specie (Labruna et al., 2011).

3.3.4. *Reproductive processes, nutrients assimilation and metabolism*

3.3.4.1. Storage/vitellogenesis

We have annotated 58 coding sequences related to storage of nutrients and vitellogenesis processes, of which 49 CDS were expressed at *A. sculptum* ovaries infected with *R. amblyommii* (OI), while 55 CDS were expressed at ONI (Table 1). Among the proteins families identified into this functional class, are those for vitellogenin, the main precursor for yolk (Denardi et al., 2004) with 39 coding sequences, followed by its receptor, with 10 CDS (Spreadsheet S1). Primary function of these proteins is the development of precursos for yolk synthesis in *A. sculptum* ovaries. These proteins develop a central role as precursor to yolk synthesis at *A. scultpum* ovaries (Denardi et al., 2004). In addition, we have also observed the up-regulation 3 vitellogenin receptors in response to rickettsial infection (Spreadsheet S1).), which have been described to play an essential role during the transovarian transmission of *Babesia gibsoni* at *H. longicornis* ovaries (Boldbaatar et al., 2008).

3.3.4.2. Nutrient transporters

The assimilation of nutrients from the blood meal to vitellogenesis is a key process to viable egg formation in tick ovaries (Hatta et al., 2010). In this context, we have identified a total of 1,392 coding sequences for nutrient transporters in the transcriptome of *A. sculptum* ovaries (OVA). One thousand two hundred and ninety seven (1,297) CDS were annotated from OI transcriptome, while 1,310 were derived from ONI (Table 1). From the OVA dataset, a total of 35 transporter families were annotated. Among those that stood out in our analysis, those related to metabolites transporters (165 CDS), amino acid/peptides transporters (144), sugar/nucleotide transporters (91), lipid transporters (52) and among others are listed in Spreadsheet S1. The expression of these transporter families in OVA after 7-10 days of the blood meal agrees with the histological observation of lipids, carbohydrates and proteins observed during the yolk formation at *A. sculptum* ovaries by Denardi et al. (2004).

Regarding the differentially expressed CDS between OI and ONI datasets, we have observed the more abundance of a several metabolites and sodium transporters in response to rickettsial infection (Spreadsheet S1). Other more abundant transporter families were those related to amino acid and peptide transporters, with 23 CDS, and sugar/nucleotide transporters (17 CDS). Despite of the down regulation of coding sequences related to carbohydrate metabolism, the corresponding overexpression of sugar/nucleotide transporters (OI versus ONI) appears to be contradictory. However, as rickettsiae are totally dependent on the host to supply pyruvate and amino acids, for nutrition and survivor (Renesto et al., 2005; Fuxelius et al., 2007), the lower concentration of these nutrients into ovary cell could be due to putative assimilation by rickettsia.

3.3.4.3. *Digestive proteases and amino acid metabolism*

The knowledge gained about the enzymes that play a role with blood digestion and vitellogenesis is a key step to the development of new strategies of the control the *A. sculptum* tick in the environment. We have discovered a subset of 431 coding sequences, of which 420 CDS were expressed in the OI transcriptome, while 412 were derived from the ONI dataset (Table 1). The digestive proteases dataset was annotated into 32 functional families, featuring metalloproteases, aspartyl protease, metalloproteases, asparaginyl proteases and among others (Spreadsheet S1). A key protein family we identified are the leucine aminopeptidases. This family has been described to be essential for the supplying of nutrients for vitelogenic ovaries in *H. longicornis* (Hatta et al., 2010). Regarding up-regulated genes at OI, we annotated 1 leucine aminopeptidase that was almost 10-fold overexpressed in response to ricketsial infection. This could suggest a positive effect of rickettsial infection on the vitellogenesis processes during oviposition.

3.3.4.4. *Carbohydrate metabolism and digestive enzymes*

We annotated 667 coding sequences related to enzymes from carbohydrates metabolism, of which 604 CDS were in the OI transcriptome, while 595 were expressed at non-infected ovaries (ONI). All these coding sequences were subcategorized into 26 gene families, according function and metabolic pathways. The most prevalent families were those related to glycolysis and gluconeogenesis (92 CDS), amino sugar metabolism (66 CDS), chitinase related (51), citrate cycle (51 CDS) and others annotated at Spreadsheet S1.

However, we did observe a down regulation of coding sequences related to catabolism of carbohydrates at OI in comparison to ONI. Among the pathways with the largest number of down regulated genes (Spreadsheet S1) were those for the citrate cycle (14 down regulated CDS), glycolysis and gluconeogenesis related (26 CDS), galactosidases (11 CDS) and pentose phosphate pathway (10 CDS). We also observed the down regulation of 52 of the 182 CDS annotated as

belonging to nucleotide metabolism, which could means a lower mitotic activity and ovarian tissue growth at the infected condition, as previously seen with the cytoskeletal cell cycle proteins (Table 1 and Spreadsheet S1). The lower expression of these pathways suggests that rickettsial infection leads to a decrease in the metabolic activities at the ovaries, especially those related to catabolic processes, necessaries to provide energy for the vitellogenesis processes. By other hand, we have verified up regulation of a several coding sequences related to chitinases in response to rickettsial infection (Spreadsheet S1). These enzymes have been described to play role during eggs hatching of *Tribolium castaneum* by hydrolysis of bonds β-1,4 of chitin molecules (Merzendorfer, 2013). The transcriptomic data of OI is related with ovaries of *A. scultum* fed during 7-10 days (early in the oviposture phase), our hypoyhesis is that overexpression of chitinases in response to *R. amblyommii* infection could to contribute to the premature hatching of eggs in infected *A. sculptum* and, for rickettsial infection of the eggs. Our hypothesis is that overexpression could lead to a thinner chitin layer of eggs facilitating the bacterial entry. Indeed, this hypothesis is corroborated by the lower mass and viability of eggs from *A. auerolatum* infected with *R. rickettsii*, with higher rates of transovarian transmission, as previously described by Labruna et al. (2011).

Regarding the coding sequences annotated as energy metabolism, it is worthy to note 165 CDS related to cytochrome P450 (Spreadsheet S1), of which 61 were more abundant in infected ovaries (OI) in comparison to ONI, which indicates an increase of the oxidative phosphorylation in response to rickettsial infection. Indeed, Denardi et al. (2004) described a great amount of micropores covering the external surface of oocytes during the vitellogenesis process in *A. sculptum* ovaries, with the function to allow the oxygenation of internal structures of the future egg. These morphological features, in association with RNAseq results, suggests a putative increase of oxygenation and, consequently, of the oxidative phosphorylation at *A. sculptum* ovaries infected with *R. amblyommii*. However, the more abundance of some components of the electron transport chain at *A. sculptum* ovaries could be related with reactive oxygen species (ROS) production in response to rickettsial infection, a defense mechanism previously observed during for *R. rickettsii* infection on mammals cells (Sahni and Rydkina, 2009).

3.3.5. *Proteins turnover, detoxification and immunity related processes*

3.3.5.1. Proteasome machinery

A total of 638 coding sequences were annotated as being related to proteasome machinery, of which 611 were expressed in infected ovaries (OI), while 602 were expressed at ONI transcriptome (Table 1). This class includes eight annotated protein families, with the largest family related to 324 ubiquitin ligases, followed by proteasome regulatory proteins (102), Vacuolar ATPases (15) and others (Spreadsheet S1). Ubiquitin ligases, that have been previously described to

participate of the infection mechanism of mammalian cells by *Rickettsia conorii*, in which the rickettsial entry into the host cell is dependent of the host ubiquitination of the the ku70 receptor (Martinez et al., 2005). In addition, Pectahmpai et al. (2014) have demonstrated the Vacuolar (V)-ATPase from *Dermancentor variabilis* ovaries is important to infection of ovaries by *R. montanensis*. We have observed several up regulated genes (OI versus ONI) for V ATPases and Ubiquitin ligases in response to rickettsial infection (Spreadsheet S1), which suggest that the infection and establishment mechanisms of *R. amblyommi* on *A. sculptum* ovaries present similar aspects with those previously described for *R. montanensis-D. variabilis* ovaries (Pectahmpai et al., 2014) and *R. conorii* on mammals cells (Martinez et al., 2005).

3.3.5.2. Detoxification related processes

From the OVA dataset**,** a total of 524 coding sequences identified as detoxification process proteins, of which 504 CDS were expressed in the OI transcriptome, and 513 were derived from ONI dataset. These coding sequences were categorized into 15 functional families (Spreadsheet S1). The families that presented the most expressed CDS in OVA were those related to sulfotransferases, glutathione S-transferase, dehydrogenases, oxidases and others (Spreadsheet S1).

In association with the overexpression of several chaperones in response to rickettsial infection, we have identified several detoxification genes overexpressed at *A. sculptum* ovaries in response to rickettsial infection (Spreadsheet S1), particularly 42 glutathione S-tranferases. These enzymes have been described to play role in the protection against oxidative damage of permethrin in *Rhipicephalus sanguineus* (Duscher et al., 2014). Previous studies have described the increase of oxidative stress in host cells during rickettsial infection in eukaryotes (Rydkina et al., 2004; Sahni & Rydkina, 2009). In addition, we observed the overexpression of several sulfotransferases (37 CDS) at OI transcriptome in comparison to ONI dataset, which have been described in a previous study to play role on anti-oxidative mechanisms in *I. scapularis* (Pichu et al., 2011). Thus, the overexpression of glutathione S-transferases and sulfotransferases at infected *A. sculptum* ovaries could be a putative defense mechanism against the potential oxidative damage of *R. amblyommii* infection at this organ.

3.3.5.3. Immunity related

We identified 98 genes encoding for immune related products, of which 90 were expressed in the OI transcriptome, and 96 CDS were derived from ONI dataset (Table 1). For better understanding, this functional class were subcategorized into 12 families (Spreadsheet S1). Among these families, the largest were those related to tumor necrosis factor receptors, with 49 CDS and toll-like receptors (11 CDS). Regarding the effect of rickettsial infection on the gene expression

profiles from OI and ONI transcriptomes (Spreadsheet S1), we observed the overexpression of four coding sequences for toll-like receptors (TLRs), which have been described to recognize molecular patterns associated to pathogens, including those from *Rickettsia* spp. (Bilak et al., 2003, Quevedo-Diaz et al., 2010). Thus, the up-regulation of this set of CDS for TLRs in infected *A. sculptum* ovaries (OI) was expected, once ticks from this group were previously infected with the *R. amblyommii*, thus, the expression of TLRs could be a response of the tick against the infection condition. However, we verified the down regulation of a several CDS related to tumor necrosis factor receptors. These receptors have been described to activate apoptotic processes in eukaryotes through NFkappaB transcription factor activation (Gravestein & Borst, 1998; Micheau & Tischopp, 2003). The down regulation of the expression of these receptors suggests a positive point regarding rickettsial infection, for avoidance of the apoptotic processes of its host tissue. Indeed, we have also observed that 2 coding sequences related to NFkappaB transcription factors were more expressed at OI in comparison to ONI (Spreadsheet S1), which could potentially corroborates this hypothesis.

3.4. Putative secreted proteins from *A. sculptum* ovaries

Despite of both transcriptomes obtained in this study being derived from ovarian tissues of *A. sculptum* females, we have annotated 1,953 coding sequences that presented signal peptide by *SignalP* analysis (Table 1). This suggests these proteins are secreted from the reproductive tract. This class of secreted proteins have been previously described from a study of the ovarian transcriptome of the triatomine bug *Rhodinius prolixius* and as a contributor for the follicle cells to oocyte content or to eggshell (Medeiros et al., 2011). We also annotated several proline-rich proteins (Spreadsheet S1) that have been previously suggested to participate on oocyte maturation in *R. prolixus* (Medeiros et al., 2011). We have also identified several proteins families described from other tick sialomes (Francischetti et al., 2008; Anatriello et al., 2010; Ribeiro et al., 2011; Karim & Ribeiro, 2011; Batista et al., 2008; Garcia et al., 2014) such as kunitz domains, serpins, mucins, cystatins, suggesting that they could play novel roles at tick ovaries (Spreadsheet S1).

3.5. Transposable elements, viral related transcripts and unknown products

A total of 194 CDS from the OVA dataset were annotated as transposable elements, of which 182 were expressed at OI transcriptome, while 179 were identified from ONI dataset. From the ten families identified by the annotation, those related to Gag related family (23 CDS), Tigger family (17 CDS), Pol polyprotein family (8), Harbinger family (5), Gypsy family (2) and others stand out (Spreadsheet S1). The presence of transposable elements at tick transcriptomes may be related with transposon incorporations in tick genomes, as described for the *Ixodes scapularis* (Gulia-Nuss et al., 2016). In this context, we have identified 14 coding sequences related homologues to viral or

transposons sequences (Table 1). The presence of viral transcripts could be related justified the presence of transposable elements transcripts at *A. sculptum* whole ovarian transcriptome (OVA).

Finally, similar to verified observations in several ixodidc ticks transcriptomes (Schwarz et al., 2013; Garcia et al., 2014; Kotsyfakis et al., 2015), we were not able to annotated a total of 872 CDS from OVA, and were classified as unknown sequences (Table 1). From this subset, a total of 84 unknown CDS were up-regulated in response to rickettsial infection (OI versus ONI), while 99 underwent down regulation. This unknown dataset could be an interesting source for further discovery of novel genes, as well as, those possibly involved at the putative mechanism of transovarian transmission of *R. amblyommii* at *A. sculptum* ovaries.

4. CONCLUSION

The rickettsial infection of the *A. sculptum* ovaries results, directly or indirectly, in down regulation of transcription of genes related to a several housekeeping processes, such as protein synthesis, especially aminoacyl-tRNAs, metabolism of carbohydrates, cytoskeleton components involved during cell cycle and replication machinery. In contrast, the up regulation of amino acid and sugar/nucleotide transporters could be related with the lower cytosolic concentration of these nutrients, due the uptake and utilization by *R. amblyommii*. This orchestrated down regulation of these processes allow us to conclude that rickettsial infection is a physiological burden for the ovarian development and oogenesis in *A. sculptum*. On the other hand, rickettsial infection could activates genes related to Arp 2/3 complex, Ubiquitin ligases and V ATPase and proteins play a role at *R. amblyommii* infection at *A. sculptum* ovaries, when comparing infected and non-infected ovaries. Rickettsial infection results in up regulation of transcripts related to transcription factor GABP, to promotes vitellogenesis in *Haemaphysalis longicornis*. The infection condition induces, directly or indirectly, the up regulation of molecular chaperones related to cellular stress. Transcripts related to cuticle proteins, that compose eggshell of tick, are also down regulated in response to infection. We conclude that rickettsial infection promotes a antagonic effect on pattern of gene expression at *A. sculptum* ovaries, leading to up regulation of genes toward a putative transovarian transmission, but resulting, at the same time, in a metabolic burden for ovaries of *A. sculptum*.

5. ACKNOWLEDGEMENTS

We thank the Foundation for Research Support of the State of Minas Gerais (FAPEMIG) for financial support. Dr. Marcelo B. Labruna, Faculty of Veterinary Medicine, University of São Paulo, for providing the inoculum of *Rickettsia amblyommii*. Dr. Pablo Juan Szabó Mathias, Laboratory of Ixodology Federal University of Uberlandia, for providing the specific pathogen free

Amblyomma sculptum specimens**.**

6. SUPLEMENTARY DATA

Spreadsheet S1 (.DOC)

7. REFERENCES

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology, 215:403–410.

ANATRIELLO, E.; RIBEIRO, J.M.; DE MIRANDA-SANTOS, I.K.; BRANDÃO, L.G.; ANDERSON, J.M.; VALENZUELA, J.G.; MARUYAMA, S.R.; SILVA, J.S.; FERREIRA, B.R. 2010. An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. BMC Genomics, 22:11-450.

ANDERSEN. L.L.; MØRK, N.; REINERT, L.S.; KOFOD-OLSEN, E.; NARITA, R.; JØRGENSEN, S.E.; SKIPPER, K.A.; HÖNING, K.; GAD, H.H.; ØSTERGAARD, L.; ØRNTOFT, T.F.; HORNUNG, V.; PALUDAN, S.R.; MIKKELSEN, J.G.; FUJITA, T.; CHRISTIANSEN, M.; HARTMANN, R.; MOGENSEN, T.H. 2015. Functional IRF3 deficiency in a patient with herpes simplex encephalitis. Journal of Experimental Medicine, 212(9):1371-9.

BARROS-BATESTI, D.M.; ARZUA, M.; BECHARA, G.H. Carrapatos de importância médico-veterinária da Região Neotropical: Um guia ilustrado para identificação de espécies. São Paulo: Vox/ICTTD-3/Butantan, 233p.

BASSÈRES, E.; COPPOTELLI, G.; PFIRRMANN, T.; ANDERSEN, J.B.; MASUCCI, M.; FRISAN, T. 2010. The ubiquitin C-terminal hydrolase UCH-L1 promotes bacterial invasion by altering the dynamics of the actin cytoskeleton. Cell Microbiology, 12(11):1622-33.

BATISTA, I.F.C.; CHUDZINSKI-TAVASSI, A.M.; FARIA, F.; SIMONS, S.M.; BARROS-BATESTTI, D.M.; LABRUNA, M.B.; LEÃO, L.I.; HO, P.L.; JUNQUEIRA-DE-AZEVEDO, I.L.M. 2008. Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acari: Ixodidae). Toxicon 51:823–34.

BILAK, H.; TAUSZIG-DELAMASURE, S.; IMLER, J.L. 2003.Toll and Toll-like receptors in Drosophila. Biochem Biochemical Society Transactions, 3:648-51.

BILLINGS, A.N.; YU, J.; TEEL, P. D.; WALKER, D.H. 1998. Detection of a spotted fever group rickettsia in *Amblyomma cajennense* (Acari : Ixodidae) in south Texas. Journal of Medical Entomology, 35:474–478.

BOLDBAATAR, D.; BATTSETSEG, B.; MATSUO, T.; HATTA ,T.; UMEMIYA-SHIRAFUJI, R.; XUAN, X.; FUJISAKI, K. 2008. Tick vitellogenin receptor reveals critical role in oocyte development and transovarial transmission of *Babesia* parasite. Biochemistry and Cell Biology, 4:331-44.

CALVO, E.; PHAM, V.M.; MARINOTTI, O.; ANDERSEN, J.F.; RIBEIRO, J.M.C. 2009. The salivary gland transcriptome of the neotropical malaria vector *Anopheles darlingi* reveals accelerated evolution of genes relevant to hematophagy. BMC Genomics 10:57.

CHMELAŘ, J.; KOTÁL, J.; KARIM, S.; KOPACEK, P.; FRANCISCHETTI, I.M.; PEDRA, J.H.; KOTSYFAKIS, M. 2015. Sialomes and Mialomes: A systems-biology view of tick tissues and tick-host interactions. Trends in Parasitology, 32(3):242-54.

DENARDI, S.E.; BECHARA, G.H.; OLIVEIRA, P.R.; NUNES, E.T.; SAITO, K.C.; MATHIAS, M.I. 2004. Morphological characterization of the ovary and vitellogenesis dynamics in the tick *Amblyomma cajennense* (Acari: Ixodidae). Veterinary Parasitology. 125(3-4):379-95.

DRUMMOND, R.O.; WHETSTONE, T.M. 1975. Oviposition of the cayena tick, *Amblyomma cajennense* (F.) in the laboratory. Annals of Entomological Society of America, v.68, n. 2, p.214-216.

DUSCHER, G.G.; GALINDO, R.C.; TICHY, A.; HUMMEL, K.; KOCAN, K.M.; DE LA FUENTE, J. 2014. Glutathione S-transferase affects permethrin detoxification in the brown dog tick, *Rhipicephalus sanguineus*. Ticks and Tick Borne Diseases, 5(3):225-33.

FENG, J.; MEYER, C.A.; WANG, Q.; LIU, J.S.; SHIRLEY LIU, X.; ZHANG, Y. 2012. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics, 28(21):2782-8.

FERNANDES, F.F.; FREITAS, E.P.S.; COSTA, A.C.; SILVA, I.G. 2005. Larvicidal potential of *Sapindus saponaria* to control the cattle tick *Boophilus microplus*. Pesquisa Agropecuária Brasileira, vol.40 no.12 Brasília.

FRANCISCHETTI, I.M.; MANS, B.J.; MENG, Z.; GUDDERRA, N.; VEENSTRA, T.D.; PHAM, V.M.; RIBEIRO, J.M. 2008. An insight into the sialome of the soft tick, *Ornithodorus parkeri*. Insect Biochemistry and Molecular Biology, 38(1):1-21.

FRIDMAN, A.; SAHA, A.; CHAN, A.; CASTEEL, D.E.; PILZ, R.B.; BOSS, G.R. 2013. Cell cycle regulation of purine synthesis by phosphoribosyl pyrophosphate and inorganic phosphate. Biochemical Journal, 454(1):91-9.

FUXELIUS, H.H.; DARBY, A.; MIN, C.K.; CHO, N.H.; ANDERSSON, S.G. 2007. The genomic and metabolic diversity of *Rickettsia*. Research in Microbiology, 158(10):745-53.

GARCIA, G.R.; GARDINASSI, L.G.; RIBEIRO, J.M.; ANATRIELLO, E.; FERREIRA, B.R.; MOREIRA, H.N.; MAFRA, C.; MARTINS, M.M.; SZABÓ, M.P.; DE MIRANDA-SANTOS, I.K.; MARUYAMA, S.R. 2014. The sialotranscriptome of *Amblyomma triste*, *Amblyomma* parvum and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. Parasites & Vectors, 7,430.

GRAVESTEIN.; L.A.; BORST.; J. 1998. Tumor necrosis factor receptor family members in the immune system. Seminars in Immunology, 10(6):423-34.

GULIA-NUSS, M.; NUSS, A.B.; MEYER, J.M.; SONENSHINE, D.E.; ROE, R.M.; WATERHOUSE, R.M.; SATTELLE, D.B.; DE LA FUENTE, J.; RIBEIRO, J.M.; MEGY, K.;THIMMAPURAM, J.; MILLER, J.R.; WALENZ, B.P.; KOREN, S.; HOSTETLER, J.B.; THIAGARAJAN, M.; JOARDAR, V.S.; HANNICK, L.I.; BIDWELL, S.; HAMMOND, M.P.; YOUNG, S.; ZENG, Q.; ABRUDAN, J.L.; ALMEIDA, F.C.; AYLLÓN, N.; BHIDE, K.; BISSINGER, B.W.; BONZON-KULICHENKO, E.; BUCKINGHAM, S.D.; CAFFREY, D.R.; CAIMANO, M.J.; CROSET, V.; DRISCOLL, T.; GILBERT, D.; GILLESPIE, J.J.; GIRALDO-CALDERÓN, G.I.; GRABOWSKI, J.M.; JIANG, D.; KHALIL, S.M.; KIM, D.; KOCAN, K.M.; KOČI, J.; KUHN, R.J.; KURTTI, T.J.; LEES, K.; LANG, E.G.; KENNEDY, R.C.; KWON, H.; PERERA, R.; QI, Y.; RADOLF, J.D.; SAKAMOTO, J.M.; SÁNCHEZ-GRACIA, A.; SEVERO, M,S.; SILVERMAN, N.; ŠIMO, L.; TOJO, M.; TORNADOR, C.; VAN ZEE, J.P.; VÁZQUEZ, J.; VIEIRA, F.G.; VILLAR, M.; WESPISER, A.R.; YANG, Y.; ZHU, J.; ARENSBURGER, P.; PIETRANTONIO, P.V.; BARKER, S.C.; SHAO, R.; ZDOBNOV, E.M.; HAUSER, F.; GRIMMELIKHUIJZEN, C.J.; PARK, Y.; ROZAS, J.; BENTON, R.; PEDRA, J.H.; NELSON, D.R.; UNGER, M.F.; TUBIO, J.M.; TU, Z.; ROBERTSON, H.M.; SHUMWAY, M.; SUTTON, G.; WORTMAN, J.R.; LAWSON, D.; WIKEL, S.K.; NENE, V.M.; FRASER, C.M.; COLLINS, F.H.;

BIRREN, B.; NELSON, K.E.; CALER, E.; HILL, C.A. 2016. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. Nature Communications, 7:10507.

GUO, Y.; RIBEIRO, J.M.C.; ANDERSON, J.M.; BOUR, S. 2009. dCAS: a desktop application for cDNA sequence annotation. Bioinformatics 25:1195–6.

HACKMAN, R.H. Structure and function in tick cuticle. 1982. Annual Review of Entomology, 27:75-95.

HATTA, T.; TSUJI, N.; MIYOSHI, T.; ISLAM, M.K.; ALIM, M.A.; YAMAJI, K.; ANI-SUZZAMAN FUJISAKI, K. 2010. Leucine aminopeptidase, HlLAP, from the ixodid tick *Haemaphysalis longicornis*, plays vital roles in the development of oocytes. Parasitology International, 59(2):286-9.

HAVUKAINEN, H.; UNDERHAUG, J.; WOLSCHIN, F.; AMDAM, G.; HALSKAU, Ø. 2012. A vitellogenin polyserine cleavage site: highly disordered conformation protected from proteolysis by phosphorylation. Journal of Experimental Biology, 215:1837-46.

HENDERSON, B.; ALLAN, E.; COATES, A.R. 2006. Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. Infection and Immunity, 74(7):3693-706.

HILLMAN, R.D.; BAKTASH, Y.M.; MARTINEZ, J.J. 2013. OmpA-mediated rickettsial adherence to and invasion of human endothelial cells is dependent upon interaction with α2β1 integrin. Cellular Microbiology, 15(5):727-41.

HOCH, M.; PANKRATZ, M.J. 1996. Control of gut development by fork head and cell signaling molecules in *Drosophila*. Mechanical. Development. 58:3–14.

JÜNGER, M.A.; RINTELEN, F.; STOCKER, H.; WASSERMAN, J.D.; VÉGH, M.; RADIMERSKI, T.; GREENBERG, M.E.; HAFEN, E. 2003. The Drosophila forkhead transcription factor FOXO mediates the reduction in cell number associated with reduced insulin signaling. Journal of Biology, 2(3):20.

KARBOWIAK, G.; BIERNAT, B.; SZEWCZYK, T.; SYTYKIEWICZ, H. 2015. The role of particular tick developmental stages in the circulation of tick-borne pathogens affecting humans in Central Europe. 1. The general pattern. Ann. Parasitology, 61(4):221-8.

KARIM, S.; RIBEIRO, J.M. 2015. An Insight into the Sialome of the Lone Star Tick, *Amblyomma americanum*, with a Glimpse on Its Time Dependent Gene Expression. PLoS One, 10(7):e0131292.

KARIM, S.; SINGH, P.; RIBEIRO, J.M. 2011. A deep insight into the sialotranscriptome of the gulf coast tick, Amblyomma maculatum. PLoS One, 6(12):e28525.

KROGH, A.; LARSSON, B.; VON HEIJNE, G.; SONNHAMMER, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology,. 305(3):567-80.

KURTTI, T.J.; MUNDERLOH, U.G.; ANDREADIS, T.G.; MAGNARELLI, L.A.; MATHER, T.N. 1996. Tick cell culture isolation of an intracellular prokaryote from the tick *Ixodes scapularis*. Journal Invertebrate Pathology, 67:318–21.

LABRUNA, M.B.; OGRZEWALSKA, M.; SOARES, J.F.; MARTINS, T.F.; SOARES, H.S.; MORAES-FILHO, J.; NIERI-BASTOS, F.A.; ALMEIDA, A.P.; PINTER, A 2011. Experimental infec-

tion of *Amblyomma aureolatum* ticks with *Rickettsia rickettsii*. Emerging Infectious Diseases, 17(5):829-34.

LABRUNA, M.B.; PACHECO, R.C.; NAVA, S.; BRANDAO, P.E.; RICHTZENHAIN, L.J.; GUGLIELMONE, A.A. 2007. Infection by Rickettsia bellii and "*Rickettsia amblyommii*" in *Amblyomma neumanni* ticks from Argentina. Microbial Ecology, 54:126-33.

LOPES, C.M.; OLIVEIRA, P.R.; HADDAD, J.P.; DOMINGUES, L.N.; PINHEIRO, R.R.; BORGES, L.M.; LABRUNA, M.B.; LEITE, R.C. 2008. Biological parameters of ticks (*Amblyomma cajennense* Fabricius, 1787) under field and laboratory conditions in Pedro Leopoldo, State of Minas Gerais, Brazil. Revista Brasileira de Parasitologia Veterinaria, 1:14-7.

LUO W, BROUWER C. 2013. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics, 29(14):1830-1.

MARTINEZ, J.J.; COSSART, P. 2004. Early signaling events involved in the entry of *Rickettsia conorii* into mammalian cells. Journal of Cell Science, 117:5097-106.

MARTINEZ, J.J.; SEVEAU, S.; VEIGA, E.; MATSUYAMA, S.; COSSART, P. 2005. Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for *Rickettsia conorii*. Cell, 123(6):1013-1023.

MARUYAMA, S.R.; ANATRIELLO, E.; ANDERSON, J.M.; RIBEIRO, J.M.; BRANDÃO, L.G.; VALENZUELA, J.G.; FERREIRA, B.R.; GARCIA, G.R.; SZABÓ, M.P.; PATEL, S.; BISHOP, R.; DE MIRANDA-SANTOS, I.K. 2010. The expression of genes coding for distinct types of glycine-rich proteins varies according to the biology of three metastriate ticks, *Rhipicephalus* (*Boophilus*) *microplus*, *Rhipicephalus sanguineus* and *Amblyomma cajennense*. BMC Genomics 11:363.

MERZENDORFER, H. 2013. Chitin synthesis inhibitors: old molecules and new developments. Insect Sci. 20:121–38.

MICHEAU O.; TSCHOPP J. 2003. Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. Cell, 114(2):181-90.

NARASIMHAN, S.; RAJEEVAN, N.; LIU, L.; ZHAO, Y.O.; HEISIG, J.; PAN, J.; EPPLER-EPSTEIN, R.; DEPONTE, K.; FISH, D.; FIKRIG, E. 2014. Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. Cell Host Microbe 15:58–71.

NIEBYLSKI, M.L.; PEACOCK, M.G.; SCHWAN, T.G. 1999. Lethal effect of *Rickettsia rickettsii* on its tick vector (*Dermacentor andersoni*). Applied and Environmental Microbiology, 65:773–8

NIGHTINGALE, K.P.; WELLINGER, R.E.; SOGO, J.M.; BECKER, P.B. Histone acetylation facilitates RNA polymerase II transcription of the *Drosophila* hsp26 gene in chromatin. EMBO Journal, 17(10):2865-76.

OLIVEIRA, C.J.F.; SÁ-NUNES, A.; FRANCISCHETTI, I.M.B.; CARREGARO, V.; ANATRIELLO, E.; SILVA, J.S.; SANTOS, I.K.F. DE M.; RIBEIRO, J.M.C.; FERREIRA, B.R. 2011. Deconstructing tick saliva: non-protein molecules with potent immunomodulatory properties. J. Biol. Chem. 286:10960–9.

PAL, U.; LI, X.; WANG, T.; MONTGOMERY, R.R.; RAMAMOORTHI, N.; DESILVA, A.M.; BAO, F.; YANG, X.; PYPAERT, M.; PRADHAN, D.; KANTOR, F.S.; TELFORD, S.; ANDERSON, J.F.; FIKRIG, E.; 2004. TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. Cell 119:457–68.

PETCHAMPAI, N.; SUNYAKUMTHORN, P.; BANAJEE, K.H.; VERHOEVE, V.I.; KEARNEY, M.T.; MACALUSO, K.R. 2015. Identification of host proteins involved in rickettsial invasion of tick cells. Infection and Immunity, 83(3):1048-55.

PETCHAMPAI.; N.; SUNYAKUMTHORN.; P.; GUILLOTTE.; M.L.; VERHOEVE.; V.I.; BANAJEE.; K.H.; KEARNEY.; M.T.; MACALUSO.; K.R. 2014. Novel Identification of *Dermacentor variabilis* Arp2/3 complex and its role in rickettsial infection of the arthropod vector. PLoS One, 14;9(4).

PICHU, S.; YALCIN, E.B.; RIBEIRO, J.M.; KING, R.S.; MATHER, T.N. 2011. Molecular characterization of novel sulfotransferases from the tick, *Ixodes scapularis*. BMC Biochemistry. 12:32.

QUEVEDO-DIAZ, M.A.; SONG, C.; XIONG, Y.; CHEN, H.; WAHL, L.M.; RADULOVIC, S.; MEDVEDEV, A.E. 2010. Involvement of TLR2 and TLR4 in cell responses to *Rickettsia akari*, Journal of Leukocyte Biology, 88(4):675-85.

RIBEIRO, J.M.; ANDERSON, J.M.; MANOUKIS, N.C.; MENG, Z.; FRANCISCHETTI, I.M. 2011. A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. BMC Genomics 12:136.

RYDKINA, E.; SAHNI, S.K.; SANTUCCI, L.A.; TURPIN, L.C.; BAGGS, R.B.; SILVERMAN, D.J. 2004. Selective modulation of antioxidant enzyme activities in host tissues during *Rickettsia conorii* infection. Microbial Pathogenesis, 36:293–301.

SCHWARZ, A.; TENZER, S.; HACKENBERG, M.; ERHART, J.; GERHOLD-AY, A.; MAZUR, J.; KUHAREV, J.; RIBEIRO, J.M.; KOTSYFAKIS, M. 2014. A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinu*s midgut and salivary glands during early attachment and feeding. Molecular & Cellular Proteomics, 13(10):2725-35.

SHENG.; Z.; XU.; J.; BAI.; H.; ZHU.; F.; PALLI.; S.R. 2011. Juvenile hormone regulates vitellogenin gene expression through insulin-like peptide signaling pathway in the red flour beetle.; *Tribolium castaneum*. The Journal of Biological Chemistry, 286(49):41924-36.

STANKE, M.; KELLER, O.; GUNDUZ, I.; HAYES, A.; WAACK, S.; MORGENSTERN, B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research. 34:W435-9.

STANKE, M.; MORGENSTERN, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research. 33:W465-7.

STANKE, M.; STEINKAMP, R.; WAACK, S.; MORGENSTERN, B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Research. 32:W309-12.

ANEXOS

ANEXOS CAPÍTULO 1: A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of *A. sculptum*

Spreadsheet S1. Functional classification and distribution of CDS expressed among the *A. scultpum* internal organs

FUNCTIONAL CLASSES	Tissue expression profiles										
	MDG	OVA	SG	M/O/S	M/O	M/S	S/O	MDG*	OVA *	SG*	TOTAL
CYTOSKELETAL	**1178**	**1107**	**669**	**622**	**413**	**2**	**30**	**141**	**42**	**15**	**1265**
Actin and related proteins	44	42	26	25	16	0	0	3	1	1	46
Actin binding proteins	56	55	25	23	27	0	2	6	3	0	61
Actin depolymerizing factors	6	4	3	3	1	0	0	2	0	0	6
Ankyrin related	75	86	54	49	25	0	5	1	7	0	87
Arp related	41	44	28	26	14	0	1	1	3	1	46
Arp 2/3 complex	16	16	12	12	3	0	0	1	1	0	17
Autophagy processes related	9	9	6	6	3	0	0	0	0	0	9
Calponin related	22	20	19	18	1	0	0	3	1	1	24
Catenin	4	3	3	3	0	0	0	1	0	0	4
CDC-42 related	6	6	4	4	2	0	0	0	0	0	6
Cell cycle processes related	117	93	51	46	44	1	1	26	2	3	123
Cohesin	2	2	1	1	1	0	0	0	0	0	2
Condensin	2	2	1	1	1	0	0	0	0	0	2
Coronin	4	5	1	1	3	0	0	0	1	0	5
Dynactins	11	10	6	6	4	0	0	1	0	0	11
Dyneins	25	26	18	15	9	0	1	1	1	2	29
Dystrobrevin	5	5	4	4	1	0	0	0	0	0	5
Dystrophin	12	12	7	7	5	0	0	0	0	0	12
FERM family	16	16	12	11	4	0	1	1	0	0	17
Filamin	21	15	10	10	5	0	0	6	0	0	21
Flagellins and others motility	63	19	15	15	4	0	0	44	0	0	63

related											
Formin	4	4	3	3	1	0	0	0	0	0	4
Keratin	4	2	1	1	1	0	0	2	0	0	4
Kinesin	77	78	39	38	37	0	1	2	2	0	80
Microtubule related	52	54	39	35	16	0	2	1	1	2	57
Muscle-like proteins	19	18	7	7	11	0	0	1	0	0	19
Myosin	138	132	81	73	46	0	5	19	8	3	154
Myotubularin	9	9	7	7	2	0	0	0	0	0	9
Nesprins	3	4	3	2	1	0	1	0	0	0	4
Nexins	36	36	24	24	12	0	0	0	0	0	36
Others related cytoskeletal	57	59	33	30	24	0	3	3	2	0	62
Septin family	3	1	1	1	0	0	0	2	0	0	3
Spectrin related	36	37	20	20	15	0	0	1	2	0	38
Spindle pole related proteins	6	6	4	4	2	0	0	0	0	0	6
Thymosin	2	1	1	1	0	0	0	1	0	0	2
Connectins	6	8	5	3	3	0	1	0	1	1	9
Tropomodulins	4	5	5	4	0	0	1	0	0	0	5
Troponins	15	15	11	11	4	0	0	0	0	0	15
Tubulins	88	92	42	36	46	0	5	6	5	1	99
WASH domain related	5	0	0	0	0	0	0	5	0	0	5
WASP related proteins	6	6	3	3	3	0	0	0	0	0	6
WD40 containing domains	51	50	34	33	16	1	0	1	1	0	52
DETOXIFICATION	**614**	**514**	**213**	**201**	**289**	**2**	**8**	**122**	**16**	**2**	**640**
Peroxidases	35	28	14	12	10	0	2	13	4	0	41
Arylsulfatase	27	29	8	7	20	0	1	0	1	0	29
Cytochrome oxidase	8	0	0	0	0	0	0	8	0	0	8
Dehydrogenases	75	52	29	27	23	1	1	24	1	0	77
Superoxide dismutase	17	15	6	6	9	0	0	2	0	0	17
Glutathione S-transferase	102	97	40	38	58	0	1	6	0	1	104
Glutathione synthetase	6	6	1	1	5	0	0	0	0	0	6

Glutathione related hydrolases	5	1	0	0	1	0	0	4	0	0	5
Multidrug resistance associated	60	42	18	16	20	1	1	23	5	0	66
Other detoxification proteins	44	33	13	12	19	0	1	13	1	0	46
Oxidases	48	36	14	13	22	0	0	13	1	1	50
Peroxidases	8	6	3	3	3	0	0	2	0	0	8
Reductases	24	13	8	8	4	0	0	12	1	0	25
Selenium binding proteins	5	4	4	4	0	0	0	1	0	0	5
Sulfotransferases	150	152	55	54	95	0	1	1	2	0	153
DIGESTIVE ENZYMES: LIPIDS	**170**	**171**	**56**	**52**	**112**	**1**	**3**	**5**	**4**	**0**	**177**
Acetyl/Butyrylcholinesterases	44	45	10	8	35	1	1	0	1	0	46
Acetyl esterases	5	4	1	1	3	0	0	1	0	0	5
Carboxyesterases	18	18	0	0	17	0	0	1	1	0	19
Cholinesterases	2	2	1	1	1	0	0	0	0	0	2
Isoamyl-acetate esterases	13	13	3	3	10	0	0	0	0	0	13
Lipases	19	19	5	5	14	0	0	0	0	0	19
Other esterases	9	9	6	6	3	0	0	0	0	0	9
Phosphodiesterases	34	35	14	12	20	0	2	2	1	0	37
Phosphoesterases	3	2	1	1	1	0	0	1	0	0	3
Thioesterases	15	16	13	13	2	0	0	0	1	0	16
Triglyceride lipase-cholesterol esterases	8	8	2	2	6	0	0	0	0	0	8
DIGESTIVE ENZYMES: PROTEASES	**417**	**431**	**182**	**175**	**230**	**0**	**7**	**12**	**19**	**0**	**443**
Aminopeptidases	14	15	9	9	5	0	0	0	1	0	15
Asparaginyl peptidases	13	14	6	6	7	0	0	0	1	0	14
Aspartyl proteases	12	12	3	3	9	0	0	0	0	0	12
ATP-dependent proteases	6	7	1	1	4	0	0	1	2	0	8
Other cysteine proteases	6	6	6	6	0	0	0	0	0	0	6
C48 family peptidases	2	2	1	1	1	0	0	0	0	0	2
C54 family peptidases	2	2	2	2	0	0	0	0	0	0	2

Carboxypeptidases	24	24	6	6	18	0	0	0	0	0	24
Cathepsin A	17	18	6	6	11	0	0	0	1	0	18
Cathepsin B	1	2	0	0	1	0	0	0	1	0	2
Cathepsin C	1	1	0	0	1	0	0	0	0	0	1
Cathepsin D	1	1	0	0	1	0	0	0	0	0	1
Cathepsin L	3	3	2	2	1	0	0	0	0	0	3
Dipeptidyl peptidases	8	8	6	6	2	0	0	0	0	0	8
Calpains cysteine proteases	5	7	3	3	2	0	0	0	2	0	7
Endopeptidases	4	6	4	2	2	0	2	0	0	0	6
Furin-like proteases	1	1	1	1	0	0	0	0	0	0	1
Membrane protease fragments	1	1	1	1	0	0	0	0	0	0	1
Insulinase superfamily	8	8	5	5	3	0	0	0	0	0	8
Tick legumains	2	2	0	0	2	0	0	0	0	0	2
Angiotensinases	5	5	0	0	5	0	0	0	0	0	5
M20 peptidases	25	26	10	9	16	0	1	0	0	0	26
Oligopeptidases	6	6	2	2	4	0	0	0	0	0	6
Mitochondrial peptidases	9	9	7	7	2	0	0	0	0	0	9
Yolk cathepsin	2	2	0	0	2	0	0	0	0	0	2
Other proteases	18	19	11	11	6	0	0	1	2	0	20
Papain family	4	4	1	1	3	0	0	0	0	0	4
Serine proteases	42	42	10	10	31	0	0	1	1	0	43
g-glutamyltranspeptidases	6	5	1	1	4	0	0	1	0	0	6
Tripeptidyl peptidase	61	58	21	21	36	0	0	4	1	0	62
Proline-X peptidases	11	12	6	6	5	0	0	0	1	0	12
Metalloproteases	97	103	51	47	46	0	4	4	6	0	107
DIGESTIVE ENZYMES: SUGAR	**142**	**145**	**70**	**69**	**71**	**0**	**1**	**2**	**4**	**0**	**147**
Amylases	17	18	8	7	10	0	1	0	0	0	18
Fucose kinase	2	2	1	1	1	0	0	0	0	0	2
Fucose synthase	1	1	1	1	0	0	0	0	0	0	1
Fucosidase	6	6	2	2	4	0	0	0	0	0	6

Fucosyltransferases	16	17	7	7	9	0	0	0	1	0	17
Glucanases	2	2	0	0	2	0	0	0	0	0	2
Glucosaminidase	8	8	6	6	2	0	0	0	0	0	8
Glucosidases	23	22	12	12	9	0	0	2	1	0	24
Hydrolases	2	2	0	0	2	0	0	0	0	0	2
Lectins	12	12	4	4	8	0	0	0	0	0	12
Glycosyl transferases	2	2	1	1	1	0	0	0	0	0	2
Maltases	8	8	4	4	4	0	0	0	0	0	8
Other sugar enzymes	12	13	9	9	3	0	0	0	1	0	13
Mannosidases	10	11	2	2	8	0	0	0	1	0	11
Mannosyl transferases	6	6	4	4	2	0	0	0	0	0	6
Melibiases	7	7	3	3	4	0	0	0	0	0	7
Oligosaccharyltransferases	8	8	6	6	2	0	0	0	0	0	8
EXTRACELLULAR MATRIX	**890**	**759**	**394**	**352**	**343**	**6**	**30**	**189**	**34**	**6**	**960**
Ankyrin related	23	23	7	7	15	0	0	1	1	0	24
Cadherin related	12	12	6	3	6	2	1	1	2	0	15
Calsyntenin	4	4	3	3	1	0	0	0	0	0	4
Catenins	11	12	7	6	4	1	0	0	2	0	13
Cell adhesion antigens	35	44	25	20	15	0	5	0	4	0	44
Cell cycle related proteins	6	6	2	2	4	0	0	0	0	0	6
Cell Wall related proteins	9	5	4	3	2	1	0	3	0	0	9
Collagen related proteins	20	21	11	8	11	0	2	1	0	1	23
Cuticle related proteins	52	50	24	24	26	0	0	2	0	0	52
EM70 antigen related	8	8	6	6	2	0	0	0	0	0	8
Enzymes related	35	22	8	5	12	0	3	18	2	0	40
Fasciclin	5	5	5	4	0	0	1	1	0	0	6
Fibrillin	5	4	2	2	2	0	0	1	0	0	5
Fibronectin	21	21	19	16	3	1	2	1	0	0	23
Focaldhesin	3	3	1	1	2	0	0	0	0	0	3
Galectin	4	4	2	2	2	0	0	0	0	0	4

Glycosyltransferases	6	1	1	1	0	0	0	5	0	0	6
Glycine-rich proteins	7	6	5	4	2	0	0	1	0	1	8
GTPases	5	5	1	1	3	0	0	1	1	0	6
Hemolysin	4	2	1	1	1	0	0	2	0	0	4
Innexin	4	5	2	2	2	0	0	0	1	0	5
Intergrin	7	10	4	4	3	0	0	0	3	0	10
Extracellular matrix kinases	8	9	5	4	4	0	1	0	0	0	9
Lipoproteins	3	0	0	0	0	0	0	3	0	0	3
Mucins	16	18	11	8	8	0	2	0	0	1	19
O-glycosyl hidrolases	2	2	1	1	1	0	0	0	0	0	2
Outer membrane antigens	31	3	2	2	1	0	0	28	0	0	31
Others related enziymes	74	67	28	27	36	0	1	11	3	0	78
Peptidoglycan related protein	4	4	2	0	2	0	2	2	0	0	6
Phosphatidylinositol glycan	4	4	2	2	2	0	0	0	0	0	4
Glycoproteins	59	61	34	32	26	0	2	1	1	0	62
Secretion related	7	4	1	1	3	0	0	3	0	0	7
Syntrophins	4	4	2	2	2	0	0	0	0	0	4
TEM7 antigen	3	3	2	2	1	0	0	0	0	0	3
Tetraspanin	18	18	11	11	7	0	0	0	0	0	18
Connectin	6	0	0	0	0	0	0	6	0	0	6
Other transmembrane proteins	342	268	136	125	122	1	7	94	14	3	366
Transmembrane channels	15	12	6	6	6	0	0	3	0	0	15
Vitronectin related	8	9	5	4	4	0	1	0	0	0	9
IMMUNE RELATED PRODUCTS	**113**	**98**	**3**	**3**	**89**	**0**	**0**	**21**	**6**	**0**	**119**
Beta-lactamase	4	2	1	1	1	0	0	2	0	0	4
Immune related antigens	2	1	0	0	1	0	0	1	0	0	2
Complement related proteins	1	1	0	0	1	0	0	0	0	0	1
Cytokine related proteins	2	2	0	0	2	0	0	0	0	0	2
Immunoglobulin binding	5	4	0	0	4	0	0	1	0	0	5
Immunoglobulins fragments	5	4	0	0	4	0	0	1	0	0	5

Interferon binding proteins	2	2	0	0	2	0	0	0	0	0	2
MHC related proteins	3	1	0	0	1	0	0	2	0	0	3
Other immune proteins	22	21	0	0	17	0	0	5	4	0	26
Penicillin binding proteins	6	0	0	0	0	0	0	6	0	0	6
Tumor necrosis facto	50	49	0	0	48	0	0	2	1	0	51
Toll-lke receptors	11	11	2	2	8	0	0	1	1	0	12
AMINO ACIDS METABOLISM	**580**	**388**	**200**	**190**	**174**	**2**	**8**	**214**	**16**	**0**	**604**
CARBOHYDRATE METABOLISM	**847**	**652**	**329**	**304**	**299**	**5**	**17**	**239**	**32**	**3**	**899**
Amino sugar metabolism	72	66	36	30	27	2	4	13	5	0	81
Carbonic anhydrase	6	6	5	5	1	0	0	0	0	0	6
Chitinase related	49	51	13	12	35	0	1	2	3	0	53
Oxidative descarboxylation	4	2	2	2	0	0	0	2	0	0	4
Fucosyltransferases	2	2	1	1	1	0	0	0	0	0	2
Galactosidases	29	26	12	12	14	0	0	3	0	0	29
Galactosyltransferases	18	18	10	10	8	0	0	0	0	0	18
Glucan related metabolism	22	17	6	4	9	0	2	9	2	0	26
Glucosidases	10	5	5	5	0	0	0	5	0	0	10
Glucogen metabolism	30	27	13	12	12	0	1	6	2	0	33
Glycosyl hydrolases	14	13	7	7	5	0	0	2	1	0	15
Glycolysis and Gluconeogenesis	153	92	47	46	41	0	1	66	4	0	158
Glycosyltransferases	30	36	17	15	14	0	2	1	5	0	37
Glyoxylate cycle related	15	9	2	2	7	0	0	6	0	0	15
Inositol metabolism	10	9	4	4	5	0	0	1	0	0	10
Lactic fermentation	5	4	1	1	3	0	0	1	0	0	5
Mannose metabolism	24	17	7	5	11	1	1	7	0	0	25
Other dehydrogenases	38	29	13	10	14	0	2	14	3	1	44
Pentoses phosphate pathway	60	41	16	16	23	0	0	21	2	0	62
Polysaccharide metabolism	4	2	0	0	2	0	0	2	0	0	4
Sulfatases	17	18	6	5	12	0	1	0	0	0	18
Tricarboxilic acid cycle	83	51	35	32	16	1	0	34	3	2	88

Others transferases	8	8	7	7	1	0	0	0	0	0	8
Trehalose assimilation	13	4	2	1	2	0	1	10	0	0	14
Xylose metabolism	6	4	2	1	3	1	0	1	0	0	6
Other carbon metabolism transcritps	125	95	60	59	33	0	1	33	2	0	128
ENERGY METABOLISM	**758**	**539**	**284**	**275**	**241**	**1**	**2**	**241**	**21**	**6**	**787**
INTERMEDIATE METABOLISM	**1720**	**1093**	**484**	**455**	**552**	**0**	**24**	**713**	**62**	**5**	**1811**
LIPID METABOLISM	**751**	**636**	**272**	**250**	**339**	**3**	**14**	**159**	**33**	**5**	**803**
NUCLEOTIDE METABOLISM	**291**	**182**	**96**	**92**	**85**	**0**	**3**	**114**	**2**	**1**	**297**
NUCLEAR EXPORT MACHNERY	**145**	**126**	**92**	**86**	**35**	**0**	**4**	**24**	**1**	**2**	**152**
Exporter ABC superfamily	7	4	4	3	1	0	0	3	0	1	8
Exportin	8	4	1	1	3	0	0	4	0	0	8
Importin	23	24	19	18	5	0	1	0	0	0	24
Karyopherin	15	16	12	11	4	0	1	0	0	0	16
Nesprins	5	5	4	2	1	0	2	2	0	0	7
Nuclear receptors related	15	13	12	11	2	0	0	2	0	1	16
Nucleoporins	42	42	29	29	12	0	0	1	1	0	43
Nuclear transport regulators	6	4	3	3	1	0	0	2	0	0	6
Other nuclear export related	4	4	2	2	2	0	0	0	0	0	4
Translocators	20	10	6	6	4	0	0	10	0	0	20
NUCLEAR REGULATION	**1351**	**1122**	**624**	**574**	**474**	**7**	**41**	**296**	**33**	**2**	**1427**
Sister chromatid related	7	9	5	4	3	0	1	0	1	0	9
Chromatin related proteins	34	33	28	26	5	0	2	3	0	0	36
Chromatin segregation proteins	19	17	12	11	6	1	0	1	0	0	19
Chromatin structural proteins	66	56	30	29	24	0	1	13	2	0	69
Cysteine/Serine-rich proteins	3	2	1	1	1	0	0	1	0	0	3
DNA polymerases	188	140	79	71	63	2	5	52	1	1	195
DNA binding proteins	118	116	64	59	49	0	4	10	4	1	127
DNA gyrase	14	0	0	0	0	0	0	14	0	0	14

DNA kinase related	13	11	7	6	5	1	0	1	0	0	13
DNA ligase	15	8	2	2	5	0	0	8	1	0	16
DNA methylase	37	35	20	18	14	0	2	5	1	0	40
DNA primase	7	4	1	1	3	0	0	3	0	0	7
Deoxyribonuclease	44	33	16	16	16	0	0	12	1	0	45
Nuclear envelope	4	4	3	3	1	0	0	0	0	0	4
Exonucleases	19	10	5	5	5	0	0	9	0	0	19
DNA glycosylase	6	3	1	1	2	0	0	3	0	0	6
H1, H2A, H3 and H4 histones	35	32	20	17	12	0	3	6	0	0	38
H2B histones	9	8	7	7	1	0	0	1	0	0	9
Histones acetylases	36	35	27	24	9	1	2	2	0	0	38
Histones deacetylases	21	26	13	10	11	0	3	0	2	0	26
Histones demethylases	9	13	9	6	2	0	3	1	2	0	14
HU family	1	1	0	0	1	0	0	0	0	0	1
Helicases	197	140	83	79	55	1	3	62	3	0	203
Histone methylases	22	24	15	13	8	0	2	1	1	0	25
Nuclear lamina related proteins	5	4	1	1	3	0	0	1	0	0	5
Leucine-rich repeat	16	19	11	10	6	0	1	0	2	0	19
Nucleases	8	7	3	3	4	0	0	1	0	0	8
Nucleolar related proteins	22	21	13	13	8	0	0	1	0	0	22
Other Nuclear proteins	126	116	70	67	41	0	3	18	5	0	134
Recombination machnery	12	3	1	1	1	0	0	10	1	0	13
Repair machnery	116	98	34	32	60	0	2	24	4	0	122
Other replication related	34	31	14	13	16	0	1	5	1	0	36
Ribonucleases	10	6	5	4	1	0	1	5	0	0	11
Septins	3	3	2	2	1	0	0	0	0	0	3
Topoisomerases	72	53	21	18	32	1	2	21	1	0	75
DNA translocases	3	1	1	1	0	0	0	2	0	0	3
PROTEIN EXPORT MACHNERY	**530**	**503**	**314**	**291**	**182**	**2**	**13**	**55**	**17**	**8**	**568**
Atlastin related	479	454	288	266	159	2	12	52	17	8	516

Synaptobrevins and related	13	14	9	8	5	0	1	0	0	0	14
Translocases related	4	4	0	0	4	0	0	0	0	0	4
TRAPP complex related	6	6	6	6	0	0	0	0	0	0	6
Vacuolar ATPases	28	25	11	11	14	0	0	3	0	0	28
PROTEIN MODIFICATION	**963**	**800**	**458**	**422**	**314**	**4**	**25**	**223**	**39**	**7**	**1034**
PROTEASOME MACHNERY	**610**	**638**	**405**	**382**	**207**	**4**	**15**	**17**	**34**	**4**	**663**
ATPases	16	15	14	13	2	0	0	1	0	1	17
Neddylation related	16	16	10	10	6	0	0	0	0	0	16
Other proteasome proteins	34	36	22	22	12	0	0	0	2	0	36
Other ubiquitin related	101	102	67	65	31	1	1	4	5	0	107
Proteasome regulatION	71	71	52	51	17	0	1	3	2	0	74
Ubiquitin ligases	302	324	188	176	117	1	9	8	22	2	335
Ubiquitin hydrolasres	61	64	43	36	22	2	4	1	2	1	68
Ubiquitin proteases	9	10	9	9	0	0	0	0	1	0	10
PROTEIN SYNTHESIS MACHNERY	**1158**	**895**	**533**	**509**	**341**	**4**	**13**	**304**	**32**	**7**	**1210**
Protein synthesis activator	6	8	6	6	0	0	0	0	2	0	8
Ribosome assembly protein	4	4	3	3	1	0	0	0	0	0	4
Ribosome biogenesis related	24	23	13	13	9	0	0	2	1	0	25
Translation elongation factors	44	33	18	17	14	1	0	12	2	0	46
Mitochondrial elongation	6	6	0	0	4	0	0	2	2	0	8
Other related enzymes	26	21	13	13	7	0	0	6	1	0	27
Translation initiation factors	112	99	67	64	32	0	2	16	1	1	116
Ribosome maturation related	3	1	1	1	0	0	0	2	0	0	3
Other translation related	30	28	18	17	8	0	1	5	2	0	33
Pre-ribosome related proteins	8	8	4	4	4	0	0	0	0	0	8
Ribosome synthesis regulators	4	3	1	1	2	0	0	1	0	0	4
Polipeptide release factors	6	3	3	3	0	0	0	3	0	0	6
Ribosome binding proteins	18	17	9	9	7	0	0	2	1	0	19
Other ribosomal structural	230	187	135	129	54	1	1	46	3	4	238

28S Ribosome strutural proteins	3	3	3	3	0	0	0	0	0	0	3
30S Ribosome strutural proteins	31	26	10	8	16	0	2	7	0	0	33
39S ribosome structural	2	2	2	1	1	0	0	0	0	1	3
40S ribosome structural	50	35	13	13	17	0	0	20	5	0	55
50S ribosome structural	42	23	5	4	15	0	1	23	3	0	46
60S ribosome structural	74	50	25	24	26	1	0	23	0	0	74
Mitochondrial ribosomes	41	43	32	31	10	0	1	0	1	0	43
Mitochondrial 28S ribosome	14	14	10	10	4	0	0	0	0	0	14
Mitochondrial 37S ribosome	0	1	1	0	0	0	1	0	0	0	1
Mitochondrial 39S ribosome	11	11	10	10	1	0	0	0	0	0	11
rRNA methylases	24	15	6	6	9	0	0	9	0	0	24
rRNA maturation related	2	1	1	0	1	1	0	0	0	0	2
Others tRNA related proteins	162	124	65	63	54	0	2	45	5	0	169
Glycyl-tRNA synthetases	9	3	2	2	1	0	0	6	0	0	9
Alanyl-tRNA synthetases	8	4	2	2	2	0	0	4	0	0	8
Leucyl-tRNA synthetases	11	8	2	2	5	0	0	4	1	0	12
Valyl-tRNA synthetases	18	6	1	1	5	0	0	12	0	0	18
Isoleucyl-tRNA synthetases	13	4	1	1	3	0	0	9	0	0	13
Methionyl-tRNA synthetases	8	6	4	4	2	0	0	2	0	0	8
Phenylalanyl-tRNA synthetases	9	5	4	4	1	0	0	4	0	0	9
Tyrosyl-tRNA synthetases	6	4	1	1	3	0	0	2	0	0	6
Tryptophanyl-tRNA synthetases	4	2	1	1	1	0	0	2	0	0	4
Cysteinyl-tRNA synthetases	8	7	4	4	3	0	0	1	0	0	8
Seryl-tRNA synthetases	12	10	6	6	3	0	0	3	1	0	13
Histidyl-tRNA synthetases	8	4	3	3	1	0	0	4	0	0	8
Threonyl-tRNA synthetases	8	3	1	1	2	0	0	5	0	0	8
Prolyl-tRNA synthetases	11	6	3	3	3	0	0	5	0	0	11
Asparaginyl-tRNA synthetases	5	2	1	1	1	0	0	3	0	0	5
Glutaminyl-tRNA synthetsases	6	5	5	4	0	0	1	2	0	0	7
Lysyl-tRNA synthetases	7	6	2	2	4	0	0	1	0	0	7
Arginyl-tRNA synthetase	5	6	3	2	2	0	1	1	1	0	7

Glutamyl-tRNA synthetsases	9	5	4	4	1	0	0	4	0	0	9
Aspartyl-tRNA synthetases	8	4	3	3	1	0	0	4	0	0	8
Other tRNA synthetase related	8	6	6	5	1	0	0	2	0	1	9
SIGNAL TRANSDUCTION	**2919**	**2725**	**1477**	**1355**	**1151**	**25**	**77**	**388**	**142**	**20**	**3158**
Tetraspanin receptor complex	11	12	6	6	5	0	0	0	1	0	12
Acetylcholine receptor	17	19	8	7	9	0	1	1	2	0	20
Odorant binding proteins	16	16	4	4	11	0	0	1	1	0	17
Other signaling related proteins	4	4	2	2	2	0	0	0	0	0	4
Serine/Threonine kinase Act	11	11	9	9	2	0	0	0	0	0	11
Apoptosis related	51	49	27	25	22	0	1	4	1	1	54
GTPase activating ARF family	28	25	18	18	6	0	0	4	1	0	29
Arrestin related	13	11	6	6	5	0	0	2	0	0	13
Calcium dependent signaling	126	124	63	60	60	1	1	5	3	1	131
cAMP signaling related	148	139	75	70	61	2	3	15	5	0	156
CDC 42 related kinases	10	9	6	6	3	0	0	1	0	0	10
Intestinal cobalamin receptor	3	3	2	2	1	0	0	0	0	0	3
Cytokine related receptors	6	7	1	1	5	0	0	0	1	0	7
Dopamine related receptors	3	3	3	3	0	0	0	0	0	0	3
GABA related receptors	6	4	5	2	1	1	1	2	0	1	8
Membrane Glicoprotein LIG-1	9	10	6	5	4	0	1	0	0	0	10
Guanylate kinases	15	15	9	9	4	0	0	2	2	0	17
G protein-couple receptors	52	54	20	15	31	1	2	5	6	2	62
Other G proteins modulators	10	7	2	2	5	0	0	3	0	0	10
GTPases	52	39	24	24	14	0	0	14	1	0	53
GTP binding proteins	22	16	12	12	3	0	0	7	1	0	23
Hippo signaling pathway related	91	91	58	55	29	0	2	7	5	1	99
Insect hormone related	9	10	3	3	6	0	0	0	1	0	10
Integrin-like receptor related	5	3	2	2	1	0	0	2	0	0	5
JAK/STAT signaling pathway	74	72	37	35	29	2	0	8	8	0	82
Other signaling related	300	202	115	107	74	1	7	118	14	0	321
MAP kinase signaling	304	308	167	155	132	0	10	17	11	2	327

mTOR signaling pathway	11	11	6	6	5	0	0	0	0	0	11
Notch signaling pathway	8	11	8	6	2	0	2	0	1	0	11
Other GTPase activators	17	17	13	12	4	1	0	0	1	0	18
Phosphatases and related	31	27	18	18	9	0	0	4	0	0	31
Other signaling related	108	105	52	47	51	2	2	8	5	1	116
Other signaling related	574	510	254	236	248	6	8	84	18	4	604
Inositol phosphate pathway	70	68	47	41	20	1	3	8	4	2	79
Rab GTPase activator	56	66	36	34	22	0	2	0	8	0	66
Rac GTPase activators	6	6	3	3	3	0	0	0	0	0	6
RaI GTPase activators	20	19	14	13	5	0	1	2	0	0	21
Ran GTPase activators	9	9	7	7	2	0	0	0	0	0	9
Rap1 signaling pathway	89	87	42	39	39	1	1	10	8	1	99
Rap GTPase activators	8	9	4	4	4	0	0	0	1	0	9
Ras signaling pathway related	253	261	146	123	108	4	16	18	14	3	286
Ras GTPase activators	28	26	14	12	11	0	2	5	1	0	31
Rho GTPase activators	55	62	34	30	22	0	4	3	6	0	65
Signalosome related	5	4	2	2	2	0	0	1	0	0	5
Two component system	8	9	6	6	2	0	0	0	1	0	9
TNF receptors related	6	6	5	5	1	0	0	0	0	0	6
TonB receptors	13	0	0	0	0	0	0	13	0	0	13
Tyrosine phosphatases	46	47	20	19	26	1	0	0	2	0	48
Wnt signaling pathway	118	118	60	51	51	1	7	15	9	1	135
STORAGE	**62**	**58**	**17**	**16**	**39**	**0**	**1**	**7**	**2**	**0**	**65**
TRANSCRIPTION FACTORS	**735**	**744**	**382**	**355**	**346**	**4**	**18**	**30**	**25**	**5**	**783**
TRANSCRIPTION MACHNERY	**1438**	**1175**	**743**	**685**	**423**	**12**	**28**	**318**	**39**	**18**	**1523**
TRANSPORTERS	**1755**	**1392**	**702**	**616**	**656**	**6**	**68**	**477**	**52**	**12**	**1887**
UNKNOWN CONSERVED	**1638**	**1303**	**435**	**362**	**785**	**7**	**54**	**484**	**102**	**12**	**1806**
SECRETED	**2033**	**1937**	**963**	**879**	**935**	**4**	**59**	**215**	**64**	**21**	**2177**
8.9 kDa family	3	3	2	2	1	0	0	0	0	0	3
Cysteine-rich proteins	6	6	3	2	3	0	0	1	1	1	8

Cystatin	7	7	4	4	3	0	0	0	0	0	7
Secreted lipases and esterases	35	36	23	20	15	0	1	0	0	2	38
Secreted sugar enzymes	26	26	17	17	9	0	0	0	0	0	26
Glycine-rich proteins	29	30	27	25	4	0	1	0	0	1	31
Hebreains	4	4	4	4	0	0	0	0	0	0	4
Secreted immune related	97	103	91	81	16	0	6	0	0	4	107
Ixodegrins	3	3	1	1	2	0	0	0	0	0	3
Ixodidins	2	2	0	0	2	0	0	0	0	0	2
Kunitz containning domains	27	28	9	9	18	0	0	0	1	0	28
Llipocalins	58	59	13	13	45	0	0	0	1	0	59
Perilipins	16	18	4	4	12	0	0	0	2	0	18
Other secreted enzymes	225	217	120	111	93	0	7	21	6	2	240
Microplusins	13	13	6	6	7	0	0	0	0	0	13
Mucins	40	40	25	25	15	0	0	0	0	0	40
5'-Nucleotidases	9	9	8	8	1	0	0	0	0	0	9
Other secreted proteins	664	681	362	334	306	3	21	21	20	4	709
Peritrophins	12	12	9	9	3	0	0	0	0	0	12
Secreted proteases	129	128	68	66	59	0	2	4	1	0	132
Proline-rich proteins	7	7	4	3	4	0	0	0	0	1	8
Serpins	48	51	22	21	27	0	1	0	2	0	51
Secreted sufotransferase	573	454	141	114	290	1	20	168	30	6	629
TRANSPOSABLE ELEMENTS	**256**	**194**	**101**	**96**	**94**	**3**	**1**	**63**	**3**	**1**	**261**
UNKNOWN	**1487**	**886**	**195**	**152**	**609**	**3**	**30**	**723**	**95**	**10**	**1622**
TOTAL	**25569**	**21230**	**10697**	**9824**	**9839**	**107**	**594**	**5799**	**973**	**172**	**27308**

Spreadsheet S2. Functional classification and distribution of CDS at least 5-Fold more abundant among the *A. scultpum* internal organs

FUNCTIONAL CLASSES	MDG vs SG (5x)	SG vs MDG (5x)	MDG vs OVA (5x)	OVA vs MDG (5x)	OVA vs SG (5x)	SG vs OVA (5x)
CYTOSKELETAL						
Actin and related proteins	12	8	2	0	9	7
Actin binding proteins	5	15	1	0	4	15
Actin depolymerizing factors	1	2	1	0	0	2
Ankyrin related	11	24	0	0	12	24
Arp related	13	9	0	0	15	8
Arp 2/3 complex	0	4	0	0	2	4
Autophagy processes related	0	2	0	0	0	3
Calponin related	2	7	1	1	2	7
Catenin	1	1	0	0	1	1
CDC-42 related	0	4	0	0	0	4
Cell cycle processes related	16	23	7	0	15	20
Cohesin	1	0	0	0	1	0
Condensin	1	0	0	0	1	0
Coronin	1	1	0	0	1	0
Dynactins	2	3	1	0	1	2
Dyneins	1	10	0	0	2	10
Dystrobrevin	1	3	0	0	0	3
Dystrophin	0	3	0	0	0	3
FERM family	1	6	0	0	1	6
Filamin	1	3	0	0	1	3
Flagellins and others motility related	7	8	6	1	2	7
Formin	0	2	0	0	0	2
Keratin	0	0	0	0	0	0
Kinesin	4	22	0	0	6	20
Microtubule related	6	19	0	0	7	18
Muscle-like proteins	1	4	0	0	2	4
Myosin	14	45	3	1	15	44
Myotubularin	0	3	0	0	0	2
Nesprins	0	1	0	0	0	1
Nexins	2	15	0	2	5	13

Others related cytoskeletal proteins	6	12	0	0	5	11
Septin family	0	0	0	0	0	0
Spectrin related	3	11	0	0	3	11
Spindle pole related proteins	0	3	0	0	1	2
Thymosin	0	0	0	0	0	0
Connectins	0	3	0	0	0	3
Tropomodulins	0	1	0	0	0	1
Troponins	5	0	0	1	7	0
Tubulins	13	17	1	1	18	10
WASH domain related	0	0	0	0	0	0
WASP related proteins	0	2	0	0	0	2
WD40 containing domains	13	19	1	0	13	17
DETOXIFICATION						
Peroxidases	3	5	0	1	4	4
Arylsulfatase	7	2	0	0	5	2
Cytochrome oxidase	3	0	3	0	0	0
Dehydrogenases	16	9	3	1	18	8
Superoxide dismutase	3	1	0	1	6	1
Glutathione S-transferase	56	12	0	4	63	9
Glutathione synthetase	5	1	0	0	4	0
Glutathione related hydrolases	2	0	2	0	0	0
Multidrug resistance associated proteins	6	5	1	0	7	5
Other detoxification proteins	4	4	0	0	4	4
Oxidases	4	10	2	2	4	8
Peroxidases	2	2	0	0	2	2
Reductases	4	5	2	0	2	5
Selenium binding proteins	2	0	1	0	1	1
Sulfotransferases	64	12	0	3	79	9
DIGESTIVE ENZYMES: LIPID METABOLISM						
Acetylcholinesterases/Butyrylcholinesterases	8	3	0	1	17	4
Acetyl esterases	0	0	0	0	0	0
Carboxyesterases	7	0	0	0	10	0
Cholinesterases	0	1	0	0	0	1
Isoamyl-acetate esterases	10	0	0	0	11	0
Lipases	2	2	0	0	5	2
Other esterases	2	3	0	0	3	3
Phosphodiesterases	5	7	1	3	8	7

Phosphoesterases	1	0	0	0	2	0
Thioesterases	3	7	0	0	5	6
Triglyceride lipase-cholesterol esterases	1	0	0	0	1	0
DIGESTIVE ENZYMES: PROTEASES						
Aminopeptidases	3	4	0	0	4	3
Asparaginyl peptidases	10	0	0	0	10	0
Aspartyl proteases	7	1	2	0	4	1
ATP-dependent proteases	3	0	0	1	3	0
Other cysteine proteases	1	2	0	0	2	2
C48 family peptidases	0	1	0	0	1	1
C54 family peptidases	1	1	0	0	1	1
Carboxypeptidases	11	2	0	0	15	2
Cathepsin A	5	1	0	0	6	2
Cathepsin B	1	0	0	0	1	0
Cathepsin C	1	0	0	0	1	0
Cathepsin D	1	0	0	0	1	0
Cathepsin L	0	0	0	0	0	0
Dipeptidyl peptidases	2	3	0	0	1	2
Calpains cysteine proteases	1	1	0	0	1	1
Endopeptidases	0	2	0	0	0	2
Furin-like proteases	0	0	0	0	0	0
Membrane protease fragments	0	1	0	0	0	0
Insulinase superfamily	0	3	0	0	0	2
Tick legumains cysteine proteases	2	0	0	0	2	0
Angiotensinases	3	0	0	0	4	0
M20 peptidases	8	2	0	0	12	2
Oligopeptidases	4	0	0	0	4	0
Mitochondrial peptidases	1	5	0	0	1	4
Yolk cathepsin	1	0	0	1	2	0
Other proteases	2	5	0	1	4	4
Papain family cysteine proteases	3	0	0	0	3	0
Serine proteases	31	5	1	2	29	5
Gamma-glutamyltranspeptidases	1	0	0	0	1	0
Tripeptidyl peptidase	18	8	0	1	23	8
Proline-X peptidases	2	2	0	0	4	2
Metalloproteases	13	19	1	4	19	19
DIGESTIVE ENZYMES: SUGAR						

Amylases	3	5	0	0	3	4
Fucose kinase	1	0	0	0	2	0
Fucose synthase	0	0	0	0	0	0
Fucosidase	4	1	0	0	5	1
Fucosyltransferases	7	2	0	1	7	2
Glucanases	0	0	0	0	0	0
Glucosaminidase	3	2	0	0	3	2
Glucosidases	6	3	2	0	6	3
Hydrolases	0	0	0	0	0	0
Lectins	2	2	0	1	3	2
Glycosyl transferases	0	0	0	0	1	0
Maltases	4	1	0	0	5	1
Other sugar enzymes	1	3	0	0	1	3
Mannosidases	1	1	0	0	2	1
Mannosyl transferases	0	3	0	0	0	2
Melibiases	2	1	0	0	2	1
Oligosaccharyltransferases	1	1	0	0	2	1
EXTRACELLULAR MATRIX						
Ankyrin related	7	3	1	0	6	3
Cadherin related	0	3	0	0	0	3
Calsyntenin	0	0	0	0	1	0
Catenins	0	3	0	0	0	3
Cell adhesion antigens	2	13	0	0	1	13
Cell cycle related proteins	2	1	0	1	2	0
Cell Wall related proteins	2	0	0	1	3	0
Collagen related proteins	2	1	0	1	3	1
Cuticle related proteins	24	1	0	27	30	0
EM70 antigen related	0	4	0	0	0	4
Enzymes related	9	2	3	2	6	2
Fasciclin	0	1	0	0	0	1
Fibrillin	1	1	0	0	3	1
Fibronectin	0	14	0	0	1	11
Focaldhesin	1	1	0	0	1	1
Galectin	2	1	0	0	2	1
Glycosyltransferases	0	1	0	0	0	1
Glycine-rich proteins	1	2	1	1	0	1
GTPases	0	1	0	0	0	1
						110

Hemolysin	0	0	0	0	0	0
Innexin	0	2	0	0	0	2
Intergrin	0	1	0	0	0	1
Extracellular matrix related kinases	2	3	0	0	2	3
Lipoproteins	0	0	0	0	0	0
Mucins	3	4	0	0	4	4
O-glycosyl hidrolases	1	1	0	0	1	1
Outer membrane antigens	3	1	3	0	0	1
Others related enziymes	17	12	2	2	16	10
Peptidoglycan related protein	0	0	0	0	1	0
Phosphatidylinositol glycan related proteins	1	1	0	0	2	0
Glycoproteins	9	16	0	2	10	15
Secretion related	0	1	0	0	0	1
Syntrophins	0	1	0	0	0	1
TEM7 antigen	0	2	0	0	0	2
Tetraspanin	2	6	0	0	3	6
Connectin	3	0	3	0	0	0
Other transmembrane proteins	51	72	9	10	50	62
Transmembrane channels related	0	2	0	0	0	2
Vitronectin related	2	0	0	0	2	0
IMMUNE RELATED PRODUCTS						
Beta-lactamase	1	0	0	0	1	0
Immune related antigens	0	0	0	0	1	0
Complement related proteins	0	0	0	0	0	0
Cytokine related proteins	0	0	0	0	0	0
Immunoglobulin binding proteins	3	0	1	0	3	0
Immunoglobulins fragments	0	0	0	0	0	0
Interferon binding proteins	1	0	0	0	1	0
MHC related proteins	1	0	1	0	1	0
Other immune proteins	5	0	0	1	3	0
Penicillin binding proteins	1	0	1	0	0	0
Tumor necrosis facto receptors	24	0	0	0	27	0
Toll-lke receptors	2	2	0	0	2	2
AMINO ACIDS METABOLISM	**93**	**85**	**38**	**6**	**80**	**73**
CARBOHYDRATE METABOLISM						
Glycolysis and Gluconeogenesis related	25	15	5	0	28	13
Tricarboxilic acid cycle	20	9	11	0	10	8

Other carbon metabolism transcritps	17	21	4	0	17	16
Amino sugar metabolism	16	16	3	0	17	14
Chitinase related	14	4	0	2	20	2
Galactosidases	14	4	1	2	14	2
Pentoses phosphate pathway	10	7	3	3	10	7
Glyoxylate cycle related	6	1	3	1	4	1
Glucogen metabolism	4	7	2	0	3	6
Glucan related metabolism	3	3	0	0	5	3
Glycosyl hydrolases	3	4	1	0	2	4
Other dehydrogenases	3	6	1	0	2	5
Sulfatases	3	5	0	0	4	4
Galactosyltransferases	2	6	0	0	4	5
Inositol metabolism	2	2	0	0	2	2
Mannose metabolism	2	4	2	0	1	4
Others transferases	2	2	0	0	2	1
Carbonic anhydrase	1	2	0	0	1	2
Glucosidases	1	2	1	0	1	3
Trehalose assimilation	1	1	1	0	0	1
Xylose metabolism	1	1	1	0	0	1
Oxidative descarboxylation	0	1	0	0	0	0
Fucosyltransferases	0	0	0	0	0	0
Glycosyltransferases	0	12	0	1	1	10
Lactic fermentation	0	0	0	0	0	0
Polysaccharide metabolism	0	0	0	0	1	0
ENERGY METABOLISM	**166**	**113**	**48**	**8**	**149**	**101**
INTERMEDIATE METABOLISM	**258**	**215**	**98**	**26**	**237**	**188**
LIPID METABOLISM	**137**	**127**	**25**	**4**	**141**	**116**
NUCLEOTIDE METABOLISM	**61**	**40**	**15**	**3**	**62**	**37**
NUCLEAR EXPORT MACHNERY						
Exporter ABC superfamily	1	3	1	0	0	3
Exportin	1	1	1	0	0	1
Importin	0	13	0	0	1	11
Karyopherin	1	6	0	0	1	5
Nesprins	1	1	1	0	0	1
Nuclear receptors related	0	5	0	0	0	5
Nucleoporins	5	17	0	1	9	15
Nuclear transport regulators	1	1	1	0	0	1

Other nuclear export related	0	2	0	0	1	2
Translocators	4	3	3	0	1	3
NUCLEAR REGULATION						
Sister chromatid related proteins	1	3	0	0	2	3
Chromatin related proteins	4	15	0	0	4	14
Chromatin segregation proteins	2	6	1	0	2	6
Chromatin structural proteins	11	16	3	0	14	16
Cysteine/Serine-rich proteins	1	1	1	0	0	1
DNA polymerases	27	37	6	0	38	32
DNA binding proteins	10	37	1	1	19	35
DNA gyrase	2	0	2	0	0	0
DNA kinase related	1	1	0	0	2	0
DNA ligase	2	0	0	0	2	0
DNA methylase	4	9	1	1	4	8
DNA primase	0	1	0	0	1	1
Deoxyribonuclease	12	9	3	2	10	6
Nuclear envelope	0	2	0	0	1	2
Exonucleases	1	1	0	0	1	1
DNA glycosylase	1	0	0	0	2	0
H1, H2A, H3 and H4 histones	6	7	1	1	6	5
H2B histones	3	0	0	0	3	0
Histones acetylases	2	20	0	0	3	18
Histones deacetylases	1	10	0	0	3	8
Histones demethylases	1	7	1	0	0	7
HU family	0	0	0	0	0	0
Helicases	19	47	7	0	22	43
Histone methylases	3	9	1	0	2	8
Nuclear lamina related proteins	0	0	0	1	2	1
Leucine-rich repeat	3	6	0	1	3	6
Nucleases	0	2	0	0	0	2
Nucleolar related proteins	1	7	0	0	2	7
Other Nuclear proteins	24	28	9	2	20	24
Recombination machnery	2	0	2	0	0	0
Repair machnery	12	19	3	0	21	15
Other replication related proteins	4	10	0	1	7	6
Ribonucleases	1	3	0	0	1	3
Septins	0	1	0	0	0	1

Topoisomerases	8	11	4	0	10	10
DNA translocases	2	1	2	0	0	1
PROTEIN EXPORT MACHNERY						
Atlastin related	0	1	0	0	0	1
Clathrin	3	9	0	0	4	9
CLIP related	0	3	0	0	0	3
Vesicle coat complex proteins	2	17	0	0	6	15
EMP70 related	0	1	0	0	0	1
Endoplasmic reticulum	2	5	0	0	2	3
Exocyst complex component	2	8	0	0	3	7
Flotilin	1	0	1	0	0	0
FYVE domain	1	7	0	0	2	9
Golgi related enzymes	0	0	0	0	0	0
ER/Golgi transport vesicle proteins	1	1	0	0	1	2
Golgi vesicle related proteins	5	9	0	1	7	9
GTPase SAR1	6	19	1	0	7	16
Guanine exchange factor	2	3	0	0	2	3
Other vesicle related proteins	11	15	3	0	9	12
SEC family and related proteins	13	11	3	0	11	9
Reticulon related proteins	1	2	0	0	1	2
Signal peptidases	0	0	0	0	0	0
SNARE related proteins	4	4	0	0	4	2
Sorting related	22	34	2	1	20	32
Synaptobrevins and related proteins	1	4	0	0	1	4
Translocases related	2	0	1	0	2	0
TRAPP complex related	2	0	0	1	2	0
Vacuolar ATPases	10	1	0	2	13	0
PROTEIN MODIFICATION						
Acetyltransferases	0	2	0	0	1	2
ATPase	10	4	7	0	4	4
Other chaperones	10	18	6	4	11	15
DnaJ	18	20	3	0	19	18
DnaK	2	2	0	0	2	2
Hsp70	4	14	1	0	5	12
Hsp90	3	1	0	0	3	1
Tubulin-related chaperones	1	3	0	0	2	3
Chaperonins	12	6	2	2	11	4

Alpha-crystalins	7	3	0	2	7	2
Galectins	12	5	0	0	10	5
Peptidyl-prolyl cis-trans isomerases	9	9	2	1	13	7
Serine proteinase inhibitor (KU family)	4	1	1	0	5	1
Alpha-macroglobulin	2	10	0	0	3	9
Methyltransferases	5	6	0	1	4	6
Others protein modification related	24	43	1	4	28	35
Phosphatases	1	4	0	0	2	4
Proteases	30	35	16	3	14	34
Signal peptidases	1	2	0	0	1	2
Thioredoxin	8	7	0	3	12	5
Ubiquitin-like hydrolases	3	11	0	0	4	10
PROTEASOME MACHNERY						
ATPases	4	4	0	0	5	4
Neddylation related	5	2	0	0	8	2
Other proteasome proteins related	7	11	0	0	9	10
Other ubiquitin related	15	37	0	1	19	34
Proteasome regulatory machnery	17	17	3	3	22	15
Ubiquitin ligases	43	90	3	2	61	81
Ubiquitin hydrolasres	7	23	0	0	10	20
Ubiquitin proteases	1	4	0	0	1	3
PROTEIN SYNTHESIS MACHNERY						
Protein synthesis activator fators	0	2	0	0	0	2
Ribosome assembly protein	0	3	0	0	0	3
Ribosome biogenesis related protein	6	5	0	1	7	3
Translation elongation factors	10	7	3	0	10	7
Mitochondrial elongation factors	1	0	0	0	1	0
Other related enzymes	4	4	0	0	6	4
Translation initiation factors	19	15	2	2	21	10
Ribosome maturation related proteins	1	1	1	0	0	1
Other translation related	2	12	1	2	4	8
Pre-ribosome related proteins	2	2	0	0	3	2
Ribosome synthesis regulators	1	0	0	1	2	0
Polipeptide release factors	0	0	0	0	2	0
Ribosome binding proteins	6	2	2	0	8	2
Other ribosomal structural proteins	77	29	19	5	75	19
28S Ribosome strutural proteins	2	1	0	0	1	0

30S Ribosome strutural proteins	4	5	2	1	5	3
39S ribosome structural proteins	0	0	0	0	0	0
40S ribosome structural proteins	10	0	1	1	10	0
50S ribosome structural proteins	10	1	9	2	6	0
60S ribosome structural proteins	23	4	6	0	16	4
Mitochondrial ribosomes structural proteins	7	15	0	0	9	6
Mitochondrial 28S ribosome structural	5	4	0	0	4	2
Mitochondrial 37S ribosome structural	0	1	0	0	0	1
Mitochondrial 39S ribosome structural	3	2	0	0	5	0
rRNA methylases	12	2	10	0	2	2
rRNA maturation related proteins	0	0	0	0	0	0
Others tRNA related proteins	29	29	6	3	37	24
Glycyl-tRNA synthetases	2	1	2	0	0	1
Alanyl-tRNA synthetases	1	2	0	0	1	2
Leucyl-tRNA synthetases	2	1	0	0	3	1
Valyl-tRNA synthetases	3	1	3	0	0	1
Isoleucyl-tRNA synthetases	2	1	1	0	1	1
Methionyl-tRNA synthetases	3	2	1	0	2	0
Phenylalanyl-tRNA synthetases	0	2	0	0	0	0
Tyrosyl-tRNA synthetases	0	1	0	0	1	1
Tryptophanyl-tRNA synthetases	1	0	1	0	2	0
Cysteinyl-tRNA synthetases	0	2	0	0	1	1
Seryl-tRNA synthetases	2	0	0	1	4	0
Histidyl-tRNA synthetases	0	2	0	0	0	2
Threonyl-tRNA synthetases	0	1	0	0	0	0
Prolyl-tRNA synthetases	1	0	0	0	3	0
Asparaginyl-tRNA synthetases	0	1	0	0	0	1
Glutaminyl-tRNA synthetsases	1	2	1	0	0	0
Lysyl-tRNA synthetases	2	1	0	0	2	0
Arginyl-tRNA synthetase	2	2	0	0	2	1
Glutamyl-tRNA synthetsases	0	2	0	0	1	2
Aspartyl-tRNA synthetases	1	0	0	0	1	0
Other tRNA synthetase related	1	3	1	0	0	2
SECRETED						
8.9 kDa family	0	0	0	1	2	0
Cysteine-rich proteins	1	1	0	1	1	1
Cystatin	4	1	0	1	4	1

Secreted lipases and esterases	10	9	0	4	13	9
Secreted sugar enzymes	12	5	0	0	13	4
Glycine-rich proteins	15	10	0	16	15	7
Hebreains	0	2	0	1	0	2
Secreted immune related products	17	39	0	5	21	38
Ixodegrins	2	1	0	0	2	1
Ixodidins	0	0	0	0	2	0
Kunitz containning domains	16	7	7	2	12	6
Llipocalins	18	5	1	15	36	3
Perilipins	10	2	1	0	9	2
Other secreted enzymes	51	49	3	13	70	45
Microplusins	4	3	0	1	6	3
Mucins	17	4	0	14	21	4
5'-Nucleotidases	1	4	0	1	2	4
Odorant binding proteins	2	2	0	0	4	2
Other secreted proteins	201	117	4	63	260	109
Peritrophins	8	0	0	4	9	0
Secreted proteases	34	17	5	5	38	12
Other serine protease inhibidors	12	11	0	1	12	8
Proline-rich proteins	0	1	0	0	1	1
Serpins	11	15	0	1	12	14
Secreted sufotransferase	8	2	0	0	11	2
TIL	2	0	0	2	4	0
Secreted unknown conserved	100	44	34	22	89	35
Ixodidae related - SignalP	10	6	0	0	14	5
SIGNAL TRANSDUCTION						
Tetraspanin receptor complex	1	3	0	0	3	2
Acetylcholine receptor	0	5	0	0	1	5
Other signaling related proteins	1	1	0	0	1	1
Serine/Threonine proteon kinase Act	0	8	0	0	1	7
Apoptosis related	8	15	1	0	10	12
GTPase activating proteins: ARF family	4	8	1	0	4	7
Arrestin related	3	1	0	0	3	1
Calcium dependent signaling proteins	26	22	0	2	32	19
cAMP signaling related	9	38	2	0	14	36
CDC 42 related kinases	1	3	0	0	1	2
Intestinal cobalamin receptor	1	1	0	0	1	0

Cytokine related receptors	2	1	0	0	2	1
Dopamine related receptors	0	2	0	0	0	2
GABA related receptors	0	1	0	0	0	1
Membrane Glicoprotein LIG-1	1	1	0	0	3	1
Guanylate kinases	0	6	0	0	0	6
G protein-couple receptors (GPCRs)	3	7	0	0	5	7
Other G proteins modulators	0	1	0	1	1	1
GTPases	8	11	1	0	8	8
GTP binding proteins	5	8	3	0	3	7
Hippo signaling pathway related	9	33	1	0	9	32
Insect hormone related	2	0	0	0	2	0
Integrin-like receptor related	1	1	0	0	1	1
JAK/STAT signaling pathway related	11	16	5	0	7	15
Other signaling related kinases	35	57	14	2	24	53
MAP kinase signaling pathway related	33	94	3	2	42	89
mTOR signaling pathway related	1	1	0	0	1	1
Notch signaling pathway related	1	5	0	0	1	5
Other GTPase activators	0	10	0	0	0	10
Phosphatases and related proteins	3	9	0	0	4	9
Other signaling related receptors	12	21	2	1	15	20
Other signaling related proteins	75	131	13	7	82	118
Inositol phosphate related pathways proteins	3	25	0	0	6	25
Rab GTPase activator	5	16	0	1	8	13
Rac GTPase activators	1	1	0	0	2	1
RaI GTPase activators	0	11	0	0	1	12
Ran GTPase activators	1	4	0	0	2	3
Rap1 signaling pathway related	5	25	1	2	5	22
Rap GTPase activators	0	2	0	0	0	2
Ras signaling pathway related	21	80	3	0	28	72
Ras GTPase activators	0	4	0	1	2	4
Rho GTPase activators	7	16	1	1	9	14
Signalosome related	1	2	1	0	0	1
Two component system related	1	3	0	0	1	3
TNF receptors and related proteins	0	3	0	0	0	2
TonB receptors	2	0	2	0	0	0
Tyrosine phosphatases	10	10	1	0	11	10
Wnt signaling pathway	11	30	0	0	15	29

STORAGE						
Other storage related	4	0	0	0	5	0
Vitellogenin	30	0	15	0	23	0
Vitellogenin receptor	2	4	0	0	3	4
TRANSPOSABLE ELEMENTS						
Gag related family	0	12	0	0	0	12
Gypsy related family	0	0	0	0	1	0
Harbinger family	0	0	0	0	0	0
IS4 family	2	0	2	0	0	0
Other transposases related products	23	6	20	0	2	5
Outcast family related	0	0	0	0	0	0
Pogo family related	0	5	0	0	0	5
Pol polyprotein family	0	6	0	0	0	6
Tickrelated families transposases	10	21	2	0	10	22
Tigger family related	0	7	0	0	2	7
TRANSCRIPTION FACTORS						
Other related transcripts	3	19	3	0	2	3
5QNCA family	0	2	0	0	0	2
NEURO-D4	1	2	0	0	1	2
AT-hook	0	0	0	0	0	0
bHLH	0	0	0	1	0	0
Bromodomain	2	1	0	0	2	1
bZIP	3	6	0	0	4	6
C2H2	8	20	0	0	13	19
C3H1	0	1	0	0	0	1
CAP	0	1	0	0	0	1
CB2B	0	0	0	0	0	0
CBF	0	1	0	0	0	1
CCAAT	0	2	0	0	0	1
DAFT1	0	0	0	0	0	0
E2F	1	1	0	0	1	1
ETS-1	4	3	4	5	5	3
Forkhead related	5	7	0	0	5	6
Forkhead-HNF3	1	0	0	0	1	0
Forkhead-LRR	3	3	0	0	3	3
GABP	1	1	0	0	1	1
General transcription factors	0	9	0	0	1	7

GT-2	0	1	0	0	0	1
Homeobox	2	6	1	0	4	5
Homeobox-LRR	3	0	3	0	0	0
Homeobox-SIP1	0	2	0	0	1	2
HCFC	0	5	0	0	1	5
HLH	0	3	0	0	0	3
HMG	0	7	0	0	2	6
HOX	3	4	0	1	3	4
IWS1	1	1	1	0	0	1
Kayak	0	2	0	0	0	2
LIM	1	1	0	0	1	1
PHOX2-ARIX	0	0	0	0	1	0
MEIS1	0	2	0	0	0	1
LRR	4	3	0	0	5	3
MADS	0	0	0	0	0	0
MRTF	0	2	0	0	0	2
Myb	2	3	1	0	3	2
Myc	0	1	0	0	0	1
NAC	1	0	0	0	1	0
NF1	0	0	0	0	0	0
NFAT	0	3	0	0	0	2
NFKappaB	0	2	0	0	1	2
NFX	1	1	0	0	1	1
Ovo	0	1	0	0	0	1
p65	0	0	0	0	0	0
PBX	0	1	0	0	0	1
PCBP	0	2	0	0	0	1
PHD	0	3	0	0	0	2
POU	0	0	0	0	0	0
SOX	0	1	0	0	0	1
SP3	0	1	0	0	0	1
STAT	0	1	0	0	1	1
TBP	3	2	0	0	3	1
TBX	1	0	0	0	2	0
TCF	0	0	0	0	0	0
TEAD	0	0	0	0	0	0
Wimged-Helix	0	0	0	0	1	0
						120

XBP-1	1	0	0	0	1	0
Zinc finger	24	55	1	1	40	47
TRANSCRIPTION MACHNERY	**186**	**398**	**43**	**12**	**205**	**329**
TRANSPORTERS						
Other transporters	9	13	3	0	6	13
Amino acid and peptides transporters	24	46	13	0	16	43
Sodium symporters	17	42	2	2	24	42
ABC superfamily	27	18	17	2	13	18
Ammonia transporters	1	0	1	0	0	0
Aquaporins	3	3	0	0	3	3
ADP/ATP translocases	1	1	0	0	1	1
Biopolymer transporters	0	0	0	0	0	0
CARGO family transporters	0	1	0	0	0	1
CLC superfamily	11	7	0	0	9	7
Ferritins	0	0	0	0	0	0
Ion transporters	29	71	12	3	27	68
Lipid transporters	10	10	1	0	14	10
Major facilitator superfamily	7	7	4	0	5	7
Metal transporters	12	10	2	0	13	11
Metabolites transporters	31	26	4	2	37	22
Other mitocondrial transporters	3	10	0	1	7	9
Mechanossensitive channels	0	0	0	0	1	0
Multidrug transporters	4	1	2	0	3	1
Nitrogen permeases	0	0	0	0	0	0
Permeases	1	0	1	0	0	0
Phosphate transporters	0	1	0	0	0	1
Porins	0	1	0	0	1	1
RND family	0	0	0	0	0	0
Solute carrier family	2	3	0	1	2	3
Sec family transporters	4	6	2	0	2	6
Sugar/nucleotide transporters	15	16	3	0	16	15
Sulfate transporters	1	3	0	1	2	2
TRAM superfamily	0	1	0	0	0	1
Transportins	0	0	0	0	0	0
Putative transporters	1	2	0	1	1	2
V ATPase subunits	2	4	0	0	3	4
Voltage dependent Ca2+ channels	6	11	1	0	5	11
						121

Vesicle transporters	2	11	0	0	4	10
Vitamin transporters	5	0	0	0	5	0
UNKNOWN CONSERVED	**183**	**186**	**86**	**13**	**170**	**178**
UNKNOWN	**308**	**70**	**212**	**28**	**133**	**64**
VIRAL	**4**	**2**	**1**	**2**	**7**	**1**
TOTAL	**4333**	**4840**	**1085**	**464**	**4438**	**4318**

ANEXOS CAPÍTULO 2: RNASeq data analysis reveals deep changes in gene expression patterns of *Amblyomma sculptum* tick midguts in response to rickettsial infection

Spreadsheet S1. Functional classification of transcripts from the whole *A. sculptum* mialome (MDG)

CLASS	MDG		MI		MNI		MI vs MNI		
HOUSEKEEPING GENES									
CYTOSKELETAL	**CDS**	**Reads**	**CDS**	**Reads**	**CDS**	**Reads**	**DEGs**	**UP**	**DOWN**
Actin and related proteins	44	190931	42	176948	24	13983	33	29	4
Actin binding proteins	56	113468	55	64975	15	48493	21	19	2
Actin depolymerizing factors	6	2111	6	1903	4	208	5	4	1
Ankyrin related	75	11194	74	10211	20	983	41	37	4
Arp related	41	239226	40	168352	24	70874	36	33	3
Arp 2/3 complex	16	11631	15	11513	9	118	11	9	2
Autophagy processes related	9	2462	9	2434	2	28	6	5	1
Calponin related	22	40482	22	39415	13	1067	20	17	3
Catenin	4	3379	4	712	1	2667	2	1	1
CDC-42 related	6	53	5	34	2	19	2	0	2
Cell cycle processes related	117	38252	96	19174	51	19078	81	55	26
Cohesin	2	1806	2	1800	1	6	2	2	0
Condensin	2	1831	2	1548	2	283	1	1	0
Coronin	4	2096	4	1910	2	186	3	3	0
Dynactins	11	186091	9	8438	7	177653	8	6	2
Dyneins	25	2821	23	2362	11	459	9	7	2
Dystrobrevin	5	5727	5	4139	1	1588	3	3	0
Dystrophin	12	210	12	207	1	3	3	3	0
FERM family	16	3659	16	3537	5	122	9	9	0
Filamin	21	9715	20	9597	13	118	12	8	4
Flagellins and others motility related	63	9059	19	8274	50	785	51	14	37
Formin	4	1303	4	1288	1	15	4	4	0
Keratin	4	88	2	41	3	47	2	0	2
Kinesin	77	12214	72	11884	25	330	33	25	8
Microtubule related	52	44992	52	36050	18	8942	31	25	6
Muscle-like proteins	19	14400	19	14214	7	186	11	10	1
Myosin	138	175916	128	144701	57	31215	81	71	10
Myotubularin	9	658	9	558	3	100	3	2	1

Nesprins	3	50	3	10	2	40	1	0	1
Nexins	36	31156	36	30716	9	440	27	27	0
Others related cytoskeletal proteins	57	27170	52	21045	16	6125	35	27	8
Septin family	3	98	3	42	2	56	2	1	1
Spectrin related	36	32092	35	30381	16	1711	18	15	3
Spindle pole related proteins	6	1619	5	1615	1	4	5	5	0
Thymosin	2	2437	2	2405	1	32	1	1	0
Connectins	6	72	6	61	2	11	0	0	0
Tropomodulins	4	238	4	222	1	16	3	3	0
Troponins	15	44683	15	43123	13	1560	15	15	0
Tubulins	88	148812	86	136763	41	12049	53	46	7
WASH domain related	5	49	0	0	5	49	4	0	4
WASP related proteins	6	408	6	389	2	19	2	2	0
WD40 containing domains	51	25058	50	24495	24	563	35	31	4
DETOXIFICATION									
Peroxidases	35	24446	30	23330	13	1116	20	15	5
Arylsulfatase	27	5343	27	5144	6	199	19	19	0
Cytochrome oxidase	8	177	2	5	8	172	6	0	6
Dehydrogenases	75	185358	52	179890	46	5468	60	41	19
Superoxide dismutase	17	12466	14	10985	8	1481	12	10	2
Glutathione S-transferase	102	1369488	96	1256643	63	112845	86	83	3
Glutathione synthetase	6	1487	6	1484	1	3	6	6	0
Glutathione related hydrolases	5	104	1	2	4	102	4	0	4
Multidrug resistance associated proteins	60	3496	38	3126	26	370	36	18	18
Other detoxification proteins	44	66875	30	65611	17	1264	25	12	13
Oxidases	48	5363	36	4975	18	388	29	16	13
Peroxidases	8	48933	5	48782	6	151	7	5	2
Reductases	24	15859	12	15326	17	533	17	8	9
Selenium binding proteins	5	13763	4	13694	2	69	5	4	1
Sulfotransferases	150	62846	149	61822	30	1024	120	118	2
DIGESTIVE ENZYMES: LIPID METABOLISM									
Acetylcholinesterases/Butyrylcholinesterases	44	7279	44	7219	4	60	21	20	1
Acetyl esterases	5	677	4	659	2	18	5	3	2
Carboxyesterases	18	37132	17	36422	7	710	14	13	1
Cholinesterases	2	306	2	306	0	0	1	1	0
Isoamyl-acetate esterases	13	48324	13	46770	10	1554	12	12	0
Lipases	19	11813	19	11374	4	439	10	10	0

Other esterases	9	2882	9	2825	5	57	8	8	0
Phosphodiesterases	34	5252	32	4733	10	519	22	18	4
Phosphoesterases	3	946	2	933	2	13	3	2	1
Thioesterases	15	4924	15	4831	7	93	14	14	0
Triglyceride lipase-cholesterol esterases	8	9053	8	8737	2	316	4	4	0
DIGESTIVE ENZYMES: PROTEASES									
Aminopeptidases	14	67263	14	61847	8	5416	13	12	1
Asparaginyl peptidases	13	202620	13	201147	10	1473	11	11	0
Aspartyl proteases	12	13567	12	13438	4	129	12	12	0
ATP-dependent proteases	6	7506	5	7488	3	18	4	4	0
Other cysteine proteases	6	5424	6	5335	3	89	6	5	1
C48 family peptidases	2	210	2	206	1	4	2	2	0
C54 family peptidases	2	1474	2	1474	0	0	2	2	0
Carboxypeptidases	24	12170	24	11922	10	248	19	19	0
Cathepsin A	17	16521	17	14511	4	2010	14	13	1
Cathepsin B	1	57549	1	57280	1	269	1	1	0
Cathepsin C	1	3524	1	3516	1	8	1	1	0
Cathepsin D	1	1851	1	1851	0	0	1	1	0
Cathepsin L	3	12298	3	11918	2	380	2	2	0
Dipeptidyl peptidases	8	14534	8	14258	4	276	7	6	1
Calpains cysteine proteases	5	14956	5	14324	5	632	5	4	1
Endopeptidases	4	2910	4	2868	3	42	2	2	0
Furin-like proteases	1	51	1	35	1	16	0	0	0
Membrane protease fragments	1	2160	1	2160	0	0	1	1	0
Insulinase superfamily	8	14168	8	13645	3	523	6	6	0
Tick legumains cysteine proteases	2	1214	2	1214	0	0	2	2	0
Angiotensinases	5	16039	5	10745	2	5294	4	4	0
M20 peptidases	25	10751	25	9425	9	1326	16	13	3
Oligopeptidases	6	3234	6	3200	4	34	6	6	0
Mitochondrial peptidases	9	1003	9	679	3	324	4	3	1
Yolk cathepsin	2	1829	2	1767	1	62	1	1	0
Other proteases	18	29019	17	27185	9	1834	14	12	2
Papain family cysteine proteases	4	14058	4	13581	3	477	4	4	0
Serine proteases	42	265824	41	261504	14	4320	33	32	1
Gamma-glutamyltranspeptidases	6	12086	5	12003	3	83	3	2	1
Tripeptidyl peptidase	61	148126	58	143230	27	4896	43	39	4
Proline-X peptidases	11	20284	11	19683	5	601	7	7	0

Metalloproteases	97	65463	92	62846	35	2617	69	62	7
DIGESTIVE ENZYMES: SUGAR									
Amylases	17	3520	17	3113	6	407	8	8	0
Fucose kinase	2	923	2	917	1	6	2	2	0
Fucose synthase	1	1815	1	1773	1	42	1	1	0
Fucosidase	6	10105	6	9293	2	812	6	6	0
Fucosyltransferases	16	4256	16	4186	3	70	11	10	1
Glucanases	2	76	2	51	1	25	1	1	0
Glucosaminidase	8	3333	8	3254	3	79	6	6	0
Glucosidases	23	24313	21	17679	10	6634	12	9	3
Hydrolases	2	26	2	26	0	0	1	1	0
Lectins	12	2629	12	2554	4	75	9	8	1
Glycosyl transferases	2	125	2	125	0	0	2	2	0
Maltases	8	39352	8	38498	5	854	6	6	0
Other sugar enzymes	12	10573	12	10353	3	220	9	9	0
Mannosidases	10	486	10	460	4	26	6	6	0
Mannosyl transferases	6	3317	6	3317	0	0	4	4	0
Melibiases	7	3926	7	3870	3	56	6	5	1
Oligosaccharyltransferases	8	32775	8	31562	7	1213	7	7	0
EXTRACELLULAR MATRIX									
Ankyrin related	23	8927	21	8159	8	768	16	13	3
Cadherin related	12	589	11	581	3	8	4	3	1
Calsyntenin	4	60	4	52	2	8	1	1	0
Catenins	11	3858	11	3573	2	285	4	4	0
Cell adhesion antigens	35	54435	35	53835	8	600	12	11	1
Cell cycle related proteins	6	972	6	972	0	0	4	4	0
Cell Wall related proteins	9	3687	5	3595	7	92	7	4	3
Collagen related proteins	20	24351	20	23869	7	482	13	11	2
Cuticle related proteins	52	159117	50	153292	24	5825	36	34	2
EM70 antigen related	8	582	8	568	2	14	2	2	0
Enzymes related	35	10411	16	6127	25	4284	29	11	18
Fasciclin	5	3487	4	3401	2	86	3	2	1
Fibrillin	5	1890	4	1827	2	63	5	4	1
Fibronectin	21	2687	19	2574	6	113	13	10	3
Focaldhesin	3	772	2	400	2	372	3	1	2
Galectin	4	11409	4	11361	1	48	4	4	0
Glycosyltransferases	6	97	1	14	5	83	6	1	5

Glycine-rich proteins	7	1576	6	953	3	623	5	3	2
GTPases	5	68	4	61	1	7	2	1	1
Hemolysin	4	11590	3	10170	3	1420	2	1	1
Innexin	4	126	4	87	2	39	4	2	2
Intergrin	7	286	7	254	1	32	4	4	0
Extracellular matrix related kinases	8	745	8	693	3	52	5	4	1
Lipoproteins	3	16	0	0	3	16	1	0	1
Mucins	16	4590	15	4452	6	138	7	6	1
O-glycosyl hidrolases	2	4467	2	4363	2	104	2	2	0
Outer membrane antigens	31	4053	3	3494	29	559	29	2	27
Others related enziymes	74	94759	64	74562	33	20197	49	37	12
Peptidoglycan related protein	4	111	2	94	3	17	3	2	1
Phosphatidylinositol glycan related proteins	4	2679	4	2673	1	6	4	4	0
Glycoproteins	59	110400	58	107950	19	2450	31	27	4
Secretion related	7	408	4	252	4	156	4	1	3
Syntrophins	4	36	4	36	0	0	1	1	0
TEM7 antigen	3	19	3	19	0	0	0	0	0
Tetraspanin	18	19485	18	18905	9	580	12	12	0
Connectin	6	200	0	0	6	200	6	0	6
Other transmembrane proteins	342	138269	253	125565	167	12704	240	147	93
Transmembrane channels related	15	685	13	655	4	30	8	6	2
Vitronectin related	8	465	8	441	2	24	3	3	0
IMMUNE RELATED PRODUCTS									
Beta-lactamase	4	4487	2	4299	3	188	2	1	1
Immune related antigens	2	119	2	114	1	5	1	1	0
Complement related proteins	1	3	1	3	0	0	0	0	0
Cytokine related proteins	2	6	2	6	0	0	0	0	0
Immunoglobulin binding proteins	5	338	4	261	3	77	2	1	1
Immunoglobulins fragments	5	61	5	15	1	46	1	0	1
Interferon binding proteins	2	3837	2	3729	1	108	1	1	0
MHC related proteins	3	175	2	50	2	125	3	1	2
Other immune proteins	22	1587	18	1514	8	73	9	7	2
Penicillin binding proteins	6	82	0	0	6	82	5	0	5
Tumor necrosis facto receptors	50	10524	47	10240	18	284	43	42	1
Toll-lke receptors	11	860	11	755	3	105	5	4	1
AMINO ACIDS METABOLISM	580	366763	367	236850	334	129913	423	218	205
CARBOHYDRATE METABOLISM									

Amino sugar metabolism	72	54984	60	52775	32	2209	47	34	13
Carbonic anhydrase	6	4933	6	4788	4	145	5	5	0
Chitinase related	49	102177	47	100684	13	1493	34	33	1
Oxidative descarboxylation	4	777	2	761	2	16	2	1	1
Fucosyltransferases	2	23	2	11	1	12	0	0	0
Galactosidases	29	41762	26	34480	13	7282	22	19	3
Galactosyltransferases	18	4544	18	4326	10	218	15	12	3
Glucan related metabolism	22	2884	13	2748	11	136	17	8	9
Glucosidases	10	4961	6	4589	8	372	8	3	5
Glucogen metabolism	30	8506	23	8042	14	464	21	15	6
Glycosyl hydrolases	14	8365	12	8150	9	215	9	7	2
Glycolysis and Gluconeogenesis related	153	328582	101	279706	97	48876	113	58	55
Glycosyltransferases	30	2035	29	1943	5	92	15	13	2
Glyoxylate cycle related	15	1784	8	1202	11	582	11	3	8
Inositol metabolism	10	902	10	868	3	34	6	6	0
Lactic fermentation	5	3252	4	3201	2	51	3	2	1
Mannose metabolism	24	22217	18	15759	10	6458	16	7	9
Other dehydrogenases	38	26306	25	25964	19	342	19	7	12
Pentoses phosphate pathway	60	41977	39	40913	34	1064	41	20	21
Polysaccharide metabolism	4	78	2	56	2	22	4	2	2
Sulfatases	17	3832	17	3746	4	86	14	13	1
Tricarboxilic acid cycle	83	80232	47	76976	55	3256	64	33	31
Others transferases	8	10781	8	10591	7	190	7	7	0
Trehalose assimilation	13	3341	3	22	10	3319	8	0	8
Xylose metabolism	6	95	5	79	1	16	4	3	1
Other carbon metabolism transcritps	125	723050	93	145753	77	577297	88	59	29
ENERGY METABOLISM	758	909324	519	747315	457	162009	619	382	237
INTERMEDIATE METABOLISM	1720	974039	1011	577627	969	396412	1203	553	650
LIPID METABOLISM	751	787049	602	587707	328	199342	493	334	159
NUCLEOTIDE METABOLISM	291	419692	177	315995	178	103697	227	121	106
NUCLEAR EXPORT MACHNERY									
Exporter ABC superfamily	7	166	4	58	4	108	4	1	3
Exportin	8	171	4	68	5	103	3	0	3
Importin	23	5728	23	5265	10	463	12	10	2
Karyopherin	15	24291	15	23838	7	453	10	10	0
Nesprins	5	87	3	49	2	38	2	1	1
Nuclear receptors related	15	5006	13	4785	6	221	10	8	2

Nucleoporins	42	10903	41	10225	14	678	30	27	3
Nuclear transport regulators	6	57	4	9	3	48	2	0	2
Other nuclear export related	4	1115	4	1044	2	71	2	2	0
Translocators	20	57081	10	776	13	56305	11	3	8
NUCLEAR REGULATION									
Sister chromatid related proteins	7	2837	7	2733	2	104	4	3	1
Chromatin related proteins	34	6339	33	5613	14	726	26	24	2
Chromatin segregation proteins	19	26182	18	23072	9	3110	14	12	2
Chromatin structural proteins	66	8328	53	7858	25	470	41	29	12
Cysteine/Serine-rich proteins	3	112	2	45	2	67	1	0	1
DNA polymerases	188	61791	139	32350	82	29441	124	75	49
DNA binding proteins	118	20988	108	19595	44	1393	67	55	12
DNA gyrase	14	238	0	0	14	238	14	0	14
DNA kinase related	13	1839	13	1682	3	157	5	4	1
DNA ligase	15	1217	7	1136	10	81	9	3	6
DNA methylase	37	5657	33	5289	11	368	22	18	4
DNA primase	7	263	4	208	3	55	5	2	3
Deoxyribonuclease	44	8045	32	6494	21	1551	34	22	12
Nuclear envelope	4	134	4	134	0	0	1	1	0
Exonucleases	19	5161	10	4988	11	173	13	7	6
DNA glycosylase	6	1047	3	1015	4	32	4	2	2
H1, H2A, H3 and H4 histones	35	61891	29	61211	21	680	27	20	7
H2B histones	9	12968	8	12534	9	434	8	7	1
Histones acetylases	36	5813	33	5530	15	283	22	18	4
Histones deacetylases	21	22035	21	12432	9	9603	13	11	2
Histones demethylases	9	3355	8	1338	2	2017	3	2	1
HU family	1	7	1	7	0	0	0	0	0
Helicases	197	58141	133	51536	107	6605	136	80	56
Histone methylases	22	6363	22	2165	4	4198	14	12	2
Nuclear lamina related proteins	5	1141	4	1107	2	34	5	4	1
Leucine-rich repeat	16	6875	16	6709	5	166	11	11	0
Nucleases	8	122	7	101	1	21	3	2	1
Nucleolar related proteins	22	1910	22	1749	7	161	15	13	2
Other Nuclear proteins	126	99812	109	38362	54	61450	88	66	22
Recombination machnery	12	159	2	7	10	152	10	0	10
Repair machnery	116	18793	92	13671	48	5122	77	53	24
Other replication related proteins	34	3967	32	3617	11	350	20	17	3

Ribonucleases	10	1692	5	1339	6	353	7	3	4
Septins	3	604	3	566	1	38	1	1	0
Topoisomerases	72	5810	51	4393	32	1417	41	19	22
DNA translocases	3	367	1	297	3	70	3	1	2
PROTEIN EXPORT MACHNERY									
Atlastin related	4	2429	4	2221	1	208	2	2	0
Clathrin	24	17469	24	17191	9	278	16	15	1
CLIP related	3	38	3	38	0	0	1	1	0
Vesicle coat complex proteins	40	24216	40	23658	17	558	23	23	0
EMP70 related	2	12	2	12	0	0	0	0	0
Endoplasmic reticulum	10	2524	9	2416	5	108	10	9	1
Exocyst complex component	15	5882	15	5797	5	85	14	14	0
Flotilin	2	38	0	0	2	38	2	0	2
FYVE domain	24	8614	21	8427	10	187	18	16	2
Golgi related enzymes	0	0	0	0	0	0	0	0	0
ER/Golgi transport vesicle proteins	6	7611	6	7085	4	526	5	5	0
Golgi vesicle related proteins	27	49303	27	48768	10	535	16	15	1
GTPase SAR1	46	32593	46	31274	24	1319	42	41	1
Guanine exchange factor	6	9327	6	8057	1	1270	4	4	0
Other vesicle related proteins	64	59047	55	56744	28	2303	47	34	13
SEC family and related proteins	67	15479	49	14951	31	528	42	23	19
Reticulon related proteins	3	1604	3	1592	1	12	3	3	0
Signal peptidases	2	23	1	5	1	18	1	0	1
SNARE related proteins	17	42990	17	40062	6	2928	12	11	1
Sorting related	117	195821	105	65438	46	130383	78	64	14
Synaptobrevins and related proteins	13	10664	12	10478	5	186	7	7	0
Translocases related	4	4294	4	4279	2	15	2	2	0
TRAPP complex related	6	29460	6	26108	5	3352	6	5	1
Vacuolar ATPases	28	44873	25	37468	13	7405	22	19	3
PROTEIN MODIFICATION									
Acetyltransferases	4	2063	4	1967	1	96	2	2	0
ATPase	57	33977	20	32398	45	1579	38	5	33
Other chaperones	65	424163	45	13366	38	410797	50	29	21
DnaJ	71	24561	64	23256	33	1305	54	45	9
DnaK	18	2842	13	2760	9	82	7	4	3
Hsp70	48	12690	40	12254	21	436	27	20	7
Hsp90	14	117564	10	54216	9	63348	8	4	4

Tubulin-related chaperones	7	2544	7	2465	2	79	4	4	0
Chaperonins	40	40813	25	39302	31	1511	36	20	16
Alpha-crystalins	14	36920	14	35634	10	1286	14	14	0
Galectins	31	23974	30	23682	12	292	24	23	1
Peptidyl-prolyl cis-trans isomerases	67	33879	56	26521	32	7358	42	31	11
Serine proteinase inhibitor (KU family)	17	32921	17	32590	7	331	10	7	3
Alpha-macroglobulin	19	6034	19	5848	6	186	13	12	1
Methyltransferases	19	8330	18	8068	5	262	11	9	2
Others protein modification related	142	135729	128	78082	59	57647	106	90	16
Phosphatases	25	25671	22	18412	7	7259	11	7	4
Proteases	232	91476	146	22941	121	68535	143	64	79
Signal peptidases	8	4418	8	4017	3	401	5	5	0
Thioredoxin	40	63238	32	54029	18	9209	32	24	8
Ubiquitin-like hydrolases	25	18447	25	15231	14	3216	20	17	3
PROTEASOME MACHNERY									
ATPases	16	14344	16	10311	12	4033	14	13	1
Neddylation related	16	216537	16	214593	8	1944	13	13	0
Other proteasome proteins related	34	93529	34	89918	18	3611	29	26	3
Other ubiquitin related	101	53867	99	52138	46	1729	70	67	3
Proteasome regulatory machnery	71	262452	69	117299	38	145153	61	56	5
Ubiquitin ligases	302	148833	298	124202	97	24631	175	158	17
Ubiquitin hydrolasres	61	19063	58	17759	18	1304	31	29	2
Ubiquitin proteases	9	8681	9	8547	5	134	6	6	0
PROTEIN SYNTHESIS MACHNERY									
Protein synthesis activator fators	6	3301	6	3175	3	126	3	1	2
Ribosome assembly protein	4	2183	4	2098	3	85	3	2	1
Ribosome biogenesis related protein	24	22585	23	20135	15	2450	21	18	3
Translation elongation factors	44	96632	35	93867	25	2765	26	17	9
Mitochondrial elongation factors	6	6911	4	6864	4	47	2	2	0
Other related enzymes	26	7860	20	7233	16	627	22	17	5
Translation initiation factors	112	174752	103	159893	66	14859	80	66	14
Ribosome maturation related proteins	3	1274	1	291	2	983	3	1	2
Other translation related	30	24716	25	19924	21	4792	23	19	4
Pre-ribosome related proteins	8	2366	8	2341	2	25	4	4	0
Ribosome synthesis regulators	4	3728	3	3294	3	434	4	3	1
Polipeptide release factors	6	23561	3	22885	5	676	6	3	3
Ribosome binding proteins	18	11058	17	10151	11	907	13	9	4

Other ribosomal structural proteins	230	4220964	202	2834676	168	1386288	198	132	66
28S Ribosome strutural proteins	3	32097	3	6836	3	25261	3	2	1
30S Ribosome strutural proteins	31	97358	24	88392	15	8966	15	8	7
39S ribosome structural proteins	2	1915	2	1914	1	1	1	1	0
40S ribosome structural proteins	50	273807	39	234430	35	39377	33	16	17
50S ribosome structural proteins	42	6154	18	4309	26	1845	30	8	22
60S ribosome structural proteins	74	846117	61	603027	57	243090	51	25	26
Mitochondrial ribosomes structural proteins	41	47674	41	44714	30	2960	38	36	2
Mitochondrial 28S ribosome structural	14	10438	14	9494	6	944	10	10	0
Mitochondrial 37S ribosome structural	0	0	0	0	0	0	0	0	0
Mitochondrial 39S ribosome structural	11	9844	11	8424	7	1420	11	11	0
rRNA methylases	24	270465	17	4663	15	265802	17	7	10
rRNA maturation related proteins	2	7	2	7	0	0	0	0	0
Others tRNA related proteins	162	45158	119	42921	80	2237	118	79	39
Glycyl-tRNA synthetases	9	13679	3	11118	8	2561	8	1	7
Alanyl-tRNA synthetases	8	3749	4	3673	7	76	6	3	3
Leucyl-tRNA synthetases	11	14219	7	14028	6	191	7	4	3
Valyl-tRNA synthetases	18	587	7	355	11	232	13	2	11
Isoleucyl-tRNA synthetases	13	320	4	163	10	157	10	2	8
Methionyl-tRNA synthetases	8	4828	6	4507	6	321	6	4	2
Phenylalanyl-tRNA synthetases	9	2664	6	2450	7	214	8	3	5
Tyrosyl-tRNA synthetases	6	157	4	131	2	26	4	2	2
Tryptophanyl-tRNA synthetases	4	81	2	47	2	34	3	1	2
Cysteinyl-tRNA synthetases	8	7669	7	7453	3	216	6	5	1
Seryl-tRNA synthetases	12	9999	11	9722	6	277	7	6	1
Histidyl-tRNA synthetases	8	7760	4	7117	7	643	6	3	3
Threonyl-tRNA synthetases	8	1521	3	1256	7	265	6	1	5
Prolyl-tRNA synthetases	11	1901	6	1778	7	123	7	3	4
Asparaginyl-tRNA synthetases	5	1608	2	1411	4	197	4	1	3
Glutaminyl-tRNA synthetsases	6	1277	4	502	3	775	4	2	2
Lysyl-tRNA synthetases	7	3494	6	3387	3	107	6	5	1
Arginyl-tRNA synthetase	5	1877	3	1850	2	27	5	3	2
Glutamyl-tRNA synthetsases	9	2662	5	2571	6	91	7	4	3
Aspartyl-tRNA synthetases	8	622	4	566	5	56	8	4	4
Other tRNA synthetase related	8	4837	6	4486	4	351	6	4	2
SIGNAL TRANSDUCTION									
Tetraspanin receptor complex	11	6922	11	6277	3	645	9	9	0

Acetylcholine receptor	17	869	16	708	4	161	9	7	2
Other signaling related proteins	4	1922	4	1898	1	24	3	3	0
Serine/Threonine proteon kinase Act	11	6928	11	6766	3	162	8	8	0
Apoptosis related	51	23888	50	15339	20	8549	34	31	3
GTPase activating proteins: ARF family	28	6315	26	6018	13	297	17	14	3
Arrestin related	13	27721	11	27015	7	706	8	7	1
Calcium dependent signaling proteins	126	74426	123	72381	51	2045	76	70	6
cAMP signaling related	148	27721	133	18056	57	9665	85	67	18
CDC 42 related kinases	10	2934	9	2761	3	173	7	7	0
Intestinal cobalamin receptor	3	1167	3	1068	2	99	3	3	0
Cytokine related receptors	6	571	6	530	2	41	6	5	1
Dopamine related receptors	3	3199	3	3129	1	70	2	2	0
GABA related receptors	6	20	6	19	1	1	0	0	0
Membrane Glicoprotein LIG-1	9	1134	9	1123	1	11	6	6	0
Guanylate kinases	15	4016	13	3177	8	839	7	5	2
G protein-couple receptors (GPCRs)	52	9815	50	8333	12	1482	15	13	2
Other G proteins modulators	10	531	8	444	4	87	4	2	2
GTPases	54	20909	39	18733	35	2176	44	25	19
GTP binding proteins	22	52525	14	42824	17	9701	21	13	8
Hippo signaling pathway related	91	23146	87	21613	33	1533	50	44	6
Insect hormone related	9	4401	8	4332	3	69	7	6	1
Integrin-like receptor related	5	1439	4	1401	4	38	4	3	1
JAK/STAT signaling pathway related	74	168357	68	13339	31	155018	38	26	12
Other signaling related kinases	300	82760	186	46659	173	36101	207	94	113
MAP kinase signaling pathway related	304	174845	290	166784	99	8061	171	147	24
mTOR signaling pathway related	11	3115	11	2844	3	271	4	3	1
Notch signaling pathway related	8	1919	8	1841	2	78	4	4	0
Other GTPase activators	17	273	17	234	2	39	4	3	1
Phosphatases and related proteins	31	57474	29	18857	12	38617	18	13	5
Other signaling related receptors	108	48267	97	32645	29	15622	52	38	14
Other signaling related proteins	574	577683	498	518132	227	59551	354	272	82
Inositol phosphate related pathways proteins	70	22154	63	21034	21	1120	41	35	6
Rab GTPase activator	56	11100	56	9127	19	1973	23	21	2
Rac GTPase activators	6	544	6	463	2	81	4	3	1
RaI GTPase activators	20	6541	18	4652	7	1889	12	9	3
Ran GTPase activators	9	3342	9	3195	4	147	6	6	0
Rap1 signaling pathway related	89	40716	81	39279	27	1437	37	26	11

Rap GTPase activators	8	37	8	34	2	3	0	0	0
Ras signaling pathway related	253	222792	236	132477	90	90315	128	108	20
Ras GTPase activators	28	4578	24	4486	10	92	6	5	1
Rho GTPase activators	55	280833	50	27185	22	253648	34	27	7
Signalosome related	5	2057	4	1987	4	70	4	3	1
Two component system related	8	8293	8	8171	3	122	3	3	0
TNF receptors and related proteins	6	2452	6	2377	2	75	5	5	0
TonB receptors	13	211	0	0	13	211	11	0	11
Tyrosine phosphatases	46	10470	44	10067	14	403	30	28	2
Wnt signaling pathway	118	17800	102	16485	40	1315	60	40	20
STORAGE	62	1278307	59	1272642	33	5665	49	43	6
TRANSCRIPTION FACTORS									
Other related transcripts	89	39472	82	37827	30	1645	52	39	13
5QNCA family	4	710	4	710	0	0	2	2	0
NEURO-D4	5	2258	5	2199	2	59	3	3	0
AT-hook	1	8	1	4	1	4	0	0	0
bHLH	5	168	5	168	0	0	2	2	0
Bromodomain	9	8884	9	6463	2	2421	5	4	1
bZIP	15	37126	15	36560	7	566	12	12	0
C2H2	62	73634	62	70097	14	3537	33	32	1
C3H1	2	38	2	14	1	24	1	0	1
CAP	2	51	2	51	0	0	2	2	0
CB2B	2	24	2	14	1	10	1	0	1
CBF	2	120	2	120	0	0	1	1	0
CCAAT	2	984	2	974	1	10	2	2	0
DAFT1	1	1	1	1	0	0	0	0	0
E2F	15	685	11	634	4	51	6	3	3
Forkhead related	20	4137	20	4036	3	101	15	14	1
Forkhead-HNF3	9	8453	9	8395	3	58	4	4	0
Forkhead-LRR	20	4014	19	3840	5	174	9	9	0
GABP	12	1089	12	1041	2	48	5	5	0
General transcription factors	13	4840	13	4805	3	35	11	10	1
GT-2	2	214	2	214	0	0	2	2	0
Homeobox	28	1516	24	1371	7	145	16	12	4
Homeobox-LRR	4	16236	2	2	4	16234	4	0	4
Homeobox-SIP1	11	279	11	247	2	32	4	4	0
HCFC	9	727	9	713	3	14	6	6	0

HLH	5	509	5	396	2	113	3	3	0
HMG	18	8962	18	8771	7	191	8	8	0
HOX	17	1391	17	1147	2	244	5	4	1
IWS1	2	6636	2	73	1	6563	2	1	1
Kayak	2	39	2	37	1	2	1	1	0
LIM	3	248	3	248	0	0	2	2	0
PHOX2-ARIX	2	90	2	90	0	0	1	1	0
MEIS1	4	637	4	637	0	0	2	2	0
LRR	22	1373	20	1126	7	247	8	6	2
MADS	3	59	3	59	0	0	2	2	0
MRTF	4	147	4	147	0	0	2	2	0
Myb	11	8516	10	2753	3	5763	9	8	1
Myc	5	362	5	345	1	17	4	4	0
NAC	3	1126	3	1073	3	53	3	3	0
NF1	4	4339	4	4326	1	13	2	2	0
NFAT	4	4509	4	4508	1	1	3	3	0
NFKappaB	4	1336	4	1328	1	8	3	3	0
NFX	3	2659	3	2348	3	311	3	3	0
Ovo	3	43	3	43	0	0	2	2	0
p65	1	17	1	17	0	0	1	1	0
PBX	4	1156	4	1096	1	60	3	3	0
PCBP	2	1092	2	1050	1	42	2	2	0
PHD	6	569	6	515	2	54	1	1	0
POU	4	32	4	26	1	6	1	1	0
SOX	3	709	3	709	0	0	1	1	0
SP3	2	440	2	434	1	6	1	1	0
STAT	6	2872	6	2853	2	19	3	3	0
TBP	7	4783	7	4687	5	96	7	7	0
TBX	4	531	4	530	1	1	3	3	0
TCF	2	7	2	7	0	0	0	0	0
TEAD	2	21	2	21	0	0	1	1	0
Wimged-Helix	1	314	1	314	0	0	1	1	0
XBP-1	2	104673	2	102861	2	1812	2	2	0
Zinc finger	226	66497	220	52922	59	13575	119	113	6
TRANSCRIPTION MACHNERY	1438	1537187	1146	1394151	742	143036	1000	717	283
TRANSPORTERS									
Other transporters	72	59307	52	18620	33	40687	44	24	20

Amino acid and peptides transporters	183	56162	136	48163	88	7999	119	68	51
Sodium symporters	176	38316	164	36828	53	1488	105	90	15
ABC superfamily	293	73293	94	32367	217	40926	218	33	185
Ammonia transporters	3	60	0	0	3	60	3	0	3
Aquaporins	10	7503	10	7346	4	157	6	6	0
ADP/ATP translocases	7	8611	7	8475	4	136	4	3	1
Biopolymer transporters	2	22	0	0	2	22	2	0	2
CARGO family transporters	4	5619	4	5406	3	213	4	4	0
CLC superfamily	25	27468	25	25017	8	2451	22	20	2
Ferritins	4	12181	2	11900	3	281	3	1	2
Ion transporters	276	44923	230	31477	97	13446	156	100	56
Lipid transporters	63	23481	49	22565	30	916	40	25	15
Major facilitator superfamily	38	2138	18	1565	25	573	29	9	20
Metal transporters	60	27953	52	27306	24	647	43	38	5
Metabolites transporters	174	74069	150	38841	52	35228	117	94	23
Other mitocondrial transporters	39	57021	39	48127	17	8894	31	30	1
Mechanossensitive channels	4	57	1	13	4	44	3	0	3
Multidrug transporters	24	8541	8	2740	17	5801	21	7	14
Nitrogen permeases	2	38	2	32	1	6	0	0	0
Permeases	6	107	1	19	6	88	5	0	5
Phosphate transporters	5	48	3	22	2	26	1	0	1
Porins	9	542	2	405	9	137	9	1	8
RND family	4	38	1	1	4	37	3	0	3
Solute carrier family	15	2356	15	2352	1	4	8	8	0
Sec family transporters	29	13226	18	12566	20	660	23	12	11
Sugar/nucleotide transporters	106	11208	89	10162	41	1046	76	55	21
Sulfate transporters	14	2030	8	1941	7	89	10	6	4
TRAM superfamily	2	187	2	163	1	24	1	1	0
Transportins	3	9	3	9	0	0	0	0	0
Putative transporters	9	1231	9	1210	3	21	5	5	0
V ATPase subunits	14	11096	12	8887	9	2209	8	5	3
Voltage dependent Ca2+ channels	38	5499	36	5299	12	200	19	15	4
Vesicle transporters	28	18008	27	17653	14	355	21	20	1
Vitamin transporters	14	2684	13	2591	6	93	10	10	0
UNKNOWN CONSERVED	1638	783171	1159	217099	701	566072	873	413	460
TRANSPOSABLE ELEMENTS									
Gag related family	23	1284	22	675	21	609	15	5	10

Gypsy related family	3	106	2	76	2	30	2	1	1
Harbinger family	5	158	5	158	0	0	4	4	0
IS4 family	6	125	0	0	6	125	6	0	6
Other transposases related products	83	30218	28	2568	66	27650	70	12	58
Outcast family related	4	107	4	23	4	84	3	0	3
Pogo family related	8	932	8	857	3	75	7	6	1
Pol polyprotein family	11	711	9	284	10	427	8	0	8
Tickrelated families transposases	97	6973	91	3764	74	3209	50	11	39
Tigger family related	16	623	16	463	9	160	11	6	5
SECRETED									
8.9 kDa family	3	98	3	98	0	0	3	3	0
Cysteine-rich proteins	6	525	5	507	4	18	4	3	1
Cystatin	7	274470	7	241032	4	33438	7	7	0
Secreted lipases and esterases	35	50051	35	49208	14	843	26	25	1
Secreted sugar enzymes	26	112692	26	107243	15	5449	26	26	0
Glycine-rich proteins	29	341770	28	339470	20	2300	28	27	1
Hebreains	4	2957	4	2922	2	35	4	4	0
Secreted immune related products	97	119743	96	117995	39	1748	77	75	2
Ixodegrins	3	41577	3	28171	2	13406	2	2	0
Ixodidins	2	65	2	65	0	0	2	2	0
Kunitz containning domains	27	86479	27	86459	2	20	24	24	0
Llipocalins	58	16631	58	16250	10	381	35	35	0
Perilipins	16	32219	16	31965	8	254	14	14	0
Other secreted enzymes	225	331725	204	269143	92	62582	165	140	25
Microplusins	13	46164	13	42618	4	3546	7	7	0
Mucins	40	114164	39	110987	20	3177	30	29	1
5'-Nucleotidases	9	4250	9	4157	2	93	7	7	0
Odorant binding proteins	16	2403	14	2280	4	123	10	8	2
Other secreted proteins	664	5496486	642	2615895	255	2880591	494	459	35
Peritrophins	12	23764	12	23565	7	199	12	12	0
Secreted proteases	86	1394111	84	1069273	39	324838	72	68	4
Other serine protease inhibidors	43	14615	42	14402	9	213	29	28	1
Proline-rich proteins	7	203	7	200	1	3	4	4	0
Serpins	48	12446	48	12332	9	114	39	38	1
Secreted sufotransferase	16	14072	16	12434	8	1638	16	15	1
TIL	7	104076	6	96484	4	7592	6	5	1
Secreted unknown conserved	494	383704	333	153237	229	230467	298	150	148

Ixodidae related - SignalP	56	19101	56	16483	9	2618	28	24	4
UNKNOWN	**1468**	**4899306**	**764**	**92754**	**873**	**4806552**	**818**	**200**	**618**
VIRAL	**19**	**8700**	**14**	**5553**	**11**	**3147**	**13**	**7**	**6**
TOTAL	**25569**	**40295145**	**20013**	**24790632**	**12116**	**15504513**	**16849**	**11167**	**5682**

ANEXOS CAPÍTULO 3: The first functional genomics of *Amblyomma sculptum* tick ovaries, under conditions of infection and non-infection with *Rickettsia amblyommii*

Spreadsheet S1. Functional classification of transcripts from the whole transcriptome of *A. sculptum* ovaries (OVA)

FUNCTIONAL CLASSES	**OVA**		**OI**		**ONI**		**OI *versus* ONI**		
CYTOSKELETAL	**CDS**	**Reads**	**CDS**	**Reads**	**CDS**	**Reads**	**DEGs**	**UP**	**Down**
Actin and related proteins	42	201829	40	99703	40	102124	15	11	4
Actin binding proteins	55	127850	47	71336	47	56484	12	8	5
Actin depolymerizing factors	4	2066	4	1055	4	1011	3	2	1
Ankyrin related	86	17174	82	7982	82	9183	16	8	8
Arp related	44	264275	42	141129	42	123105	23	15	9
Arp 2/3 complex	16	18652	16	9210	16	9442	4	4	0
Autophagy processes related	9	2243	9	1139	9	1104	3	3	0
Calponin related	20	67454	19	35016	19	32435	12	9	3
Catenin	3	1706	3	827	3	879	0	0	0
CDC-42 related	6	24	4	10	4	11	0	0	0
Cell cycle processes related	93	36503	86	15596	86	20882	27	10	17
Cohesin	2	2925	2	1381	2	1544	0	0	0
Condensin	2	2365	2	1206	2	1159	2	2	0
Coronin	5	3720	4	1728	4	1987	2	1	1
Dynactins	10	12253	8	5980	8	6267	2	2	0
Dyneins	26	5044	24	2123	24	2915	6	4	2
Dystrobrevin	5	5625	5	2844	5	2781	2	2	0

Dystrophin	12	268	11	113	11	153	2	1	1
FERM family	16	4707	16	2162	16	2545	5	2	3
Filamin	15	12393	14	6605	14	5781	6	5	1
Flagellins and others motility related	19	11662	19	5667	19	5995	5	3	2
Formin	4	2789	4	1159	4	1630	2	0	2
Keratin	2	49	2	26	2	23	0	0	0
Kinesin	78	20411	74	9712	74	10683	17	11	6
Microtubule related	54	51517	52	26349	52	25165	17	10	7
Muscle-like proteins	18	27609	16	14684	16	12918	3	3	0
Myosin	132	358099	128	191201	128	166889	48	32	16
Myotubularin	9	992	8	560	8	431	2	2	0
Nesprins	4	66	4	41	4	25	1	1	0
Nexins	36	42162	35	21269	35	20889	14	6	8
Others related cytoskeletal proteins	59	17806	53	7400	53	10371	13	7	7
Septin family	1	29	1	16	1	13	0	0	0
Spectrin related	37	40675	35	21515	35	19156	14	11	3
Spindle pole related proteins	6	3643	6	1468	6	2175	3	0	3
Thymosin	1	3843	1	1955	1	1888	1	1	0
Connectins	8	144	8	100	8	44	3	3	0
Tropomodulins	5	587	5	288	5	299	0	0	0
Troponins	15	79969	15	39177	15	40792	12	9	3
Tubulins	92	344404	79	158116	79	186253	34	10	25
WASH domain related	0	0	0	0	0	0	0	0	0
WASP related proteins	6	1011	6	515	6	496	1	1	0
WD40 containing domains	50	36508	50	16393	50	20115	22	8	14
DETOXIFICATION									
Peroxidases	28	38130	25	16854	25	21266	9	4	5
Arylsulfatase	29	7084	28	3019	28	4061	16	4	12
Cytochrome oxidase	0	0	0	0	0	0	0	0	0
Dehydrogenases	52	199231	52	93729	52	105502	26	11	15
Superoxide dismutase	15	29999	14	13719	14	16264	8	3	6

Glutathione S-transferase	97	1337591	96	772867	96	564722	66	42	24
Glutathione synthetase	6	2434	6	1255	6	1179	3	2	1
Glutathione related hydrolases	1	1	1	1	0	0	0	0	0
Multidrug resistance associated proteins	42	10195	39	6321	39	3868	10	6	4
Other detoxification proteins	33	68192	30	33872	30	34310	6	5	1
Oxidases	36	4911	34	1573	34	3327	13	7	7
Peroxidases	6	43570	4	24491	4	19061	4	2	3
Reductases	13	45938	11	18264	11	27666	6	1	5
Selenium binding proteins	4	8735	4	4880	4	3855	3	2	1
Sulfotransferases	152	74750	151	34721	151	40028	73	37	36
DIGESTIVE ENZYMES: LIPID METABOLISM									
Acetylcholinesterases/Butyrylcholinesterases	45	8062	42	3880	42	4170	16	6	10
Acetyl esterases	4	887	4	413	4	474	0	0	0
Carboxyesterases	18	71120	15	32049	15	39060	11	6	5
Cholinesterases	2	134	2	93	2	41	1	1	0
Isoamyl-acetate esterases	13	35683	13	24930	13	10753	10	6	4
Lipases	19	10524	19	5168	19	5356	9	5	4
Other esterases	9	4168	9	1894	9	2274	6	2	4
Phosphodiesterases	35	8428	32	4302	32	4117	14	7	7
Phosphoesterases	2	1802	2	769	2	1033	2	1	1
Thioesterases	16	10163	15	4304	15	5850	7	2	5
Triglyceride lipase-cholesterol esterases	8	6446	7	3602	7	2794	2	1	2
DIGESTIVE ENZYMES: PROTEASES									
Aminopeptidases	15	74968	15	35764	15	39204	8	5	3
Asparaginyl peptidases	14	207859	14	94707	14	113152	10	5	5
Aspartyl proteases	12	15150	12	5714	12	9436	11	1	10
ATP-dependent proteases	7	9496	7	4957	7	4539	5	2	3
Other cysteine proteases	6	11478	6	5255	6	6223	3	1	2
C48 family peptidases	2	747	2	276	2	471	2	0	2
C54 family peptidases	2	2236	2	1133	2	1103	1	1	0
Carboxypeptidases	24	14196	24	7920	24	6276	16	8	8

Cathepsin A	18	28937	17	11145	17	17782	10	2	9
Cathepsin B	2	63399	2	25195	2	38204	1	0	1
Cathepsin C	1	2568	1	1174	1	1394	0	0	0
Cathepsin D	1	750	1	349	1	401	0	0	0
Cathepsin L	3	19760	3	10076	3	9684	1	1	0
Dipeptidyl peptidases	8	21143	8	10303	8	10840	4	3	1
Calpains cysteine proteases	7	22856	7	9940	7	12916	3	2	1
Endopeptidases	6	3674	5	1943	5	1723	2	2	0
Furin-like proteases	1	49	1	27	1	22	0	0	0
Membrane protease fragments	1	5343	1	1956	1	3387	1	0	1
Insulinase superfamily	8	21188	8	10157	8	11031	6	2	4
Tick legumains cysteine proteases	2	648	1	4	1	554	1	0	2
Angiotensinases	5	26469	4	10809	4	15637	4	2	3
M20 peptidases	26	8976	26	3600	26	5376	10	5	5
Oligopeptidases	6	3114	6	1591	6	1523	2	1	1
Mitochondrial peptidases	9	1695	8	701	8	992	5	2	3
Yolk cathepsin	2	581	1	39	1	455	1	0	2
Other proteases	19	49111	19	26792	19	22319	6	4	2
Papain family cysteine proteases	4	18389	4	7864	4	10525	3	2	1
Serine proteases	42	185845	42	104639	42	81206	29	15	14
Gamma-glutamyltranspeptidases	5	13603	5	6256	5	7347	3	2	1
Tripeptidyl peptidase	58	124584	56	74037	56	50539	30	19	11
Proline-X peptidases	12	26418	12	12456	12	13962	4	2	2
Metalloproteases	103	74734	100	34511	100	40214	41	18	23
DIGESTIVE ENZYMES: SUGAR									
Amylases	18	6923	17	3086	17	3826	5	2	4
Fucose kinase	2	1739	2	864	2	875	0	0	0
Fucose synthase	1	3748	1	1340	1	2408	1	0	1
Fucosidase	6	10929	6	5211	6	5718	3	1	2
Fucosyltransferases	17	9948	17	4039	17	5909	6	0	6
Glucanases	2	95	2	61	2	34	1	1	0

Glucosaminidase	8	4010	8	1832	8	2178	3	1	2
Glucosidases	22	24145	22	12887	22	11258	9	4	5
Hydrolases	2	43	2	16	2	27	0	0	0
Lectins	12	4518	12	1992	12	2526	5	2	3
Glycosyl transferases	2	398	2	158	2	240	2	1	1
Maltases	8	39839	8	21325	8	18514	5	2	3
Other sugar enzymes	13	21661	13	9041	13	12620	5	0	5
Mannosidases	11	1309	11	435	11	874	2	0	2
Mannosyl transferases	6	4825	6	2193	6	2632	1	0	1
Melibiases	7	4261	6	1884	6	2360	3	1	3
Oligosaccharyltransferases	8	56720	8	25739	8	30981	7	2	5
EXTRACELLULAR MATRIX									
Ankyrin related	23	13964	21	5654	21	8303	7	3	4
Cadherin related	12	740	9	397	9	335	1	1	0
Calsyntenin	4	111	4	64	4	47	2	2	0
Catenins	12	8042	12	3917	12	4125	3	3	0
Cell adhesion antigens	44	47719	38	26857	38	20830	11	6	6
Cell cycle related proteins	6	3210	6	1259	6	1951	3	1	2
Cell Wall related proteins	5	6486	5	3579	5	2907	3	3	0
Collagen related proteins	21	111466	20	45878	20	65587	10	6	4
Cuticle related proteins	50	1799050	50	962403	50	836647	28	12	16
EM70 antigen related	8	915	8	448	8	467	1	1	0
Enzymes related	23	17916	20	8582	20	9124	6	2	5
Fasciclin	5	7671	5	5290	5	2381	2	2	0
Fibrillin	4	1993	4	985	4	1008	3	1	2
Fibronectin	21	6233	21	2543	21	3690	8	4	4
Focaldhesin	3	1030	2	457	2	571	0	0	0
Galectin	4	58136	4	26003	4	32133	4	2	2
Glycosyltransferases	1	26	1	9	1	17	0	0	0
Glycine-rich proteins	6	3652	6	1457	6	2195	2	1	1
GTPases	5	265	3	99	3	164	1	0	1

Hemolysin	2	17841	2	8447	2	9394	0	0	0
Innexin	5	338	3	154	3	172	0	0	0
Intergrin	10	373	10	205	10	168	1	1	0
Extracellular matrix related kinases	9	1363	9	829	9	534	4	4	0
Mucins	18	13641	16	6798	16	6840	6	5	1
O-glycosyl hidrolases	2	4814	2	2374	2	2440	2	1	1
Outer membrane antigens	3	5434	3	2348	3	3086	2	0	2
Others related enziymes	67	109884	65	61195	65	48686	20	11	9
Peptidoglycan related protein	4	264	2	121	2	137	0	0	0
Phosphatidylinositol glycan related proteins	4	2628	4	824	4	1804	3	1	2
Glycoproteins	61	90805	59	64255	59	26545	19	11	8
Secretion related	4	472	3	232	3	227	0	0	1
Syntrophins	4	95	4	40	4	55	0	0	0
TEM7 antigen	3	29	3	16	3	13	0	0	0
Tetraspanin	18	19485	18	9923	18	9562	7	4	3
Connectin	0	0	0	0	0	0	0	0	0
Other transmembrane proteins	268	544395	250	296041	250	248262	89	40	51
Transmembrane channels related	12	659	11	336	11	317	3	3	0
Vitronectin related	9	1277	8	399	8	868	2	0	3
IMMUNE RELATED PRODUCTS									
Beta-lactamase	2	6938	2	3792	2	3146	1	1	0
Immune related antigens	1	545	1	248	1	297	0	0	0
Complement related proteins	1	1	0	0	0	0	0	0	0
Cytokine related proteins	2	7	1	1	1	2	0	0	0
Immunoglobulin binding proteins	4	528	4	217	4	311	2	0	2
Immunoglobulins fragments	4	96	3	32	3	49	1	0	2
Interferon binding proteins	2	5273	2	2443	2	2830	1	1	0
MHC related proteins	1	498	1	115	1	383	1	0	1
Other immune proteins	21	2513	19	817	19	1634	4	2	3
Penicillin binding proteins	0	0	0	0	0	0	0	0	0
Tumor necrosis facto receptors	49	16809	47	6920	47	9883	22	4	18

Toll-lke receptors	11	586	10	316	10	268	5	4	1
AMINO ACIDS METABOLISM	**388**	**406683**	**372**	**192108**	**372**	**214497**	**154**	**64**	**93**
CARBOHYDRATE METABOLISM									
Amino sugar metabolism	66	65663	63	31360	63	34296	29	13	16
Carbonic anhydrase	6	10244	6	4571	6	5673	4	1	3
Chitinase related	51	124086	46	74496	46	49570	22	18	4
Oxidative descarboxylation	2	2120	2	855	2	1265	1	0	1
Fucosyltransferases	2	30	2	9	2	21	0	0	0
Galactosidases	26	55224	25	24373	25	30849	14	3	11
Galactosyltransferases	18	7701	18	3686	18	4015	6	4	2
Glucan related metabolism	17	3887	12	1824	12	2032	4	1	5
Glucosidases	5	10841	5	4445	5	6396	2	0	2
Glucogen metabolism	27	12881	26	5845	26	7006	11	5	7
Glycosyl hydrolases	13	10580	13	4875	13	5705	6	2	4
Glycolysis and Gluconeogenesis related	92	480536	89	223991	89	256522	46	21	26
Glycosyltransferases	36	3149	33	1327	33	1805	7	1	7
Glyoxylate cycle related	9	2124	7	1023	7	1097	1	0	1
Inositol metabolism	9	1581	9	691	9	890	3	1	2
Lactic fermentation	4	3659	4	1547	4	2112	1	0	1
Mannose metabolism	17	39047	13	19114	13	19925	7	4	3
Other dehydrogenases	29	28175	25	14178	25	13949	6	4	3
Pentoses phosphate pathway	41	94527	39	43321	39	51202	17	7	10
Polysaccharide metabolism	2	154	1	39	1	98	1	0	2
Sulfatases	18	5054	18	2263	18	2791	9	4	5
Tricarboxilic acid cycle	51	141531	50	64777	50	76751	20	6	14
Others transferases	8	19645	8	9245	8	10400	6	3	3
Trehalose assimilation	4	78	4	29	4	49	0	0	0
Xylose metabolism	4	101	4	35	4	66	1	0	1
Other carbon metabolism transcritps	95	204387	91	108788	91	95584	39	17	22
ENERGY METABOLISM	**539**	**1371403**	**524**	**661971**	**524**	**709331**	**274**	**135**	**142**
INTERMEDIATE METABOLISM	**1093**	**926779**	**1022**	**455915**	**1022**	**470339**	**380**	**156**	**234**

LIPID METABOLISM	**636**	**848038**	**595**	**414012**	**595**	**433782**	**235**	**110**	**131**
NUCLEOTIDE METABOLISM	**182**	**420828**	**173**	**205844**	**173**	**214941**	**78**	**27**	**52**
NUCLEAR EXPORT MACHNERY									
Exporter ABC superfamily	4	174	3	75	3	95	1	1	0
Exportin	4	100	3	48	3	50	0	0	0
Importin	24	11266	20	4552	20	6706	9	1	8
Karyopherin	16	40877	16	18885	16	21992	7	2	5
Nesprins	5	82	4	54	4	26	1	1	0
Nuclear receptors related	13	6953	13	3666	13	3287	4	4	0
Nucleoporins	42	32210	38	12023	38	20171	18	3	15
Nuclear transport regulators	4	32	3	12	3	15	0	0	0
Other nuclear export related	4	2029	4	914	4	1115	0	0	0
Translocators	10	1629	9	694	9	931	3	1	2
NUCLEAR REGULATION									
Sister chromatid related proteins	9	5118	9	2641	9	2477	2	2	0
Chromatin related proteins	33	11698	32	5078	32	6608	14	6	9
Chromatin segregation proteins	17	51825	17	25598	17	26227	8	5	3
Chromatin structural proteins	56	14027	55	5880	55	8142	23	6	17
Cysteine/Serine-rich proteins	2	74	2	32	2	42	0	0	0
DNA polymerases	140	60952	137	25846	137	35093	57	12	45
DNA binding proteins	116	39793	111	17161	111	22587	38	13	26
DNA gyrase	0	0	0	0	0	0	0	0	0
DNA kinase related	11	2993	11	1456	11	1537	3	2	1
DNA ligase	8	2029	7	899	7	1127	2	0	2
DNA methylase	35	11938	33	5037	33	6895	8	0	8
DNA primase	4	521	4	213	4	308	2	0	2
Deoxyribonuclease	33	11487	33	5073	33	6414	10	6	4
Nuclear envelope	4	345	4	133	4	212	1	0	1
Exonucleases	10	7752	10	2826	10	4926	2	0	2
DNA glycosylase	3	1299	3	682	3	617	2	1	1
H1, H2A, H3 and H4 histones	32	122817	32	50750	32	72067	17	5	12

H2B histones	8	46872	8	17749	8	29123	6	1	5
Histones acetylases	35	11436	34	4899	34	6535	9	0	9
Histones deacetylases	26	18921	26	9018	26	9903	7	3	4
Histones demethylases	13	2913	10	1332	10	1572	1	0	1
HU family	1	2	1	2	0	0	0	0	0
Helicases	140	93174	133	42453	133	50690	44	12	33
Histone methylases	24	4499	23	1941	23	2556	4	1	3
Nuclear lamina related proteins	4	1554	4	595	4	959	4	0	4
Leucine-rich repeat	19	11991	18	6221	18	5767	8	3	5
Nucleases	7	358	6	112	6	234	1	0	2
Nucleolar related proteins	21	4578	20	1725	20	2838	6	0	7
Other Nuclear proteins	116	103182	112	42119	112	61054	36	10	26
Recombination machnery	3	12	2	8	2	3	0	0	0
Repair machnery	98	27521	94	12242	94	15259	31	6	25
Other replication related proteins	31	10252	30	3801	30	6447	14	0	14
Ribonucleases	6	2454	6	1168	6	1286	0	0	0
Septins	3	923	3	465	3	458	2	1	1
Topoisomerases	53	9257	48	4060	48	5188	13	3	10
DNA translocases	1	503	1	211	1	292	0	0	0
PROTEIN EXPORT MACHNERY									
Atlastin related	4	3037	3	1421	3	1592	0	0	1
Clathrin	25	24784	25	11748	25	13036	9	3	6
CLIP related	3	72	3	31	3	41	0	0	0
Vesicle coat complex proteins	40	37752	38	17880	38	19866	14	7	7
EMP70 related	3	61	3	31	3	30	0	0	0
Endoplasmic reticulum	9	7481	9	3158	9	4323	6	2	4
Exocyst complex component	15	9663	15	4362	15	5301	4	2	2
Flotilin	0	0	0	0	0	0	0	0	0
FYVE domain	21	12108	20	6112	20	5995	10	7	3
Golgi related enzymes	2	3	2	3	0	0	0	0	0
ER/Golgi transport vesicle proteins	7	8996	7	4872	7	4124	6	4	2

Golgi vesicle related proteins	29	32829	27	21231	27	11591	14	6	8
GTPase SAR1	47	47483	46	21660	46	25819	20	7	13
Guanine exchange factor	8	7246	8	2477	8	4769	3	0	3
Other vesicle related proteins	55	62244	50	33768	50	28456	17	8	9
SEC family and related proteins	51	20041	47	10241	47	9787	14	8	6
Reticulon related proteins	3	1845	3	922	3	923	1	1	0
Signal peptidases	1	3	1	1	1	2	0	0	0
SNARE related proteins	19	55816	18	20001	18	35809	9	5	4
Sorting related	112	131445	103	57889	103	73541	34	17	17
Synaptobrevins and related proteins	14	10875	14	4694	14	6181	4	3	1
Translocases related	4	1206	4	694	4	512	1	1	0
TRAPP complex related	6	99116	6	52296	6	46820	5	5	0
Vacuolar ATPases	25	104692	24	59193	24	45498	14	7	7
PROTEIN MODIFICATION									
Acetyltransferases	4	2798	3	1273	3	1521	2	0	2
ATPase	22	33955	17	18780	17	15164	5	3	2
Other chaperones	48	33040	48	13857	48	19183	22	12	10
DnaJ	68	55515	65	22532	65	32978	32	12	20
DnaK	15	4760	14	2093	14	2654	4	2	3
Hsp70	48	29698	46	12180	46	17514	18	7	11
Hsp90	12	111531	10	49793	10	61723	3	0	4
Tubulin-related chaperones	7	5542	6	2598	6	2935	1	1	0
Chaperonins	28	158981	28	84732	28	74249	18	9	9
Alpha-crystalins	15	50273	15	25035	15	25238	9	4	5
Galectins	32	20234	28	12229	28	7991	16	14	2
Peptidyl-prolyl cis-trans isomerases	57	56890	55	24958	55	31900	24	7	18
Serine proteinase inhibitor (KU family)	17	9158	14	2863	14	6287	4	2	2
Alpha-macroglobulin	20	10664	18	4173	18	6482	10	2	8
Methyltransferases	20	26035	19	16120	19	9911	5	2	3
Others protein modification related	134	162858	131	92701	131	70147	52	20	32
Phosphatases	24	32586	18	17883	18	14687	3	2	1

Proteases	158	34654	139	15546	139	19033	34	19	17
Signal peptidases	9	8407	8	3794	8	4608	4	2	2
Thioredoxin	36	104931	34	48983	34	55865	19	10	10
Ubiquitin-like hydrolases	26	30052	26	12845	26	17207	12	1	11
PROTEASOME MACHNERY									
ATPases	15	19790	15	9131	15	10659	8	3	5
Neddylation related	16	411782	16	191136	16	220646	8	5	3
Other proteasome proteins related	36	90305	35	52324	35	37979	16	8	8
Other ubiquitin related	102	80815	99	38596	99	42190	33	14	20
Proteasome regulatory machnery	71	208292	71	100777	71	107515	43	21	22
Ubiquitin ligases	324	179027	307	82044	307	96867	94	43	52
Ubiquitin hydrolasres	64	34604	58	15659	58	18931	14	6	8
Ubiquitin proteases	10	15238	10	7097	10	8141	1	0	1
PROTEIN SYNTHESIS MACHNERY									
Protein synthesis activator fators	8	5734	8	2781	8	2953	1	1	0
Ribosome assembly protein	4	3876	4	1632	4	2244	2	0	2
Ribosome biogenesis related protein	23	42398	23	17530	23	24868	15	2	13
Translation elongation factors	33	144273	28	67653	28	76598	12	6	7
Mitochondrial elongation factors	6	4356	6	2389	6	1967	2	1	1
Other related enzymes	21	13859	21	6018	21	7841	8	1	7
Translation initiation factors	99	312382	98	145059	98	167321	54	29	25
Ribosome maturation related proteins	1	945	1	367	1	578	1	0	1
Other translation related	28	47263	27	18691	27	28568	16	2	14
Pre-ribosome related proteins	8	5561	7	2165	7	3395	3	0	3
Ribosome synthesis regulators	3	7438	3	3459	3	3979	2	1	1
Polipeptide release factors	3	43316	3	18934	3	24382	3	0	3
Ribosome binding proteins	17	14839	17	6968	17	7871	5	2	3
Other ribosomal structural proteins	187	5612072	181	2688832	181	2923225	120	61	59
28S Ribosome strutural proteins	3	18083	3	6398	3	11685	2	0	2
30S Ribosome strutural proteins	26	199708	24	93074	24	106619	10	4	7
39S ribosome structural proteins	2	3453	2	1713	2	1740	2	2	0

40S ribosome structural proteins	35	429570	29	210856	29	218682	15	9	8
50S ribosome structural proteins	23	4806	21	1945	21	2834	8	2	7
60S ribosome structural proteins	50	1034965	41	502047	41	532873	23	13	12
Mitochondrial ribosomes structural proteins	43	105716	43	45395	43	60321	30	9	21
Mitochondrial 28S ribosome structural proteins	14	20331	14	9025	14	11306	9	2	7
Mitochondrial 37S ribosome structural proteins	1	4	1	4	0	0	0	0	0
Mitochondrial 39S ribosome structural proteins	11	22422	11	9322	11	13100	6	0	6
rRNA methylases	15	8705	14	3937	14	4767	6	3	3
rRNA maturation related proteins	1	9	1	6	1	3	0	0	0
Others tRNA related proteins	124	76998	118	33892	118	43028	51	9	44
Glycyl-tRNA synthetases	3	19595	3	8718	3	10877	1	0	1
Alanyl-tRNA synthetases	4	6429	4	2550	4	3879	3	0	3
Leucyl-tRNA synthetases	8	27987	8	11577	8	16410	3	1	2
Valyl-tRNA synthetases	6	1147	4	440	4	697	2	1	1
Isoleucyl-tRNA synthetases	4	189	4	108	4	81	1	1	0
Methionyl-tRNA synthetases	6	11683	6	5271	6	6412	3	0	3
Phenylalanyl-tRNA synthetases	5	4973	5	2035	5	2938	4	1	3
Tyrosyl-tRNA synthetases	4	403	4	241	4	162	1	1	0
Tryptophanyl-tRNA synthetases	2	217	2	85	2	132	1	0	1
Cysteinyl-tRNA synthetases	7	13610	7	6302	7	7308	3	1	2
Seryl-tRNA synthetases	10	11856	10	5913	10	5943	5	2	3
Histidyl-tRNA synthetases	4	11780	4	5282	4	6498	3	0	3
Threonyl-tRNA synthetases	3	2247	3	1134	3	1113	1	1	0
Prolyl-tRNA synthetases	6	4278	6	1891	6	2387	4	1	3
Asparaginyl-tRNA synthetases	2	4586	2	2003	2	2583	1	0	1
Glutaminyl-tRNA synthetsases	5	973	4	396	4	541	1	0	2
Lysyl-tRNA synthetases	6	6475	6	2733	6	3742	4	1	3
Arginyl-tRNA synthetase	6	4717	6	2092	6	2625	3	1	2
Glutamyl-tRNA synthetsases	5	4385	5	1851	5	2534	3	1	2
Aspartyl-tRNA synthetases	4	903	4	387	4	516	1	0	1
Other tRNA synthetase related	6	10446	6	4109	6	6337	4	0	4

SECRETED									
8.9 kDa family	3	782	3	471	3	311	2	1	1
Cysteine-rich proteins	6	1002	5	397	5	601	3	1	2
Cystatin	7	222908	7	117747	7	105161	5	3	2
Secreted lipases and esterases	36	136197	35	73485	35	62708	22	13	9
Secreted sugar enzymes	26	118266	26	57480	26	60786	20	11	9
Glycine-rich proteins	30	2175540	30	1265294	30	910246	25	16	9
Hebreains	4	10618	4	5757	4	4861	3	1	2
Secreted immune related products	103	155747	99	88244	99	67481	55	22	34
Ixodegrins	3	26500	3	17021	3	9479	3	3	0
Ixodidins	2	261	2	133	2	128	0	0	0
Kunitz containning domains	28	21505	21	7947	21	12866	9	5	11
Llipocalins	59	40382	58	11327	58	28998	32	8	25
Perilipins	18	28259	18	15258	18	13001	10	8	2
Other secreted enzymes	217	643607	203	317990	203	325568	109	53	57
Microplusins	13	80259	13	30410	13	49849	6	3	3
Mucins	40	773938	40	421608	40	352330	30	17	13
5'-Nucleotidases	9	5341	9	2667	9	2674	4	3	1
Odorant binding proteins	16	4001	16	1800	16	2201	4	1	3
Other secreted proteins	681	6023735	650	3481058	650	2541955	364	191	183
Peritrophins	12	106019	12	59053	12	46966	10	8	2
Secreted proteases	85	923522	83	397111	83	526388	57	26	33
Other serine protease inhibidors	43	42159	43	18977	43	23182	21	10	11
Proline-rich proteins	7	511	6	272	6	235	1	1	0
Serpins	51	22754	48	10843	48	11864	26	18	9
Secreted sufotransferase	17	33715	17	14459	17	19256	10	6	4
TIL	6	97901	6	49361	6	48540	4	2	2
Secreted unknown conserved	372	378314	321	179184	321	198640	107	59	54
Ixodidae related - SignalP	59	18177	53	8740	53	9362	14	8	7
SIGNAL TRANSDUCTION									
Tetraspanin receptor complex	12	11066	11	5680	11	5385	6	2	4

Acetylcholine receptor	19	1971	18	958	18	1009	2	1	1
Other signaling related proteins	4	4827	4	1981	4	2846	3	1	2
Serine/Threonine proteon kinase Act	11	8137	11	3886	11	4251	6	2	4
Apoptosis related	49	25338	47	11512	47	13816	12	5	7
GTPase activating proteins: ARF family	25	12219	22	5667	22	6543	9	6	3
Arrestin related	11	26548	10	14902	10	11645	4	4	0
Calcium dependent signaling proteins	124	139671	117	57659	117	81982	50	26	25
cAMP signaling related	139	32773	126	14922	126	17803	40	22	19
CDC 42 related kinases	9	7775	9	4253	9	3522	3	2	1
Intestinal cobalamin receptor	3	1492	3	832	3	660	2	2	0
Cytokine related receptors	7	1277	7	534	7	743	2	1	1
Dopamine related receptors	3	3089	3	1555	3	1534	2	1	1
GABA related receptors	4	28	2	12	2	8	0	0	0
Membrane Glicoprotein LIG-1	10	2684	10	1646	10	1038	5	3	2
Guanylate kinases	15	5125	14	2510	14	2612	5	3	2
G protein-couple receptors (GPCRs)	54	15327	48	5792	48	9503	9	6	4
Other G proteins modulators	7	981	4	298	4	672	3	0	3
GTPases	39	33132	39	14580	39	18552	21	6	15
GTP binding proteins	16	59440	15	28369	15	31058	9	3	7
Hippo signaling pathway related	91	39933	89	18231	89	21696	26	13	13
Insect hormone related	10	4692	10	2638	10	2054	6	5	1
Integrin-like receptor related	3	2718	3	1479	3	1239	2	2	0
JAK/STAT signaling pathway related	72	22637	67	12475	67	10145	16	7	9
Other signaling related kinases	202	83315	190	37552	190	45728	64	23	41
MAP kinase signaling pathway related	308	146233	295	72345	295	73842	84	39	46
mTOR signaling pathway related	11	4313	11	2095	11	2218	1	1	0
Notch signaling pathway related	11	2373	8	1535	8	824	2	1	1
Other GTPase activators	17	390	17	165	17	225	2	1	1
Phosphatases and related proteins	27	34775	27	22178	27	12597	6	4	2
Other signaling related receptors	105	72328	101	36371	101	35949	26	11	15
Other signaling related proteins	510	775397	477	380882	477	394370	166	70	101

Inositol phosphate related pathways proteins	68	33354	65	15595	65	17747	18	6	12
Rab GTPase activator	66	19262	61	8213	61	11034	14	5	9
Rac GTPase activators	6	1082	6	502	6	580	1	0	1
RaI GTPase activators	19	7318	18	3302	18	4009	9	3	6
Ran GTPase activators	9	8964	9	3554	9	5410	4	1	3
Rap1 signaling pathway related	87	57408	81	29636	81	27756	15	7	8
Rap GTPase activators	9	68	9	43	9	25	0	0	0
Ras signaling pathway related	261	212906	237	111707	237	101141	71	39	32
Ras GTPase activators	26	51341	19	35859	19	15461	3	2	1
Rho GTPase activators	62	46940	62	27583	62	19357	18	7	11
Signalosome related	4	4549	4	1918	4	2631	2	0	2
Two component system related	9	11174	9	5336	9	5838	3	1	2
TNF receptors and related proteins	6	3617	6	1584	6	2033	4	1	3
TonB receptors	0	0	0	0	0	0	0	0	0
Tyrosine phosphatases	47	15100	43	7105	43	7987	14	7	7
Wnt signaling pathway	118	23109	114	10926	114	12158	34	18	17
STORAGE									
Other storage related	9	112862	7	59460	7	53362	5	3	3
Vitellogenin	39	774885	33	396808	33	378052	23	6	18
Vitellogenin receptor	10	3318	9	1511	9	1801	5	3	2
TRANSPOSABLE ELEMENTS									
Gag related family	23	1068	23	660	23	408	3	3	0
Gypsy related family	2	361	2	78	2	283	1	0	1
Harbinger family	5	439	5	206	5	233	3	1	2
IS4 family	0	0	0	0	0	0	0	0	0
Other transposases related products	28	4348	25	1717	25	2621	8	0	8
Outcast family related	3	152	3	43	3	109	2	1	1
Pogo family related	8	2361	8	982	8	1379	5	2	3
Pol polyprotein family	11	484	11	237	11	247	0	0	0
Tickrelated families transposases	97	6271	89	3187	89	3048	21	13	10
Tigger family related	17	1287	16	740	16	541	5	2	3

TRANSCRIPTION FACTORS									
Other related transcripts	86	60847	83	27322	83	33517	28	12	16
5QNCA family	4	1508	4	754	4	754	0	0	0
NEURO-D4	5	3260	5	1359	5	1901	2	0	2
AT-hook	0	0	0	0	0	0	0	0	0
bHLH	5	359	5	119	5	240	1	0	1
Bromodomain	10	12120	9	5677	9	6442	3	2	1
bZIP	15	108959	15	32787	15	76172	8	4	4
C2H2	65	85671	63	49087	63	36576	19	8	11
C3H1	2	34	2	17	2	17	0	0	0
CAP	3	107	3	60	3	47	1	1	0
CB2B	2	9	1	4	1	4	0	0	0
CBF	2	308	1	101	1	204	1	0	1
CCAAT	2	2007	2	903	2	1104	1	0	1
DAFT1	2	6	2	5	2	1	0	0	0
E2F	14	1748	13	605	13	1140	2	1	1
Forkhead related	20	5806	20	2617	20	3189	9	4	5
Forkhead-HNF3	9	13329	8	6608	8	6720	4	3	1
Forkhead-LRR	19	4228	16	2049	16	2169	7	4	3
GABP	12	1030	12	578	12	452	6	6	0
General transcription factors	13	10644	13	4341	13	6303	7	1	6
GT-2	2	528	2	260	2	268	0	0	0
Homeobox	28	3116	26	1293	26	1818	5	2	3
Homeobox-LRR	1	1	1	1	0	0	0	0	0
Homeobox-SIP1	13	757	12	207	12	545	3	1	2
HCFC	9	1370	9	597	9	773	1	1	0
HLH	5	1458	5	604	5	854	4	2	2
HMG	18	7658	16	4136	16	3517	6	4	2
HOX	18	4121	15	2099	15	1996	4	2	3
IWS1	1	657	1	226	1	431	1	0	1
Kayak	2	49	2	24	2	25	0	0	0

LIM	3	436	3	222	3	214	2	1	1
PHOX2-ARIX	2	382	2	78	2	304	1	0	1
MEIS1	3	1400	3	619	3	781	1	0	1
LRR	21	2024	21	855	21	1169	7	3	4
MADS	3	107	3	45	3	62	0	0	0
MRTF	4	179	4	98	4	81	0	0	0
Myb	11	6007	11	2473	11	3534	3	0	3
Myc	5	680	5	312	5	368	2	1	1
NAC	3	3619	3	863	3	2756	2	0	2
NF1	4	7754	4	3485	4	4269	1	0	1
NFAT	4	8507	4	3928	4	4579	3	2	1
NFKappaB	4	2452	4	912	4	1540	2	0	2
NFX	3	5753	3	3450	3	2303	3	1	2
Ovo	3	32	3	17	3	15	0	0	0
p65	1	27	1	14	1	13	0	0	0
PBX	4	1965	4	872	4	1093	1	0	1
PCBP	2	2478	2	1117	2	1361	0	0	0
PHD	7	2907	7	1048	7	1859	2	1	1
POU	4	29	3	19	3	9	0	0	0
SOX	3	765	3	304	3	461	1	0	1
SP3	2	678	2	517	2	161	1	1	0
STAT	6	6099	6	2868	6	3231	2	1	1
TBP	7	8486	7	3918	7	4568	5	1	4
TBX	5	784	3	396	3	382	1	1	0
TCF	2	20	2	11	2	9	0	0	0
TEAD	2	27	2	6	2	21	1	0	1
Wimged-Helix	1	549	1	205	1	344	1	0	1
XBP-1	2	88270	2	60670	2	27600	2	2	0
Zinc finger	236	99434	223	47803	223	51579	64	23	42
TRANSCRIPTION MACHNERY	**1175**	**2848746**	**1123**	**1429080**	**1123**	**1419443**	**492**	**171**	**327**
TRANSPORTERS									

Other transporters	56	33718	54	16050	54	17663	19	12	7
Amino acid and peptides transporters	144	94268	132	42389	132	51836	46	23	24
Sodium symporters	177	53955	170	26143	170	27775	52	35	18
ABC superfamily	105	57858	93	27784	93	29993	26	11	17
Ammonia transporters	0	0	0	0	0	0	0	0	0
Aquaporins	11	13423	10	5863	10	7559	2	0	2
ADP/ATP translocases	8	17287	5	7414	5	9834	3	1	3
Biopolymer transporters	0	0	0	0	0	0	0	0	0
CARGO family transporters	6	9865	6	4503	6	5362	1	0	1
CLC superfamily	26	24848	25	13392	25	11454	13	11	2
Ferritins	2	14102	1	8044	1	6056	1	1	0
Ion transporters	258	47181	238	24862	238	22137	61	37	29
Lipid transporters	52	30089	49	15561	49	14482	22	12	11
Major facilitator superfamily	23	3660	20	1745	20	1912	2	1	1
Metal transporters	53	44533	51	20864	51	23663	20	6	14
Metabolites transporters	165	65754	152	34656	152	31028	79	53	29
Other mitocondrial transporters	38	80556	38	37782	38	42774	14	4	10
Mechanossensitive channels	1	79	1	14	1	65	1	0	1
Multidrug transporters	9	3121	9	1346	9	1775	3	0	3
Nitrogen permeases	2	84	2	54	2	30	1	1	0
Permeases	1	103	1	44	1	59	0	0	0
Phosphate transporters	4	53	4	27	4	26	0	0	0
Porins	3	1149	2	404	2	744	1	0	1
RND family	1	2	1	2	0	0	0	0	0
Solute carrier family	19	1605	15	868	15	628	3	2	3
Sec family transporters	18	18811	18	8176	18	10635	6	2	4
Sugar/nucleotide transporters	91	18881	88	8564	88	10302	39	17	22
Sulfate transporters	10	7841	9	3222	9	4618	5	0	5
TRAM superfamily	2	186	1	111	1	73	1	1	0
Transportins	3	29	3	15	3	14	0	0	0
Putative transporters	10	2087	10	816	10	1271	5	2	3

V ATPase subunits	12	19592	12	9627	12	9965	6	3	3
Voltage dependent Ca2+ channels	41	5145	37	3010	37	2125	12	9	3
Vesicle transporters	26	29049	25	13916	25	15129	12	6	6
Vitamin transporters	15	3299	15	1702	15	1597	4	4	0
UNKNOWN CONSERVED	**1303**	**507109**	**1178**	**273812**	**1178**	**231898**	**279**	**125**	**165**
UNKNOWN	**872**	**353014**	**775**	**194086**	**775**	**158399**	**175**	**84**	**99**
VIRAL	**14**	**10376**	**13**	**4872**	**13**	**5502**	**7**	**3**	**4**
TOTAL	**21230**	**45259807**	**19943**	**23079436**	**19936**	**22170987**	**7690**	**3486**	**4368**