

GABRIEL GUIMARÃES GOMES

**INTERAÇÃO FUNCIONAL, EXPRESSÃO DE GENES E FILOGENIA ENTRE
PRÓTEÍNAS DE MEMBRANA DE *Rickettsia* spp. E CARRAPATOS**

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*.

Orientador: Cláudio Lísias Mafra de Siqueira

Coorientadore: Tiago Antônio de Oliveira Mendes

VIÇOSA MINAS GERAIS

2019

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

G633i
2019
Gomes, Gabriel Guimarães, 1985-
Interação funcional, expressão de genes e filogenia entre
proteínas de membrana de *Rickettsia* spp. e carrapatos: / Gabriel
Guimarães Gomes. – Viçosa, MG, 2019.
389f.: il. (algumas color.).

Inclui anexos.

Orientador: Cláudio Lísias Mafra de Siqueira.
Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Bioquímica e Biologia Molecular, 2019.
Referências bibliográficas: f.318-360.

1. Carrapatos. 2. *Rickettsia*. 3. Proteínas. 4. Filogenia.
I. Siqueira, Cláudio Lísias Mafra de. II. Universidade Federal de
Viçosa. Departamento de Bioquímica e Biologia Molecular.
Programa de Pós-Graduação em Bioquímica Aplicada.
III. Título.

CDD 22 ed. 579.327

Bibliotecário(a) responsável: Renata de Fátima Alves CRB6/2578

GABRIEL GUIMARÃES GOMES

**INTERAÇÃO FUNCIONAL, EXPRESSÃO DE GENES E FILOGENIA ENTRE
PROTEÍNAS DE MEMBRANA DE *Rickettsia* spp. E CARRAPATOS**

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: 02 de julho de 2019.

Assentimento:

Gabriel Guimarães Gomes
Autor

Cláudio Lísias Mafra de Siqueira
Orientador

DEDICATÓRIA

Dedico este trabalho de investigação
a minha amada Família.

*“Anti-social behaviour is a trait of intelligence
in world full of conformists”*

Nikola Tesla

BIOGRAFICA DO AUTOR

Gabriel Guimarães Gomes, nascido em 14 de agosto de 1985, na cidade de Ponte Nova, estado de Minas Gerais, Brasil, filho de Adão Daniel Gomes & Ana Regina Toledo Guimarães Gomes. Estudou na Escola Municipal Joaquim Machado Guimarães do jardim a 4ª série, no colégio Salesiano Don Helvécio da 5ª ao 1º ano do ensino médio e no Colégio Equipe de Ponte Nova, aonde terminou o 2º e 3º ano do ensino médio, agraciado com bolsas de estudos, boa educação e ajuda dos pais. No ano de 2006, ingressou no Bacharelado em Ciências Biológicas pela Universidade Federal de Ouro Preto (UFOP), obtendo o título em 4 de abril de 2011. Durante a graduação, exerceu diversas atividades acadêmicas sendo bolsista de iniciação científica pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), no laboratório de Epidemiologia Molecular do Núcleo de Pesquisa em Ciências Biologia (LEM/NUPEB) em projeto intitulado: Mamíferos comuns de áreas urbanas no ciclo epidemiológico da febre maculosa, sob orientação do Prof. Dr. Márcio Antonio Moreira Galvão, docente da Escola de Farmácia, Departamento de Medicina. Em agosto de 2011, iniciou suas atividades no Mestrado pelo Programa de Pós-graduação em Bioquímica Agrícola, do Departamento de Bioquímica e Biologia Molecular (DBB), Universidade Federal de Viçosa (UFV). Exercendo pesquisas no Laboratório de Parasitologia e Epidemiologia Molecular (LAPEM), com o projeto intitulado: Modelo epidemiológico para avaliação do risco biológico potencializado por ações antrópicas na ocorrência de infecções veiculadas por carrapato. Este projeto teve como apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), sob orientação do Prof. Dr. Cláudio Lísias Mafra de Siqueira. Em outubro de 2014, iniciou suas atividades no Doutorado pelo Programa de Pós-graduação em Bioquímica Aplicada, podendo pelo DBB / UFV continuar exercendo atividades de pesquisas no LAPEM, com o projeto intitulado: Avaliação topológica, ranqueamento e co-evolução entre a interação proteína-proteína no ovário de *Amblyomma sculptum* (BERLESE, 1888) e *Rickettsia amblyommatis*. Este projeto teve como apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), sob orientação do Prof. Dr. Cláudio Lísias Mafra de Siqueira e coorientação do Prof. Dr. Tiago Antônio de Oliveira Mendes.

AGRADECIMENTOS

Primeiramente agradeço a Deus com fé e a minha família. Numa ordem cronológica, agradeço aos meus familiares e amigos de infância de Ponte Nova MG. Aos irmãos pasargadeanos de Ouro Preto MG e aos republicanos de Viçosa MG. Esta história não se repete! Às universidades federais UFOP e UFV, meus colegas de curso e pesquisa, amigos de departamentos e aos meus professores incentivadores. Ao professor Márcio Galvão, pela primeira oportunidade de experiência no mundo acadêmico científico. Ao programa de Pós-Graduação em Bioquímica Aplicada do DBB / UFV. À CAPES pelo amparo financeiro durante a permanência no município de Viçosa. Ao LAPEM / UFV pelo suporte técnico e científico. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – por meio do projeto, número 188595 e registro 20806267360. Agradeço os colegas de trabalho, em especial, à Raquel Ligeiro, doutoranda do programa, pelo apoio com as pesquisas. Ao querido professor Cláudio Mafra. “Não importa a fase, um mestre é para a vida toda. Eis um discípulo”. **MUITO OBRIGADO!** Ao professor Tiago Mendes, foi um privilégio assistir suas aulas - muito importantes para o meu trabalho. Grato pelas discussões e esclarecimentos quanto às teorias e técnicas que pude aprender!

RESUMO

GOMES, Gabriel Guimarães, M.Sc, Universidade Federal de Viçosa, julho de 2019. **Interação funcional, expressão de genes e filogenia entre proteínas de membrana de *Rickettsia* spp. e carrapatos.** Orientador: Cláudio Lísias Mafra de Siqueira. Coorientador: Tiago Antônio de Oliveira Mendes.

A interface da relação parasito-hospedeiro entre carrapatos e espécies de *Rickettsia* endossimbióticas é mediada por interações e mecanismos moleculares, ocorrendo através de genes e proteínas potencialmente codificadas nos genomas de ambos os organismos. Tratando-se de uma relação em nível molecular, buscamos prever as interações, ativação da expressão de genes e coevolução entre os repertórios de proteínas destes organismos. Para avaliar a hipótese de interação, utilizamos sequências de proteínas de membrana devido a maior probabilidade de participar do processo de infecção. Estas sequências foram submetidas a abordagem de reconstrução do interatoma adotando como modelo, as proteínas provenientes dos genomas do *Ixodes scapularis* em resposta a infecção por *Rickettsia amblyommatis*. Com o auxílio de bancos de dados de interação proteína-proteína, utilizamos de métodos de alinhamentos de sequências e predição de domínios proteicos. Contra a base de dados PEIMAP, adotou-se do BLASTP. Para a base de dados SCOP, utilizada pelo PSIMAP, adotou-se o PSI-BLAST. Para a base de dados iPfam, utilizada pelo PFAM, adotou-se modelos ocultos de Markov. A partir disto, realizamos um cálculo de confiança assumindo independência entre os quatro métodos. Estes resultados foram concatenados e filtrados nos resultados do interatoma para a construção do interatoma de proteínas de membrana. As proteínas ortólogas e conservadas na rede do interatoma de proteínas de membrana foram avaliadas com o uso do aplicativo OrthoMcl adotando o BLASTP contra as proteínas do genoma de *R. amblyommatis* contra proteínas de outras espécies de *Rickettsia*. As proteínas ortólogas de membrana foram avaliadas quanto as características evolutivas através do método do Maximum Likelihood Tree para as inferências e construção de cladogramas e árvores de hipóteses filogenéticas. Em seguida, os resultados das redes do interatoma de proteínas de membrana e ortólogas serviram de base para realizamos a anotação de proteínas putativas provenientes de sequências de transcritos adquiridas com a montagem *de novo* do transcriptoma de glândulas salivares, intestinos e ovários de carrapatos. Os resultados das redes apresentaram estruturas fortemente modulares com altos números de componentes conexos. Esta análise de anotação das proteínas de interface da membrana de *R. amblyommatis* permitiu identificar famílias de proteínas sabidamente envolvidas na modulação da fisiologia da célula do carrapato *I. scapularis*. Os resultados obtidos para as análises

filogenéticas destas proteínas sugerem haver divergências entre as espécies de *Rickettsia* que infectam carrapatos em relação a outras espécies encontradas em insetos, ou em outros hospedeiros correlacionados. A ocorrência da interação, os aspectos evolutivos e os níveis de expressão diferenciada de proteínas sugerem mecanismos moleculares conservados com maior capacidade de ocasionar mudanças nos padrões de expressão para as espécies de *Rickettsia* específicas ao carrapato em determinados órgãos e ou grupos de hospedeiros hematófagos com os quais, elas puderam coevoluir.

Palavras-chave: Carrapatos. *Rickettsia*. Interatoma de Proteínas. Expressão de Genes. Filogenia de Proteínas de Membrana.

ABSTRACT

GOMES, Gabriel Guimarães, M.Sc, Universidade Federal de Viçosa, July, 2019. **Functional interaction, genes expression and phylogeny between *Rickettsia* spp. membrane proteins and ticks.** Adviser: Cláudio Lísias Mafra de Siqueira. Co-adviser: Tiago Antônio de Oliveira Mendes

In the coevolutionary process, the interface of the host-parasite interaction between ticks and endosymbiotic *Rickettsia* species is mediated by molecular mechanism interactions involving both organisms, occurring through genes and potentially codified proteins. Dealing with this relationship on a molecular level, we aim to assess the interactions and co-evolution between the protein repertoires of these organisms. In order to assess the predicted interaction between the membrane protein repertoires related to the infection process, we propose to adopt the interatome approach to determine the pattern, as the proteins derived from *Ixodes scapularis*'s genomes in response to infection by *Rickettsia amblyommatis*. With the aid of databases of the protein-protein interaction, we used sequences alignment and domain prediction. Against the PEIMAP database, the BLASTP program was utilized. For the SCOP database, used by PSIMAP, the PSI-BLAST was adopted. For the iPfam database, used by PFAM, the hidden Markov was used. We performed a confidence calculation assuming independence among the four methods and assigned a new rate to each pair of interactions. These results were concatenated and the interatome results were filtered for the construction of the membrane protein interatome. The orthologous and conserved proteins in the interatom network of membrane proteins were evaluated using the OrthoMcl application adopting the BLASTP against the proteins of the genome of *R. amblyommatis* versus proteins of other species of *Rickettsia*. The orthologous membrane proteins were evaluated for the evolutionary traits by the Maximum Likelihood Tree method for inferences and construction of cladograms and phylogenetic trees by adding several groups of species of *Rickettsia*. to the initial model. Afterwards, the interatomic networks of membrane proteins and orthologs results served as the basis for the annotation of putative proteins. For achieve this goal we performed a new assemblage for the *Amblyomma aureolatum* and *Amblyomma sculptum* species in response to *Rickettsia* spp. infection to examine the level of expression and abundance of data from the transcriptome of salivary glands, intestines and ovaries. The results of the networks presented strongly modular structures with high numbers of related components. This annotation analysis of the membrane interface proteins of *R. amblyommatis* allowed to identify families of proteins known to be involved in the modulation of the cell physiology of the *I. scapularis* tick. The

results obtained for the phylogenetic analysis of these proteins suggest that there are divergences between species of *Rickettsia*, which infect ticks with regard to other species found in insects, or in other correlated hosts. The occurrence of interaction, evolutionary aspects and levels of differentiated expression of proteins revealed conserved molecular mechanisms with greater capacity to cause changes in expression patterns for *Rickettsia* spp. specific to the tick or particular groups of hosts which they could co-evolve with.

Keywords: Ticks. *Rickettsia*. Protein Interaction. Gene Expression. Membrane Protein Phylogeny.

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

2DG	Glicólise 2-desoxiglicose,
ABC	<i>ATP-binding cassette transporter</i>
ADR1	<i>Rickettsia</i> adhesin
Ago	Proteína argonauta
Ank	Motivo de repetição de anquirina
Arp2/3	<i>Actin-Related Proteins</i>
ATs	Sistema de secreção autotransporte
BIND	<i>Biomolecular Interaction Network Data-base IntAct</i>
BioGrid	<i>The Biological General Repository for Interaction Datasets</i>
BLAST	<i>The Basic Local Alignment Search Tool</i>
BLOSUM	<i>Blocks of Amino Acid Substitution Matrix</i>
COG	Consortio de ontologia gênica
COPI	Complexo de revestimento de vesícula Ap1
COPII	Complexo de revestimento de vesícula Ap2
COPIII	Complexo de revestimento de vesícula Ap3
CYT	Citoplasmático
Dcr	Proteína dicer
DIP	<i>Data-base of Interacting Proteins</i>
DlgA	Proteína do disco largo
DTN	Rede de ditirosina
DUOX	Dual oxidase
ECR	Centro de reciclagem endossomal
EE	Endossomo precoce
FDR	<i>False Discovery Rate</i>
GO	<i>Gene Ontology</i>
GPCR	Receptor acoplado à proteína G
HDAC1	Histona desacetilase 1
HGT	Transferência horizontal de genes
Hk	Histidina quinase
HMEs	Enzimas modificadoras de histonas
HMM	<i>Hidden Markov Model</i>
HPRD	<i>Human Protein Reference Database</i>
HSP	<i>High-scoring Segment Pairs</i>
Hsp	Proteína de choque térmico
IMD	Via da imunodeficiência
ISE6	Células embrionárias de carrapato da espécie <i>I. scapularis</i>
JAK-STAT	Janus Kinase transdutor de sinal e ativador da transcrição
LE	Endossomo tardio
LPS	Lipopolissacárido
LRR	Domínios de repetição rico em leucina
MCL	<i>Markov Clustering algorithm</i>
MINT	<i>Molecular INTeraction database</i>
MVB	Corpo multivesicular
NCBI	<i>National Center for Biotechnology Information</i>
NOX	Espécies oxidativas de nitrogênio
Pb	Pares de bases
PDB	<i>Protein Data Bank</i>
PI3K	Via fosfoinositido-3-quinase

PPI	<i>Protein-Protein Interaction</i>
PSIMAP	<i>Protein Structural Interactome map</i>
RAGE	<i>Rickettsiales Amplified Genetic Element</i>
RARP-1	<i>Retinoic Acid-Responsive Proline-Rich Protein 1</i>
REIS	<i>Rickettsia endosymbiont Ixodes scapularis</i>
RER	Retículo endoplasmático rugoso
RISC	Complexo ribonucleico silenciador
ROS	Espécies reativas de oxigênio
RRE	Endossomos de reciclagem rápida
Sec-TolC	Sistemas Sec translocon
SER	Endossomos de reciclagem lenta
SGBD	Sistema de gerenciamento de banco de dados
SNG	Sequenciamento de nova geração
SQL	<i>Structured Query Language</i>
SRA	<i>Sequence Read Archive</i>
STRING	<i>Search Tool for the Retrieval of Interacting Genes/Proteins</i>
T1SS	Sistema de secreção do tipo I
T3SS	Sistema de secreção do tipo III
T4SS	Sistema de secreção do tipo IV
T5SS	Sistema de secreção do tipo V
Tat	Sistema de translocação de duas argininas
TCA	Ciclo de ácido tricarboxílico
TLR	Receptores Toll-like
TMH	Hélice transmembrana n-terminal
TPM	<i>Transcripts Per Kilobase Million</i>
Tpr	Domínio de repetição de tetratricopeptídeo
TrxR	Proteína tioredoxina redutase
WASP	Proteína da síndrome Wiskott-Aldrich

LISTA DE FIGURAS

- Figura 1 Relato de microrganismos próximos a *Rickettsia* spp. (ciclo vermelho). coletado na hemolinfa (B) de uma ninfa de carrapato Ixodídeo da espécie *Cornupalpatum burmanicum*, já extinto, preservado em âmbar (A) encontrado na Península Dominicana. 37
- Figura 2 Distribuição geográfica atual de bactérias do gênero *Rickettsia* transmitidas por espécies de carrapatos ao redor do mundo. 40
- Figura 3 Distribuição geográfica, filogeográfica e especiação do carrapato do complexo *A. cajennense* na América e nos biomas Brasileiros. Destaque (a): Formação e distribuição de quatro clados entre populações de carrapatos do complexo *A. cajennense*. Destaque (b): Distribuição geográfica do complexo *A. cajennense* após redescritção na América. Destaque (c): Distribuição geográfica do complexo *A. cajennense* no Brasil correlacionada aos fatores ambientais florestais e do clima. 44
- Figura 4 Mapa geográfico dos EUA demonstrando a distribuição geográfica para as espécies de carrapatos, em especial, *I. scapularis* e *Amblyomma americanum*. Destaque para a área com forte sobreposição de incidência para ambas, as espécies. Fonte: CDC. (2015) 46
- Figura 5 Secretomas de bactérias encontrados em *Rickettsia* com destaque para os sistemas de secreção Sec-dependente contendo o sistema de secreção do tipo V (T5SS) e o Sec-Tolc, além dos sistemas de secreção Sec-independentes contendo o sistema de duas argininas (Tat), o sistema de secreção do tipo I (T1SS) e do sistema de secreção do tipo IV (T4SS) sendo este último, o menos conhecido para o gênero. 49
- Figura 6 Fluxograma para as análises de similaridade dos dados de sequência de proteínas dos genomas provenientes da bactéria da espécie *R. amblyommatis* ♦ e do carrapato da espécie *I. scapularis* ◇, fazendo o uso de multialinhamentos de sequências e de dados de domínios de interação de proteínas em banco de dados do interatoma. Destaque para os alinhamentos e parâmetros avaliados em cada banco dando montagem do interatoma parcial utilizando de cada banco. 58
- Figura 7 Fluxograma contendo os passos consequentes após a análise de similaridade para as montagens dos interatomas parciais com cada par de interação em cada banco. Destaque para o cálculo do *score* final para cada interatoma parcial. 69
- Figura 8 Fluxograma para o tratamento dos interatomas parciais e obtenção do interatoma interespecífico entre as proteínas de *R. amblyommatis* e *I. scapularis*. Destaque para o cálculo do *score* combinado para cada par de interação combinando os *scores* finais por interatoma parcial em cada banco. 87
- Figura 9 Fluxograma das análises de predição para as proteínas de membrana provenientes do genoma da espécie *R. amblyommatis*, sendo estas mesmas preditas como secretadas, excretadas e / ou de superfície. 93
- Figura 10 Fluxograma representativo para o uso do aplicativo OrthoMCL 2.0 e análise de sequências de proteínas ortógas entre as proteínas provenientes do genoma de *R. amblyommatis* e *Rickettsia* spp., que infectam carrapatos e insetos com destaque para os passos e os parâmetros adotados. 104

- Figura 11 Caixa de comando do pacote Bingo 3.0.3 adotado com o uso do aplicativo Cytoscape 3.7 para o teste hipergeométrico e da taxa de descoberta do erro FDR, e dos processos biológicos de ontologia gênica (GO). 120
- Figura 12 Fluxograma para as análises de qualidade dos dados brutos dos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp., fazendo o uso do aplicativo FastQc 0.11.5. Destaque para os parâmetros avaliados dos sequenciamentos. 122
- Figura 13 Fluxograma contendo os passos conseguintes para análise de qualidade dos dados brutos e trimagem de transcriptoma utilizando do software Trimmomatic 0.36 e em seguida novamente o FastQC 0.11.5. Destaque para os parâmetros avaliados da trimagem. 125
- Figura 14 Fluxograma para a montagem “*de novo*” dos dados trimados e de boa qualidade dos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp., fazendo o uso do aplicativo Trinity 2.5.1 Destaque para os parâmetros avaliados do sequenciamentos. 128
- Figura 15 Fluxograma para o cálculo dos níveis de abundância da transcrição para as sequências de RNA montadas dos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp., fazendo o uso do aplicativo Kallisto 0.43.0. Destaque para os parâmetros avaliados dos sequenciamentos. 130
- Figura 16 Redes consenso de interações potenciais de alta confiabilidade entre proteínas do interatoma interespecífico de proteínas geral (Dados suplementares S1e) construído para o modelo das espécies *R. amblyommatis* vs *I. scapularis*, com o uso do aplicativo Cytoscape 3.7 adotando do modulo ModLand. Em (A), todas as interações possíveis. Em (B), destaque para as proteínas de *R. amblyommatis* (vermelho), interagindo com as proteínas de *I. scapularis* (azul). Em (C), destaque para os grupos de proteínas mais bem conectados entre si (módulos), totalizando 19 grupos significativos e distintos (conforme as cores). 168
- Figura 17 Vias de sinalização envolvendo os módulos de secreção, transporte, metilação e fosforilação obtidos por meio da análise de visualização das redes de interação do interatoma de proteínas de membrana de *R. amblyommatis* vs *I. scapularis* fazendo o uso do aplicativo Cytoscape 3.3.4, com a adoção dos pacotes ModuLand 2.0 para clusterizar a rede, e do módulo BINGO 3.0.3 para realizar as análises do enriquecimento hipergeométrico, utilizando o Teste de correção de FDR. Em círculos de cor laranja e amarelo, tem-se: Destaque (a) para os processos celulares envolvidos com a secreção de proteínas, localização e transporte com p-valor = 2,9350E-5. Destaque (b) para os processos celulares envolvendo o transporte mediado por vesículas com p-valor= 7,8875E-10. Destaque (c) para os processos celulares envolvidos com a metilação do DNA com p-valor = 1.3812E-3. Destaque (d) para os processos celulares envolvidos com a fosforilação de proteínas, com p-valor = 2.5491E-16 174
- Figura 18 Fluxograma representativo da presença de espécies de bactérias do gênero *Rickettsia*, em gêneros e espécies específicas de carrapatos Ixodídeos, incluindo indivíduos patogênicos e não patogênicos construídos após revisão literária e com o uso do aplicativo Yed para a plotagem dos dados. 198
- Figura 19 Árvore filogenética para a família das sequências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID

- KJV62536.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 203
- Figura 20 Árvore filogenética para a família das sequências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62591.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia com o uso do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 205
- Figura 21 Árvore filogenética para a família das sequências de proteínas do sistema de secreção do tipo III T3SS com destaque para a proteína ID KJV62562.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 206
- Figura 22 Árvore filogenética para a família das sequências de proteínas contendo (Ftsk) com destaque para a proteína ID KJV62594.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*... 208
- Figura 23 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo II/IV (T2T4SS) com destaque para a proteína ID KJV62252.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 211
- Figura 24 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62535.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 214
- Figura 25 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62505.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 217
- Figura 26 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62253.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 220
- Figura 27 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID

- KJV62597.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 223
- Figura 28 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62563.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 226
- Figura 29 Árvore filogenética para a família das sequências de proteínas contendo domínios repetitivos de anquirina (Ank) com destaque para a proteína ID KJV61977.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 228
- Figura 30 Árvore filogenética para a família das sequências de proteínas ribossomais metiltransferase (L11) com destaque para a proteína ID KJV62584.1 e KJV62665.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*..... 230
- Figura 31 Modelo de vias hospedeiras ativadas durante a invasão de *Rickettsia* e *Orientia* em células humanas. Na invasão da espécie *Rickettsia parkeri* em células hospedeiras de humanos a proteína Rho GTPase coopera com as proteínas Cdc42 e Rac1 na polimerização de actina, desempenhando papéis paralelos. .. 243
- Figura 32 Esquema representativo dos compartimentos e vias de secreção. Destaque (a), lisossomal, vacuolar e endossomos, indicando o sistema de transportes de proteínas de revestimento de vesículas do tipo Ap1 (vermelho), Ap2 (azul) e Clatrin (amarelo). Destaque (b), subcomplexos Sec23p.Sec24p e Sec13p.Sec31p. Destaque (c), o complexo de revestimento de vesícula Ap2 necessita de uma pequena GTPase Sar1p..... 246
- Figura 33 Modelo que resume a interação do patógeno *A. marginale* (AmV) (vacúolo) em células de carrapato da espécie *D. andersoni* sob condições de pH neutro com GTPases Rab nas vias endocitária e secretora. A inserção na parte inferior direita mostra a maturação do patógeno em um fagolisossomo multiorganelo com LAMP1, Rab7 e catepsina L. cG, cis-Golgi; mG, medial-Golgi; tG, trans-Golgi; proteína V-ATPase GP 58kD, proteína GolK 58K; SG, grânulo de secreção. 248
- Figura 34 Mecanismos moleculares potencialmente implicados na modulação do metabolismo de carrapatos infectados por *A. phagocytophilum*. Destaque para a reprogramação transcricional e na modulação do proteoma. 256
- Figura 35 A ubiquitina é ativada pela enzima ativadora da ubiquitina (E1), transferida para a enzima conjugadora da ubiquitina (E2) e finalmente transferida para os substratos pelas ligases da ubiquitina (E3). As proteínas da ubiquitina são reconhecidas e degradadas pelo proteassoma 26s. O complexo E3 ligase consiste na variação de F-box de Skp1, Rbx1, e uma proteína Cullin. 263

- Figura 36 Modelo da via do trânsito de heme na célula digestora de carrapato da espécie *Rhipicephalus (B.) microplus*. Este modelo proposto integra dados sobre o papel de transportadores ABC e resultados de relatórios anteriores descrevendo a captação de hemoglobina, seguido pela exportação de heme da vesícula digestiva para a citosol e formação de heme agregados no hemosomo. 266
- Figura 37 Modelo de produção dos miRNAs pela RNA Pol II e sua passagem pela membrana. (A) Destaque para a exportação nuclear mediada pela Exportin 5 (XPO5) e Ran-GTP. (B) A segunda clivagem é realizada no citoplasma pela enzima Dicer RNase III como parte de um complexo proteico que produz o miRNA 269
- Figura 38 Modelo do perfil metabólico de células ISE6 do carrapato *I. scapularis* infectadas por *A. phagocytophilum* (bactérias sem LPS) em hemócitos versus o modelo dos macrófagos M1 de mamíferos em resposta a outras bactérias intracelulares com LPS. Altos níveis de hexoquinase (HXK, verde) e fosfofrutoquinase (PFK, verde) estão associados com alta glicólise aeróbica e aumento da produção de lactato em macrófagos M1 de mamíferos. Os níveis e atividade de HXK do carrapato diminuíram (vermelho), e os níveis de PFK do carrapato não se alteraram (negros) em resposta à infecção que foi associada com níveis mais baixos de lactato em células ISE6 infectadas. 272
- Figura 39 Modelo de uracil ou peptidoglicano (PG) liberado pelos patógenos para induzir a via de sinalização Hh através de um mecanismo desconhecido, para manter altos nível de Cad99C na membrana apical. Moléculas PG de bactérias comensais também pode desempenhar um papel na manutenção do nível basal apropriado de Cad99C sob uma condição convencional antes da infecção do patógeno. Quando o uracil é reconhecido por um receptor acoplado à proteína G (GPCR), os endossomos de sinalização dependentes de Cad99C/PLCb/PKC facilitam a amplificação do sinal, induzindo a mobilização de Ca²⁺ dependente de PLCb para a geração de ROS dependente de DUOX..... 276
- Figura 40 Modelo de diferentes mecanismos para a inibição de apoptose de células hospedeiras por patógenos. N, núcleo; M, mitocôndria; ER, retículo endoplasmático; Cyt c, Citocromo c; ASK1, quinase 1 reguladora de sinal de apoptose; MKK, proteína quinase ativada por mitógeno; HDAC1, histona desacetilase 1; Anka, Ankyrin A; PEPCK, Fosfoenolpiruvato Carboxiquinase; FOXO, caixa de garfo O; Hid, Involução da cabeça defeituosa; JNK, junho amino-terminal quinases; Casp caspases..... 280
- Figura 41 Fluxograma de proteínas ortólogas de membrana para as espécies de *Rickettsia* que infectam somente carrapatos. Destaque para as diferenças de patogenicidade e grupos de classificação taxonômica entre as espécies e a presença das proteínas, incluindo espécime do gênero *Orientia*, sendo este um Grupo Ancestral irmão do gênero *Rickettsia*..... 283
- Figura 42 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo motivos repetitivos de tetratricopeptídeos (Tpr) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62536.1 pertencente a espécie *R. amblyommatidis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura..... 289

- Figura 43 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo motivos repetitivos de tetratricopeptídeos (Tpr) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62591.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura..... 290
- Figura 44 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas T3SS, podendo conter motivos repetitivos de tetratricopeptídeos (Tpr) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62562.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura..... 291
- Figura 45 Fluxograma representativo para as espécies de *Rickettsia* do grupo ancestral Belli, do grupo Canadense, do grupo das Febres Maculosas transmitidas por carrapatos com pouco potencial patogênico e, ou, espécimes endossimbionticas que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtidos com o uso do aplicativo *Yed Grafics*. Destaque para as proteínas Tpr ID KJV62536 no círculo (azul) e ID KJV62591 no círculo (amarelo) e para a proteína T3SS ID KJV62562 no círculo (rosa) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via de sinalização conservada para transporte e secreção via GTPase..... 292
- Figura 46 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas recombinantes podendo ser responsáveis pelos movimentos de translocação e conjugação do DNA FtskSpoIIIE, com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62594.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura. 293
- Figura 47 Fluxograma representativo para as espécies de *Rickettsia* do grupo ancestral Belli, do grupo Canadense, do grupo das Febres Maculosas transmitidas por carrapatos com pouco potencial patogênico e, ou, espécimes endossimbionticas que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtidos com o uso do aplicativo *Yed Grafics*. Destaque para a proteína Ftsk KJV62594 no círculo (laranja) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui esta proteína essencial para uma via conservada de proteínas recombinantes conjugadora da família SpoIIIE. 294
- Figura 48 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas do sistema de secreção do tipo II/IV, com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62252.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura. 295
- Figura 49 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo proteínas do sistema de secreção

do tipo IV (T4SS) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para as sequências de proteínas ID KJV62253.1, KJV62505.1, KJV62503.1, KJV62535.1 e KJV62563.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diversas espécies de carrapatos conforme relatados na literatura..... 297

Figura 50 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas do sistema de secreção do tipo IV, com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62597.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura 299

Figura 51 Fluxograma representativo para as espécies de *Rickettsia* do grupo ancestral Belli, do grupo Canadense, do grupo das Febres Maculosas transmitidas por carrapatos com pouco potencial patogênico e, ou, espécimes endossimbionticas que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtidos com o uso do aplicativo *Yed Grafics*. Destaque para as proteínas T2T4SS ID KJV62252.1 no círculo (verde) e T4SS ID KJV62253.1, KJV62505.1, KJV62503.1, KJV62535.1, KJV62563.1 no círculo (roxo) e ID KJV62597.1 no círculo (azul) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via conservada de sinalização para transporte e secreção..... 301

Figura 52 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo domínios repetidos de anquirina (Ank) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV619771.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diversas espécies de carrapatos conforme relatados na literatura. 302

Figura 53 Fluxograma representativo para as espécies de *Rickettsia* do grupo das Febres Maculosas com potencial com potencial patogênico e, ou, espécimes endossimbionticas, transmitidas por carrapatos que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtido com o uso do aplicativo *Yed Grafics*. Destaque para as proteínas contendo domínios repetitivos de anquirina (Ank) no círculo (vermelho) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via conservada de sinalização para fosforilação e metabolismo de aminoácidos, apoptose e epigenia. 303

Figura 54 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas ribossomais metiltransferase (L11) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para as sequências de proteínas IDs KJV62584.1 e KJV62556.1 pertencentes a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diversas espécies de carrapatos, conforme relatados na literatura..... 305

Figura 55 Fluxograma representativo para as espécies de *Rickettsia* do grupo Belli e do grupo das Febres Maculosas sem potencial patogênico e, ou, espécimes

endossimbionticas, sem nenhum caso humano descrito, podendo ser transmitidas por carrapatos em diversas partes do mundo, obtidos com o uso do aplicativo *Yed Graphics*. Destaque para as proteínas ribossomais metiltransferase (L11) no círculo (verde) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via de sinalização conservada para a metilação do DNA. 306

Figura 56 A simbiose de invertebrados e a associação com artrópodes apareceram precocemente em membros de Rickettsiales, como *Anaplasma*, *Ehrlichia*, *Wolbachia*, *NeoRickettsia* e *Orientia*. Destaque para as espécies de *Rickettsia* que se relacionaram com protistas unicelulares, como espécies de *Hydra* surgindo anteriormente do que a espécie *R. bellii*. SFG, Grupo das Febres Maculosas; TG, Grupo do Tifo; PSE, elementos palindrômicos de *Rickettsia*.. 309

Figura 57 Eventos evolutivos que moldaram os plasmídeos de *Rickettsia* e *Orientia*. Os genes utilizados foram hsp, dnaA, sca12, transposase 1 e 2, fosfolipase, traD, repetição de leucina, helix-turn-helix e timidilato quinase. 314

LISTA DE TABELAS

Tabela 1 Métodos de análises que comprovam a interação entre proteínas disponibilizados no banco de dados do Peimap, com atribuição da confiabilidade para cada tipo de validação experimental.	78
Tabela 2 Parâmetros e passos adotados para a análise de averiguação da função molecular e o processo biológico associado às proteínas provenientes do genoma do carrapato da espécie <i>I. scapularis</i>	117
Tabela 3 Espécies de <i>Rickettsia.</i> , estirpes, e seus genomas de referência contendo os IDs taxonômicos.....	138
Tabela 4 Resumo dos resultados das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , utilizando do BLASTP contra as sequências de proteínas depositados no banco de dados do Peimap.	142
Tabela 5 Resumo dos resultados preliminares das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , utilizando do PSIBLAST contra as sequências de proteínas depositados no banco de dados do Psimap.....	143
Tabela 6 Resumo dos resultados preliminares das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , utilizando do HMM contra as sequências de proteínas depositados no banco de dados do Pfam.	144
Tabela 7 Resumo dos resultados preliminares das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , utilizando do BLASTP contra as sequências de proteínas depositados no banco de dados do String.	144
Tabela 8 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , provenientes do BLASTP contra os ids dos domínios das sequências de proteínas depositados no banco de dados do Peimap, contendo os métodos com os quais, estes pares de interação tenham sido validados.....	146
Tabela 9 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , provenientes do PSIBLAST contra os ids dos domínios das sequências de proteínas depositados no banco de dados do Psimap, contendo os valores de distâncias com os quais, estes pares de interação tenham sido preditos.	146
Tabela 10 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , provenientes do HMM contra os ids dos domínios das sequências de proteínas depositados no banco de dados do Pfam, contendo os valores dos dois $> bit scores$ com os quais, estes pares de interação tenham sido preditos..	147
Tabela 11 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, <i>R. amblyommatis</i> e <i>I. scapularis</i> , provenientes do BLASTP contra os ids dos domínios das sequências de proteínas depositados no banco de dados do String, contendo os	

valores de confiança com os quais, estes pares de interação tenham sido preditos.	147
Tabela 12 Resumo dos resultados do cálculo do <i>score</i> final normalizado para os pares de proteínas que interagem entre <i>R. amblyommatis</i> e <i>I. scapularis</i> , presente no banco de dados do Peimap.	149
Tabela 13 Resumo dos resultados do cálculo do <i>score</i> final normalizado para os pares de proteínas que interagem entre <i>R. amblyommatis</i> e <i>I. scapularis</i> , presente no banco de dados do Psimap.	149
Tabela 14 Resumo dos resultados do cálculo do <i>score</i> final normalizado para os pares de proteínas que interagem entre <i>R. amblyommatis</i> e <i>I. scapularis</i> , presente no banco de dados do Pfam.	150
Tabela 15 Resumo dos resultados do cálculo do <i>score</i> final para os pares de proteínas que interagem entre <i>R. amblyommatis</i> e <i>I. scapularis</i> , presente no banco de dados do String.	150
Tabela 16 Resumo dos resultados do cálculo do <i>score</i> combinado com valores $\geq 0,7$ de <i>cutoff</i> para os pares de proteínas que interagem entre <i>R. amblyommatis</i> e <i>I. scapularis</i> , presente no banco de dados do interatoma.	151
Tabela 17 Resumo dos resultados de predição com o uso do aplicativo Signalp 4.1 para as sequências de proteínas provenientes do genoma da espécie <i>R. amblyommatis</i>	152
Tabela 18 Resumo dos resultados de predição com o uso do aplicativo Tmhmm 2.0c para as sequências de proteínas provenientes do genoma da espécie <i>R. amblyommatis</i>	153
Tabela 19 Resumo dos resultados de predição com o uso do aplicativo Lipop 1.0a para as sequências de proteínas provenientes do genoma da espécie <i>R. amblyommatis</i>	153
Tabela 20 Resumo dos resultados de predição com o uso do aplicativo Psortb 3.0.4 para as sequências de proteínas provenientes do genoma da espécie <i>R. amblyommatis</i>	154
Tabela 21 Resumo dos resultados de predição com o uso do aplicativo Phobius 1.01 para as sequências de proteínas provenientes do genoma da espécie <i>R. amblyommatis</i>	155
Tabela 22 Resultados de predição e integração das sequências de proteínas secretadas, excretadas e ou de superfície da espécie <i>R. amblyommatis</i>	155
Tabela 23 Resumo dos resultados de predição e integração das sequências de proteínas secretadas, excretadas e ou de superfície filtradas no interatoma interespecífico de proteínas geral entre as espécies <i>R. amblyommatis</i> vs <i>I. scapularis</i>	157
Tabela 24 Grupos de cluster de ortólogos preditos para proteínas pertencentes as espécies de <i>Rickettsia</i> que infectam carrapatos e insetos.	159
Tabela 25 Resumo dos resultados de integração e filtro dos ortólogos para a espécie <i>R. amblyommatis</i> e outras <i>Rickettsia</i> spp. que infectam os carrapatos no interatoma de proteínas de membrana.	160
Tabela 26 Resumo dos resultados da análise de similaridade das sequências de proteínas de membrana ortólogas para a espécie <i>R. amblyommatis</i> e outras <i>Rickettsia</i> spp. que infectam os carrapatos adotando o alinhamento do TBLASTn.	161

Tabela 27	Resumo dos resultados da análise de similaridade das sequências de proteínas de membrana ortólogas para a espécie <i>R. amblyommatis</i> e outras <i>Rickettsia</i> spp. que infectam os carrapatos adotando o alinhamento do BLASTp.	162
Tabela 28	Módulos funcionais relacionados as interações entre proteínas provenientes dos genomas de <i>R. amblyommatis</i> vs <i>I. scapularis</i> , de membrana para esta bactéria contendo sequências ortologas para <i>Rickettsia</i> spp. que infectam carrapatos sustentadas nos bancos do interatoma obtidos com o uso do aplicativo Cytoscape 3.7 adotando o modulo BINGO 3.0.3.....	171
Tabela 29	Resumo da análise de ontologia gênica obtidos com o pacote BINGO 3.0.3 realizada para as proteínas do carrapato que compuseram os modulos funcionais de secreção e transporte obtidos com o modulo Modland e visualizado no aplicativo Cytoscape 3.7.	172
Tabela 30	Quadro dos resultados parciais das análises de qualidade e trimagem para o sequenciamento dos dados de RNA-seq de espécies de carrapatos <i>Amblyomma</i> spp. infectados e, ou, não por espécies de bactéria <i>Rickettsia</i> spp., com o uso dos aplicativos Fastqc e Trimmomatic. Destaque para o número de <i>reads</i> e as diferenças entre estes valores em (Dados suplementares (S)), após os procedimentos iniciais de qualidade e o final com os dados trimados para a obtenção de sequências de alta qualidade.	185
Tabela 31	Resumo dos resultados preliminares da análise da montagem de novo de dados de RNA-seq com destaque para os arquivos de saída Trinity.Fasta contendo as sequências de RNA montadas.	186
Tabela 32	Resumo dos resultados preliminares da análise do cálculo de abundância e TPM com a indexação do arquivo Trinity.Fasta obtido com uso do aplicativo Trinity 2.5.1 contendo as sequências de RNA ao arquivo transcriptoma.idx obtido com uso no aplicativo Kallisto 0.43.0.	187
Tabela 33	Resumo dos resultados da análise de similaridade para as sequências de proteínas do <i>I. scapularis</i> obtidas dos interatomas contra as sequências de RNA referentes aos transcriptomas montados para órgãos de carrapatos <i>Amblyomma</i> spp. infectados ou não por <i>Rickettsia</i> spp.....	189
Tabela 34	Resumo dos resultados de <i>fold change</i> obtidos com a divisão dos valores de TPM inf por TPM ã inf, destacando os IDs dos transcritos de <i>Amblyomma</i> spp. similares às proteínas de <i>I. scapularis</i> , sendo regulados de forma positiva, negativa ou sem níveis de expressão detectado para uma definição.....	191
Tabela 35	Quadro de resumo da anotação e expressão diferencial para o conjunto de dados de RNA-seq abordando a montagem de novo para cada espécie de carrapato <i>Amblyomma</i> spp. e em determinado órgão infectado ou não por <i>Rickettsia</i> spp., sendo (sg) glândulas salivares, (mg) intestino e (ov) ovários.	191
Tabela 36	Quadro de comparação da anotação e expressão diferencial entre intestinos para o conjunto de dados de RNA-seq das espécies de carrapato <i>A. aureolatum</i> (Aa_Mg) versus <i>A. sculptum</i> (As_Mg), ambos infectados ou não por <i>R. rickettsii</i> adotando o modelo do interatoma de proteínas de membrana e ortólogas.	195
Tabela 37	Comparação de comparação da anotação e expressão diferencial entre intestinos para o conjunto de dados de RNA-seq da espécie de carrapato <i>A. sculptum</i>	

	infectados ou não por <i>R. rickettsii</i> (Rr_Mg) versus <i>R. amblyommatis</i> (Ra_Mg), adotando o modelo do interatoma de proteínas de membrana e ortólogos.....	196
Tabela 38	Resumo dos resultados de integração e filtro dos ortólogos para a espécie <i>R. amblyommatis</i> e outras <i>Rickettsia</i> spp. que infectam os carrapatos no interatoma de proteínas de membrana.	200
Tabela 39	Quadro de legendas para a árvore filogenética da família de sequências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62536.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio do multialinhamentos de sequências e da análise de filogenia.	204
Tabela 40	Quadro de legendas para a árvore filogenética da família de sequências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62591.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio do multialinhamento de sequências e da análise de filogenia.	205
Tabela 41	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção do tipo III T3SS com destaque para a proteína ID KJV62562.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.....	208
Tabela 42	Quadro de legendas para a árvore filogenética da família de sequências de proteínas recombinantes responsáveis pelos movimentos de translocação e conjugação do DNA (Ftsk/SpoIIIE) com destaque para a proteína ID KJV62594.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio do multialinhamentos de sequências e da análise de filogenia.	209
Tabela 43	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T2T4SS com destaque para a proteína ID KJV62252.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.	212
Tabela 44	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62535.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.	215
Tabela 45	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62505.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.	218
Tabela 46	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62253.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio do multialinhamentos de sequências e da análise de filogenia.	221

Tabela 47	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62597.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências da análise de filogenia.	224
Tabela 48	Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62563.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.	227
Tabela 49	Quadro de legendas para a árvore filogenética da família de sequências de proteínas contendo domínios repetitivos de anquirina (Ank) com destaque para a proteína ID KJV61977.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.	230
Tabela 50	Quadro de legendas para a árvore filogenética da família de sequências de proteínas ribossomais metiltransferase (L11) com destaque para a proteína ID KJV62584.1 e KJV62556.1 pertencente a espécie <i>R. amblyommatis</i> Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.	231

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Identificação do problema	29
1.2	Justificativa.....	32
1.3	Objetivos.....	33
1.3.1	Objetivo Geral	33
1.3.1.1	Interatoma de proteínas	33
1.3.1.2	Expressão diferencial de genes.....	34
1.3.1.3	Filogenia de proteínas.....	35
2	REVISÃO BIBLIOGRÁFICA.....	36
2.1	Relações ecológicas e evolutivas entre carrapatos e <i>Rickettsia</i>	36
2.2	Genoma funcional	46
2.2.1	Carrapato	46
2.2.2	<i>Rickettsia</i>	48
2.3	Interatoma de proteínas	50
2.4	Coevolução de proteínas.....	53
3	METODOLOGIA.....	56
3.1	Construção do Interatoma.....	56
3.1.1	Análise de similaridade	57
3.1.2	Montagem do Interatoma	68
3.1.2.1	Interatomas interespecíficos	71
3.1.3	Cálculos de confiança do modelo do interatoma.....	75
3.1.3.1	Cálculo do <i>score</i> final.....	76
3.1.3.2	Cálculo do <i>score</i> combinado	86
3.2	Identificação de potenciais proteínas de membrana.....	92
3.2.1	Integração das proteínas de membrana.....	101
3.2.2	Interatoma de proteínas de membrana.....	102
3.3	Detecção de proteínas ortólogas.....	103
3.3.1	Clusters de proteínas ortologas de <i>Rickettsia</i> spp. que infectam carrapatos versus insetos.....	106
3.3.2	Interatoma de proteínas de membrana e ortólogas de <i>Rickettsia</i> spp. que infectam carrapatos.....	107
3.3.3	Alinhamentos das proteínas ortólogas.....	108
3.3.4	Análise evolutiva das proteínas de membrana ortólogas	109
3.3.4.1	Construção de árvores filogenéticas	115

3.4	Identificação de interações conservadas e específicas	116
3.5	Visualização de redes de interação e análise de ontologia gênica.....	117
3.6	Obtenção de dados de RNA-seq para os transcriptomas de carrapatos.....	120
3.6.1	Análise e controle da qualidade dos dados do sequenciamento	121
3.6.2	Análise de qualidade e trimagem de dados do sequenciamento.....	124
3.6.3	Montagem de novo das sequências de dados de RNA-seq	128
3.6.4	Cálculo dos níveis de abundância da transcrição	129
3.6.4.1	Indexação das sequências de RNA com o transcriptoma	131
3.6.5	Integração das sequências dos transcritos nos interatomas	132
3.6.5.1	Análise da abundância e níveis de expressão da transcrição.....	134
3.7	Evolução de proteínas ortólogas e de membrana	137
4	RESULTADOS	142
4.1	Construção do interatoma geral.....	142
4.1.1	Análise de similaridade das proteínas	142
4.2	Montagem do Interatoma	145
4.2.1	Interatoma interespecífico	145
4.3	Cálculos de confiança do modelo.....	148
4.3.1	Cálculo do <i>score</i> final.....	148
4.3.2	Cálculo do <i>score</i> combinado	151
4.4	Construção do interatoma de proteínas de membrana.....	152
4.4.1	Proteínas secretadas, excretadas e ou de superfície para <i>R. amblyommatis</i>	152
4.4.2	Interatoma de proteínas de membrana.....	157
4.5	Construção do interatoma de proteínas de membrana ortólogas.....	158
4.5.1	Predição de proteínas ortólogas de <i>Rickettsia</i> spp., que infectam os carrapatos	158
4.5.2	Interatoma de proteínas de membrana ortólogas.....	159
4.5.2.1	Análise de similaridade das proteínas ortólogas de membrana	160
4.5.3	Aspectos evolutivos das proteínas ortólogas de membrana	163
4.5.3.1	Alinhamentos das proteínas ortólogas de membrana de <i>Rickettsia</i> spp.	163
4.6	Identificação de interações conservadas.....	166
4.7	Visualização das redes.....	167
4.8	Montagem dos transcriptomas de carrapatos <i>Amblyomma</i> spp.....	181
4.8.1	Qualidade e trimagem das <i>reads</i> dos transcriptomas	182
4.8.1.1	Qualidade do sequenciamento	182
4.8.1.2	Trimagem das <i>reads</i> do sequenciamento.....	183
4.8.2	Montagem “ <i>de novo</i> ” dos transcriptomas de <i>Amblyomma</i> spp.	186

4.8.3	Análise dos níveis de abundância e indexação da transcrição	187
4.9	Análises de expressão diferencial.....	188
4.9.1	Similaridade entre as sequências de transcritos de <i>Amblyomma</i> versus as proteínas do <i>I. scapularis</i> presentes no interatoma de membrana.....	189
4.9.2	Níveis de expressão dos transcritos	190
4.9.3	Expressão de transcritos coexpressos	194
4.10	Revisão de literatura para os relatos que confirmem o isolamento e identificação de espécies de <i>Rickettsia</i> e seus hospedeiros carrapatos	196
4.11	Evolução de proteínas de membrana de <i>Rickettsia</i> spp.....	199
4.11.1	Análise de similaridade	199
4.11.2	Árvores filogenéticas das proteínas de membrana e ortólogas do interatoma ...	200
4.11.3	Árvores filogenéticas para as proteínas de membrana de <i>Rickettsia</i> spp., que infectam somente os carrapatos.....	202
5	DISCUSSÃO	233
5.1	Interatomas	233
5.2	Expressão diferencial de transcritos	238
5.2.1	Secreção e transporte	240
5.2.2	Resposta contra vírus.....	267
5.2.3	Oxidação.....	270
5.2.4	Proteólise	277
5.2.5	Fosforilação	280
5.3	Evolução de proteínas de <i>Rickettsia</i> secretadas, excretadas e ou de superfície	281
6	CONCLUSÃO.....	317
	REFERÊNCIAS BIBLIOGRÁFICA	318
	DADOS SUPLEMENTARES	361
	ANEXOS	363

1 INTRODUÇÃO

1.1 Identificação do problema

Com o crescente aumento no número de genomas sequenciados em bancos de dados públicos, tais como o *GenBANK* (<https://www.ncbi.nlm.nih.gov/genbank/>), ocorreu o incremento no desenvolvimento de ferramentas de bioinformática destinadas a comparar os repertórios de genes e proteínas entre as espécies, incluindo a identificação de genes ortólogos e parálogos (ALTENHOFF et al., 2012), o que tem sido os elementos-chave na anotação do genoma e estudos evolutivos (ALTENHOFF; GLOVER; DESSIMOZ, 2019; KOONIN, 2005). Com a advento das plataformas de sequenciamento de nova geração (SNG) foram geradas informações sobre sequências gênicas em maior escala do que o sequenciamento de Sanger, economizando tempo e custo por bases. Assim, devido ao seu elevado poder e rapidez no sequenciamento de grandes quantidades de sequências de DNA ou RNA (cDNA), tem sido possível a obtenção de um grande número de sequências gênicas ativamente expressas em um dado órgão ou tecido de um organismo qualquer, em uma dada situação fisiológica (MARDIS, 2008; ROTHBERG; LEAMON, 2008).

Ao combinar ferramentas de bioinformática com o crescente número de sequências genômicas, estudos recentes têm fornecido diversas evidências, sugerindo, por exemplo, que genomas bacterianos são mosaicos, incluindo genes compartilhados por todos os isolados de um grupo de interesse (núcleo do genoma), bem como contando com estirpes específicas com genes parcialmente em comum (WELCH et al., 2002). A soma do genoma consenso e os demais genes dentro do grupo são definidos como pangenoma (TETTELIN et al., 2008). Essa estratégia de análise tem se destacado no estudo da evolução e filogenia ao estruturar, comparar e agrupar dados dos genomas, utilizando por princípio a distância de dissimilaridade e similaridade entre os perfis pangenômicos preestabelecido (SNIPEN; USSERY, 2010), sendo os melhores métodos de agrupamento aqueles que produzem a menor distância dentro do grupo e a máxima distância entre grupos formados, o que já foi realizado para algumas bactérias do gênero *Rickettsia* transmitidas por carrapatos (DUAN et al., 2014).

Para os artrópodes hematófagos vetores de doenças, um futuro pangenoma traz consigo a possibilidade do desenvolvimento de vacinas de amplo espectro baseadas nas proteínas conservadas durante a evolução dos artrópodes (MEJIA; BISHOP; TITUS, 2006). Entretanto, a seleção de epítomos em proteínas de artrópodes candidatos a vacinas têm como problema o

polimorfismo antigênico durante a sua concepção, pois algumas sequências de moléculas polimórficas de artrópodes têm apresentado significância geográfica restrita (YIN et al., 2004).

Apesar de grandes lacunas ainda existirem na compreensão da fisiologia e mecanismos moleculares associados a hematofagia de artrópodes, grandes avanços foram feitos neste campo através de análise genômica funcional das glândulas salivares e intestino médio de artrópodes hematófagos (MEJIA; BISHOP; TITUS, 2006).

A variação espacial de interações entre parasitas e seus hospedeiros é a força por trás dos aspectos da seleção natural neste processo coevolucionário (THOMPSON, 1994) e na geração de diversidade biológica (THOMPSON, 1999), sendo o estado de simpatia do parasita aos hospedeiros um estado de adaptação de fundamental destaque para a hipótese da prevalência da reprodução sexual (LIVELY, 1996), permitindo aos mesmos que sejam passados de geração em geração via materna. Estas interações entre *loci* intragenômicos codificadores em parasito são fundamentais para determinar o resultado da coevolução (HAMILTON; BOOTS; PATERSON, 2005). Segundo estes autores, espera-se um efeito desproporcional sobre a aptidão do parasito quando estes *loci* interagem sinergisticamente, o que favorece o genótipo de parasito generalista, com efeitos independentes sobre o *fitness* ou antagonistas, levando a diminuição dos efeitos sobre sua aptidão, onde os genótipos do parasita especializado são favorecidos.

Atualmente, os mecanismos genéticos subjacentes à expressão sequencial de membros individuais de uma família multigene em sialomas de carrapatos são desconhecidos (CHMELLAŘ et al., 2016), notando-se o mesmo fato para outros órgãos, como singânglios e ovários, sendo, no entanto, elementos transponíveis e genes de origem viral consistentemente detectados, sugerindo a ocorrência de mudanças ativas na estrutura genômica do carrapato (KARIM; SINGH; RIBEIRO, 2011; KOTSYFAKIS et al., 2015). Estes fenômenos, juntamente com a influência das histonas nas modificações da expressão gênica observada em estudos para a espécie *Amblyomma maculatum* (ADAMSON et al., 2013), permitem supor que a transcrição sequencial possui regulação epigenética nestes ácaros (KOTSYFAKIS et al., 2015). Os estudos da relação de carrapatos infectados por *Rickettsia* spp. sugerem significativas alterações na expressão de genes constitutivos, reforçando a hipótese sobre a manipulação da maquinaria de expressão gênica do carrapato por esta bactéria (MOREIRA, 2016).

A infecção por *Rickettsia amblyommatis* sugere provocar um efeito antagonista para a infecção na expressão de genes nos órgãos dos carrapatos *Amblyomma* spp. em relação a infecção no intestino, levando a uma regulação positiva de genes putativos relacionados com a transmissão transovariana, resultando, ao mesmo tempo, em uma carga metabólica para este órgão do carrapato (MOREIRA, 2016). Este potencial do genoma de *Rickettsia* em poder codificar vários sistemas de secreção são homólogas com as vias de secreção de proteínas caracterizados em outras bactérias. Por exemplo, pode-se citar os sistemas *Sec translocon* (Sec-TolC) e o sistema de secreção Sec-dependente tipo V (T5SS) (GILLESPIE et al., 2014). Em contraste com os sistemas de secreção de proteínas e vias descritas, uma substancial falta de conservação é vista dentro de muitas famílias de proteínas secretoras tais como SCA5, Sca4, ADR1, ADR2, TlyC, PLD e RARP-1 que carecem de genes em um ou mais genomas de *Rickettsia* (GILLESPIE et al., 2014). Segundo estes autores, a maioria destes genes apresenta codificação variável de proteínas ricas em motivos repetidos e outras com regiões hipervariáveis, sugestivos de evolução rápida e provavelmente impulsionada por pressões seletivas de adaptação na célula hospedeira.

A abordagem do interatoma entre dados de sequência de proteínas em outros patógenos intracelulares tais como *Leishmania* spp. versus as proteínas do genoma humano tem apontado pares de proteínas que supostamente interagem com chance de traços marcados em ambos organismos (FLÓREZ et al., 2010; GONG et al., 2005; MERING, 2003; SONNHAMMER, 1998). Segundo estes autores, avaliando multialinhamentos de sequências, tem sido identificado domínios e proteínas de diversos organismos consideradas com potencial para interagir, quando a presença de ambas ocorre entre os dois domínios para os bancos de dados do interatoma.

Estes achados de proteínas e complexos com as suas possíveis interações em hipótese poderia ter passado por um processo coevolucionário moldado pela relação parasito-hospedeiro. Esta interação, em nível molecular e sua história evolucionária, deve ser aferida quanto à funcionalidade deste sistema, averiguando o interatoma entre a interface destes organismos, os aspectos coevolutivos e a expressão diferencial destes genes e suas proteínas que possivelmente seriam encontradas em módulos funcionais de vias fisiológicas.

Dessa forma, tem-se a hipótese de que ocorre interação proteína-proteínas e suas famílias em nível de resíduo, domínios e complexos, podendo ser diferente entre as espécies de hospedeiros com os quais coevoluem. Os dados de transcriptoma de *Amblyomma* spp. sugerem

que exista interações entre proteínas para ambos os organismos considerando a espécie do carrapato e da bactéria podendo haver distintos grupos funcionais de vias moleculares sendo ativados.

Como objetivo principal, este estudo buscou prever as proteínas de *Rickettsia* spp. que interagem com as proteínas do carrapato, prevendo quais são secretadas, excretadas e/ou de superfície, e destas quais proteínas são ortólogas para as espécies que infectam carrapatos versus as que infectam insetos. A construção de modelos de interatoma, filtrados à partir de proteínas de membranas escolhendo sequências ortólogas, pode ser aprofundada quando utilizamos de dados da expressão diferencial dos genes que as compõe, sobre os módulos funcionais significativos. Dessa forma, pudemos prever quais pares de proteínas interagem, coevoluem e ativam genes, podendo estes sofrer regulações positivas e ou negativas. Ao final, buscamos demonstrar os processos ocultos ocorridos para o estabelecimento desta relação perante a interface coevolucionária, esclarecendo a funcionalidade deste sistema para obter alvos que serão destinados para validações futuras por técnicas bioquímicas e moleculares.

Com isto, visamos a busca por correlação auxiliando na elucidação da relação parasito-hospedeiro, em especial de patógenos veiculados por carrapatos, sendo este um passo para o desenvolvimento de estratégias de combate e controle destes patógenos.

1.2 Justificativa

O interesse no desenvolvimento de novos medicamentos e vacinas leva em consideração, no caso de doenças infecciosas, que a via alvo destes seja essencial para a sobrevivência do parasita e específicos, de forma que o alvo não deve ter um homólogo. No entanto, não há consenso ainda sobre os melhores alvos biológicos de essencialidade, bem como o nível de expressão e localização subcelular destes. Contudo, estes pressupostos muitas vezes não levam em consideração a complexidade da rede de interações biológicas subjacentes entre estas proteínas.

As abordagens baseadas no estudo do interatoma na seleção de melhores complexos de proteínas que interagem e na coevolução deles, podem ser considerados como parte dos métodos para detectar interações e as relações funcionais entre proteínas usando informação genômica, conhecido em conjunto como métodos de "base de contexto". Esta situação é uma reminiscência de redes ecológicas em que a interação entre espécie ocorre em condições

ambientais, onde mudanças no espaço e no tempo e no qual coevoluiram, considerando os aspectos dinâmicos destas interações.

Neste sentido, o estudo aqui proposto visou identificar as possíveis vias de alvos e inferir a respeito dos processos biológicos realizados por proteínas preditas, com a descrição funcional destas.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo deste estudo foi prever e comparar modelos de infecção sob abordagem do interatoma entre proteínas do carrapato *Ixodes scapularis* e *Rickettsia amblyommatis*, abordando os aspectos moleculares e funcionais ocorridos no estabelecimento desta interação putativa, a coevolução destas proteínas, seus níveis de expressão e evolução entre diferentes espécies de *Rickettsia*.

1.3.1.1 Interatoma de proteínas

- Construir a rede geral de interação de proteína-proteína entre a bactéria da espécie *R. amblyommatis* versus o carrapato da espécie *I. scapularis* à partir das sequências de proteínas codificadas pelos genomas destes organismos;
- Prever as potenciais proteínas de membrana de *R. amblyommatis* secretadas/excretadas e/ou de superfície à partir das sequências de proteínas codificadas pelo genoma deste organismo;
- Construir a rede de interação de proteínas de membrana secretadas/excretadas e/ou de superfície preditas para a espécie *R. amblyommatis* pela filtragem contra as proteínas da bactéria que formaram pares de interação de proteínas preditos no interatoma geral entre *R. amblyommatis* vs *I. scapularis*;
- Prever as proteínas ortólogas para a espécie *R. amblyommatis* e as espécies de *Rickettsia* que infectam os carrapatos versus as que infectam os insetos, à partir das sequências de proteínas codificadas pelos genomas destes organismos;

- Construir o interatoma de proteínas de membrana ortólogas preditos para a espécie *R. amblyommatis* e as espécies de *Rickettsia* que infectam os carrapatos pela filtragem contra as proteínas da bactéria que formaram pares de interação de proteínas preditos no interatoma de proteínas de membrana entre *R. amblyommatis* vs *I. scapularis*;
- Realizar alinhamentos e a análise de evolução das proteínas da espécie *R. amblyommatis* ortólogas para as espécies de *Rickettsia* que infectam os carrapatos presentes no interatoma de proteínas de membrana, versus as sequências gênicas e das proteínas codificadas pelos genomas destes organismos;
- Realizar a anotação de ontologia gênica e a visualização das redes e dos módulos funcionais significativos para as proteínas do *I. scapularis* que potencialmente participem de interações entre as proteínas de *R. amblyommatis*.

1.3.1.2 Expressão diferencial de genes

- Avaliar a qualidade do sequenciamento e a trimagem das *reads* à partir dos dados de RNA -seq para as espécies de carrapato *Amblyomma* spp., *Amblyomma aureolatum* e *Amblyomma sculptum* (Complexo *A. cajennense*), infectados e não infectados por espécies de *Rickettsia* do Grupo das Febres Maculosas;
- Construir os *transcriptômas* por “montagem *de novo*” para a obtenção das *reads* montadas e as sequências de RNA em formato Fasta das espécies de carrapato *Amblyomma* spp., *A. aureolatum* e *A. sculptum* (Complexo *A. cajennense*), infectados e não infectados por espécies de *Rickettsia* do Grupo das Febres Maculosas;
- Avaliar os níveis de abundância dos transcritos e a indexação dos transcritos a todas as sequências de RNA dos *transcriptomas* montados para a expressão diferencial e o cálculo dos níveis de abundância destes transcritos das espécies de carrapato *Amblyomma* spp., *A. aureolatum* e *A. sculptum* (Complexo *A. cajennense*), infectados e não infectados por espécies de *Rickettsia* do Grupo das Febres Maculosas;
- Integrar os resultados dos *transcriptômas* ao interatoma de proteínas de membrana e ortólogas para a espécie *R. amblyommatis* versus *I. scapularis* à partir da filtragem nos módulos funcionais significativos destes interatomas para a anotação das proteínas possivelmente expressas;

- Comparar os resultados dos transcritos anotados em cada módulo funcional entre as espécies *A. aureolatum* e *A. sculptum* (Complexo *A. cajennense*) sob a mesma condição de infecção por *Rickettsia rickettsii*, uma espécie patogênica;
- Comparar os resultados dos transcritos anotados em cada módulo funcional para a espécie de carrapato *A. sculptum* (Complexo *A. cajennense*) sob condições de infecção para um mesmo órgão por duas espécies diferentes, *R. rickettsii* (patogênica) e *R. amblyommatis* (não patogênica).

1.3.1.3 Filogenia de proteínas

- Realizar revisões bibliográficas para compreender as relações ecológicas e evolutivas de bactérias do gênero *Rickettsia* com seus possíveis hospedeiros carrapatos, à partir não apenas de dados epidemiológicos para as rickettsioses, mas principalmente de estudos moleculares que relataram por meio de técnicas moleculares a presença destes microrganismos, que tiveram a taxonomia comprovada por meio do isolamento microbiológico e do sequenciamento do seu genoma;
- Realizar alinhamentos e a análise de evolução das proteínas da espécie *R. amblyommatis* presentes no interatoma de proteínas de membrana, versus as sequências de aminoácidos das proteínas codificadas pelos genomas de distintas espécies de distintas espécies de *Rickettsia* do Grupo das Febres Maculosas, Grupo do Tifo, Grupo de Transição e do Grupo Ancestral além de outras bactérias do grupo das *Rickettsiales*, que infectam os carrapatos, insetos e ácaros;
- Construir árvores filogenéticas e evolutivas para as sequências de aminoácidos das proteínas e suas famílias
- Realizar o cálculo de médias das árvores filogenéticas para as proteínas de membrana presentes no interatoma separando aquelas, as quais as proteínas de *Rickettsia* spp. infectam são ortólogas para espécies que infectam os carrapatos e divergem das proteínas das espécies que infectam os insetos ou outros artrópodes.

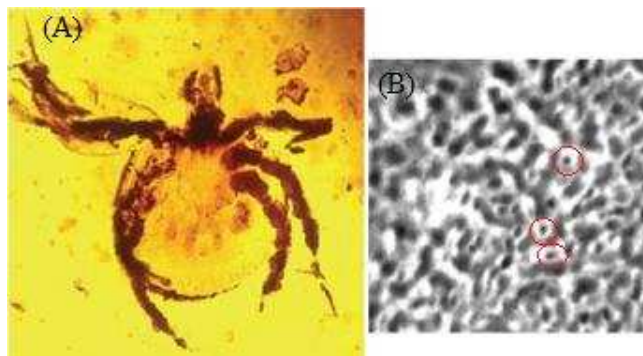
2 REVISÃO BIBLIOLIOGRÁFICA

2.1 Relações ecológicas e evolutivas entre carrapatos e *Rickettsia*

Os carrapatos estão entre os principais vetores de patógenos ao homem, situando-se atrás apenas dos culicídeos (DURDEN, 2006; JONGEJAN; UILENBERG, 2004), sendo as bactérias pertencentes ao gênero *Rickettsia* os principais agentes infecciosos transmitidos por estes ácaros, ocasionando surtos em todo mundo (JONES et al., 2008). Estes artrópodes pertencentes à Classe Arachnida, Ordem Acari são classificados em três famílias: Argasidae, Ixodidae e Nuttalliellidae, sendo as duas primeiras cosmopolitas e a última restrita à África do Sul e Tanzânia, (BEDFORD, 1931; KEIRANS; CLIFFORD; CORWIN, 1976). Os membros do gênero *Rickettsia* são classificados na Ordem *Rickettsiales*, família *Rickettsiaceae* e divididos em quatro grupos: Grupo Ancestral, Grupo do Tifo, Grupo de Transição e Grupo das Febres Maculosas (GILLESPIE et al., 2008). As espécies de *Rickettsia* são bactérias Gram-negativas, intracelulares obrigatórias (LA SCOLA; RAOULT, 1997), apresentando formato de cocóides, cocobacilares ou pleomórficos, medindo 0,3 a 0,6 μm de largura, retendo fucsina quando coradas pelo método de Gimenez (GIMÉNEZ, 1964).

Os registros fósseis de carrapatos da família Ixodidae, conhecidos como carrapatos duros, remontam ao início do Cretáceo (POINAR; BROWN, 2003; POINAR; BUCKLEY, 2008). Em hipótese, estes ácaros poderiam ter adquirido microrganismos, seguido por um período de coevolução correlacionado ao estabelecimento do seu ciclo de alimentação por hematofagia sobre hospedeiros vertebrados e, conseqüentemente, desenvolvimento de sua competência como vetores de patógenos (POINAR, 2014). Presume-se, que a origem dos carrapatos tenha ocorrido entre a radiação de mamíferos e aves (ESCALANTE; AYALA, 1995), podendo ter parasitado estes organismos num elo de especiação ter tido os dinossauros como possíveis hospedeiros (JACOBSON, 2008) (Figura 1).

Figura 1 Relato de microrganismos próximos a *Rickettsia* spp. (ciclo vermelho). coletado na hemolinfa (B) de uma ninfa de carrapato Ixodídeo da espécie *Cornupalpatum burmanicum*, já extinto, preservado em âmbar (A) encontrado na Península Dominicana.



Fonte: Poinar et al. (2014).

Na Figura 1, verifica-se a presença de organismos próximos à *Rickettsia* no registro fóssil de carrapato do espécime *Cornupalpatum burmanicum*, já extinto e que predominou durante o Cretáceo (POINAR, 2014), são evidências de coevolução. Atualmente as espécies de *Rickettsia* e seus hospedeiros são responsáveis por manterem esse microparasita no ambiente silvestre, aonde seu ciclo de transmissão dá-se por passagem transovariana, venérea (BURGDORFER; VARMA, 1967a, 1967b; PAROLA; PADDOCK; RAOULT, 2005a), ou transestadial para as novas gerações de carrapatos com a passagem de um vetor para outro via um hospedeiro vertebrado infectado (MCDADE; NEWHOUSE, 1986). No entanto, são os artrópodes, ao invés dos hospedeiros vertebrados, que ditam a composição de comunidades bacterianas em carrapatos, podendo esta variação ser resultado de outros fatores, tais como interações interespecíficas, entre as bactérias, ou eventos de resultados estocásticos, sendo a dinâmica da transmissão de microrganismos infecciosos considerada como a força propulsora da dinâmica de populações de patógenos (HAWLENA et al., 2013).

A evolução dos patógenos veiculados por carrapatos seria conduzida então por barreiras biológicas distintas, mas direcionadas e limitadas por condicionantes ambientais, onde são relatadas correlações para várias espécies de carrapatos, com a temperatura e a umidade influenciando diversas etapas do ciclo de vida, p. ex., mortalidade, tempo de procura de hospedeiro e tempo da muda, e, conseqüentemente, em suas abundâncias (RANDOLPH, 2004b). Atualmente, a classificação filogenética e a identificação de comunidades de bactérias endossimbiontes em carrapatos, em parte se devem aos condicionantes da infecção, como a espécie, o sexo, o estágio de desenvolvimento e o nível de ingurgitamento do hospedeiro (GUTIÉRREZ et al., 2014; HAWLENA et al., 2013; HEISE; ELSHAHED; LITTLE, 2010;

MORENO et al., 2006). Estudos têm incluído também o período de coleta com autocorrelação temporal (TELFER et al., 2010), o que sugere haver uma forte sintonia do patógeno com a dinâmica populacional de seu hospedeiro carrapato (RANDOLPH, 2004a), podendo as espécies de *Rickettsia* usar sinais ambientais, no caso a temperatura, para sincronizar sua virulência com os períodos de máxima disponibilidade desses ácaros (GOMES, 2014).

Durante seu ciclo de vida, estas bactérias são submetidas a alterações dramáticas nos diferentes ambientes a que têm de se adaptar, com a temperatura dos ambientes nos quais estão se desenvolvendo, o que pode variar amplamente ao alternarem entre um vetor invertebrado e um hospedeiro vertebrado, sugerindo uma provável expressão diferenciada de seus genes (KONKEL; TILLY, 2000). As bactérias do gênero *Rickettsia* são capazes de se multiplicar em células dos intestinos, ovários, glândulas salivares, túbulos de Malpighi e hemolinfa, penetrando nas células hospedeiras de invertebrados através de receptores específicos de membrana (WALKER, 1989), podendo ainda se multiplicar no interior das células epiteliais de vasos sanguíneos de hospedeiros vertebrados por divisão binária, sendo mantidas no carrapato quando em prolongada ausência de hematofagia por sucessivas gerações, tendo sua virulência reativada por fatores nutricionais presentes no sangue de hospedeiros vertebrados, ou quando da exposição à temperatura adequada em torno de 28°C (RAOULT; ROUX, 1997; WALKER, 1989).

Em nível molecular, estudos indicaram que a temperatura pode modular o perfil transcriptômico, causando indução genômica em várias espécies de *Rickettsia* em carrapatos (DREHER-LESNICK et al., 2008; ELLISON et al., 2009; GALLETTI et al., 2016) e em culturas de células (POLICASTRO et al., 1997). Estes mesmos autores sugerem que além da temperatura, outros condicionantes ambientais, como pH e osmolaridade, também tem sido reportado como importantes para as bactérias sobreviverem e prosperarem em seus diversos nichos em resposta a um estímulo ambiental, sendo que genes coordenadamente regulados por um sistema regulador, muitas vezes contribuem direta ou indiretamente para a virulência do organismo. De certa forma, as espécies de bactérias endossimbióticas possuem este status por ser o mais sofisticado em que o hospedeiro e simbiote estão integrados uma entidade biológica coerente não poderia sobreviver sem a parceria (RICHARDSON, 2010).

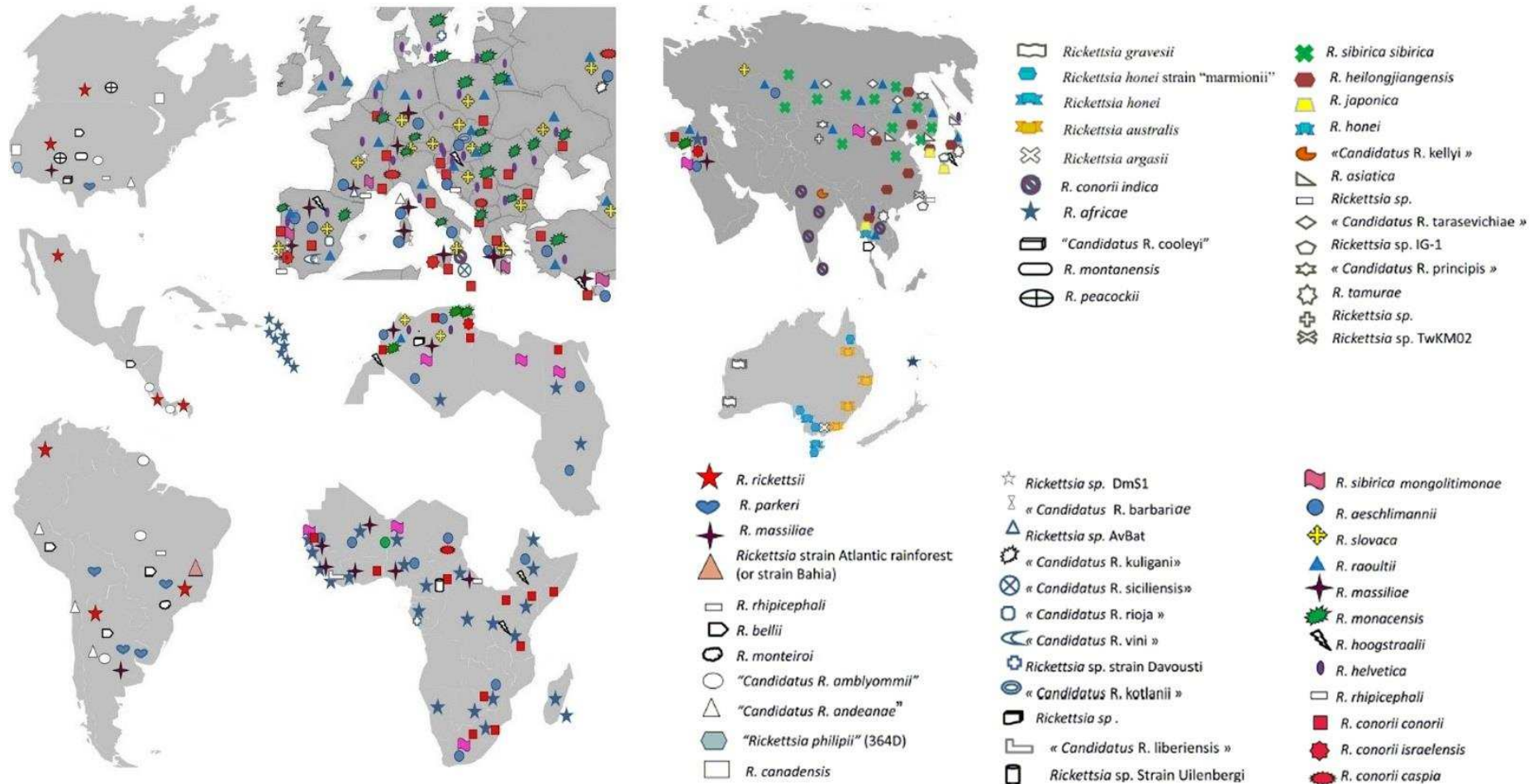
Não obstante a isto, a fisiologia dos carrapatos hospedeiros podem vir a ser modulados pela *Rickettsia*, ocasionando, conseqüentemente, na geração futura, a diminuição da população, com a postura de ovos inviáveis, reduzindo desta maneira sua taxa de fertilidade

(BURGDORFER; BRINTON, 1975). Modelos experimentais de infecção colocaram em cheque o efeito letal da relação entre *Rickettsia conorii* e seu hospedeiro invertebrado, a espécie de carrapato *Rhiphicephalus sanguineus* (LEVIN et al., 2009; SOCOLOVSKI et al., 2009a), sugerindo que as razões para a não redução em *fitness* mostraram-se alheias à origem geográfica do carrapato, diminuindo a aptidão deste patógeno (MATSUMOTO et al., 2005). No entanto, o tamanho reduzido do genoma com a perda de genes em bactérias tem sido relacionado ao seu estilo como a força motriz na adaptação de *Rickettsia* e outras bactérias para viver dentro de células eucarióticas (MERHEJ et al., 2009; MERHEJ; RAOULT, 2011). Isto sugere que a degradação do genoma em *Rickettsia* estaria associada ao aumento de virulência (FOURNIER et al., 2009), sendo esta evolução redutiva marcada por uma grande proporção de sequências notadamente não codificadoras, resultando em genes com funções degradadas em vias biossintéticas (PAROLA et al., 2013).

Estudos com base na genômica comparativa demonstraram que a especialização de bactérias em células eucarióticas está associada com uma perda massiva de genes (MERHEJ et al., 2009) e, conseqüentemente, a perda de "fatores de virulência" (LESCOT et al., 2008). Esta especialização também está associada com a perda da tradução de genes, constituindo um passo fundamental na evolução, em direção a uma associação mais íntima com o hospedeiro (MERHEJ et al., 2009). A redução do número de *operons* ribossomais também seria outro fator da especiação de genes que codificam RNAr em bactérias intracelulares obrigatórias, possuindo menos cópias destes genes, em relação às bactérias de vida livre (KLAPPENBACH; DUNBAR; SCHMIDT, 2000). Outro fator desta especiação foi a perda de um grupo de 100 genes permitindo serem intracelulares obrigatórias, com uma evolução convergente estável e de irreversível associação com o hospedeiro (MERHEJ et al., 2009). Com a mais extrema forma de deterioração do genoma observada para o estado de alopatria de bactérias associadas aos seus hospedeiros, particularmente endosimbiontes, mantidos em um nicho intracelular estático e pouco exigente (MERHEJ et al., 2009; MORAN, 2002; MORAN; WERNEGREN, 2000).

Atualmente, registram-se em torno de 136 espécies de carrapatos interagindo com, pelo menos, 26 espécies validadas de *Rickettsia*, em diferentes continentes ao redor do planeta, com exceção do continente Antártico, no qual não há relatos da presença de carrapatos infectados com bactérias do gênero *Rickettsia* (PAROLA et al., 2013) (Figura 2).

Figura 2 Distribuição geográfica atual de bactérias do gênero *Rickettsia* transmitidas por espécies de carrapatos ao redor do mundo.



Fonte: Parola et. (2013).

Na Figura 2, verifica-se que a relação entre linhagens de patógenos transmitidos por carrapatos (PAROLA et al., 2013), sugere haver um passo evolutivo posterior a hematofagia, em que estes invadiram as glândulas salivares, de onde poderiam ser transferidos para os vertebrados (POINAR, 2014). De certa forma, algumas outras espécies simbióticas de carrapatos e outros artrópodes hematófagos, como *Wolbachia*, são benéficas e ajudam na digestão de refeições de sangue (POINAR, 2014). Isto sugere que o mutualismo nutricional obrigatório entre artrópodes-bactéria está entre as formas mais sofisticadas de simbiose, em que o hospedeiro e simbiote estão integrados em uma entidade biológica coerente e incapazes de sobreviver sem a parceria (NIKOH et al., 2014).

Originalmente, essas bactérias simbióticas obrigatórias devem ter sido derivadas de bactérias de vida livre, onde o mutualismo obrigatório foi altamente especializado, originando associações menos especializadas (POINAR, 2014). Através da relação íntima ao longo do tempo evolutivo, estes e outros genomas endosimbionticos foram reduzidos drasticamente, perdendo muitos genes necessários para a vida independente e simplificados para as funções biológicas específicas para apoiar os seus hospedeiros, tendo adquirido novas propriedades biológicas através destes microrganismos que desempenharam papéis importantes na adaptação, evolução e diversificação de artrópodes e outros organismos (RICHARDSON, 2010).

O gênero *Rickettsia* é um grupo extraordinariamente diversificado com espécies conhecidas em plantas, amebas, ciliados, artrópodes, anelídeos, sanguessugas e animais vertebrados (PERLMAN; HUNTER; ZCHORI-FEIN, 2006), havendo uma linhagem derivada, que contém diversas espécies conhecidas, associadas à artrópodes, como: carrapatos, ácaros, pulgas e piolhos, sendo organizadas nos grupos Grupo Ancestral, Grupo do Tifo, Grupo de Transição e Grupo das Febres Maculosas (GILLESPIE et al., 2008). Os genomas de *Rickettsia* possuem, em média, entre 1,1 – 1,5 (Mb) contendo aproximadamente em torno de 900 – 1,500 genes, refletindo na plasticidade e probabilidade de diversidade de estratégias para infectarem e viverem em seus hospedeiros (MERHEJ; RAOULT, 2011), sendo de fundamental papel na coevolução entre estas bactérias e seus vetores na transmissão interespecies (KANG et al., 2014). Percebe-se que a degradação do genoma, especialmente a perda de genes, tem sido uma força motriz na adaptação de *Rickettsia* e outras bactérias intracelulares para viverem dentro das células eucarióticas (MERHEJ et al., 2009).

Para os organismos do gênero *Rickettsia*, um dos paradigmas mais intrigantes de evolução reductiva foi o ocorrido para a espécie *Rickettsia prowazekii*, a mais virulenta do gênero pertencente ao Grupo do Tifo (MERHEJ; RAOULT, 2011), onde, surpreendentemente, 24% do seu pequeno genoma é composto de pseudogenes e de DNA não codificantes (ANDERSSON et al., 1998). Apesar deste fato, o sequenciamento do genoma de diversas espécies de *Rickettsia* não possibilitou a identificação de quaisquer genes de virulência identificável, e somente uma área de plasticidade genômica responsável pela restauração da virulência em *R. prowazekii*, representando um exemplo de mutação adaptativa por este agente patogênico (BECHAH et al., 2010). Neste sentido, a motilidade intracelular não existe em *R. prowazekii*, sendo esta espécie transmitida via aerossóis provenientes das fezes de piolhos, diferentemente da picada destes vetores. Por outro lado, a espécie *Rickettsia typhi*, pertencente ao Grupo do Tifo e menos virulenta, é móvel no citoplasma e estreitamente relacionada a *R. prowazekii*. De maneira distinta, a motilidade de *Rickettsia rickettsii*, pertencente ao Grupo das Febres Maculosas, requer atividade de duas proteínas riquetsiais, SCA2 e ricka, funcionando em conjunto, o que sugere que estas duas proteínas podem ser fatores de virulência de *R. rickettsii* (MERHEJ; RAOULT, 2011).

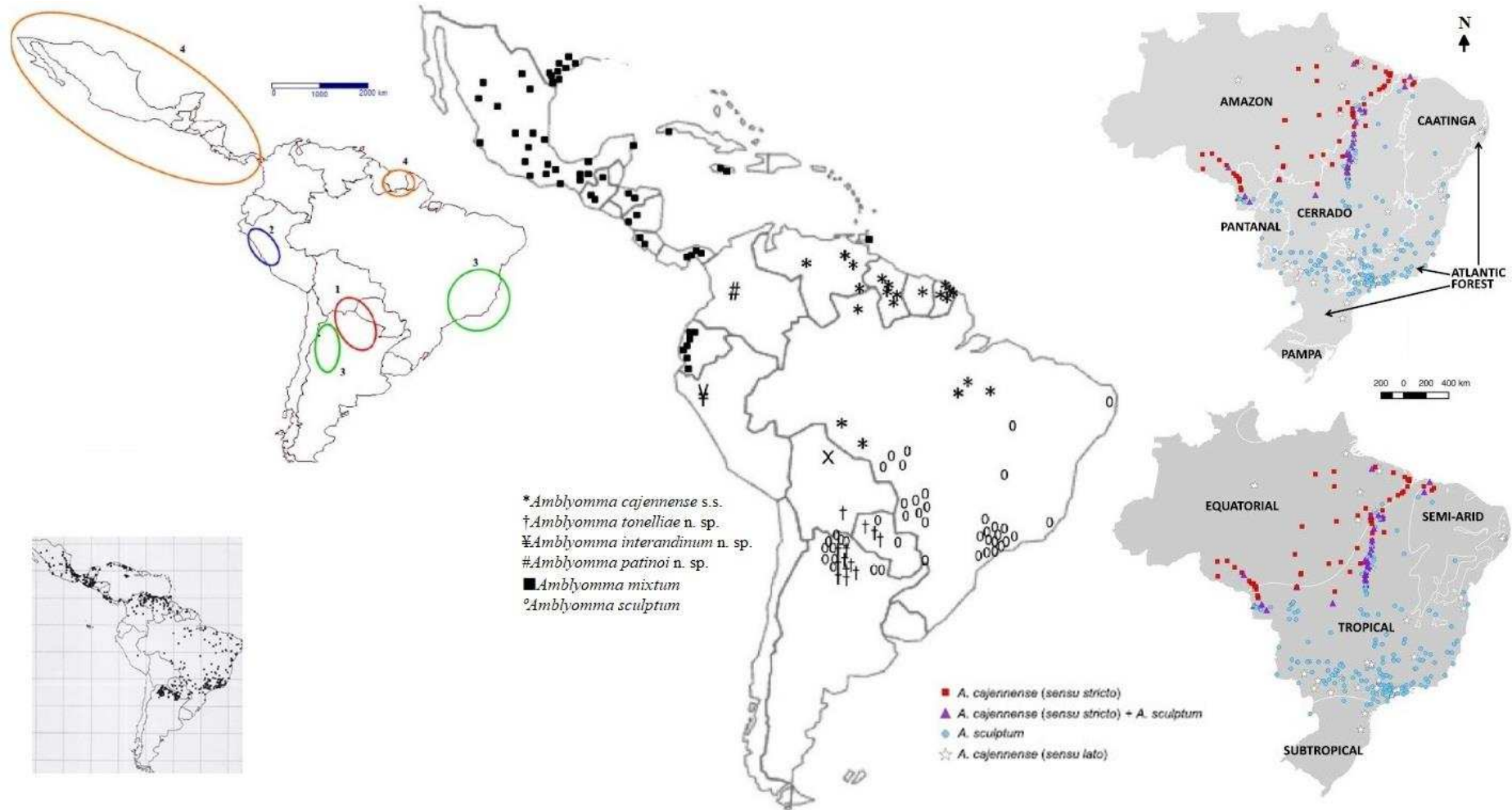
Esta observação sugere que a mobilidade não é um fator de virulência em si, mas pode ser encontrado em patógenos como parte de um repertório de virulência. A motilidade é útil para *R. typhi*, *R. rickettsii* e algumas riquetsias não patogênicas, mas não para *R. prowazekii*. Muitos genes que codificam proteínas envolvidas na biossíntese de aminoácidos foram perdidos no genoma de *R. prowazekii*, o que ocasionou diminuição da capacidade de tradução afetando fatores de regulação de tradução, onde tais perdas poderiam alterar mecanismos que regulam a invasão, replicação e processos de transmissão, além de induzir um fenótipo virulento (ANDERSSON et al., 1998). Curiosamente, estudo recente comparando *Rickettsia africae* e *R. rickettsii* em carrapatos demonstrou que a perda de genes essenciais, tais como genes reguladores em *R. rickettsii*, seria um possível fator envolvido no desenvolvimento da patogenicidade (FOURNIER et al., 2009). Neste contexto, a realização de um repertório de genes relativamente importante por processos de recombinação e reparação pode resultar da necessidade de oferecer proteção contra a resposta imune do hospedeiro contra o influxo de elementos genéticos móveis (MORAN, 2002). No caso de outras espécies, a taxa estimada de perda de gene foi altamente variável, onde *Rickettsia felis* pertencente ao Grupo das Febres Maculosas, em pulgas, e *Rickettsia bellii* pertencente ao Grupo Ancestral, sofrem menos perdas de genes do que outras espécies de *Rickettsia* (BLANC et al., 2007).

Para os carrapatos, graças a esta maior cobertura da análise do transcriptoma por SNG e as novas abordagens computacionais da bioinformática, análises filogenéticas tem fornecido mais informações sobre a evolução de famílias de proteínas multigenes em glândulas salivares de carrapatos (CHMELÁŘ et al., 2016). Provavelmente, estes genes múltiplos se originam à partir de ancestrais comuns ao sofrerem eventos de duplicação de genes e mutações subsequentes com a evolução divergente, onde estudos comprovaram taxas aceleradas de mutação em famílias multigênicas, revelando entre as espécies de carrapatos avaliadas que certos *clados* da árvore filogenética foram preferencialmente ocupados por membros da mesma família provenientes de uma única espécie de carrapato (KARIM; SINGH; RIBEIRO, 2011; SCHWARZ et al., 2013). Em outras palavras, a maioria das duplicações gênicas e, portanto, a maioria das proteínas provenientes de famílias multigenes possui em comum carrapato ancestral de linhagem específica (MANS et al., 2008).

Em outras abordagens filogenéticas, o estudo do rDNA 12S e 16S, e dos *loci* de microssatélites, revelou que, pelo menos, duas espécies diferentes de carrapatos partilham o nome *Rhipicephalus (Boophilus) microplus*, de forma que estas espécies podem ser redefinidas como *Rhipicephalus microplus* (Canestrini, 1887) (para as estirpes americanas e africanas) e, provavelmente, o antigo *Rhipicephalus australis* Fuller, 1899 (para as estirpes australianas), o qual necessita ser redescrito (LABRUNA et al., 2009). Outros indícios revelaram que a espécie *Amblyomma cajennense* é composta por um complexo de espécies, as quais se encontram agrupadas desde a década de 1950, sob um mesmo táxon (ARAGÃO; FONSECA, 1961; NAVA et al., 2014). Cruzamentos experimentais entre populações de *A. cajennense* provenientes de diferentes regiões geográficas (Salta, Argentina; Villeta, Colômbia; São Paulo e Rondônia, Brasil), resultaram num baixo rendimento de nascimento de larvas (LABRUNA et al., 2011; MASTROPAOLO et al., 2011). Por meio de análises moleculares para a estruturação genética do *A. cajennense*, avaliando genes mitocondriais, sugere-se que na maioria dos países da América do Sul, e alguns da América Central, o carrapato *A. cajennense* seria um complexo possuindo mais de uma espécie, visto a presença de vários haplótipos distintos nas diferentes regiões em que este carrapato foi coletado (BEATI et al., 2012; BURKMAN, 2009).

Em 2014, este carrapato foi redescrito, com a validação taxonômica de duas espécies que tinham sido levados à sinonímia no passado, *Amblyomma mixtum* e *Amblyomma sculptum*, e a descrição e definição de três novas espécies, *Amblyomma tonelliae* n. sp., *Amblyomma interandinum* n. sp. e *Amblyomma patinoi* n. sp. (NAVA et al., 2014) (Figura 3).

Figura 3 Distribuição geográfica, filogeográfica e especiação do carrapato do complexo *A. cajennense* na América e nos biomas Brasileiros. Destaque (a): Formação e distribuição de quatro clados entre populações de carrapatos do complexo *A. cajennense*. Destaque (b): Distribuição geográfica do complexo *A. cajennense* após redefinição na América. Destaque (c): Distribuição geográfica do complexo *A. cajennense* no Brasil correlacionada aos fatores ambientais florestais e do clima.

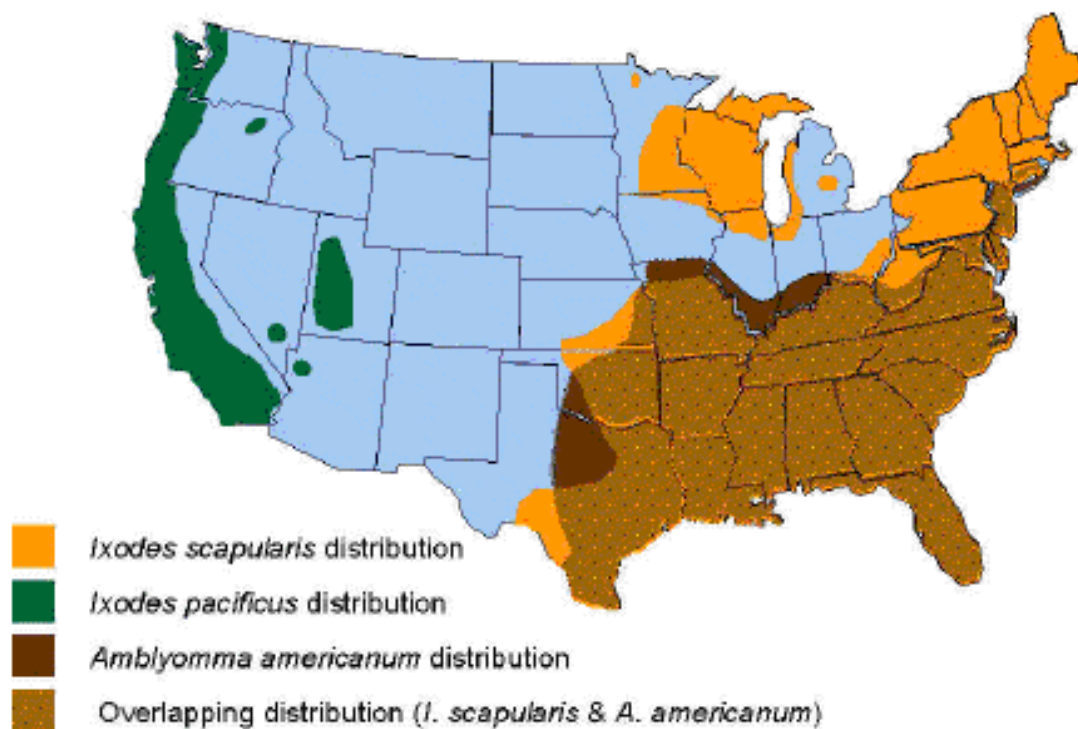


Fonte: Destaque (a), Barros-Batestti et al. (2006); Destaque (b), Mazioli (2012); Destaque (c) Nava et al. (2014) e os Destaques (d) e (e) Martins et al. (2015).

Na Figura 3, verifica-se, que populações de *A. cajennense* distribuídas na América (BARROS-BATTESTI; ARZUA; BECHARA, 2006) estão tendo fluxo gênico entre elas, expandindo-se demograficamente (NAVA et al., 2014), com a diversidade genética encontrada corroborando este fluxo ao apresentar baixos valores de diversidade haplotípica (0,0204 e 0,1393) e diversidade nucleotídica (0,0008 e 0,0030) para os clados formados: Clado 1 e Clado 2, sugerindo com a presença de *singletons* (haplótipos únicos) o surgimento de possíveis novas populações, em simpatria e com a alopatria podendo estar se iniciando (MAZIOLI, 2011). No país, o complexo do *A. cajennense* é representado por apenas duas espécies, *A. cajennense* (sensu stricto) e o *Amblyomma sculptum* (Complexo *A. cajennense*), embora essas espécies tenham áreas de distribuição distintas sendo encontradas em simpatria em algumas regiões de transição dos biomas brasileiros (MARTINS et al., 2016). A suspeita da ocorrência de um complexo aumenta quando estudos filogenéticos usando marcadores distintos (MARRELLI et al., 2007) e genomas mitocondriais completos (BURGER et al., 2012) indicam que o gênero *Amblyomma* ainda não está bem definido filogeneticamente, de forma que o uso de mais marcadores moleculares, tais como região controle, citocromo oxidase, 18S poderão auxiliar no melhor entendimento da história evolutiva, bem como na estruturação genética das espécies de carrapatos estudadas (LABRUNA et al., 2009, 2011; MAZIOLI, 2011).

Os carrapatos do gênero *Ixodes* e *Amblyomma* são infectados por uma rica microbiota de *Rickettsia*. Na América do Norte, o carrapato *I. scapularis* é um importante vetor de patógenos bacterianos, não sendo associado à transmissão de agentes patogênicos do gênero *Rickettsia* (MORENO et al., 2006). No entanto, as espécies *R. endosymbiontic* *I. scapularis*, conhecida como REIS e *R. bucheneri* infectam o *I. scapularis*, dando a este ácaro o título de real hospedeiro com a passagem transovariana por esta espécie de *Rickettsia* (KURTTI et al., 2015). A espécie *R. amblyommatis*, que possuía o status de “*Candidatus Rickettsia amblyommi*” (KARPATHY et al., 2016), atualmente tem sido relatada em carrapatos da espécie *I. scapularis* (MAYS et al., 2014; TROUT FRYXELL et al., 2015a) (Figura 4).

Figura 4 Mapa geográfico dos EUA demonstrando a distribuição geográfica para as espécies de carrapatos, em especial, *I. scapularis* e *Amblyomma americanum*. Destaque para a área com forte sobreposição de incidência para ambas, as espécies. Fonte: CDC. (2015)



Fonte: CDC. (2015)

Na Figura 4, verifica-se que a região geográfica sugere o encontro da espécie de bactéria *R. amblyommatis* com os carrapatos da espécie *I. scapularis* podendo estar sendo adquirida com a ingestão de sangue durante o ato de ingurgitamento e hematofagia deste ácaro sob espécies de hospedeiros vertebrados, que habitam as mesmas áreas de distribuição destas populações. Indivíduos de *I. scapularis*, coexistem próximos às áreas dominantes por espécies de *Amblyomma*, especialmente o *Amblyomma americanum*, além de outros carrapatos do complexo do *A. cajennense* que possuem a infecção relatada por *R. amblyommatis*.

2.2 Genoma funcional

2.2.1 Carrapato

Diversos estudos têm caracterizado a transcrição dos genes em ambos sexos de carrapatos e seus estágios de desenvolvimento, sendo averiguados singânglio, órgãos reprodutores ovários e túbulos seminíferos, glândulas salivares e intestino médio, considerando diferentes

tratamentos, incluindo condições de infecção ou não infecção por espécies de *Rickettsia*. Em fêmeas de *I. scapularis* verifica-se a presença de 15 genes putativos semelhantes aos encontrados em *Dermacentor variabilis* (EGEKWU et al., 2014). Na caracterização desta expressão em *D. variabilis*, sob condições de jejum e alimentados parcialmente, utilizando carrapatos fêmeas virgens e acasaladas, obteve-se um total de 14.881 sequências *contigs*, com um tamanho médio de 229 pb (BISSINGER et al., 2011). Em ambos os estudos relata-se a presença de hormônios de eclosão, os quais também são encontrados em singânglios de insetos durante a regulação do comportamento e desenvolvimento da cutícula de quitina associada com a muda, sugerindo que este sistema possa ser utilizado em carrapatos para regular a hematofagia durante a expansão da cutícula e o desenvolvimento relacionado com a reprodução do sexo feminino durante o ingurgitamento sob hospedeiro vertebrado (BISSINGER et al., 2011; EGEKWU et al., 2014).

No transcriptoma da biologia reprodutiva de carrapatos machos alimentados da espécie *D. variabilis*, revelaram-se 12.804 *contigs*, com 3.866 transcritos em 73 categorias diferentes, incluindo espermatogênese (SONENSHINE et al., 2011). Avaliando-se os perfis transcriptômicos de glândulas salivares de *I. scapularis*, para amostras coletadas em tempos diferentes de 18 e 24 h, após o início da hematofagia, em hospedeiro vertebrado, verifica-se, pelo menos, 20 genes, sendo este achado duas vezes mais abundante do que o esperado (RIBEIRO et al., 2006). A expressão dependente do tempo foi recentemente confirmada para *I. ricinus* na análise da transcrição de proteínas metaloproteases contendo o domínio Kunitz e lipocalinas, sugerindo que os carrapatos mais alimentados tiveram 1.447 genes expressos nas glândulas salivares, sendo que 1.135 codificavam proteínas secretadas (KOTSYFAKIS et al., 2015). Estes autores revelaram que o tempo da alimentação, elevou em pelo menos 10 vezes, a expressão destes genes, sendo observados 8.300 genes. Em outro estudo, a discrepância revelada por análises de sistemas biológicos entre os dados de transcriptômica e proteômica das glândulas salivares e intestino médio de *I. ricinus* fêmeas e ninfas, permitiu predizer 1.510 genes expressos, havendo 373 proteínas mais abundantes nas glândulas salivares do que no intestino médio, entretanto, com 217 proteínas reguladas positivamente de forma significativa no intestino médio, contra apenas 93 proteínas nas glândulas salivares (SCHWARZ et al., 2014).

A avaliação individualizada dos transcriptomas do intestino médio e do ovário de carrapato da espécie *A. sculptum* (Complexo *A. cajennense*), no tocante à infecção e não

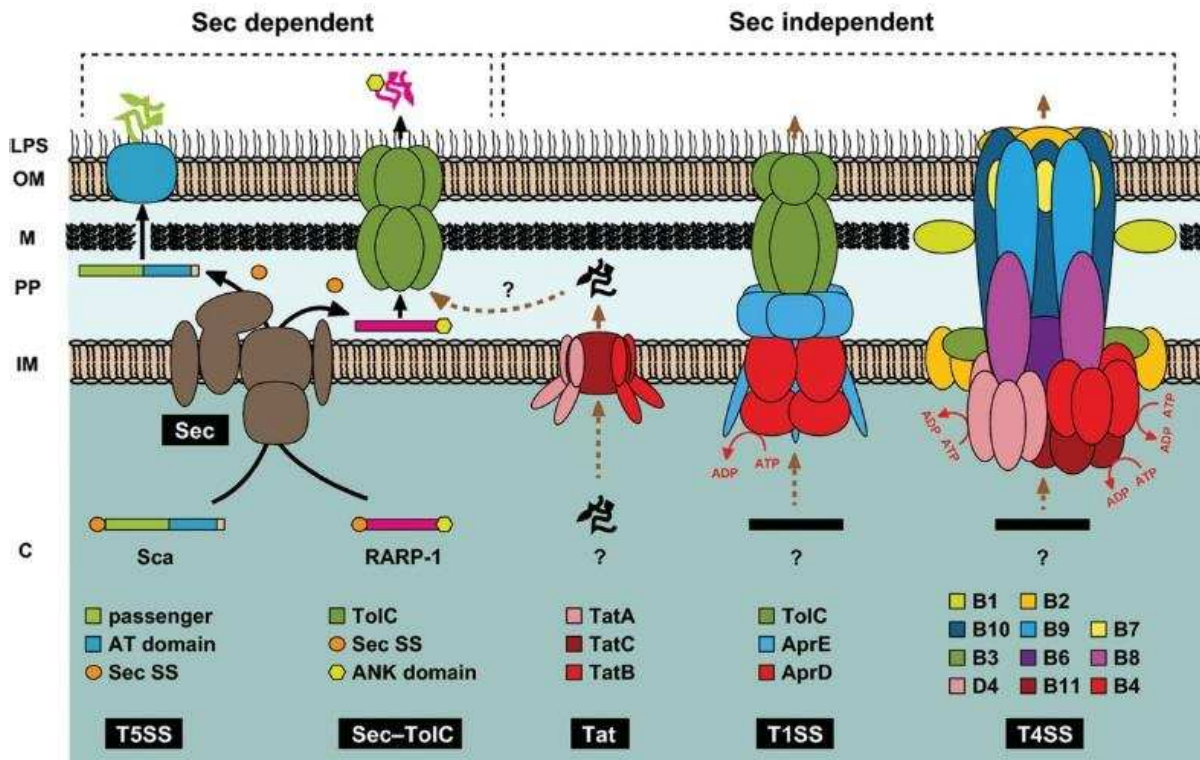
infecção por *R. amblyommatis* (infectados vs não infectados), revelou um padrão oposto entre os dois órgãos, com quase todos os processos celulares no intestino médio altamente expressos em resposta à infecção, enquanto a maioria dos processos celulares em ovário foi regulada negativamente em resposta à infecção por esta espécie de bactéria, corroborando outros estudos *in vitro* utilizando espécimes diferentes de carrapatos infectados com diferentes *Rickettsia* spp., os quais observaram *in vivo* com foco após postura de ovos de carrapatos infectados, apresentado queda das funções fisiológicas de gêmeas ingurgitadas e queda no rendimento da oviposição (MOREIRA, 2016).

Para este estudo, temos que a expressão gênica em carrapatos da espécie *A. sculptum* (Complexo *A. cajennense*) foi significativamente alterada pela expressão de genes constitutivos em resposta a diferentes condições fisiológicas e da infecção por *R. amblyommatis* (MOREIRA, 2016). Sobretudo, estes dados revelaram aspectos desta expressão relacionados com o controle da expressão gênica, síntese proteica, vias metabólicas centrais e componentes da maquinaria do citoesqueleto, reforçando a hipótese sobre a manipulação da maquinaria de expressão gênica do carrapato por esta bactéria.

2.2.2 *Rickettsia*

O genoma de *Rickettsia* pode codificar vários sistemas de secreção que são homólogos com as vias de secreção de proteínas caracterizadas em outras bactérias. Todos estes sistemas são essenciais para orquestrar a vida dentro de células eucarióticas, sendo os mais conhecidos os sistemas dependentes Sec-TolC e o T5SS (GILLESPIE et al., 2014) (Figura 5).

Figura 5 Secretomas de bactérias encontrados em *Rickettsia* com destaque para os sistemas de secreção Sec-dependente contendo o sistema de secreção do tipo V (T5SS) e o Sec-TolC, além dos sistemas de secreção Sec-independentes contendo o sistema de duas argininas (Tat), o sistema de secreção do tipo I (T1SS) e do sistema de secreção do tipo IV (T4SS) sendo este último, o menos conhecido para o gênero.



Fonte: Gillespie et al. (2015).

Na Figura 5, verifica-se que o sistema autotransporte (Ats), constituído por antígenos que dominam a superfície da célula de *Rickettsia*, possui uma gama de interações com moléculas da célula hospedeira (GILLESPIE et al., 2014). Segundo estes autores, a interação do Sec-TolC com outras vias de secreção, bem como as funções dos sistemas de secreção Sec-independentes e o sistema de duas argininas *twin-arginine translocation* (Tat), continuam a ser mal compreendidos para os aspectos da biologia de *Rickettsia*, em contraste com os sistemas de secreção de proteínas e vias descritas, com uma substancial falta de conservação vista dentro de muitas famílias de proteínas secretoras, onde as famílias: SCA5, SCA4, ADR1, ADR2, TlyC, PLD, e RARP-1 carecem de genes em um ou mais genomas deste gênero de bactéria.

A maioria destes genes apresentam sequências variáveis codificadoras de proteínas ricas em motivos repetidos e outras com regiões hipervariáveis, sugestivos de evolução rápida e

provavelmente impulsionada por pressões seletivas de acolhimento do hospedeiro (GILLESPIE et al., 2014). Segundo estes autores, apesar de serem codificadas em todos os genomas de bactérias do gênero *Rickettsia*, as famílias das proteínas SCA5, SCA4 e RARP-1, são hipervariáveis em comprimento e composição de sequências, além de proteínas PLD que também é hipervariável na composição, enquanto ADR1 e ADR2 possuem mínima variação dentro da sequência com porções expostas à superfície das suas estruturas em formato de beta-barril, deixando a TlyC como a mais conservada proteína de secreção (77% de identidade em *R. typhi* e *R. bellii*).

2.3 Interatoma de proteínas

As proteínas em sua forma nativa não fornecem uma compreensão da sua funcionalidade e o comportamento celular, exigindo correlacionar suas interações, gerando uma rede tridimensional de alta complexidade estrutural, ao adotar a análise topológica com métricas de conectividade e centralização que permite identificar proteínas essenciais na rede proteica predita (FLÓREZ et al., 2010). Nestes estudos, visam-se detectar aglomerados enriquecidos de ontologia gênica (GO) (CAMON, 2004), para determinar as possíveis vias de alvos, inferindo a respeito dos processos biológicos realizados por proteínas desconhecidas e putativas, sendo um ponto de partida para a validação experimental por ensaios *in vitro* e posterior descoberta de novos alvos (FLÓREZ et al., 2010).

As métricas desta análise para inferir a interação de redes de proteínas utilizam-se de diversos bancos de dados do interatoma, entre eles, o banco de dados PSIMAP e o iPfam, os quais buscam evidenciar potenciais interações baseadas na identificação de domínios estruturais do SCOP (HUBBARD et al., 1997) e de sequência do Pfam (FINN, 2006; SONNHAMMER, 1998) que interagem entre si, em estruturas tridimensionais de proteínas depositadas no PDB (BERNSTEIN et al., 1978; KOURANOV, 2006a). Segundo estes autores, ambos os bancos de dados permitem observarmos as interações de proteína ao nível de família-família e de proteína-proteína individuais, abrangendo informações desta interação de ambos níveis, os níveis de fusão dos genes em interação e a interação física dentro de complexos ou proteínas de vários domínios.

O banco PSIMAP possui o servidor web PSIBase (GONG et al., 2005), que contém várias informações para a interação entre domínio-domínio, proteína-proteína, mapa da

interação e os níveis de proteína, superfamília e famílias, interface de interação entre proteínas e previsão estrutural de domínios averiguando homologia à partir de sequências depositadas no PDB (BERNSTEIN et al., 1978). Ainda, o PSIBase utiliza de dois bancos de dados e dois algoritmos para atribuir o domínio estrutural, sendo o SCOP (HUBBARD et al., 1997), com sua base de dados (TEICHMANN et al., 2000) e o PSI-BLAST (ALTSCHUL et al., 1997), fazendo uso de modelo oculto de Markov (EDDY, 1996).

O banco iPfam difere da maioria dos outros bancos de interatoma ao incluir os motivos conservados e as regiões menos conservadas, como inserções e deleções, fazendo o uso de modelo oculto de Markov (SONNHAMMER, 1998), permitindo inserções e supressões “deleções” e, ainda, facilitando a anotação de domínios completos, apresentando maior significância e sensibilidade (EDDY, 1996; KROGH et al., 1994). Segundo estes autores, cada família de domínio de proteína possui três importantes arquivos, sendo o primeiro denominado de sementes, tratando-se de um alinhamento múltiplo de um conjunto representativo das sequências, servindo de base para originarmos aos outros dois arquivos. O banco PEIMAP é um banco de dados que incorpora informação de interação de pares de proteína de seis bancos de dados experimentais: *Data-base of Interacting Proteins* (DIP) (SALWINSKI, 2004; XENARIOS, 2001) *Biomolecular Interaction Network Data-base* (BIND) (BADER, 2003), *Molecular Interaction Database* (IntAct) (HERMJAKOB, 2004; KERRIEN et al., 2012), *Molecular INTeraction database* (MINT) (ZANZONI et al., 2002), *A High Performance RDF Database* (HPRD) e *A General Repository for Interaction Datasets* (BioGrid) (STARK, 2006).

Algumas abordagens baseadas em coevolução podem ser consideradas como parte dos métodos para detectar interações e as relações funcionais entre proteínas usando informação genômica, conhecido em conjunto como métodos de “base de contexto” (HARRINGTON; JENSEN; BORK, 2008; WASS; DAVID; STERNBERG, 2011), proporcionando uma alternativa ortogonal para a previsão mais tradicional da função baseado em homologia para semelhança de árvores genealógicas (MERING, 2003; SZKLARCZYK et al., 2011). Segundo estes autores, tem se firmado o uso do banco de dados STRING, o qual faz uso de uma única pontuação para a estrutura da árvore com base em valores de referência para os diferentes tipos de associações contra um conjunto de referência comum integrado em um único valor de confiança por previsão.

As interações proteína-proteína não estão limitadas por aspectos físicos obrigatórios, podendo também interagir indiretamente ao compartilharem um mesmo substrato numa via

metabólica, regulando uns aos outros transcricionalmente, ou através da participação de vias de proteínas-multigenes (SZKLARCZYK et al., 2011). Para prever associações funcionais os métodos de inferência são utilizados para esta análise de contexto, as quais baseiam-se no fato de que proteínas funcionalmente associadas são codificadas por genes que partilham semelhantes pressões de seleção e precisam ser mantidos juntos e regulados em conjunto, de tal modo que as proteínas codificadas podem interagir ao mesmo tempo e no mesmo lugar na célula, codificando um combinado polipeptídeo (GALPERIN; KOONIN, 2000; HUYNEN et al., 2000; MARCOTTE, 1999).

Uma importante diferença é que as proteínas fusionadas não são reconhecidas por homologia, mas sim por ortologia das partes fundidas a outra (SNEL; BORK; HUYNEN, 2000; YANAI; DERTI; DELISI, 2001). Neste sentido, a co-ocorrência de proteínas em mapas metabólicos de banco de dados permite presumir-se que estas estão interagindo funcionalmente, diferentemente daquelas que ocorrem em outros diferentes mapas (SZKLARCZYK et al., 2011). Segundo estes autores, a simples contagem dos eventos é insuficiente, sendo necessário pontuar, normalizar e possuir um número de espécies que possuem em seus genomas estes genes envolvidos. Para obter informações sobre genomas, os genes e proteínas codificadas, o banco STRING depende dos proteomas anotados e mantidos pelo Swiss-Prot (BAIROCH, 2000), sendo essencial para as previsões e para a atribuição de equivalência funcional dos genes em todos os genomas, informação esta derivada do banco de dados de proteínas ortólogas *Clusters of Orthologous Groups of Proteins* (COG) (TATUSOV, 2001).

Uma outra abordagem considera a formação de complexos para interações entre proteína-proteína (PPis) e a identificação do complexo mais provável entre milhares de outros (ranqueamento de sequências) (MISHRA, 2012). Estas análises utilizam-se de métodos de acoplagem da interface proteína-proteína, com a predição de resíduos de aminoácidos susceptíveis na interface e a medição da área de superfície oculta dos complexos. Aqui, considera-se as distinções entre os complexos idênticos (homo complexo) e complexos não idênticos (hetero complexo) (JONES; THORNTON, 1996), além de complexos obrigatórios e não-obrigatórios (MISHRA, 2012). De certa forma, os complexos não-obrigatórios são aqueles hetero complexos, em que os parceiros que interagem não são complexos colocalizados inicialmente, sendo transitórios entre os complexos permanentes, dependendo diretamente do tempo de vida (NOOREN, 2003).

Segundo este autor, os complexos distintos entre proteína-proteína diferem em termos de complementariedades de superfície, funções estéricas, eletrostática, hidrofóbica, forças de ligação de hidrogênio, área de superfície acessível, propensão de resíduo e planaridade. Desta forma, o estudo de validação *docking* utiliza-se dos aplicativos *Optimal Area* (ODA) e ZDOCK2.3 para identificar corretamente a interface de ligação de alguns complexos em estruturas experimentais (CHEN; LI; WENG, 2003). Aqui, utiliza-se de aplicativos computacionais, tais como HHsearch/HHpred (SODING, 2005) em SWISS-MODEL (ARNOLD et al., 2006) e ClusPro (COMEAU et al., 2004) para o ranqueamento de proteína-proteína não ligadas do MetaPPISP para a predição de interface entre resíduo e do aplicativo Racer superfície para o cálculo de área de superfície ocultas (MISHRA, 2012).

2.4 Coevolução de proteínas

A coevolução refere-se às alterações coordenadas que ocorrem em pares de organismos ou biomoléculas, mantendo ou refinando as interações funcionais entre os mesmos, ocorrendo mudanças na sequência, entre genes e os seus produtos de proteína, sendo essencial no desempenho fisiológico e no sucesso reprodutivo, indicando assim, um importante e generalizado determinante da aptidão (DE JUAN; PAZOS; VALENCIA, 2013). Segundo estes autores, embora a coevolução ocorra entre várias biomoléculas, a maioria das ferramentas recentes de bioinformática se concentram nas proteínas, sendo amplamente individualizadas ao teor de resíduos de aminoácidos e aqueles, que procuram ao nível de proteínas, baseado nos princípios da filogenia molecular.

As ferramentas em nível de resíduos consideram a independência na variação de grupos de amino ácidos numa sequência, utilizando do alinhamento múltiplo de sequência para uma determinada família de proteínas homólogas na busca por mutações correlacionadas (FITCH; MARKOWITZ, 1970), visando identificar locais de interação interproteínas (PAZOS et al., 1997; TRESS et al., 2005; YEANG; HAUSSLER, 2007), enquanto outros métodos procuram grupos maiores de resíduos, especificamente coconservados dentro de uma subfamília de proteínas em particular, podendo identificar resíduos que definem as propriedades funcionais e especificidade de um substrato ligante a uma determinada enzima (DE JUAN; PAZOS; VALENCIA, 2013).

As mutações correlacionadas sugerem mudanças compensatórias, ocorrendo entre os resíduos em diferentes posições, podendo estes estarem próximos em contato direto, agindo juntos em catalises ou, até mesmo, em locais de ligação para manter a estabilidade da proteína, sua função e enovelamento (GÖBEL et al., 1994; NEHER, 1994). Segundo estes autores, o estudo da coevolução entre resíduos de diferentes proteínas sugere o uso destes como preditores das superfícies que interagem nas interfaces de proteínas, indicando pares de posições de aminoácidos interdependentes ou padrões semelhantes de substituição, os quais podem ser construídos com o uso de uma matriz de substituição de aminoácidos précalculada, tendo sua semelhança avaliada por uma correlação linear.

As ferramentas em nível de proteínas permitem a construção de árvores filogenéticas para muitas famílias de proteínas (GOH et al., 2000; PAZOS; VALENCIA, 2001). A coevolução entre espécies, tais como a interação parasita-hospedeiro, predador-presa e simbiote-anfitriões, em muitos casos apresenta árvores filogenéticas semelhantes para a coevolução destas relações (DE JUAN; PAZOS; VALENCIA, 2013). Da mesma forma, a coevolução molecular entre proteínas, que são frutos de fenômenos físicos ou devido às interações funcionais, frequentemente resulta em correspondentes semelhanças para árvores genealógicas, permitindo através da similaridade, identificar parceiros de interação para uma determinada proteína, tais como seus pares de receptores ligantes (FRYXELL, 1996; PAZOS; VALENCIA, 2001).

Abordagens baseadas em coevolução como parte dos métodos para detectar interações e as relações funcionais entre proteínas usando informação genômica, são conhecidas em conjunto como métodos de "base de contexto" (HARRINGTON; JENSEN; BORK, 2008; WASS; DAVID; STERNBERG, 2011), como discutido no tópico anterior de interatoma. Baseado em homologia para semelhança de árvores genealógica (MERING, 2003), esta abordagem, chamada de MirrorTree (KANN et al., 2009), avalia a correlação física entre árvores semelhantes e a interação funcional, prevendo o potencial das interações proteína-proteína em uma escala genômica (PAZOS; VALENCIA, 2001). Outras abordagens derivadas preveem pares específicos de duas proteínas parálogas que interagem numa mesma família de proteína, considerando famílias de proteínas e correspondentes entre diferentes membros das duas famílias (DE JUAN; PAZOS; VALENCIA, 2013).

Em adição, as árvores genealógicas podem ser usadas para identificar os domínios que interagem em duas proteínas, caso estes domínios exibam sinais fortes de coevolução, não

interagindo com outros domínios dentro da mesma proteína (JOTHI et al., 2006), de forma que as árvores para as interfaces entre proteínas mostram mais intensamente a coevolução do que o resto das proteínas (KANN et al., 2009). Ambos os autores sugerem que, em muitos casos, a coevolução entre proteínas que interagem é um fenômeno local, sendo circunscrita aos resíduos em regiões particulares, podendo identificar a interação entre resíduos dentro de proteínas (DE JUAN; PAZOS; VALENCIA, 2013).

Proteínas que realizam funções comuns presentes ou ausentes nos mesmos organismos podem ser consideradas um caso extremo de coevolução, sendo este fenômeno explorado por alguns métodos para prever a interação entre as famílias de proteínas funcionalmente relacionadas (PELLEGRINI et al., 1999). Quanto aos métodos MirrorTree (KANN et al., 2009), o processo de especiação subjacente introduz um viés importante, exigindo a imposição de fortes restrições sobre o conjunto de genes estudados, devido ao fato de que muitas sequências de proteínas que desempenham uma função comum podem estar simultaneamente em bactérias e archaea, mas ausentes em eucariontes (DE JUAN; PAZOS; VALENCIA, 2013).

Segundo estes autores, no entanto, a coevolução de intra e interproteínas são dois fenômenos relacionados que envolvem forças semelhantes de interação física e restrições evolutivas, podendo ser explicada por um conjunto de pares de resíduos interproteína que coevoluíram. De fato, vários métodos para a detecção de resíduos que coevoluíram também têm sido utilizados para detectar resíduos das interfaces de interações interproteína (PAZOS; VALENCIA, 2001; TRESS et al., 2005; YEANG; HAUSSLER, 2007). Embora historicamente separadas, métodos para o estudo da coevolução inter e intraproteína possuem aplicações intimamente relacionadas e compartilham conceitos metodológicos semelhantes (DE JUAN; PAZOS; VALENCIA, 2013).

Neste sentido, a utilização do conjunto de métodos de "base de contexto" (HARRINGTON; JENSEN; BORK, 2008; MERING, 2003; WASS; DAVID; STERNBERG, 2011), aqui proposto como estratégia para o estudo tanto dos aspectos do interatoma, como do estudo coevolutivo de proteínas da interface carrapato vs *Rickettsia*, são os mais sugestivos para o norteamento destas análises para o projeto proposto visando estudar dentro da rede de interação os mapas de ontologia gênica proposto para a verificação e anotação da funcionalidade destes pares que interagem podendo conter respostas da coevolução entre estes organismos.

3 METODOLOGIA

Para executar este projeto foi utilizado o Cluster hospedado no Departamento de Tecnologia da Informação da Universidade Federal de Viçosa UFV, adotando a plataforma do supercomputador Júpiter, podendo ser acessado no *link*: (http://www.novoscursos.ufv.br/diretorias/ufv/dti/www/?page_id=744). Para utilizar o cluster, foi necessário transformar os bancos de dados de genoma, proteoma e transcriptoma em bancos de dados biológicos internos parecidos com os do NCBI com o uso de documento formatdb que consiste num programa para esta formatação, sendo obrigatório construir os *scripts* no formato pbs para a submissão dos trabalhos. O uso destes *scripts* sendo obrigatório, visto no cluster não ser permitido rodar análises de trabalhos na linha de comando. Neste sentido, foi preciso desenvolver diferentes *scripts* em linguagem pbs, perl e bin bash para submetermos a maioria das análises e realizar este trabalho. Este uso obrigatório consiste em desenvolver o ambiente de trabalho no Cluster. Através do aplicativo WinSCP pudemos transferir e conferir os nossos dados diretamente na plataforma do Júpiter e com o uso do programa Putty, orquestramos a execução deste projeto, utilizando pacotes e serviços inseridos previamente no Cluster, com o uso de linguagens computacionais e comandos de execução Linux.

3.1 Construção do Interatoma

Como modelo para a realização deste estudo e construção do interatoma utilizamos de 1821 proteínas do genoma ASM96467v1 ID 1359164 da espécie de bactéria *Rickettsia amblyommatis* estirpe *Ac/Pa* (KARPATHY et al., 2016), baixado no banco de dados *Ensembl* bactérias (http://bacteria.ensembl.org/Rickettsia_amblyommatis_str_ac_pa/Info/Index) e de 21486 proteínas do genoma IscaW1 ID 6945 do *Ixodes scapularis* (GULIA-NUSS et al., 2016), baixado no banco de dados *Ensembl* metazoa (http://metazoa.ensembl.org/Ixodes_scapularis/Info/Index). Com depósito no GenBank pelo número de acesso ABB010000000, o projeto do sequenciamento do genoma do carrapato da espécie *I. scapularis* colônia Wikel, número PRJNA16232 pôde ser acessado (<https://www.ncbi.nlm.nih.gov/nucleotide/ABB010000000>).

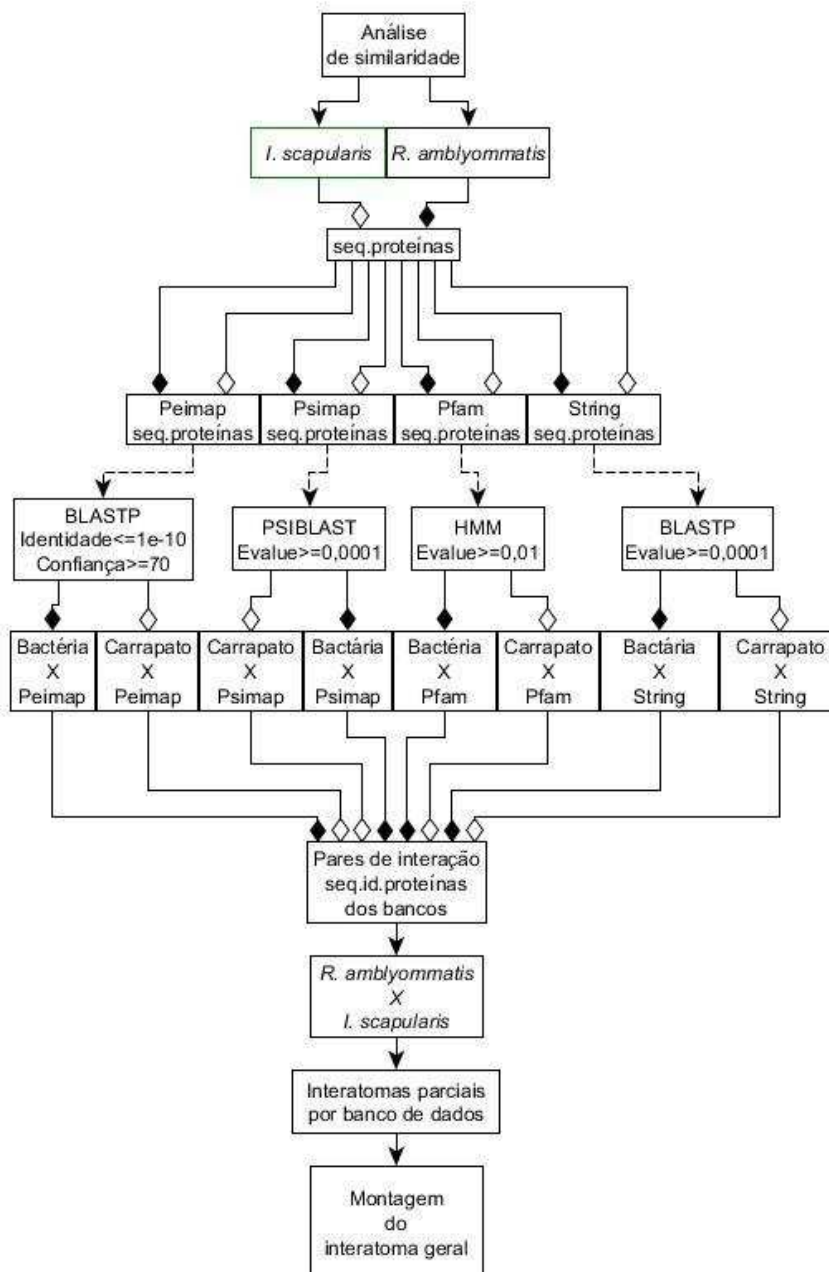
Sob a abordagem do interatoma, o estudo da interação foi realizado utilizando de dados de interação entre proteínas e famílias de domínios depositados nos bancos de dados (*Protein Experimental Interactome MAP*) PEIMAP (FLÓREZ et al., 2010; KIM et al., 2008); (*Search*

Tool for the Retrieval of Interacting Genes/Proteins) STRING (<https://string-db.org/>) (MERING, 2003); (*Protein Structural Interactome MAP*) PSIMAP (<https://sourceforge.net/projects/psisoft/>) (PARK et al., 2005), e do (*Protein Families database*) Pfam (<https://pfam.xfam.org/>) (FINN; MARSHALL; BATEMAN, 2005), apresentando o número de interações entre proteínas ultrapassando 2 bilhões, 1 milhão, 500 mil e 13 mil com o número de pares de domínios ultrapassando 400 mil, 6 milhões, 60 mil e 3 mil conforme respectivos bancos.

3.1.1 Análise de similaridade

Para a construção do interatoma, utilizamos das sequências de aminoácidos das proteínas provenientes do genoma da bactéria da espécie *R. amblyommatis* e do carrapato da espécie *I. scapularis*. Inicialmente, realizamos a análise de similaridade com estes repertórios de proteínas, separadamente para cada organismo contra as sequências de proteínas depositadas para os quatro bancos de dados do interatoma (Figura 6).

Figura 6 Fluxograma para as análises de similaridade dos dados de sequência de proteínas dos genomas provenientes da bactéria da espécie *R. amblyommatis* ♦ e do carrapato da espécie *I. scapularis* ◇, fazendo o uso de multialinhamentos de seqüências e de dados de domínios de interação de proteínas em banco de dados do interatoma. Destaque para os alinhamentos e parâmetros avaliados em cada banco dando montagem do interatoma parcial utilizando de cada banco.



Fonte: Autor deste trabalho.

Na Figura 6, verifica-se os passos e o esboço realizados para as análises de similaridade. Estas foram a base inicial para a montagem do interatoma parcial e do geral, de forma que cada conjunto de proteínas, tanto as provenientes do genoma de *R. amblyommatis* como do *I.*

scapularis foram submetidas à alinhamentos de sequências contra as sequências de proteínas de dados de bancos que constituem o interatoma.

Peimap x Similaridade

Para o alinhamento local contra a base de dados de proteínas depositadas no PEIMAP, utilizou-se do BLASTP (ALTSCHUL et al., 1997). As interações de proteínas significativas foram consideradas com alinhamentos com um mínimo de 40% de identidade e 70% de cobertura (FLÓREZ et al., 2010). O alinhamento local com o uso do BLASTP (*Basic Local Alignment Search Tool*) para sequências de proteínas forneceu a pontuação máxima entre qualquer par de subsequências de proteínas depositadas no banco de dados do Peimap em relação as sequências comparadas, no caso, contra as sequências de proteínas do carrapato e da bactéria desejáveis para se identificar trechos altamente conservados entre estes dois genomas.

Por meio do uso do BLASTP entre sequências de proteínas, obteve-se o arquivo de saída contendo o número de acesso para a sequência alinhada, uma descrição breve da sequência, o *score* máximo resultante de um único HSPs (*High-scoring Segment Pairs*), o *score* total resultante da soma de HSPs, a *Query coverage* com a porcentagem da sequência submetida ao programa coberta pelo alinhamento, o *e-value* que é o parâmetro de confiança e a máxima identidade obtida por um HSP. O *expected value (e-value)* é o parâmetro de confiança do alinhamento. Onde $E = \text{expect value}$, m e n tamanho das sequências alinhadas S' *score* normalizado. O *P-value* (s) é a probabilidade de obter um *score* tão bom ou melhor que (s) puramente por chance em um banco de dados aleatório, do mesmo tamanho e com a mesma composição de bases, enquanto o *E-value* (s) é o número de *hits* com *score* tão bom ou melhor que puramente por chance em um banco de dados aleatório, do mesmo tamanho e com a mesma composição de bases, sendo um número real não negativo.

No entanto, a interpretação de um resultado nem sempre é trivial, muitas vezes depende do que o usuário busca como a definição de função, motivos conservados e tamanho da sequência submetida, sendo que sequências muito pequenas nunca obterão um *e-value* muito baixo mesmo obtendo um alinhamento perfeito. O fato de termos um *e-value* significativo não implica que necessariamente podemos postular a função de uma proteína, pois em alguns casos isso reflete meramente a conservação de algum motivo ou domínio que podem estar presentes em proteínas com funções diversas.

Este alinhamento local é útil para encontrar um gene em um genoma, identificar éxons, identificar domínios proteicos ou identificar possíveis homólogos em um banco de dados. Aqui, o BLAST é um algoritmo heurístico e útil para alinhamento entre sequências de tamanhos diferentes, para sequências com apenas alguns trechos conservados, para a consulta de sequências em banco de dados biológicos. A pontuação e os alinhamentos, demonstram a similaridade como um grau de semelhança de sequências permitindo inferência de homologia sob, por exemplo, um gene com ancestral comum (ALTSCHUL et al., 1997).

Com a finalidade de avaliarmos a similaridade entre as sequências de proteínas destes organismos versus as sequências de proteínas do banco de dados Peimap, adotando do BLASTP desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=6,mem=6gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python
blastall -p blastp -i (arquivo 1a) -d (arquivo 2) -a 6 -F T -m8 -o (arquivo de saída 3a)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2) -a 6 -F T -m8 -o (arquivo de saída 3b)
#Fim do script
```

Neste *script* pbs, os arquivos 1a e 1b consistem nas listas de sequências de aminoácidos das proteínas no formato Fasta para a bactéria e o carrapato, respectivamente e submetidos a análise de similaridade adotando o BLASTP contra o arquivo 2 de sequências de aminoácidos de proteínas depositadas no banco de dados do Peimap no formato Fasta para gerarem os arquivos de saída 3a e 3b, contendo os resultados do alinhamento para ambos os organismos (Dados não publicados). Para selecionarmos as sequências de proteínas de *R. amblyommatidis* e do *I. scapularis* com maior similaridade das sequências dos domínios do banco de dados do Peimap, adotando o nível de confiança ≥ 70 e a identidade $\leq 1e-10$ do BLASTP desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
```

```

#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE="arquivo 1"
OUTPUT_FILE="${INPUT_FILE}-2col"
perl script.peimap.id.pl "$INPUT_FILE" > "$OUTPUT_FILE"
cat "$OUTPUT_FILE" | grep PEI_ | sort | uniq > "${OUTPUT_FILE}-no-repeat"
echo "Removendo arquivo '${OUTPUT_FILE}'"
rm -v ${OUTPUT_FILE}
#Fim do script

```

```

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
my $confianca;
my $identidade;
my $idpeimap;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @linha = split (" ", $linhas[$i]);
#pega a primeira coluna
$id = $linha[0];
#pega a segunda coluna
$idpeimap = $linha[1];
#pega a quarta coluna
$confianca = $linha[2];
#pega a decima primeira coluna
$identidade = $linha[10];
if ($confianca >= 70 && $identidade <= 1e-10){
print "$linha[0]\t$linha[1]\n";
}
}
#Fim do script

```

No *script* pbs, o arquivo 1 consiste nas entradas para os arquivos 3a e 3b contendo os alinhamentos para as sequências de aminoácidos das proteínas dos organismos e seus similares, tratando-se dos domínios de proteínas do banco de dados do Peimap. O *script* perl, chamado dentro do *script* pbs, foi programado para se obter os IDs das sequências de proteínas e os seus similares conforme os valores de confiança e identidade adotados para o BLASTP. Além disso, com base nos pares de interação do banco do Peimap, somente as proteínas similares contendo nos IDs o código identificador PEI_, foram selecionadas, porque estas proteínas possuem pares de interação com validação comprovada por métodos bioquímicos e moleculares.

Psimap x Similaridade

Para o alinhamento contra a base de dados SCOP (HUBBARD et al., 1997), utilizada pelo PSIMAP, fizemos o uso do PSI-BLAST (ALTSCHUL et al., 1997), considerando *E-value* de 1×10^{-4} (FLÓREZ et al., 2010; KIM et al., 2008), para prever interações de proteínas. O PSIMAP é um banco de dados que infere interações baseadas na identificação de domínios estruturais do (LO CONTE, 2000). O PSI-BLAST, inicia-se como um BLASTP e atua na recuperação de sequências de proteínas contendo similaridades com a sequência de proteína inserida no alinhamento. Este algoritmo faz o uso das sequências que obtiveram um *score* acima de um certo limite para criar uma matriz (*Position-Specific Score Matrix*) resultantes do BLASTP.

A matriz tende a fornecer *score* mais altos para regiões conservadas dentro de famílias de proteínas e *score* baixos para regiões pouco conservadas. Uma nova busca é realizada com esta matriz e com os novos alinhamentos formando uma nova matriz, o que se repetirá até o final para todas as sequências de proteínas testadas. O PSI-BLAST é um BLASTP interativo no qual uma matriz da família BLOSUM (*Blocks of Amino Acid Substitution Matrix*), após a primeira interação, é refeita com base nos alinhamentos entre as proteínas resultantes da consulta. Na medida que o alinhamento se estende, uma posição conservada no alinhamento recebe um *score* alto e uma posição não conservada um *score* baixo, sendo útil para encontrar membros distantes de famílias de proteínas, podendo identificar famílias de antígenos de membrana, por exemplo.

A matriz BLOSUM (HENIKOFF; HENIKOFF, 2000) é conhecida como alinhamentos múltiplos de blocos conservados de aminoácidos sem a presença de *gaps* de vários grupos de proteínas relacionadas, apresentando o cálculo de frequência de substituição de cada par de aminoácido. Para adquirir sequências com o máximo x% de identidade, são utilizadas as matrizes BLOSUM62 que é padrão para alinhamento de proteínas como o BLAST e a matriz BLOSUM85. Uma matriz BLOSUM_k considera apenas sequências com não menos do que k% de divergência entre suas bases, sendo sua fórmula geral definida por: $BLOSUM_k(i, j) = \log_2 \frac{freq_k(a_i, a_j)}{freq_k(a_i) \cdot freq_k(a_j)}$.

Com a finalidade de avaliarmos a similaridade entre as sequências de proteínas destes organismos versus as sequências de proteínas do banco de dados Psimap, adotando do PSIBLAST desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (Nome do job)
#PBS -l nodes=1:ppn=4,mem=6gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python
blastpgp -i (arquivo 1a) -d (arquivo 2) -a 4 -F T -s T -j 5 -m8 -o (arquivo de saída 3a)
blastpgp -i (arquivo 1b) -d (arquivo 2) -a 4 -F T -s T -j 5 -m8 -o (arquivo de saída 3b)
#Fim do script
```

Neste *script* pbs, os arquivos 1a e 1b consistem nas listas de sequências de aminoácidos das proteínas no formato Fasta para a bactéria e o carrapato, respectivamente e submetidas ao PSIBLAST contra o arquivo 2 contendo as sequências de proteínas no formato Fasta do banco de dados Scop, gerando os arquivos de saída 3a e 3b (Dados não publicados). Para obtermos as sequências de proteínas de *R. amblyommatis* e do *I. scapularis* similares aos domínios do banco de dados do Psimap, adotamos o $\text{eval} \leq 0,0001$ do PSIBLAST. Para esta finalidade desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE="(arquivo 1)"
OUTPUT_FILE="${INPUT_FILE}-2col"
perl script.scop.id.pl "$INPUT_FILE" > "$OUTPUT_FILE"
cat "$OUTPUT_FILE" | sort | uniq > "${OUTPUT_FILE}-no-repeat"
echo "Removendo arquivo '${OUTPUT_FILE}'"
rm -v "${OUTPUT_FILE}"
#Fim do script
```

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO, "<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
```



```

my $evalue;
my $idscope;
for (my $i = 0; $i <= $#linhas; $i++) {
  chomp ($linhas[$i]);
  my @linha = split (" ", $linhas[$i]);
  # Pega a primeira coluna
  $id = $linha[0];
  # Pega a segunda coluna
  $idscope = $linha[1];
  # Pega a decima primeira coluna
  $evalue = $linha[10];
  if ($evalue <= 0.0001) {
    print "$id\t$idscope\t$evalue\n";
  }
}
#Fim do script

```

No *script* pbs, o arquivo 1 consiste nas entradas para os arquivos 3a e 3b contendo os alinhamentos para as sequências de aminoácidos das proteínas dos organismos e os seus similares, tratando-se dos domínios de proteínas do banco de dados do Psimap. O *script* perl, chamado dentro do *script* pbs, foi programado para se obter os IDs das sequências de proteínas e os seus similares conforme o valor do *e-value* adotado para o PSIBLAST.

Pfam x Similaridade

O iPfam é um banco de dados que inferem interações baseado em sequência do Pfam (FINN, 2006), que interagem entre si em estruturas tridimensionais de proteínas depositadas no PDB (BERNSTEIN et al., 1978; KOURANOV, 2006). Para identificar a presença de domínios no iPfam, utilizamos a ferramenta HMM pfam que adota análise de modelos ocultos de Markov para sequências biológicas (KROGH et al., 1994).

Para avaliarmos a similaridade entre as sequências de proteínas destes organismos versus as sequências de proteínas do banco de dados Pfam, adotamos do HMM. Com esta finalidade desenvolvemos o *script* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=6,mem=200mb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh

```

```

#Carrega o módulo opt-python
module load hmmer/3.1b2
module load perl/5.24.0
module load pfam_scan/1.6
pfam_scan.pl -dir $PFAM_DATA -e_seq 0.05 -cpu $PBS_NP -Fasta (arquivo 1a) -outfile (arquivo de saída 2a)
pfam_scan.pl -dir $PFAM_DATA -e_seq 0.05 -cpu $PBS_NP -Fasta (arquivo 1b) -outfile (arquivo de saída 2b)
#Fim do script

```

Neste *script* pbs, temos os arquivos 1a e 1b referentes as sequências de aminoácidos das proteínas no formato Fasta para a bactéria e o carrapato, respectivamente submetidas ao HMM contra o arquivo de sequências de proteínas PFAM_DATA no formato Fasta do banco de dados do PFAM gerando os arquivos de saída 2a e 2b (Dados não publicados). Para obtermos as sequências de proteínas de *Rickettsia* e do *I. scapularis* similares aos domínios do banco de dados do Pfam, adotamos o *e-value* $\leq 0,01$ do HMM. Para esta finalidade desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE="(arquivo 1)"
OUTPUT_FILE="${INPUT_FILE}-2col"
perl script.pfam.id.pl "$INPUT_FILE" > "$OUTPUT_FILE"
cat "$OUTPUT_FILE" | sort | uniq > "${OUTPUT_FILE}-no-repeat"
echo "Removendo arquivo '${OUTPUT_FILE}'"
rm -v ${OUTPUT_FILE}
#Fim do script

```

```

#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
use Text::Trim qw(trim);
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
my $evalue;
my $idpfam;
my $bit score;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @linha = split (" ", $linhas[$i]);

```

```

#ignorar linha em branco
if(trim($linhas[$i]) eq ""){
#print "Linha $i: Blank line. Ignoring...\n";
next;
}
#ignorar linhas que começam com #
if ($linhas[$i] =~ /^#){
#print "Linha $i: Comment line. Ignoring...\n";
next;
}
# Pega a primeira coluna
$id = $linha[0];
# Pega a sexta coluna
$idpfam = $linha[5];
$idpfam =~ /(\w+)\.w+\/;
$idpfam = $1;
# Pega a decima terceira coluna
$value = $linha[12];
# pega a decima segunda coluna
$bit score = $linha[11];
if ( $value <= 0.01 ){
print "$id\t$idpfam\t$bit score\n";
}
}
}
#Fim do script

```

No *script* pbs, o arquivo 1 consiste nas entradas para os arquivos 2a e 2b contendo os alinhamentos para as sequências de aminoácidos das proteínas dos organismos e os seus similares, tratando-se dos domínios de proteínas do banco de dados do Pfam. O *script* perl, chamado dentro do *script* pbs, foi programado para se obter os IDs das sequências de proteínas e os seus similares conforme o valor do *e-value* adotado para o HMM.

String x Similaridade

Finalizando, fizemos o uso do banco de dados STRING, considerando *cutoff* ≥ 700 , na previsão de associações funcionais com métodos de inferência (SZKLARCZYK et al., 2011), baseando na coocorrência de proteínas em mapas metabólicos do banco de dados SCOP e de proteomas anotados e mantidos pelo Swiss-Prot (BAIROCH, 2000), atribuindo equivalência funcional para os genes derivada do banco COG (TATUSOV, 2001).

Para avaliarmos a similaridade entre as sequências de proteínas destes organismos versus as sequências de proteínas do banco de dados String, adotamos do BLASTP. Com esta finalidade desenvolvemos o *script* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job

```

```

#PBS -N (nome do job)
#PBS -l nodes=1:ppn=4,mem=6gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python
blastall -p blastp -i (arquivo 1a) -d (arquivo 2) -a 4 -F T -m8 -o (arquivo de saída 3a)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2) -a 4 -F T -m8 -o (arquivo de saída 3b)
#Fim do script

```

Neste *script* pbs, os arquivos 1a e 1b referentes as sequências de aminoácidos das proteínas no formato Fasta para a bactéria e o carrapato, respectivamente e submetidos ao BLASTP contra o arquivo 2 contendo as sequências de proteínas no formato Fasta do banco de dados do String, gerando os arquivos de saída 3a e 3b (Dados não publicados). Para obtermos as sequências de proteínas de *Rickettsia* e do *I. scapularis* similares aos domínios do banco de dados do String, adotamos da identidade $\leq 0,0001$ do BLASTP. Com esta finalidade desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1 gb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE="(arquivo 1)"
OUTPUT_FILE="${INPUT_FILE}-2col"
perl script.string.id.pl "$INPUT_FILE" > "$OUTPUT_FILE"
cat "$OUTPUT_FILE" | sort | uniq > "${OUTPUT_FILE}-no-repeat"
echo "Removendo arquivo '${OUTPUT_FILE}'"
rm -v ${OUTPUT_FILE}
#Fim do script

```

```

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
my $confianca;
my $identidade;

```

```

my $idpeimap;
for (my $i = 0; $i <= $#linhas; $i++) {
  chomp ($linhas[$i]);
  my @linha = split (" ", $linhas[$i]);
  #pega a primeira coluna
  $id = $linha[0];
  #pega a segunda coluna
  $idpeimap = $linha[1];
  #pega a quarta coluna
  $confianca = $linha[2];
  #pega a decima primeira coluna
  $identidade = $linha[10];
  if ($identidade <= 0.0001) {
    print "$linha[0]\t$linha[1]\n";
  }
}
#Fim do script

```

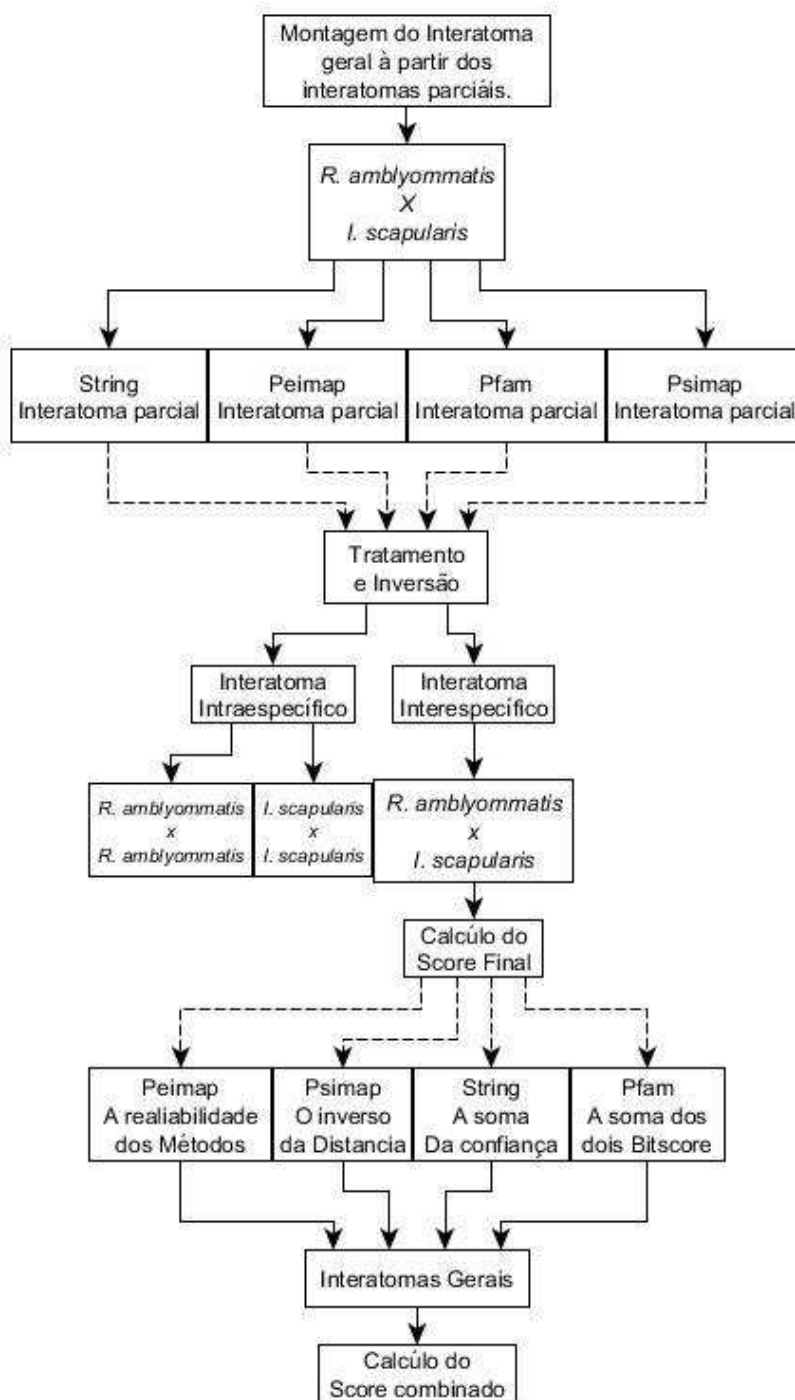
No *script* pbs, o arquivo 1 consiste nas entradas para os arquivos 3a e 3b contendo os resultados dos alinhamentos para as sequências de aminoácidos das proteínas dos organismos e os seus similares, tratando-se dos domínios de proteínas do banco de dados do String. O *script* perl, chamado dentro do *script* pbs, foi programado para se obter os IDs das sequências de proteínas e os seus similares conforme o valor de identidade adotado para o BLASTP.

3.1.2 Montagem do Interatoma

Na identificação de domínios, as proteínas de *R. amblyommatis* e do *I. scapularis* foram consideradas similares contra os pares de domínios dos bancos de dados com potencial para interagir se as interações entre os dois domínios estavam presentes nos bancos de dados do interatoma. Para a construção dos interatomas parciais de cada banco e do interatoma geral entre os quatro bancos, utilizamos dos resultados das análises de similaridade.

Consideramos, os pares de sequências de proteínas similares para a bactéria da espécie *R. amblyommatis* e para o carrapato da espécie *I. scapularis* entre os domínios de sequências de proteínas dos bancos de dados conforme cada análise individual (Figura 7).

Figura 7 Fluxograma contendo os passos consequentes após a análise de similaridade para as montagens dos interatomos parciais com cada par de interação em cada banco. Destaque para o cálculo do *score* final para cada interatoma parcial.



Fonte: Autor deste trabalho.

Na Figura 7, verifica-se os passos realizados após a realização das análises de similaridades. Nisto, as proteínas de *R. amblyommatidis* que formaram pares similares com as sequências dos bancos de dados foram interagidas com as proteínas do *I. scapularis* que

formaram pares de similaridade com as mesmas sequências destes bancos. Os pares de proteínas formados com o uso dos pares de domínios que interagem obtido em cada banco sugerem a ocorrência da interação entre estes organismos. Para esta finalidade desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=200mb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email),
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
perl interatoma_interespecifico.pl (arquivo 1a) (arquivo 2a) (arquivo 3a)
perl interatoma_interespecifico.pl (arquivo 1b) (arquivo 2b) (arquivo 3b)
perl interatoma_interespecifico.pl (arquivo 1c) (arquivo 2c) (arquivo 3c)
perl interatoma_interespecifico.pl (arquivo 1d) (arquivo 2c) (arquivo 3d)
#Fim do script

#!/usr/bin/perl
#como rodar: perl interatoma_interespecifico.pl
$peimapfile = $ARGV[0];
$blastfile = $ARGV[1];
$humanblast = $ARGV[2];
$outputfile = $ARGV[3];
open (PEIMAP, $peimapfile) || die ("PEIMAPInválido!");
open (BLAST, $blastfile) || die ("BlastInválido!");
open (HUMANBLAST, $humanblast) || die ("HumanInválido!");
open (OUTPUT, ">$outputfile"); #abertura do arquivo de saída
@peimap = <PEIMAP>;
@humanblast = <HUMANBLAST>;
@blast = <BLAST>;
close PEIMAP;
close BLAST;
close HUMANBLAST;
for ($i = 0; $i < @peimap; $i++) {
chomp $peimap[$i];
@linhapei = split (/s+/, $peimap[$i]); #separação da linha do peimap em colunas
@par1 = ();
@par2 = ();
$aux = 0;
$aux2 = 0;
for ($i2 = 0; $i2 < @blast; $i2++) {
chomp $blast[$i2];
@linhabla = split (/s+/, $blast[$i2]); #separação da linha do blast em coluna
if ($linhapei[0] =~ $linhabla[1]) {
$par1[$aux] = $linhabla[0];
$aux++;
}
if ($linhapei[1] =~ $linhabla[1]) {
$par2[$aux2] = $linhabla[0];
$aux2++;
}
}
}
```

```

for ($i2 = 0; $i2 < @humanblast; $i2++) {
chomp $humanblast[$i2];
@linhabla = split (/s+/, $humanblast[$i2]); #separação da linha do blast em coluna
if ($linhapei[0] =~ $linhabla[1]) {
$par1[$aux] = $linhabla[0];
$aux++;
}
if ($linhapei[1] =~ $linhabla[1]) {
$par2[$aux2] = $linhabla[0];
$aux2++;
}
}
$teste1 = @par1;
$teste2 = @par2;
if ($teste1 != 0 && $teste2 != 0) {
for ($aux = 0; $aux < @par1; $aux++) {
for ($aux2 = 0; $aux2 < @par2; $aux2++) {
print OUTPUT "$par1[$aux]\t$par2[$aux2]\t$peimap[$i]\n";
}
}
}
exit;
#Fim do script

```

No *script* pbs, os arquivos 1a-d consistem nos IDs para os pares de sequências de interação proteína-proteína presente nos bancos de dados do interatoma (Peimap, Pfam, Psimap e String). Os arquivos 2a-d e 3a-d, contém os IDs de identificação para as sequências de proteínas de *R. amblyommatis* e do *I. scapularis*, respectivamente contendo também os IDs das sequências de proteínas similares, obtidos por meio da análise de similaridade, conforme o Tópico 3.1.1, respectivamente para cada banco. O *script* perl, chamado dentro do *script* pbs, foi programado para filtrar os pares de sequências de proteínas de *R. amblyommatis* e do *I. scapularis* apresentando o potencial para interagir conforme os pares de interação de proteína-proteína similares depositados nos bancos de dados do interatoma (Dados não publicados).

3.1.2.1 Interatomas interespecíficos

Os arquivos de saída do interatoma parcial são referentes aos dados obtidos de interação proteína-proteína entre estes organismos utilizando dos quatro bancos de dados do Peimap, Pfam, Psimap e String, respectivamente. Todos estes arquivos possuem em comum interações intraespecífica e interespecífica entre as proteínas de ambos os organismos *R. amblyommatis* e *I. scapularis*, havendo a interação entre proteínas dos próprios organismos e entre eles. No entanto, para o estudo de interações entre as proteínas da bactéria *R. amblyommatis* e do carrapato da espécie *I. scapularis*, foi realizado um filtro em todos os resultados, com a remoção

de proteínas com interações intraespecíficas e a inversão dos pares de interação que contém as sequências de proteínas do carrapato interagindo com as da bactéria, exclusivamente.

Inversão e limpeza das interações intraespecíficas

A inversão foi realizada para organizarmos os pares de interação, colocando as proteínas da bactéria interagindo com as proteínas do carrapato e a remoção de proteínas com interações intraespecíficas.

Peimap

Com a finalidade de preservar os pares de interação e os formatos de saída dos arquivos do interatoma construído por meio do uso do banco de dados do Peimap, desenvolvemos o *script* abaixo:

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @coluna = split ("\t", $linhas[$i]);
#print "$coluna[0]\n";
if (($coluna[0]=~/^K/)&&($coluna[1]=~/^I/)){
print "$coluna[0]\;$coluna[1]\;$coluna[2]\;$coluna[3]\;$coluna[4]\n";
}
elsif (($coluna[0]=~/^I/)&&($coluna[1]=~/^K/)){
print "$coluna[1]\;$coluna[0]\;$coluna[3]\;$coluna[2]\;$coluna[4]\n";
}
}
#fim
```

Neste *script* pbs, foi programado a limpeza de interações intraespecíficas e a inversão das interações na ordem bactéria-carrapato. O arquivo 1 contém os dados obtidos do interatoma parcial para os pares de interação entre as sequências de proteínas de *R. amblyommatis* e do *I. scapularis* com o uso do Peimap. Estes pares apresentaram o potencial para interação conforme os pares de interação proteína-proteína e os métodos experimentais de validação da interação, com os quais estas foram comprovadas *in vivo* em outras abordagens e estudos.

Pfam

Com a finalidade de preservar os pares de interação e os formatos de saída dos arquivos do interatoma construído por meio do uso do banco de dados do Pfam, desenvolvemos o *script* abaixo:

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @coluna = split ("\t", $linhas[$i]);
#print "$coluna[0]\n";
if (($coluna[0]=~/^K/)&&($coluna[2]=~/^I/)){
print "$coluna[0]\;$coluna[1]\;$coluna[2]\;$coluna[3]\;$coluna[4]\;$coluna[5]\n";
}
elsif (($coluna[0]=~/^I/)&&($coluna[2]=~/^K/)){
print "$coluna[2]\;$coluna[3]\;$coluna[0]\;$coluna[1]\;$coluna[5]\;$coluna[4]\n";
}
}
#fim
```

Neste *script* pbs, foi programado a limpeza de interações intraespecíficas e a inversão das interações na ordem bactéria-carrapato. O arquivo 1 contém os dados obtidos do interatoma parcial para os pares de interação entre as sequências de proteínas de *R. amblyommatis* e do *I. scapularis* com o uso do Pfam. Estes pares apresentaram o potencial para interação conforme os pares de interação proteína-proteína contendo os valores dos dois *Bit score*.

Psimap

Com a finalidade de preservar os pares de interação e os formatos de saída dos arquivos do interatoma construído por meio do uso do banco de dados do Psimap, desenvolvemos o *script* abaixo:

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
```

```

my @linhas = <ARQUIVO>;
close ARQUIVO;
for (my $i = 0; $i <= $#linhas; $i++) {
chomp ($linhas[$i]);
my @coluna = split ("t", $linhas[$i]);
#print "$coluna[0]\n";
if (($coluna[0]=~/^K/)&&($coluna[1]=~/^I/)){
print "$coluna[0]\;$coluna[1]\;$coluna[2]\;$coluna[3]\;$coluna[4]\n";
}
elsif (($coluna[0]=~/^I/)&&($coluna[1]=~/^K/)){
print "$coluna[1]\;$coluna[0]\;$coluna[3]\;$coluna[2]\;$coluna[4]\n";
}
}
#fim

```

Neste *script* pbs, foi programado a limpeza de interações intraespecíficas e a inversão das interações na ordem bactéria-carrapato. O arquivo 1 contém os dados obtidos do interatoma parcial para os pares de interação entre as sequências de proteínas de *R. amblyommatis* e do *I. scapularis* com o uso do Psimap. Estes pares apresentaram o potencial para interação conforme os pares de interação proteína-proteína contendo os valores das distâncias entre estes pares.

String

Com a finalidade de preservar os pares de interação e os formatos de saída dos arquivos do interatoma construído por meio do uso do banco de dados do String, desenvolvemos o *script* abaixo:

```

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
for (my $i = 0; $i <= $#linhas; $i++) {
chomp ($linhas[$i]);
my @coluna = split (";", $linhas[$i]);
#print "$coluna[0]\n";
if (($coluna[0]=~/^K/)&&($coluna[1]=~/^I/)){
print "$coluna[0]\;$coluna[1]\;$coluna[2]\;$coluna[3]\;$coluna[4]\n";
}
elsif (($coluna[0]=~/^I/)&&($coluna[1]=~/^K/)){
print "$coluna[1]\;$coluna[0]\;$coluna[3]\;$coluna[2]\;$coluna[4]\n";
}
}
#fim

```

Neste *script* pbs, foi programado a limpeza de interações intraespecíficas e a inversão das interações na ordem bactéria-carrapato. O arquivo 1 contém os dados obtidos do interatoma parcial para os pares de interação entre as sequências de proteínas de *R. amblyommatis* e do *I. scapularis* com o uso do String. Estes pares apresentaram o potencial para interação conforme os pares de interação proteína-proteína contendo os valores de confiança entre estes pares.

3.1.3 Cálculos de confiança do modelo do interatoma

Para integrar os dados de predição, foi realizado um cálculo de confiança (FLÓREZ et al., 2010) que assume independência entre os quatro métodos e é calculado de acordo com a fórmula:

$$score\ combinado = 1 - \prod_{i \in M} (1 - R_i)^n$$

Onde *score combinado* é um valor de confiança final, M é o conjunto de métodos utilizados (Peimap, Psimap, Pfam e String), R_i é o valor do *score final* para cada interação pelo método i , n é o número de interações preditas pelo método i . O valor do *score final* do Peimap reflete a confiabilidade de cada método experimental para validar as interações proteína-proteína, sendo previamente estimado denominada cálculo de confiabilidade (FLÓREZ et al., 2010). Para cada par de interação, este valor do *score final* foi obtido pela soma dos valores de confiabilidade, caso a interação ocorra e seja validada experimentalmente por mais de um método, e quando esta ocorre mais de uma vez para as mesmas proteínas, apresentando o número para esta frequência.

O valor de *score final* para cada par de interação proteína-proteína do Pfam é a soma da soma entre os dois “bits *score*” da assinatura dos dois domínios da interação. Para cada par de proteína com potencial de interação, este valor de *score final* foi obtido pela soma dos valores da soma entre os bits *score*, caso a interação ocorra mais de uma vez para as mesmas proteínas, apresentando o número para esta frequência. O valor de *score final* para cada par de interação proteína-proteína do Psimap é a soma do inverso da distância entre os pares de domínios estruturais que interagem, depositados no PDB. Para cada par de proteína com potencial de interação, este valor de *score final* foi obtido pela soma dos valores do inverso da distância entre os pares de domínios estruturais, caso a interação ocorra mais de uma vez para as mesmas proteínas, apresentando o número para esta frequência.

O valor de *score* final para cada par de interação proteína-proteína do String, que adota o uso de métodos de “base de contexto” para detectar interações e as relações funcionais entre proteínas usando informação genômica, conforme homologia para semelhança de árvores genealógica é o valor de confiança. Para cada par de proteína com potencial de interação, este valor de *score* final foi obtido pela soma dos valores de confiança, caso a interação ocorra mais de uma vez para as mesmas proteínas, apresentando o número para esta frequência.

3.1.3.1 Cálculo do *score* final

Peimap x *Score* final

Para o cálculo do *score* final atribuído aos pares de proteínas que interagem relacionados aos pares de interação proteína-proteína depositados no banco de dados do Peimap, o valor de confiança deste banco reflete a confiabilidade de cada método experimental para detectar interações entre proteínas e foi previamente estimado denominada cálculo de confiabilidade (CHUA; SUNG; WONG, 2006). Para a determinação da confiabilidade dos métodos do Peimap foi feito o uso de valores entre 0 e 1 referentes aos métodos utilizados para as validações experimentais destas interações. Com estes valores e a detecção dos métodos do Peimap foi possível curar uma lista contendo os métodos e estes valores

Confiabilidade dos Métodos do Peimap

A confiabilidade de cada método experimental para detectar interações entre proteínas foi previamente estimada sendo denominada cálculo de confiabilidade (FLÓREZ et al., 2010). Para a determinação destes valores foi realizado um filtro obtendo do interatoma construído com o Peimap os métodos utilizados para as validações experimentais que validaram as interações de proteínas preditas em trabalhos anteriores.

Detecção dos Métodos

Com a finalidade de preservar os métodos para cada par de interação obtido do arquivo do interatoma construído, desenvolvemos o *script* abaixo:

```
#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#Realiza o carregamento do arquivo com 'inter Não Redundantes'
my $entrada = $ARGV[0];
open ARQUIVO,"<$entrada";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my %métodos = ();
for (my $i = 0; $i <= $#linhas; $i++){
my $linha = $linhas[$i];
chomp ($linha);
my @colunas = split (";", $linha);
my @methods_inter = split (/[,;]/, $colunas[4]);
for (my $m = 0; $m <= $#methods_inter; $m++){
my $method = $methods_inter[$m];
if($method ne "null"){
$métodos{$method} = 1;
}
}
}
while( my( $key, $value ) = each %métodos ){
print "$key\n";
}
#Fim do script
```

Valores de Reliabilidade

Com os valores de confiabilidade sugerido por diversos autores e com o filtro realizado para a obtenção dos métodos do Peimap foi possível curar uma lista contendo os métodos e estes valores. Com a finalidade de construir uma lista de valores de confiança para cada método, desenvolvemos o *script* abaixo:

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $file1 = $ARGV[1];
open ARQUIVO,"<$file1";
my @linhas1 = <ARQUIVO>;
close ARQUIVO;
my $métodos;
my $reabilidade;
my $metodos1;
for (my $i = 0; $i <= $#linhas; $i++ ){
chomp ($linhas[$i]);
my @linha = split (";", $linhas[$i]);
```

```

for (my $j = 0; $j <= $#linhas1; $j++){
chomp ($linhas1[$j]);
my @linha1 = $linhas1[$j];
#pega a primeira coluna
$métodos = $linha[0];
#pega a segunda coluna
$reabilidade = $linha[1];
#pega a primeira coluna
$metodos1 = $linha1[0];
if ($metodos1 eq $métodos){
print "$linha1[0]\t$linha[1]\n";
if($metodos1 ne $métodos){
print "$linha1[0]\n";
}
}
}
}
#Fim do script

```

No primeiro *script* perl, pudemos extrair os métodos de validação experimental para cada par de interação do interatoma parcial do Peimap. No segundo *script* perl, plotamos para cada método o valor de confiabilidade sugerido para a validação da interação correspondendo entre valores de 0 e 1 (Tabela 1).

Tabela 1 Métodos de análises que comprovam a interação entre proteínas disponibilizados no banco de dados do Peimap, com atribuição da confiabilidade para cada tipo de validação experimental.

Métodos experimentais	Confiabilidade
2h fragment pooling	0.455904
3 hybrid method	0.265407
adenylate cyclase	0.666667
affinity chrom	0.823077
fluorescence-activated cell sorting	0.455904
fluorescence-anisotropy	0.265407
fluorescence imaging	0.5
fluorescence microscopy	0.5
fluorescence polarization spectroscopy	0.265407
in vitro	0.265407
in vivo	0.265407
ion exchange chrom	0.823077
ion exchange chromatography	0.823077
isothermal titration calorimetry	0.265407
Itc	0.265407
kinase htrf	0.666667
lambda phage	0.455904
lambda phage display	0.455904
light scattering	0.265407
light-scattering	0.265407
Mappit	0.265407
mass-spectrometry	0.265407
mass spectrometry studies of complexes	0.265407
x-ray diffraction	0.265407
yeast display	0.265407

Na Tabela 1, verifica-se o resumo da lista dos métodos e o valor de confiabilidade atribuído a cada técnica para a validação da interação presente no banco do Peimap. A partir disto, para o cálculo do *score* final desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1500mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
ARQUIVO_PEIMAP_UNICOS=(arquivo 1a)
ARQUIVO_PEIMAP_COMPLETO=(arquivo 1b)
ARQUIVO_RELIABILIDADE=(arquivo 1c)
ARQUIVO_SAIDA=(arquivo de saída 1d)
perl script.formula.int.peimap.score.final.pl $ARQUIVO_PEIMAP_UNICOS
$ARQUIVO_PEIMAP_COMPLETO $ARQUIVO_RELIABILIDADE > $ARQUIVO_SAIDA
#Fim do script
```

```
#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#Realiza o carregamento do arquivo com 'inter Nao Redundantes'
my $inter_nred = $ARGV[0];
open ARQUIVO,"<$inter_nred";
my @linhas_inter_nred = <ARQUIVO>;
close ARQUIVO;
#Realiza o carregamento do arquivo com 'inter Completas com Informação do método'
my $inter_comp = $ARGV[1];
open ARQUIVO,"<$inter_comp";
my @linhas_inter_comp = <ARQUIVO>;
close ARQUIVO;
#Realiza o carregamento do arquivo com 'Metodo x Valor'
my $confiabilidade = $ARGV[2];
open ARQUIVO,"<$confiabilidade";
my @linhas_confabilidade = <ARQUIVO>;
close ARQUIVO;
#print "Lendo arquivos '$ARGV[0]', '$ARGV[1]', '$ARGV[2]'\n\n";
my %confiabilidade = ();
for (my $i_conf = 0; $i_conf <= $#linhas_confabilidade; $i_conf++){
my $linha = $linhas_confabilidade[$i_conf];
chomp ($linha);
my @colunas = split(";", $linha);
#if ($linha =~ /(.*?)s+(\d*\.?\d+)/isg){
my $metodo = $colunas[0];
my $valor = $colunas[1];
$confiabilidade{$metodo} = $valor;
#print "$confiabilidade{$metodo}\n";
#}
}
# Array que contara a frequencia de cada interação com cada metodo
my %cont_freq_inter = ();
```



```

my %cont_score_inter = ();
# Percorre interações sem repetição
for (my $i_nred = 0; $i_nred <= $#linhas_inter_nred; $i_nred++){
my $linha = $linhas_inter_nred[$i_nred];
chomp ($linha);
my @inter_nred = split (";", $linha);
#print "$inter_nred[0],$inter_nred[1]\n";
# Inicializa o contador para cada interação não redundante
$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]} = 0;
$cont_score_inter{$inter_nred[0]}{$inter_nred[1]} = 0;
# percorre interações completo
for (my $i_comp = 0; $i_comp <= $#linhas_inter_comp; $i_comp++){
my $linha = $linhas_inter_comp[$i_comp];
chomp ($linha);
if ($linha =~ /(.*);(.*)\;(.*);(.*)\;(.*);(.*)/isg){
my $inter_comp_col1 = $1;
my $inter_comp_col2 = $2;
my $inter_comp_col5 = $5;
if(($inter_nred[0] eq $inter_comp_col1) and ($inter_nred[1] eq $inter_comp_col2)){
my @methods_inter = split (/[;]/, $inter_comp_col5);
for (my $m = 0; $m <= $#methods_inter; $m++){
my $method = $methods_inter[$m];
if($method ne "null"){
#print "$inter_comp_col5\n";
#print "$methods_inter[0]\n";
#print "$confiabilidade{$method}\n";
$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]} ++;
$cont_score_inter{$inter_nred[0]}{$inter_nred[1]} += $confiabilidade{$method};
}
}
}
}
}
}
Print
"$inter_nred[0];$inter_nred[1];$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]};$cont_score_inter{$inter_nred[0]}{$inter_nred[1]}\n";
}
#Fim do script

```

No *script* pbs, o arquivo 1a contém os pares de IDs das proteínas que interagem entre a bactéria da espécie *R. amblyommatis* e o de carrapato da espécie *I. scapularis*, enquanto o arquivo 1b contém o arquivo do interatoma parcial contendo os pares de IDs das proteínas que interagem, os pares de IDs da interação do banco do Peimap e os métodos de validação para cada par de interação. O arquivo 1c contém os valores de confiabilidade declarados para cada método conforme sugerido por diversos autores que comprovem a interação entre os pares de interação proteína-proteína deste banco de dados. O Arquivo de saída 1d contém os ids dos pares de proteínas que interagem com o valor de *score* final para cada par de interação. No *script* perl, chamado dentro do *script* pbs, foi programado para obter o cálculo do *score* final para cada par de interação formados entre estes organismos.

Pfam x Score final

Para o cálculo do *score* final atribuído aos pares de proteínas que interagem relacionados aos pares de domínios depositados no banco de dados do Pfam, adotamos o valor de confiança do iPfam que é a soma entre os dois “*bits score*” da assinatura dos domínios do Pfam do par proteínas com potencial interação. Para esta finalidade desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=3gb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
##PBS -M (email)
#PBS -j oe
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
ARQUIVO_UNICOS=(arquivo 1a)
ARQUIVO_COMPLETO=(arquivo 1b)
ARQUIVO_SAIDA=(arquivo 1c)
echo "PWD=`pwd`"
START=$(date +%s)
perl script.formula.pfam.score.final.pl $ARQUIVO_UNICOS $ARQUIVO_COMPLETO >
$ARQUIVO_SAIDA
END=$(date +%s)
DIFF=$(( $END - $START ))
echo "It took $DIFF seconds"
#Fim do script

#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#Realiza o carregamento do arquivo com 'inter Nao Redundantes'
my $inter_nred = $ARGV[0];
open ARQUIVO,"<$inter_nred";
my @linhas_inter_nred = <ARQUIVO>;
close ARQUIVO;
#Realiza o carregamento do arquivo com 'inter Completas com Informação do metodo'
my $inter_comp = $ARGV[1];
open ARQUIVO,"<$inter_comp";
my @linhas_inter_comp = <ARQUIVO>;
close ARQUIVO;
#print "Lendo arquivos '$ARGV[0]', '$ARGV[1]'\n\n";
# Array que contara a frequencia de cada interação com cada metodo
my %cont_freq_inter = ();
my %sum_bit_score_inter = ();
my @inter_nred = [];
#my $min_dist_global = $inf;
# Percorre interações sem repetição
for (my $i = 0; $i <= $#linhas_inter_nred; $i++){
my $linha = $linhas_inter_nred[$i];
chomp ($linha);
@inter_nred = split (";", $linha);
```

```

# Inicializa o contador para cada interação não redundante
$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]} = 0;
$sum_bit_score_inter{$inter_nred[0]}{$inter_nred[1]} = 0;
}
# percorre interações completo
for (my $j = 0; $j <= $#linhas_inter_comp; $j++){
my $linha = $linhas_inter_comp[$j];
chomp ($linha);
if ($linha =~ /(.*);(.*)\;(.*);(.*)\;(.*);(.*)/isg){
my $inter_1 = $1;
my $bit_score_1 = $2;
my $inter_2 = $3;
my $bit_score_2 = $4;
#print ("inter_1 = $inter_1, bit score_1 = $bit score_1, inter_2 = $inter_2, bit score_2 =
$bit score_2\n");
$sum_bit_score_inter{$inter_1}{$inter_2} += ($bit_score_1 + $bit_score_2);
$cont_freq_inter{$inter_1}{$inter_2} ++;
}
}
for (my $i = 0; $i <= $#linhas_inter_nred; $i++){
my $linha = $linhas_inter_nred[$i];
chomp ($linha);
@inter_nred = split (";", $linha);
print"$inter_nred[0];$inter_nred[1];$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]};$sum_bit_score_
inter{$inter_nred[0]}{$inter_nred[1]}\n";
}
}
#Fim do script

```

No *script* pbs, o arquivo 1a contém os pares de IDs das proteínas que interagem entre a bactéria da espécie *R. amblyommatis* e o de carrapato da espécie *I. scapularis*, enquanto o arquivo 1b contém o arquivo do interatoma parcial contendo os pares de IDs das proteínas que interagem, os pares de IDs da interação do banco do Pfam e os valores dos dois *bit score* para cada par de interação. O Arquivo de saída 1c contém os ids dos pares de proteínas que interagem com o valor de *score* final para cada par de interação. No *script* perl, chamado dentro do *script* pbs, foi programado para obter o cálculo do *score* final para cada par de interação formados entre estes organismos.

Psimap x Score final

Para o cálculo do *score* final atribuído aos pares de proteínas que interagem relacionados aos pares de domínios depositados no banco de dados do Psimap, adotamos o valor de confiança do PSIMAP que é o inverso da distância entre os pares de domínios estruturais que estão interagindo no PDB. Para esta finalidade desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)

```

```

#PBS -l nodes=1:ppn=1,mem=3gb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
##PBS -M (email)
#PBS -j oe
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
ARQUIVO_UNICOS=(arquivo 1a)
ARQUIVO_COMPLETO=(arquivo 1b)
ARQUIVO_SAIDA=saida_(arquivo de saída 1c)
echo "PWD=`pwd`"
START=$(date +%s)
perl script.formula.psimap.score.final.pl $ARQUIVO_UNICOS $ARQUIVO_COMPLETO >
$ARQUIVO_SAIDA
END=$(date +%s)
DIFF=$(( $END - $START ))
echo "It took $DIFF seconds"
#Fim do script

```

```

#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#Realiza o carregamento do arquivo com 'inter Nao Redundantes'
my $inter_nred = $ARGV[0];
open ARQUIVO,"<$inter_nred";
my @linhas_inter_nred = <ARQUIVO>;
close ARQUIVO;
#Realiza o carregamento do arquivo com 'inter Completas com Informação do metodo'
my $inter_comp = $ARGV[1];
open ARQUIVO,"<$inter_comp";
my @linhas_inter_comp = <ARQUIVO>;
close ARQUIVO;
#print "Lendo arquivos '$ARGV[0]', '$ARGV[1]'\n\n";
# Array que contara a frequencia de cada interação com cada metodo
my %cont_freq_inter = ();
my %min_dist_inter = ();
my $inf = 9**9**9;
my @inter_nred = [];
my $min_dist_global = $inf;
# Percorre interações sem repetição
for (my $i = 0; $i <= $#linhas_inter_nred; $i++){
my $linha = $linhas_inter_nred[$i];
chomp ($linha);
@inter_nred = split (";", $linha);
# Inicializa o contador para cada interação não redundante
$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]} = 0;
$min_dist_inter{$inter_nred[0]}{$inter_nred[1]} = $inf;
}
# percorre interações completo
for (my $j = 0; $j <= $#linhas_inter_comp; $j++){
my $linha = $linhas_inter_comp[$j];
chomp ($linha);
if ($linha =~ /(.*);(.*)\;(.*);(.*)\;(.*);(.*)/isg){
my $inter_1 = $1;
my $inter_2 = $2;
my $dist_inter = $5;
if (($dist_inter < $min_dist_global) && $dist_inter != 0){

```

```

$min_dist_global = $dist_inter;
}
if($dist_inter < $min_dist_inter{$inter_1}{$inter_2}){
$min_dist_inter{$inter_1}{$inter_2} = $dist_inter;
}
$cont_freq_inter{$inter_1}{$inter_2} ++;
}
}
#foreach $inter_A(@item){
for (my $i = 0; $i <= $#linhas_inter_nred; $i++){
my $linha = $linhas_inter_nred[$i];
chomp ($linha);
@inter_nred = split (";", $linha);
my $dist_inter = $min_dist_inter{$inter_nred[0]}{$inter_nred[1]};
my $inverse_min_dist_inter = -1;
if ($dist_inter == 0){
$inverse_min_dist_inter = (1/$min_dist_global) + 1.0;
#print "inverse_min_dist_inter = $inverse_min_dist_inter\n";
}else{
$inverse_min_dist_inter = (1/$dist_inter);
}
print
"$inter_nred[0];$inter_nred[1];$cont_freq_inter{$inter_nred[0]}{$inter_nred[1]};$min_dist_inter{$inter_nred[0]}{$inter_nred[1]};$inverse_min_dist_inter\n";
}
#print "min_dist_global = $min_dist_global\n";
#Fim do script

```

No *script* pbs, o arquivo 1a contém os pares de IDs das proteínas que interagem entre a bactéria da espécie *R. amblyommatidis* e o de carrapato da espécie *I. scapularis*, enquanto o arquivo 1b contém o arquivo do interatoma parcial contendo os pares de IDs das proteínas que interagem, os pares de IDs da interação do banco do Psimap e os valores das distâncias para cada par de interação. O Arquivo de saída 1c contém os ids dos pares de proteínas que interagem com o valor de *score* final para cada par de interação. No *script* perl, chamado dentro do *script* pbs, foi programado para obter o cálculo do *score* final para cada par de interação formados entre estes organismos.

String x Score final

Para o cálculo do *score* final atribuído aos pares de proteínas que interagem relacionados aos pares de domínios depositados no banco de dados do String, adotamos a soma da confiança para um mesmo par de interação e a soma da frequência desta. Com esta finalidade desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serão utilizados pelo job
#PBS -N (nome do job)

```

```

#PBS -l nodes=1:ppn=1,mem=1500mb
#PBS -m ae
#mensagem será enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
ARQUIVO_STRING_UNICOS=(arquivo 1a)
ARQUIVO_STRING_COMPLETO=(arquivo 1b)
ARQUIVO_SAIDA=(arquivo 1c)
perl script.formula.string.score.final.pl $ARQUIVO_STRING_UNICOS
$ARQUIVO_STRING_COMPLETO > $ARQUIVO_SAIDA
#Fim do script

#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
#arquivo sem redundância "saida.ambly.int.n.red.string"
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
#arquivo completo "saida.ambly.int.string"
my $file1 = $ARGV[1];
open ARQUIVO,"<$file1";
my @linhas1 = <ARQUIVO>;
close ARQUIVO;
my $cont_freq_inter = 0;
my $cont_score_inter = 0;
#Percorre interação n.red
for (my $i = 0; $i <= $#linhas; $i = $i + 1){ #analisa cada linha do arquivo sem redundância
chomp ($linhas[$i]);
$cont_freq_inter = 0;
$cont_score_inter = 0;
#Percorre interação completo
for (my $t = 0; $t <= $#linhas1; $t = $t + 1){ #analisa cada linha do arquivo completo
chomp ($linhas1[$t]);
if ($linhas1[$t]=~/^$linhas[$i]/){
$cont_freq_inter++;
#print "$linhas1[$t]\n";
my @colunas = split (";", $linhas1[$t]);
$cont_score_inter = $cont_score_inter + $colunas[4];
}
}
print "$linhas[$i]\;$cont_freq_inter\;$cont_score_inter\n";
}
#Fim do script

```

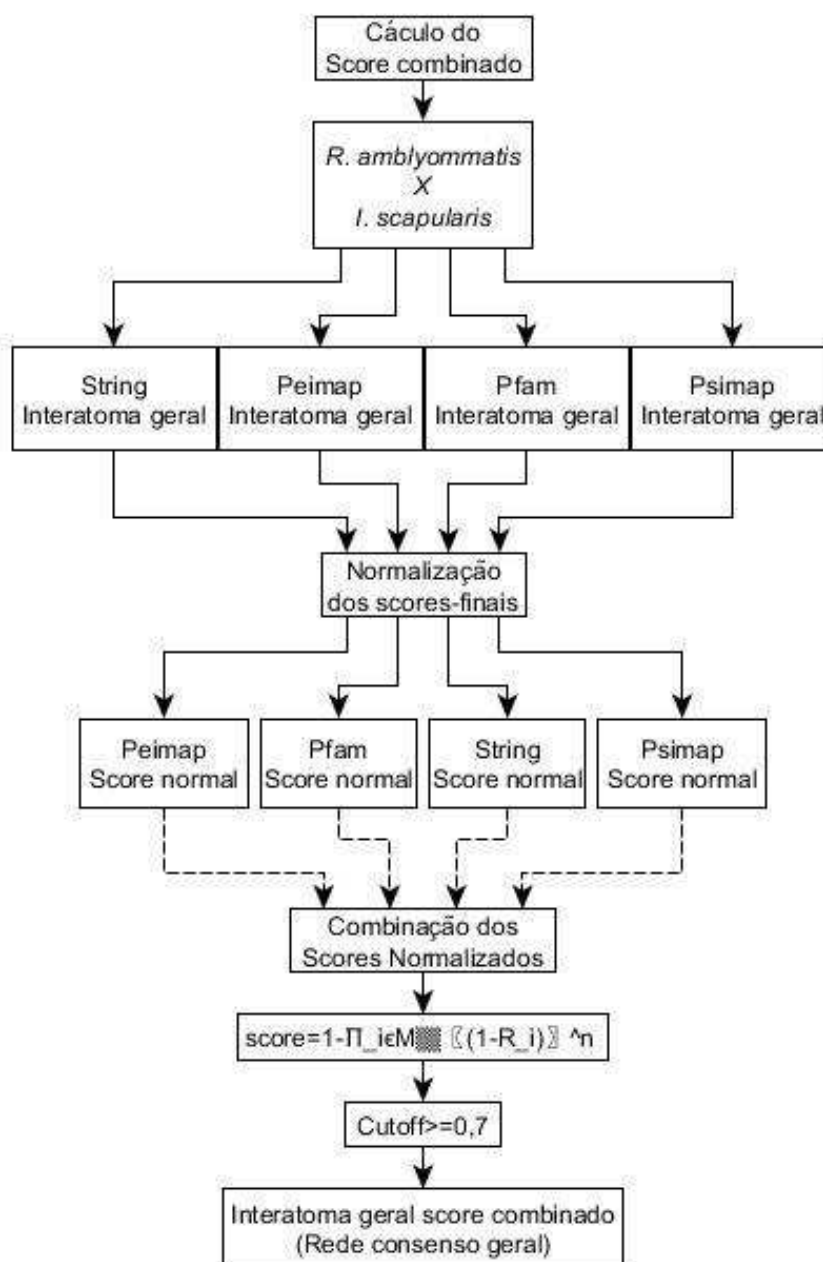
No *script* pbs, o arquivo 1a contém os pares de IDs das proteínas que interagem entre a bactéria da espécie *R. amblyommatidis* e o de carrapato da espécie *I. scapularis*, enquanto o arquivo 1b contém o arquivo do interatoma parcial contendo os pares de IDs das proteínas que interagem, os pares de IDs da interação do banco do String e os valores de confiança para cada par de interação. O Arquivo de saída 1c contém os ids dos pares de proteínas que interagem

com o valor de *score* final para cada par de interação. No *script* perl, chamado dentro do *script* pbs, foi programado para obter o cálculo do *score* final para cada par de interação formados entre estes organismos.

3.1.3.2 Cálculo do *score* combinado

Para o cálculo do *score* combinado, tem-se a fórmula do interatoma atribuindo o valor de confiança final, com a combinação dos valores dos *scores* finais de cada par de interação entre as proteínas de *R. amblyommatis* e do *I. scapularis* por banco do interatoma avaliado (Figura 8).

Figura 8 Fluxograma para o tratamento dos interatomos parciais e obtenção do interatoma interespecífico entre as proteínas de *R. amblyommatis* e *I. scapularis*. Destaque para o cálculo do *score* combinado para cada par de interação combinando os *scores* finais por interatoma parcial em cada banco.



Fonte: Autor deste trabalho.

Na Figura 8, verifica-se os passos realizados para a análise do *score* combinado. O valor de confiança final foi combinando e normalizado gerando valores entre 0 e 1. Interações preditas foram consideradas de alta confiabilidade para valores de confiança combinada acima de 0,7 [146].

Normalização do *score* final

Para cada par de interação obtidos com o uso dos quatro banco de dados, o valor do *score* final foi normalizado, gerando valores entre 0 e 1. Com esta finalidade desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1 gb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE_1=(arquivo 1a).peimp.score.final
INPUT_FILE_2=(arquivo 1b).psimp.score.final
INPUT_FILE_3=(arquivo 1c).pfam.score.final
INPUT_FILE_4=(arquivo 1d).string.score.final
OUTPUT_FILE_1="${INPUT_FILE_1}.normal"
OUTPUT_FILE_2="${INPUT_FILE_2}.normal"
OUTPUT_FILE_3="${INPUT_FILE_3}.normal"
OUTPUT_FILE_4="${INPUT_FILE_4}.normal"
perl script.normal.score.pl ${INPUT_FILE_1}.txt > ${OUTPUT_FILE_1}
mv -v ${OUTPUT_FILE_1} .
perl script.normal.score.pl ${INPUT_FILE_2}.txt > ${OUTPUT_FILE_2}
mv -v ${OUTPUT_FILE_2} .
perl script.normal.score.pl ${INPUT_FILE_3}.txt > ${OUTPUT_FILE_3}
mv -v ${OUTPUT_FILE_3} .
perl script.normal.score.pl ${INPUT_FILE_4}.txt > ${OUTPUT_FILE_4}
mv -v ${OUTPUT_FILE_4} .
#Fim do script

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id1;# proteina 1
my $id2;# proteina 2
my $frequencia;
my $score;
#my $normal = $score / $maiorvalor;
my $inf = 9**9**9;
my $maior = -1*$inf;
#identifica o maior score
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @colunas = split (/[\s+;|/, $linhas[$i]);
#pega a quarta primeira coluna
```

```

$score = $colunas[3];
if ($score > $maior){
$maior = $score;
}
}
#Calcula o score corrigido
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @colunas = split (/[\s+;]/, $linhas[$i]);
#pega a primeira coluna
$id1 = $colunas[0];
#pega a segunda coluna
$id2 = $colunas[1];
#pega a terceira coluna
$frequencia = $colunas[2];
#pega a quarta primeira coluna
$score = $colunas[3];
my $score_corrigido = $score/$maior;
print "$id1;$id2;$frequencia;$score_corrigido\n";
}
#Fim do script

```

No *script* pbs, os arquivos 1a-d contém os pares únicos dos IDs das sequências de proteínas que interagem entre a bactéria da espécie *R. amblyommatis* e o carrapato da espécie *I. scapularis*, contendo os valores da frequência da interação e do *score* final, conforme obtido com o uso de cada banco de dados do interatoma, sendo o Peimap, Pfam, Psimap e String, respectivamente. No *script* perl, chamado dentro do *script* pbs, foi programado a normalização do *score* final. Esta fórmula tem por base, o valor da divisão entre o *score* final calculado para o par de interação formado entre as proteínas destes organismos, sendo dividido pelo valor do maior *score* final observado, em cada um dos quatro bancos do interatoma adotados.

Combinação do *score* final

Logo em seguida a normalização dos valores de *score* final, foi realizado a combinação destes dados gerando um valor de *score* combinado para cada par de interação. Após, o cálculo do *score* combinado para estes pares de interação preditos entre as proteínas de *R. amblyommatis* e do *I. scapularis* foram consideradas de alta confiabilidade para valores de confiança combinada (cutoff) $\geq 0,7$. Para esta finalidade desenvolvemos o *script* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=500mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)

```

```

#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE_1=(arquivo 1a).pfam.score.final.normal
INPUT_FILE_2=(arquivo 1b).string.score.final.normal
INPUT_FILE_3=(arquivo 1c).peimap.score.final.normal
INPUT_FILE_4=(arquivo 1d).psimap.score.final.normal
OUTPUT_FILE="(arquivo de saída 1e)"
perl script.score.combinado.pl ${INPUT_FILE_1} ${INPUT_FILE_2} ${INPUT_FILE_3}
${INPUT_FILE_4} > ${OUTPUT_FILE}
#Fim do script

```

```

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my %freq_interacao_banco = ();
my %score_interacao_banco = ();
my %interacoes_unicas = ();
#Contabilizar a frequência e score de cada banco
for (my $i = 0; $i <= $#ARGV; $i++){
my $banco = $ARGV[$i];
#print "Lendo banco '$ARGV[$i]'\n";
open ARQUIVO,"<$ARGV[$i]";
my @linhas_banco = <ARQUIVO>;
close ARQUIVO;
for (my $j = 0; $j <= $#linhas_banco; $j++){
chomp ($linhas_banco[$j]);
my @colunas = split (/;/, $linhas_banco[$j]);
my $id1 = $colunas[0];
my $id2 = $colunas[1];
my $freq = $colunas[2];
my $score = $colunas[3];
my $inter = "$id1;$id2";
$freq_interacao_banco{$inter}{$banco}=$freq;
$score_interacao_banco{$inter}{$banco}=$score;
$interacoes_unicas{$inter} = 1;#qualquer valor. Apenas para registro
}
}
print "ID1;ID2;SCORE_COMBINADO;";
for (my $i = 0; $i <= $#ARGV; $i++){
print "$ARGV[$i]";
if ($i <= ($#ARGV - 1)){
print ";";
}
}
print "\n";
#Para cada interação não redundante
while( my( $inter, $value ) = each %interacoes_unicas ){
my $score_combinado_inter = 1;
my %inter_ocorreu_banco = ();
for (my $i = 0; $i <= $#ARGV; $i++){
my $banco = $ARGV[$i];
my $score = $score_interacao_banco{$inter}{$banco};
my $freq = $freq_interacao_banco{$inter}{$banco};
my $score_calc_inter;
if($score){

```

```

$score_calc_inter = (1 - $score)**($freq);
$inter_ocorreu_banco{$banco} = 1;
}else{
$score_calc_inter = 1;
$inter_ocorreu_banco{$banco} = 0;
}
$score_combinado_inter = ($score_combinado_inter * $score_calc_inter);
}
$score_combinado_inter = (1 - $score_combinado_inter);
#Gera a saída
print "$inter;$score_combinado_inter;";
for (my $i = 0; $i <= $#ARGV; $i++){
my $banco = $ARGV[$i];
print "$inter_ocorreu_banco{$ARGV[$i]}";
if ($i <= ($#ARGV - 1)){
print ",";
}
}
print "\n";
}
#Fim do script

```

No *script* pbs, os arquivos 1a-d contém os pares únicos dos IDs das sequências de proteínas que interagem entre a bactéria da espécie *R. amblyommatis* e o carrapato da espécie *I. scapularis*, contendo os valores da frequência da interação e do *score* final normalizado, obtidos com o uso de banco de dados do interatoma, sendo o Peimap, Pfam, Psimap e String, respectivamente. No *script* perl, chamado dentro do *script* pbs, foi programado a combinação do *score* final normalizado. Esta fórmula tem por base, o cálculo de confiança final do interatoma construído para cada par de proteínas que interagem. Após, o cálculo do *score* combinado para cada par de interação predito entre as proteínas de *R. amblyommatis* e do *I. scapularis* foram consideradas de alta confiabilidade aqueles pares que obtiveram valores de *score* combinado com (cutoff) $\geq 0,7$. Para esta finalidade desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1 gb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE="(arquivo 1).score_combinado"
OUTPUT_FILE="${INPUT_FILE}-4col"
perl script.score.cutoff.frequencia.pl ${INPUT_FILE}.csv > ${OUTPUT_FILE}
cat ${OUTPUT_FILE} | sort > ${OUTPUT_FILE}-cutoff.frequencia.csv
echo "Removendo arquivo '${OUTPUT_FILE}'"
rm -v ${OUTPUT_FILE}
#Fim do script

```

```

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id1;
my $id2;
my $score_normal;
my $pfam;
my $peimap;
my $psimap;
my $string;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @coluna = split (";", $linhas[$i]);
#pega a primeira coluna
$id1 = $coluna[0];
#pega a segunda coluna
$id2 = $coluna[1];
#pega a terceira coluna
$score_normal = $coluna[2];
#pega a quarta coluna
$pfam = $coluna[3];
#pega a quinta coluna
$peimap = $coluna[4];
#pega a sexta coluna
$psimap = $coluna[5];
#pega a sétima coluna
$string = $coluna[6];
if ($score_normal >= 0.7){
print "$coluna[3]\;$coluna[4]\;$coluna[5]\;$coluna[6]\n";
}
}
#Fim do script

```

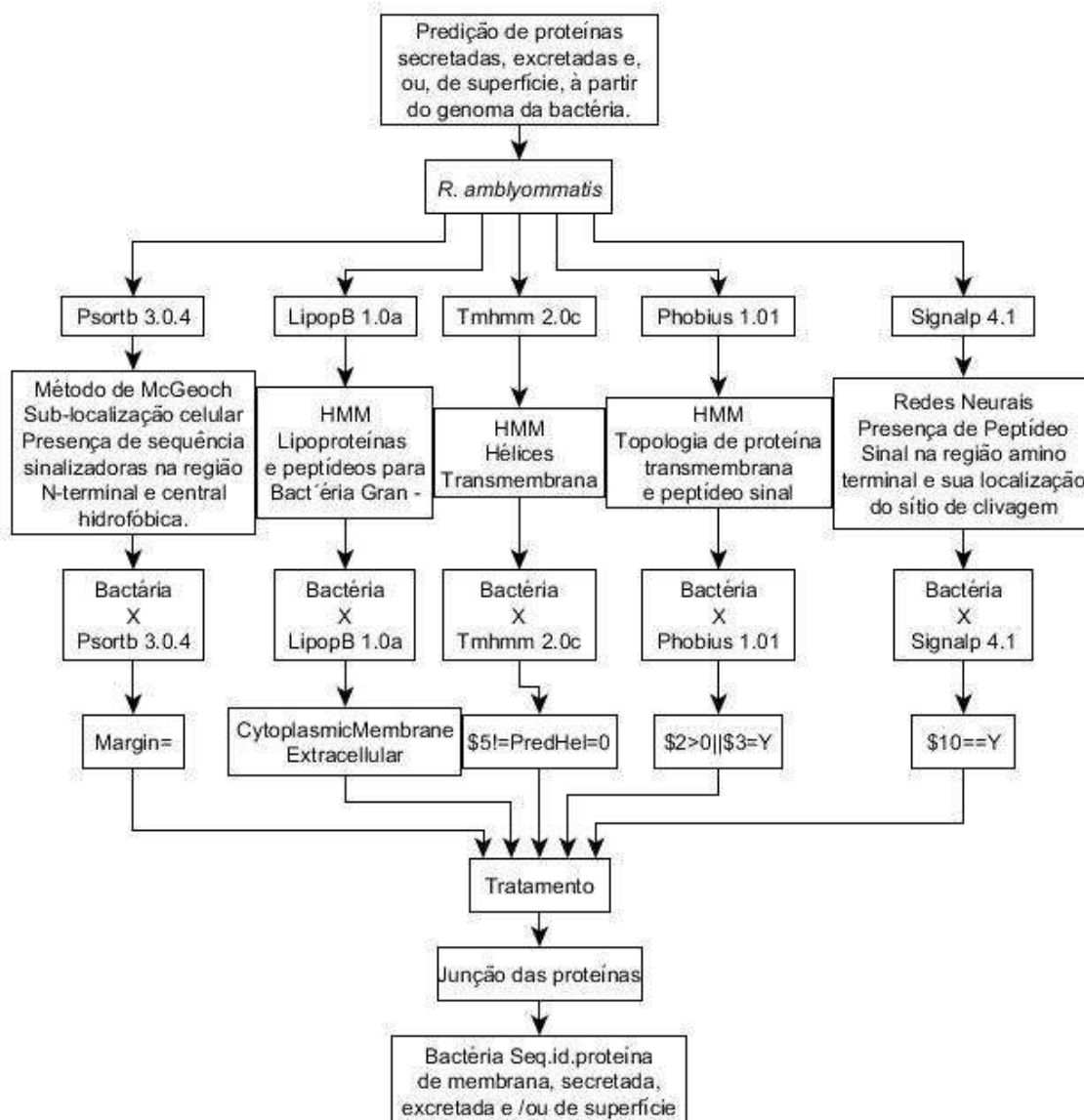
No *script* pbs, o arquivo 1 contém uma lista dos IDs para os pares de interação entre as proteínas com os valores do *score* combinado e a distribuição binomial da ocorrência da interação por banco do interatoma. No *script* perl, chamado dentro do *script* pbs, temos programado o corte de confiança para o *score* combinado sendo significativas as interações entre as proteínas de *R. amblyommatidis* e o *I. scapularis* apresentando o valor de cutoff $\geq 0,7$.

3.2 Identificação de potenciais proteínas de membrana

Visando a construção do interatoma entre as proteínas de membrana de *R. amblyommatidis* que interagem com as proteínas do *I. scapularis*, foram avaliadas todas as sequências de proteínas

da bactéria provenientes do genoma fazendo o uso de cinco aplicativos SignalP 4.1, Tmhmm 2.0c, LipopB 1.0a, Psortb 3.0.4 e Phobius 1.01 para a predição de proteínas secretadas, excretadas e/ou de superfície podendo ter sua sub-localização descrita (Figura 9).

Figura 9 Fluxograma das análises de predição para as proteínas de membrana provenientes do genoma da espécie *R. amblyommatidis*, sendo estas mesmas preditas como secretadas, excretadas e / ou de superfície.



Fonte: Autor deste trabalho

Na Figura 9, verifica-se o resumo dos passos realizados para as análises de predição de proteínas de membrana de *R. amblyommatidis* com o uso destes aplicativos, e conforme a utilidade de cada método, com destaque para os parâmetros utilizados.

SignalP

Com o aplicativo O SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP-4.1/>) (PETERSEN et al., 2011), que adota uma combinação de diferentes redes neurais, predizemos a presença do peptídeo sinal na região amino terminal e sua localização no sítio de clivagem que medeia a translocação de uma proteína do citoplasma para a membrana. O uso das redes neurais no SignalP 4.1, produzem três pontuações de saída para cada posição na sequência de entrada, onde temos: (i) o *score C*, que se refere ao valor de pontuação bruta do sítio de clivagem, (ii) o *score S*, que se refere ao valor de pontuação do peptídeo sinal e (iii) o *score Y*, que se refere a pontuação combinada do sítio de clivagem.

No arquivo de saída das redes neurais o parâmetro do *score C* tem a função de distinguir os locais de clivagem de peptídeos sinal de todo os demais, possuindo um valor alto numa posição após o local do sítio de clivagem localizada no primeiro resíduo na proteína madura. O parâmetro do *score S* tem a função de distinguir as posições de peptídeos sinal de posições sem peptídeos sinal na parte madura das proteínas. O parâmetro do *score Y* tem a função de uma combinação (média geométrica) do parâmetro do *score C* e a inclinação do parâmetro do *score S*, o que garante uma melhor predição do sítio de clivagem do que apenas o *score C* bruto.

No entanto, várias pontuações do *score C* com alto pico podem ser encontradas em uma sequência, onde apenas uma indica o verdadeiro local do sítio de clivagem. O *score Y* distingue os picos do *score C* escolhendo aquele em que a inclinação do *score S* é íngreme. A pontuação média do *score S* para o possível peptídeo sinal ocorre da posição 1 para a posição imediatamente antes da pontuação máxima do *score Y*. O *score D* (*escore* de discriminação) é uma média ponderada da média do *score S* e da máxima pontuação do *score Y*. Esta é a pontuação usada para discriminar peptídeos sinal de peptídeos não-sinal.

Para as proteínas não secretoras, todas as pontuações representadas na saída do SignalP 4.1 são idealmente baixas e próximas ao valor negativo de 0,1. Para mais informações e instruções a respeito do uso deste aplicativo, seu tutorial pode ser acessado (<http://www.cbs.dtu.dk/services/SignalP-4.1/instructions.php>). Neste sentido, com a finalidade de predizermos as sequências de proteínas de membrana de *R. amblyommatidis* para o uso deste aplicativo desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=4,mem=200mb
#PBS -m ae
#mensagem sera enviada: a abort; b begin; e end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo
module load signalp/4.1
signalp -f short -t gram- (arquivo 1a) > (arquivo 1b)
INPUT_FILE=(arquivo 1b)
OUTPUT_FILE="${INPUT_FILE}_id"
cat ${INPUT_FILE} | awk '{if($10=="Y")print$1}' > ${OUTPUT_FILE}
#Fim do script
```

Neste *script* pbs, o arquivo 1a contém às sequências de proteínas provenientes do genoma de *R. amblyommatidis*, submetidas a análise de proteínas de membrana através do software Signalp 4.1. O arquivo 1b final contém os IDs de identificação referentes às sequências de proteínas de membrana, obtidas por meio dos parâmetros do *score* Y, apresentando características de proteínas de membranas com a presença do peptídeo sinal na região amino terminal e sua localização do sítio de clivagem que media a translocação de uma proteína do citoplasma para a membrana.

TMHMM

Com o aplicativo TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (SONNHAMMER; KROGH, 1998), utilizando um modelo oculto de Markov (KROGH et al., 1994), identificamos potenciais hélices transmembrana em sequências proteicas. No arquivo de saída, uma linha é produzida para cada proteína sem gráficos. Cada linha começa com o identificador de sequência e, em seguida, os seguintes parâmetros: (i) (len) que se refere ao tamanho do comprimento da sequência de proteínas; (ii) (ExpAA), que se refere ao número esperado de hélices transmembranas de aminoácidos; (iii) (First60), que se refere ao número esperado de aminoácidos nas hélices transmembranas para os primeiros 60 aminoácidos da sequência da proteína; (iv) (PredHel), que se refere ao número de hélices transmembranas previstas por N-best, e (v) (Topologia), que se refere a topologia prevista pelo N-best.

A topologia é dada como a posição das hélices transmembranas separadas por 'i' se o laço “loop” estiver do lado de dentro ou 'o' se estiver do lado de fora. O exemplo 'i7-29o44-

66i87-109o' significa que ele começa por dentro, tem uma TMH preditas nas posições 7 a 29, do lado de fora e depois uma TMH na posição 44-66 etc. Os segmentos de TM preditos na região n-terminal, em algum momento, acabam sendo peptídeos sinal. Um dos erros mais comuns do programa é reverter a direção das proteínas em segmento da TM, não sendo útil para predizer se uma proteína não é de transmembrana, sendo citoplasmática ou não.

Para mais informações e instruções a respeito do uso deste aplicativo, o tutorial pode ser acessado no *link* (<http://www.cbs.dtu.dk/services/TMHMM/TMHMM2.0b.guide.php>). Neste sentido, com a finalidade de predizermos as sequências de proteínas de membrana de *R. amblyommatidis* para o uso deste aplicativo desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=200mb
#PBS -m ae
#mensagem sera enviada: a abort; b begin; e end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo
module load tmhmm/2.0c
tmhmm --short (arquivo 1a) > (arquivo 1b)
INPUT_FILE_a=(arquivo 1a)
OUTPUT_FILE_a="${INPUT_FILE_a}_id"
cat ${INPUT_FILE_a} | awk -F "\t" '{if($5!="PredHel=0")print$1}' > ${OUTPUT_FILE_a}
#Fim do script
```

Neste *script* pbs, o arquivo 1a contém às sequências de proteínas provenientes do genoma de *R. amblyommatidis*, submetidas a análise de proteínas de membrana através do aplicativo Tmhmm 2.0c. O arquivo 1b final contém os ids de identificação referentes às sequências de proteínas de membrana, obtidas por meio do parâmetro PredHel=0, que se refere ao número de hélices transmembranas preditas por *N-best*, apresentando características de potenciais hélices transmembrana.

LipoP

Com o aplicativo LipoP 1.0a (<http://www.cbs.dtu.dk/services/LipoP/>) (JUNCKER et al., 2003), que adotada o modelo oculto de Markov, foram preditas a presença de lipoproteínas e peptídeos para bactéria Gram-negativas. O aplicativo LipoP 1.0 prediz a presença de lipoproteínas e

discrimina entre peptídeos sinal de lipoproteína, de outros peptídeos sinal e hélices de membrana n-terminal em bactérias Gram-negativas.

Na primeira linha do arquivo de saída, como exemplo, tem-se o ID da proteína ANIA_NEIGO, SpII *score* = 29.6052, margem = 11.2327, clivagem = 18-19 e Pos + 2 = G é a melhor predição para uma lipoproteína com um local de clivagem entre o aminoácido 18 e 19 e o aminoácido G (glicina) na posição +2 após o local de clivagem. No arquivo de saída a segunda linha fornece o valor de corte adotado, no exemplo tem-se o Corte = -3. As colunas contêm: (i) o ID de identificação da sequência; (ii) o tipo de predição, sendo que o (Best means) significa a classe de pontuação mais alta, (Margin) indica a diferença entre os dois melhores *score*, (Class) classifica o *score* de outras classes e as linhas de sinal nas sequências de proteínas que contêm sítios de clivagem preditos.

Outros tipos de recursos, como (iii) a localização na sequência para linhas com uma predição de classe, é sempre 1. Para os sítios de clivagem, é o último aminoácido do peptídeo sinal em relação ao sítio de clivagem predito. A localização indica o primeiro aminoácido após o sítio de clivagem; (iv) o (*Score*) para (Margin), é a diferença entre os dois melhores *score* da classe. Para os locais de clivagem, o contexto ± 5 é mostrado após o #, e para os locais de clivagem de lipoproteínas, o aminoácido na posição +2 é mostrado o que pode determinar se a lipoproteína está ligada à membrana interna ou externa.

As quatro classes são preditas como: SpI o peptídeo sinal (peptidase sinal I), SpII: o peptídeo sinal de lipoproteína (peptidase sinal II), TMH a hélice transmembrana n-terminal. No entanto, essa não é uma predição muito confiável e deve ser testada. Esta parte do modelo existe principalmente para evitar que as hélices das membranas sejam falsamente preditas como peptídeos sinal. Por último, CYT citoplasmático, significando apenas todo o resto, apresentando sempre uma mesma pontuação. Estes sinais são preditos como CleavI, indicando os locais de clivagem para (sinal peptidase I); e CleavII indicando, os locais de clivagem para (sinal peptidase II).

Interpretando o arquivo de saída, este demonstra que a margem (Margin), é a diferença entre os dois melhores *score* da predição, correlacionando-se com o número de peptídeos sinal falsamente preditos. Para mais informações e instruções a respeito do uso deste aplicativo, o tutorial pode ser acessado no *link* (<http://www.cbs.dtu.dk/services/LipoP/LipoP1.0server.html>).

Neste sentido, com a finalidade de predizermos as sequências de proteínas de membrana de *R. amblyommatis* para o uso deste aplicativo desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do Job)
#PBS -l nodes=1:ppn=2,mem=100mb
#PBS -m ae
#mensagem sera enviada: a abort; b begin; e end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo
module load lipop/1.0a
LipoP (arquivo 1a) > (arquivo 1b)
INPUT_FILE_a=(arquivo 1b)
OUTPUT_FILE_a="${INPUT_FILE_a}_id"
cat ${INPUT_FILE_a} | grep 'margin' | grep 'KJV' | awk '{print $2}' > ${OUTPUT_FILE_a}
#Fim do script
```

Neste *script* pbs, o arquivo 1a contém as sequências de proteínas provenientes do genoma de *R. amblyommatis*, submetidas a análise de proteínas de membrana através do software Lipop 1.0a. O arquivo 1b final contém os IDs de identificação referentes às sequências de proteínas de membrana obtidas pelo parâmetro (Margin), que indica a diferença entre os dois melhores *score*, com características de lipoproteínas e possíveis peptídeos para bactéria Gram negativas.

Psort

Com o aplicativo Psortb 3.0.4 (<https://www.psort.org/psortb/>), que adota o método de McGeoch (MCGEOCH, 1985) modificado (NAKAI; KANEHISA, 1991, 1996), predizemos a sub-localização celular de proteínas por meio da presença das sequências sinalizadoras sugerindo a região carregada positivamente N-terminal (N- região) e a região central hidrofóbica (região H).

No arquivo de saída encontram-se delimitado por tabulações (formato conciso), o ID das sequências de proteínas, uma em cada linha, com 3 colunas: (i) SeqId contém as informações da linha de definição de arquivo FASTA, (ii) a localização contém a predição final do sítio de localização ou desconhecido, não sendo um sítio com pontuação acima de 7,5 e *score* contém o valor de confiança associado a este sítio de localização. Para mais informações

e instruções a respeito do uso deste aplicativo, o tutorial pode ser acessado no *link* (<https://www.psort.org/documentation/index.html#using>).

Neste sentido, com a finalidade de predizermos as sequências de proteínas de membrana de *R. amblyommatidis* para o uso deste aplicativo desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=200mb
#PBS -m ae
#mensagem sera enviada: a abort; b begin; e end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo
module load psortb/3.0.4
psort -p -o terse (arquivo 1a) > (arquivo 1b)
INPUT_FILE_a=s(arquivo 1b)
OUTPUT_FILE_a="${INPUT_FILE_a}_id"
cat ${INPUT_FILE_a} | grep 'CytoplasmicMembrane' | awk -F " " '{print$1}' > ${OUTPUT_FILE_a}
cat ${INPUT_FILE_a} | grep 'Extracellular' | awk -F " " '{print$1}' >> ${OUTPUT_FILE_a}
#Fim do script
```

Neste *script* pbs, o arquivo 1a contém às sequências de aminoácidos das proteínas em formato Fasta provenientes do genoma de *R. amblyommatidis*, submetidas a análise de proteínas de membrana através do software Psortb 3.0.4. O arquivo 1b final contém os IDs de identificação referentes às sequências de proteínas de membrana obtidas por meio dos parâmetros *CytoplasmicMembrane* e *Extracellular*, apresentando características de proteínas de membranas citoplasmáticas e extracelulares, com a sub-localização celular havendo a presença de sequências sinalizadoras sugerindo a região carregada positivamente N-terminal (N- região) e a região central hidrofóbica (região H).

Phobius

Com o aplicativo Phobius 1.01 (<http://phobius.sbc.su.se/>) que adota o modelo oculto de Markov, predizemos uma topologia de proteína transmembrana combinada a um preditor de peptídeo sinal numa série de estados interligados (KÄLL; KROGH; SONNHAMMER, 2004). O aplicativo Phobius 1.01 fornece como arquivo de saída uma lista da localização das hélices transmembranas preditas, a localização predita das regiões intermediárias de loop e o peptídeo sinal.

A predição fornece a localização e orientação mais prováveis das hélices transmembranas na sequência por meio de um algoritmo chamado *N-best* ou *I-best* que é a soma de todos os caminhos através do modelo com a mesma localização e direção das hélices preditas. Através do gráfico das probabilidades posteriores de peptídeo citoplasmático, não citoplasmático, hélice TM e peptídeo sinal. Pode-se ver possíveis hélices TM fracas que não foram preditas, e podemos ter uma ideia da certeza de cada segmento na predição.

No arquivo de saída no formato curto, uma linha é produzida para cada proteína sem gráficos. Cada linha começa com o ID da sequência e, em seguida, os seguintes campos: (i) TM contém o número de segmentos transmembranas preditos; (ii) SP contém os indicadores Y / N se um peptídeo sinal foi predito ou não, e (iii) a Predição que contém a topologia predita para a sequência da proteína

A topologia é dada como a posição das hélices transmembranas separadas por 'i' se o laço "loop" estiver do lado citoplasmático ou 'o' se estiver no lado não citoplasmático. Um peptídeo sinal é dado pela posição de sua região h separada por um e ac, e pela posição do último aminoácido no peptídeo sinal e o primeiro da proteína madura separado por um /. No exemplo, n4-19c24 / 25o219-238i250-269o281-302i322-342o372-391i422-439o451-476i significa que se prevê que a proteína contenha um peptídeo de sinal com uma região h entre a posição 4 e 19 que é clivada entre a posição 24 e 25. É seguida por uma alça não citoplasmática e um segmento da TM entre as posições 219 e 238, que é seguida por uma alça citoplasmática.

Para mais informações e instruções a respeito do uso deste aplicativo, o tutorial pode ser acessado no *link* (<http://phobius.sbc.su.se/instructions.html>). Neste sentido, com a finalidade de predizermos as sequências de proteínas de membrana de *R. amblyommatidis* para o uso deste aplicativo desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=2,mem=100mb
#PBS -m ae
#mensagem sera enviada: a abort; b begin; e end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo
module load phobius/1.01
phobius -short (arquivo 1a) > (arquivo 1b)
INPUT_FILE_a=(arquivo 1b)
```

```

OUTPUT_FILE_a="${INPUT_FILE_a}_id"
cat ${INPUT_FILE_a} | grep 'KJV' | awk '{if ($2>0||$3=Y) print $1}' > ${OUTPUT_FILE_a}
#Fim do script

```

Neste *script* pbs, o arquivo 1a contém as sequências de proteínas provenientes do genoma de *R. amblyommatis*, submetidas a análise de proteínas de membrana através do aplicativo Phobius 1.01. O arquivo 1b final contém os IDs referentes às sequências de proteínas de membrana, obtidas por meio dos parâmetros TM>0, que contém o número de segmentos transmembranas preditos e do SP=Y, indicando Y (YES) se um peptídeo sinal foi predito ou N (NO), caso não o seja, com características de topologia de proteína transmembrana combinada a um preditor de peptídeo sinal.

3.2.1 Integração das proteínas de membrana

As proteínas preditas de membrana de *R. amblyommatis* utilizando dos cinco aplicativos foram integradas e organizadas em arquivo único, contendo somente os IDs de identificação e sem repetição destes. Neste sentido, com a finalidade de integrarmos as sequências de proteínas de membrana de *R. amblyommatis* desenvolvemos o *script* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=500mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
INPUT_FILE_1a="(arquivo 1a).ambly_lipop_id"
INPUT_FILE_1b="(arquivo 1b).ambly_phobius_id"
INPUT_FILE_1c="(arquivo 1c).ambly_psort_id"
INPUT_FILE_1d="(arquivo 1d).ambly_signalp_id"
INPUT_FILE_1e="(arquivo 1e).ambly_tmhmm_id"
OUTPUT_FILE_2a="(arquivo de saída 2a).ambly_protein-combinado"
OUTPUT_FILE_2b="(arquivo de saída 2b).ambly_protein-combinado_id"
cat ${INPUT_FILE_1a} ${INPUT_FILE_1b} ${INPUT_FILE_1c} ${INPUT_FILE_1d}
${INPUT_FILE_1e} | awk '{print$0}' > ${OUTPUT_FILE_2a}
cat ${OUTPUT_FILE_2a} | sort | uniq > ${OUTPUT_FILE_2b}
#fim do script

```

Neste *script* pbs, os arquivos 1a-e contém os IDs de identificação das sequências de proteínas de membrana para a bactéria da espécie *R. amblyommatis* obtidas com o uso dos aplicativos SignalP 4.1, Tmhmm 2.0c, Lipop1.0a, Psortb 3.0.4 e Phobius 1.01. O arquivo de

saída 2a contém os IDs totais das sequências de proteínas de membrana, enquanto o arquivo de saída 2b contém os IDs únicos para estas sequências preditas.

3.2.2 Interatoma de proteínas de membrana

Para a construção do interatoma de proteínas de membrana, foi realizado um filtro utilizando das proteínas preditas e integradas de membrana de *R. amblyommatis* contra o interatoma construído no Tópico , contendo os pares de interação entre a bactéria e o carrapato obtido após o cálculo do *score* combinado com valores de cutoff $\geq 0,7$. Neste sentido, com a finalidade de construirmos o interatoma com as proteínas de membrana preditas da bactéria que interagem com as sequências de proteínas do carrapato desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=500mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
#ambly
INPUT_FILE_1a="(Arquivo 1) saida_rick.ambly_protein-combinado_id"
INPUT_FILE_1b="(Arquivo 2) rede_id.ambly.vs.ixodes_teste2"
OUTPUT_FILE_1c="(Arquivo 3a) rede_id.ambly.vs.ixodes_teste2_protein.membrane"
OUTPUT_FILE_1e="(Arquivo 3b) rede_id.ambly.vs.ixodes_teste2_protein.membrane-ambly-no-red"
OUTPUT_FILE_1f="(Arquivo 3c) rede_id.ambly.vs.ixodes_teste2_protein.membrane-ixodes-no-red"
perl script.protein.membrane.vs.interatoma.pl ${INPUT_FILE_1a} ${INPUT_FILE_1b} >
${OUTPUT_FILE_1c}
cat ${OUTPUT_FILE_1c} | grep "KJV" | awk -F ";" '{print $1}' | sort | uniq > ${OUTPUT_FILE_1e}
cat ${OUTPUT_FILE_1c} | grep "ISCW" | awk -F ";" '{print $2}' | sort | uniq > ${OUTPUT_FILE_1f}
#Fim!

#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
#arquivo sem redundância localizacao_subcelular_rick "rick.ambly.localizacao.id.final-no-red"
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
#arquivo completo interatoma "saida_ambly_vs_ixodes_score_combinado.2col-id.cutoff.inter.csv"
my $file1 = $ARGV[1];
open ARQUIVO,"<$file1";
my @linhas1 = <ARQUIVO>;
```

```

close ARQUIVO;
my $id_rick;
my $id_car;
#Percorre localizacao_subcelular rick n.red
for (my $i = 0; $i <= $#linhas; $i++){ #analisa cada linha do arquivo sem redundância
chomp ($linhas[$i]);
my @cols = split (/s+/, $linhas[$i]);
#print "$cols[0]\n";
#Pecorre interação completo interatoma
for (my $t = 0; $t <= $#linhas1; $t++){ #analisa cada linha do arquivo completo
chomp ($linhas1[$t]);
my @cols1 = split (";", $linhas1[$t]);
#pega a primeira coluna
$id_rick = $cols1[0];
#pega a segunda coluna
$id_car = $cols1[1];
#print "$cols1[1]\n";
if (($cols1[0] eq ($cols[0]))){
print "$cols[0]\;$cols1[1]\n";
}
}
}
}
#Fim do script

```

No *script* pbs, o arquivo 1 contém os IDs das sequências de proteínas preditas de membrana à partir do genoma de *R. amblyommatis*, obtidas com o uso dos aplicativos SignalP 4.1, Tmhmm 2.0c, Lipop1.0a, Psortb 3.0.4 e Phobius 1.01, respectivamente. Enquanto, o arquivo 2 contém a rede do interatoma geral construído para as proteínas de *R. amblyommatis* que interagem com as proteínas do *I. scapularis*. No *script* perl, chamado dentro do *script* pbs, temos programado o filtro do arquivo 1 no arquivo 2 para formarem os arquivos de saída 3a-c. O arquivo de saída 3a contém os pares de interação com os IDs totais das sequências de proteínas de membrana da bactéria que interagem com as sequências de proteínas do carrapato, enquanto os arquivos de saída 3b e 3c contém separadamente os IDs únicos para as sequências de proteínas de cada organismos, sendo bactéria e carrapato, respectivamente.

3.3 Detecção de proteínas ortólogas

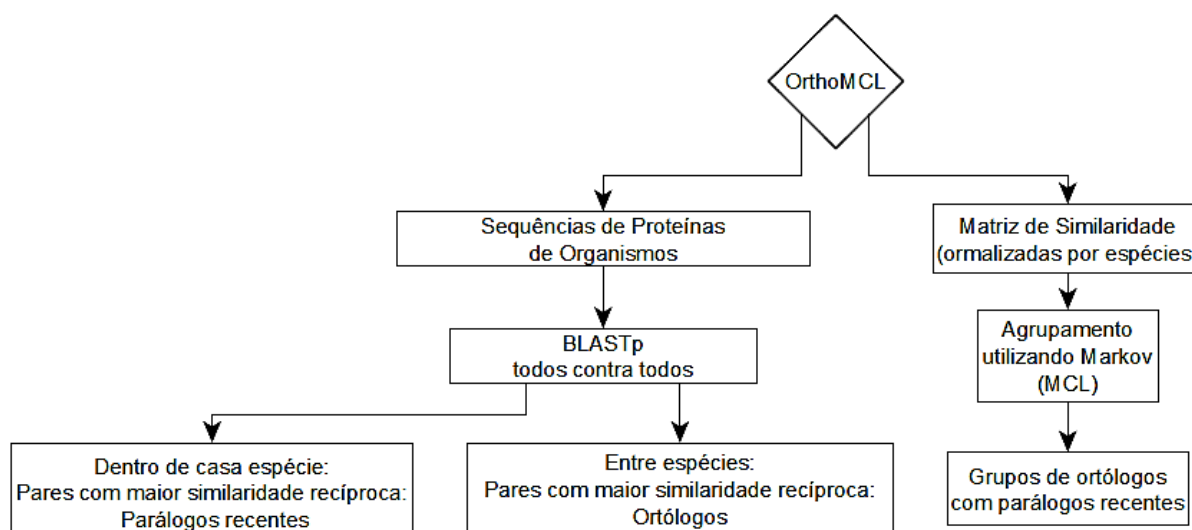
Para predizermos as proteínas ortólogas entre *R. amblyommatis* versus outras *Rickettsia* spp. que infectam carrapatos e especificamente o *I. scapularis* e espécies que infectam insetos, visando a construção do interatoma de proteínas de membrana ortólogas fizemos o uso do aplicativo OrthoMCL 2.0 (<https://orthomcl.org/orthomcl/>) (LI, LI.; CHRISTIAN, J. STOECKERT, JR., DAVID, 2003). Esta análise foi implementada com o uso do banco de dados MySQL utilizando de vários BLASTP em escalas genômicas, com uma combinação da abordagem de *bidirectional best hit* (OVERBEEK et al., 1999) e algoritmo de *Markov cluster*

(DONGEN, 2000) que foi extensivamente validado para aplicação em genomas de eucariotos (FENG, CHEN, AARON. J, MACKEY, CHRISTIAN. J, STOECKERT JR, DAVID. S, 2006).

O OrthoMCL 2.0 é um software desenvolvido para realizar o agrupamento de sequências de proteínas ortólogas e parálogas em escala genômica e determinar um perfil filogenético dos organismos pesquisados, sendo uma ferramenta bastante experimentada para anotação de genomas eucarióticos cujo algoritmo de análise é utilizado para o agrupamento de sequências de proteínas previamente processadas. Este algoritmo identifica os melhores *best hits* recíprocos entre pares, dentro do mesmo genoma, como potenciais parálogos e melhores *best hits* recíprocos em dois genomas como potenciais pares ortólogos, de forma que as proteínas relacionadas estão interligadas por um gráfico de similaridade.

Primeiro, foi realizado o alinhamento local de todas as sequências de proteínas contra todas as proteínas de alta qualidade do conjunto de organismos usando o programa BLASTp (ALTSCHUL et al., 1997). O arquivo de saída do BLAST foi utilizado como entrada do programa OrthoMCL 2.0 (Figura 10).

Figura 10 Fluxograma representativo para o uso do aplicativo OrthoMCL 2.0 e análise de sequências de proteínas ortógas entre as proteínas provenientes do genoma de *R. amblyommatis* e *Rickettsia* spp., que infectam carrapatos e insetos com destaque para os passos e os parâmetros adotados.



Fonte: Autor deste trabalho.

Na Figura 10, verifica-se um fluxograma para o uso do aplicativo OrthoMCL 2.0 e os seguintes passos: (i) sendo baseados nos pesos entre cada par de proteínas e por isso são normalizados antes da execução do MCL e (ii), de modo a corrigir as diferenças na distância

evolutiva. Este processo é análogo à avaliação manual dos agrupamentos ortólogos de grupos de proteínas (COGs).

Para identificação de potenciais proteínas ortólogas em *Rickettsia* spp. que infectam carrapatos versus insetos, inicialmente com o uso do programa OrthoMCL, pôde-se aferir o número dos grupos de ortólogos identificados, compartilhados e específicos para as espécies que infectam estes diferentes grupos de artrópode hematófagos, entre as proteínas provenientes dos genomas de seis espécies de *Rickettsia*. Entre as espécies de *Rickettsia* utilizadas para esta análise, quatro possuem a capacidade de infectar carrapatos e duas de infectar insetos.

A distribuição de grupos de ortólogos foi avaliada baseada no perfil de infectividade para carrapatos e entre espécies da bactéria para avaliar o impacto do ambiente onde os parasitos se desenvolvem no hospedeiro artrópode hematófago na evolução do conteúdo proteico dos organismos e conseqüentemente identificar as sequências de proteínas que potencialmente participam de mecanismos de infecção.

Para mais informações e instruções a respeito do uso deste aplicativo, o tutorial pode ser acessado no *link* (<https://orthomcl.org/common/downloads/software/v2.0/UserGuide.txt>). Neste sentido, com a finalidade de predizermos a ocorrência de sequência de proteínas ortólogas entre a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam carrapatos e insetos desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python
blastall -p blastp -i goodProteins.Fasta -d (arquivo pasta ortologos) -a 3 -m 8 -o blast.final
#-p (proteinas)
#-i (arquivo de entrada)
#-d (database formatado)
#-a 3 (número de processadores. Quando coloca o -a tem que colocar o mesmo número de processadores)
#-m 8 (saída tabulada)
#Fim do script
```

Neste *script* pbs, o arquivo goodProteins.fast contém as sequências de proteínas provenientes dos genomas das espécies de *Rickettsia*, submetidas inicialmente ao aplicativo

OrthoMcl 2.0. O arquivo de saída contém o resultado do BLASTp final para as sequências ortólogas contendo o arquivo grupos.txt dentro do arquivo de pasta console ortólogos.

3.3.1 Clusters de proteínas ortólogas de *Rickettsia* spp. que infectam carrapatos versus insetos

Neste sentido, com a finalidade de dividirmos e agruparmos as sequências de proteínas ortólogas entre a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam carrapatos e insetos desenvolvemos o *script* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome job)
#PBS -l nodes=1:ppn=1,mem=1gb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE_1a="(Arquivo 1a) grupo.txt"
OUTPUT_FILE_1b="(Arquivo 1b) saida_rick.car_vs_ixodes_cluster_ortologos_id-2col"
OUTPUT_FILE_1c="(Arquivo 1c) saida_rick.ins_vs_ixodes_cluster_ortologos_id-2col"
#carrapatos
cat ${INPUT_FILE_1a} | grep "amb" | grep "end" | grep "buc" | grep "mon" | grep -v "fel" | grep -v "pro" >
${OUTPUT_FILE_1b}
#insetos
cat ${INPUT_FILE_1a} | grep "fel" | grep "pro" | grep "amb" | grep -v "end" | grep -v "buc" | grep -v "mon" >
${OUTPUT_FILE_1c}
#Fim do script
```

Neste *script* pbs, o arquivo 1a contém os grupos de ortólogos com os IDs das sequências de proteínas provenientes dos genomas de *R. amblyommatis* e de outras cinco espécies de *Rickettsia*, obtidas com o uso do aplicativo OrthoMcl 2.0. Para a extração dos clusters contendo separadamente as proteínas das espécies de *Rickettsia*, que infectam os carrapatos em relação as que infectam os insetos, utilizamos dos códigos de identificação fornecidos para o conjunto de proteínas analisadas na análise de ortologia inicial conforme o nome das bactérias: *R. amblyommatis* (amb), *R. endonsymbiontic* *I. scapularis* (end), *R. buchannerii* (buc), *R. monanensce* (mon), *R. felis* (fel) e *R. prowazekii* (pro).

O arquivo de saída 1b contém os grupos de ortólogos e os IDs totais das sequências de proteínas de *Rickettsia* spp. que infectam somente os carrapatos, enquanto o arquivo de saída

Ic contém os grupos e os IDs totais das sequências de proteínas de *Rickettsia* spp. que infectam em parte, os carrapatos e insetos.

3.3.2 Interatoma de proteínas de membrana e ortólogos de *Rickettsia* spp. que infectam carrapatos

Para a construção do interatoma das proteínas de membrana que possuem grupos de ortólogos entre as espécies de *Rickettsia* que infectam carrapatos versus as espécies que infectam insetos, foram realizados filtros utilizando destes grupos de ortólogos preditos contra o interatoma de proteínas de membrana construído para a bactéria da espécie *R. amblyommatis* versus o carrapato da espécie *I. scapularis*, conforme demonstrado no Tópico 3.3.1. Neste sentido, com a finalidade de construirmos o interatoma de proteínas de membrana de ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam em especial os carrapatos, desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=1gb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
INPUT_FILE_1a="(Arquivo 1) "
INPUT_FILE_1b="(Arquivo 2) "
OUTPUT_FILE_1c="(Arquivo 3) membrane_cluster.ortologos_4col_id"
perl script.rick.car.ortologos.pl ${INPUT_FILE_1a} ${INPUT_FILE_1b} > ${OUTPUT_FILE_1c}
#fim do script
```

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO1,"<$file";
my @linhas1 = <ARQUIVO1>;
close ARQUIVO1;
my $file1 = $ARGV[1];
open ARQUIVO2,"<$file1";
my @linhas2 = <ARQUIVO2>;
close ARQUIVO2;
for (my $i = 0; $i <= $#linhas1; $i++){
chomp ($linhas1[$i]);
```

```

my @cols1 = split (/\;/, $linhas1[$i]);
#print "$cols1[0]\n";
for (my $j = 0; $j <= $#linhas2; $j++){
chomp ($linhas2[$j]);
my @cols2 = split (/\s+/, $linhas2[$j]);
#print "$cols2[2]\n";
for (my $x = 1; $x <= $#cols2; $x++){
my @cols3 = split (/\|/, $cols2[$x]);
#print "$cols3[1]\n";
if ($cols3[1] eq $cols1[0]){
print "$cols2[0]\t$cols1[0]\t$cols1[1]\n";
}
}
}
}
}
#fim do script

```

No *script* pbs, o arquivo 1 contém os pares de IDs da rede do interatoma de membrana construído para as proteínas de *R. amblyommatis* que interagem com as proteínas do *I. scapularis*. Enquanto, o arquivo 2 contém os grupos de ortólogos preditos para as proteínas de *R. amblyommatis* e outras *Rickettsia* spp. que infectam somente os carrapatos. No *script* perl, chamado dentro do *script* pbs, temos programa o filtro do arquivo 1 no arquivo 2 para formar o arquivo de saída 3. O arquivo de saída 3a contém os pares de interação com os IDs das sequências de proteínas de membrana da bactéria que interagem com as sequências de proteínas do carrapato, juntamente com o grupos ortólogos e os IDs das sequências de proteínas das outras espécies de *Rickettsia* que infectam os carrapatos.

3.3.3 Alinhamentos das proteínas ortólogas

Entre as espécies que infectam carrapatos, os cluster 1050 e 1816 foram os únicos presentes na rede do deste interatoma de membrana. Estas sequências de proteínas de membrana ortólogas foram submetidas ao alinhamento múltiplos de sequências ClustalW (<https://www.genome.jp/tools-bin/clustalw>) (LARKIN et al., 2007) e o Clustal Ômega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (SIEVERS et al., 2011) para estabelecer relações filogenéticas entre diferentes proteínas. Com isto, podemos predizer domínios conservados, incluindo resíduos críticos para a função da proteína e comparar a proteína de interesse de forma mais detalhada com membros da mesma família desta.

Para isto, fizemos o uso do alinhamento progressivo (Clustal), sendo este um método heurístico desenvolvido de modo a fazer um alinhamento progressivo de sequências baseado na sua proximidade em uma árvore filogenética baseada em alinhamentos par a par. Apesar

deste tipo de método gerar bons alinhamentos em tempo razoável, ele não garante que o melhor alinhamento possível será obtido. Este alinhamento global é útil quando as sequências têm tamanhos próximos.

3.3.4 Análise evolutiva das proteínas de membrana ortólogas

TBLASTn

As sequências de proteínas de membrana ortólogas para espécies de *Rickettsia* que infectam carrapatos foram submetidas ao alinhamento local com adoção do TBLASTn contra as sequências de aminoácidos proveniente dos genomas das espécies de *R. amblyommatis* além de outras que constituíram para a construção do interatoma de proteínas de membrana ortólogas.

Neste sentido, com a finalidade de avaliarmos quanto as características evolutivas das proteínas de membrana de ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam carrapatos, além das que espécies que infectam os insetos desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python
#formatdb -i Rickettsia_amblyommatis_str_ac_pa.ASM96467v1.dna.toplevel.fa -p F
#formatdb -i Rickettsia_prowazekii_str_breinl.ASM36740v1.dna.toplevel.fa -p F
#formatdb -i Rickettsia_monacensis.RMONA_1.dna.toplevel.fa -p F
#formatdb -i Rickettsia_felis_str_pedreira.ASM96466v1.dna.toplevel.fa -p F
#formatdb -i Rickettsia_endosymbiont_of_ixodes_scapularis.ASM16073v1.dna.toplevel.fa -p F
#formatdb -i Rickettsia_buchneri.REISMNv1.dna.toplevel.fa -p F
#formatdb -i seq_cluster1050_amb -p T
#formatdb -i seq_cluster1816_amb -p T
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2a) -a 6 -b 1 -F T -m 8 -o (arquivo 3a)
blastall -p tblastn -i (arquivo 1b) -d (arquivo 2a) -a 6 -b 1 -F T -m 8 -o (arquivo 3b)
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2b) -a 6 -b 1 -F T -m 8 -o (arquivo 3c)
blastall -p tblastn -i (arquivo 1b) -d (arquivo 2b) -a 6 -b 1 -F T -m 8 -o (arquivo 3d)
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2c) -a 6 -b 1 -F T -m 8 -o (arquivo 3e)
blastall -p tblastn -i (arquivo 1b) -d (arquivo 2c) -a 6 -b 1 -F T -m 8 -o (arquivo 3f)
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2d) -a 6 -b 1 -F T -m 8 -o (arquivo 3g)
blastall -p tblastn -i (arquivo 1b) -d (arquivo 2d) -a 6 -b 1 -F T -m 8 -o (arquivo 3h)
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2e) -a 6 -b 1 -F T -m 8 -o (arquivo 3i)
```

```

blastall -p tblastn -i (arquivo 1b) -d (arquivo 2e) -a 6 -b 1 -F T -m 8 -o (arquivo 3j)
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2f) -a 6 -b 1 -F T -m 8 -o (arquivo 3k)
blastall -p tblastn -i (arquivo 1b) -d (arquivo 2f) -a 6 -b 1 -F T -m 8 -o (arquivo 3l)
#-n (nucleotídeo - bases nitrogenadas)
#-p (proteínas - aminoácido)
#-i (arquivo de entrada)
#-d (database formatado)
#-a 3 (número de processadores. Quando coloca o -a tem que colocar o mesmo número de processadores)
#-m 8 (saída tabulada)
#Fim do script

```

Neste *script* pbs, os arquivos 1a e 1b se referem as sequências de proteínas agrupadas nos clusters ortólogos 1050 e 1816, respectivamente. Estas sequências de aminoácidos de proteínas de membrana e ortólogas para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam os carrapatos foram submetidas a análise de similaridade contra os arquivos 2a-f das sequências de nucleotídeos provenientes dos genomas das mesmas seis espécies de *Rickettsia*, utilizando do alinhamento adotado pelo TBLASTn. Os pares de arquivos de saída 3a e 3b, 3c e 3d, 3e e 3f, 3g e 3h, 3i e 3j, 3k e 3l, contém os resultados do alinhamento para as seis espécies de *Rickettsia* separadas e agrupadas em ambos os cluster 1050 e 1816, respectivamente.

Tratamento TBLASTn

Com a finalidade de avaliarmos as características evolutivas das proteínas de membrana de ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam carrapatos, além das que espécies que infectam os insetos, após a análise de similaridade com o uso do TBLASTn, foram adotados os dois valores de $> \textit{bit score}$. Para isto desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
#Carrega o módulo opt-python
INPUT_FILE_1="(arquivo 3a)"
INPUT_FILE_2="(arquivo 3b)"
INPUT_FILE_3="(arquivo 3c)"
INPUT_FILE_4="(arquivo 3d)"
INPUT_FILE_5="(arquivo 3e)"

```

```

INPUT_FILE_6="(arquivo 3f)"
INPUT_FILE_7="(arquivo 3g)"
INPUT_FILE_8="(arquivo 3h)"
INPUT_FILE_9="(arquivo 3i)"
INPUT_FILE_10="(arquivo 3j)"
INPUT_FILE_11="(arquivo 3k)"
INPUT_FILE_12="(arquivo 3l)"
#cluster 1050
OUTPUT_FILE_1="(arquivo 3a)"
OUTPUT_FILE_2="(arquivo 3c)"
OUTPUT_FILE_3="(arquivo 3e)"
OUTPUT_FILE_4="(arquivo 3g)"
OUTPUT_FILE_5="(arquivo 3i) "
OUTPUT_FILE_6="(arquivo 3k)"
OUTPUT_FILE_a="(arquivo 4a)"
#cluster 1816
OUTPUT_FILE_1="(arquivo 3b)"
OUTPUT_FILE_2="(arquivo 3d)"
OUTPUT_FILE_3="(arquivo 3f)"
OUTPUT_FILE_4="(arquivo 3h)"
OUTPUT_FILE_5="(arquivo 3j) "
OUTPUT_FILE_6="(arquivo 3l)"
OUTPUT_FILE_b="(arquivo 4b)"
perl script.blast.cutoof.pl ${INPUT_FILE_1} > ${OUTPUT_FILE_1}
perl script.blast.cutoof.pl ${INPUT_FILE_2} > ${OUTPUT_FILE_2}
perl script.blast.cutoof.pl ${INPUT_FILE_3} > ${OUTPUT_FILE_3}
perl script.blast.cutoof.pl ${INPUT_FILE_4} > ${OUTPUT_FILE_4}
perl script.blast.cutoof.pl ${INPUT_FILE_5} > ${OUTPUT_FILE_5}
perl script.blast.cutoof.pl ${INPUT_FILE_6} > ${OUTPUT_FILE_6}
cat ${OUTPUT_FILE_1} ${OUTPUT_FILE_2} ${OUTPUT_FILE_3} ${OUTPUT_FILE_4}
${OUTPUT_FILE_5} ${OUTPUT_FILE_6} | sort | uniq >> ${OUTPUT_FILE_a}
perl script.blast.cutoof.pl ${INPUT_FILE_7} > ${OUTPUT_FILE_7}
perl script.blast.cutoof.pl ${INPUT_FILE_8} > ${OUTPUT_FILE_8}
perl script.blast.cutoof.pl ${INPUT_FILE_9} > ${OUTPUT_FILE_9}
perl script.blast.cutoof.pl ${INPUT_FILE_10} > ${OUTPUT_FILE_10}
perl script.blast.cutoof.pl ${INPUT_FILE_11} > ${OUTPUT_FILE_11}
perl script.blast.cutoof.pl ${INPUT_FILE_12} > ${OUTPUT_FILE_12}
cat ${OUTPUT_FILE_7} ${OUTPUT_FILE_8} ${OUTPUT_FILE_9} ${OUTPUT_FILE_10}
${OUTPUT_FILE_11} ${OUTPUT_FILE_12} | sort | uniq >> ${OUTPUT_FILE_b}
#fim do script

```

```

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
my $confianca;
my $identidade;
my $idblast;
my $value;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas [$i]);
my @cols1 = split ("\t", $linhas[$i]);
#print "$cols1[0]\n";
#pega a primeira coluna

```



```

Sid = $cols1[0];
#pega a segunda coluna
Sidblast = $cols1[1];
#pega a quarta coluna
$confianca = $cols1[2];
#pega a decima primeira
$evalue = $cols1[10];
#pega a decima segunda coluna
$identidade = $cols1[11];
if ($cols1[2] >= 70 && $cols1[10] <= 0.0001){
print "$cols1[0]\t$cols1[1]\t$cols1[2]\t$cols1[10]\n";
}
}
}

```

No *script* pbs, encontram as entradas para os pares de arquivos 3ab-kl, que se referem aos resultados dos alinhamentos do TBLASTn para a similaridade das sequências de proteínas de membrana ortólogas agrupadas nos clusters 1050 e 1816 contra as sequências de genes provenientes dos genomas para *R. amblyommatis* e outras cinco espécies de *Rickettsia*. No *script* perl, chamado dentro do *script* pbs, foi programado para filtrar as sequências similares adotando os dois valores de *> bit score*. Nos arquivos de saída para os pares 3ab-kl, encontram-se estes resultados separadamente para cada espécie de *Rickettsia*. Os arquivos 4a e 4b correspondem ao conjunto destas sequências de nucleotídeos similares para cada conjunto de grupo ortólogo, sendo os cluster 1050 e 1816, respectivamente.

BLASTp

Além disso, estes ortólogos foram submetidos ao BLASTP contra as sequências de aminoácidos de proteínas provenientes dos genomas das mesmas bactérias, além das sequências de proteínas do interatoma de proteínas de membrana para a espécie *R. amblyommatis*. Neste sentido, com a finalidade de avaliarmos quanto as características evolutivas das proteínas de membrana de ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam carrapatos, além das que espécies que infectam os insetos desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python

```

```

blastall -p blastp -i (arquivo 1a) -d (arquivo 2a) -a 6 -b 1 -F T -m 8 -o (arquivo 3a)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2a) -a 6 -b 1 -F T -m 8 -o (arquivo 3b)
blastall -p blastp -i (arquivo 1a) -d (arquivo 2b) -a 6 -b 1 -F T -m 8 -o (arquivo 3c)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2b) -a 6 -b 1 -F T -m 8 -o (arquivo 3d)
blastall -p blastp -i (arquivo 1a) -d (arquivo 2c) -a 6 -b 1 -F T -m 8 -o (arquivo 3e)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2c) -a 6 -b 1 -F T -m 8 -o (arquivo 3f)
blastall -p blastp -i (arquivo 1a) -d (arquivo 2d) -a 6 -b 1 -F T -m 8 -o (arquivo 3g)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2d) -a 6 -b 1 -F T -m 8 -o (arquivo 3h)
blastall -p blastp -i (arquivo 1a) -d (arquivo 2e) -a 6 -b 1 -F T -m 8 -o (arquivo 3i)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2e) -a 6 -b 1 -F T -m 8 -o (arquivo 3j)
blastall -p blastp -i (arquivo 1a) -d (arquivo 2f) -a 6 -b 1 -F T -m 8 -o (arquivo 3k)
blastall -p blastp -i (arquivo 1b) -d (arquivo 2f) -a 6 -b 1 -F T -m 8 -o (arquivo 3l)
#-n (nucleotídeo - bases nitrogenadas)
#-p (proteínas - aminoácido)
#-i (arquivo de entrada)
#-d (database formatado)
#-a 3 (número de processadores. Quando coloca o -a tem que colocar o mesmo número de processadores)
#-m 8 (saída tabulada)
#Fim do script

```

Neste *script* pbs, os arquivos 1a e 1b contém as sequências de proteínas agrupadas no cluster 1050 e 1816, respectivamente. Estas sequências de aminoácidos de proteínas de membrana e ortólogas para a espécie *R. amblyommatis* e as outras mesmas quatro *Rickettsia* spp. que infectam carrapatos foram submetidas a análise de similaridade contra os arquivos 2a-f das sequências de proteínas provenientes dos genomas de seis espécies de *Rickettsia*, utilizando do alinhamento adotado pelo BLASTp. Os pares de arquivos de saída 3a e 3b, 3c e 3d, 3e e 3f, 3g e 3h, 3i e 3j, 3k e 3l, contém os resultados do alinhamento para as seis espécies de *Rickettsia* separadas e agrupadas em ambos os cluster 1050 e 1816, respectivamente.

Tratamento BLASTp

Com a finalidade de avaliarmos quanto as características evolutivas das proteínas de membrana de ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam carrapatos, além das que espécies que infectam os insetos, após a análise de similaridade com o uso do BLASTp, foram adotados o nível de confiança ≥ 70 e identidade 0,0001. Para isto desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR

```

```

#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
#Carrega o módulo opt-python
INPUT_FILE_1="(arquivo 3a)"
INPUT_FILE_2="(arquivo 3b)"
INPUT_FILE_3="(arquivo 3c)"
INPUT_FILE_4="(arquivo 3d)"
INPUT_FILE_5="(arquivo 3e)"
INPUT_FILE_6="(arquivo 3f)"
INPUT_FILE_7="(arquivo 3g)"
INPUT_FILE_8="(arquivo 3h)"
INPUT_FILE_9="(arquivo 3i)"
INPUT_FILE_10="(arquivo 3j)"
INPUT_FILE_11="(arquivo 3k)"
INPUT_FILE_12="(arquivo 3l)"
#cluster 1050
OUTPUT_FILE_1="(arquivo 3a)"
OUTPUT_FILE_2="(arquivo 3c)"
OUTPUT_FILE_3="(arquivo 3e)"
OUTPUT_FILE_4="(arquivo 3g)"
OUTPUT_FILE_5="(arquivo 3i)"
OUTPUT_FILE_6="(arquivo 3k)"
OUTPUT_FILE_a="(arquivo 4a)"
#cluster 1816
OUTPUT_FILE_1="(arquivo 3b)"
OUTPUT_FILE_2="(arquivo 3d)"
OUTPUT_FILE_3="(arquivo 3f)"
OUTPUT_FILE_4="(arquivo 3h)"
OUTPUT_FILE_5="(arquivo 3j)"
OUTPUT_FILE_6="(arquivo 3l)"
OUTPUT_FILE_b="(arquivo 4b)"
perl script.blast.cutoof.pl ${INPUT_FILE_1} > ${OUTPUT_FILE_1}
perl script.blast.cutoof.pl ${INPUT_FILE_2} > ${OUTPUT_FILE_2}
perl script.blast.cutoof.pl ${INPUT_FILE_3} > ${OUTPUT_FILE_3}
perl script.blast.cutoof.pl ${INPUT_FILE_4} > ${OUTPUT_FILE_4}
perl script.blast.cutoof.pl ${INPUT_FILE_5} > ${OUTPUT_FILE_5}
perl script.blast.cutoof.pl ${INPUT_FILE_6} > ${OUTPUT_FILE_6}
cat ${OUTPUT_FILE_1} ${OUTPUT_FILE_2} ${OUTPUT_FILE_3} ${OUTPUT_FILE_4}
${OUTPUT_FILE_5} ${OUTPUT_FILE_6} | sort | uniq >> ${OUTPUT_FILE_a}
perl script.blast.cutoof.pl ${INPUT_FILE_7} > ${OUTPUT_FILE_7}
perl script.blast.cutoof.pl ${INPUT_FILE_8} > ${OUTPUT_FILE_8}
perl script.blast.cutoof.pl ${INPUT_FILE_9} > ${OUTPUT_FILE_9}
perl script.blast.cutoof.pl ${INPUT_FILE_10} > ${OUTPUT_FILE_10}
perl script.blast.cutoof.pl ${INPUT_FILE_11} > ${OUTPUT_FILE_11}
perl script.blast.cutoof.pl ${INPUT_FILE_12} > ${OUTPUT_FILE_12}
cat ${OUTPUT_FILE_7} ${OUTPUT_FILE_8} ${OUTPUT_FILE_9} ${OUTPUT_FILE_10}
${OUTPUT_FILE_11} ${OUTPUT_FILE_12} | sort | uniq >> ${OUTPUT_FILE_b}
#fim do script

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;

```

```

my $confianca;
my $identidade;
my $idblast;
my $value;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas [$i]);
my @cols1 = split ("\t", $linhas[$i]);
#print "$cols1[0]\n";
#pega a primeira coluna
$id = $cols1[0];
#pega a segunda coluna
$idblast = $cols1[1];
#pega a quarta coluna
$confianca = $cols1[2];
#pega a decima primeira
$value = $cols1[10];
#pega a decima segunda coluna
$identidade = $cols1[11];
if ($cols1[2] >= 70 && $cols1[10] <= 0.0001){
print "$cols1[0]\t$cols1[1]\t$cols1[2]\t$cols1[10]\n";
}}

```

No *script* pbs, encontram as entradas para os pares de arquivos 3ab-kl, que se referem aos resultados dos BLASTp para a similaridade das sequências de proteínas de membrana ortólogas agrupadas nos clusters 1050 e 1816 contra as sequências de genes provenientes dos genomas para *R. amblyommatis* e outras cinco espécies de *Rickettsia*. No *script* perl, chamado dentro do *script* pbs, foi programado para filtrar as sequências similares adotando o nível de confiança ≥ 70 e identidade 0,0001. Nos arquivos de saída para os pares 3ab-kl, encontram-se estes resultados separadamente para cada espécie de *Rickettsia*. Os arquivos 4a e 4b correspondem ao conjunto destas sequências de nucleótidos similares para cada conjunto de grupo ortólogo, sendo os cluster 1050 e 1816, respectivamente.

3.3.4.1 Construção de árvores filogenéticas

Por meio do aplicativo *Reverse* (<http://reverse-complement.com/>), pudemos inverter sequências 5' – 3' em 3' – 5' para a obtenção das sequências das proteínas na ordem correta de leitura com o uso do aplicativo *Translate* (<http://insilico.ehu.es/translate/>), para os resultados obtidos com o uso do TBLASTn. Após a análise com o uso do BLASTp, pudemos avaliar os aspectos evolutivos das sequências genicas e proteicas obtidas com estes alinhamentos para serem utilizadas no aplicativo MEGA 7.26 (<https://www.megasoftware.net/>) (KUMAR; STECHER; TAMURA, 2016). Por meio deste aplicativo foram realizados múltiplos alinhamentos de sequências com o uso do ClustalW considerando penalidade de 10 para regiões de gap, 0,2 para regiões de gap estendidas, e 4 para regiões separadas entre gaps, adotando do método do

Maximum Likelihood Tree (FU; LI, 1993) para um teste filogenético com 1000 repetições *boots-trap* (FELSENSTEIN, 1985).

3.4 Identificação de interações conservadas e específicas

Para a averiguação da função molecular e o processo biológico associado às proteínas provenientes do genoma do carrapato da espécie *I. scapularis*, foram realizadas as anotações de proteína e anotação de ontologia gênica (GO) (CAMON, 2004). Com o uso do aplicativo AgBase-GOanna 2.2 (<https://agbase.arizona.edu/cgi-bin/tools/GOanna.cgi>) (MCCARTHY et al., 2006), realizou-se a análise de enriquecimento hipergeométrico e o teste FDR para múltiplas hipóteses, considerando as sequências de proteínas pertencentes à todos os grupos funcionais para os quais elas forem assinaladas e classificadas com múltiplas assinaturas de GO no COG.

O GOanna 2.2 permite que os usuários adicionem rapidamente mais anotações de GO a seus dados, transferindo anotações de GO com base na homologia de sequência, do que aquelas recuperadas pelo GORetriever transferindo anotações de GO de produtos genéticos anotados em outras espécies com base em pesquisas padrão do BLAST em bancos de dados que contêm sequências anotadas de GO (MCCARTHY et al., 2006). Segundo estes autores, o GOanna 2.2 retorna os alinhamentos de sequência e suas anotações GO, sendo que os usuários podem filtrar manualmente esses dados para definir boas correspondências de alinhamento exatamente como fariam em qualquer pesquisa do BLAST.

O arquivo de entrada contém o acesso as sequências de proteínas provenientes de carrapato da espécie *I. scapularis* em formato Fasta. O arquivo de entrada contendo as sequências de proteínas do carrapato foi submetido aos seguintes parâmetros (Tabela 2).

Tabela 2 Parâmetros e passos adotados para a análise de averiguação da função molecular e o processo biológico associado às proteínas provenientes do genoma do carrapato da espécie *I. scapularis*.

Passos	Parâmetros
1	Blastp (programa)
2	Sequência e/ou arquivo a ser analisado
3	Data base (uniprot)
4	Expect evalue 10e-10
5	Matriz (BLOSUM 62)
6	Gap Costs, Existence 11, Extision 1
7	Word size 6
8	Filter low complexity
9	Nbr of target segs: 1 alvo
10	Pct identidade filter 70% query coverage filter 70%
11	Blast results format and TSV format
12	Blast Html and Tsu format
13	Type of evidence to return: marcar todos, exceto obsolete evidence
14	Blast

Na Tabela 2, verifica-se o resumo do tutorial para o uso do AgBase-GOanna 2.2, (https://agbase.arizona.edu/Education/NABDAWorkshop/tutorial_4/tutorial_4.htm). Os três arquivos de saída e o que eles contêm podem ser vistos em: (i) o arquivo Excel contendo as entradas, suas correspondências e as anotações GO para as correspondências; (ii) o arquivo align.html contendo os alinhamentos, e (iii) o arquivo sliminput.txt contendo um resumo dos resultados para as anotações GO.

3.5 Visualização de redes de interação e análise de ontologia gênica

Após a construção do interatoma geral, filtragem do interatoma de proteínas de membrana e interatoma de proteínas de membrana ortólogas, foram construídas redes-consenso contendo todas as interações preditas para a interface *R. amblyommatis* vs *I. scapularis* e visualizadas com o uso do aplicativo Cytoscape 3.7 (<https://cytoscape.org/>) (SAITO et al., 2012). Para darmos início a construção e visualização de redes do interatoma, primeiramente demos a entrada do arquivo acessando; o *link file import - Network file*: (arquivo obtido contendo a rede de interação com os IDs de proteínas da bactéria da espécie *R. amblyommatis* versus o carrapato da espécie *I. scapularis*). Para ajustar e separar as duas colunas do arquivo contendo na primeira as proteínas da bactéria e na segunda as proteínas do carrapato, com as quais foram formados os pares de interação, demos início ao comando do aplicativo para separar as duas colunas por tabulação acessando: o *link advanced options - File Import Options: (semicolon)* para tabular as colunas. Para a editar as colunas demos início ao comando do aplicativo para definir na rede

os nós de fonte acessando o *link target*: bolinha cor verde (nó de entrada) para as proteínas da bactéria e na bolinha cor laranja (nó de saída) para as proteínas do carrapato.

Para a otimização da rede demos início ao comando do aplicativo para visualização da topologia obtida acessando; o *link Style node – Shape (ellipse)* para formatar os nós em elipses; *Label* (ID das sequências) removido para compactar as estruturas de interação entre os nós e facilitar a visualização dos módulos e complexos de proteínas formados; *Height*: 30 para definir o tamanho dos nós, *Width*; 30 para definir o tamanho das arestas entre os nós; e *Fill color* para dar a cor azul para todos os nós. Para escolher o tipo de linha demos início ao comando do aplicativo acessando: no *link Style edge - Line type* (Marquee Equal Dash) para que as linhas sejam tracejadas. Para apagar proteínas fora da rede que não fizeram partes dos maiores complexos de interação obtidos demos início ao comando do aplicativo acessando; no *link Detach view* para abrir a rede em outra janela, aonde pode-se com o comando *ctrl* do computador, selecionar, e em seguida deletar estes nós de proteínas separadas das redes principais com o comando delete do computador.

Para selecionar as proteínas de *R. amblyommatidis*, utilizou-se do comando do aplicativo acessando; no *link Select nodes - From id list file* para inserir uma lista contendo os IDs únicos das proteínas da bactéria que interagiram entre as proteínas do carrapato extraídas dos dados suplementares S9, S12 e S14, separadamente para cada rede obtida. Para marcar na rede em cor azul as proteínas da bactéria, construiu-se uma nova coluna para as proteínas selecionadas, dando início ao comando do aplicativo acessando; no *link Table painel* sinal de (+), *new single* (colum) para a nova coluna e (integer) para a sua integração na rede. Para ajustar a nova coluna, a primeira atribuiu-se a nomeada (amb) em referência as proteínas da bactéria da espécie *R. amblyommatidis*.

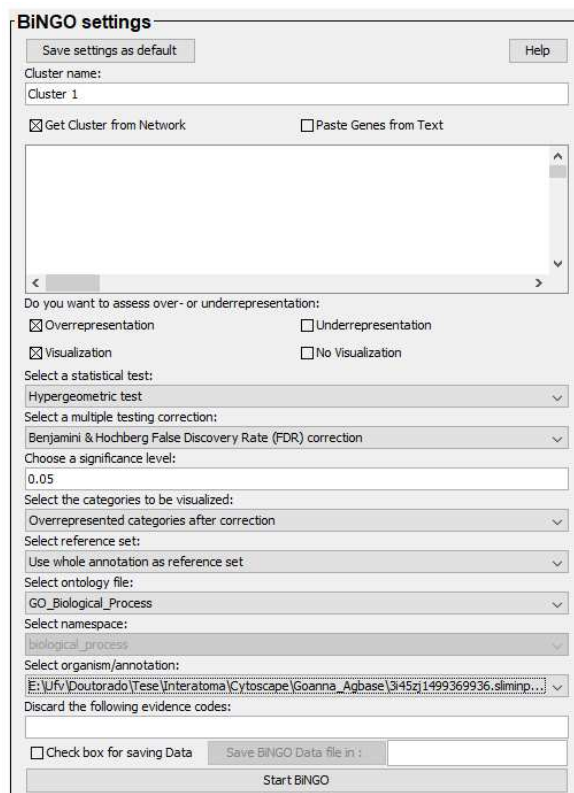
Logo em seguida, atribui-se um valor representativo “1” para sua inserção na rede, atribuindo de forma correta à todas as proteínas da bactéria fazendo uso do comando do lado direito do *mouse* do computador dando acesso; ao *link Apply to selected nodes*. Para destacar na rede as proteínas da bactéria utilizou-se das atribuições inseridas com a nova coluna dando início ao comando do aplicativo acessando; no *link Style - Node (Fill color)* para selecionar a coluna inserida à partir do nome (amb). Em seguida no *link Mapping type* selecionou-se o parâmetro (*Discrete mapping*) para a marcação das proteínas de *R. amblyommatidis* na rede acessando na caixa do comando aberta o valor atribuído "1", utilizando do comando do lado direito do *mouse* do computador, com acesso do contendo para explorar e escolher a cor para

as proteínas da bactéria que apareceram na rede dando destaque entre as proteínas do carrapato, sendo representada pelos nós em cor vermelha.

Para a identificação de módulos biologicamente funcionais foi adotado o pacote ModuLand 2.0 (<http://apps.cytoscape.org/apps/moduland20>), que utiliza o método de afinidade de propagação para identificar coerentes módulos dentro de uma rede (FREY; DUECK, 2007). O termo de ontologia gênica mais significativo para cada módulo foi identificado com o uso do método de distribuição hipergeométrica, utilizando o Teste FDR implementado no pacote Bingo 3.0.3 (<http://apps.cytoscape.org/apps/bingo>) (MAERE; HEYMANS; KUIPER, 2005). No aplicativo Cytoscape 3.7, demos início a esta etapa com o acessando os pacotes e comandos; no *link Apps* para solicitar os pacotes Moduland 2.0 e o Bingo 3.0.3.

Na caixa de comando do pacote Moduland 2.0 após sua instalação demos início a análise de afinidade de propagação acessando; no *link Starting module on current network - (Run)* para clusterizar a rede do interatoma visualizada. Após esta modulação, na caixa de comando do pacote Bingo 3.0.3 após sua instalação demos início ao teste do método de distribuição hipergeométrica para avaliar os termos de ontologia gênica mais significativos para os cluster obtidos (Figura 11).

Figura 11 Caixa de comando do pacote Bingo 3.0.3 adotado com o uso do aplicativo Cytoscape 3.7 para o teste hipergeométrico e da taxa de descoberta do erro FDR, e dos processos biológicos de ontologia gênica (GO).



Fonte: pacote Bingo 3.0.3 adotado com o uso do aplicativo Cytoscape 3.7

Na Figura 11, o arquivo 3i45zj1499369936.sliminput.txt, trata do resumo da análise de ontologia genica (GO) realizada com o uso do aplicativo Goanna AgBase 2.2 conforme descrito no Tópico 3.4. Este arquivo foi ajustado e utilizado para o teste de FDR e de enriquecimento hipergeométrico para a obtenção dos resultados. Conforme os dados obtidos com o uso do pacote Bingo 3.0.3, selecionamos as proteínas que compuseram os módulos funcionais significativos.

3.6 Obtenção de dados de RNA-seq para os transcriptomas de carrapatos

Para a anotação funcional de potenciais transcritos, adotamos a “montagem *de novo*” para avaliámos a expressão diferencial, seus níveis de transcrição e a abundância de transcritos. Para a reconstrução e comparação, utilizamos de dados de espécies de carrapatos *Amblyomma* spp. baixados do NCBI (<https://www.ncbi.nlm.nih.gov/sra/?term=>), acessando o *Sequence Read Archive* (SRA). Com esta finalidade realizamos os downloads destes arquivos contendo os

dados de RNA-seq foram realizados em programas Linux diretamente na plataforma do Júpiter, fazendo uso do aplicativo `fastq_dump 2.9.0` diretamente na linha de comando do cluster: “./fastq_dump 2.9.0 --split-3 (SRR número do arquivo RNA-seq)”.

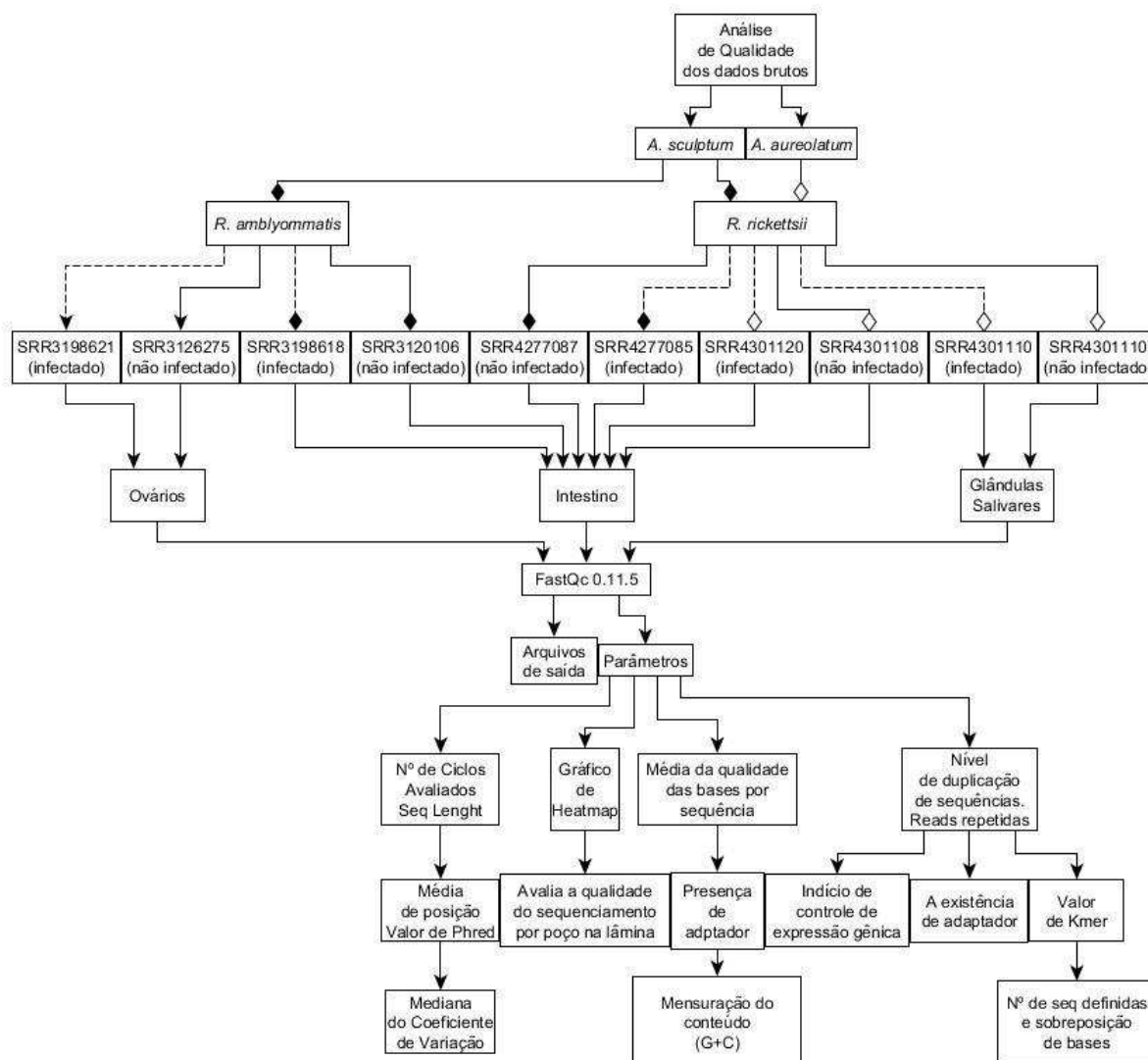
Nisto, foram adotados os dados do transcriptoma do carrapato das espécies *Amblyomma sculptum* (complexo *Amblyomma cajennense*) em resposta a infecção por *R. amblyommatis* ou não em ambos, os órgãos intestino e ovários. Para este carrapato utilizamos dos dados do intestino SRR3198618 (infectado) e SRR3120106 (não infectado), e dos dados dos ovários SRR3198621 (infectado) e SRR3126275 (não-infectado) pertencentes ao bio projeto número de acesso [PRJNA309641/](#) ID: 309641 (MOREIRA, 2016).

Além destes transcriptomas, foram adotados também os dados de RNA-seq do carrapato da espécie *A. sculptum* em resposta à infecção por *R. rickettsii* no intestino, sendo SRR4277085 (infectado) e SRR4277087 (não infectado) pertencentes ao bio projeto número de acesso [PRJNA343654/](#) ID: 343654 (ESTEVES et al., 2017), e da espécie *Amblyomma aureolatum* em resposta a infecção por *R. rickettsii* em glândula salivar, sendo SRR4301110 (infectado) e SRR4301100 (não infectado) e do intestino, sendo SRR4301120 (infectado) e SRR4301108 (não infectado) pertencentes ao bio projeto número de acesso [PRJNA344771/](#) ID: 344771 (MARTINS et al., 2017).

3.6.1 Análise e controle da qualidade dos dados do sequenciamento

Os dados brutos de RNA-seq foram destinados ao controle de qualidade para a remoção de adaptadores e sequências de baixa qualidade aumentando assim, a confiabilidade dos resultados, utilizando do aplicativo FastQC 0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) (ANDREWS, 2010) para análise de qualidade (Figura 12).

Figura 12 Fluxograma para as análises de qualidade dos dados brutos dos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp., fazendo o uso do aplicativo FastQc 0.11.5. Destaque para os parâmetros avaliados dos sequenciamentos.



Fonte: Autor deste trabalho.

Na figura 12, verifica-se o fluxograma contendo os passos para a análise de qualidade e controle de dados de RNA-seq. Dessa forma é possível avaliar o número de ciclos de sequenciamento que amplificaram através do gráfico da *Seq Length*, aonde temos a média de posição com o valor de *phred*, que analisa a mediana do coeficiente de variação possuindo a tendência de piorar a qualidade do sequenciamento na medida que avança ao longo do eixo da posição de cada base. No entanto, nem todas as posições estão presentes e o *reverse* ocorre logo depois do *forward*, sendo geralmente pior.

Através do gráfico do *Heatmap*, apresentando no eixo x a posição da base na lâmina e no eixo y a posição da lâmina, pudemos avaliar a qualidade do sequenciamento por poço na lâmina, de forma que a cor azul escuro indica que não tem variação de fluorescência, enquanto as partes de cor amarelo e vermelho indicam mais de uma fluorescência podendo haver erros.

A Média da qualidade por sequência, a qualidade das bases que entrou: saber se algum dos reagentes tinha defeito (início ruim) adaptador presente, a mensuração do conteúdo de G+C através da equação Conteúdo teórico = média/desvio padrão, o Conteúdo real, sendo necessário entender o Conteúdo de base N, de forma que, não saber qual nucleotídeo é um indício de contaminação, podendo ser DNA de outro organismo, ou quantidade inferior ao desejado.

Nesta análise espera-se que a distância do tamanho das *reads* esteja certa conforme a escolha do sequenciador. Pode-se também, avaliar o nível de duplicação de sequência, de forma que o transcriptoma pode ter *reads* muitas repetidas. Nisto, espera-se encontrar regiões duplicadas que é indício que teve controle da expressão gênica. Além disso, em amostras de sequências duplicadas avalia-se a existência de adaptador e através do conteúdo de *Kmer*, ou seja, o número de sequências definidas na busca indica quanto de sobreposição de bases ocorre.

Sendo assim, o valor de *Kmer* baixo, indica quão fácil será a montagem do transcriptoma, havendo a chance de artefatos na medida que aumenta o *Kmer* como ocorre no final da sequência. Para mais informações o tutorial pode ser acessado no [link \(https://www.mybiosoftware.com/fastqc-0-10-0-quality-control-tool-high-throughput-sequence-data.html\)](https://www.mybiosoftware.com/fastqc-0-10-0-quality-control-tool-high-throughput-sequence-data.html).

Com a finalidade de avaliarmos a qualidade do sequenciamento das sequências RNA-seq para dados de carrapatos *Amblyomma* spp. infectados ou não por espécies de *Rickettsia* que infectam carrapatos, em diferentes órgãos com o uso do aplicativo Fastqc 0.11.5 desenvolvemos o *script* abaixo:

```
#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=900mb
#PBS -m ae
#mensagem sera enviada: a - abort; b - begin; e - end
#PBS -M (email)
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega os módulos
module load fastqc/0.11.5
```

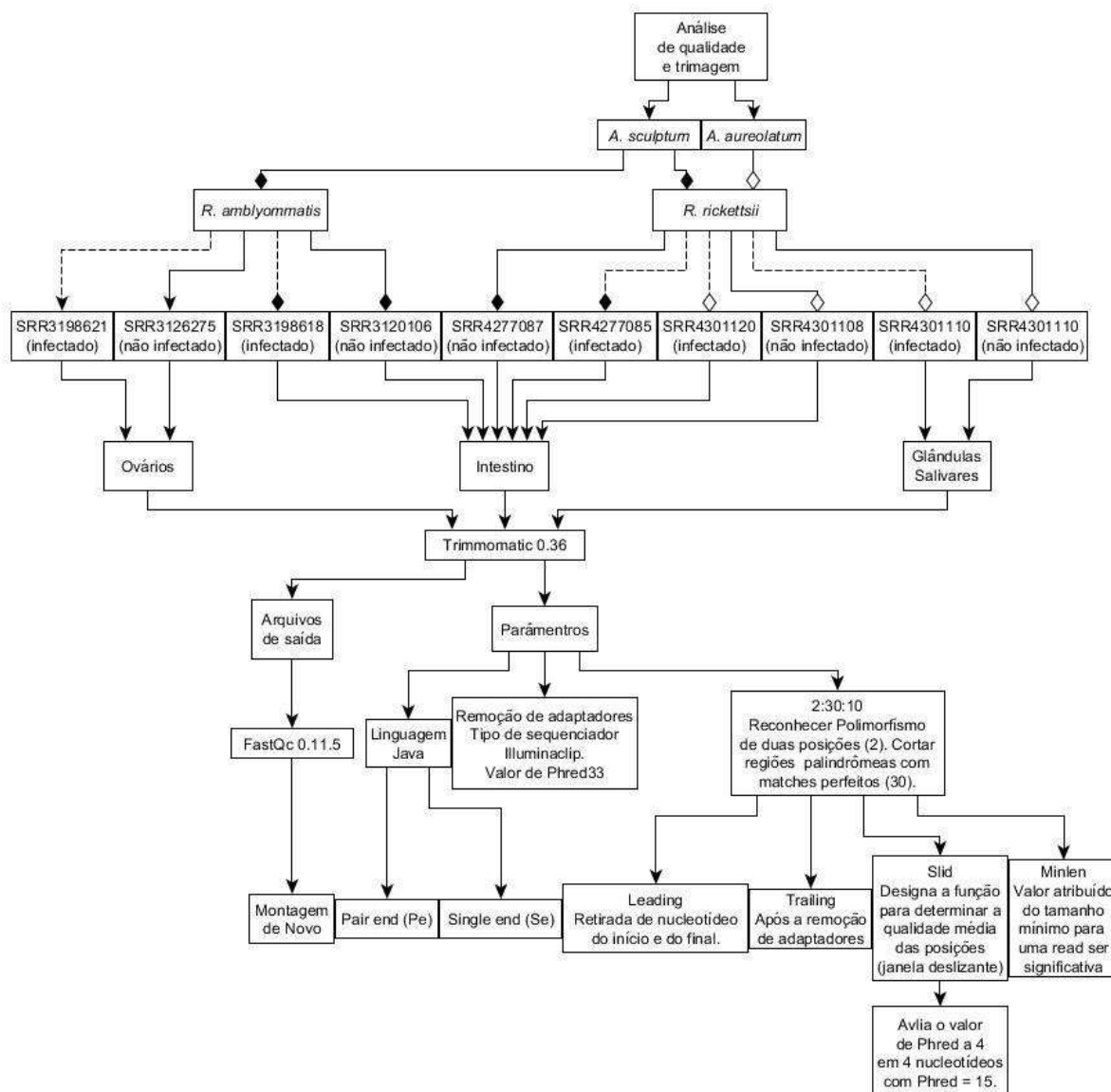
```
fastqc (arquivo 1a)  
fastqc (arquivo 1b)  
#Fim do Script
```

Neste *script* pbs, os pares de arquivos a-b contém os dados de RNA-seq provenientes de diferentes órgãos de espécies de carrapatos *Amblyomma* spp. infectados por *Rickettsia* spp. ou não, submetidas a análise de qualidade do sequenciamento com o uso do aplicativo Fastqc 0.11.5. Os arquivos de saída contém as *reads* e os *links* dos parâmetros avaliados.

3.6.2 Análise de qualidade e trimagem de dados do sequenciamento

Para a análise de trimagem e qualidade das *reads* do sequenciamento, fez-se o uso do Trimmomatic 0.36 (<http://www.usadellab.org/cms/?page=trimmomatic>) (BOLGER; LOHSE; USADEL, 2014) e em seguida o FastQC 0.11.5 novamente (ANDREWS, 2010). Dessa forma, realizamos a análise de trimagem das sequências e a remoção de adaptadores dos dados do sequenciamento dos transcriptomas de carrapatos medindo diversos parâmetros (Figura 13).

Figura 13 Fluxograma contendo os passos consequentes para análise de qualidade dos dados brutos e trimagem de transcriptoma utilizando do software Trimmomatic 0.36 e em seguida novamente o FastQC 0.11.5. Destaque para os parâmetros avaliados da trimagem.



Fonte: Autor deste trabalho.

Na Figura 13, verifica-se os passos para a análise de trimagem das *reads* através do aplicativo Trimmomatic 0.36 com os parâmetros avaliados. Este aplicativo utiliza de linguagem Java, sendo útil na trimagem de dados de transcriptomas, podendo trabalhar com arquivos de RNA-seq em formato *pair end* (Pe) e ou *single end* (Se), variando conforme o tipo de sequenciamento. O comando valor de Phred33 reconhece o sequenciador Illumina. Através do comando Illuminaclip:TruSeq3-PE.fa:2:30:10, estabeleceram as seguintes funções: reconhecer polimorfismo de duas posições “2”, cortar regiões palíndromas que possuem matches perfeitos “30” e remover as seqüências de adaptadores:

```

>PrefixPE/1
TACACTCTTTCCCTACACGACGCTCTTCCGATCT
>PrefixPE/2
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
>PE1
TACACTCTTTCCCTACACGACGCTCTTCCGATCT
>PE1_rc
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
>PE2
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
>PE2_rc
AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

```

Os comandos LEADING:3 e TRAILING:3, designaram a função para a retirada de nucleotídeo do início e do final, após a remoção dos adaptadores, enquanto o comando SLIDINGWINDOW:4:15, designou a função para determinarmos a qualidade média das posições (janela deslizante), avaliando o valor de *phred* a cada 4 em 4 nucleotídeos com *phred* = 15. O comando MINLEN:50 é um valor atribuído do tamanho mínimo para uma *read* ser significativa.

Para mais informações a respeito do uso do aplicativo Trimmomatic 0.36, o tutorial pode ser acessado no *link* (<http://gensoft.pasteur.fr/docs/Trimmomatic/0.36/>). Com a finalidade de avaliarmos a qualidade do sequenciamento, com a remoção de adaptadores e a trimagem das sequências de RNA-seq para dados de carrapatos *Amblyommas* spp. infectados ou não por espécies de *Rickettsia* que infectam carrapatos, em diferentes órgãos com o uso deste aplicativo desenvolvemos o *script* abaixo:

```

#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=6gb
#PBS -m ae
#mensagem sera enviada: a - abort; b - begin; e - end
#PBS -M (email)
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega os módulos
module load trimmomatic/0.36
java -jar ${TRIMMOMATIC_PREFIX}/trimmomatic.jar SE -phred33 (arquivo 1a) (arquivo de saída 1b)
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
MINLEN:50
java -jar ${TRIMMOMATIC_PREFIX}/trimmomatic.jar SE -phred33 (arquivo 2a) (arquivo de saída 2b)
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
MINLEN:50
java -jar ${TRIMMOMATIC_PREFIX}/trimmomatic.jar PE -phred33 (arquivo 3a) (arquivo 4a) (arquivo
de saída pair 3b) (arquivo de saída Unpair 3c) (arquivo de saída pair 4b) (arquivo de saída Unpair
4c) ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
MINLEN:50
#Fim do script

```

Neste *script* pbs, os arquivos 1a e 2a contém os dados brutos de RNA-seq para às sequências de transcritos *single end* provenientes do sequenciamento de transcriptoma de espécies de carrapatos *Amblyomma* spp. infectados e não infectados por espécies de bactéria *Rickettsia* spp. Para ilustrar, os arquivos 3a e 4a se referem aos dados brutos de RNA-seq para as sequências de *transcritos pair end* provenientes do sequenciamento de transcriptoma do carrapato. Nisto, tanto para dados *single end* ou *pair end* a análise de trimagem das *reads* foram realizadas através do aplicativo Trimmomatic 0.36. Os arquivos de saída 1b e 2b contém as *reads* trimadas em formato Fastq para os dados *single end*, enquanto os arquivos de saída 3b-c e 4b-c, contém as *reads* trimadas em formato Fastq para os dados *pair end*.

Fastqc_Trimmomatic

Com a finalidade de avaliarmos a qualidade das *reads* provenientes do sequenciamento de RNA-seq após a análise de trimagem com a retirada de possíveis adaptadores com uso do Trimmomatic 0.36, fizemos o uso do aplicativo Fastqc 0.11.5. Para isto, desenvolvemos o *script* abaixo:

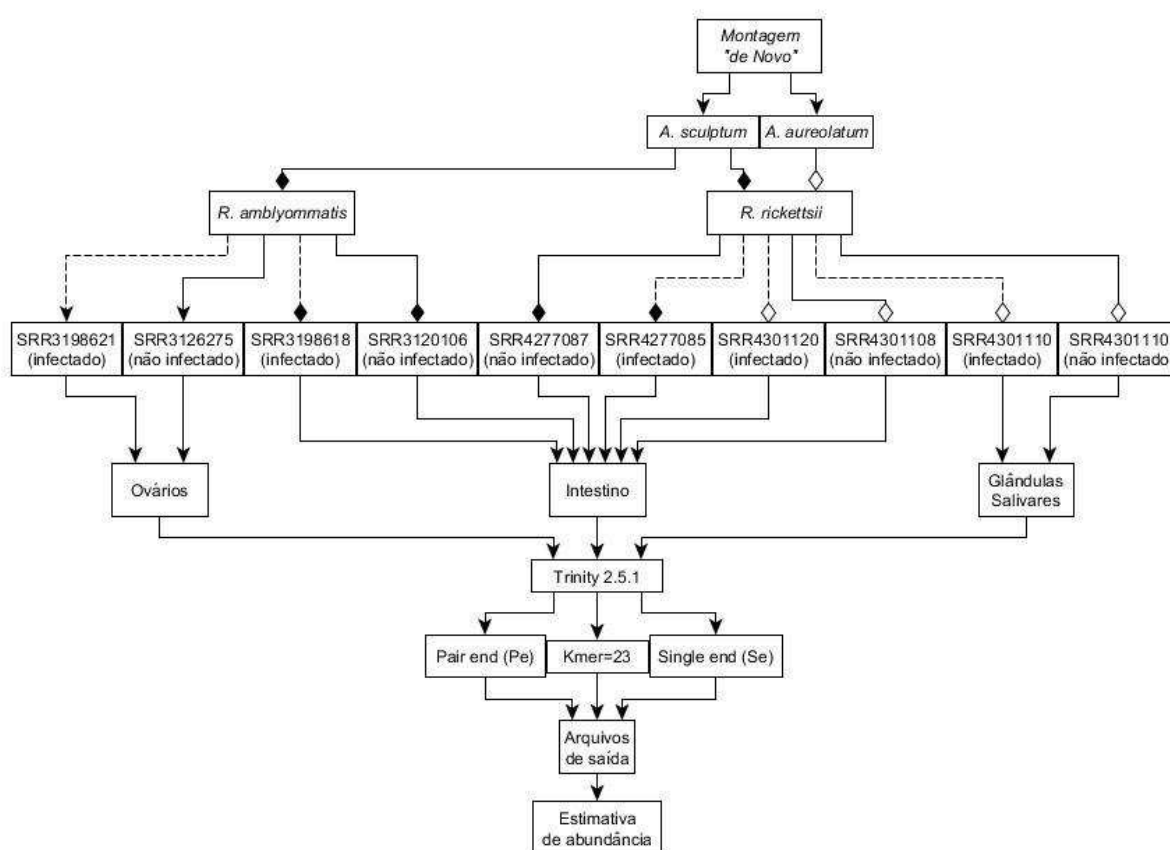
```
#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=2,mem=300mb
#PBS -m ae
#mensagem sera enviada: a - abort; b - begin; e - end
#PBS -M (email)
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega os módulos
module load fastqc/0.11.5
fastqc (arquivo 1a)
fastqc (arquivo 1b)
#Fim do Script
```

Neste *script* pbs, os arquivos 1a-b contém às *reads* provenientes dos dados de RNA-seq trimadas com os possíveis adaptadores removidos com o uso do aplicativo Trimmomatic 0.36. Em seguida estas sequências foram submetidas novamente a análise de qualidade através do aplicativo Fastqc 0.11.5, gerando os arquivos de saída contendo as *reads* e os *links* dos parâmetros avaliados.

3.6.3 Montagem de novo das seqüências de dados de RNA-seq

Para a montagem *de novo* utilizou-se do aplicativo Trinity 2.5.1 (<https://github.com/trinityRNA-seq/trinityRNA-seq/wiki>) (GRABHERR et al., 2011), após a remoção dos adaptadores das seqüências brutas e a seleção das *reads* com alta qualidade (Figura 14).

Figura 14 Fluxograma para a montagem “*de novo*” dos dados trimados e de boa qualidade dos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp., fazendo o uso do aplicativo Trinity 2.5.1. Destaque para os parâmetros avaliados do sequenciamentos.



Fonte: Autor deste trabalho.

Na Figura 14, verifica-se os passos para a análise da “montagem *de novo*” das *reads* através do aplicativo Trinity 2.5.1 com os parâmetros de construção. Este aplicativo pode trabalhar com arquivos de RNA-seq em formato *pair end* (Pe) e ou *single end* (Se), variando conforme o tipo de sequenciamento.

Para mais informações a respeito do uso do aplicativo Trinity 2.5.1, o tutorial pode ser acessado no *link* (<https://github.com/trinityRNA-seq/trinityRNA-seq/wiki/Running-Trinity>).

Com a finalidade de montarmos as *reads* do sequenciamento pela análise da montagem de novo de dados de RNA-seq para os carrapatos *Amblyomma* spp. infectados ou não por espécies de *Rickettsia* que infectam carrapatos, em diferentes órgãos com o uso deste aplicativo desenvolvemos o *script* abaixo:

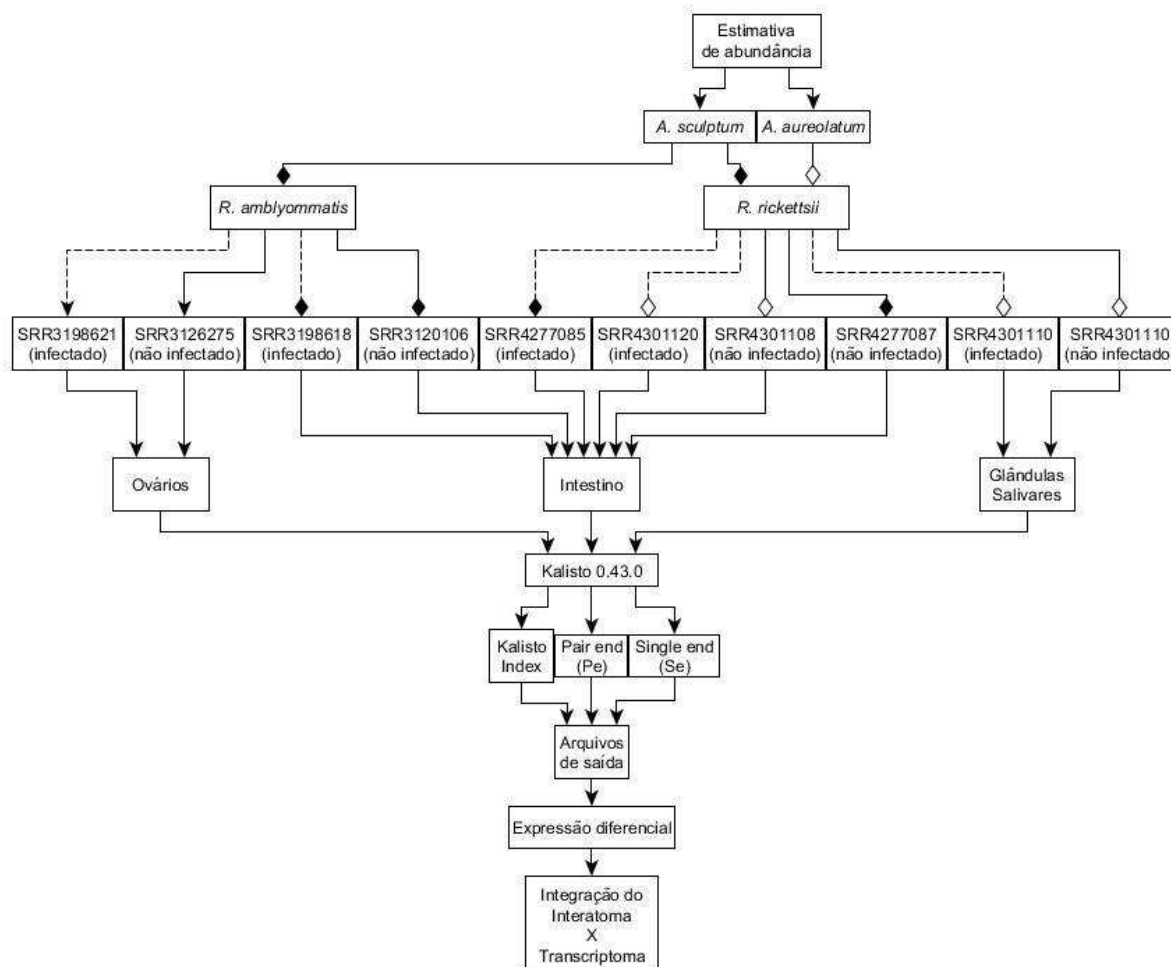
```
#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=15,mem=15gb,walltime=72:00:00
#PBS -q qtime
#PBS -m ae
#mensagem sera enviada: a - abort; b - begin; e - end
#PBS -M (email)
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega os módulos
module load trinity/2.5.1
module load perl/5.26.1
cd $PBS_O_WORKDIR
perl $TRINITY_HOME/Trinity.pl --seqType fq --max_memory 15G --single (arquivo 1a) --CPU $PBS_NP
--output (arquivo de saída pasta 1 montagem-k23) --KMER_SIZE 23 1>(arquivo de saída 1b)
perl $TRINITY_HOME/Trinity.pl --seqType fq --max_memory 15G --single (arquivo 2a) --CPU $PBS_NP
--output (arquivo de saída pasta 2 montagem -k23) --KMER_SIZE 23 1>(arquivo de saída 2b)
SRR4301120_1.trinity.k23.txt
#Fim do script
```

Neste *script* pbs, os arquivos 1a e 2a contém às sequências das *reads* provenientes do carrapato infectados ou não pela bactéria, em formato *single end*, obtidas após a análise de trimagem adotando o valor de *phred* = 15 com o uso do aplicativo Trimmomatic 0.36, para serem submetidas a montagem de novo com *kmer* = 23 com a obtenção das sequências de RNA com o uso do aplicativo Trinity 2.5.1. Os arquivos de saída pasta 1 e 2 são os consoles que receberam os arquivos trinity.Fasta contendo as sequências de RNA montadas. Os arquivos de saída 1b e 2b contêm os resultados do algoritmo do aplicativo Trinity 2.5.1 em formato Txt. Após montagem das *reads*, os arquivos de saída 1b e 2b fazem referência e fornecem dentro dos consoles, os arquivos com as sequências de RNA em formato Fasta.

3.6.4 Cálculo dos níveis de abundância da transcrição

Os transcritos montados foram alinhados para o cálculo dos níveis de abundância da transcrição de modo a avaliar as mudanças de expressão gênica utilizando o aplicativo Kallisto 0.43.0 (<https://pachterlab.github.io/kallisto/>) (BRAY et al., 2016) nos órgãos internos com ou sem a infecção (Figura 15).

Figura 15 Fluxograma para o cálculo dos níveis de abundância da transcrição para as sequências de RNA montadas dos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp., fazendo o uso do aplicativo Kallisto 0.43.0. Destaque para os parâmetros avaliados dos sequenciamentos.



Fonte: Autor deste trabalho.

Na Figura 15, verifica-se os passos para o cálculo dos níveis de abundância da transcrição por meio do uso do aplicativo Kallisto 0.43.0 com os parâmetros da análise. Este aplicativo pode trabalhar com arquivos de RNA-seq em formato *pair end* (Pe) e ou *single end* (Se), variando conforme o tipo de sequenciamento. Para mais informações a respeito do uso do aplicativo Kallisto 0.43.0, o tutorial pode ser acessado no [link \(https://pachterlab.github.io/kallisto/starting\)](https://pachterlab.github.io/kallisto/starting).

Com a finalidade de realizarmos o cálculo dos níveis de abundância da transcrição das *reads* montadas para os dados de RNA-seq de carrapatos *Amblyomma* spp. infectados ou não por espécies de *Rickettsia* que infectam carrapatos, em diferentes órgãos com o uso deste aplicativo desenvolvemos o *script* abaixo:

```
#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=2,mem=2gb,walltime=72:00:00
#PBS -q qtime
#PBS -m ae
#mensagem sera enviada: a - abort; b - begin; e - end
#PBS -M (email)
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega os módulos
module load kallisto/0.43.0
kallisto quant --index (arquivo de saída transcriptoma.idx 2a) --output-dir (arquivo pasta de saída a) --
bootstrap-samples 100 --single -l 200 -s 20 (arquivo 1a)
kallisto quant --index (arquivo de saída transcriptoma.idx 2b) --output-dir (arquivo pasta de saída b) --
bootstrap-samples 100 --single -l 200 -s 20 (arquivo 1b)
#Fim do script
```

Neste *script* pbs, os arquivos 1a e 1b contém às sequências das *reads* provenientes do carrapato infectados ou não pela bactéria, em formato *single end*, obtidas após a análise de qualidade e trimagem adotando o valor de *phred* = 15 com o uso do aplicativo Trimmomatic 0.36 e do Fastqc 0.11.5. Os arquivos de saída, pasta a e b são os consoles que receberam os arquivos transcriptoma.idx 1b e 2b contendo os resultados do algoritmo do aplicativo Kallisto 0.43.0, os quais foram utilizados para indexar os arquivos Trinity.Fasta à todo o transcriptoma para este conjunto de dados.

3.6.4.1 Indexação das sequências de RNA com o transcriptoma

Com a finalidade de realizarmos a cálculo e obtenção dos níveis de abundância da transcrição das *reads* montadas para os dados de RNA-seq de carrapatos *Amblyommas* spp. infectados ou não por espécies de *Rickettsia* em diferentes órgãos, realizamos a indexação dos arquivos de saída contendo as sequências de RNA em formato Fasta, obtidas com o uso do Trinity 2.5.1 e os arquivos transcriptoma.idx, obtidos com o uso do aplicativo Kallisto 0.43.0, desenvolvemos o *script* abaixo:

```
#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N index.pbs
#PBS -l nodes=1:ppn=1,mem=3gb,walltime=72:00:00
#PBS -q qtime
#PBS -m ae
#mensagem sera enviada: a - abort; b - begin; e - end
#PBS -M (email)
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
```

```

source /etc/profile.d/modules.sh
#Carrega os módulos
module load kallisto/0.43.0
#indexado o Trinity.Fasta com o todo transcriptoma
kallisto index -i (arquivo 1a transcriptoma.idx) (arquivo 1b Trinity.Fasta)
#Fim do script

```

Neste *script* pbs, o arquivo 1a se refere ao transcriptoma.idx, obtido com uso do aplicativo Kallisto 0.43.0. Estes arquivos foram indexados ao arquivo 2b, que contém as sequências de RNA em formato Fasta obtidas com uso do Trinity 2.5.1. Nos arquivos de saída para esta análise, foram originados três arquivos de saída para cada conjunto de dados de transcriptoma de *Amblyomma* spp. infectado e não infectado por espécies de *Rickettsia* contendo em especial os valores dos níveis de abundância dos transcritos.

3.6.5 Integração das sequências dos transcritos nos interatomas

Para a análise da expressão diferencial por meio do cálculo dos valores de *fold change* e averiguação da regulação da expressão dos transcriptomas montados utilizamos das proteínas de carrapato da espécie *I. scapularis*, com as quais formaram pares de interação de proteínas com *R. amblyommatis* obtidos anteriormente com a construção dos interatomas e da análise de averiguação de ontologia genética e visualização das redes.

Com isto, averiguamos as sequências por análise de similaridade fazendo o uso do TBLASTn, adotando as sequências de proteínas do carrapato presentes nos interatomas de membrana, em especial aquelas, que participam de módulos funcionais significativos: secreção, transporte, fosforilação, metilação, defesa contra vírus, oxidação e proteólise. Estas sequências de proteínas do *I. scapularis* foram extraídas dos dados suplementares (S19) obtidas anteriormente com uso do aplicativo Cytoscape 3.6.0, e foram submetidas a estes multialinhamentos de sequência contra as sequências de RNA em formato Fasta (Dados não publicados) obtidas com o uso do aplicativo Trinity 2.5.1.

Dessa forma, obtivemos em parte o interatoma funcional putativo para as espécies de carrapatos *Amblyomma* spp. adotando as sequências dos transcriptomas montados similares as sequências de proteínas de carrapato da espécie *I. scapularis* presentes nos interatomas de proteínas de membrana. Para esta finalidade desenvolvemos os *scripts* abaixo:

```

#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job

```

```

#PBS -N tblastn_protein.ixodes.interaction.rick.amb_vs_transcriptomas.pbs
#PBS -l nodes=1:ppn=6,mem=1gb,walltime=72:00:00
#PBS -q qtime
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
#Carrega o módulo opt-python
blastall -p tblastn -i (arquivo 1a) -d (arquivo 2a) -a 6 -b 1 -F T -m 8 -o (arquivo de saída 3a)
blastall -p tblastn -i (arquivo 1b) -d (arquivo 2b) -a 6 -b 1 -F T -m 8 -o (arquivo de saída 3b)
INPUT_FILE_1="(arquivo 1a)"
INPUT_FILE_2="(arquivo 1b)"
INPUT_FILE_3="(arquivo 2a)"
INPUT_FILE_4="(arquivo 2b)"
OUTPUT_FILE_1="(arquivo 3a)"
OUTPUT_FILE_2="(arquivo 3b)"
INPUT_FILE_a="(arquivo de saída 3a)"
INPUT_FILE_b="(arquivo de saída 3b)"
OUTPUT_FILE_a="(arquivo 4a)"
OUTPUT_FILE_b="(arquivo 4b)"
perl script.blast.cutoof.pl ${INPUT_FILE_a} > ${OUTPUT_FILE_a}
perl script.blast.cutoof.pl ${INPUT_FILE_b} > ${OUTPUT_FILE_b}
#-n (nucleotídeo - bases nitrogenadas)
#-p (proteínas - aminoácido)
#-i (arquivo de entrada)
#-d (database formatado)
#-a 6 (número de processadores. Quando coloca o -a tem que colocar o mesmo número de processadores)
#-m 8 (saída tabulada)
#Fim do script

#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
my $confianca;
my $identidade;
my $idblast;
my $value;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas [$i]);
my @cols1 = split ("\t", $linhas[$i]);
#print "$cols1[0]\n";
#pega a primeira coluna
$id = $cols1[0];
#pega a segunda coluna
$idblast = $cols1[1];
#pega a quarta coluna
$confianca = $cols1[2];
#pega a decima primeira
$value = $cols1[10];

```

```
#pega a decima segunda coluna
$identidade = $cols1[11];
if ($cols1[2] >= 70 && $cols1[10] <= 0.0001){
print "$cols1[0]\t$cols1[1]\t$cols1[2]\t$cols1[10]\n";
}
}
```

No *script* pbs, o arquivo 1a se refere às sequências de proteínas do carrapato da espécie *I. scapularis* obtidas através dos interatomos, em especial as, que constituíram os módulos funcionais significativos (Dados suplementares S19), sendo submetidas ao TBLASTn contra os arquivos 2a e 2b contendo as sequências de RNA de transcritos em formato Fasta (Dados não publicados), obtidas para cada conjunto de dados, infectados e não infectados por *Rickettsia* spp., respectivamente.

Os arquivos de saída 3a e 3b contêm os resultados do alinhamento com os parâmetros avaliados (Dados não publicados). Com o uso do *script* perl, adotando o nível de confiança ≥ 70 e identidade $\geq 0,0001$, filtramos as sequências similares. Nos arquivos de saída 4a e 4b, encontram-se os resultados dos alinhamentos com uso do TBLASTn para a similaridade das sequências de transcritos dos dados de transcriptoma infectados e não infectados para espécie de *Amblyomma* spp. e seus órgãos, para o conjunto de proteínas do *I. scapularis* de seis módulos funcionais significativos.

3.6.5.1 Análise da abundância e níveis de expressão da transcrição

Para análise de expressão diferencial foram adotados os resultados obtidos com uso do TBLASTn para a análise de similaridade, tratando-se dos arquivos dos dados suplementares S35ab-S39ab contendo os IDs das sequências de proteínas similares do carrapato da espécie *I. scapularis* e das sequências de RNA obtidas, anteriormente com o uso do Trinity 2.5.1. Com o uso destes arquivos, foram obtidos os dados dos transcriptomas integrados aos interatomos de membrana e ortólogos.

Nisto, foram obtidos os resultados da TPM para estes transcritos e as proteínas similares por meio de um filtro contra os arquivos contendo os dados suplementares S30ab-S34ab, obtidos anteriormente conforme os dados de transcrição e o uso do aplicativo Kallisto 0.43.0. Para esta finalidade desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#diretivas PBS. Especificam recursos que serao utilizados pelo job
```

```

#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=3g,walltime=72:00:00
#PBS -q qtime
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
#tnblast_trinity
INPUT_FILE_1a="(arquivo 1a)"
INPUT_FILE_1b="(arquivo 1b)"
INPUT_FILE_2a="(arquivo 2a)"
INPUT_FILE_2b="(arquivo 2b)"
#Resultado
OUTPUT_FILE_1a="(arquivo de saída 1a)"
OUTPUT_FILE_1b="(arquivo de saída 1b)"
OUTPUT_FILE_3a="(arquivo de saída 3a)"
OUTPUT_FILE_4a="(arquivo de saída 4a) teste_module4_vs_mg_fold.change.txt"
cat ${INPUT_FILE_1a} | awk -F "\t" '{print $1"\t"$2}' | sort | uniq > ${OUTPUT_FILE_1a}
cat ${INPUT_FILE_1b} | awk -F "\t" '{print $1"\t"$2}' | sort | uniq > ${OUTPUT_FILE_1b}
cat ${OUTPUT_FILE_1a} ${OUTPUT_FILE_1b} | sort | uniq >> ${OUTPUT_FILE_3a}
perl script.abundance2.pl ${OUTPUT_FILE_3a} ${INPUT_FILE_2a} ${INPUT_FILE_2b} >
${OUTPUT_FILE_4a}
#Fim do script

#!/usr/bin/perl
#carregar os módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
#arquivo sem redundância "abundance_total_id_non_red.txt"
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
#arquivo completo "ov_inf"
my $file1 = $ARGV[1];
open ARQUIVO,"<$file1";
my @linhas1 = <ARQUIVO>;
close ARQUIVO;
#arquivo completo "ov_non"
my $file2 = $ARGV[2];
open ARQUIVO,"<$file2";
my @linhas2 = <ARQUIVO>;
close ARQUIVO;
#Percorre abundance_total_id_non_red
my $id_abundance;
my $id_mid_inf;
my $tpm_mid_inf;
my $id_mid_non;
my $tpm_mid_non;
#Abrindo a linha
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas[$i]);
my @cols = split ("\t", $linhas[$i]);
#pega a primeira coluna
$id_abundance = $cols[1];

```


A integração dos dados do transcriptoma as proteínas presentes nos módulos funcionais do interatoma de membrana, gerou um arquivo de saída 4a contendo os IDs para cada uma das sequências de proteínas do carrapato *I. scapularis* com seus pares de IDs das sequências de RNA similares e os valores de TPM dos transcritos. Após a obtenção destes arquivos contendo os IDs de identificação das proteínas do *I. scapularis*, os IDs de identificação das sequências de RNA providas dos transcriptomas de *Amblyomma* spp. obtidas com o uso do Trinity, contendo os valores de TPM para cada transcrito, foram calculados os valores de *fold change* com o uso de gráficos construídos no aplicativo do Excel.

Com isto, foi realizado a divisão dos valores de TPM para os dados de transcritos em condições de infecções sob os dados de transcritos em condições de não infecção. Esta divisão pode ser obtida com a fórmula:

$$Fold\ change = \frac{TPM_{inf}}{TPM_{\bar{inf}}}$$

Após logaritmizar os valores de *fold change* na base Log2, foi realizado o cálculo da média para estes valores quando mais de uma sequência de transcrito foram similares e obtidas para uma mesma sequência de proteínas do carrapato *I. scapularis*. Logo em seguida, averiguamos se os transcritos são regulados de forma positiva quando os valores de *fold change* $\log_2 \geq 1$, e se regulados de forma negativa quando os valores de *fold change* $\log_2 \leq -1$.

3.7 Evolução de proteínas ortólogas e de membrana

Para análise filogenética das sequências de proteínas de membrana da espécie *R. amblyommatis* que interagem entre as sequências de proteínas do carrapato da espécie *I. scapularis*, foram adotados o multialinhamento de sequências com o uso do BLASTP contra as sequências proteicas de aminoácidos provenientes de 75 genomas de bactérias intracelulares obrigatórias, que infectam artrópodes, obtidas de depósitos no banco de dados *Ensembl* (<http://bacteria.ensembl.org/>) (Tabela 3).

Tabela 3 Espécies de *Rickettsia*., estirpes, e seus genomas de referência contendo os IDs taxonômicos.

Espécie	Estirpe	Genoma	ID Taxonômico	
<i>A. phagophitophilum</i>	-	AnapV1	948	
<i>E. chafeensis</i>	Heartland	ASM63281v1	1249651	
<i>N. risticii</i>	Illinois	ASM2252v1	434131	
<i>O. tsugamuch</i>	Karp	ASM96458v1	1359185	
<i>R. africae</i>	ESF-5	ASM2300v1	347255	
<i>R. akari</i>	Hartford	ASM1820v1	293614	
<i>R. amblyommatis</i>	Ac/Pa	ASM96467v1	1359164	
	Darkwater	ASM96499v1	1359166	
	GAT-30V	ASM28405v1	1105111	
<i>R. argasii</i>	T170-B	ASM96518v1	1268837	
<i>R. asemboensis</i>	-	ASM82812v2	1068590	
<i>R. australis</i>	Cutlack	ASM28415v1	1105110	
<i>R. bellii</i>	OSU 85-389	ASM1824v1	391896	
	RML369-C	ASM1238v1	336407	
	RML An4	ASM96500v1	1359193	
	RML Mogi	ASM96504v1	1359194	
<i>R. buchneri</i>	-	REISMNV1	1462938	
<i>R. canadenses</i>	CA410	ASM28391v1	1105107	
	McKiel	ASM1434v1	293613	
<i>R. conorii</i>	Malish 7	ASM702v1	272944	
<i>R. endosymbiont</i>	<i>I. pacificus</i>	ASM96515v1	1133329	
	<i>I. scapularis</i>	ASM16073v1	444612	
	<i>P. fluctus</i>	ASM160263v1	1462733	
<i>R. felis</i>	-	ASM80450v1	42862	
	LSU	ASM80452v1	1359195	
	Pedreira	ASM96466v1	1359196	
	URRWXCal2	ASM1214v1	315456	
<i>R. heilongjiangensis</i>	54	ASM22120v1	1032845	
<i>R. hoogstraalii</i>	RCCE3	ASM96484v1	1359197	
<i>R. japonica</i>	YH	ASM28359v1	652620	
<i>R. massiliae</i>	MTU5	ASM1662v1	416276	
	AZT80	ASM28385v1	1105112	
<i>R. monacensis</i>	-	RMONA_1	109232	
<i>R. montanensis</i>	OSU 85-930	ASM28417v1	1105114	
<i>R. parkeri</i>	AT#24	ASM96507v1	1359202	
	Grand Bay	ASM96508v1	1359203	
	Portsmouth	ASM28419v1	1105108	
	Tate's Hell	ASM96514v1	1359189	
<i>R. peacockii</i>	Rustic	ASM2152v1	562019	
<i>R. philipii</i>	364D	ASM28399v1	481009	
	Breinl	ASM36740v1	1290428	
<i>R. prowazekii</i>	BuV67-CWPP	ASM27720v1	1105096	
	Cairo 3	P_2008_03_27_11_29_25_runAssembly	1290430	
	Chernikova	ASM27716v1	1105094	
	Dachau	ASM27722v1	1105097	
	GvF12	P_2008_04_01_09_56_31_runAssembly	1290429	
	GvV257	ASM27724v1	1105098	
	Katsinyian	ASM27718v1	1105095	
	Madrid E	ASM19573v1	272947	
	NMRC Madrid E	ASM36390v1	1290427	
	Rp22	ASM2278v1	449216	
	RpGvF24	ASM27726v1	1105099	
	<i>R. raoultii</i>	-	ASM94095v1	369822
	<i>R. rhipicephali</i>	-	ASM144247v1	33992
3-7-female6-CWPP		ASM28407v1	1105113	
Ect		ASM96490v1	1359199	
<i>R. rickettsii</i>	Arizona	ASM28379v1	1105105	
	Brazil (RPN)	ASM28395v1	1105104	
	Colombia	ASM28377v1	1105102	
	Hauke	ASM28383v1	1105103	
	Hino	ASM28381v1	1105100	
	Hlp#2	ASM28393v1	1105101	
	Iowa	ASM1744v3	452659	
	Morgan	ASM83154v1	1337396	
	R	ASM83152v1	1338411	
	Sheila Smith'	ASM1822v1	392021	
<i>R. sibirica</i>	246	ASM16693v1	272951	

<i>R. slovacica</i>	13-B	ASM23784v1	941638
	D-CWPP	ASM25236v1	1105109
<i>R. spp</i>	Wb	ASM170792v1	1851512
<i>R. typhi</i>	B9991CWPP	ASM27730v1	1003202
	TH1527	ASM27728v1	1003201
	Wilmington	ASM804v1	257363
<i>W. endosymbiotic</i>	<i>C. quinquefasciatus</i> JHB	ASM15673v1	569881
<i>W. pipientis</i>	-	ASM175266v1	955

Na Tabela 3, verifica-se os nomes das espécies de *Rickettsia*, que possuem o genoma sequenciado e proteoma predito com o número de acesso, o *taxonomy* ID.

Neste sentido, com a finalidade de avaliarmos as sequências de proteínas de membrana de *R. amblyommatis* que interagiram com as sequências de proteínas do carrapato *I. scapularis* quanto as características evolutivas e filogenéticas, desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
#Carrega o módulo opt-python
blastall -p blastp -i (arquivo 1a) -d (arquivo 2a) -a 6 -b 1 -F T -m 8 -o (arquivo 3a)
#-n (nucleotídeo - bases nitrogenadas)
#-p (proteinas - aminoácido)
#-i (arquivo de entrada)
#-d (database formatado)
#-a 3 (número de processadores. Quando coloca o -a tem que colocar o mesmo número de processadores)
#-m 8 (saída tabulada)
#Fim do script
```

Neste *script* pbs, o arquivo 1a se refere as sequências de proteínas de membrana de *R. amblyommatis* que interagiram com as proteínas do *I. scapularis*. Estas sequências de aminoácidos de proteínas de membrana e ortólogas foram submetidas a análise de similaridade contra os arquivos 2a das sequências de proteínas provenientes dos genomas de 75 espécies de bactéria, incluindo 26 espécies de *Rickettsia* isoladas e reconhecidas de forma taxonômica (PAROLA et al., 2013).

Com isso, foram realizados multialinhamentos de sequências com o uso do BLASTp, gerando os arquivos de saída 3a, contendo os resultados destes alinhamentos para cada genoma.

Tratamento BLASTp

Com a finalidade de avaliarmos a similaridade das características evolutivas das proteínas de membrana de ortólogos para a espécie *R. amblyommatis* contra espécies de *Rickettsia* que infectam carrapatos, outras espécies do gênero *Rickettsia* que infectam os insetos (pulga, piolho, mosca e ácaro), além de outras bactérias pertencentes às famílias *Rickettsiales* e *Anaplasmatacea*, após a análise de similaridade com o uso do BLASTp, foram adotados o nível de confiança ≥ 70 e identidade 0,0001. Para isto desenvolvemos os *scripts* abaixo:

```
#!/bin/bash
#Diretivas PBS. Especificam recursos que serao utilizados pelo job
#PBS -N (nome do job)
#PBS -l nodes=1:ppn=1,mem=20mb
#PBS -m ae
#mensagem sera enviada: a-abort; b-begin; e-end
#PBS -M (email)
#Altera o diretório corrente para o diretório de onde foi submetido o script
cd $PBS_O_WORKDIR
#Habilita o carregamento de módulos
source /etc/profile.d/modules.sh
module load perl/5.24.0
#Carrega o módulo opt-python
INPUT_FILE_1="(arquivo 3a)"
OUTPUT_FILE_1="(arquivo 3b)"
perl script.blast.cutoof.pl ${INPUT_FILE_1} > ${OUTPUT_FILE_1}
cat ${OUTPUT_FILE_1} | sort | uniq >> ${OUTPUT_FILE_a}
#fim do script
```

```
#!/usr/bin/perl
#carregar módulos:
use warnings;
use strict;
#carrega o arquivo e declara variáveis:
my $file = $ARGV[0];
open ARQUIVO,"<$file";
my @linhas = <ARQUIVO>;
close ARQUIVO;
my $id;
my $confianca;
my $identidade;
my $idblast;
my $value;
for (my $i = 0; $i <= $#linhas; $i++){
chomp ($linhas [$i]);
my @cols1 = split ("\t", $linhas[$i]);
#print "$cols1[0]\n";
#pega a primeira coluna
$id = $cols1[0];
#pega a segunda coluna
$idblast = $cols1[1];
#pega a quarta coluna
$confianca = $cols1[2];
#pega a decima primeira
$value = $cols1[10];
```

```
#pega a decima segunda coluna
$identidade = $cols1[11];
if ($cols1[2] >= 70 && $cols1[10] <= 0.0001){
print "$cols1[0]\t$cols1[1]\t$cols1[2]\t$cols1[10]\n";
}
}
```

No *script* pbs, encontram as entradas para os arquivos de saída 3a, que se referem aos resultados dos BLASTp para a similaridade das sequências de proteínas de membrana ortólogas contra as proteínas provenientes de 75 genomas de bactérias. No *script* perl, chamado dentro do *script* pbs, foi programado para filtrar as sequências similares adotando o nível de confiança ≥ 70 e identidade 0,0001. Os arquivos de saída 3b, foram concatenados, em especial os resultados do alinhamento para as sequências de proteínas que constituem o interatoma de proteínas de membrana de *R. amblyommatis* contra as sequências de proteínas similares provenientes dos 29 genomas das espécies de *Rickettsia*.

Após isto, utilizando do aplicativo MEGA 7.26 (KUMAR; STECHER; TAMURA, 2016) foram realizados múltiplos alinhamentos de sequências com o aplicativo ClustalW considerando penalidade de 10 para regiões de gap, 0,2 para regiões de gap estendidas, e 4 para regiões separadas entre gaps, adotando do método do *Maximum Likelihood Tree* (FU; LI, 1993) para um teste filogenético com 1000 repetições *boots-trap* (FELSENSTEIN, 1985).

4 RESULTADOS

4.1 Construção do interatoma geral

Para a construção do interatoma geral, realizamos diversos multialinhamento de sequências visando uma análise de similaridade inicial para as proteínas provenientes dos genomas da bactéria da espécie *R. amblyommatis* e do carrapato da espécie *I. scapularis* contra as sequências depositadas nos bancos de dados do Peimap, Psimap, Pfam e String (Tabela 4-7) (Dados suplementares S1ab-S4ab). A análise de similaridade buscou identificar as sequências de proteínas presentes nos genomas destes organismos e os seus similares, que estavam presentes nos bancos de dados do interatoma, conforme apresentado no Tópico 3.1.1.

4.1.1 Análise de similaridade das proteínas

Para as análises de similaridade das proteínas destes organismos versus o banco de dados do Peimap, com base nos valores de confiança e identidade do BLASTP obtivemos no arquivo de saída os IDs de identificação das sequências e os seus similares, além dos IDs de identificação (PEI_) referentes ao banco do Peimap. Neste banco, as sequências de proteínas com este identificador possivelmente possui um par de interação ou mais, e supostamente teve a validação da interação comprovada por métodos bioquímicos e moleculares (Tabela 4).

Tabela 4 Resumo dos resultados das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, utilizando do BLASTP contra as sequências de proteínas depositados no banco de dados do Peimap.

Similaridade	ID_proteína	ID_proteína_Peimap	Confiança \geq 70	Identidade \leq 1e-10
<i>R. amblyommatis</i>	KJV61003.1	PEI_820920N	98.08	0.0
	KJV61004.1	PEI_389750N	95.48	3e-133
	KJV61005.1	PEI_1012609N	94.94	0.0
	KJV61007.1	PEI_964375N	97.46	0.0
	KJV61010.1	PEI_1137377N	90.94	0.0
<i>I. scapularis</i>	ISCW000012-PA.1	PEI_1143549N	73.80	0.0
	ISCW000014-PA.1	PEI_411930N	82.82	1e-92
	ISCW000018-PA.1	PEI_296080N	79.78	1e-95
	ISCW000018-PA.1	PEI_309886N	75.98	3e-77
	ISCW000067-PA.2	PEI_1099698N	80.00	1e-12

Na Tabela 4, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas similares encontrados no banco de dados do interatoma do Peimap. As análises de similaridade contra as proteínas do banco de dados do Peimap para as proteínas de *R.*

amblyommatis ([Dados suplementares S1a](#)) como para o *I. scapularis* ([Dados suplementares S1b](#)), podendo ser acessados nestes *links*.

Para as análises de similaridade das proteínas destes organismos versus o banco de dados do Psimap, com base nos valores do *E-value* do PSIBLAST obtivemos no arquivo de saída os IDs das sequências de aminoácidos e os seus similares, além dos IDs das proteínas referentes ao banco do Scop utilizado como base de dados do Psimap. Neste banco, as sequências de proteínas com este identificador possivelmente possui um par de interação ou mais, presente neste banco de interação de domínios estruturais de proteínas (Tabela 5).

Tabela 5 Resumo dos resultados preliminares das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, utilizando do PSIBLAST contra as sequências de proteínas depositados no banco de dados do Psimap.

Similaridade	ID_proteína	ID_proteína_Psimap	<i>E-value</i> ≤ 0,0001
<i>R. amblyommatis</i>	KJV61004.1	c.37.1.0	6e-36
	KJV61004.1	c.37.1.12	1e-04
	KJV61004.1	c.37.1.0	6e-36
	KJV61006.1	d.145.1.4	9e-19
	KJV61011.1	d.218.1.8	1e-36
<i>I. scapularis</i>	ISCW000002-PA.1	d.26.1.1	9e-32
	ISCW000003-PA.1	d.14.1.5	1e-21
	ISCW000002-PA.1	d.26.1.1	9e-32
	ISCW000003-PA.1	d.14.1.5	1e-21
	ISCW000002-PA.1	d.26.1.1	9e-32

Na Tabela 5, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas similares encontrados no banco de dados do interatoma do Psimap. As análises de similaridade contra as proteínas do banco de dados do Psimap para as proteínas de *R. amblyommatis* ([Dados suplementares S2a](#)) como para o *I. scapularis* ([Dados suplementares S2b](#)), podendo ser acessados nestes *links*.

Para as análises de similaridade das proteínas destes organismos versus o banco de dados do Pfam, com base nos valores do *E-value* do HMM obtivemos no arquivo de saída os IDs das sequências de aminoácidos e os seus similares, além dos IDs das proteínas referentes ao banco do iPfam utilizado como base de dados do Pfam. Neste banco, as sequências de proteínas com este identificador possivelmente possui um par de interação ou mais, presente neste banco de interação contendo motivos conservados e estruturas tridimensionais de proteínas (Tabela 6).

Tabela 6 Resumo dos resultados preliminares das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, utilizando do HMM contra as sequências de proteínas depositados no banco de dados do Pfam.

Similaridade	ID_proteína	ID_proteína_Pfam	Bit score \geq 70	E-value \leq 1e-10
<i>R. amblyommatis</i>	KJV61002.1	PF07690	146.6	7.9e-43
	KJV61003.1	PF12704	62.6	5.7e-17
	KJV61003.1	PF02687	50.9	1.6e-13
	KJV61004.1	PF00005	104.7	4.9e-30
	KJV61004.1	PF07066	15.9	0.0095
<i>I. scapularis</i>	ISCW000002-PA.1	PF00254	71.9	3.9e-20
	ISCW000003-PA.1	PF00288	38.7	8e-10
	ISCW000003-PA.1	PF08544	51.9	6.6e-14
	ISCW000003-PA.1	PF10509	68.7	2.1e-19
	ISCW000006-PA.1	PF02014	47.2	3.2e-12

Na Tabela 6, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas similares encontrados no banco de dados do interatoma do Pfam. As análises de similaridade contra as proteínas do banco de dados do Pfam para as proteínas de *R. amblyommatis* ([Dados suplementares S3a](#)) como para o *I. scapularis* ([Dados suplementares S3b](#)), podendo ser acessados nestes *links*.

Para as análises de similaridade das proteínas destes organismos versus o banco de dados do String, com base nos valores de identidade do BLASTP obtivemos no arquivo de saída os IDs das sequências de aminoácidos e os seus similares, além dos IDs das proteínas referentes ao banco do String. Neste banco, as sequências de proteínas com este identificador possivelmente possui um par de interação ou mais, presente neste banco de interação contendo mapas funcionais de proteínas (Tabela 7).

Tabela 7 Resumo dos resultados preliminares das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, utilizando do BLASTP contra as sequências de proteínas depositados no banco de dados do String.

Similaridade	ID_proteína	ID_proteína_String	Identidade \leq 0,01
<i>R. amblyommatis</i>	KJV61002.1	1001530.PMSV_1042	7e-10
	KJV61002.1	1001530.PMSV_4113	1e-09
	KJV61002.1	1005048.CFU_4229	5e-09
	KJV61003.1	999541.bgla_1g27830	5e-73
	KJV61004.1	1001530.PMSV_623	2e-54
<i>I. scapularis</i>	ISCW000001-PA.1	6945.ISCW020438-PA	9e-09
	ISCW000002-PA.1	1000565.METUNv1_03896	2e-17
	ISCW000003-PA.1	1001530.PMSV_1919	3e-67
	ISCW000003-PA.1	100226.SCO3136	3e-55
	ISCW000003-PA.1	1005058.UMN179_01896	6e-59

Na Tabela 7, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas similares encontrados no banco de dados do interatoma do String. As análises de

similaridade contra as proteínas do banco de dados do String para as proteínas de *R. amblyommatis* ([Dados suplementares S4a](#)) como para o *I. scapularis* ([Dados suplementares S4b](#)), podendo ser acessados nestes *links*.

Os resultados obtidos (Dados suplementares S1ab-S4ab), revelaram a ocorrência de 238, 1397, 1690 e 217668 sequências de proteínas para a bactéria, havendo 314606, 314606, 2288 e 273682 proteínas para o carrapato similares as sequências de proteínas depositadas nas bases dos bancos do interatoma Peimap, Psimap, Pfam e String, respectivamente.

4.2 Montagem do Interatoma

Para a montagem dos interatomos parciais, foram selecionadas as proteínas da bactéria da espécie *R. amblyommatis* e do carrapato da espécie *I. scapularis*, que possuem pares de sequências de proteínas similares, obtidas com o uso de multialinhamentos de sequências contra as sequências de proteínas depositados nos bancos de dados do interatoma. Os IDs de identificação dos pares de sequências de proteínas similares obtidos nos arquivos (Dados suplementares S1a-S4a) para a bactéria e (Dados suplementares S1b-S4b) para o carrapato foram interagidos por meio dos IDs de identificação dos pares de interação pré-estabelecidos entre as proteínas nas bases dos bancos de dados do interatoma, conforme o Tópico 3.1.2.

Consideradas com potencial para interagir, estas proteínas formaram as interações entre a bactéria e o carrapato baseadas nos pares de domínios presentes nos bancos de dados do Peimap, Psimap, Pfam e String. Os resultados obtidos (Dados não publicados), revelaram a ocorrência de 23138, 6880360, 1006036 e 30093 interações totais, contendo 8630, 145951, 495013 e 2194 interações únicas entre a bactéria e o carrapato, respectivamente.

4.2.1 Interatoma interespecífico

Os pares de interação obtidos para as sequências de proteínas da bactéria da espécie *R. amblyommatis* e de carrapato da espécie *I. scapularis*, sugerem a ocorrência da interação entre estes organismos, sustentadas nos bancos do interatoma. Os interatomos parciais montados possuem interações intraespecífica e interespecífica, preditas entre as proteínas de *R.*

amblyommatis e *I. scapularis*, podendo ser estudadas as interações entre as sequências de proteínas dos próprios organismos, e entre estes organismos.

No entanto, para o estudo de interações entre as proteínas da bactéria *R. amblyommatis* e do carrapato da espécie *I. scapularis* (patossistema), foram considerados apenas, os pares de proteínas que apresentaram interações interespecíficas, ocorrendo entre estas diferentes espécies (Tabelas 8-11) (Dados suplementares S1c-S4c). Após a limpeza de interações intraespecíficas e a inversão das interações obtivemos os pares que interagem na ordem bactéria-carrapato para o banco do Peimap (Tabela 8).

Tabela 8 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, provenientes do BLASTP contra os ids dos domínios das sequências de proteínas depositados no banco de dados do Peimap, contendo os métodos com os quais, estes pares de interação tenham sido validados.

ID_proteína <i>R. amblyommatis</i>	ID_proteína <i>I. scapularis</i>	ID1 Bactéria_Peimap	ID2 Carrapato_Peimap	Método
KJV61832.1	ISCW016079-PA.1	PEI_24644N	PEI_991914N	two hybrid
KJV61832.1	ISCW016080-PA.1	PEI_24644N	PEI_991914N	two hybrid
KJV62406.1	ISCW014417-PA.1	PEI_260968N	PEI_449695N	anti bait coip
KJV61174.1	ISCW016907-PA.1	PEI_925335N	PEI_266128N	null,two hybrid
KJV61833.1	ISCW020749-PA.1	PEI_441503N	PEI_301980N	two hybrid

Na Tabela 8, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas que interagem, referindo-se ao interatoma parcial encontrado no banco de dados do Peimap, ([Dados suplementares S1c](#)), podendo ser acessado neste [link](#).

Após a limpeza de interações intraespecíficas e a inversão das interações obtivemos os pares que interagem na ordem bactéria-carrapato para o banco do Psimap (Tabela 9).

Tabela 9 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, provenientes do PSIBLAST contra os ids dos domínios das sequências de proteínas depositados no banco de dados do Psimap, contendo os valores de distâncias com os quais, estes pares de interação tenham sido preditos.

ID_proteína <i>R. amblyommatis</i>	ID_proteína <i>I. scapularis</i>	ID1 Bactéria_Psimap	ID2 Carrapato_Psimap	Distância
KJV61004.1	ISCW000200-PA.1	c.37.1.12	c.37.1.12	2.05755
KJV61004.1	ISCW000200-PA.1	c.37.1.12	c.37.1.12	2.12306
KJV61004.1	ISCW009589-PA.1	c.37.1.12	c.37.1.12	3.48821
KJV61018.1	ISCW000019-PA.1	c.47.1.10	c.3.1.2	1.31858
KJV61041.1	ISCW003547-PA.1	c.37.1.11	c.37.1.11	2.32577

Na Tabela 9, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas que interagem encontrados nos bancos de dados do Psimap ([Dados suplementares S2c](#)), podendo ser acessado neste *link*.

Após a limpeza de interações intraespecíficas e a inversão das interações obtivemos os pares que interagem na ordem bactéria-carrapato para o banco do Pfam (Tabela 10).

Tabela 10 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, provenientes do HMM contra os ids dos domínios das sequências de proteínas depositados no banco de dados do Pfam, contendo os valores dos dois > *bit scores* com os quais, estes pares de interação tenham sido preditos.

ID_proteína <i>R. amblyommatis</i>	ID1 <i>Bit score</i>	ID_proteína <i>I. scapularis</i>	ID2 <i>Bit score</i>	ID1 Bactéria_Pfam	ID2 Carrapato_Pfam
KJV61004.1	104.7	ISCW000565-PA.1	38.6	PF00005	PF00005
KJV61004.1	104.7	ISCW000765-PA.1	114.0	PF00005	PF00005
KJV61004.1	104.7	ISCW005559-PA.1	29.2	PF00005	PF00664
KJV61005.1	25.6	ISCW023133-PA.1	81.3	PF02902	PF00240
KJV61006.1	22.7	ISCW001657-PA.1	328.1	PF00571	PF00478

Na Tabela 10, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas que interagem, referindo-se ao interatoma parcial encontrado no banco de dados do Pfam ([Dados suplementares S3c](#)), podendo ser acessado neste *link*.

Após a limpeza de interações intraespecíficas e a inversão das interações obtivemos os pares que interagem na ordem bactéria-carrapato para o banco do String (Tabela 11).

Tabela 11 Resumo dos resultados do interatoma utilizando das análises de similaridades entre as sequências de proteínas de ambos organismos, *R. amblyommatis* e *I. scapularis*, provenientes do BLASTP contra os ids dos domínios das sequências de proteínas depositados no banco de dados do String, contendo os valores de confiança com os quais, estes pares de interação tenham sido preditos.

ID_proteína <i>R. amblyommatis</i>	ID_proteína <i>I. scapularis</i>	ID1 Bactéria_String	ID2 Carrapato_String	Confiança
KJV61085.1	ISCW020779-PA.1	394.NGR_c31670	394.NGR_c11900	971
KJV61107.1	ISCW016070-PA.1	394.NGR_c16470	394.NGR_c31110	952
KJV61107.1	ISCW017192-PA.1	394.NGR_c16470	394.NGR_c35070	937
KJV61166.1	ISCW016022-PA.1	394.NGR_c10580	394.NGR_c05240	919
KJV61174.1	ISCW002558-PA.1	3055.EDP04740	3055.EDP07903	827

Na Tabela 11, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas que interagem encontrados nos bancos de dados do String ([Dados suplementares S4c](#)), podendo ser acessado neste *link*.

Os resultados obtidos (Dados suplementares S1c-S4c), revelaram a ocorrência de 166, 645855, 20305 e 171 interações totais, respectivamente similares as sequências de proteínas depositadas nas bases dos bancos do interatoma Peimap, Psimap, Pfam e String.

4.3 Cálculos de confiança do modelo

Os resultados obtidos para os interatomos parciais interespecíficos construídos e montados (Dados suplementares S1c-S4c), foram utilizados para o cálculo do *score* final para cada par de interação, e por banco de dados do interatoma, conforme apresentado no Tópico 3.1.3, sendo normalizados entre valores de 0 a 1, conforme apresentado no Tópico 3.1.4 (Tabelas 12-15) (Dados suplementares S1d-S4d), visando o cálculo do *score* combinado. Os valores do $> score$ final obtido para cada par de interação foi dividido pelo valor do $> score$ observado para cada banco de dados.

Para o cálculo do *score* combinado, utilizamos da fórmula do interatoma para atribuir o valor de confiança final adotando dos valores do *score* final normalizado e cada par de interação entre as proteínas de *R. amblyommatis* e do *I. scapularis* por banco do interatoma avaliado, conforme o Tópico 3.1.4. Os resultados do interatoma interespecífico obtidos apresentaram interações significativas com valores de *cuttof* ≥ 0.7 (Tabela 16) (Dados suplementares S1e).

4.3.1 Cálculo do *score* final

O cálculo do *score* final para a base de dados do Peimap relacionado a cada par de proteínas que interagem entre *R. amblyommatis* e o *I. scapularis* foi obtido com base no valor de confiabilidade de cada método experimental para detectar interações entre proteínas e foi previamente estimado para o cálculo de reabilidade, sendo normalizado em seguida para o cálculo do *score* combinado (Tabela 12).

Tabela 12 Resumo dos resultados do cálculo do *score* final normalizado para os pares de proteínas que interagem entre *R. amblyommatis* e *I. scapularis*, presente no banco de dados do Peimap.

ID_proteína <i>R. amblyommatis</i>	ID_proteína <i>I. scapularis</i>	Frequência da interação	Score final (ScF)	(ScF) Normal 0-1
KJV61130.1	ISCW020780-PA.1	2	1.347377	0,200217161
KJV61174.1	ISCW000908-PA.1	4	1.809795	0,268931425
KJV61174.1	ISCW024910-PA.1	1	0.823077	0,132470862
KJV61447.1	ISCW000555-PA.1	3	0.796221	0,118316631
KJV61830.1	ISCW016007-PA.1	1	0.265407	0,039438877

Na Tabela 12, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas destes organismos que interagem, representadas pelos IDs de identificação e com os valores de *score* final e normalizado obtidos por meio das análises de similaridade contra o banco de dados do Peimap ([Dados suplementares S1d](#)), podendo ser acessado neste *link*.

O cálculo do *score* final para a base de dados do Psimap relacionado a cada par de proteínas que interagem entre *R. amblyommatis* e o *I. scapularis*, foi obtido com base no valor de confiança deste banco que é o inverso da distância entre os pares de domínios estruturais que estão interagindo no PDB, sendo em seguida normalizado para o cálculo do *score* combinado (Tabela 13).

Tabela 13 Resumo dos resultados do cálculo do *score* final normalizado para os pares de proteínas que interagem entre *R. amblyommatis* e *I. scapularis*, presente no banco de dados do Psimap.

ID_proteína <i>R. amblyommatis</i>	ID_proteína <i>I. scapularis</i>	Frequência da interação	Score final (ScF)	(ScF) Normal 0-1
KJV61004.1	ISCW000200-PA.1	48	0,486014920658064	0,598406808
KJV61004.1	ISCW000565-PA.1	48	0,486014920658064	0,598406808
KJV61018.1	ISCW009623-PA.1	16	0,758391603088171	0,383488736
KJV61024.1	ISCW002352-PA.1	11	0,763230602494238	0,381057358
KJV61026.1	ISCW000069-PA.1	8	0,406139200149459	0,716095952

Na Tabela 13, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas destes organismos que interagem, representadas pelos IDs de identificação e com os valores de *score* final e normalizado encontrados no Banco do Psimap ([Dados suplementares S2d](#)), podendo ser acessado neste *link*.

O cálculo do *score* final para a base de dados do Pfam relacionado a cada par de proteínas que interagem entre *R. amblyommatis* e o *I. scapularis*, foi obtido com base no valor de confiança do iPfam que é a soma entre os dois “*bits score*” da assinatura dos domínios do Pfam do par proteínas com potencial interação, sendo em seguida normalizado para o cálculo do *score* combinado (Tabela 14).

Tabela 14 Resumo dos resultados do cálculo do *score* final normalizado para os pares de proteínas que interagem entre *R. amblyommatidis* e *I. scapularis*, presente no banco de dados do Pfam.

ID_proteína <i>R. amblyommatidis</i>	ID_proteína <i>I. scapularis</i>	Frequência da interação	Score final (ScF)	(ScF) Normal 0-1
KJV61004.1	ISCW000565-PA.1	1	143.3	0,016559582
KJV61004.1	ISCW000765-PA.1	2	401.5	0,046396875
KJV61005.1	ISCW023133-PA.1	1	106.9	0,01235324
KJV61006.1	ISCW001657-PA.1	2	706.7	0,081665434
KJV61011.1	ISCW010794-PA.1	1	142.7	0,016490247

Na Tabela 14, verifica-se um resumo dos resultados preliminares obtidos para os pares de proteínas destes organismos que interagem, representadas pelos IDs de identificação e com os valores de *score* final normalizado encontrados no Banco do Pfam ([Dados suplementares S3d](#)), podendo ser acessado neste *link*.

O cálculo do *score* final para a base de dados do String relacionado a cada par de proteínas que interagem entre *R. amblyommatidis* e o *I. scapularis*, foi obtido com base na soma da confiança para um mesmo par de interação e a soma da frequência desta, sendo em seguida normalizado para o cálculo do score combinado (Tabela 15).

Tabela 15 Resumo dos resultados do cálculo do *score* final para os pares de proteínas que interagem entre *R. amblyommatidis* e *I. scapularis*, presente no banco de dados do String.

ID_Proteína <i>R. amblyommatidis</i>	ID_Proteína <i>I. scapularis</i>	Frequência da interação	Score final (ScF)	(ScF) Normal 0-1
KJV61085.1	ISCW020779-PA.1	1	971	0,04769858
KJV61107.1	ISCW016070-PA.1	1	952	0,04676524
KJV61107.1	ISCW017192-PA.1	1	937	0,046028393
KJV61166.1	ISCW016022-PA.1	1	919	0,045144176
KJV61174.1	ISCW002558-PA.1	9	8215	0,403546692

Na Tabela 15, verifica-se um resumo dos resultados preliminares obtidos para a predição das interações entre os pares de proteínas destes organismos, representadas pelos IDs de identificação, contendo a frequência da formação da interação e os valores do cálculo do *score* final normalizado para cada par de domínios presente no Banco do String ([Dados suplementares S4d](#)), podendo ser acessado neste *link*.

Os resultados obtidos (Dados suplementares S1d-S4d), revelaram a ocorrência de 185, 645855, 20305 e 171 interações totais, contendo 115, 13551, 9931 e 39 interações únicas, respectivamente similares as sequências de proteínas depositadas nas bases dos bancos do interatoma Peimap, Psimap, Pfam e String.

4.3.2 Cálculo do *score* combinado

Para o cálculo do *score* combinado foi utilizado dos resultados do *score* final normalizado entre valores de 0 a 1 para cada par de interação obtido entre as proteínas da bactéria da espécie *R. amblyommatis* e de carrapato da espécie *I. scapularis* (Dados suplementares S1d-S4d). A combinação do *score* final para cada par de proteínas deu origem ao modelo do interatoma geral interespecífico entre as sequências de proteínas de *R. amblyommatis* versus *I. scapularis* (Dados não apresentados). Por meio da combinação matemática do *score* final normalizado para cada par de interação, obtivemos uma rede consenso do interatoma geral interespecífico entre estes organismos, considerando o valor de *cutoff* $\geq 0,7$ (Dados suplementares S1e) (Tabela 16).

Tabela 16 Resumo dos resultados do cálculo do *score* combinado com valores $\geq 0,7$ de *cutoff* para os pares de proteínas que interagem entre *R. amblyommatis* e *I. scapularis*, presente no banco de dados do interatoma.

ID proteína <i>R. amblyommatis</i>	ID proteína <i>I. scapularis</i>	<i>Score</i> combinado	Peimap	Psimap	Pfam	String
KJV62157.1	ISCW016714-PA.1	0.700324958753441	0	0	1	0
KJV61971.1	ISCW007313-PA.1	0.701315409694645	0	1	1	0
KJV61887.1	ISCW019069-PA.1	0.702032924044534	0	0	1	0
KJV61321.1	ISCW007313-PA.1	0.703682762168888	0	1	1	0
KJV61821.1	ISCW023102-PA.1	0.703851778219922	0	0	1	0

Na Tabela 16, verifica-se um resumo dos resultados obtidos para os pares de proteínas que interagem, sendo representadas pelos IDs das proteínas e com os valores de *score* combinado para cada par de interação e a distribuição binomial da frequência de sua ocorrência em cada banco de dados, tratando-se do interatoma interespecífico geral ([Dados suplementares S1e](#)), podendo ser acessado neste *link*.

Os pares de proteínas formados com o uso dos pares de domínios que interagem obtido em cada banco após os cálculos do *score* final sua normalização e a combinação sugeriram a ocorrência da interação entre proteínas destes organismos. O número das interações interespecíficas entre as proteínas dos genomas completos de *R. amblyommatis* versus *I. scapularis* sustentadas nos bancos de dados do interatoma, revelaram a ocorrência de 13151 interações totais, 387 proteínas *R. amblyommatis* versus 1252 proteínas do carrapato *I. scapularis*.

4.4 Construção do interatoma de proteínas de membrana

Para a construção de interatoma de proteínas de membrana, foram realizadas predições para obtermos as proteínas de *R. amblyommatis* secretadas, excretadas e/ou de superfície, conforme o Tópico 3.2. Estas sequências foram obtidas por meio do uso dos aplicativos SignalP 4.1, Tmhmm 2.0c, LipopB 1.0a, Psortb 3.0.4 e Phobius 1.01 (Tabelas 17-21) (Dados suplementares S1f-S5f) e após a integração destas proteínas (Tabela 22).

Para a construção do interatoma das proteínas de membrana (Dados suplementares S1g) (Tabela 23), foi realizado um filtro, conforme o Tópico 3.2.1, utilizando das proteínas preditas de membrana de *R. amblyommatis* como um filtro contra o interatoma geral intraespecífico (Dados suplementares S1e) contendo os pares de interação de proteínas entre a bactéria da espécie *R. amblyommatis* e o carrapato da espécie *I. scapularis*, obtidos após o cálculo do *score* combinado

4.4.1 Proteínas secretadas, excretadas e ou de superfície para *R. amblyommatis*

Com o uso do aplicativo SignalP 4.1, obteve-se o arquivo de saída contendo os IDs referentes às sequências de proteínas de membrana, adotando o parâmetro do *score* Y. Estas sequências de proteínas apresentaram características de proteínas de membranas com a presença do peptídeo sinal na região amino terminal e sua localização do sítio de clivagem que media a translocação de uma proteína do citoplasma para a membrana (Tabela 17).

Tabela 17 Resumo dos resultados de predição com o uso do aplicativo Signalp 4.1 para as sequências de proteínas provenientes do genoma da espécie *R. amblyommatis*.

ID proteína <i>R. amblyommatis</i>	Cmax	pos	Ymax	Pos	Smax	pos	Smean	D	?	Dmaxcut	Networks-use
KJV61002.1	0.112	20	0.103	27	0.133	21	0.080	0.094	N	0.510	SignalP-TM
KJV61003.1	0.109	51	0.149	26	0.366	23	0.184	0.162	N	0.510	SignalP-TM
KJV61010.1	0.201	35	0.379	35	0.962	17	0.831	0.592	Y	0.570	SignalP-noTM
KJV61016.1	0.557	23	0.687	23	0.921	21	0.852	0.765	Y	0.570	SignalP-noTM
KJV61061.1	0.371	27	0.542	27	0.876	19	0.731	0.631	Y	0.570	SignalP-noTM

Na Tabela 17, verifica-se um resumo dos resultados preliminares obtidos para as proteínas preditas com o uso do aplicativo Signalp 4.1 ([Dados suplementares S1f](#)), podendo ser acessado neste *link*.

Com o uso do aplicativo Tmhmm 2.0c, obteve-se o arquivo de saída contendo os IDs de referentes às sequências de proteínas de membrana, adotando o parâmetro do *PredHel*=0. Estas sequências de proteínas apresentaram características de proteínas contendo potenciais hélices transmembranas preditas por *N-best* (Tabela 18).

Tabela 18 Resumo dos resultados de predição com o uso do aplicativo Tmhmm 2.0c para as sequências de proteínas provenientes do genoma da espécie *R. amblyommatidis*.

ID proteína <i>R. amblyommatidis</i>	len	ExpAA	First60	PredHel	Topology
KJV61002.1	382	238.63	22.54	11	o30-52i59-81o85-107i120-142o146-168i194-216o229-251i260-282o287-309i322-344o348-370i
KJV61003.1	415	95.01	23.08	4	i23-45o274-296i316-338o381-403i
KJV61004.1	221	0.08	0.07	0	O
KJV61005.1	632	0.00	0.00	0	O
KJV61006.1	413	62.61	10.09	3	i51-73o78-97i109-131°

Na Tabela 18, verifica-se um resumo dos resultados preliminares obtidos para as proteínas de membrana preditas com o uso do aplicativo Tmhmm 2.0c ([Dados suplementares S2f](#)), podendo ser acessado neste *link*.

Com o uso do aplicativo LipoP 1.0a, obteve-se o arquivo de saída contendo os IDs de referentes às sequências de proteínas de membrana, adotando o parâmetro do *Margin*. Estas sequências de proteínas apresentaram características de lipoproteínas e possíveis peptídeos sinal para bactéria Gram negativas (Tabela 19).

Tabela 19 Resumo dos resultados de predição com o uso do aplicativo Lipop 1.0a para as sequências de proteínas provenientes do genoma da espécie *R. amblyommatidis*.

```
# KJV61002.1 CYT score=-0.200913
# Cut-off=-3
KJV61002.1 LipoP1.0:Best CYT 1 1 -0.200913
# NO PLOT made - less than 4 putative cleavage sites predicted
# KJV61003.1 SpI score=2.43321 margin=0.32802 cleavage=41-42
# Cut-off=-3
KJV61003.1 LipoP1.0:Best SpI 1 1 2.43321
KJV61003.1 LipoP1.0:Margin SpI 1 1 0.32802
KJV61003.1 LipoP1.0:Class TMH 1 1 2.10519
KJV61003.1 LipoP1.0:Class CYT 1 1 -0.200913
KJV61003.1 LipoP1.0:Signal CleavI 41 42 1.82876 # IGAAA|LIVVM
KJV61003.1 LipoP1.0:Signal CleavI 40 41 -0.343023 # MIGAA|ALIVV
KJV61003.1 LipoP1.0:Signal CleavI 39 40 -1.51164 # VMIGA|AALIV
KJV61003.1 LipoP1.0:Signal CleavI 43 44 -1.65201 # AAALI|VVMSV
KJV61003.1 LipoP1.0:Signal CleavI 42 43 -2.08645 # GAAAL|IVVMS
# KJV61004.1 CYT score=-0.200913
# Cut-off=-3
KJV61004.1 LipoP1.0:Best CYT 1 1 -0.200913
# NO PLOT made - less than 4 putative cleavage sites predicted
# KJV61005.1 CYT score=-0.200913
```

Cut-off=-3

Na Tabela 19, verifica-se um resumo dos resultados preliminares obtidos para as sequências de proteínas de membrana preditas com o uso do aplicativo LipoP 1.0a ([Dados suplementares S3f](#)), podendo ser acessado neste *link*.

Com o uso do aplicativo Psortb 3.0.4, obteve-se o arquivo de saída contendo os IDs referentes às sequências de proteínas de membrana, adotando os parâmetros dos *CytoplasmicMembrane* e *Extracellular*. Estas sequências de proteínas apresentaram características de proteínas de membranas citoplasmáticas e extracelulares, com a sub-localização célula havendo a presença de sequências sinalizadoras sugerindo a região carregada positivamente N-terminal (N- região) e a região central hidrofóbica (região H) (Tabela 20).

Tabela 20 Resumo dos resultados de predição com o uso do aplicativo Psortb 3.0.4 para as sequências de proteínas provenientes do genoma da espécie *R. amblyommatidis*.

ID proteína <i>R. amblyommatidis</i>	Localization	Score
KJV61002.1	CytoplasmicMembrane	10.00
KJV61003.1	CytoplasmicMembrane	10.00
KJV61004.1	CytoplasmicMembrane	9.96
KJV61005.1	Unknown	2.50
KJV61006.1	CytoplasmicMembrane	9.96
KJV61007.1	CytoplasmicMembrane	10.00
KJV61008.1	CytoplasmicMembrane	9.55
KJV61009.1	CytoplasmicMembrane	9.55
KJV61010.1	Extracellular	8.91
KJV61011.1	Unknown	2.50
KJV61012.1	Cytoplasmic	9.97
KJV61013.1	Unknown	2.50
KJV61014.1	Unknown	2.50
KJV61015.1	Cytoplasmic	7.50
KJV61016.1	Unknown	2.50

Na Tabela 20, verifica-se um resumo dos resultados obtidos para as sequências de proteínas de membrana preditas com o uso do aplicativo Psortb 3.0.4 ([Dados suplementares S4f](#)), podendo ser acessado neste *link*.

Com o uso do aplicativo Phobius 1.01, obteve-se o arquivo de saída contendo os IDs de identificação referentes às sequências de proteínas de membrana, adotando os parâmetros da $TM > 0$, que contém o número de segmentos transmembranas preditos e do $SP = Y$, indicando Y (YES) se um peptídeo sinal foi predito ou N (NO), caso não o seja. Estas sequências de proteínas apresentaram características de proteínas com características de topologia de proteína transmembrana combinada a um preditor de peptídeo sinal (Tabela 21).

Tabela 21 Resumo dos resultados de predição com o uso do aplicativo Phobius 1.01 para as sequências de proteínas provenientes do genoma da espécie *R. amblyommatidis*.

ID proteína <i>R. amblyommatidis</i>	TM	SP	PREDICTION
KJV61010.1	0	Y	n11-22c34/35°
KJV61016.1	0	Y	n4-17c22/23°
KJV61022.1	0	Y	n7-16c21/22°
KJV61024.1	0	0	O
KJV61027.1	1	0	i12-34°
KJV61043.1	1	Y	n6-17c25/26o201-219i

Na Tabela 21, verifica-se um resumo dos resultados das sequências de proteínas de membrana preditas com o uso do aplicativo Phobius 1.01 ([Dados suplementares S5f](#)), podendo ser acessado neste *link*.

Com o uso dos aplicativos Tmhmm 2.0c, LipopB 1.0a, SignalP 4.1, Phobius 1.01 e Psortb 3.0.4 foram obtidas 362, 334, 44, 363 e 462 sequências de proteínas (Dados suplementares S1f-S5f), respectivamente. Estes dados de sequências preditas, foram integradas e organizadas em arquivo único, contendo somente os IDs das sequências de proteínas preditas, sem repetição destes (Tabela 22).

Tabela 22 Resultados de predição e integração das sequências de proteínas secretadas, excretadas e ou de superfície da espécie *R. amblyommatidis*.

ID proteína de membrana <i>R. amblyommatidis</i>					KJV62208.1	KJV62515.1
KJV61002.1	KJV61223.1	KJV61468.1	KJV61741.1	KJV61972.1	KJV62216.1	KJV62531.1
KJV61003.1	KJV61224.1	KJV61471.1	KJV61747.1	KJV61973.1	KJV62219.1	KJV62535.1
KJV61004.1	KJV61225.1	KJV61474.1	KJV61751.1	KJV61974.1	KJV62223.1	KJV62536.1
KJV61006.1	KJV61230.1	KJV61479.1	KJV61753.1	KJV61977.1	KJV62226.1	KJV62537.1
KJV61007.1	KJV61232.1	KJV61482.1	KJV61754.1	KJV61984.1	KJV62230.1	KJV62541.1
KJV61008.1	KJV61233.1	KJV61483.1	KJV61755.1	KJV61986.1	KJV62235.1	KJV62544.1
KJV61009.1	KJV61235.1	KJV61484.1	KJV61757.1	KJV61990.1	KJV62242.1	KJV62545.1
KJV61010.1	KJV61236.1	KJV61485.1	KJV61759.1	KJV61993.1	KJV62243.1	KJV62547.1
KJV61016.1	KJV61237.1	KJV61486.1	KJV61760.1	KJV61994.1	KJV62244.1	KJV62548.1
KJV61020.1	KJV61238.1	KJV61487.1	KJV61762.1	KJV61996.1	KJV62245.1	KJV62549.1
KJV61021.1	KJV61240.1	KJV61488.1	KJV61765.1	KJV62004.1	KJV62246.1	KJV62550.1
KJV61022.1	KJV61247.1	KJV61498.1	KJV61770.1	KJV62007.1	KJV62247.1	KJV62551.1
KJV61025.1	KJV61248.1	KJV61499.1	KJV61773.1	KJV62011.1	KJV62252.1	KJV62554.1
KJV61027.1	KJV61255.1	KJV61500.1	KJV61774.1	KJV62013.1	KJV62253.1	KJV62556.1
KJV61029.1	KJV61270.1	KJV61503.1	KJV61775.1	KJV62016.1	KJV62258.1	KJV62557.1
KJV61031.1	KJV61277.1	KJV61504.1	KJV61776.1	KJV62022.1	KJV62261.1	KJV62561.1
KJV61032.1	KJV61278.1	KJV61510.1	KJV61777.1	KJV62024.1	KJV62264.1	KJV62562.1
KJV61035.1	KJV61279.1	KJV61511.1	KJV61778.1	KJV62025.1	KJV62265.1	KJV62563.1
KJV61043.1	KJV61280.1	KJV61516.1	KJV61779.1	KJV62026.1	KJV62268.1	KJV62570.1
KJV61045.1	KJV61281.1	KJV61517.1	KJV61783.1	KJV62029.1	KJV62269.1	KJV62573.1
KJV61046.1	KJV61286.1	KJV61518.1	KJV61784.1	KJV62030.1	KJV62271.1	KJV62575.1
KJV61050.1	KJV61289.1	KJV61520.1	KJV61785.1	KJV62031.1	KJV62272.1	KJV62576.1
KJV61051.1	KJV61290.1	KJV61521.1	KJV61792.1	KJV62034.1	KJV62275.1	KJV62578.1
KJV61052.1	KJV61291.1	KJV61522.1	KJV61795.1	KJV62035.1	KJV62276.1	KJV62583.1
KJV61054.1	KJV61294.1	KJV61523.1	KJV61797.1	KJV62036.1	KJV62277.1	KJV62584.1
KJV61057.1	KJV61297.1	KJV61524.1	KJV61808.1	KJV62038.1	KJV62282.1	KJV62586.1

KJV61058.1	KJV61298.1	KJV61530.1	KJV61811.1	KJV62040.1	KJV62283.1	KJV62591.1
KJV61059.1	KJV61302.1	KJV61534.1	KJV61817.1	KJV62043.1	KJV62290.1	KJV62593.1
KJV61061.1	KJV61304.1	KJV61536.1	KJV61818.1	KJV62044.1	KJV62298.1	KJV62594.1
KJV61063.1	KJV61305.1	KJV61538.1	KJV61820.1	KJV62045.1	KJV62299.1	KJV62597.1
KJV61066.1	KJV61307.1	KJV61546.1	KJV61822.1	KJV62046.1	KJV62311.1	KJV62602.1
KJV61067.1	KJV61314.1	KJV61547.1	KJV61824.1	KJV62047.1	KJV62312.1	KJV62616.1
KJV61072.1	KJV61317.1	KJV61548.1	KJV61825.1	KJV62048.1	KJV62314.1	KJV62632.1
KJV61081.1	KJV61318.1	KJV61551.1	KJV61826.1	KJV62049.1	KJV62315.1	KJV62634.1
KJV61082.1	KJV61319.1	KJV61552.1	KJV61827.1	KJV62051.1	KJV62320.1	KJV62635.1
KJV61083.1	KJV61328.1	KJV61553.1	KJV61829.1	KJV62052.1	KJV62323.1	KJV62636.1
KJV61093.1	KJV61332.1	KJV61554.1	KJV61830.1	KJV62053.1	KJV62324.1	KJV62645.1
KJV61100.1	KJV61336.1	KJV61558.1	KJV61831.1	KJV62054.1	KJV62326.1	KJV62648.1
KJV61102.1	KJV61338.1	KJV61559.1	KJV61832.1	KJV62056.1	KJV62329.1	KJV62652.1
KJV61103.1	KJV61343.1	KJV61574.1	KJV61833.1	KJV62057.1	KJV62333.1	KJV62653.1
KJV61104.1	KJV61344.1	KJV61577.1	KJV61834.1	KJV62058.1	KJV62334.1	KJV62654.1
KJV61106.1	KJV61349.1	KJV61587.1	KJV61835.1	KJV62059.1	KJV62335.1	KJV62655.1
KJV61110.1	KJV61351.1	KJV61589.1	KJV61837.1	KJV62060.1	KJV62336.1	KJV62658.1
KJV61117.1	KJV61355.1	KJV61590.1	KJV61841.1	KJV62061.1	KJV62337.1	KJV62659.1
KJV61118.1	KJV61357.1	KJV61592.1	KJV61844.1	KJV62064.1	KJV62338.1	KJV62668.1
KJV61120.1	KJV61360.1	KJV61593.1	KJV61845.1	KJV62066.1	KJV62345.1	KJV62671.1
KJV61122.1	KJV61361.1	KJV61597.1	KJV61847.1	KJV62069.1	KJV62349.1	KJV62672.1
KJV61124.1	KJV61363.1	KJV61601.1	KJV61850.1	KJV62075.1	KJV62369.1	KJV62675.1
KJV61126.1	KJV61364.1	KJV61602.1	KJV61852.1	KJV62080.1	KJV62370.1	KJV62677.1
KJV61127.1	KJV61365.1	KJV61605.1	KJV61854.1	KJV62081.1	KJV62371.1	KJV62679.1
KJV61128.1	KJV61366.1	KJV61608.1	KJV61855.1	KJV62083.1	KJV62373.1	KJV62687.1
KJV61136.1	KJV61372.1	KJV61614.1	KJV61857.1	KJV62089.1	KJV62374.1	KJV62689.1
KJV61137.1	KJV61378.1	KJV61623.1	KJV61858.1	KJV62097.1	KJV62375.1	KJV62691.1
KJV61138.1	KJV61379.1	KJV61631.1	KJV61859.1	KJV62098.1	KJV62378.1	KJV62693.1
KJV61139.1	KJV61380.1	KJV61635.1	KJV61860.1	KJV62100.1	KJV62381.1	KJV62697.1
KJV61144.1	KJV61382.1	KJV61636.1	KJV61874.1	KJV62101.1	KJV62382.1	KJV62698.1
KJV61145.1	KJV61385.1	KJV61638.1	KJV61876.1	KJV62102.1	KJV62388.1	KJV62699.1
KJV61146.1	KJV61387.1	KJV61641.1	KJV61877.1	KJV62104.1	KJV62397.1	KJV62700.1
KJV61148.1	KJV61392.1	KJV61644.1	KJV61878.1	KJV62105.1	KJV62398.1	KJV62707.1
KJV61149.1	KJV61396.1	KJV61645.1	KJV61882.1	KJV62109.1	KJV62413.1	KJV62708.1
KJV61150.1	KJV61397.1	KJV61646.1	KJV61884.1	KJV62115.1	KJV62416.1	KJV62712.1
KJV61151.1	KJV61398.1	KJV61651.1	KJV61885.1	KJV62116.1	KJV62417.1	KJV62715.1
KJV61152.1	KJV61400.1	KJV61657.1	KJV61887.1	KJV62118.1	KJV62418.1	KJV62745.1
KJV61157.1	KJV61401.1	KJV61663.1	KJV61890.1	KJV62119.1	KJV62422.1	KJV62746.1
KJV61158.1	KJV61402.1	KJV61672.1	KJV61895.1	KJV62121.1	KJV62423.1	KJV62749.1
KJV61159.1	KJV61412.1	KJV61679.1	KJV61897.1	KJV62122.1	KJV62425.1	KJV62750.1
KJV61160.1	KJV61413.1	KJV61680.1	KJV61900.1	KJV62128.1	KJV62426.1	KJV62756.1
KJV61161.1	KJV61414.1	KJV61681.1	KJV61901.1	KJV62130.1	KJV62427.1	KJV62759.1
KJV61163.1	KJV61418.1	KJV61682.1	KJV61909.1	KJV62138.1	KJV62431.1	KJV62762.1
KJV61164.1	KJV61425.1	KJV61693.1	KJV61910.1	KJV62139.1	KJV62440.1	KJV62763.1
KJV61165.1	KJV61426.1	KJV61695.1	KJV61914.1	KJV62140.1	KJV62441.1	KJV62765.1
KJV61167.1	KJV61428.1	KJV61696.1	KJV61919.1	KJV62142.1	KJV62442.1	KJV62766.1
KJV61168.1	KJV61432.1	KJV61699.1	KJV61925.1	KJV62143.1	KJV62448.1	KJV62776.1
KJV61170.1	KJV61437.1	KJV61703.1	KJV61928.1	KJV62145.1	KJV62451.1	KJV62783.1
KJV61174.1	KJV61439.1	KJV61706.1	KJV61929.1	KJV62147.1	KJV62458.1	KJV62786.1
KJV61175.1	KJV61443.1	KJV61707.1	KJV61933.1	KJV62149.1	KJV62460.1	KJV62789.1
KJV61182.1	KJV61445.1	KJV61708.1	KJV61934.1	KJV62150.1	KJV62462.1	KJV62791.1
KJV61185.1	KJV61447.1	KJV61711.1	KJV61935.1	KJV62151.1	KJV62473.1	KJV62792.1
KJV61186.1	KJV61448.1	KJV61713.1	KJV61937.1	KJV62159.1	KJV62474.1	KJV62796.1
KJV61194.1	KJV61450.1	KJV61716.1	KJV61943.1	KJV62161.1	KJV62475.1	KJV62797.1
KJV61195.1	KJV61451.1	KJV61717.1	KJV61945.1	KJV62168.1	KJV62480.1	KJV62799.1
KJV61200.1	KJV61452.1	KJV61720.1	KJV61950.1	KJV62169.1	KJV62483.1	KJV62804.1
KJV61203.1	KJV61453.1	KJV61721.1	KJV61951.1	KJV62176.1	KJV62487.1	KJV62809.1
KJV61206.1	KJV61454.1	KJV61726.1	KJV61953.1	KJV62180.1	KJV62493.1	KJV62811.1

KJV61208.1	KJV61455.1	KJV61727.1	KJV61958.1	KJV62182.1	KJV62496.1	KJV62812.1
KJV61209.1	KJV61456.1	KJV61728.1	KJV61959.1	KJV62189.1	KJV62502.1	KJV62813.1
KJV61215.1	KJV61457.1	KJV61735.1	KJV61960.1	KJV62195.1	KJV62504.1	KJV62818.1
KJV61220.1	KJV61461.1	KJV61738.1	KJV61961.1	KJV62197.1	KJV62505.1	KJV62819.1
KJV61221.1	KJV61463.1	KJV61739.1	KJV61968.1	KJV62200.1	KJV62511.1	KJV62820.1
KJV61222.1	KJV61464.1	KJV61740.1	KJV61969.1	KJV62205.1	KJV62512.1	KJV62821.1

Na Tabela 22, verifica-se os resultados obtidos para as sequências de proteínas de membrana para a espécie *R. amblyommatis* integradas, que foram preditas com o uso dos aplicativos de predição de proteínas de membrana, revelando no total 632 proteínas únicas de membrana presentes no genoma desta bactéria, secretadas, excretadas e ou de superfície.

4.4.2 Interatoma de proteínas de membrana

Para a obtenção do interatoma de proteínas de membrana, o arquivo contendo os IDs das sequências de proteínas do interatoma interespecífico geral após o *score* combinado (Dados suplementares S1e) foi filtrado no arquivo contendo os IDs das sequências de proteínas de membrana de *R. amblyommatis* preditas e integradas (Tabela 22) para formarem no arquivo de saída os pares de interação com os IDs totais das sequências de proteínas de membrana da bactéria que interagem com as sequências de proteínas do carrapato (Dados suplementares S1g) (Tabela 23).

Tabela 23 Resumo dos resultados de predição e integração das sequências de proteínas secretadas, excretadas e ou de superfície filtradas no interatoma interespecífico de proteínas geral entre as espécies *R. amblyommatis* vs *I. scapularis*.

<i>Rickettsia amblyommatis</i>		<i>Ixodes scapularis</i>	
Id proteína	Nome	Id proteína	Nome
KJV61004.1	ABC transporter family protein	ISCW000200-PA.1	Adenylate kinase isoenzyme, putative
KJV61004.1	ABC transporter family protein	ISCW000565-PA.1	ABC transporter, putative
KJV61004.1	ABC transporter family protein	ISCW000765-PA.1	Half-ABC transporter, putative
KJV61004.1	ABC transporter family protein	ISCW000841-PA.1	Multidrug resistance protein, putative
KJV61004.1	ABC transporter family protein	ISCW001227-PA.1	Putative uncharacterized protein
KJV61004.1	ABC transporter family protein	ISCW001366-PA.1	Adenylate kinase isoenzyme, putative
KJV61004.1	ABC transporter family protein	ISCW001681-PA.1	ABC transporter, putative (Fragment)
KJV61004.1	ABC transporter family protein	ISCW002474-PA.1	ABC transporter, putative

Na Tabela 23, verifica-se o resumo para os resultados obtidos dos pares de proteínas de membrana da bactéria que interagem com as proteínas do carrapato, sendo representadas pelos IDs de identificação e o nome da proteína para cada par de interação ([Dados suplementares S1g](#)), podendo ser acessado neste *link*.

O interatoma de proteínas de membrana revelaram uma rede consenso de interações interespecíficas entre as sequências de proteínas de membrana de *R. amblyommatis* versus *I. scapularis*, ocorrendo 3456 interações únicas, preditas entre 91 proteínas da bactéria que interagiram com 980 proteínas do carrapato.

4.5 Construção do interatoma de proteínas de membrana ortólogas

Para a construção do interatoma de proteínas de membrana contendo grupos ortólogos, o interatoma de proteínas de membrana (Dados suplementares S1g) foi utilizado como um filtro. Dessa forma, os ortólogos foram obtidos anteriormente por meio de predições utilizando o OrthoMCL 2.0, conforme o Tópico 3.3 (Dados não publicados). Esta análise revelou a ocorrência de 36 grupos ortólogos de sequências de proteínas para as espécies de *Rickettsia* que infectam os carrapatos separadamente de 10 grupos ortólogos para as que infectam os insetos (Tabela 24).

Estes dados após serem filtrados no interatoma de proteínas de membrana de *R. amblyommatis*, conforme apresentado no Tópico 3.3.1, revelaram a presença de 2 clusters de proteínas ortólogas (cluster 1050 e 1816) pertencentes unicamente, as espécies de *Rickettsia* que infectam os carrapatos. Estes resultados revelaram uma rede consenso apresentando os pares de interação do interatoma das proteínas de membrana ortólogas para *R. amblyommatis* e as espécies de *Rickettsia* que infectam os carrapatos (Tabela 25) (Dados suplementares S1h).

4.5.1 Predição de proteínas ortólogas de *Rickettsia* spp., que infectam os carrapatos

Com o uso do aplicativo OrthoMCL 2.0, foi obtido o arquivo de saída contendo os grupos de ortólogos e os IDs totais das sequências de proteínas de *Rickettsia* spp. que infectam os carrapatos e em parte, carrapatos e insetos (Tabela 24).

Tabela 24 Grupos de cluster de ortólogos preditos para proteínas pertencentes as espécies de *Rickettsia* que infectam carrapatos e insetos.

Cluster <i>Rickettsia</i> Carrapatos (amb, end, buc e mon)				Cluster <i>Rickettsia</i> Insetos (amb, fel e pro)	
cluster1000	cluster1718	cluster2012	cluster2032	cluster1818	cluster2176
cluster1005	cluster1807	cluster2016	cluster2033	cluster1826	cluster2198
cluster1006	cluster1816	cluster2018	cluster2042	cluster2144	cluster2202
cluster1011	cluster1880	cluster2019	cluster2048	cluster2146	cluster2223
cluster1018	cluster1995	cluster2020	cluster2051	cluster2166	cluster2229
cluster1050	cluster1998	cluster2024	cluster2061	cluster2170	
cluster1052	cluster2002	cluster2029	cluster2064		
cluster1061	cluster2006	cluster2030	cluster2067		
cluster1134	cluster2009	cluster2031	cluster2072		

Na Tabela 24, verifica-se os resultados preliminares obtidos com o uso do OrthoMCL 2.0 sugerindo haver diferentes mecanismos de invasão celular e/ou sobrevivência ao ambiente intracelular que utiliza destes grupos de proteínas diferentes. Considerando apenas grupos de ortólogos presentes nas espécies de *Rickettsia* que infectam os carrapatos agrupadas em uma mesma classe, a maior parte dos grupos de ortólogos são compartilhados entre as quatro bactérias analisadas *R. monanense*, *R. endosymbiotic*, *R. buchneri* e *R. amblyommatis*, totalizaram 36 clusteres. Enquanto, entre a espécie *R. amblyommatis* e as duas espécies *R. prowazekii* e *R. felis*, que infectam os insetos, apenas 10 grupos de ortólogos são compartilhados.

4.5.2 Interatoma de proteínas de membrana ortólogas

Não foi observado a formação de grupo de proteínas ortólogas compartilhado entre todas as espécies de *Rickettsia* utilizadas. Os resultados obtidos dos clusters ortólogos (Tabela 24), foram utilizados para a construção do interatoma de proteína ortólogos de membrana, com o filtro entre estes dados e o interatoma de proteínas de membrana (Dados suplementares S1g). O arquivo de saída contém os pares de interação com os IDs das sequências de proteínas de membrana da bactéria que interagem com as sequências de proteínas do carrapato, juntamente com o grupos ortólogos e os IDs das sequências de proteínas, com os quais possui relação com os pares de IDs de outras espécies de *Rickettsia* que infectam os carrapatos (Tabela 25).

Tabela 25 Resumo dos resultados de integração e filtro dos ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam os carrapatos no interatoma de proteínas de membrana.

Cluster ortólogos	<i>Rickettsia amblyommatis</i>		<i>Ixodes scapularis</i>	
	Id proteína	Nome	Id proteína	Nome
cluster1816	KJV62536.1	TPR repeat family Protein	ISCW000002-PA.1	FKBP-type peptidyl-prolyl cis-trans isomerase
cluster1816	KJV62536.1	TPR repeat family protein	ISCW000086-PA.1	Translation elongation fator
cluster1816	KJV62536.1	TPR repeat family protein	ISCW004349-PA.1	GTPase Rho, putative
cluster1050:	KJV62578.1	His Kinase A	ISCW008384-PA.1	Two component histidine kinase

Na Tabela 25, verifica-se os resultados para os pares de proteínas de membrana que interagem no interatoma e possui ortólogos para a *R. amblyommatis* e outras espécies de *Rickettsia* que infectam os carrapatos, sendo representadas pelos IDs de identificação e seus grupos de ortólogos preditos com o uso do aplicativo OrthoMCL 2.0 ([Dados suplementares S1h](#)), podendo ser acessado neste *link*.

Não foram observados a presença de proteínas ortólogas na rede da espécie *R. amblyommatis* para os clusters de espécies de *Rickettsia* que infectam os insetos. O número de interações entre as proteínas de membrana ortólogas de *R. amblyommatis* para espécies de *Rickettsia* que infectam os carrapatos sustentadas nos bancos de dados do interatoma, revelaram 100 interações únicas, preditas entre 2 proteínas da bactéria que interagiram com 100 proteínas do carrapato para os clusters 1816 e 1050 preditos utilizando do OrthoMcl 2.0.

4.5.2.1 Análise de similaridade das proteínas ortólogas de membrana

Com o uso do TBLASTn contra as sequências de nucleotídeos e do BLASTp contra as sequências de aminoácidos de proteínas provenientes dos genomas das espécies *R. amblyommatis*, *R. endonssibiontic*, *I. scapularis*, *R. buchenneri* e *R. monanencis*, além das espécies *R. felis* e *R. prowazekii*, que infectam os insetos, conforme apresentando no Tópico 3.3.2. Para a seleção e obtenção das sequências gênicas similares, foi adotado do valor $> bit\ score$ do alinhamento, revelando as sequências de nucleotídeos presentes nos genomas das espécies referentes às sequências de proteínas ortólogas presentes nos cluster 1050 e 1816 (Tabela 26).

Tabela 26 Resumo dos resultados da análise de similaridade das sequências de proteínas de membrana ortólogas para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam os carrapatos adotando o alinhamento do TBLASTn.

Cluster	ID_proteína <i>R. amblyommatis</i>	<i>Rickettsia</i> spp.	ID_gene <i>Rickettsia</i> spp.	Início da sequência	Final da sequência	Valor de > <i>Bit score</i>
1050	KJV62261.1	<i>R. prowazekii</i>	Chromosome	480272	479523	50,8
1050	KJV62261.1	<i>R. felis</i>	rfPED.contig.0_1	1263301	1262663	50,4
1050	KJV62261.1	<i>R. monanensce</i>	Chromosome	923790	922846	594
1050	KJV62261.1	<i>R. endosymbiotic</i>	Chromosome	731400	730456	599
1050	KJV62261.1	<i>R. buchneri</i>	Contig124	7153	8097	599
1050	KJV62261.1	<i>R. amblyommatis</i>	raACPA.contig.0_1	995518	994562	649
1050	KJV62578.1	<i>R. prowazekii</i>	Chromosome	480161	479529	57,8
1050	KJV62578.1	<i>R. felis</i>	rfPED.contig.0_1	1263301	1262663	57
1050	KJV62578.1	<i>R. monanensce</i>	Chromosome	923790	922843	596
1050	KJV62578.1	<i>R. endosymbiotic</i>	Chromosome	731400	730453	592
1050	KJV62578.1	<i>R. buchneri</i>	Contig124	7153	8100	592
1050	KJV62578.1	<i>R. amblyommatis</i>	raACPA.contig.0_1	1253921	1252974	617
1816	KJV62536.1	<i>R. prowazekii</i>	Chromosome	677776	678240	25,8
1816	KJV62536.1	<i>R. felis</i>	rfPED.contig.0_1	673041	673925	165
1816	KJV62536.1	<i>R. monanensce</i>	Chromosome	934765	935460	394
1816	KJV62536.1	<i>R. endosymbiotic</i>	Chromosome	714978	716144	707
1816	KJV62536.1	<i>R. buchneri</i>	Contig124	23575	22409	707
1816	KJV62536.1	<i>R. amblyommatis</i>	raACPA.contig.0_1	1202298	1201132	707

Na Tabela 26, verifica-se os resultados do alinhamento obtidos com o uso do TBLASTn para os pares de proteínas de membrana que interagem no interatoma e possui ortólogos para a *R. amblyommatis* e outras espécies de *Rickettsia* que infectam os carrapatos. As sequências representadas pelos IDs das proteínas de *R. amblyommatis*, os IDs das sequências de nucleotídeos de genes das proteínas similares de outras *Rickettsia* spp. e os valores de > *Bit score* para os grupos de ortólogos 1050 e 1816.

Com o uso do BLASTp foram obtidas as sequências de aminoácidos de proteínas similares para cada conjunto de grupo ortólogo obtidas nos cluster 1050 e 1816 entre as espécies de *Rickettsia* que infectam os carrapatos, acrescentando as espécies que infectam os insetos. . Para a seleção e obtenção das sequências proteicas similares, foi adotado os valores de Confiança \geq 70 e da Indentidade \leq 1e-10 do alinhamento, revelando as sequências de nucleotídeos presentes nos genomas das espécies que possuem às sequências de proteínas ortólogas presentes nos cluster 1050 e 1816 (Tabela 27).

Tabela 27 Resumo dos resultados da análise de similaridade das sequências de proteínas de membrana ortólogas para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam os carrapatos adotando o alinhamento do BLASTp.

Cluster	ID_proteína <i>R. amblyommatis</i>	<i>Rickettsia</i> spp.	ID_proteína <i>Rickettsia</i> spp.	Confiança>=70	Identidade<=1e-10
1050	KJV62261.1	<i>R. monanensce</i>	CEO17519	92.14	0.0
1050	KJV62261.1	<i>R. monanensce</i>	CEO17558	91.82	0.0
1050	KJV62261.1	<i>R. endosymbiotic</i>	EER21700	92.77	0.0
1050	KJV62261.1	<i>R. endosymbiotic</i>	EER21904	92.86	0.0
1050	KJV62261.1	<i>R. endosymbiotic</i>	EER22377	89.72	5,00E-67
1050	KJV62261.1	<i>R. buchneri</i>	KDO02624	92.77	0.0
1050	KJV62261.1	<i>R. buchneri</i>	KDO02998	93.75	4,00E-25
1050	KJV62261.1	<i>R. buchneri</i>	KDO02999	88.68	1,00E-30
1050	KJV62261.1	<i>R. amblyommatis</i>	KJV62261.1	100.00	0.0
1050	KJV62261.1	<i>R. amblyommatis</i>	KJV62578.1	92.14	0.0
1050	KJV62578.1	<i>R. monanensce</i>	CEO17519	96.52	0.0
1050	KJV62578.1	<i>R. monanensce</i>	CEO17558	96.20	0.0
1050	KJV62578.1	<i>R. endosymbiotic</i>	EER21700	96.20	0.0
1050	KJV62578.1	<i>R. endosymbiotic</i>	EER21904	93.57	0.0
1050	KJV62578.1	<i>R. endosymbiotic</i>	EER22377	91.67	3,00E-58
1050	KJV62578.1	<i>R. buchneri</i>	KDO02624	96.20	0.0
1050	KJV62578.1	<i>R. buchneri</i>	KDO02998	100.00	1,00E-17
1050	KJV62578.1	<i>R. buchneri</i>	KDO02999	76.19	2,00E-30
1050	KJV62578.1	<i>R. amblyommatis</i>	KJV62261.1	92.11	0.0
1050	KJV62578.1	<i>R. amblyommatis</i>	KJV62578.1	100.00	0.0
1816	KJV62536.1	<i>R. monanensce</i>	CEO17529	85.49	4,00E-117
1816	KJV62536.1	<i>R. endosymbiotic</i>	EER21220	76.09	8,00E-44
1816	KJV62536.1	<i>R. endosymbiotic</i>	EER21691	100.00	0.0
1816	KJV62536.1	<i>R. endosymbiotic</i>	EER21723	74.42	2,00E-15
1816	KJV62536.1	<i>R. endosymbiotic</i>	EER22348	78.41	0.0
1816	KJV62536.1	<i>R. buchneri</i>	KDO02305	76.09	7,00E-44
1816	KJV62536.1	<i>R. buchneri</i>	KDO02633	100.00	0.0
1816	KJV62536.1	<i>R. buchneri</i>	KDO03258	74.42	1,00E-15
1816	KJV62536.1	<i>R. amblyommatis</i>	KJV62251.1	82.08	3,00E-96
1816	KJV62536.1	<i>R. amblyommatis</i>	KJV62536.1	100.00	0.0
1816	KJV62536.1	<i>R. amblyommatis</i>	KJV62562.1	70.69	2,00E-168

Na Tabela 27, verifica-se os resultados do alinhamento obtidos com o uso do BLASTp para os pares de proteínas de membrana que interagem no interatoma e possui ortólogos para a *R. amblyommatis* e outras espécies de *Rickettsia* que infectam os carrapatos. As sequências representadas pelos IDs das proteínas de *R. amblyommatis*, os IDs das sequências de aminoácidos das proteínas similares de outras *Rickettsia* spp. e os valores de Confiança e Identidade para os grupos de ortólogos 1050 e 1816.

4.5.3 Aspectos evolutivos das proteínas ortólogas de membrana

As sequências de nucleotídeos obtidas com o uso do TBLASTn (Dados não publicados) como similares para as proteínas ortólogas presentes nos cluster 1050 e 1816, foram invertidas na ordem de leitura correta 5' – 3' em 3' – 5, com o uso do aplicativo *Reverse* (Dados não publicados). Para a obtenção das sequências de aminoácidos das proteínas, estas sequências de nucleotídeos foram traduzidas com o uso do aplicativo *Translate* (Dados não publicados), conforme o Tópico 3.3.3.

4.5.3.1 Alinhamentos das proteínas ortólogas de membrana de *Rickettsia* spp.

Com o uso de alinhamentos de sequência Clustal, obteve-se em parte os aspectos evolutivos das relações filogenéticas entre as sequências de proteínas ortólogas das espécies de *Rickettsia* que infectam os carrapatos presentes no interatoma de proteínas de membrana. As sequências de proteínas preditas para os clusters 1050 e 1816 (Dados não publicados), foram submetidas a diversos alinhamentos múltiplos de sequência, conforme apresentado no Tópico 3.3.2.

Dessa forma por meio de alinhamentos múltiplos de sequências pudemos avaliar os aspectos evolutivos das sequências genicas e proteicas com o uso dos aplicativos ClustalW e Clustal Ômega (Anexo 1). Os alinhamentos obtidos sugeriram os domínios conservados, incluindo resíduos críticos para a função das proteínas, comparando as proteínas de interesse de forma mais detalhada com os membros da mesma família.

Após isto, pudemos avaliar os aspectos filogenéticos das sequências genicas e proteicas obtidas para serem utilizadas no aplicativo MEGA 7.26 adotando do método do *Maximum Likelihood Tree*, para obtermos os resultados das arvores filogenéticas de ortólogos obtidos com o uso do aplicativo OrthoMCL 2.0, em especial as sequências presentes no interatoma de proteínas de membrana de *R. amblyommatis* contendo sequências proteicas e gênicas similares de outras *Rickettsia* spp. (Anexo 2).

No Anexo 3, verifica-se as árvores filogenéticas para as proteínas similares do cluster ortólogo 1050, com destaque para as duas proteínas KJV62578.1 e KJV62261.1 *His kinase A domain protein*, e no cluster ortólogo 1816 com destaque para a proteína KJV62536.1 *Tetratricopeptide repeat protein family*, apresentando divergências entre os clados formados pelas sequências de aminoácidos de proteínas e das sequências gênicas de nucleotídeos

pertencentes as espécies de *Rickettsia* que infectam os carrapatos, em relação as espécies que infectam os insetos.

A ocorrência de proteínas ortológicas no interatoma de proteínas de membrana para a espécie *R. amblyommatis* que infectam os carrapatos versus insetos, sugerem diferenças entre as espécies comparadas *R. amblyommatis*, *R. endonsimbiontic I. scapularis*, *R. bucheneri* e *R. monascensis* versus *R. prowazeki* e *R. felis*, respectivamente.

Cluster 1050

Na árvore filogenética (a), verifica-se os resultados obtidos para o estudo evolutivo das sequências de aminoácidos da proteína KJV62578.1 e o seu ortólogo KJV62261.1 pertencente a espécie *R. amblyommatis* contra o seu próprio genoma, alinhando-se com as suas respectivas sequências gênicas (amb1-amb1 e amb2-amb2), além das sequências gênicas das outras cinco espécies de *Rickettsia*, sendo: (amb1-buc1 e amb2-buc2), (amb1-end1 e amb2-end2), (amb1-mon1 e amb2-mon2), (amb1-fel1 e amb2-fel2) e (amb1-pro1 e amb2-pro2), provenientes dos genomas das espécies *R. endonsimbiontic I. scapularis*, *R. bucheneri* e *R. monascensis*, que infectam os carrapatos versus *R. prowazeki* e *R. felis*, que infectam insetos, respectivamente.

Na árvore filogenética (b), verifica-se os resultados para o estudo evolutivo das sequências de aminoácidos das proteínas KJV62578.1 e o seu ortólogo KJV62261.1 pertencentes a espécie *R. amblyommatis*, as proteínas EER21700 e seu o parólogo EER21904, pertencentes a espécie *R. endonsimbiontic I. scapularis*, a proteína KDO02624 pertencente a espécie *R. bucheneri*, as proteínas CEO17519 e seu parólogo CEO17558, pertencente a espécie *R. monanensci*, respectivamente, alinhando-se com as sequências de proteínas obtidas com a tradução das sequências gênicas. Enquanto isto, as sequências de proteínas traduzidas das sequências gênicas (amb1-fel1 e amb2-fel1) e (amb1-pro1 e amb2-pro2) obtidas também pelo alinhamento do TBLASTn, e pertencentes as espécies *R. felis* e *R. prowazekii*, respectivamente, se ramificaram separadamente, sugerindo a divergência entre as espécies de *Rickettsia* que infectam carrapato em relação as que infectam os insetos.

Na árvore filogenética (c), verifica-se os resultados evolutivos obtidos para as sequências de aminoácidos das proteínas ortólogas preditas com o uso do aplicativo OrthoMcl 2.0, juntamente com as sequências de proteínas similares obtidas com o uso do BLASTp, sendo

as proteínas KJV62578.1 e o seu ortólogo KJV62261.1 pertencentes a espécie *R. amblyommatis*, as proteínas EER21700 e os seus parálogos EER21904 e EER22377, pertencentes a espécie *R. endonsimbiontic I. scapularis*, a proteína KDO02624 e os seus parálogos KDO02998 e KDO02999 pertencentes a espécie *R. buchneri*, as proteínas CEO17519 e o seu parólogo CEO17558, pertencente a espécie *R. monanensci*, respectivamente, alinhando-se com as sequências de proteínas obtidas com a tradução das sequências gênicas. Nesta árvore foi possível observarmos apenas a presença de sequências de proteínas provenientes de espécies de *Rickettsia*, que infectam os carrapatos.

Na árvore filogenética (d), verifica-se os resultados gerais evolutivos obtidos para as sequências de aminoácidos das proteínas ortólogas preditas com o uso do aplicativo OrthoMcl, juntamente com as sequências de proteínas similares obtidas com o uso do BLASTp, e das sequências de proteínas traduzidas das sequências gênicas provenientes do uso do TBLASTn, sendo.

Cluster 1816

Na árvore filogenética (e), verifica-se os resultados do estudo evolutivo para as sequências de aminoácidos da proteína KJV62536.1 pertencente à *R. amblyommatis* contra o seu próprio genoma, alinhando-se com a sua respectiva sequência gênica (amb-amb), além das sequências gênicas das outras cinco espécies de *Rickettsia*, sendo (amb-buc), (amb-end), (amb-mon), (amb-fel) e (amb-pro), provenientes dos genomas das espécies *R. endonsimbiontic I. scapularis*, *R. buchneri* e *R. monascensis*, que infectam os carrapatos versus *R. prowazeki* e *R. felis*, que infectam insetos, respectivamente.

Na árvore filogenética (f), verifica-se os resultados do estudo evolutivo para as sequências de aminoácidos das proteínas ortólogas preditas com o uso do aplicativo OrthoMcl, sendo KJV62536.1 pertencentes a espécie *R. amblyommatis*, EER21691 e seu parólogo EER22448, pertencente a espécie *R. endonsimbiontic I. scapularis*, KDO02633 pertencente a espécie *R. buchneri*, e CEO17529 pertencente a espécie *R. monanensci* respectivamente, alinhando-se com as sequências de proteínas obtidas com a tradução das sequências gênicas. Enquanto isto, as sequências gênicas (amb-fel) e (amb-pro) obtidas também pelo alinhamento do TBLASTn, e pertencentes as espécies *R. felis* e *R. prowazekii*, respectivamente, se

ramificaram separadamente sugerindo a divergência entre as espécies de *Rickettsia* que infectam carrapato em relação as que infectam os insetos.

Na árvore filogenética (g), verifica-se os resultados do estudo evolutivo para as sequências de aminoácidos das proteínas ortólogas preditas com o uso do aplicativo OrthoMcl, junto com as sequências de proteínas similares obtidas com o uso do BLASTp, sendo: KJV62536.1 e o seus parálogos KJV62251.1 e KJV62562.1 pertencentes a espécie *R. amblyommatis*, a proteína EER21691 e os seus parálogos EER22348, EER21723 e EER21220, pertencentes a espécie *R. endosymbiotic I. scapularis*, a proteína KDO02633 e os seus parálogos KDO02305 e KDO03258 pertencentes a espécie *R. buchneri*, a proteína CEO17529 pertencente a espécie *R. monanensci*. Nesta árvore filogenética observamos apenas a presença de sequências de proteínas provenientes de espécies de *Rickettsia* que infectam os carrapatos.

Na árvore filogenética (h), verifica-se os resultados gerais do estudo evolutivo obtidos para as sequências de proteínas ortólogas preditas, junto com as sequências de proteínas similares obtidas com o uso do BLASTp, e das sequências de proteínas traduzidas das sequências gênicas provenientes do uso do TBLASTn.

4.6 Identificação de interações conservadas

Com o aplicativo AgBase-GOanna 2.2 pudemos averiguar as anotações de proteína e anotações de ontologia gênica no Uniprot, descrevendo as funções moleculares e os processos biológicos de proteínas destas interfaces de interação, conforme apresentado no Tópico 3.4. Os arquivos de saída (Dados suplementares S1i-S3i) e o que eles contêm podem ser acessados nestes *links*: o arquivo Excel contendo as entradas, suas correspondências e as anotações de ontologia gênica GO para as correspondências ([Dados suplementares S1i](#)); o arquivo Align.html contendo os alinhamentos ([Dados suplementares S2i](#)); e o arquivo Sliminput.txt contendo um resumo dos resultados para as anotações GO ([Dados suplementares S3i](#)).

No arquivo de saída Excel para cada entrada, o GOanna 2.2 reporta: o ID da proteína de entrada; o ID do arquivo de entrada; o cabeçalho dos acessos da sequência de proteína; o *Hit* das correspondências BLAST da entrada; e o nome da proteína e suas sequências correspondentes contendo as informações estatísticas sobre a correspondência. Isso é repetido

para o número de sequências de destino escolhido ao executar o GOanna 2.2 (o padrão é de 10 correspondências) (Dados suplementares S1i).

No arquivo de saída Aling.html, a primeira entrada foi a sequência de proteína de carrapato da espécie: *I. scapularis*: >ISCW000012-PA.1 data base 7227 |Spectrin alpha chain, *Drosophila melanogaster* (fruit fly), apresentando como primeira correspondência a proteína: >P13395 alpha spectrin, com a qual no alinhamento BLAST, obteve-se a pontuação do *score*=9293 e o *E-value*=0.0 indicando um bom pareamento das bases de aminoácidos (Dados suplementares S2i).

No arquivo de saída Sliminput.txt contendo um resumo dos resultados para as anotações GO, observou-se quais sequências de proteínas provenientes do genoma do carrapato *I. scapularis*, compuseram vias moleculares da célula deste ácaro, sendo agrupadas conforme as letras: B processo biológico, C Compartimento celular, F função molecular (Dados suplementares S3i). Estes resultados foram utilizados para as análises de enriquecimento hipergeométrico servindo como base para o teste de FDR, e para isto sendo modificado por curadoria manual e formatado substituindo o /GO por espaço = espaço, e as letras por nada. A partir disto, servindo como base, pode-se visualizar as redes do interatoma interespecífico entre as proteínas da bactéria *R. amblyommatis* que interagiram com as proteínas do carrapato *I. scapularis*.

4.7 Visualização das redes

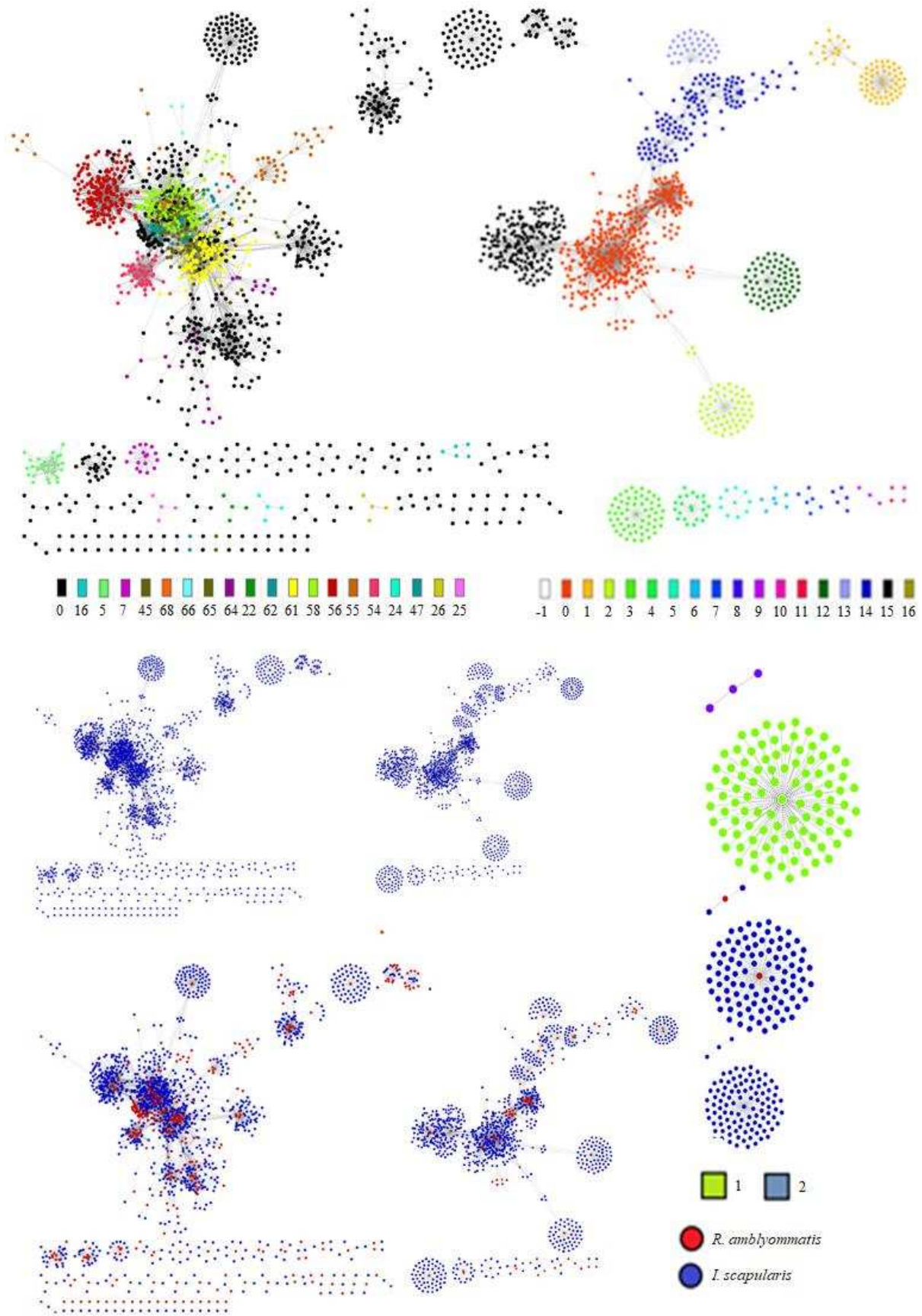
Para a visualização das redes de interação fazendo o uso do aplicativo Cytoscape 3.7, foram adotados o ModuLand 2.0 para clusterizar a rede, e do módulo BINGO 3.0.3 para realizar as análises do enriquecimento hipergeométrico, utilizando o Teste de correção de FDR, conforme apresentado no Tópico 3.5. O arquivo (Dados suplementares S3i), que se trata do resumo da análise de ontologia genica (GO) realizada com o uso do aplicativo Goanna AgBase 2.2 conforme descrito no Tópico 3.4, foi ajustado e utilizado para o teste de FDR e de enriquecimento hipergeométrico.

A visualização das redes consenso para o interatoma interespecífico de proteínas geral (Dados suplementares S1e), o interatoma de proteínas de membrana (Dados suplementares S1g) e de proteínas ortólogas (Dados suplementares S1h) permitiu observarmos os aspectos deste (patossistema) entre as proteínas da bactéria da espécie *R. amblyommatis* e do carrapato da espécie *I. scapularis* (Figura 16).

Figura 16 Redes consenso de interações potenciais de alta confiabilidade entre proteínas do interatoma interespecífico de proteínas geral (Dados suplementares S1e) construído para o modelo das espécies *R. amblyommatis* vs *I. scapularis*, com o uso do aplicativo Cytoscape 3.7 adotando do modulo ModLand. Em (A), todas as interações possíveis. Em (B), destaque para as proteínas de *R. amblyommatis* (vermelho), interagindo com as proteínas de *I. scapularis* (azul). Em (C), destaque para os grupos de proteínas mais bem conectados entre si (módulos), totalizando 19 grupos significativos e distintos (conforme as cores).

Redes consenso de interações potenciais de alta confiabilidade entre as proteínas do interatoma de proteínas de membrana construído (Dados suplementares S1g) para o modelo das espécies *R. amblyommatis* vs *I. scapularis*, com o uso do aplicativo Cytoscape 3.7 adotando do modulo ModLand. Em (D), todas as interações possíveis. Em (E), destaque para as proteínas de *R. amblyommatis* (vermelho), interagindo com as proteínas de *I. scapularis* (azul). Em (F), destaque para os grupos de proteínas mais bem conectados entre si (módulos), totalizando 5 grupos funcionais significativos e distintos (conforme as cores).

Redes consenso de interações potenciais de alta confiabilidade entre as proteínas do interatoma de proteínas de membrana e ortólogas para espécie de *Rickettsia* que infectam os carrapatos (Dados suplementares S1h) construído para o modelo das espécies *R. amblyommatis* vs *I. scapularis*. Em (G), todas as interações possíveis (azul). Em (H), destaque para as proteínas de *R. amblyommatis* (vermelho) interagindo com as proteínas de *I. scapularis* (azul). Em (I), destaque para os grupos de proteínas modularizadas, totalizando 2 grupos funcionais distintos com apenas um significativo (verde).



Fonte: Autor deste trabalho.

Na Figura 16, verifica-se que a maior proporção de proteínas da interface *R. amblyommatis* que interagem com *I. scapularis* pode ser associada aos seguintes fatores: (i) uma maior quantidade de proteínas codificadas nos seus genomas, (ii) o alto número de isoformas para uma mesma proteína que são representadas como vértices diferentes na rede (iii) e, ou, enriquecimento de interações entre genes de ambos os organismos presente nos bancos de dados utilizados.

As análises de anotação das proteínas de interface permitiram identificar famílias de proteínas sabidamente envolvidas na modulação da fisiologia do *I. scapularis* para os três domínios: Componente celular (C), refere-se às partes de uma célula ou do ambiente extracelular; Função molecular (F): refere-se as atividades elementares de um gene ao nível molecular, tais como atividade de ligação ou atividade catalítica; Processo biológico (P), refere-se ao conjunto de eventos moleculares pertinentes para o funcionamento das unidades de estar integradas: células, tecidos, órgãos e organismos (Tabela 28).

Tabela 28 Módulos funcionais relacionados as interações entre proteínas provenientes dos genomas de *R. amblyommatis* vs *I. scapularis*, de membrana para esta bactéria contendo sequências ortólogas para *Rickettsia* spp. que infectam carrapatos sustentadas nos bancos do interatoma obtidos com o uso do aplicativo Cytoscape 3.7 adotando o modulo BINGO 3.0.3.

Interatomas	Nº de Módulos significativos	Módulo funcional	P valor	Domínios de produtos gênicos
Proteínas do genoma total	19	4 (5)	1,3812E-3	Regulação através de metilação do DNA
		(6) (7)	1,3812E-3	Biossíntese de proteínas que dependem de cofatores íons metálicos
		(22)	1,3812E-3	Regulação da pressão sanguínea (vasodilatadores)
		16	4,1436E-3	Ativação de chaperonas e dobragem de proteínas
		24	1,1462E-5	Processo metabólico de pirimidinas
		25	1,3812E-3	Processo metabólico de oxaloacetato
		26	2,5217E-4	Ciclo do ácido cítrico
		45	2,7624E-3	Processo metabólico de citrate
		47	1,3812E-3	Processo de biossíntese de pirimidina desoxiribonucleotídeo
		54	2,0673E-7	Processo metabólico de proteínas
		55	1,4152E-5	Processo de biossíntese de RNA (transcrição)
		56	1,1077E-14	Fosforilação de amino ácidos
		58	4,5292E-6	Tradução e alongação de proteínas
		61	3,1917E-4	Resposta de defesa por vírus
		62	1,8112E-5	Transporte mediado por GTPase sinal
		64	1,5384E-5	Processo de biossíntese de nucleotídeo
		65	3,6851E-5	Secreção de proteínas
66	1,3812E-3	Controle cerebral de migração celular		
68	1,1462E-5	Regulação negativa da divisão celular		
Proteínas de Membrana	5	0	2,9350E-5	Secreção celular
		4	1,3812E-3	Regulação de metilação do DNA
		13	3,8268E-7	Processo catabólico de proteínas (proteólises)
		14	3,1158E-7	Oxidação reductiva
		15	2,5491E-16	Fosforilação de amino ácidos
Proteínas de membrana ortólogas	1	1	7,8875E-10	Transporte mediado por GTPase sinal

Na Tabela 28, verifica-se os resultados das redes consenso de interação entre as proteínas da bactéria da espécie *R. amblyommatis* com as sequências de proteínas do carrapato da espécie *I. scapularis*, obtidos com o uso do pacote BINGO 3.0.3 sugerindo a ocorrência de diferentes módulos funcionais significativos.

Dentre estas vias moleculares preditas para os módulos significativos entre as proteínas de membrana da bactéria da espécie *R. amblyommatis* versus as proteínas do carrapato da espécie *I. scapularis*, foram obtidas proteínas que compõem os módulos de transporte, secreção, resposta contra vírus, fosforilação, metilação, oxidação e proteólise (Tabela 29).

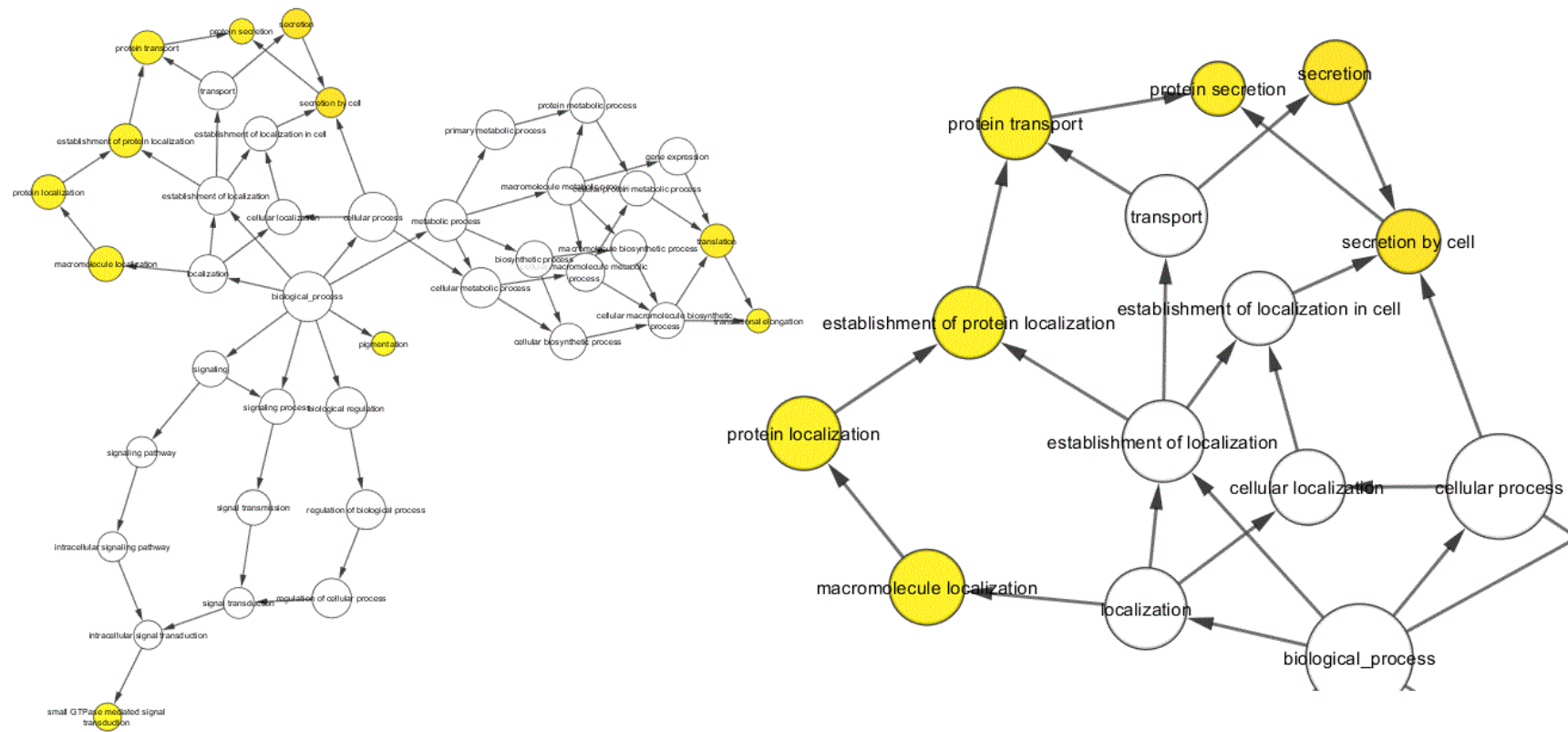
Tabela 29 Resumo da análise de ontologia gênica obtidos com o pacote BINGO 3.0.3 realizada para as proteínas do carrapato que compuseram os módulos funcionais de secreção e transporte obtidos com o módulo Modland e visualizado no aplicativo Cytoscape 3.7.

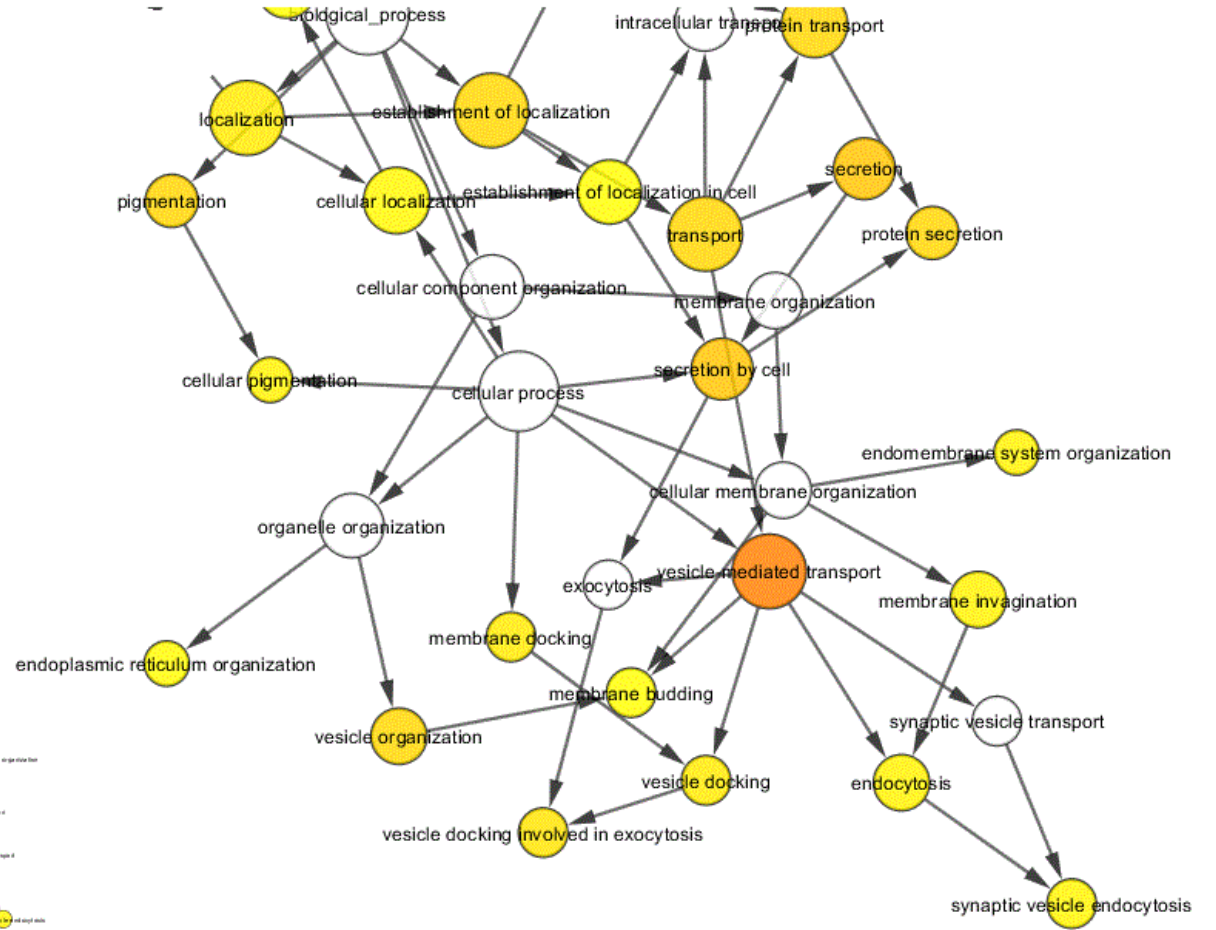
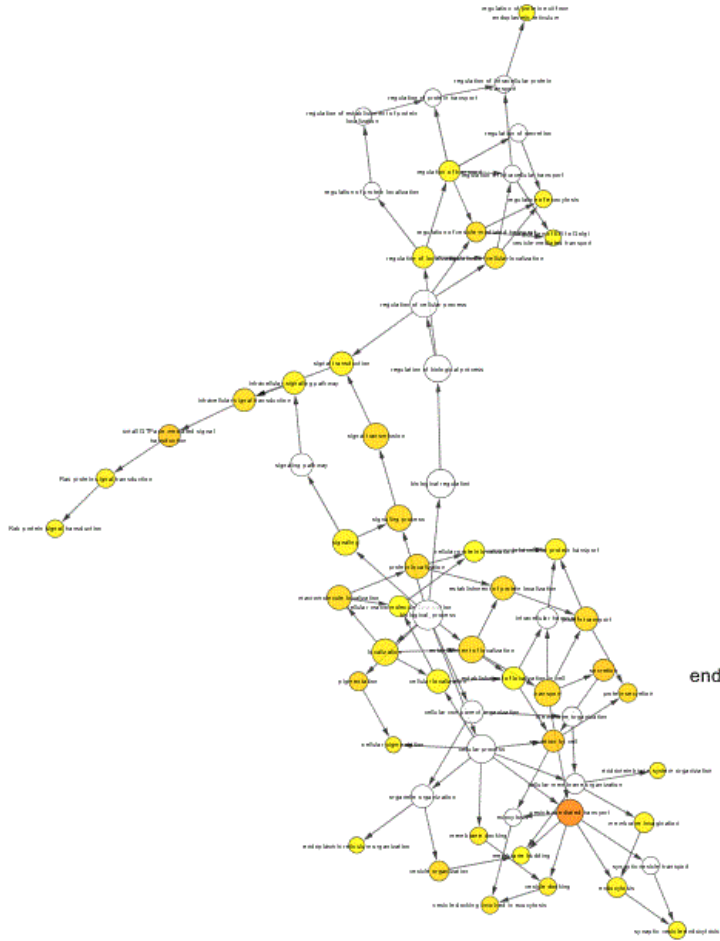
Modulo	GO-ID	p-value	corr p-value	X	n	X	N	Description	Genes in test set
Transporte	16192	7.8875E-10	5.9866E-7	13	80	18	724	vesicle-mediated transport	ISCW004922-PA.1, ISCW000919-PA.1, ISCW003614-PA.1, ISCW005274-PA.1, ISCW001188-PA.1, ISCW001856-PA.1, ISCW007610-PA.1, ISCW001857-PA.1, ISCW002107-PA.1, ISCW003929-PA.1, ISCW006433-PA.1
	7264	8.0387E-7	2.6709E-4	7	26	18	724	small GTPase mediated signal transduction	ISCW007159-PA.1, ISCW007610-PA.1, ISCW000919-PA.1, ISCW003614-PA.1, ISCW002107-PA.1, ISCW001188-PA.1, ISCW009274-PA.1
Secreção	32940	1.4076E-6	2.6709E-4	7	28	18	724	secretion by cell	ISCW007610-PA.1, ISCW004922-PA.1, ISCW001857-PA.1, ISCW000919-PA.1, ISCW003929-PA.1, ISCW006433-PA.1, ISCW001856-PA.1
Metilação	44030	1.3812E-3	2.9098E-2	1	1	1	724	regulation of DNA methylation	ISCW001109-PA.1
	51567	2.7624E-3	2.9098E-2	1	2	1	724	histone H3-K9 methylation	ISCW001109-PA.1
Fosforilação	6468	2.5491E-16	1.8864E-13	16	53	20	724	protein amino acid phosphorylation	ISCW007259-PA.1 ISCW002105-PA.1 ISCW004102-PA.1 ISCW002700-PA.1 ISCW005428-PA.1 ISCW000295-PA.1 ISCW000603-PA.1 ISCW009144-PA.1 ISCW001031-PA.1 ISCW005620-PA.1 ISCW000964-PA.1 ISCW008769-PA.1 ISCW001795-PA.1 ISCW007936-PA.1 ISCW008605-PA.1 ISCW003529-PA.1
	16310	6.3460E-14	1.1740E-11	16	72	20	724	phosphorylation	ISCW007259-PA.1 ISCW002105-PA.1 ISCW004102-PA.1 ISCW002700-PA.1 ISCW005428-PA.1 ISCW000295-PA.1 ISCW000603-PA.1 ISCW009144-PA.1 ISCW001031-PA.1 ISCW005620-PA.1 ISCW000964-PA.1 ISCW008769-PA.1 ISCW001795-PA.1 ISCW007936-PA.1 ISCW008605-PA.1 ISCW003529-PA.1

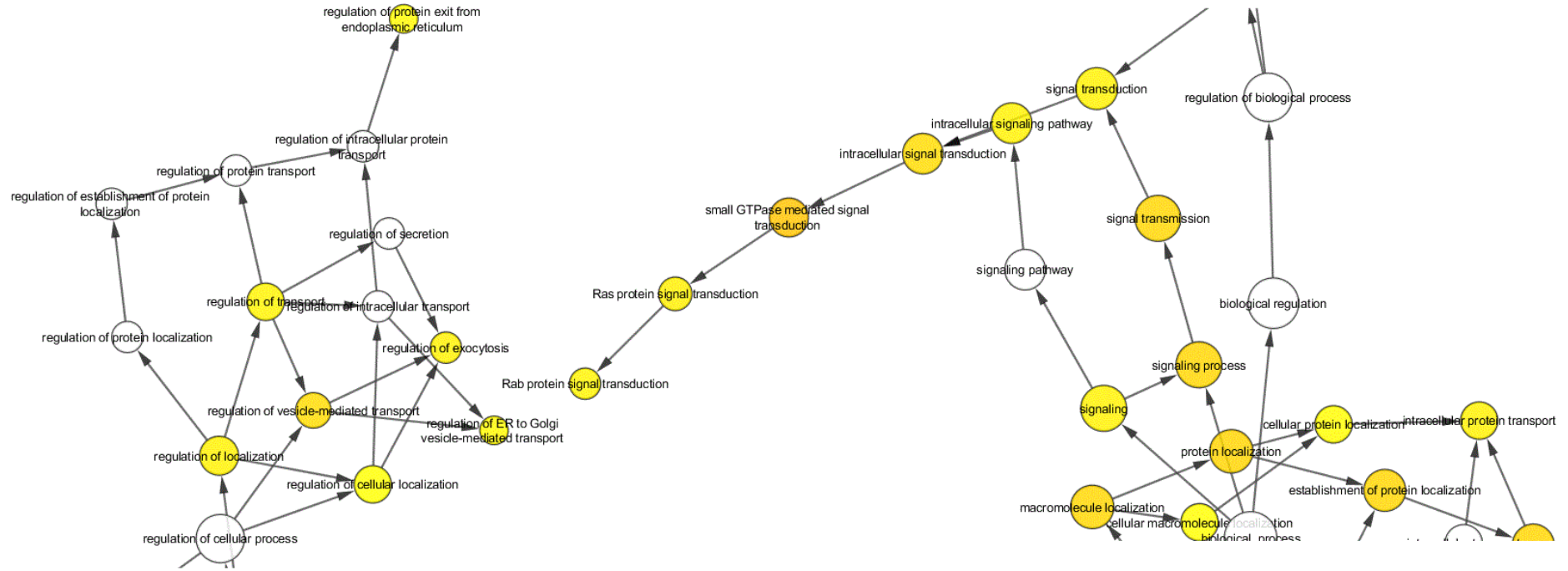
Na Tabela 29, verifica-se um resumo dos resultados de ontologia gênica obtidos por meio do uso e anotação da rede consenso de interação para as proteínas de membrana da espécie *R. amblyommatis* que interagem com as sequências de proteínas do carrapato *I. scapularis*, obtidos com o uso do pacote BINGO 3.0.3. Estes resultados sugerem a presença de diferentes módulos funcionais significativos com destaque para os módulos 0, 4, 13, 14 e 15 obtidos por meio do interatoma de proteínas de membrana ([Dados suplementares S1j](#)), podendo ser visualizados e representados conforme a participação destes, para comporem os componentes da célula, os processos celulares e ou a localização subcelular sendo ativados pela interação (Figura 17).

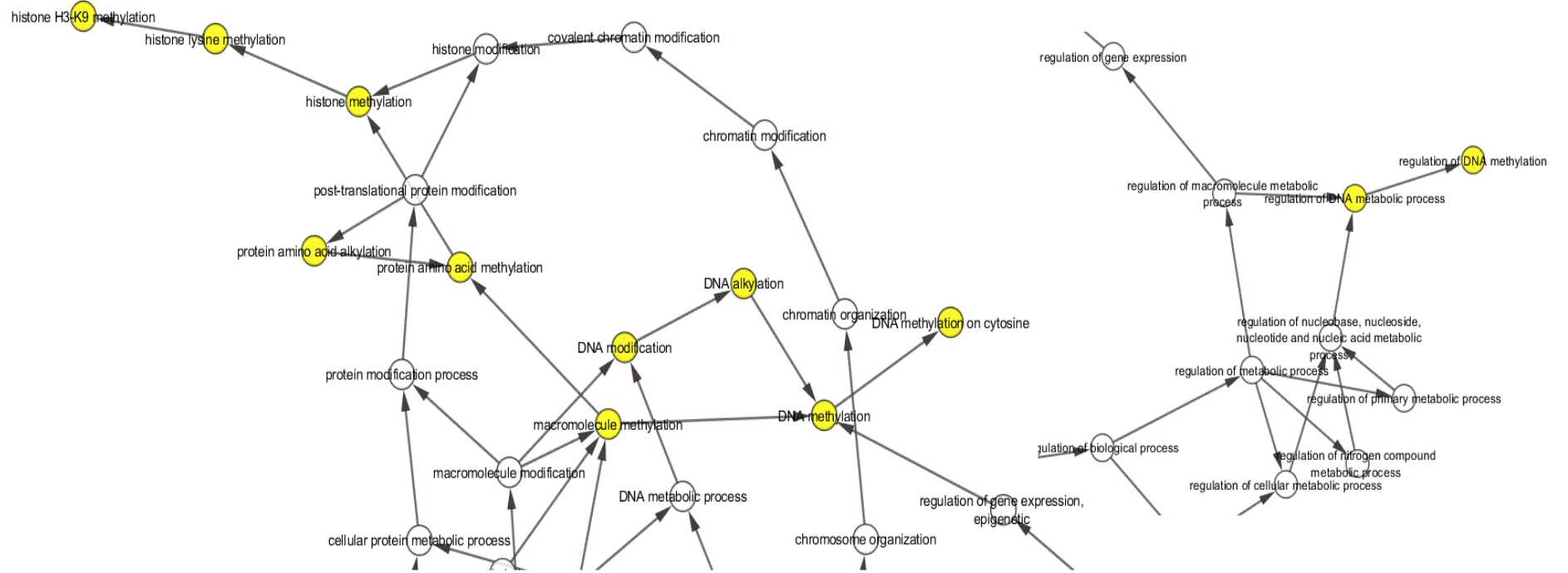
Figura 17 Vias de sinalização envolvendo os módulos de secreção, transporte, metilação e fosforilação obtidos por meio da análise de visualização das redes de interação do interatoma de proteínas de membrana de *R. amblyommatidis* vs *I. scapularis* fazendo o uso do aplicativo Cytoscape 3.3.4, com a adoção dos pacotes ModuLand 2.0 para clusterizar a rede, e do módulo BINGO 3.0.3 para realizar as análises do enriquecimento hipergeométrico, utilizando o Teste de correção de FDR. Em círculos de cor laranja e amarelo, tem-se: Destaque (a) para os processos celulares envolvidos com a secreção de proteínas, localização e transporte com $p\text{-valor} = 2,9350E-5$. Destaque (b) para os processos celulares envolvendo o transporte mediado por vesículas com $p\text{-valor} = 7,8875E-10$. Destaque (c) para os processos celulares envolvidos com a metilação do DNA com $p\text{-valor} = 1.3812E-3$. Destaque (d) para os processos celulares envolvidos com a fosforilação de proteínas, com $p\text{-valor} = 2.5491E-16$

Módulo de secreção e transporte;

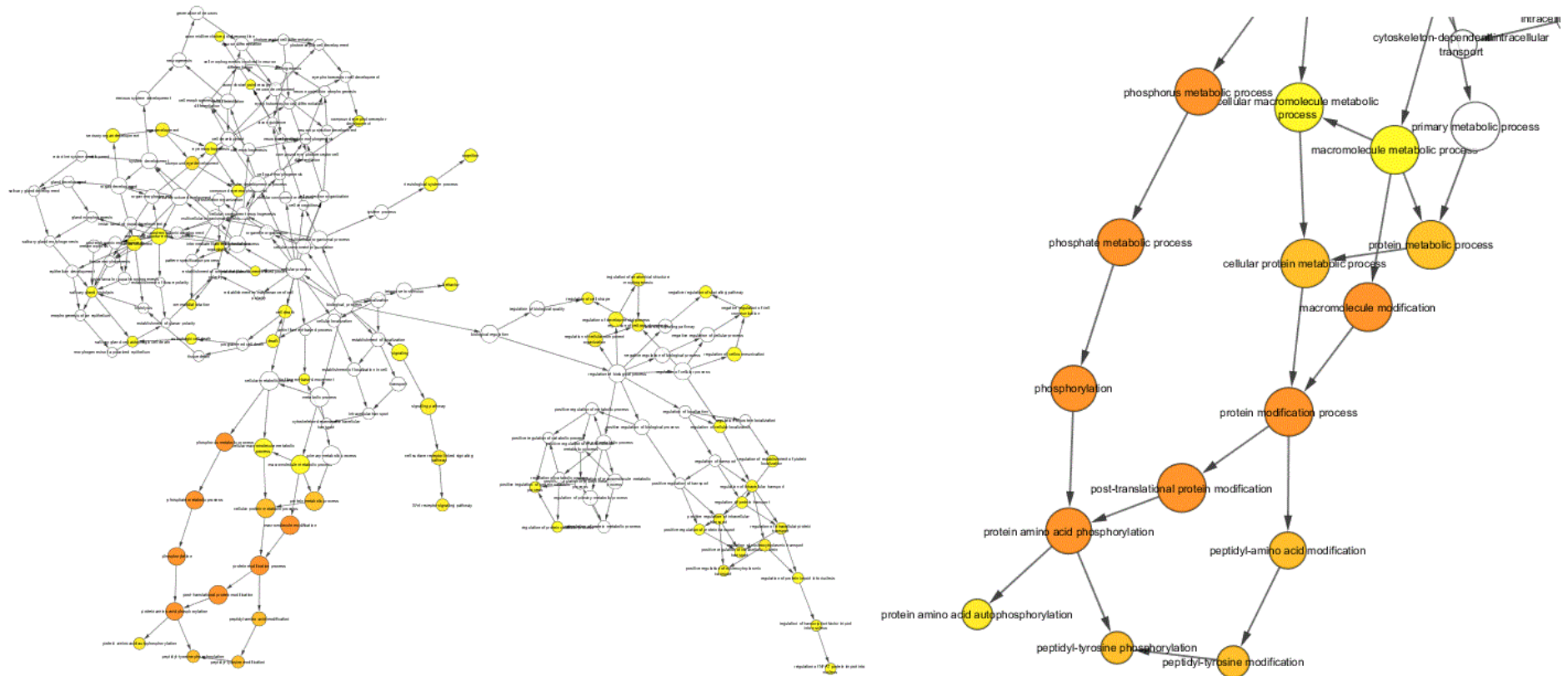








Módulo de fosforilação;



Fonte: Autor deste trabalho.

Na Figura 17, verifica-se que os resultados sugeriram haver distintos módulos funcionais significativos que participam de vias moleculares descritas para as principais proteínas do carrapato presentes nestes seis módulos significativos. No arquivo dos (Dados suplementares S1j), o conjunto de proteínas nos módulos funcionais significativos compuseram no total, pelo menos 45 proteínas da bactéria interagindo com 385 proteínas do carrapato formando 2,275 pares de interação para o módulo de secreção celular. Este foi o maior módulo encontrado, aonde foram observados a interação de proteínas de membrana de *Rickettsia* com diversos domínios de produtos de gênicos do carrapato.

Dentre as vias moleculares preditas obtidas para as proteínas do carrapato que compuseram estes módulos, obteve-se, algumas proteínas relacionadas com a secreção célula, com a interação de proteínas contendo motivos repetição de tetratricopeptídeo Tpr e do T4SS, que diz respeito, aos processos de translação e alongamento na tradução de proteínas, transporte de proteínas, regulação, formação e organização de vesículas mediando o transporte e a exocitose de proteínas e outras moléculas; regulação da localização, estabilização e organização de proteínas endomembranas em função de pequenas GTPase, como Rab e Ras podendo estas, sinalizarem a tradução mediando eventos de invaginação com a formação de sinapse entre as vesículas endocitárias levando a fusão das membranas.

Para o módulo de metilação foram observadas duas proteínas da bactéria contra 21 do carrapato formando 20 pares de interação. Dentre as vias moleculares obtidas para as proteínas do carrapato que compuseram este módulo, obteve-se, aqueles relacionados com a regulação de DNA, metilação de histonas H3-K9, alquilação de DNA e aminoácidos de proteínas, além do desenvolvimento do sistema digestivo e pancreático atuando na morfogênese de cutículas e retinas neurais.

Para o módulo de fosforilação, foram observados a ocorrência de quatro proteínas da bactéria contra 172 do carrapato formando 337 pares de interação. Dentre as vias moleculares obtidas para as proteínas do carrapato que compuseram este módulo, obteve-se, aqueles relacionados com o metabolismo de proteínas e aminoácidos por meio da fosforilação, com a interação de proteínas contendo motivos repetidos de Anquirina (Ank) e proteínas fosfatases, que diz respeito ao processo de modificações pós-translacionais de proteínas, regulação positiva do nucleoplasma e transporte nuclear, regulação de fatores de transcrição, morfogêneses, processos intermediários da organização de citoesqueleto e de morte celular.

Para o módulo de proteólise foram observadas uma única proteína da bactéria contra 34 do carrapato formando 34 pares de interação. Dentre as vias moleculares obtidas para as proteínas do carrapato que compuseram este módulo, obteve-se aqueles relacionados com os processos celulares de modificação e catabolismos de proteínas dependente de Ubiquitina.

Para o módulo de oxidação foram observadas 16 proteínas da bactéria contra 112 do carrapato formando 302 pares de interação. Dentre as vias moleculares obtidas para as proteínas do carrapato que compuseram este módulo, obteve-se, aqueles relacionados com os processos de respiração celular, produção de ATP e energia, atuando na cadeia de transportes de elétrons, com oxidação redutiva e fosforilação oxidativa, homeostase com processos metabólicos oxidativos, além de respostas de defesa contra vírus dentre outras vias moleculares.

As proteínas que compuseram os módulos significativos obtidos com o uso do pacote Bingo 3.0.3 (Dados suplementares S1j), foram selecionadas para a integração dos transcritos obtidos com o uso da montagem “*de novo*”, no Tópico 3.6. Tratando-se, das sequências de proteínas do carrapato *I. scapularis*, com as quais as proteínas da bactéria da espécie *R. amblyommatis* poderia interagir ativando módulos funcionais e vias de proteínas presentes nos interatoma de proteínas de membrana e ortólogas foram as bases para a análise final dos níveis de expressão e regulação dos transcritos dos dados de carrapatos *Amblyomma* spp. infectados e, ou, não por *Rickettsia* spp.

4.8 Montagem dos transcriptomas de carrapatos *Amblyomma* spp.

Neste estudo, foi adotado a abordagem da montagem *de novo* para os dados do transcriptomas provenientes de RNA-seq de espécies de carrapatos *Amblyomma* spp. infectados e não infectados por bactérias de espécies de *Rickettsia*, conforme descrito no Tópico 3.6. Inicialmente, obtiveram-se as sequências gênicas de alta qualidade e confiabilidade por meio da análise de qualidade dos sequenciamentos (Dados suplementares S1k-S4k) com o uso do aplicativo FastQC 0.11.5, conforme o Tópico 3.6.1. Em seguida foi realizado a análise de trimagem com a retirada de possíveis adaptadores das *reads* fazendo o uso dos aplicativos Trimmomatic 0.36 e novamente do aplicativo FastQC 0.11.5 para averiguarmos a qualidade do sequenciamento com os dados trimados (Dados não publicados), conforme apresentando no Tópico 3.6.2, (Dados suplementares S11-S41).

Para a montagem “*de novo*” foi utilizado os arquivos das *reads* trimadas (Dados não publicados), obtidas por meio dos aplicativos Trimmomatic 0.36, associados aos arquivos dos (Dados suplementares S1m-S4m) obtidos com o uso do aplicativo FastQC 0.11.5. Com o uso do aplicativo Trinity 2.5.1, obtivemos as sequências de RNA dos transcritos (Dados não publicados), conforme apresentado no Tópico 3.6.3. Com o uso do aplicativo Kallisto 0.43.0, os transcritos montados foram alinhados para obtermos os níveis de transcrição e abundância, destes, conforme apresentado no Tópico 3.6.4, contento os valores de TPM e os níveis de abundância dos transcritos para cada conjunto de dados de transcriptomas montados (Dados suplementares S1n-S4n).

4.8.1 Qualidade e trimagem das *reads* dos transcriptomas

4.8.1.1 Qualidade do sequenciamento

Com o uso do aplicativo Fastqc 0.11.5., foi obtido os resultados das análises de qualidade das *reads* de diferentes órgãos de carrapatos *Amblyomma* spp. infectados e, ou, não infectados por *Rickettsia* spp.. Os arquivos de saída contêm as *reads* (Dados não publicados) e os *links* dos parâmetros avaliados. Nos resultados da análise de qualidade inicial, observou-se a ocorrência de diferentes *reads* para os dados de RNA-seq *single end* (Se), havendo nas glândulas salivares do carrapato *A. aureolatum* 50,475,296 *reads* no arquivo (SRR4301110_1_fastqc) ([Dados suplementares S1k](#)) quando infectado e 59,846,982 *reads* no arquivo (SRR4301100_1_fastqc) ([Dados suplementares S1k'](#)), quando não infectado pela espécie *R. rickettsii*.

No intestino do carrapato desta mesma espécie, observou-se a ocorrência de 45,877,417 *reads* no arquivo (SRR4301120_1_fastqc) ([Dados suplementares S2k](#)) quando infectado e 86,468,328 *reads* no arquivo (SRR4301108_1_fastqc) ([Dados suplementares S2k'](#)) quando não infectado por esta mesma espécie de bactéria. No intestino do carrapato *A. sculptum*, observou-se a ocorrência de 42,180,348 *reads* no arquivo (SRR4277085_1_fastqc) ([Dados suplementares S3k](#)) quando infectado, 86,829,754 *reads* no arquivo (SRR4277086_1_fastqc) ([Dados suplementares S3k'](#)) e 62,613,298 no arquivo (SRR4277087_1_fastqc) ([Dados suplementares S3k''](#)) quando não infectado por esta mesma espécie de bactéria.

Nos resultados da análise de qualidade inicial, observou-se a ocorrência de diferentes *reads* para os dados de RNA-seq *pair end* (Pe), havendo no intestino do carrapato *A. sculptum*

24,604,909 *reads* no arquivo (SRR3198618_1_fastqc) ([Dados suplementares S4k](#)) quando infectado, 26,550,466 *reads* no arquivo (SRR3120106_1_fastqc) ([Dados suplementares S4k'](#)) e 26,550,466 *reads* no arquivo (SRR3120106_2_fastqc) ([Dados suplementares S4k''](#)), quando não infectado pela espécie *R. amblyommatis*. Nos ovários do carrapato desta mesma espécie, observou-se a ocorrência de 23,294,597 *reads* no arquivo (SRR3198621_1_fastqc) ([Dados suplementares S5k](#)) 22,726,084 *reads* no arquivo (SRR3126275_1_fastqc) ([Dados suplementares S5k'](#)) e 22,726,084 *reads* no arquivo (SRR3126275_2_fastqc) ([Dados suplementares S5k''](#)).

4.8.1.2 Trimagem das *reads* do sequenciamento

Com o uso do aplicativo Trimmomatic 0.36, os arquivos contendo às *reads* provenientes dos dados de RNA-seq foram trimadas com a remoção dos possíveis adaptadores. Em seguida estas sequências foram submetidas novamente a análise de qualidade através do aplicativo Fastqc 0.11.5, gerando os arquivos de saída contendo as *reads* (Dados não publicados) e os *links* dos parâmetros avaliados. Nos resultados da análise de qualidade inicial, observou-se a ocorrência de diferentes *reads* para os dados de RNA-seq *single end* (Se), havendo nas glândulas salivares do carrapato *A. aureolatum* 49,118,00 *reads* no arquivo (SRR4301110_1.clean15_pair.fastqc) ([Dados suplementares S11](#)) quando infectado e 58,342,330 *reads* no arquivo (SRR4301100_1.clean15_pair.fastqc) ([Dados suplementares S11'](#)) quando não infectado pela espécie *R. rickettsii*.

No intestino do carrapato desta mesma espécie, observou-se a ocorrência de 44,638,813 *reads* no arquivo (SRR4301120_1.clean15_pair.fastqc) ([Dados suplementares S21](#)) quando infectado e 84,201,603 *reads* no arquivo (SRR4301108_1.clean15_pair.fastqc) ([Dados suplementares S21'](#)) quando não infectado por esta mesma espécie de bactéria. No intestino do carrapato *A. sculptum*, observou-se a ocorrência de 41,088,975 *reads* no arquivo (SRR4277085_1.clean15_pair.fastqc) ([Dados suplementares S31](#)) quando infectado, 84,544,156 *reads* no arquivo (SRR4277086_1.clean15_pair.fastqc) ([Dados suplementares S31'](#)) e 60,935,922 *reads* no arquivo (SRR4277087_1.clean15_pair.fastqc) ([Dados suplementares S31''](#)) quando não infectado por esta mesma espécie de bactéria.

Nos resultados da análise de qualidade inicial, observou-se a ocorrência de diferentes *reads* para os dados de RNA-seq *pair end* (Pe), havendo no intestino do carrapato *A. sculptum*

24,408,257 *reads* no arquivo (SRR3198618_1.clean15_pair.fastqc) ([Dados suplementares S41](#)) quando infectado, 25,066,766 *reads* no arquivo (SRR3120106_1.clean15_pair.fastqc) ([Dados suplementares S411](#)) com 994,486 *reads* no arquivo (SRR3120106_1.clean15_Unpair.fastqc) ([Dados suplementares S411'](#)) e 25,066,766 *reads* no arquivo (SRR3120106_2.clean15_pair.fastqc) ([Dados suplementares S412](#)) com 444,417 *reads* no arquivo (SRR3120106_2.clean15_Unpair.fastqc) ([Dados suplementares S412'](#)) quando não infectado pela espécie *R. amblyommatis*.

Nos ovários do carrapato desta mesma espécie, observou-se a ocorrência de 23,151,084 *reads* no arquivo (SRR3198621_1_clean15_pair.fastqc) ([Dados suplementares S51](#)) quando infectado, 21,598,841 *reads* no arquivo (SRR3126275_1.clean15_pair.fastqc) ([Dados suplementares S511](#)) com 727,070 *reads* no arquivo (SRR3126275_1.clean15_Unpair.fastqc) ([Dados suplementares S511'](#)) e 21,598,841 *reads* no arquivo (SRR3126275_2.clean15_pair.fastqc) ([Dados suplementares S512](#)) com 358,442 *reads* no arquivo (SRR3126275_2.clean15_Unpair.fastqc) ([Dados suplementares S512'](#)) quando não infectos pela mesma espécie de bactéria.

Os resultados dos valores da análise de qualidade das *reads* inicial obtidos com o uso do aplicativo Fastqc 0.11.5, seguido da análise de trimagem com o uso do aplicativo Trimmomatic 0.36 e novamente a análise de qualidade sob os dados trimados com o uso do Fastqc, sugeriu a melhoria dos dados com a remoção de adaptadores e de *reads* de baixa qualidade sendo possível comparar as diferenças nos valores do total destas sequências entre o antes e o depois. Estes conjuntos de dados de transcriptomas de *Amblyomma* spp. infectados e, ou, não por *Rickettsia* spp. garantiram a escolha de sequências apresentando os melhores níveis de qualidade para a análise da montagem “*de novo*” (Tabela 30).

Tabela 30 Quadro dos resultados parciais das análises de qualidade e trimagem para o sequenciamento dos dados de RNA-seq de espécies de carrapatos *Amblyomma* spp. infectados e, ou, não por espécies de bactéria *Rickettsia* spp., com o uso dos aplicativos Fastqc e Trimmomatic. Destaque para o número de *reads* e as diferenças entre estes valores em (Dados suplementares (S)), após os procedimentos iniciais de qualidade e o final com os dados trimados para a obtenção de seqüências de alta qualidade.

Espécie de <i>Amblyomma</i>	Espécie de <i>Rickettsia</i>	Órgão	Seq	Total de <i>Reads</i> transcriptomas carrapatos infectados					Total de <i>Reads</i> transcriptomas carrapatos não infectados				
				(Dados S-k) Valores após a análise inicial de qualidade das <i>reads</i> .		(Dados S-l) Valores após a análise de trimagem e da qualidade final das <i>reads</i>		Número de <i>reads</i> removidas	(Dados S-k) Valores após a análise inicial de qualidade das <i>reads</i>		(Dados S-l) Valores após a análise de trimagem e da qualidade final das <i>reads</i>		Número de <i>reads</i> removidas
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Glândulas salivares	<i>Single end</i>	S1k	50,475,296	S1L	49,118,004	1.357.292	S1k'	59,846,982	S1l'	58,342,330	1.504.652
		Intestino	<i>Single end</i>	S2k	45,877,417	S2L	44,638,813	1.238.604	S2k'	86,468,328	S2l'	84,201,603	2.266.725
		Intestino	<i>Single end</i>	S3k	42,180,348	S3L	41,088,975	1.091.373	S3k'	86,829,754	S3l'	84,544,156	2.285.598
<i>A. sculptum</i>	<i>R. amblyommatis</i>	Intestino	<i>Pair end</i>	S4k	24,604,909	S4L	24,408,257	196.652	S4k'	26,550,466	S4l1	25,066,766	1.483.700
									S4k''		S4l2		1.483.700
		Ovários	<i>Pair end</i>	S5k	23,294,597	S5L	23,151,084	143.513	S5k'	22,726,084	S5l1	21,598,841	1.127.243
									S5k''	22,726,084	S5l2	21,598,841	1.127.243

Na Tabela 30, verifica-se o resumo dos resultados preliminares obtidos após a análise de qualidade e de trimagem do sequenciamento com o uso do aplicativo Trimmomatic 0.36., estimando o total de sequências brutas geradas com o uso do aplicativo Fastqc 0.11.5. para os dados de RNA-seq obtidos de diferentes órgãos de espécies de carrapatos *Amblyomma* spp. infectados ou não por espécies de bactérias *Rickettsia* spp., sendo representados os valores dos (Dados suplementares S-k e S-l (s)) sugerindo as diferenças (em cor laranja) para o número de *reads* removidas entre os dados brutos e os dados tratados, podendo ser acessados por meio dos *links* apresentados acima.

4.8.2 Montagem “de novo” dos transcriptomas de *Amblyomma* spp.

Com o uso do aplicativo Trinity 2.5.1, obteve-se nos arquivos de saída arquivos trinity.Fasta contendo as sequências de RNA montadas (Dados não publicados) (Tabela 31).

Tabela 31 Resumo dos resultados preliminares da análise da montagem de novo de dados de RNA-seq com destaque para os arquivos de saída Trinity.Fasta contendo as sequências de RNA montadas.

Sequência de nucleotídeo no formato Fasta contendo o Trinity ID
<pre>>TRINITY_DN8_c0_g1_i1 len=915 path=[895:0-914] [-1, 895, -2] CGCTGCCAAGATACATAGCCACTACTCCGAGTTTTCTCCAAGGTTGCCTCCTACGCGGAAGACGTC AAGCTCCACGTCGC CAACCCGGATGTA CTGAAACAGACCTTCCCTGTCTGGAGAGCTATCTGTCCCTTCTTCGGGAGAGCTACGGATCCACTAT CGAATCAGTAAACTTCAGGAACGATTACGAGAGTGTTAGGCAGCAAGTGAATGCTTGGGTCGAGAAAGCCACCGAATCA AAGATAAAGGACCTCTTGCCCACTGGGAGCGTAGATGCCCTCACCACCCTGATCCTGGTAAAGCTATCTACTTCAAAGGCT CCTGGGAATCACCCTTCGATCCAGACGCCACATGTAGTTCTGATTTCCACCTCGACTCGAAGAACAAGAAAGAAGTGGAC ATGATGTACAAGAAGAAAGCTTACAAGATGTGCACCAACGACGAAGTGGAGTAGCCGCAGTCGAGATCCCTTACCGCGG TGGCAAGACTTCCATGGTCTGCTCCTTCCCAACGATATCGAAGGGCTGTCCAAACTTGAGGACCTCATGACAGCTACAAA ACTTGC GGACGTGTTGAAAAACCTTTGCACCCGGCTGGATGTTGAGCTTTATCTCCCAAATTC AAGCTAGAACAGGCGAT TTTCTCAAGGAAACCTTGCAAGCTATGGGGGTCAAAGACTTTTTCACTACGGACGCTGACTTATCTGCCATCACTGAGGC GGAAGAGCTGATGGCTTCTGAAGTGATCCACAAAGCCTTTGTAGAAAGTCGATGAGAAGGGCACAGAAGCCGCCGCCGCG ACCGCTGTGATGGTGGCGGCTGCTGCTGACGTCCGGCTATTGTGCCAGTGACTTACAAATTTGTGGTTCGATCGGCCCTTC ATGTTCTTAATTCATAGTCACGATCCAGG</pre>
<pre>>TRINITY_DN8_c1_g1_i1 len=243 path=[223:0-242] [-1, 223, -2] TTTTTGGGTTCAAGCTTTATCTAGTGCACCAAATACTGTTACATAGTACAATGAAATGGCACATTACCAGACGTTGAGAT GGAAAGCACCCACACTGCGAGAGGCATGCTCTTCAGGGCATGAACAGAGCTTTGGAGAAGTTCACAATTCACGAACGGA GCCCATGAAAAGCACGACTTCGGGATCATGACTACGAATTAGGAACAAGAAAGGCCGATCCACCACAAATTTGTAAGTCA CTG</pre>
<pre>>TRINITY_DN8_c2_g1_i1 len=412 path=[1:0-411] [-1, 1, -2] AGTACATCCGGTTGGCGACGTGCAGTATGACGTACAGGAGAGTAGCCCGGAAGCTGTTCTTTGAGGAACTTGTAAGTGT CGATGGTGGGACTGTTCTTCCCTCACGTGGAGGGCCTCGCACAGTTGCTTGGCTGTTTGATTTCCGGGCTCCGGCCAGTACC ATGGAGAGAGCCCGTAAAATGCTGAACCGGTGAGAAGAAGATATTCTCATGTTTCTCGACTGCATTAGTTCGACAGATGCTT GTGCAGGTCCACGGAGAATGCGATGAGTGAAGTGCCGAAGTCCGCTCCATGGTATGGTACGTCTGAAAAAGCTAGGTCT CCAGTAGCGACGTAAACAGGTTACGCCCCACGATCGGGCGCAGTCAAAACTATGAAAGTTGTTTTCTTGGCAAGGCCG TCGCATGCGCT</pre>

Na Tabela 31, verifica-se um resumo das sequências de RNA para os transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por espécies de *Rickettsia*, em formato Fasta provenientes da análise da montagem “de novo” com uso do aplicativo Trinity 2.5.1 (Dados não publicados).

4.8.3 Análise dos níveis de abundância e indexação da transcrição

Com o uso do aplicativo Kallisto 0.43.0, obteve-se o arquivo de saída transcriptoma.idx para cada conjunto de dados (Dados não publicados). Estes arquivos foram indexados ao arquivo contendo as sequências de RNA em formato Fasta obtidas com uso do Trinity 2.5.1. Foram originados três arquivos de saída, entre estes, os dados do nível de abundância e TPM para os transcritos montados (Tabela 32).

Tabela 32 Resumo dos resultados preliminares da análise do cálculo de abundância e TPM com a indexação do arquivo Trinity.Fasta obtido com uso do aplicativo Trinity 2.5.1 contendo as sequências de RNA ao arquivo transcriptoma.idx obtido com uso no aplicativo Kallisto 0.43.0.

Target_ID	Length	Eff_length	Est_counts	TPM
TRINITY_DN8_c0_g1_i1	946	747	252	5.98088
TRINITY_DN78_c0_g1_i1	7221	7022	8364	21.1173
TRINITY_DN51_c0_g1_i1	3324	3125	63223	358.682
TRINITY_DN53_c0_g1_i1	1412	1213	1194	17.4513

Na Tabela 32, verifica-se um resumo do cálculo dos níveis de abundância da transcrição contendo, os IDs das sequências de RNA, que foram obtidas com uso com aplicativo Trinity 2.5.1 e os valores de TPM montados. Os resultados desta análise, sugeriu a ocorrência de diferentes sequências de RNA para os dados de transcriptoma montados de RNA-seq *single end* (Se), havendo nas glândulas salivares do carrapato *A. aureolatum* a ocorrência de 76,736 sequências de transcritos no arquivo (SRR4301110_abundance.tsv) ([Dados suplementares S1m](#)) quando infectado e 74,434 sequências de transcritos no arquivo (SRR4301100_abundance.tsv) ([Dados suplementares S1m'](#)) quando não infectado pela espécie *R. rickettsii*.

No intestino do carrapato desta mesma espécie, observou-se a ocorrência de 61,278 sequências de transcritos no arquivo (SRR4301120_abundance.tsv) ([Dados suplementares S2m](#)) quando infectado e 73,994 sequências de transcritos no arquivo (SRR4301108_abundance.tsv) ([Dados suplementares S2m'](#)) quando não infectado por esta mesma espécie de bactéria. No intestino do carrapato *A. sculptum*, observou-se a ocorrência de 65,277 sequências de transcritos no arquivo (SRR4277085_abundance.tsv) ([Dados suplementares S3m](#)) quando infectado, 67,208 sequências de transcritos no arquivo (SRR4277086_abundance.tsv) ([Dados suplementares S3m'](#)) e 54,687 sequências de transcritos no arquivo (SRR4277087_abundance.tsv) ([Dados suplementares S3m''](#)) quando não infectado por esta mesma espécie de bactéria.

Nos resultados da análise de qualidade inicial, observou-se a ocorrência de diferentes *reads* para os dados de RNA-seq *single end e pair end* (Pe), havendo no intestino do carrapato *A. sculptum* a ocorrência de 89,641 sequências de transcritos no arquivo (SRR3198618_abundance.tsv) ([Dados suplementares S4m](#)) quando infectado e 40,996 sequências de transcritos nos arquivos (SRR3120106_abundance.tsv) ([Dados suplementares S4m'](#)) quando não infectado pela espécie *R. amblyommatis*. Nos ovários do carrapato desta mesma espécie, observou-se a ocorrência de 93,260 sequências de transcritos no arquivo (SRR3198621_abundance.tsv) ([Dados suplementares S5m](#)) e 54,421 sequências de transcritos no arquivo (SRR3126275_abundance.tsv) ([Dados suplementares S5m''](#)).

4.9 Análises de expressão diferencial

As sequências de RNA em formato Fasta obtidas à partir da montagem “*de novo*” com o uso do aplicativo Trinity (Dados não publicados) provenientes dos dados dos transcriptomas de espécies de carrapatos do gênero *Amblyomma* infectados e, ou, não infectados por espécies de bactéria do gênero *Rickettsia*, juntamente com os IDs dos transcritos e os valores da TPM obtidos com o uso do aplicativo Kallisto (Dados suplementares S1m-S5m), foram utilizados para análise de expressão diferencial.

Para isto, foi realizada a integração das sequências de nucleotídeos dos transcritos aos interatomas construídos, adotando das sequências de aminoácidos das proteínas de carrapato da espécie *I. scapularis*, especialmente aquelas, com as quais, as proteínas de membrana e ortologas da bactéria da espécie *R. amblyommatis* formaram pares de interação, estando ainda presentes nos módulos significativos (Dados suplementares S1j).

Primeiramente foi realizado a análise de similaridade entre as sequências de nucleotídeos dos transcritos obtidos contra as sequências de proteínas do carrapato envolvidas nas vias preditas para a secreção, transporte, fosforilação, metilação, defesa contra vírus, oxidação e proteólise, presentes no interatoma de proteínas de membrana fazendo uso de multialinhamentos de sequências.

4.9.1 Similaridade entre as sequências de transcritos de *Amblyomma* versus as proteínas do *I. scapularis* presentes no interatoma de membrana

Para integrarmos as sequências de RNA montados (Dados não publicados) ao interatoma, utilizamos as sequências de aminoácidos das proteínas de *I. scapularis* presentes no interatoma das proteínas de membrana e ortólogas de *R. amblyommatis* (Dados suplementares S1j). Entre, as proteínas escolhidas, adotamos aquelas que fizeram parte de módulos funcionais significativos, extraíndo as proteínas do carrapato presentes nos cluster de secreção, transporte, fosforilação, metilação, defesa contra vírus, oxidação e proteólise, obtidos com o aplicativo Cytoscape 3.7 adotando o modulo BINGO 3.0.3.

À partir disto, a integração dos transcritos ao transcriptoma, pode ser realizado fazendo o uso do TBLASTn, para a análise de similaridade destas sequências de proteínas do *I. scapularis* (Dados suplementares S1j) contra as sequências de RNA (Dados não publicados) obtidas para os transcriptomas de espécies de carrapatos *Amblyomma* spp., com o uso do aplicativo Trinity 2.5.1. Os resultados observados revelaram os pares de IDs das sequências de RNA similares entre as sequências de aminoácidos de proteínas do carrapato que constituíram os módulos funcionais do interatoma de proteínas de membrana predito (Tabela 33).

Tabela 33 Resumo dos resultados da análise de similaridade para as sequências de proteínas do *I. scapularis* obtidas dos interatomos contra as sequências de RNA referentes aos transcriptomas montados para órgãos de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia* spp..

Modulos	ID proteína <i>I. scapularis</i>	ID-trinity-RNA-seq <i>Amblyomma</i> spp.	Confiança \geq 70	Identidade \leq 1e-10
	ISCW000002-PA.1	TRINITY_DN1034_c0_g1_i1	79.85	3e-70
	ISCW000002-PA.1	TRINITY_DN30393_c0_g1_i1	79.85	2e-70
	ISCW000086-PA.1	TRINITY_DN16087_c0_g1_i1	70.76	0.0
	ISCW000086-PA.1	TRINITY_DN43626_c0_g1_i1	70.76	0.0
	ISCW000128-PA.1	TRINITY_DN1752_c0_g1_i1	79.25	0.0

Na Tabela 33, verifica-se um resumo do TBLASTn para as sequências de proteínas do carrapato presentes nos módulos funcionais do interatoma de proteínas de membrana contra as sequências de transcritos de RNA dos transcriptomas montados para carrapatos *Amblyomma* spp. infectados e não infectados por *Rickettsia* spp. em diferentes órgãos ([Dados suplementares S1n](#)), podendo ser vistos neste [link](#).

4.9.2 Níveis de expressão dos transcritos

Os resultados da análise de similaridade (Dados suplementares S1n) contém os IDs das sequências de proteínas do carrapato da espécie *I. scapularis* que compuseram os seis principais módulos significativos do interatoma de proteínas de membrana predito e os seus similares, obtidos para cada conjunto de dados de transcritos de RNA dos transcriptomas montados para os carrapatos *Amblyomma* spp. infectados e, ou, não infectados por *Rickettsia* spp. em diferentes órgãos. Com os transcriptomas integrados ao interatoma de proteínas de membrana e ortólogos, obteve-se os resultados de TPM pareados para cada transcrito similar as proteínas do *I. scapularis*.

Para a integração final dos dados, cada par de ID das proteínas do *I. scapularis*, que possuem ID similar as sequências de transcritos montadas, foi concatenado gerando uma lista de coluna única com apenas uma linha para cada identificador de proteína. A integração dos dados do transcriptoma as proteínas presentes nos módulos funcionais do interatoma de membrana, gerou um arquivo contendo os IDs para cada uma das sequências de proteínas do carrapato *I. scapularis* com seus pares de IDs das sequências de RNA similares e os valores de TPM dos transcritos.

As sequências destes transcritos foram similares as sequências de proteínas do carrapato avaliadas por meio do TBLASTn., também interagiram com as proteínas de membrana e ortólogas de *R. amblyommatis* obtidas para seis diferentes módulos funcionais significativos oriundos dos interatomas de proteínas de membrana (Dados não publicados). Após a obtenção destes arquivos contendo os IDs das proteínas do *I. scapularis*, os IDs das sequências de RNA providas dos transcriptomas de *Amblyomma* spp. obtidas com o uso do Trinity 2.0, contendo os valores de TPM para cada transcrito, foram calculados os valores de *fold change* com o uso de gráficos construídos no aplicativo do Excel.

Neste sentido, foi realizado a divisão dos valores de TPM para os dados de transcritos de carrapatos *Amblyomma* spp. em condições de infecção sob os dados de transcritos em condições de não infecção por bactérias *Rickettsia* spp. Os valores de *fold change* foram obtidos pela divisão simples com o uso da fórmula matemática:

$$Fold\ change = \frac{TPM_{inf}}{TPM_{\bar{inf}}}$$

Após logaritmizar os valores de *fold change* na base Log2, foi realizado o cálculo da média para os valores quando mais de uma sequência de transcrito foram similares e obtidas para uma mesma sequência de proteínas do carrapato *I. scapularis*. Logo em seguida, averiguamos se os transcritos são regulados de forma positiva “up” quando os valores de *fold change* $\log_2 \geq 1$, e se regulados de forma negativa “down” quando os valores de *fold change* $\log_2 \leq -1$ (Tabela 34).

Tabela 34 Resumo dos resultados de *fold change* obtidos com a divisão dos valores de TPM inf por TPM ñ inf, destacando os IDs dos transcritos de *Amblyomma* spp. similares às proteínas de *I. scapularis*, sendo regulados de forma positiva, negativa ou sem níveis de expressão detectado para uma definição.

ID proteína <i>I. scapularis</i>	ID-RNA-seq <i>Amblyomma</i> spp.	TPM inf	TPM n.inf	<i>Fold.change</i>	Log2 F.c	Média
ISCW000002-PA.1	DN50051_c0_g1_i1	115,223	1,31054	87,9202466	6,458123528	0,151859
ISCW000002-PA.1	DN55224_c0_g1_i1	1,73696	123,723	0,0140391	-6,15440538	
ISCW000086-PA.1	DN13259_c0_g1_i1	8,38115	7,51545	1,11518938	0,157288723	2,002651
ISCW000086-PA.1	DN65255_c0_g1_i1	21,8704	1,51876	14,4001686	3,848013794	
ISCW000128-PA.1	DN26104_c0_g1_i1	0	7,46028	0	#NÚM!	-

Na Tabela 34, verifica-se um resumo dos IDs de identificação das sequências de RNA transcritas nos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia*, com destaque para os valores de *fold change* e a média calculada para a definirmos, se a expressão destes transcritos é regulada de forma positiva ou negativa, podendo ainda, não haver níveis satisfatórios para esta mensuração ([Dados suplementares S10](#)), podendo ser visto neste [link](#).

Estes resultados sugerem mudanças de expressão gênica para os carrapatos com ou sem a infecção. Os resultados obtidos para a anotação de proteínas preditas baseado na expressão diferencial de genes presentes nas células dos órgãos glândulas salivares, intestino e ovários de espécies de *Amblyomma*, sugeriram haver diferenças para cada conjunto de dados, extraído as sequências com a média do valor de *fold change* $\log_2 \geq 2$ ou ≤ -2 (Tabela 35).

Tabela 35 Quadro de resumo da anotação e expressão diferencial para o conjunto de dados de RNA-seq abordando a montagem de novo para cada espécie de carrapato *Amblyomma* spp. e em determinado órgão infectado ou não por *Rickettsia* spp., sendo (sg) glândulas salivares, (mg) intestino e (ov) ovários.

<i>Amblyomma</i> spp. & <i>Rickettsia</i> spp.	Órgão	Módulo	Nº de Proteínas	Reg (+)	Reg (-)
<i>A. aureolatum</i> & <i>R. rickettsii</i>	(sg)	transporte	36	15	21
		Secreção	131	71	60
		Metilação	3	2	1

		Proteólise	12	7	5
		Oxidação	38	15	23
		Fosforilação	61	29	32
	(mg)	transporte	30	26	4
		Secreção	111	89	22
		Metilação	2	2	0
		Proteólise	15	11	4
		Oxidação	32	25	7
		Fosforilação	58	46	12
<i>A. sculptum</i> & <i>R. rickettsii</i>	(mg)	transporte	30	6	24
		Secreção	116	47	69
		Metilação	0	0	0
		Proteólise	12	5	7
		Oxidação	23	12	11
		Fosforilação	54	35	19
<i>A. sculptum</i> & <i>R. amblyommatis</i>	(mg)	transporte	6	2	4
		Secreção	33	11	22
		Metilação	0	0	0
		Proteólise	5	3	2
		Oxidação	15	3	12
		Fosforilação	9	3	6
	(ov)	transporte	27	11	16
		Secreção	83	36	47
		Metilação	1	1	0
		Proteólise	9	3	6
		Oxidação	23	10	13
		Fosforilação	33	10	23

Na Tabela 35, verificasse todo o conjunto de dados avaliados, sendo anotados o número de proteínas que apresentaram níveis de abundância de transcritos e divididos entre as proteínas que apresentaram regulação positiva com valor de *fold change* na base Log2 $\geq 2,5$ das que apresentaram regulação negativa com valor de *fold change* na base Log2 $\leq -2,5$ para cada módulo funcional significativo preditos no interatoma de proteínas de membrana e ortólogas.

De certa forma, estas diferenças de expressão e na abundância dos transcritos para o conjunto de dados dos transcriptomas das espécies *A. sculptum* em resposta a infecção por *R. amblyommatis* no intestino e ovário, *A. sculptum* em resposta a infecção por *R. rickettsii* no intestino, e *A. aureolatum* em resposta a infecção por *R. rickettsii* em glândulas salivares e intestino são inerentes aos órgãos. Estes resultados revelaram a capacidade de espécies de *Rickettsia* na atividade desta expressão de genes podendo estar correlacionadas com o potencial de proteínas expressas nos seus genomas pertencentes a membrana, sendo estas secretadas, excretadas e, ou, de superfície.

As proteínas com maior nível de expressão considerando o valor de *fold change* na base $\text{Log}_2 \geq 2,5$ para uma expressão positiva e o valor de *fold change* na base $\text{Log}_2 \leq -2,5$ quanto a expressão for negativa. Nas glândulas do carrapato da espécie de carrapato *A. aureolatum* infectado ou não por *R. rickettsii*, os resultados para os níveis de expressão e abundância de transcritos de proteínas preditas no módulo de secreção, destacaram-se 28 proteínas apresentando regulação negativa e 31 apresentando regulação positiva. Para as proteínas preditas no módulo de transporte que interagiu com proteínas ortólogas de *Rickettsia* spp. utilizadas inicialmente no modelo evolutivo, destacaram-se 14 apresentando regulação negativa e seis proteínas apresentando regulação positiva. Para o módulo de metilação, destaca-se uma proteína apresentando regulação negativa e duas apresentando regulação positiva. Para o módulo de proteólise, destacaram-se quatro proteínas apresentando regulação negativa e quatro apresentando regulação positiva. Para o módulo de oxidação, destacam-se oito sequências apresentando regulação negativa e cinco apresentando regulação positiva. Para o módulo de fosforilação, destacam-se 13 sequências apresentando regulação negativa e oito apresentando regulação positiva.

No intestino, entre as proteínas transcritas relacionadas com o módulo de secreção, destacaram-se oito apresentando regulação negativa e 41 proteínas apresentando regulação positiva. Para o módulo de transporte, destacaram-se quatro apresentando regulação negativa e 11 proteínas apresentando regulação positiva. Para o módulo de metilação, ambas proteínas foram reguladas de forma positiva. Para o módulo de proteólise, destacaram-se setes sequências apresentando regulação positiva e três apresentando regulação negativa. Para o módulo de oxidação, destacaram-se três sequências apresentando regulação negativa e 11 apresentando regulação positiva. Para o módulo de fosforilação, destacaram-se cinco sequências apresentando regulação negativa e 13 sequências apresentando regulação positiva.

No intestino da espécie de carrapato *A. sculptum* infectado ou não por *R. rickettsii*, entre as proteínas transcritas relacionadas com o módulo de secreção, destacaram-se 29 sequências apresentando regulação negativa e nove apresentando regulação positiva. Para o módulo de transporte, destacam-se 13 sequências apresentando regulação negativa e duas apresentando regulação positiva. Para o módulo de metilação não houve resultados a serem demonstrados. Para o módulo de proteólise, destacaram-se sete sequências apresentando regulação negativa e cinco apresentando regulação positiva. Para o módulo de oxidação, destacaram-se cinco sequências apresentando regulação negativa e sete apresentando regulação positiva. Para o

módulo de fosforilação, destacaram-se nove sequências apresentando regulação negativa e oito apresentando regulação positiva.

No intestino da espécie de carrapato *A. sculptum* infectado ou não por *R. amblyommatidis*, entre as proteínas transcritas relacionadas com o módulo de secreção, destacam-se 12 proteínas reguladas de forma negativa e 11 reguladas de forma positiva. Para o módulo de transporte, destacaram-se uma sequência apresentando regulação negativa e duas apresentando regulação positiva. Para o módulo de metilação não houve resultados a serem demonstrados. Para o módulo de proteólise, destacaram-se uma sequência apresentando regulação negativa e uma apresentando regulação positiva. Para o módulo de oxidação, destacaram-se dez sequências apresentando regulação negativa e uma apresentando regulação positiva. Para o módulo de fosforilação, destacaram-se três sequências apresentando regulação negativa e três apresentando regulação positiva.

Nos ovários, entre as proteínas relacionadas com o módulo de secreção, destacaram-se 22 sequências apresentando regulação positiva e 27 apresentando regulação negativa. Para as proteínas transcritas relacionadas com o módulo de transporte, destacam-se dez sequências apresentando regulação negativa e dez apresentando regulação positiva. Para o módulo de metilação foi observado a ocorrência de apenas uma proteína transcrita e regulada positivamente nos ovários, não havendo proteínas transcritas para o intestino. Para o módulo de proteólise, destacam-se seis sequências apresentando regulação negativa e três apresentando regulação positiva. Para o módulo de oxidação, destacaram-se oito sequências apresentando regulação negativa e sete apresentando regulação positiva. Para o módulo de fosforilação, destacaram-se dez sequências apresentando regulação negativa e dez apresentando regulação positiva.

4.9.3 Expressão de transcritos coexpressos

Para a anotação da expressão de transcritos coexpressos, foram selecionados e comparados os dados da média calculada dos valores de *fold change* na base Log₂ para o intestino de carrapato das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*, e da mesma espécie *A. sculptum* infectado por *R. amblyommatidis*. Os transcritos obtidos apresentaram a regulação positiva com os valores de *fold change* na base Log₂ ≥ 2 e com regulação negativa da

expressão com valores de *fold change* na base Log2 ≤ -2 . Os resultados para as glândulas salivares e ovários não foram discutidos (Dados não apresentados).

Para esta discussão, consideramos os resultados da expressão de transcritos para as sequências de proteínas principalmente coexpressas para o intestino das espécies de *Amblyomma* infectados por *Rickettsia* spp. (Anexos 3 e 4). Com os resultados da anotação da expressão de transcritos de carrapato das espécies *A. aureolatum* e *A. sculptum* em condições de infecção e não infecção pela espécie de bactéria *R. rickettsii*, pudemos comparar os níveis da expressão e o comportamento da regulação entre os carrapatos e em consequência da infecção por esta espécie de bactéria, conhecida por sua alta patogenicidade para humanos.

Com isto, averiguamos as similaridades e diferenças pontuais da expressão diferencial de transcritos entre as espécies de carrapatos no conjunto de proteínas preditas para cada módulo funcional significativo do interatoma de membrana, podendo haver diferenças entre as espécies de *Amblyomma* (Tabela 36).

Tabela 36 Quadro de comparação da anotação e expressão diferencial entre intestinos para o conjunto de dados de RNA-seq das espécies de carrapato *A. aureolatum* (Aa_Mg) versus *A. sculptum* (As_Mg), ambos infectados ou não por *R. rickettsii* adotando o modelo do interatoma de proteínas de membrana e ortólogas.

Módulo	Nº de proteínas	Aa_Mg versus As_Mg			
		Aa (+)	As (-)	Aa (+) & As (-)	Aa (-) & As (+)
Transporte	18	1	1	15	1
Secreção	59	16	9	31	3
Metilação	0	0	0	0	0
Proteólise	8	3	1	3	1
Oxidação	17	8	1	7	1
Fosforilação	32	18	5	7	2
Total	134	46	17	63	8

Na Tabela 36, verifica-se o resumo de todo o conjunto de dados avaliados para cada módulo predito sugerindo entre as vias moleculares, diferenças entre as proteínas que apresentaram regulação positiva versus as que apresentaram regulação negativa entre os transcritos expressos nos intestinos de ambas as espécies de carrapatos *Amblyomma* spp. infectados por *R. rickettsii*.

Com os resultados da anotação da expressão de transcritos do intestino de carrapato da espécie *A. sculptum* em condições de diferentes infecções e não pelas espécies de bactérias *R. rickettsii* (Patogênica) e *R. amblyommatis* (Sem casos humanos reportados), comparamos entre estas bactérias os níveis de expressão e abundância de sequências de proteínas putativas para

este órgão. Com isto, pudemos averiguar as similaridades e diferenças pontuais da expressão diferencial de transcritos e potenciais proteínas putativas no mesmo órgão para os módulos funcionais significativos preditos nos interatomos de proteínas de membrana e ortólogas, podendo haver diferenças entre as espécies de *Rickettsia* (Tabela 37).

Tabela 37 Comparação de comparação da anotação e expressão diferencial entre intestinos para o conjunto de dados de RNA-seq da espécie de carrapato *A. sculptum* infectados ou não por *R. rickettsii* (Rr_Mg) versus *R. amblyommatis* (Ra_Mg), adotando o modelo do interatoma de proteínas de membrana e ortólogas.

Módulo	Nº de proteínas	Rr_Mg versus Ra_Mg			
		Rr (+)	Ra (-)	Rr(+) & Ra(-)	Rr(-) & Ra(+)
Transporte	3	0	1	2	0
Secreção	20	2	10	6	2
Metilação	0	0	0	0	0
Proteólise	2	1	1	0	0
Oxidação	8	1	3	4	0
Fosforilação	5	0	4	0	1
Total	38	4	19	12	3

Na Tabela 37, verifica-se um resumo de todo o conjunto de dados avaliados para cada módulo predito sugerindo entre as vias moleculares, diferenças entre as proteínas que apresentaram regulação positiva versus as que apresentaram regulação negativa. Nestes resultados, observamos que a expressão dos transcritos coexpressos, demonstram haver diferenças no intestino entre as espécies de carrapatos *Amblyomma* spp. infectados por *R. rickettsii*. Da mesma forma, haveria diferenças entre as espécies de *Rickettsia* infectando o intestino de carrapato da espécie *A. sculptum*, refletindo em parte os dados obtidos.

4.10 Revisão de literatura para os relatos que confirmem o isolamento e identificação de espécies de *Rickettsia* e seus hospedeiros carrapatos

Para compreendermos as relações ecológicas e evolutivas de bactérias do gênero *Rickettsia* spp. com seus possíveis hospedeiros, realizamos uma revisão bibliográfica, à partir não apenas de dados epidemiológicos para as rickettsioses, mas principalmente de estudos moleculares que relataram por meio de técnicas moleculares a presença destes microorganismos. É de extrema importância afirmar a respeito de relatos da infecção de carrapatos ou qualquer outro hospedeiro artrópode com a identificação da presença de pelo menos de regiões de genes específicos e de sequências do DNA de *Rickettsia*, utilizando de marcadores moleculares.

Além disso, a distribuição geográfica das espécies previamente conhecidas tem aumentado substancialmente com o uso cada vez mais difundido dessas técnicas em diferentes partes do mundo, o que tem contribuído para os estudos epidemiológicos ecológicos da infecção. A análise da sequência de nucleotídeos de determinados genes constitui um método bastante apropriado para identificação e análise filogenética de diferentes espécies de *Rickettsia*. No entanto, para esta revisão foram selecionados somente os estudos que confirmaram o isolamento destas bactérias em seus hospedeiros carrapatos, averiguando também se possuem o genoma sequenciado com depósito em banco de dados do NCBI (Anexo 5). Diversos autores averiguaram a ocorrência de espécies que infectam carrapatos (Figura 18).

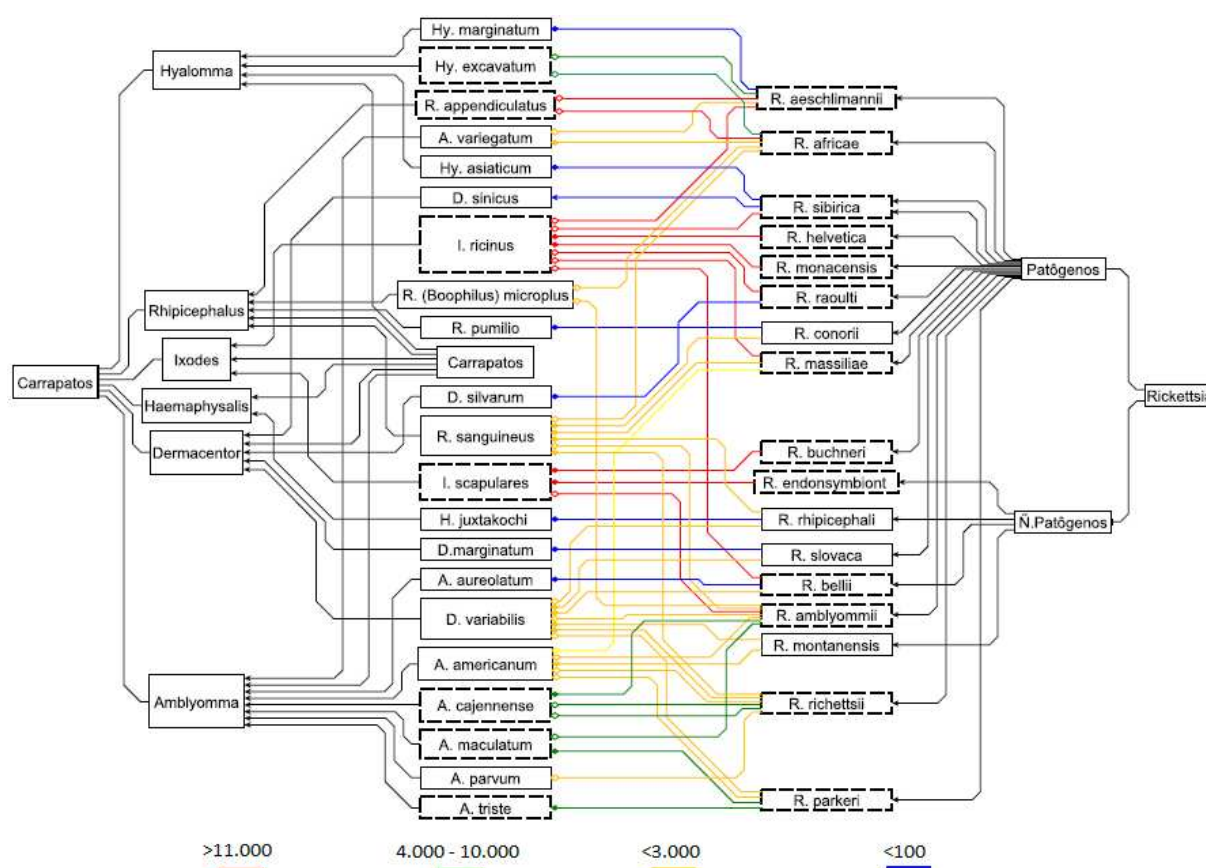
Figura 18 Fluxograma representativo da presença de espécies de bactérias do gênero *Rickettsia*, em gêneros e espécies específicas de carrapatos Ixodídeos, incluindo indivíduos patogênicos e não patogênicos construídos após revisão literária e com o uso do aplicativo *Yed* para a plotagem dos dados.

Destaque (a), verifica-se as espécies de *Rickettsia* com o isolamento comprovado diretamente de carrapato hospedeiro, conforme o fluxograma demonstra (direita para a esquerda) para a bactéria (retângulos) em (seta com ponta fechada).

Destaque (b), verifica-se as espécies de *Rickettsia* que foram identificadas a partir de marcadores moleculares, conforme o fluxograma demonstra (direita para a esquerda) para a bactéria (retângulos) em (seta com ponta aberta).

Destaque (c), verifica-se as espécies de *Rickettsia* com o genoma sequenciado e as espécies de carrapatos que possuem entre 4.000 - 10.000 proteínas podendo ser > do que 11.000 sequências, conforme o fluxograma demonstra (direita para a esquerda) para a bactéria (retângulos tracejados) em (seta com ponta aberta e fechada) e para o carrapato (retângulos tracejados) em (setas cores verde e vermelho).

Destaque (d), verifica-se as demais espécies de carrapatos que possuem um número de proteínas >3000 podendo ser inferior ou <100 sequências interligados com as espécies de *Rickettsia*, conforme o fluxograma demonstra (direita para a esquerda) para a bactéria (retângulos não tracejados) em (seta com ponta aberta e fechada) e para o carrapato (retângulos não tracejados) em (setas cores amarelo e azul).



Fonte: Autor deste trabalho.

Na Figura 18, verifica-se as interações ecológicas essenciais para que possamos inferir as relações filogenéticas sob as possíveis previsões das interações de proteínas no interatoma, utilizando as proteínas do carrapato da espécie *I. scapularis*, servindo também de base para a construção dos transcriptomas e da aproximação dos aspectos filogenéticos entre as proteínas

das espécies de *Rickettsia* que infectam os carrapatos. A espécie *R. endosymbiotic I. scapularis*, conhecida como REIS foi descrita como *R. buchenneri* e infecta o *I. scapularis*, dando a este ácaro o título de real hospedeiro com a passagem transovariana por esta espécie de *Rickettsia* [89].

A escolha de ambas as espécies de *Rickettsia* contribuiu para compreendermos estado evolutivo destas espécies que estão menos distantes filogeneticamente entre elas, podendo ajudar a entender a relação parasita hospedeiro estabelecida entre as espécies que infectam os carrapatos, especialmente as espécies que compõem o Grupo das Febres Maculosas, como a espécie patogênica *R. rickettsii* e a espécie *R. amblyommatis* com patogenicidade reduzida. Neste sentido, construímos este fluxograma para dar suporte as análises evolutivas e filogenéticas para as proteínas de membrana de e das demais proteínas ortólogas presentes no interaroma de membrana.

4.11 Evolução de proteínas de membrana de *Rickettsia* spp.

4.11.1 Análise de similaridade

Para o aprofundamento das análises evolutivas de proteínas de membrana e das demais proteínas ortólogas presentes no interaroma de membrana, adotamos o multialinhamento de sequências com o uso do BLASTp contra as sequências de aminoácidos de proteínas provenientes de 75 genomas de bactérias gram negativas, podendo ser visto na Tabela 35, conforme apresentado no Tópico 3.7.

No total, foram avaliados as sequências de proteínas de membrana de *R. amblyommatis* contra as proteínas de 36 espécies de bactéria, incluindo 26 espécies de *Rickettsia*. Entre estas, consideramos os genomas de algumas estirpes para as espécies de *Rickettsia*, incluindo espécimes dos quatro grupos de classificação definidos até o momento, aonde a maior riqueza destas microorganismos pertencem ao grupo das febres Maculosas e se concentra nas espécies que infectam os carrapatos *Ixodideos*, entre outras que infectam os carrapatos *Argasideos*, espécies que infectam ácaros, além das espécies que infectam os insetos como: pulga, piolhos e moscas. Neste estudo, também consideramos as sequências de aminoácidos de proteínas dos genomas para algumas bactérias *Rickettsiales*, que infectam os carrapatos, utilizando de 2 espécimes representando os gêneros (*Anaplasma* e *Ehrlichia*), duas espécies de *Wolbachia*, que

infectam mosquitos, uma espécie de *Orientia*, que infectam ácaros e uma espécie de *NeoRickettsia*, que infecta trematódeos.

Dessa forma, foram realizados vários multialinhamentos de sequências adotando do BLASTp. para a análise de similaridade das sequências de proteínas de membrana e ortólogas contra as proteínas provenientes de 75 genomas de bactérias. Os resultados do alinhamento contém os IDs para as sequências de proteínas que constituem o interatoma de proteínas de membrana de *R. amblyommatis* contra as sequências de proteínas similares provenientes dos 26 genomas das espécies de *Rickettsia* (Tabela 38).

Tabela 38 Resumo dos resultados de integração e filtro dos ortólogos para a espécie *R. amblyommatis* e outras *Rickettsia* spp. que infectam os carrapatos no interatoma de proteínas de membrana.

ID_proteína <i>R. amblyommatis</i>	ID_gene <i>Rickettsia</i> spp.	ID_genoma	<i>Rickettsia</i> spp.	Estirpe	Confiança ≥70	Identidade ≤1e-10
KJV61004.1	ACP53817	ASM2300v1	<i>R. africae</i>	ESF-5	95,93	1,00E-137
KJV61004.1	ABV75336	ASM1820v1	<i>R. akari</i>	-	93,67	3,00E-136
KJV61004.1	AFC70184	ASM28405v1	<i>R. amblyommatis</i>	GAT-30V	100	1,00E-144
KJV61004.1	KJV61004.1	ASM96467v1	<i>R. amblyommatis</i>	Ac/Pa	100	1,00E-144
KJV61004.1	KJV90703	ASM96499v1	<i>R. amblyommatis</i>	Dark Water	95,77	4,00E-33

Na Tabela 38, verifica-se um resumo do BLASTp para os pares de proteínas de membrana que interagem no interatoma para a *R. amblyommatis* contra as sequências de proteínas provenientes de genomas de outras bactérias e espécies de *Rickettsia*, sendo representadas pelos ID das proteínas, o ID da proteína similar e os valores de confiança e identidade ([Dados suplementares S1p](#)), podendo ser vistos neste [link](#).

4.11.2 Árvores filogenéticas das proteínas de membrana e ortólogas do interatoma

No total, foram obtidas 91 árvores filogenéticas para as proteínas obtidas nos Dados suplementares S1v com o uso do aplicativo Mega (Dados não publicados), sendo este o número de proteínas de *R. amblyommatis* que interagiram com as proteínas do *I. scapularis*. Na grande maioria das árvores filogenéticas, foi observado o agrupamento entre as proteínas que pertencem as espécies *R. belli* e *Rickettsia canadenses*, do Grupo Ancestral que infectam os carrapatos, aparecendo separadas em clados próprios com a presença de outras estirpes das mesmas espécies. Estes grupos foram seguidos do Grupo do Tifo, apresentando separadamente as proteínas para as espécies *R. prowazekii* e *R. typhi*, que infectam ácaros e especialmente piolhos.

Além destes grupos bem definidos, averiguamos a presença de proteínas do grupo de Transição com as espécies *R. akari*, *R. australis* e *R. felis*, além das espécies *R. hoogstraalii* e *R. asemboensis*. Este grupo apresenta bactérias que infectam ácaros, Ixodideo, Argasideo e pulgas. Foi possível averiguar uma divergência entre a formação dos clados à partir das proteínas deste grupo, demonstrando haver uma transição entre os grupos das espécies de *Rickettsia* do Grupo do Tifo para o Grupo das Febres Maculosas ao incluir proteínas da espécie *R. felis*, que infecta pulgas, e é descrita como uma das causadoras desta doença.

As demais sequências de proteínas abrangeram a maior diversidade e riqueza de espécies de *Rickettsia* que infectam os carrapatos pertencentes ao Grupo das Febres Maculosas. Dentre as espécies que pertencem ao Grupo das Febres Maculosas, existe uma separação entre as espécies patogênicas e não patogênicas. Neste estudo observamos que após a formação do clado das proteínas do grupo de Transição, nota-se a ocorrência do clado entre as proteínas pertencentes as bactérias que infectam especificamente o *I. scapularis* com a presença de *R. endonssibionte I. scapularis* e *R. bucheneri*, acompanhadas de *R. endonssibionte I. pacificus* além da *R. monacensis* isolada em *I. ricinus*, aonde a maioria não é patogênica, ou com poucos casos humanos relatados.

Esta última, pertence a uma classificação antiga antes denominada grupo da *Rickettsia helvetica*. Prosseguindo, observamos a ocorrência das proteínas das espécies *R. montana*, *R. rhipicephalis*, *R. raoult*, *R. amblyommatis* e *R. massilea*. Esta última anteriormente dava nome ao grupo. No entanto, na ponta dos cladogramas observamos o restante das proteínas das espécies incluindo *R. japonica*, *R. argasi*, *R. heilongjiangensis*, *R. slova*, *R. sibirica*, *R. africae*, *R. peacocki*, *R. conorii*, *R. parkeri* e *R. rickettsii*, sendo esta última a mais patogênica do grupo das Febres Maculosas.

No entanto, observamos o agrupamento de uma espécie de *Rickettsia* que infecta a mosca branca filogeneticamente próxima de *R. belli* do Grupo Ancestral, e uma outra espécie de *Rickettsia endosymbiotic* que infecta piolho filogeneticamente próxima de *R. rickettsii* do Grupo das Febre Maculosas. Para finalizarmos, algumas árvores apresentaram em seus clados as proteínas mais conservadas filogeneticamente próxima de *R. belli* do Grupo Ancestral seguido de outras bactérias *Rickettsiales* e *Anaplasma* com a presença separada de clado único para *Orientia tsugamich*, e *NeoRickettsia risticii*, além de *Anaplasma phagophitophilum*, *Ehrlichia chafeensis*, *Wolbachia pipiensis* e *Wolbachia endosymbiotic*.

Das 91 proteínas identificadas na interface da interação entre a bactéria da espécie *R. amblyommatis* e o carrapato da espécie *I. scapularis*, foi possível reconstruir 82 árvores filogenéticas devido à remoção de redundância de função gênica para algumas proteínas (Dados não apresentados). Destas, 54 (65,85%) foram capazes de discriminar espécies de *Rickettsia* que possuem habilidade de infectar carrapatos. Além disso, 13 (15,85%) árvores continham apenas espécies de *Rickettsia* capazes de infectar os carrapatos. Esses resultados indicam que a metodologia de construção de interatomas, baseando ao final do uso de métodos de contexto como o proporcionado pelo banco de dados do String pode auxiliar na identificação de potenciais alvos de interação e que podem estar submetidas a pressões evolutivas específicas de acordo com o hospedeiro de interação.

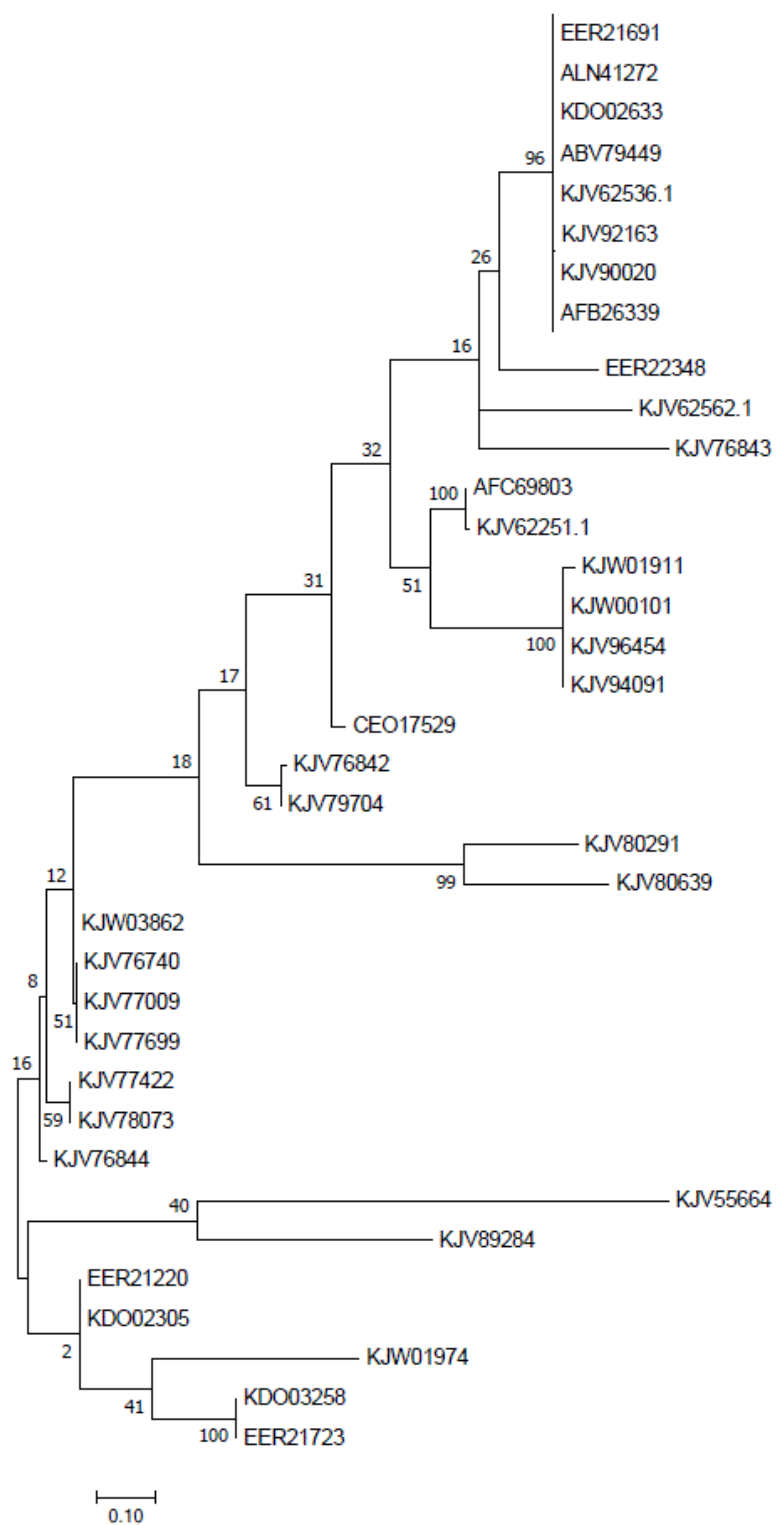
4.11.3 Árvores filogenéticas para as proteínas de membrana de *Rickettsia* spp., que infectam somente os carrapatos

Os resultados indicaram que 13 (15,85%) das árvores filogenéticas, dentre as 82 obtidas continham apenas espécies de *Rickettsia* capazes de infectar os carrapatos. Os cladogramas apresentados para as árvores filogenéticas relacionadas com as sequências de proteínas, sendo; a proteína contendo domínios de repetição de tetratricopeptideo (Tpr) KJV62536, KJV62591.1 e do sistema de secreção do tipo III (T3SS) KJV62562.1; as proteínas KJV62253.1, KJV62505.1, KJV62535.1, KJV62563.1 e KJV62597.1, referentes as proteínas responsáveis pela formação do sistema de secreção do tipo IV (T4SS); a proteína KJV61977.1, relacionada com motivos de repetição de anquiria (Ank); e as proteínas ribossomal L11 metiltransferase KJV62584.1 e KJV62556.1.

Dentre estas árvores filogenéticas, nota-se somente a ocorrência de espécies de *Rickettsia* que infectam os carrapatos Ixodídeos e Argasídeos, incluindo espécime do gênero *Orientia*, que infecta ácaros. Nestas árvores filogenéticas, predominaram a presença de proteínas de espécies que infectam carrapatos não tendo a presença de espécies que infectam pulgas e ou piolhos, além da ocorrência da espécie *Orientia tsugamush* que parasita ácaros. Dependendo da árvore, não se nota espécies de todos os grupos de *Rickettsia*.

A proteína Tpr ID KJV625361.1 apresentou uma árvore filogenética contendo sequências de proteínas similares, podendo ser ortólogas para espécies de *Rickettsia* que infectam somente os carrapatos (Figura 19).

Figura 19 Árvore filogenética para a família das seqüências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62536.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 19, verifica-se que dentre as proteínas similares para a proteína ID KJV62536.1, observou-se a predominância de sequências de proteínas da família contendo motivos repetitivos de tetratricopeptídeos (Tpr) (Tabela 39).

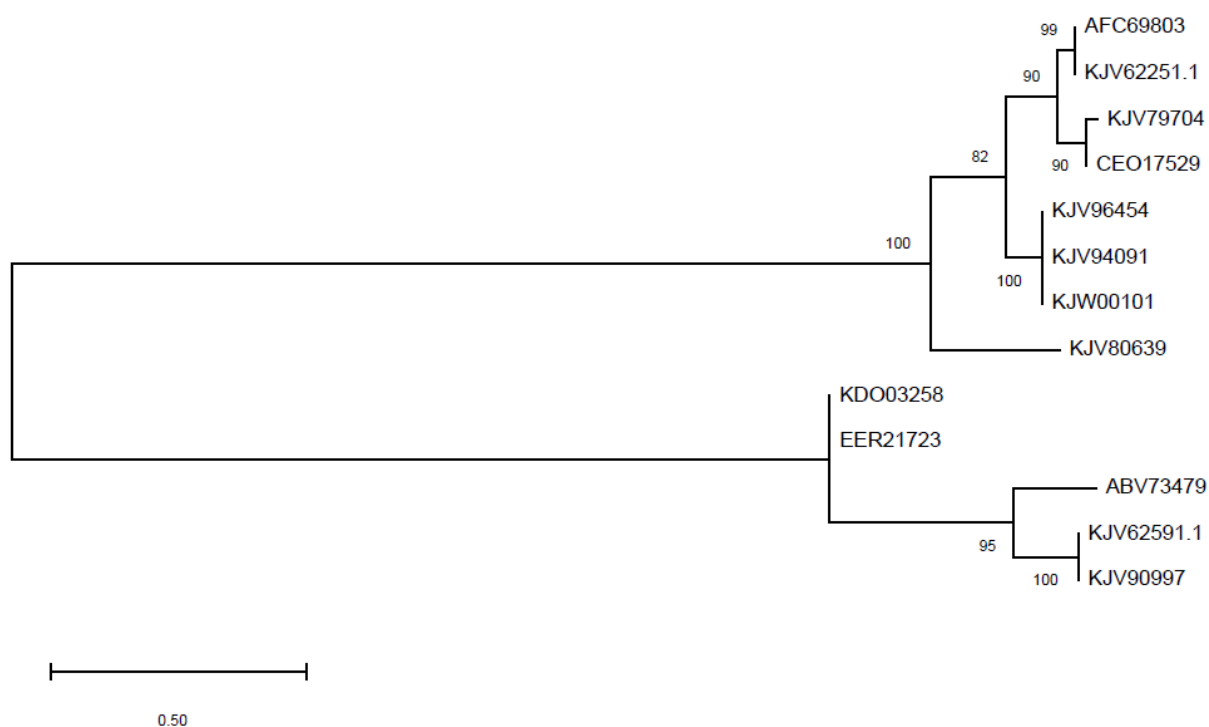
Tabela 39 Quadro de legendas para a árvore filogenética da família de sequências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62536.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio do multalinhamentos de sequências e da análise de filogenia.

Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
EER21691	ASM16073v1	TPR Tetratricopeptide repeat Family protein n	<i>R. endosymbiontis</i>	-	Ixodídeo
ALN41272	ASM144247v1	Hypothetical protein	<i>R. rhipicephali</i>	-	Ixodídeo
KDO02633	REISMNv1	Lipoprotein Nlpl	<i>R. buchneri</i>	ISO7	Ixodídeo
ABV79449	ASM1824v1	Tetratricopeptide repeat Family protein	<i>R. belli</i>	OSU 85-389	Ixodídeo
KJV62536.1	ASM96467v1	TPR Tetratricopeptide repeat Family protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV92163	ASM96504v1	TPR Tetratricopeptide repeat Family protein n	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV90020	ASM96500v1	TPR Tetratricopeptide repeat Family protein n	<i>R. belli</i>	RML An4	Ixodídeo
AFB26339	ASM28399v1	Hypothetical protein	<i>R. philipe</i>	364D	Ixodídeo
EER22348	ASM16073v1	TPR Tetratricopeptide repeat Family protein n	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV62562.1	ASM96467v1	D	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV76843	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
AFC69803	ASM28405v1	Hypothetical protein	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV62251.1	ASM96467v1	TPR Tetratricopeptide repeat Family protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJW01911	ASM96518v1	TPR Tetratricopeptide repeat Family protein n	<i>R. argasi</i>	T170-B	Argasídeo
KJW00101	ASM96514v1	TPR Tetratricopeptide repeat Family protein n	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJV96454	ASM96507v1	TPR Tetratricopeptide repeat Family protein n	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV94091	ASM96508v1	TPR Tetratricopeptide repeat Family protein n	<i>R. parkeri</i>	Grand Bay	Ixodídeo
CEO17529	RMONA1	Tetratricopeptide repeat Family protein	<i>R. monacensis</i>	IrR/Munich	Ixodídeo
KJV76842	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79704	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80291	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80639	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJW03862	ASM96518v1	TPR Tetratricopeptide repeat Family protein n	<i>R. argasi</i>	T170-B	Argasídeo
KJV76740	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77009	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77699	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77422	ASM96484v1	FAT domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78073	ASM96484v1	FAT domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76844	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV55664	ASM96458v1	TPR Tetratricopeptide repeat Family protein n	<i>O. tsugamuch</i>	-	Ácaro
KJV89284	ASM96500v1	TPR Tetratricopeptide repeat Family protein n	<i>R. belli</i>	RML An4	Ixodídeo
EER21220	ASM16073v1	Tetratricopeptide repeat Family protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KDO02305	REISMNv1	Putative O-linked N-acetylglucosay	<i>R. buchneri</i>	ISO7	Ixodídeo
KJW01974	ASM96518v1	TPR Tetratricopeptide repeat Family protein n	<i>R. argasi</i>	T170-B	Argasídeo
KDO03258	REISMNv1	Hypothetical protein	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21723	ASM16073v1	TPR Tetratricopeptide repeat Family protein n	<i>R. endosymbiontis</i>	-	Ixodídeo

Na Tabela 39, verifica-se a proteína Tpr ID KJV62536.1 e suas sequências de proteínas similares, os IDs dos genomas de referência das bactérias, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas sequências pertencem.

A proteína Tpr ID KJV62591.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 20).

Figura 20 Árvore filogenética para a família das seqüências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62591.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia com o uso do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 20, verifica-se que dentre as proteínas similares para a proteína ID KJV62591.1, observou-se a predominância de seqüências de proteínas da família contendo motivos repetitivos de tetratricopeptídeos (Tpr) (Tabela 40).

Tabela 40 Quadro de legendas para a árvore filogenética da família de seqüências de proteínas contendo motivos repetitivos de tetratricopeptídeo (Tpr) com destaque para a proteína ID KJV62591.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio do multialinhamento de seqüências e da análise de filogenia.

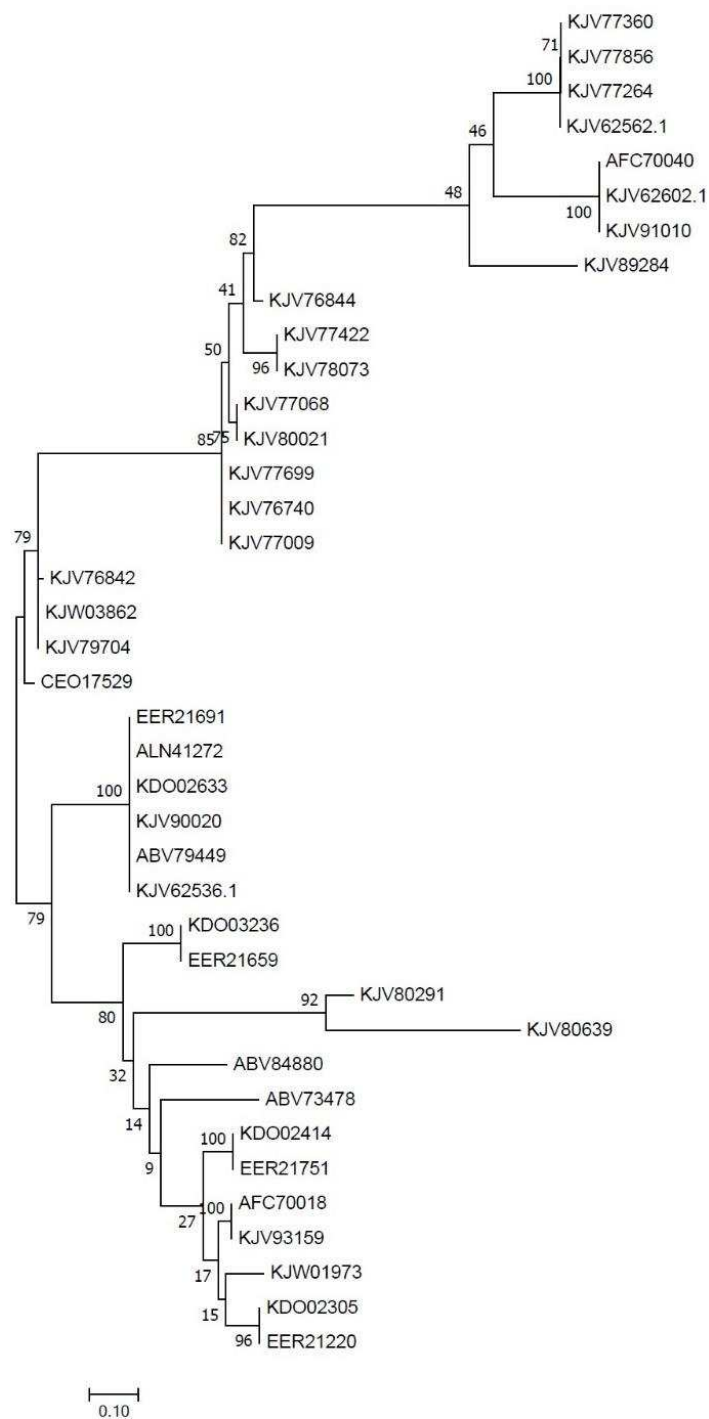
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
AFC69803	ASM28405v1	Hypothetical protein	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV62251.1	ASM96467v1	TPR Tetratricopeptide repeat Family protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV79704	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
CEO17529	RMONA1	TPR Tetratricopeptide repeat Family protein	<i>R. monacensis</i>	IrR/Munich	Ixodídeo
KJV96454	ASM96507v1	TPR Tetratricopeptide repeat Family protein n	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV94091	ASM96508v1	TPR Tetratricopeptide repeat Family protein n	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJW00101	ASM96514v1	TPR Tetratricopeptide repeat Family protein n	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJV80639	ASM96484v1	TPR Tetratricopeptide repeat Family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KDO03258	REISMNv1	Hypothetical protein	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21723	ASM16073v1	TPR Tetratricopeptide repeat Family protein n	<i>R. endosymbiontis</i>	-	Ixodídeo
ABV73479	ASM1434v1	TPR Tetratricopeptide repeat Family protein n	<i>R. canadenses</i>	McKiel	Ixodídeo
KJV62591.1	ASM96467v1	TPR Tetratricopeptide repeat Family protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV90997	ASM96499v1	TPR Tetratricopeptide repeat Family protein n	<i>R. amblyommatis</i>	Dark Water	Ixodídeo

Na Tabela 40, verifica-se a proteína ID KJV62591.1 e suas sequências de proteínas similares, os IDs dos genomas de referência das bactérias, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas sequências pertencem.

A proteína T3SS KJV62562.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 21).

Figura 21 Árvore filogenética para a família das sequências de proteínas do sistema de secreção do tipo III T3SS com destaque para a proteína ID KJV62562.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares

obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 21, verifica-se que dentre as proteínas similares para a proteína ID KJV62562.1, observou-se predominantemente sequências de proteínas da família contendo motivos repetitivos de tetratricopeptídeos (Tpr) (Tabela 41).

Tabela 41 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção do tipo III T3SS com destaque para a proteína ID KJV62562.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

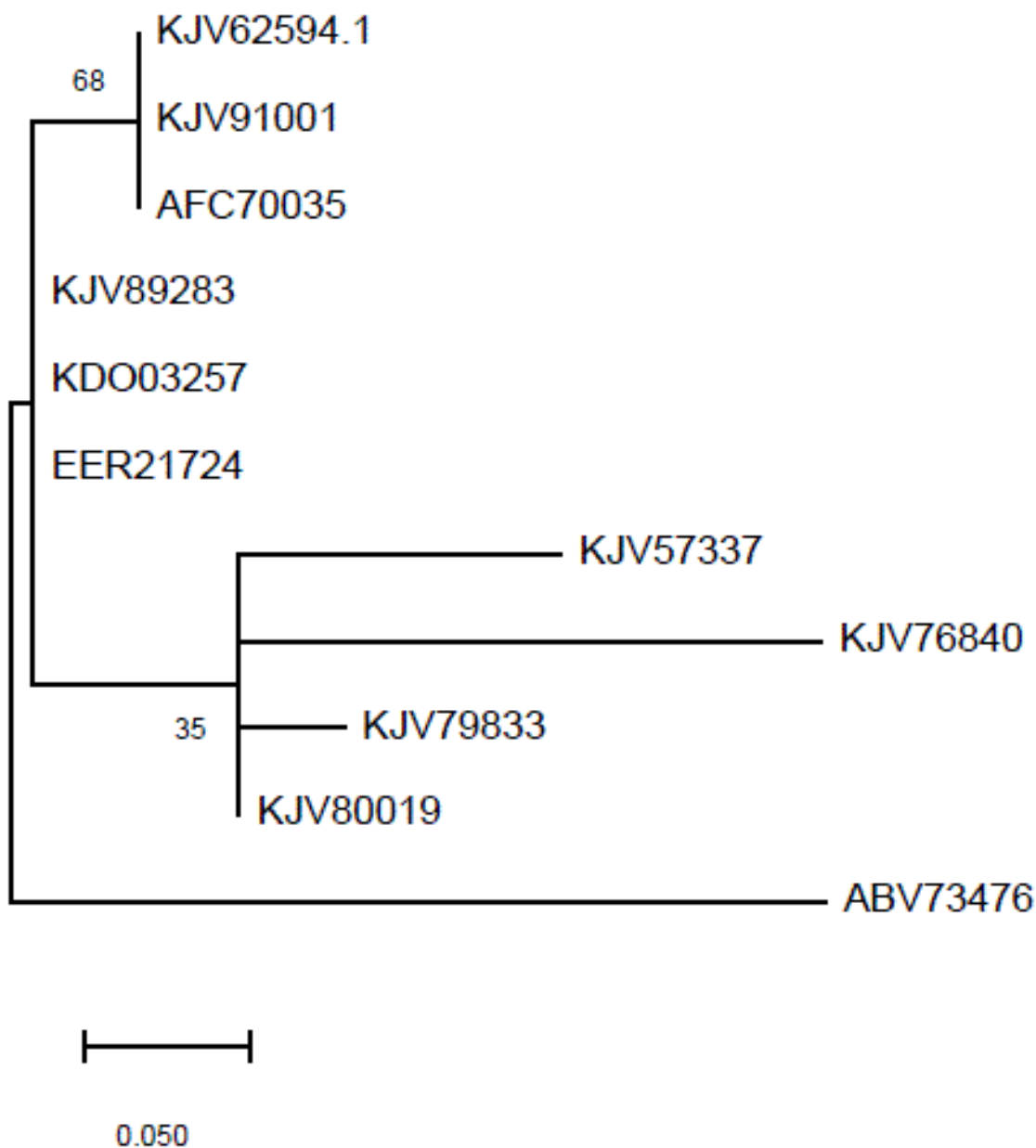
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
KJV77360	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77856	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77264	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62562.1	ASM96467v1	Type III secretion low calcium response chaperone	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
AFC70040	ASM28405v1	Hypothetical protein	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV62602.1	ASM96467v1	Tetratricopeptide repeat family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV91010	ASM96499v1	Tetratricopeptide repeat family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV89284	ASM96500v1	Tetratricopeptide repeat family protein	<i>R. belli</i>	RML An4	Ixodídeo
KJV77068	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80021	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77422	ASM96484v1	Fat domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78073	ASM96484v1	Fat domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76844	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77699	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76740	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77009	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76842	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV03862	ASM96518v1	Tetratricopeptide repeat family protein	<i>R. argasi</i>	T170-B	Argasídeo
KJV79704	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
CEO17529	RMONA1	Tetratricopeptide repeat family protein	<i>R. monacensis</i>	IrR/Munich	Ixodídeo
EER21691	ASM16073v1	Tetratricopeptide repeat	<i>R. endosymbiontis</i>	-	Ixodídeo
ALN41272	ASM144247v1	Hypothetical protein	<i>R. rhipicephali</i>	-	Ixodídeo
KDO02633	REISMNv1	Lipoprotein NlpI	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV90020	ASM96500v1	Tetratricopeptide repeat family protein	<i>R. belli</i>	RML An4	Ixodídeo
ABV79449	ASM1824v1	Tetratricopeptide repeat family protein	<i>R. belli</i>	OSU 85-389	Ixodídeo
KJV62536.1	ASM96467v1	Tetratricopeptide repeat family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KDO03236	REISMNv1	Lipoprotein NlpI	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21659	ASM16073v1	Tetratricopeptide repeat family protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV80291	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80639	ASM96484v1	Tetratricopeptide repeat family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
ABV84880	ASM1662v1	Tetratricopeptide repeat family protein	<i>R. massilea</i>	MTU5	Ixodídeo
ABV73478	ASM1434v1	Tetratricopeptide repeat family protein	<i>R. canadensis</i>	McKiel	Ixodídeo
KDO02414	REISMNv1	Lipoprotein NlpI	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21751	ASM16073v1	Tetratricopeptide repeat family protein	<i>R. endosymbiontis</i>	-	Ixodídeo
AFC70018	ASM28405v1	Hypothetical protein	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV93159	ASM96499v1	Tetratricopeptide repeat family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV01973	ASM96518v1	Tetratricopeptide repeat family protein	<i>R. argasi</i>	T170-B	Argasídeo
KDO02305	REISMNv1	O-linked -N-acetylglucosamine transferase	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21220	ASM16073v1	Tetratricopeptide repeat family protein	<i>R. endosymbiontis</i>	-	Ixodídeo

Na Tabela 41, verifica-se a proteína ID KJV62562.1 e suas sequências de proteínas similares, os IDs dos genomas de referência das bactérias, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, as quais estas sequências pertencem.

A proteína Ftsk ID KJV62594.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 22).

Figura 22 Árvore filogenética para a família das sequências de proteínas contendo (Ftsk) com destaque para a proteína ID KJV62594.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do

BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho

Na Figura 22, verifica-se que dentre as proteínas similares para a proteína Ftsk ID KJV62594.1, observou-se predominantemente sequências de proteínas da família contendo sistemas de secreção do tipo IV T4SS (Tabela 42).

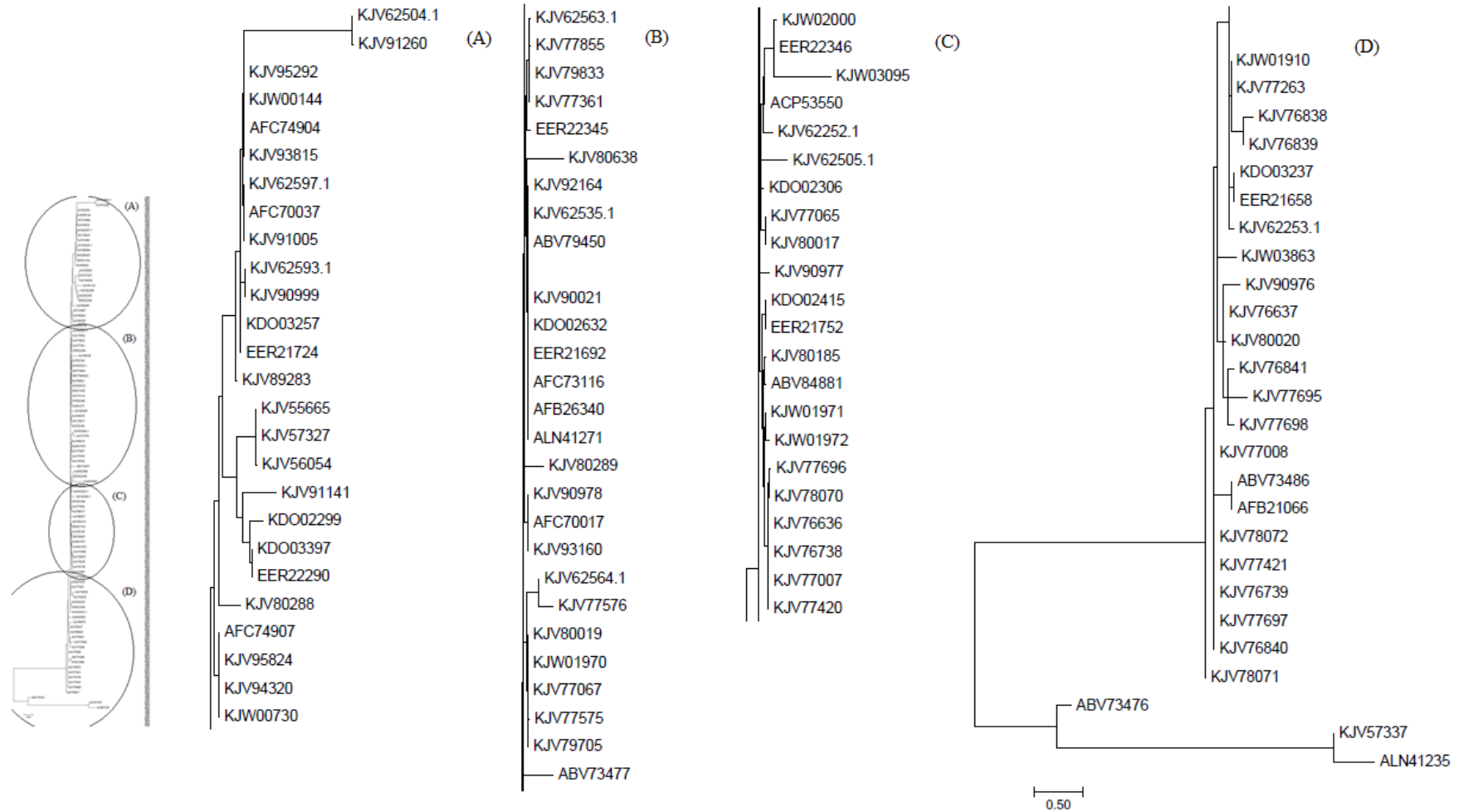
Tabela 42 Quadro de legendas para a árvore filogenética da família de sequências de proteínas recombinantes responsáveis pelos movimentos de translocação e conjugação do DNA (Ftsk/SpoIIIE) com destaque para a proteína ID KJV62594.1 pertencente a espécie *R. amblyommatidis* Ac/Pa e os seus similares obtidos por meio do multialinhamentos de sequências e da análise de filogenia.

Proteína	Genoma	Nome	<i>Rickettsia</i> spp	Estirpe	Hospedeiro
KJV62594.1	ASM96467v1	Ftsk / SpoIIIE family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV91001	ASM96499v1	Ftsk / SpoIIIE family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
AFC70035	ASM28405v1	Conjugative transfer protein Trad	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV89283	ASM96500v1	Type IV secretory syst Conjugative DNA transfer	<i>R. belli</i>	RML An4	Ixodídeo
KDO03257	REISMNv1	DNA transport protein TraD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21724	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV57337	ASM96458v1	Type II/IV secretory system family protein	<i>O. tsugamuch</i>	-	Ácaro
KJV76840	ASM96484v1	Ftsk / SpoIIIE family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79833	ASM96484v1	Type IV secretory syst Conjugative DNA transfer	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80019	ASM96484v1	Type IV secretory syst Conjugative DNA transfer	<i>R. hoogstraalii</i>	RCCE3	Argasídeo

Na Tabela 42, verifica-se a proteína Ftsk ID KJV62594.1 e suas seqüências de proteínas similares, os IDs dos genomas de referência das bactérias, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas seqüências pertencem.

A proteína T2T4SS ID KJV62252.1 apresentou uma árvore filogenética contendo seqüências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 23).

Figura 23 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo II/IV (T2T4SS) com destaque para a proteína ID KJV62252.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho

Na Figura 23, verifica-se que dentre as proteínas similares para a proteína ID KJV62252.1, observou-se predominantemente a presença de sequências de proteínas da família responsável pela formação do sistema de secreção do tipo II e IV (Tabela 43).

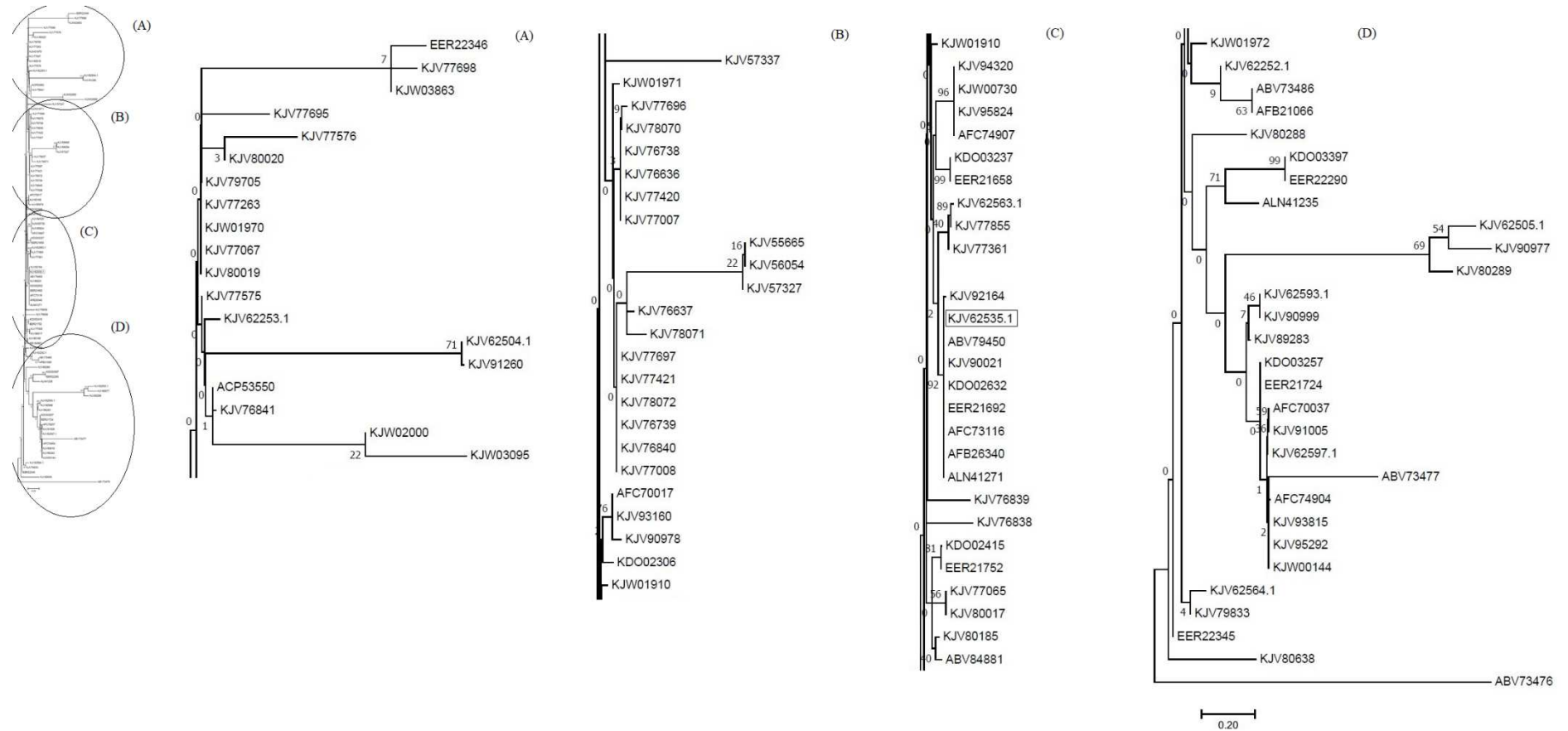
Tabela 43 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T2T4SS com destaque para a proteína ID KJV62252.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
ABV73476	ASM1434v1	Putative conjugative transfer protein TraD	<i>R. canadensis</i>	McKiel	Ixodídeo
ALN41235	ASM144247v1	Conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
ALN41271	ASM144247v1	Conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
EER21658	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21692	ASM16073v1	Putative conjugative transfer protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21752	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER22345	ASM16073v1	Protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
EER22346	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
ABV84881	ASM1662v1	Putative conjugative transfer protein TraD	<i>R. massilea</i>	MTU5	Ixodídeo
ABV79450	ASM1824v1	Putative conjugative transfer protein TraD	<i>R. belli</i>	OSU 85-389	Ixodídeo
AFB26340	ASM28399v1	Conjugative transfer protein TraD	<i>R. philipe</i>	364D	Ixodídeo
AFC70017	ASM28405v1	conjugative transfer protein TraD	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
AFC73116	ASM28417v1	Conjugative transfer protein TraD	<i>R. montana</i>	OSU85-930	Ixodídeo
AFC74907	ASM28419v1	Conjugative transfer protein TraD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
KJV57337	ASM96458v1	Type II/IV secretory system family protein	<i>O. tsugamuch</i>	-	Mite
KJV62252.1	ASM96467v1	Type II/IV secretory system family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62505.1	ASM96467v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62535.1	ASM96467v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62563.1	ASM96467v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV76636	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE19	Argasídeo
KJV76637	ASM96484v1	F sex factor N terminal family protein	<i>R. hoogstraalii</i>	RCCE18	Argasídeo
KJV76738	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE17	Argasídeo
KJV76739	ASM96484v1	F sex factor N terminal family protein	<i>R. hoogstraalii</i>	RCCE16	Argasídeo
KJV76840	ASM96484v1	ftsK/SpoIIIE family protein	<i>R. hoogstraalii</i>	RCCE15	Argasídeo
KJV77007	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE14	Argasídeo
KJV77008	ASM96484v1	F sex factor N terminal family protein	<i>R. hoogstraalii</i>	RCCE13	Argasídeo
KJV77067	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE12	Argasídeo
KJV77361	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE11	Argasídeo
KJV77420	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE10	Argasídeo
KJV77421	ASM96484v1	F sex factor N terminal family protein	<i>R. hoogstraalii</i>	RCCE9	Argasídeo
KJV77575	ASM96484v1	Type IV secretory system coupling DNA-binding domain protein	<i>R. hoogstraalii</i>	RCCE8	Argasídeo
KJV77697	ASM96484v1	ftsK/SpoIIIE family protein	<i>R. hoogstraalii</i>	RCCE7	Argasídeo
KJV77855	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE6	Argasídeo
KJV78071	ASM96484v1	Type IV secretory system coupling DNA-binding domain protein	<i>R. hoogstraalii</i>	RCCE5	Argasídeo
KJV78072	ASM96484v1	F sex factor N terminal family protein	<i>R. hoogstraalii</i>	RCCE4	Argasídeo
KJV79705	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79833	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE2	Argasídeo
KJV80019	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE1	Argasídeo
KJV80020	ASM96484v1	F sex factor N terminal family protein	<i>R. hoogstraalii</i>	RCCE0	Argasídeo
KJV80185	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE1	Argasídeo
KJV80289	ASM96484v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. hoogstraalii</i>	RCCE2	Argasídeo
KJV90977	ASM96499v1	Type IV secretory system coupling DNA-binding domain protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV90978	ASM96499v1	Type II/IV secretory system family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV93160	ASM96499v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV89283	ASM96500v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. belli</i>	RML An2	Ixodídeo
KJV90021	ASM96500v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. belli</i>	RML An3	Ixodídeo
KJV92164	ASM96504v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV95824	ASM96507v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. parkeri</i>	AT#23	Ixodídeo
KJV94320	ASM96508v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJW00730	ASM96514v1	Type IV secretory system conjugative DNA transfer family protein	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJW01910	ASM96518v1	F sex factor N terminal family protein	<i>R. argasi</i>	T170-B	Argasídeo
KJW01971	ASM96518v1	Type IV secretory system coupling DNA-binding domain protein	<i>R. argasi</i>	T170-B	Argasídeo
KJW01972	ASM96518v1	Type II/IV secretory system family protein	<i>R. argasi</i>	T170-B	Argasídeo
KJW03863	ASM96518v1	Type II/IV secretory system family protein	<i>R. argasi</i>	T170-B	Argasídeo
KDO02306	REISMNv1	DNA transport protein TraD	<i>R. buchneri</i>	ISO3	Ixodídeo
KDO02415	REISMNv1	DNA transport protein TraD	<i>R. buchneri</i>	ISO4	Ixodídeo
KDO02632	REISMNv1	DNA transport protein TraD	<i>R. buchneri</i>	ISO5	Ixodídeo
KDO03237	REISMNv1	DNA transport protein TraD	<i>R. buchneri</i>	ISO6	Ixodídeo

Na Tabela 43, observa-se a proteína ID KJV62252.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertencem.

A proteína T4SS ID KJV62535.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 24).

Figura 24 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62535.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 24, verifica-se que dentre as proteínas similares para a proteína ID KJV62535.1, observou-se predominantemente a presença de sequências de proteínas da família responsável pela formação do sistema de secreção do tipo IV (Tabela 44).

Tabela 44 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62535.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

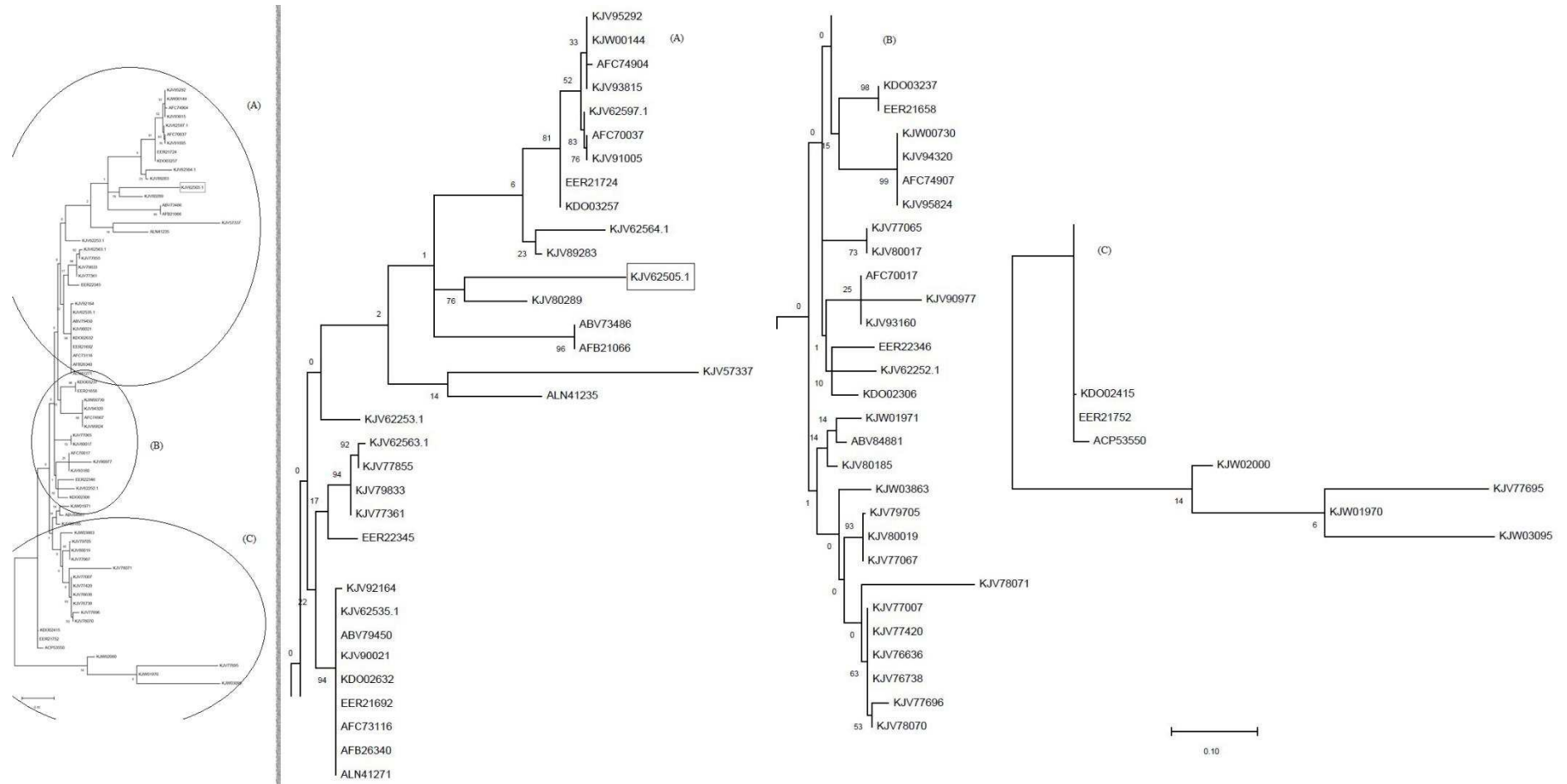
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
ABV73476	ASM1434v1	Putative conjugative transfer D	<i>R. canadensis</i>	McKiel	Ixodídeo
ALN41235	ASM144247v1	Conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
ALN41271	ASM144247v1	Conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
EER21658	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21724	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21752	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER22290	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
AFB26340	ASM28399v1	Conjugative transfer protein D	<i>R. philipe</i>	364D	Ixodídeo
AFC70017	ASM28405v1	Conjugative transfer protein D	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
AFC73116	ASM28417v1	Conjugative transfer protein D	<i>R. montanna</i>	OSU85-930	Ixodídeo
AFC74907	ASM28419v1	Conjugative transfer protein D	<i>R. parkeri</i>	Portsmouth	Ixodídeo
EER22346	ASM16073v1	Conjugative transfer protein N	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV62593.1	ASM96467v1	D	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KDO02306	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02415	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02632	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03237	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03257	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03397	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV76637	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76739	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76841	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77008	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77421	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77698	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78072	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80020	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80288	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80638	ASM96484v1	F sex factor N terminal N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV91260	ASM96499v1	F sex factor N terminal N	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJW01910	ASM96518v1	F sex factor N terminal N	<i>R. argasi</i>	T170-B	Argasídeo
KJV76840	ASM96484v1	ftsK/SpoIIIE family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77697	ASM96484v1	ftsK/SpoIIIE family protein n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62504.1	ASM96467v1	N	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62505.1	ASM96467v1	N	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62535.1	ASM96467v1	N	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62563.1	ASM96467v1	N	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62564.1	ASM96467v1	N	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62597.1	ASM96467v1	N	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
EER22345	ASM16073v1	Protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
AFC70037	ASM28405v1	Putative conjugative transfer D	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV90999	ASM96499v1	Putative conD	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
ABV84881	ASM1662v1	Putative conjugative D	<i>R. massilea</i>	MTU5	Ixodídeo
ACP53550	ASM2300v1	Putative conjugative D	<i>R. africae</i>	ESF-5	Ixodídeo
KJV77576	ASM96484v1	Putative conjugative traD	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
ABV73477	ASM1434v1	Putative conjugative transfer D	<i>R. canadensis</i>	McKiel	Ixodídeo
ABV73486	ASM1434v1	Putative conjugative transfer D	<i>R. canadensis</i>	McKiel	Ixodídeo
ABV79450	ASM1824v1	Putative conjugative transfer D	<i>R. belli</i>	OSU 85-389	Ixodídeo
AFB21066	ASM28391v1	Putative conjugative transfer D	<i>R. canadensis</i>	CA410	Ixodídeo
AFC74904	ASM28419v1	Putative conjugative transfer D	<i>R. parkeri</i>	Portsmouth	Ixodídeo

EER21692	ASM16073v1	Putative conjugative transfer D	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV77263	ASM96484v1	Putative conjugative transfer n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62252.1	ASM96467v1	Tn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62253.1	ASM96467v1	Tn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV57337	ASM96458v1	Type II/IV secretory system N	<i>O. tsugamuch</i>	-	Mite
KJV90978	ASM96499v1	Type II/IV secretory system N	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJW01972	ASM96518v1	Type II/IV secretory system N	<i>R. argasi</i>	T170-B	Argasídeo
KJW03863	ASM96518v1	Type II/IV secretory system N	<i>R. argasi</i>	T170-B	Argasídeo
KJV56054	ASM96458v1	Type IV secretory system Cn	<i>O. tsugamuch</i>	-	Mite
KJV55665	ASM96458v1	Type IV secretory system N	<i>O. tsugamuch</i>	-	Mite
KJV57327	ASM96458v1	Type IV secretory system N	<i>O. tsugamuch</i>	-	Mite
KJV76636	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76738	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76838	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76839	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77007	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77065	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77067	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77361	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77420	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77575	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77695	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77696	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77855	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78070	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78071	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79705	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79833	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80017	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80019	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80185	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80289	ASM96484v1	Type IV secretory system N	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV89283	ASM96500v1	Type IV secretory system N	<i>R. belli</i>	RML An4	Ixodídeo
KJV90021	ASM96500v1	Type IV secretory system N	<i>R. belli</i>	RML An4	Ixodídeo
KJV90977	ASM96499v1	Type IV secretory system N	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV91005	ASM96499v1	Type IV secretory system N	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV92164	ASM96504v1	Type IV secretory system N	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV93160	ASM96499v1	Type IV secretory system N	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV93815	ASM96508v1	Type IV secretory system N	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJV94320	ASM96508v1	Type IV secretory system N	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJV95292	ASM96507v1	Type IV secretory system N	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV95824	ASM96507v1	Type IV secretory system N	<i>R. parkeri</i>	AT#24	Ixodídeo
KJW00144	ASM96514v1	Type IV secretory system N	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJW00730	ASM96514v1	Type IV secretory system N	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJW01970	ASM96518v1	Type IV secretory system N	<i>R. argasi</i>	T170-B	Argasídeo
KJW01971	ASM96518v1	Type IV secretory system N	<i>R. argasi</i>	T170-B	Argasídeo
KJW02000	ASM96518v1	Type IV secretory system N	<i>R. argasi</i>	T170-B	Argasídeo
KJW03095	ASM96515v1	Type IV secretory system N	<i>R. endosymbiontis</i>	-	Ixodídeo

Na Tabela 44, verifica-se a proteína ID KJV62535.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertencem.

A proteína T4SS ID KJV62505.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 25).

Figura 25 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62505.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 25, dentre as proteínas similares para a proteína ID KJV62505.1, observou-se predominantemente a presença de sequências de proteínas da família responsável pela formação do sistema de secreção do tipo IV (Tabela 45).

Tabela 45 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62505.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

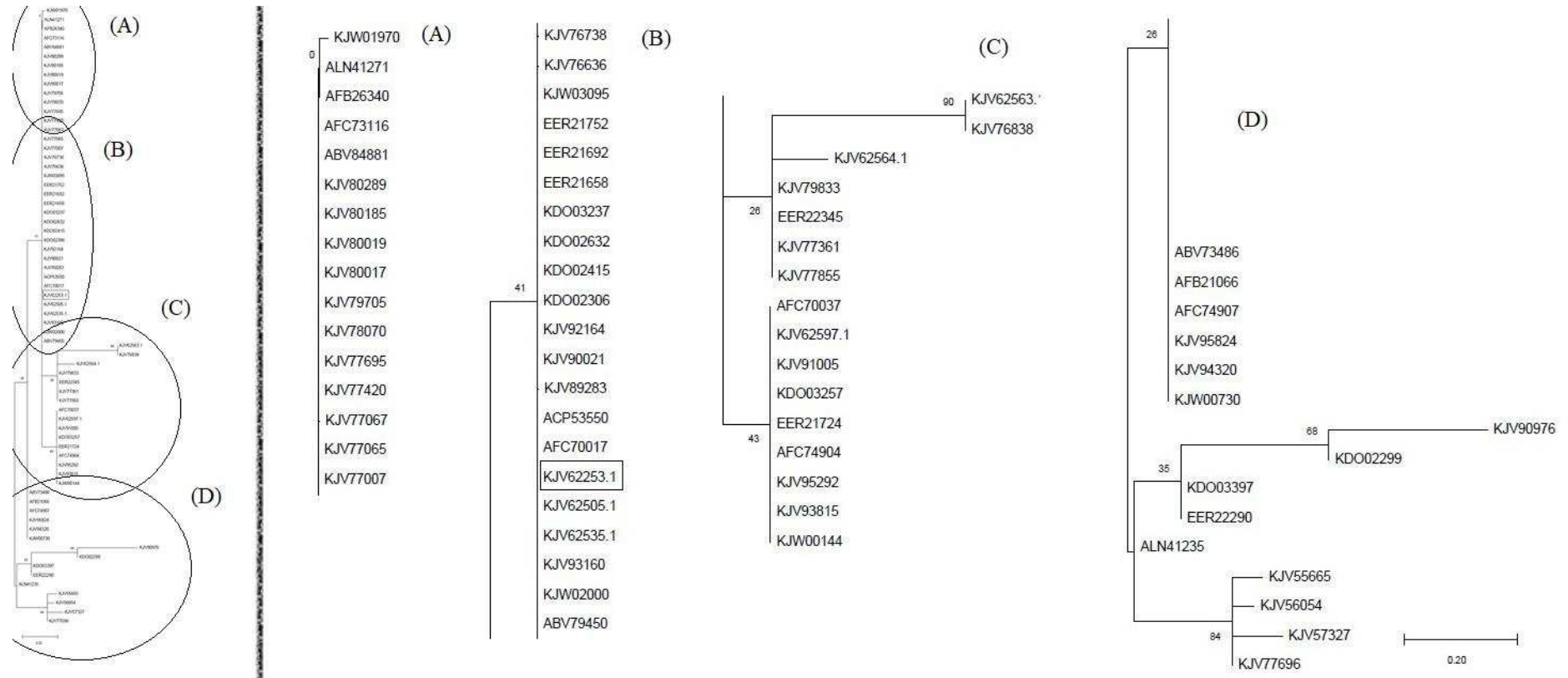
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
KJV95292	ASM96507v1	Type IV secretory systn	<i>R. parkeri</i>	AT#24	Ixodídeo
KJW00144	ASM96514v1	Type IV secretory systn	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
AFC74904	ASM28419v1	Putative conjugative transfeD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
KJV93815	ASM96508v1	Type IV secretory systn	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJV62597.1	ASM96467v1	Protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
AFC70037	ASM28405v1	Putative conjugative transfeD	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV91005	ASM96499v1	Type IV secretory systn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
EER21724	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KDO03257	REISMNv1	D 2 dDNA transpoD	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV62564.1	ASM96467v1	Protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV89283	ASM96500v1	Type IV secretory systn	<i>R. belli</i>	RML An4	Ixodídeo
KJV62505.1	ASM96467v1	protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV80289	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
ABV73486	ASM1434v1	Putative conjugative transfeD	<i>R. canadensis</i>	McKiel	Ixodídeo
AFB21066	ASM28391v1	Putative conjugative transfeD	<i>R. canadensis</i>	CA410	Ixodídeo
KJV57337	ASM96458v1	Type II/IV secretion syn	<i>O. tsugamuch</i>	-	Mite
ALN41235	ASM144247v1	Conjugative transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
KJV62253.1	ASM96467v1	Protein coding dtn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62563.1	ASM96467v1	protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV77855	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79833	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77361	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
EER22345	ASM16073v1	Protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV92164	ASM96504v1	Type IV secretory systn	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV62535.1	ASM96467v1	Protein coding description	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
ABV79450	ASM1824v1	Putative conjugative transfeD	<i>R. belli</i>	OSU 85-389	Ixodídeo
KJV90021	ASM96500v1	Type IV secretory systn	<i>R. belli</i>	RML An4	Ixodídeo
KDO02632	REISMNv1	raD 4 dDNA transD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21692	ASM16073v1	Putative conjugative transfeD	<i>R. endosymbiontis</i>	-	Ixodídeo
AFC73116	ASM28417v1	Conjugative transfer proteinD	<i>R. montanna</i>	OSU85-930	Ixodídeo
AFB26340	ASM28399v1	Conjugative transfer proteinD	<i>R. philipe</i>	364D	Ixodídeo
ALN41271	ASM144247v1	Conjugative transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
KDO03237	REISMNv1	aD 3 dDNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21658	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJW00730	ASM96514v1	Type IV secretory systn	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJV94320	ASM96508v1	Type IV secretory systn	<i>R. parkeri</i>	Grand Bay	Ixodídeo
AFC74907	ASM28419v1	Conjugative transfer proteinD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
KJV95824	ASM96507v1	Type IV secretory systn	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV77065	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80017	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
AFC70017	ASM28405v1	Conjugative transfer proteinD	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV90977	ASM96499v1	Type IV secretory systn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV93160	ASM96499v1	Type IV secretory systn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
EER22346	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV62252.1	ASM96467v1	protein coding dtn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KDO02306	REISMNv1	6 dDNA transportD	<i>R. buchneri</i>	ISO7	Ixodídeo
KJW01971	ASM96518v1	Type IV secretory systn	<i>R. argasi</i>	T170-B	Argasídeo
ABV84881	ASM1662v1	F dPutative coD	<i>R. massilea</i>	MTU5	Ixodídeo
KJV80185	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJW03863	ASM96518v1	type II/IV secretn	<i>R. argasi</i>	T170-B	Argasídeo
KJV79705	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80019	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77067	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78071	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77007	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77420	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo

KJV76636	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76738	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77696	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78070	ASM96484v1	Type IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KDO02415	REISMNv1	aD 5 dDNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21752	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
ACP53550	ASM2300v1	PutativeD	<i>R. africae</i>	ESF-5	Ixodídeo
KJW02000	ASM96518v1	Typr IV secretory systn	<i>R. argasi</i>	T170-B	Argasídeo
KJV77695	ASM96484v1	Typr IV secretory systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJW01970	ASM96518v1	Typr IV secretory systn	<i>R. argasi</i>	T170-B	Argasídeo
KJW03095	ASM96515v1	Typr IV secretory systn	<i>R. endosymbiontis</i>	-	Ixodídeo

Na Tabela 45, verifica-se a proteína ID KJV62505.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertence.

A proteína ID KJV62597.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 26).

Figura 26 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62253.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 26, verifica-se que dentre as proteínas similares para a proteína ID KJV62253.1, observou-se predominantemente a presença de sequências de proteínas da família responsável pela formação do sistema de secreção do tipo IV (Tabela 46).

Tabela 46 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62253.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio do multialinhamentos de sequências e da análise de filogenia.

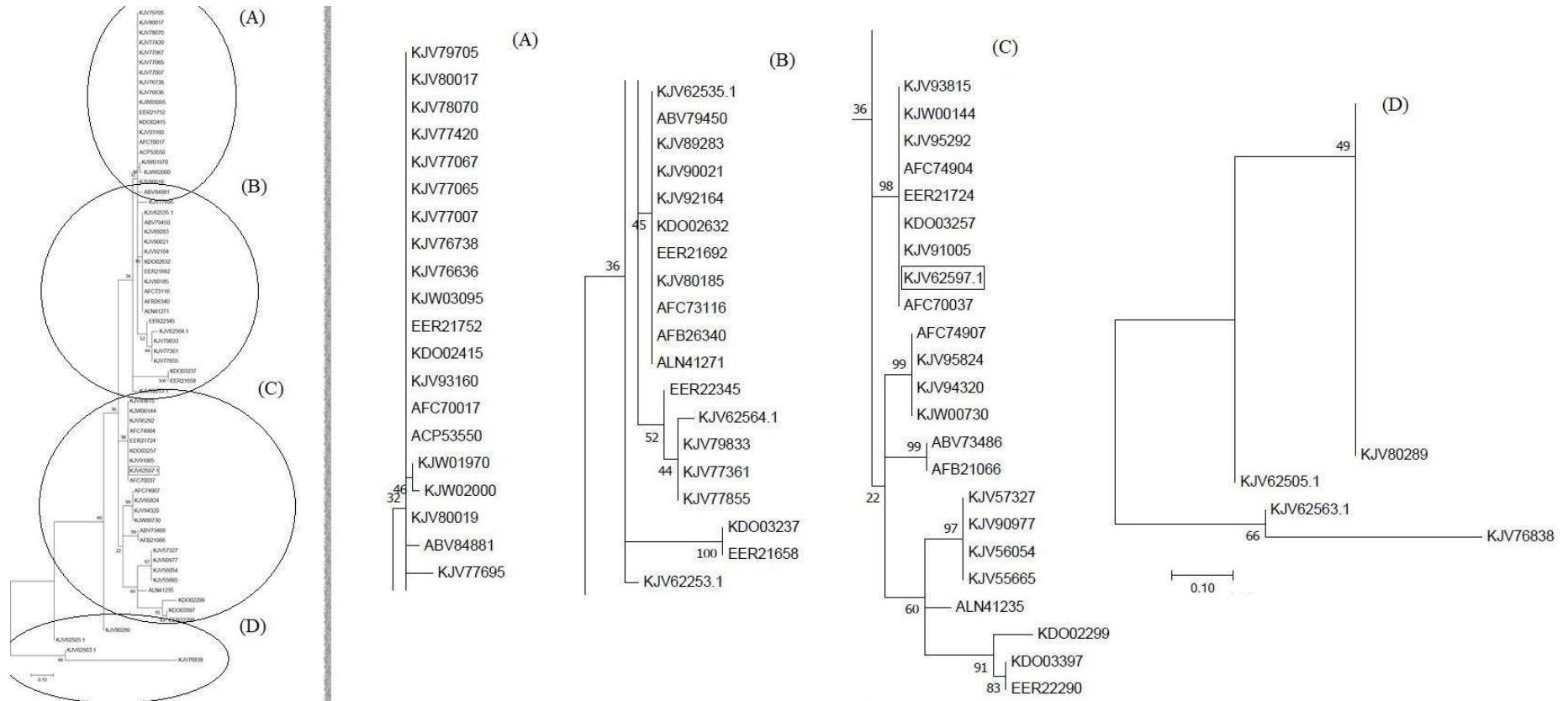
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
KJW01970	ASM96518v1	Conjugative transfer proteinD	<i>R. argasi</i>	T170-B	Argasídeo
ALN41271	ASM144247v1	Conjugugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
AFB26340	ASM28399v1	Type IV secretion-systn	<i>R. philipe</i>	364D	Ixodídeo
AFC73116	ASM28417v1	Conjugative transfer proteinD	<i>R. montanna</i>	OSU85-930	Ixodídeo
ABV84881	ASM1662v1	Conjugative transfer proteinD (Putative)	<i>R. massilea</i>	MTU5	Ixodídeo
KJV80289	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80185	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80019	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80017	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79705	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78070	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77695	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77420	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77067	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77065	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77007	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76738	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76636	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJW03095	ASM96515v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21752	ASM16073v1	Conjugative transfeD (putative)	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21692	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21658	ASM16073v1	DNA transportD	<i>R. endosymbiontis</i>	-	Ixodídeo
KDO03237	REISMNv1	DNA transportD	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02632	REISMNv1	DNA transportD	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02415	REISMNv1	DNA transportD	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02306	REISMNv1	Type IV secretion-systn	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV92164	ASM96504v1	Type IV secretion-systn	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV90021	ASM96500v1	PutativeD	<i>R. belli</i>	RML An4	Ixodídeo
KJV89283	ASM96500v1	Conjugative transfer protD	<i>R. belli</i>	RML An4	Ixodídeo
ACP53550	ASM2300v1	tn	<i>R. africae</i>	ESF-5	Ixodídeo
AFC70017	ASM28405v1	e:protein coding descriptionn	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV62253.1	ASM96467v1	pe:protein coding description	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62505.1	ASM96467v1	Type IV secretn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62535.1	ASM96467v1	tion:Type IV secretion-sn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV93160	ASM96499v1	Putative Conjugative transD	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJW02000	ASM96518v1	Type IV secretory n	<i>R. argasi</i>	T170-B	Argasídeo
ABV79450	ASM1824v1	Putative Conjugative tran(2\)	<i>R. belli</i>	OSU 85-389	Ixodídeo
KJV62563.1	ASM96467v1	protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV76838	ASM96484v1	Type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62564.1	ASM96467v1	protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV79833	ASM96484v1	Type IV secretory syn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
EER22345	ASM16073v1	protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV77361	ASM96484v1	Type IV secretory syn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77855	ASM96484v1	Type IV secretory syn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
AFC70037	ASM28405v1	putative Conjugative transD	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV62597.1	ASM96467v1	protein coding descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV91005	ASM96499v1	Type IV secrn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo

KDO03257	REISMNv1	DNA transpoD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21724	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
AFC74904	ASM28419v1	putative Conjugative transFD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
KJV95292	ASM96507v1	Type IV secretory n	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV93815	ASM96508v1	Type IV secretory systn	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJW00144	ASM96514v1	Type IV secretoryn	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
ABV73486	ASM1434v1	Putative Conjugative transfeD	<i>R. canadensis</i>	McKiel	Ixodídeo
AFB21066	ASM28391v1	putative Conjugative transFD	<i>R. canadensis</i>	CA410	Ixodídeo
AFC74907	ASM28419v1	Conjugative transfer proteiD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
KJV95824	ASM96507v1	Type IV secretory n	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV94320	ASM96508v1	Type IV secretory systn	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJW00730	ASM96514v1	Type IV secretoryn	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJV90976	ASM96499v1	hypothetical prn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KDO02299	REISMNv1	DNA transport prD	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03397	REISMNv1	DNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER22290	ASM16073v1	Conjugative transfer protn	<i>R. endosymbiontis</i>	-	Ixodídeo
ALN41235	ASM144247v1	conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
KJV55665	ASM96458v1	Type IV secretion-sysn	<i>O. tsugamuch</i>	-	Mite
KJV56054	ASM96458v1	Type IV secretory system Cn	<i>O. tsugamuch</i>	-	Mite
KJV57327	ASM96458v1	Type IV secretory sysn	<i>O. tsugamuch</i>	-	Mite
KJV77696	ASM96484v1	Type IV secretion-systn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo

Na Tabela 46, verifica-se a proteína ID KJV62253.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertence.

A proteína ID KJV62597.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 27).

Figura 27 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62597.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 27, verifica-se que dentre as proteínas similares para a proteína ID KJV62597.1, observou-se predominantemente a presença de sequências de proteínas da família responsável pela formação do sistema de secreção do tipo IV (Tabela 47).

Tabela 47 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62597.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências da análise de filogenia.

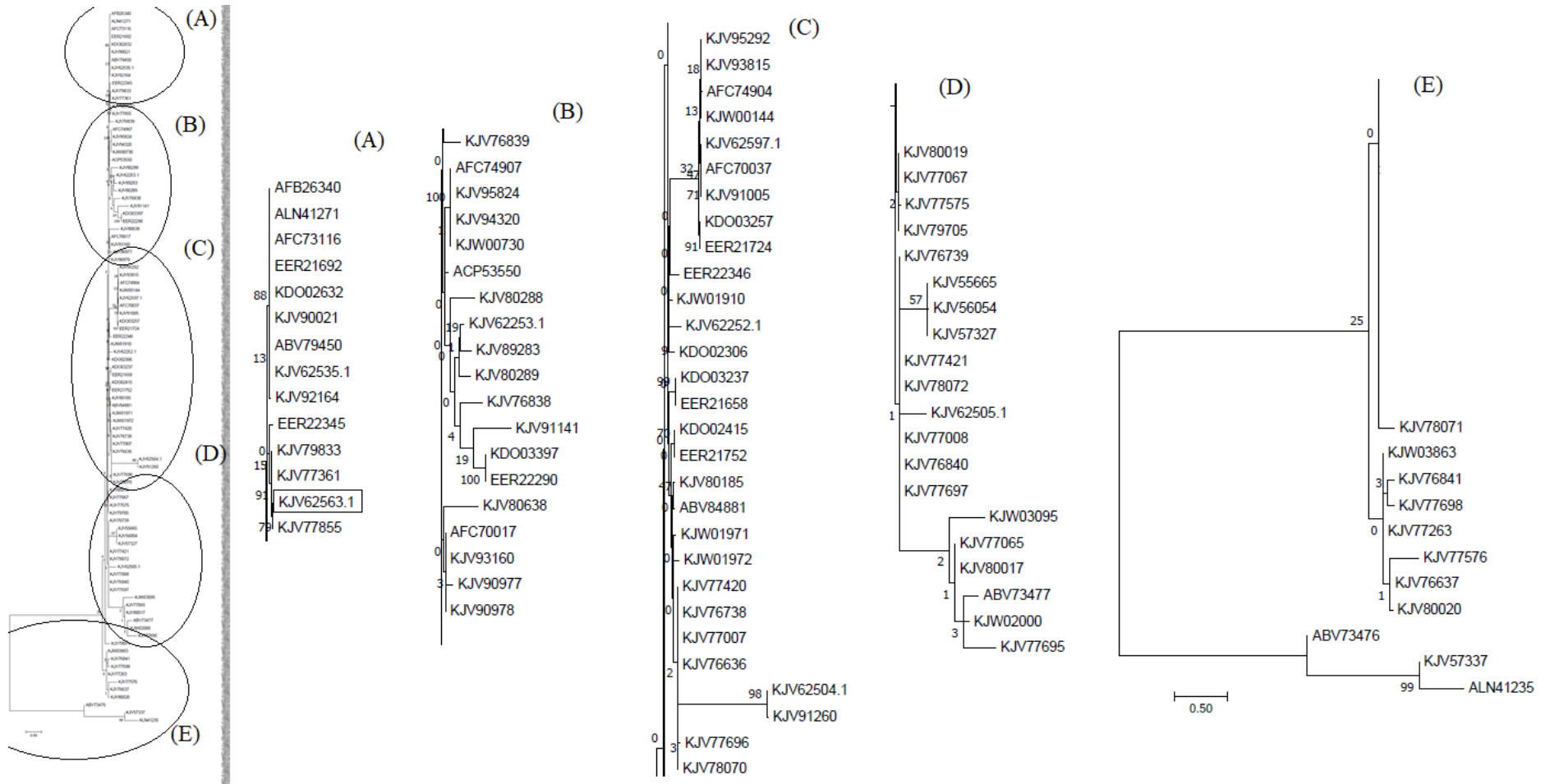
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
KJV79705	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80017	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78070	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77420	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77067	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77065	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77007	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76738	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76636	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJW03095	ASM96515v1	type IV secretion-sysn	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21752	ASM16073v1	conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KDO02415	REISMNv1	DNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV93160	ASM96499v1	type IV secretion-sysn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
AFC70017	ASM28405v1	conjugative transfer protD	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
ACP53550	ASM2300v1	PutativeD	<i>R. africae</i>	ESF-5	Ixodídeo
KJW01970	ASM96518v1	type IV secretion-sysn	<i>R. argasi</i>	T170-B	Argasídeo
KJW02000	ASM96518v1	type IV secretion-sysn	<i>R. argasi</i>	T170-B	Argasídeo
KJV80019	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
ABV84881	ASM1662v1	Putative coD	<i>R. massilea</i>	MTU5	Ixodídeo
KJV77695	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62535.1	ASM96467v1	protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
ABV79450	ASM1824v1	Putative conjugative transfD	<i>R. belli</i>	OSU 85-389	Ixodídeo
KJV89283	ASM96500v1	type IV secretion-sysn	<i>R. belli</i>	RML An4	Ixodídeo
KJV90021	ASM96500v1	type IV secretion-sysn	<i>R. belli</i>	RML An4	Ixodídeo
KJV92164	ASM96504v1	type IV secretion-sysn	<i>R. belli</i>	RML Mogi	Ixodídeo
KDO02632	REISMNv1	DNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21692	ASM16073v1	putative conjugative transfeD	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV80185	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
AFC73116	ASM28417v1	conjugative transfer proteinD	<i>R. montanna</i>	OSU85-930	Ixodídeo
AFB26340	ASM28399v1	conjugative transfer proteinD	<i>R. philipe</i>	364D	Ixodídeo
ALN41271	ASM144247v1	conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
EER22345	ASM16073v1	protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV62564.1	ASM96467v1	protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV79833	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77361	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77855	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KDO03237	REISMNv1	DNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER21658	ASM16073v1	conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV62253.1	ASM96467v1	protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV93815	ASM96508v1	type IV secretion-sysn	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJW00144	ASM96514v1	type IV secretion-sysn	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJV95292	ASM96507v1	type IV secretion-sysn	<i>R. parkeri</i>	AT#24	Ixodídeo
AFC74904	ASM28419v1	putative conjugative transfD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
EER21724	ASM16073v1	conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KDO03257	REISMNv1	DNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV91005	ASM96499v1	type IV secretion-sysn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV62597.1	ASM96467v1	protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
AFC70037	ASM28405v1	putative conjugative transD	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
AFC74907	ASM28419v1	conjugative transfer proteiD	<i>R. parkeri</i>	Portsmouth	Ixodídeo
KJV95824	ASM96507v1	type IV secretion-sysn	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV94320	ASM96508v1	type IV secretion-sysn	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJW00730	ASM96514v1	type IV secretion-sysn	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
ABV73486	ASM1434v1	Putative conjugative transfeD	<i>R. canadenses</i>	McKiel	Ixodídeo
AFB21066	ASM28391v1	putative conjugative transfD	<i>R. canadenses</i>	CA410	Ixodídeo
KJV90977	ASM96499v1	type IV secretion-sysn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
ALN41235	ASM144247v1	conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo

KDO02299	REISMNv1	DNA transport prD	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03397	REISMNv1	DNA transpD	<i>R. buchneri</i>	ISO7	Ixodídeo
EER22290	ASM16073v1	conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV80289	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62505.1	ASM96467v1	protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62563.1	ASM96467v1	protein n	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV76838	ASM96484v1	type IV secretion-sysn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo

Na Tabela 47, verifica-se a proteína ID KJV62597.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertence.

A proteína ID KJV62563.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 28).

Figura 28 Árvore filogenética para a família das sequências de proteínas relacionadas ao sistema de secreção do tipo IV (T4SS) com destaque para a proteína ID KJV62563.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 28 verifica-se que dentre as proteínas similares para a proteína ID KJV62563.1, observou-se predominantemente a presença de sequências de proteínas da família responsável pela formação do sistema de secreção do tipo IV (Tabela 48).

Tabela 48 Quadro de legendas para a árvore filogenética da família de sequências de proteínas do sistema de secreção T4SS com destaque para a proteína ID KJV62563.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
ABV73476	ASM1434v1	Putative conjugative transfeD	<i>R. canadensis</i>	McKiel	Ixodídeo
ABV73477	ASM1434v1	Putative conjugative transfeD	<i>R. canadensis</i>	McKiel	Ixodídeo
ABV79450	ASM1824v1	Putative conjugative transfeD	<i>R. belli</i>	OSU 85-389	Ixodídeo
ABV79450	ASM1824v1	Putative conjugative transfeD	<i>R. belli</i>	OSU 85-389	Ixodídeo
ABV84881	ASM1662v1	F Putative coD	<i>R. massilea</i>	MTU5	Ixodídeo
ACP53550	ASM2300v1	F Putative coD	<i>R. africae</i>	ESF-5	Ixodídeo
AFB26340	ASM28399v1	Conjugative transfer protein D	<i>R. philipe</i>	364D	Ixodídeo
AFC70017	ASM28405v1	Conjugative transfer protein D	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
AFC70037	ASM28405v1	Putative conjugative transfer protein D	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
AFC73116	ASM28417v1	Conjugative transfer protein D	<i>R. montanna</i>	OSU85-930	Ixodídeo
AFC74904	ASM28419v1	Putative conjugative transfer protein D	<i>R. parkeri</i>	Portsmouth	Ixodídeo
AFC74907	ASM28419v1	Conjugative transfer protein D	<i>R. parkeri</i>	Portsmouth	Ixodídeo
ALN41235	ASM144247v1	Conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
ALN41271	ASM144247v1	Conjugal transfer protein	<i>R. rhipicephali</i>	-	Ixodídeo
EER21658	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21692	ASM16073v1	Putative conjugative transfer protein D	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21724	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER21752	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER22290	ASM16073v1	Conjugative transfer protein	<i>R. endosymbiontis</i>	-	Ixodídeo
EER22345	ASM16073v1	Protein TraD	<i>R. endosymbiontis</i>	-	Ixodídeo
EER22346	ASM16073v1	Conjugative transfer protein N	<i>R. endosymbiontis</i>	-	Ixodídeo
KDO02306	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02415	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO02632	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03237	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03257	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KDO03397	REISMNv1	DNA transport D	<i>R. buchneri</i>	ISO7	Ixodídeo
KJV55665	ASM96458v1	Type IV secretory system n	<i>O. tsugamuch</i>	-	Mite
KJV56054	ASM96458v1	Type IV secretory system Cn	<i>O. tsugamuch</i>	-	Mite
KJV57327	ASM96458v1	Type IV secretory system n	<i>O. tsugamuch</i>	-	Mite
KJV57337	ASM96458v1	Type II/IV secretory system n	<i>O. tsugamuch</i>	-	Mite
KJV62252.1	ASM96467v1	tn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62253.1	ASM96467v1	tn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62504.1	ASM96467v1	Descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62505.1	ASM96467v1	Descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62535.1	ASM96467v1	Description	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62563.1	ASM96467v1	Descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV62597.1	ASM96467v1	Descriptionn	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV76636	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76637	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76738	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76739	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76838	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76839	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76840	ASM96484v1	ftsK/SpoIIIE family pn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76841	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77007	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77008	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77065	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77067	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo

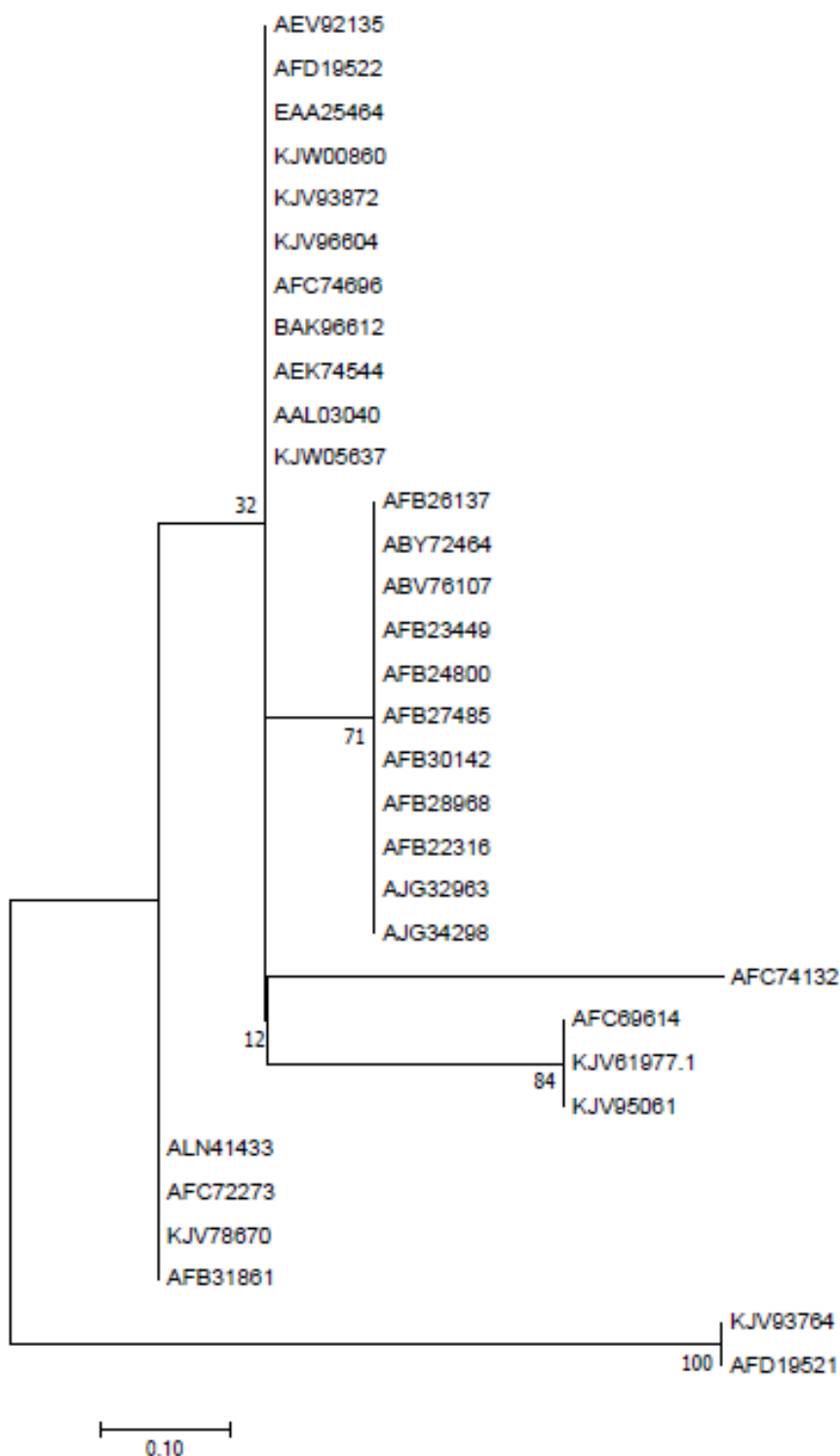
KJV77263	ASM96484v1	Putative conjugative transfer n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77361	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77420	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77421	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77575	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77576	ASM96484v1	Putative conjugative TrD	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77695	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77696	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77697	ASM96484v1	ftsK/SpoIIIE family pn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77698	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV77855	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78070	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78071	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV78072	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79705	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV79833	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80017	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80019	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80020	ASM96484v1	F sex factor N termin	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80185	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80288	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80289	ASM96484v1	Type IV secretory system n	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80638	ASM96484v1	F sex factor N terminn	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV89283	ASM96500v1	Type IV secretory system n	<i>R. belli</i>	RML An4	Ixodídeo
KJV90021	ASM96500v1	Type IV secretory system n	<i>R. belli</i>	RML An4	Ixodídeo
KJV90977	ASM96499v1	Type IV secretory system n	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV90978	ASM96499v1	Type II/IV secretory system n	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV91005	ASM96499v1	Type IV secretory system n	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV91141	ASM96504v1	Type IV secretory system n	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV91260	ASM96499v1	F sex factor N terminn	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV92164	ASM96504v1	Type IV secretory system n	<i>R. belli</i>	RML Mogi	Ixodídeo
KJV93160	ASM96499v1	Type IV secretory system n	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV93815	ASM96508v1	Type IV secretory system n	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJV94320	ASM96508v1	Type IV secretory system n	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJV95292	ASM96507v1	Type IV secretory system n	<i>R. parkeri</i>	AT#24	Ixodídeo
KJV95824	ASM96507v1	Type IV secretory system n	<i>R. parkeri</i>	AT#24	Ixodídeo
KJW00144	ASM96514v1	Type IV secretory system n	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJW00730	ASM96514v1	Type IV secretory system n	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJW01910	ASM96518v1	F sex factor N terminan	<i>R. argasi</i>	T170-B	Argasídeo
KJW01971	ASM96518v1	Type IV secretory system n	<i>R. argasi</i>	T170-B	Argasídeo
KJW01972	ASM96518v1	Type II/IV secretory system n	<i>R. argasi</i>	T170-B	Argasídeo
KJW02000	ASM96518v1	Type IV secretory system n	<i>R. argasi</i>	T170-B	Argasídeo
KJW03095	ASM96515v1	Type IV secretory system n	<i>R. endosymbiontis</i>	-	Ixodídeo
KJW03863	ASM96518v1	Type II/IV secretory system n	<i>R. argasi</i>	T170-B	Argasídeo

Na Tabela 48, verifica-se a proteína ID KJV62563.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertence.

A proteína ID KJV62594.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 29).

Figura 29 Árvore filogenética para a família das sequências de proteínas contendo domínios repetitivos de anquirina (Ank) com destaque para a proteína ID KJV61977.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os

seus similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 29, verifica-se que dentre as proteínas similares para a proteína ID KJV61977.1, observou-se predominantemente sequências de proteínas da família contendo domínios de repetição de anquirina (Ank), havendo muitas sequências hipotéticas (Tabela 49).

Tabela 49 Quadro de legendas para a árvore filogenética da família de sequências de proteínas contendo domínios repetitivos de anquirina (Ank) com destaque para a proteína ID KJV61977.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

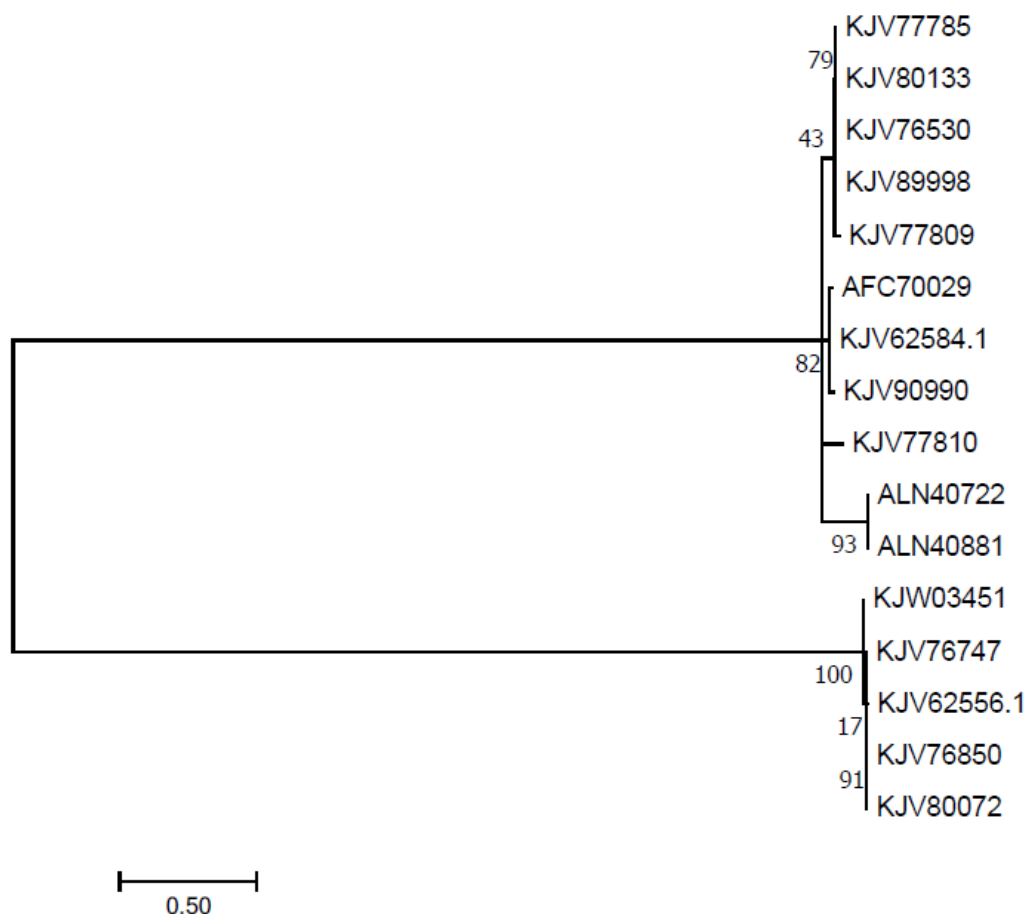
Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estirpe	Hospedeiro
AEV92135	ASM23784v1	Ankyrin repeats Family	<i>R. slova</i>	13B	Ixodídeo
AFD19522	ASM25236v1	Ankyrin repeats Family	<i>R. slova</i>	D-CWPP	Ixodídeo
EAA25464	ASM16693v1	Hypothetical protein	<i>R. sibirica</i>	subsp. sibirica 246	Ixodídeo
KJW00860	ASM96514v1	Ankyrin repeats family	<i>R. parkeri</i>	Tate's Hell	Ixodídeo
KJV93872	ASM96508v1	Ankyrin repeats family	<i>R. parkeri</i>	Grand Bay	Ixodídeo
KJV96604	ASM96507v1	Ankyrin repeats family	<i>R. parkeri</i>	AT#24	Ixodídeo
AFC74696	ASM28419v1	Hypothetical protein	<i>R. parkeri</i>	Portsmouth	Ixodídeo
BAK96612	ASM28359v1	Hypothetical protein	<i>R. japonica</i>	YH	Ixodídeo
AEK74544	ASM22120v1	Hypothetical protein	<i>R. heilongjiangensis</i>	54	Ixodídeo
AAL03040	ASM702v1	Unkown	<i>R. conorii</i>	subsp. conorii Malish 7	Ixodídeo
KJW05637	ASM96518v1	Ankyrin repeats family protein	<i>R. argasi</i>	T170-B	Argasídeo
AFB26137	ASM28399v1	Hypothetical protein	<i>R. philipe</i>	364D	Ixodídeo
ABY72464	ASM1744v3	Hypothetical protein	<i>R. rickettsii</i>	Iowa	Ixodídeo
ABV76107	ASM1822v1	Hypothetical protein	<i>R. rickettsii</i>	Sheila Smith	Ixodídeo
AFB23449	ASM28377v1	Hypothetical protein	<i>R. rickettsii</i>	Colombia	Ixodídeo
AFB24800	ASM28379v1	Hypothetical protein	<i>R. rickettsii</i>	Arizona	Ixodídeo
AFB27485	ASM28381v1	Hypothetical protein	<i>R. rickettsii</i>	Hino	Ixodídeo
AFB30142	ASM28383v1	Hypothetical protein	<i>R. rickettsii</i>	Hauke	Ixodídeo
AFB28968	ASM28393v1	Hypothetical protein	<i>R. rickettsii</i>	Hep#2	Ixodídeo
AFB22316	ASM28395v1	Hypothetical protein	<i>R. rickettsii</i>	Brazil str Brazil (RPN)	Ixodídeo
AJG32963	ASM83152v1	Ankyrin repeats family protein	<i>R. rickettsii</i>	R	Ixodídeo
AJG34298	ASM83154v1	Ankyrin repeats family protein	<i>R. rickettsii</i>	Morgan	Ixodídeo
AFC74132	ASM28417v1	Hypothetical protein	<i>R. montanna</i>	OSU85-930	Ixodídeo
KJV61977.1	ASM96467v1	Ankyrin repeats family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
AFC69614	ASM28405v1	Hypothetical protein	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV95061	ASM96499v1	Ankyrin repeats family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
AFB31861	ASM28385v1	Hypothetical protein	<i>R. massilea</i>	AZT80	Ixodídeo
ALN41433	ASM144247v1	Hypothetical protein	<i>R. rhipicephali</i>	-	Ixodídeo
AFC72273	ASM28407v1	Hypothetical protein	<i>R. rhipicephali</i>	3-7-female 6-CWPP	Ixodídeo
KJV78670	ASM96490v1	Ankyrin repeats family protein	<i>R. rhipicephali</i>	Ect	Ixodídeo
KJV93764	ASM96508v1	Hypothetical protein	<i>R. parkeri</i>	Grand Bay	Ixodídeo
AFD19521	ASM25236v1	Hypothetical protein	<i>R. slova</i>	D-CWPP	Ixodídeo

Na Tabela 49, verifica-se a proteína Ank ID KJV61977.1 e suas sequências de proteínas similares, os IDs dos genomas de referência das bactérias, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas sequências pertencem.

As proteínas L11 KJV62584.1 e KJV62556.1 apresentaram uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 30).

Figura 30 Árvore filogenética para a família das sequências de proteínas ribossomais metiltransferase (L11) com destaque para a proteína ID KJV62584.1 e KJV62665.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus

similares obtidos com o uso do BLASTP e da análise de filogenia por meio do aplicativo Mega, adotando da Máxima verossimilhança e alinhamentos ClustalW para 100 topologias de *Bootstrap*.



Fonte: Autor deste trabalho.

Na Figura 30, verifica-se que dentre as proteínas similares para as proteínas ID KJV62584.1 e KJV62556.1, observou-se predominantemente sequências de proteínas da família ribossomal L11 metiltransferase (Tabela 50).

Tabela 50 Quadro de legendas para a árvore filogenética da família de sequências de proteínas ribossomais metiltransferase (L11) com destaque para a proteína ID KJV62584.1 e KJV62556.1 pertencente a espécie *R. amblyommatis* Ac/Pa e os seus similares obtidos por meio de multialinhamentos de sequências e da análise de filogenia.

Proteína	Genoma	Nome	<i>Rickettsia</i> spp.	Estípe	Hospedeiro
KJV77785	ASM96484v1	Ribosomal L11 methyltransferase family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80133	ASM96484v1	Ribosomal L11 methyltransferase family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV76530	ASM96484v1	Ribosomal L11 methyltransferase family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV89998	ASM96500v1	Ribosomal L11 methyltransferase family protein	<i>R. belli</i>	RML An4	Ixodídeo
KJV77809	ASM96484v1	HemK domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
AFC70029	ASM28405v1	HemK	<i>R. amblyommatis</i>	GAT-30V	Ixodídeo
KJV62584.1	ASM96467v1	Ribosomal L11 methyltransferase family protein	<i>R. amblyommatis</i>	Ac/Pa	Ixodídeo
KJV90990	ASM96499v1	Ribosomal L11 methyltransferase family protein	<i>R. amblyommatis</i>	Dark Water	Ixodídeo
KJV77810	ASM96484v1	HemK domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
ALN40722	ASM144247v1	Modification methylase Hemk	<i>R. rhipicephali</i>	-	Ixodídeo
ALN40881	ASM144247v1	Modification methylase Hemk	<i>R. rhipicephali</i>	-	Ixodídeo

KJW03451	ASM96515v1	Ribosomal L11 methyltransferase family protein	<i>R. endosymbiontis</i>	-	Ixodídeo
KJV76747	ASM96484v1	HemK domain protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV62556.1	ASM96467v1	Ribosomal L11 methyltransferase family protein	<i>R. amblyommatidis</i>	Ac/Pa	Ixodídeo
KJV76850	ASM96484v1	Ribosomal L11 methyltransferase family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo
KJV80072	ASM96484v1	Ribosomal L11 methyltransferase family protein	<i>R. hoogstraalii</i>	RCCE3	Argasídeo

Na Tabela 50, verifica-se as proteínas IDs KJV62584.1 e KJV62556.1 com suas sequências de proteínas similares, os IDs dos genomas de referência, os nomes das proteínas, as espécies de *Rickettsia* e estirpes, das quais estas pertence.

5 DISCUSSÃO

5.1 Interatomas

Os resultados da análise de anotação das proteínas de interface foram obtidos por meio da construção de modelos de interação entre as proteínas dos genomas adotados. O uso das proteínas da bactéria *R. amblyommatidis* versus a espécie de carrapato *I. scapularis*, sugeriu participação de famílias de proteínas, sabidamente envolvidas na modulação da fisiologia da célula do hospedeiro. Estes resultados sugeriram a possibilidade da ocorrência de 13,151 interações interespecíficas entre as proteínas destes organismos no interatoma geral (Tabela 16) (Dados suplementares S1e). Com o uso das proteínas de membrana preditas foi possível separar dentro da rede consenso as proteínas secretadas, excretadas e ou de superfície, sendo esta estratégia base para a construção do interatoma de proteínas de membrana, totalizando uma rede consenso com 3,456 interações (Tabela 23) (Dados suplementares S1g).

Com as análises de predição de ortologia filtramos dentro da rede de proteínas de membrana algumas proteínas ortólogas entre as espécies *R. amblyommatidis*, *R. buchenneri*, *R. endossimbionte*, *R. monascencis*, *R. felis* e *R. prowazekii*. À partir destes resultados, esperávamos encontrar entre as proteínas ortólogas módulos e, ou, funções conservadas com a possibilidade da ocorrência de divergências entre as espécies de *Rickettsia* que infectam carrapatos em relação as que infectam insetos. No geral, estes resultados revelaram a ocorrência de duas proteínas ortólogas, apresentando 100 pares de interação entre *R. amblyommatidis* e *I. scapularis* na rede consenso formada (Tabela 25) (Dados suplementares S1h).

Por meio de alinhamentos múltiplos de sequências pudemos avaliar os aspectos evolutivos das sequências ortólogas de genicas e proteicas para as espécies de *Rickettsia* que infectam os carrapatos (Anexo 1). Os alinhamentos obtidos sugeriram os domínios conservados, incluindo resíduos críticos para a função das proteínas, comparando as proteínas de interesse de forma mais detalhada com os membros da mesma família.

Os resultados das arvores filogenéticas de ortólogos para as proteínas similares do cluster ortólogo 1050, com destaque para as duas proteínas KJV62578.1 e KJV62261.1 *His kinase A domain protein*, e no cluster ortólogo 1816 com destaque para a proteína KJV62536.1 *Tetratricopeptide repeat protein family*, presentes no interatoma de proteínas de membrana para a espécie *R. amblyommatidis* que infectam os carrapatos versus insetos, sugeriram diferenças entre as espécies comparadas (Anexo 2).

A ontologia gênica para o estudo de módulos funcionais ativados pela interação, possibilitou observarmos os aspectos deste (patossistema) entre as proteínas da bactéria da espécie *R. amblyommatis* e do carrapato da espécie *I. scapularis*, com a visualização das redes do interatoma de proteínas geral, o interatoma consenso de proteínas de membrana e o interatoma de proteínas ortólogas (Figura 16). A análise de anotação das proteínas de interface permitiu identificar famílias de proteínas sabidamente envolvidas na modulação da fisiologia do *I. scapularis* para os três domínios: Componente celular (C), Função molecular (F) e Processo biológico (P) (Tabela 28).

Estes resultados sugerem a presença de diferentes módulos funcionais significativos com destaque para os módulos 0 secreção celular, 4 metilação, 13 proteólise, 14 oxidação e 15 fosforilação obtidos por meio do interatoma de proteínas de membrana com destaque para o módulo 1 transporte intracelular obtido por meio do interatoma de proteínas de membrana ortólogas de *Rickettsia* spp. que infectam os carrapatos. Este enriquecimento funcional realizado com o uso do BINGO, pode importar o mapa da anotação funcional das proteínas do *I. scapularis* do Gene Ontology (GO), enriquecido com o aplicativo AgBase-GOanna permitindo atribuímos o processo biológico e a anotação de ontologia gênica mais significativo para cada módulo funcional de proteínas preditas carrapatos (Dados suplementares S1j) (Figura 17) (Tabela 29).

Diversos estudos têm relatado o funcionamento do secretoma de *Rickettsia* e como ocorre a possível interação entre este repertório de proteínas e vias sustentadas pela célula e proteínas do hospedeiro (GILLESPIE et al., 2014). De certa forma, é necessário estudar a essencialidade destes alvos encontrados. No geral, diversos estudos adotando a biologia de sistemas e ferramentas de bioinformática ao abordarem os conceitos do interatoma de proteínas sugerem a utilização de evidências experimentais que confirmam a essencialidade de alguns dos alvos preditos (FLÓREZ et al., 2010; KIM et al., 2008; MENDES et al., 2013).

De forma indireta ou não, a validação destes alvos essenciais na rede de proteínas tem sido um passo importante após a descoberta de vias moleculares funcionais. Podem ser citados como exemplos estudos quanto a interação entre proteínas da bactéria *Xanthomonas oryzae* versus *Oryza sativa* (KIM et al., 2008), para patógenos que afetam humanos como a bactéria *Mycobacterium tuberculosis* versus *Homo sapiens* (RAMAN; YETURU; CHANDRA, 2008) e também para o protozoário *Leishmania major* versus *H. sapiens* (FLÓREZ et al., 2010; MENDES et al., 2013).

Potencialmente estudos do interatoma podem ser um ponto de partida para futuras redes experimentais de interação de proteínas na relação parasito hospedeiro de patógenos e carrapatos. Mesmo considerando o fato que muitas interações podem exigir modificações pós-traducionais, dificultando assim, a validação experimental na análise de enriquecimento de ontologia gênica e predição de função para os módulos clusterizados, estudos já evidenciaram que a detecção de estruturas modulares dentro de uma rede biológica fornece *insights* sobre a organização funcional de processos celulares em organismos vivos (RIVES; GALITSKI, 2003).

A detecção de cluster combinada com a análise de enriquecimento hipergeométrico permite prever a função biológica de proteínas a esta associadas, onde estudos tem atribuído o compartilhamento de pelo menos uma função similar entre 70 a 80% das proteínas em um cluster (SHARAN; ULITSKY; SHAMIR, 2007). Neste estudo foi possível atribuir que as proteínas vizinhas mais próximas presentes nas redes de interação de proteínas também estão frequentemente envolvidas em processos similares. Isto, implica que qualquer proteína não classificada poderia ter a função provisoriamente atribuída à função de seus vizinhos (SHARAN; ULITSKY; SHAMIR, 2007; TITZ; SCHLESNER; UETZ, 2004).

Os resultados de predição de proteínas de membrana secretadas, excretadas e, ou, de superfície, quando concatenadas e filtradas na interface do interatoma geral possibilitaram averiguarmos a ocorrência de interações entre as proteínas da membrana de *R. amblyommatis* versus as proteínas de *I. scapularis*. A rede consenso obtida à partir do interatoma de proteínas de membrana sugeriu haver a ocorrência de interação entre diversas proteínas supostamente relacionadas com as que compõem os sistemas de secreção em *Rickettsia*, entre outras funcionalidades preditas, baseando no mapa da topologia do sistema clusterizados.

O enriquecimento funcional para o módulo de secreção apresentou a maior abundância de interações e riqueza de proteínas em comparação com os demais módulos preditos. A rede do interatoma de proteínas de membrana sugeriu a ocorrência de interações entre as proteínas com funções essenciais e fortes moduladores da célula do carrapato. Os resultados de interação entre as proteínas de *R. amblyommatis* versus *I. scapularis* demonstraram a participação de diversas proteínas essenciais que fazem parte de dois sistemas de secreção dependentes do sistema Sec translocom.

As bactérias do gênero *Rickettsia* possuem o potencial de codificarem vários sistemas de secreção homólogos às vias de secreção de outras bactérias que apesar do reduzido tamanho de seus genomas, conseqüentemente passam a se utilizar de metabólitos processados no ambiente celular de seus hospedeiros com a participação de inúmeros substratos (GILLESPIE et al., 2014). Segundo estes autores, os genomas bacterianos, possuem diversos complexos de secreção de proteínas, onde a proteína Ftsy possui como função o reconhecimento de partículas de reconhecimento de sinal atuando na regulação da tradução ribossomal na região intermembrar do sistema de secreção tipo Sec-translocom.

A proteína Ftsy é responsável pela ancoragem do complexo ribossômico ao complexo SecYEG na via co-translacional de Sec-translocon (EGEA et al., 2004), atuando na hidrólise de GTP (FOCIA, 2004) para impulsionar o ribossomo a passar para fora do espaço intermembranar (AKOPIAN et al., 2013). O centro de translação de proteínas no sistema Sec-translocon é mediado por uma insertase YidC, além de outras proteínas de membrana localizadas na região intermembrana e independentes do canal dividido em duas vias, sendo a via co-translacional e a via pós-translacional (GILLESPIE et al., 2014). Segundo estes autores, ambas as vias são dependentes ou independentes de partículas de reconhecimento de sinal e de fator desencadeante, o que sugere a ocorrência de uma ligação competitiva na cadeia nascente do ribossomo.

No gênero *Rickettsia*, o sistema de secreção do tipo V (T5SS) que é regulado por fator desencadeante na via co-translacional, é composto por uma superfamília de antígenos celulares de superfície conservados na maioria dos genomas, sendo Sca0 (anteriormente RompA), Sca1, Sca2, Sca4 (anteriormente PS 120 ou proteína D) (GILLESPIE et al., 2012) e Sca5 (anteriormente RompB) (SEARS et al., 2012). No entanto, as proteínas Sca3 são codificados apenas dentro dos genomas de espécies de *Rickettsia* do Grupo do Tifo e para *R. felis*, sendo associadas exclusivamente com os insetos vetores, excluindo as espécies de *Rickettsia* que infectam os carrapatos e outros ácaros, atuando ainda na regulação da atividade de YidC (GILLESPIE et al., 2014).

O *sec translocom* possui também o sistema Sec-TolC apresentando a proteína RARP-1, independente de ATP para promover a translocação de moléculas pelo canal TolC (KAUR et al., 2012). Segundo estes autores, esta proteína é dependente de domínios de anquirina Ank sendo altamente conservadas em *Rickettsia*, sendo relacionada com uma rota alternativa para RARP-1, a qual necessitaria da presença de ABC transporte do tipo I. Neste processo, a

hidrólise de ATP no citoplasma bacteriano forneceria a energia para empurrar RARP-1 através do canal TolC (GILLESPIE et al., 2014). Segundo estes autores o estudo desta via também pode auxiliar na elucidação da presença de interação de proteínas que compõem os sistemas de secreção independentes de *Sec translocom*. Para o Tat, os genomas de *Rickettsia* podem conter genes que codificam TatA e TatC, ocorrendo na sequência da subunidade ferro-enxofre do complexo de proteínas *ubiquinol-citocromo c redutase* ou no complexo *citocromo b-c1*, conhecida como PetA (GILLESPIE et al., 2014). Apesar da presença de sistema Tat, estudos de bioinformática identificaram poucos substratos de Tat em genomas de *Rickettsia* (NUÑEZ; SORIA; FARBER, 2012).

A interação de proteínas da bactéria versus o carrapato com presença das proteínas ID KJV61777.1 e ID KJV61783.1 estão estritamente relacionadas com o ubiquinol-citocromo C redutase além de transportadores de arginina ID KJV61302.1. No entanto, sabe-se que nos genomas de *Rickettsia* são codificados componentes do sistema de secreção tipo T1SS, sendo este, outro sistema de secreção capaz de fornecer a energia mais provável para translocar RARP-1 através da membrana externa. Neste estudo, observou-se que a proteína ID KJV61887.1 relacionada ao sistema T1SS apresentou pares de interação com 65 proteínas do carrapato. No T1SS, a máquina de secreção é responsável por exportar proteínas do citoplasma para o meio extracelular associadas à virulência como toxinas, lipases, proteases (ZGURSKAYA et al., 2011), fornecendo a energia promovida por proteínas ABC transporte com a hidrólise de ATP (HOLLAND; SCHMITT; YOUNG, 2005).

Além destes sistemas, o sistema de secreção do tipo IV (T4SS) formado por complexos macromoleculares podendo secretar proteínas, DNA e nucleoproteínas através do envelope celular bacteriano, são atualmente um dos principais fatores da diversificação bacteriana e responsáveis pela propagação horizontal de resistência antimicrobiana de genes de virulência envolvidos com moléculas sinalizadoras e reguladoras para sua sobrevivência (CASCALES; CHRISTIE, 2003). Diversos sistemas de secreção T4SS contêm uma ATPase ligada à membrana (famílias VirD4 ou TcpA) denominada proteína de acoplamento do tipo IV, fornecendo energia para processos como a secreção de substratos, regulação de canal e deslocamento de pilina da região intermembranar (KERR; CHRISTIE, 2010).

Neste estudo, observou-se a ocorrência de interação para as proteínas essenciais KJV62535.1, KJV62505.1, KJV62563.1, KJV62597.1, KJV62252.1 relacionadas ao sistema de secreção T4SS apresentou pares de interação com 65 proteínas do carrapato. O T4SS

compartilham genes sintéticos entre *Rickettsia* e de *Orientia*, sendo o complexo denominado RAGE (*Rickettsiales Amplified Genetic Element*) composto de genes deste sistema, bem como de genes que codificam histidina quinases Hk contendo domínios de repetições de anquirina Ank e de tetratricopeptídeo Tpr (CHO et al., 2007).

Neste estudo, observou-se ocorrência de interação destas proteínas relacionadas com proteínas His quinases, como a Ank ID KJV61717.1 e Tpr ID KJV62536.1 e KJV62591.1. O sequenciamento do genoma de *R. endosymbiotic I. scapularis* (REIS) revelou haver nove RAGES completas (ou quase completas), em oposição a um único RAGES presente em *R. bellii* (GILLESPIE et al., 2012). Segundo estes autores, sabe-se que dois dos quatro plasmídeos transportados por REIS (pREIS1 e pREIS3) foram capazes de codificar RAGES, com uma estimativa filogenética indicando múltiplas invasões destes elementos gênicos dentro dos genomas de *Rickettsiales* ao longo da evolução.

Além de genes sintéticos do T4SS, os RAGES de *Rickettsia* e de *Orientia* compartilham genes comuns que codificam DNA metiltransferases (classe D12), helicases de DNA tipo PolC, hidrolases e sintetases de resposta restrita, além de proteínas contendo os domínios His e Tpr (GILLESPIE et al., 2014). Estas duas proteínas estão envolvidos em uma série de funções, incluindo a atividade de acoplamento, o ciclo celular, transcrição, regulação gênica, transdução de sinal e transporte de proteínas (BLATCH; LÄSSLE, 1999; LI; MAHAJAN; TSAI, 2006).

De certa forma, várias destas proteínas estão envolvidas em processos acoplados ao movimento destas membranas para a secreção de proteínas. Entre os módulos de oxidação, fosforilação, metilação, proteólise e especialmente o transporte pudemos anotar distintas proteínas que fazem parte destes sistemas, associadas com diversos pares de interação na rede do interatoma.

5.2 Expressão diferencial de transcritos

Os resultados para a anotação da transcrição de transcritos e proteínas putativas sugeriu que as espécies de *Rickettsia*., como muitos outros microrganismos endossimbiontes, podem desempenhar um papel benéfico ou não na fisiologia do carrapato hospedeiro. A estrutura da rede consenso de proteínas, que os conectam em vias mediadas por interações e mecanismos

moleculares, pode ocorrer através de genes e proteínas potencialmente codificadas nos genomas de ambos os organismos.

Os resultados obtidos para os distintos módulos funcionais putativos em função das interações de proteínas entre ambos os organismos interagidos, traz as principais proteínas do carrapato presentes nestes seis módulos significativos (Dados suplementares S1j). Estas, foram a base para o estudo da expressão e regulação de transcritos relacionados aos módulos significativos de interação com as proteínas de membrana da bactéria da espécie *R. amblyommatis*.

A obtenção destes transcritos para os módulos funcionais significativos por análise de similaridade adotando o multialinhamento de sequências TBLASTn, se deu após as análises de qualidade e trimagem do sequenciamento, a montagem “*de novo*” e cálculo dos níveis de abundância, (Tabela 33) (Dados suplementares S1n). Antes disso, a integração dos dados de sequência de RNA dos transcriptomas montados (Dados não publicados) e os resultados da TPM dos transcritos (Dados suplementares S1m-S5m) com as proteínas do interatoma de membrana (Dados suplementares S1j), em especial as proteínas do carrapato que clusterizam na formação de módulos funcionais e significativos sugerem a possibilidade de ocorrer uma resposta imune do carrapato contribuindo para o desenvolvimento de estratégias, e interferir na persistência e na transmissão de espécies destas bactérias, (Tabela 34) (Dados suplementares S1o).

Estes resultados para a anotação e expressão diferencial para o conjunto de dados de RNA-seq abordando a montagem “*de novo*” para cada espécie de carrapato *Amblyomma* spp. e em determinado órgão infectado ou não por espécies de *Rickettsia* com foco neste estudo em glândulas salivares, intestino e ovários (Tabela 35). As diferenças de expressão e na abundância dos transcritos para o conjunto de dados dos transcriptomas da espécie *A. sculptum* em resposta a infecção por *R. amblyommatis* no intestino e ovário, *A. sculptum* em resposta a infecção por *R. rickettsii* no intestino e *A. aureolatum* em resposta a infecção por *R. rickettsii* em glândulas salivares e intestino são inerentes aos órgãos.

Estes resultados revelaram que a capacidade de espécies de *Rickettsia* na atividade desta expressão de genes pode estar correlacionada com o potencial de proteínas expressas nos seus genomas pertencentes a membrana, sendo estas secretadas, excretadas e, ou, de superfície, que ativariam as proteínas do carrapato presentes nos módulos funcionais significativos. Na

anotação da expressão de transcritos coexpressos, foram selecionados e comparados os dados da média calculada dos valores de *fold change* na base Log₂ para o intestino de carrapato das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*, e a espécie *A. sculptum* infectado por *R. amblyommatis*.

Os transcritos obtidos apresentaram a regulação positiva com os valores de *fold change* na base Log₂ ≥ 2 e com regulação negativa da expressão com valores de *fold change* na base Log₂ ≤ -2 . Os resultados para as glândulas salivares e ovários não foram discutidos (Dados não publicados).

Para esta discussão, consideramos os resultados da expressão de transcritos para as sequências de proteínas principalmente coexpressas para o intestino das espécies de *Amblyomma* infectados por *Rickettsia* spp., (Anexo 3) e (Anexo 4). Com isto, averiguamos as similaridades e diferenças pontuais da expressão diferencial de transcritos entre as espécies de carrapatos no conjunto de proteínas preditas para cada módulo funcional significativo do interatoma de membrana, podendo haver diferenças entre as espécies de *Amblyomma* conforme a espécie de *Rickettsia* que o infecta, (Tabelas 36 e 37).

5.2.1 Secreção e transporte

GTPases e proteínas putativamente relacionadas com adesão, invasão, fagocitose e movimento intracelular e intercelular na célula do hospedeiro.

No intestino do carrapato da espécie *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para a sequência da proteína GTPase Rho sendo regulada de forma positiva, havendo também, a sequência de actina acetiltransferase subunidade A (NAT1) e da GTPase Rho Cdc42 sendo reguladas de forma positivas. No intestino do carrapato da espécie *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para a sequência das proteínas GTPases Rho e Cdc42 sendo reguladas de forma negativa. No intestino do carrapato da espécie *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para a sequência da proteína GTPase Rho sendo regulada de forma negativa.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, não se observou a ocorrência da coexpressão de transcritos para a sequência da proteína GTPases Rho.

No entanto, entre os intestinos do carrapato *A. sculptum* infectados por *R. rickettsii* versus *R. amblyommatis*, observou-se a ocorrência da coexpressão para uma sequência de proteína GTPase, sendo regulada de forma negativa na presença de *R. amblyommatis* e de forma positiva na presença de *R. rickettsii*.

A superfamília de proteínas GTPase Ras compreende cinco subfamílias estruturalmente distintas que incluem Ras, Rho, Rab, Sar1 Arf e Ran. Estas são responsáveis por distintas funções na regulação de uma variedade de processos celulares, tais como; proliferação celular, diferenciação celular, organização do citoesqueleto, transporte e tráfico de proteínas.

O processo de fagocitose é evolutivamente conservado em artrópodes (SIDERI et al., 2008) e desempenha um papel crítico na resposta imune de defesa celular contra vários patógenos (ELROD-ERICKSON; MISHRA; SCHNEIDER, 2000; INOUE et al., 2001). No genoma do carrapato da espécie *I. scapularis* foram averiguadas a ocorrência de cinco sequências de nucleotídeos com potencial para a expressão de pequenas GTPases Ras pertencentes à família Rho, que desempenha um papel central na fagocitose (BOKOCH, 2005; ETIENNE-MANNEVILLE; HALL, 2002), com pelo menos 33 potenciais proteínas relacionadas com a sua via fagocitária (SMITH; PAL, 2014).

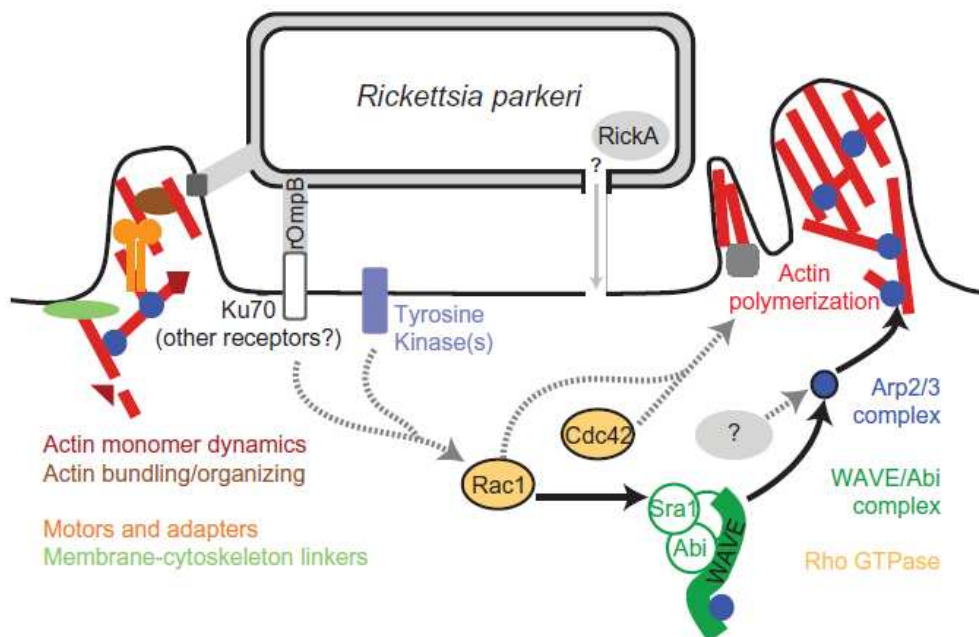
Estudos da capacidade de invasão de células por algumas espécies de *Rickettsia* do Grupo das Febres Maculosas sugeriram que as espécies *Rickettsia parkeri*, *Rickettsia conorii* e *Rickettsia montanensis* invadem rapidamente uma variedade de células dependendo da polimerização de actina do hospedeiro e da atividade de tirosina quinase. Em células de mamíferos, a infecção das células do hospedeiro por espécie de *Rickettsia* baseia-se na aderência da bactéria na membrana celular seguida da invasão por meio do processo de fagocitose induzida (WALKER, 1984; WALKER; WINKLER, 1978). As espécies de *Rickettsia* do Grupo das Febres Maculosas possuem dois tipos bem caracterizados de proteínas de aderência, conhecidas como OmpA e OmpB, enquanto a proteína OmpA não é encontrada em espécies no Grupo do Tifo, sendo intimamente identificada como uma proteína essencial na adesão celular por *R. rickettsii* (LI; WALKER, 1998). As proteínas codificadas para Sca0 (OmpA), Sca1, Sca2 e Sca5 (OmpB), desempenham importantes papéis na adesão e invasão destas bactérias (CHAN; RILEY; MARTINEZ, 2010), exceto na *R. conorii* que se liga com células epiteliais e endoteliais de mamíferos sem a proteína Sca1, apresentar este papel no processo de invasão (RILEY et al., 2010).

A proteína Sca2 expressa em *R. conorii* e é conservada entre muitas espécies do Grupo da Febre Maculosa, participando na mediação de interações com células-alvo de mamífero (CARDWELL; MARTINEZ, 2009). Em *R. parkeri* desempenha um papel como um mediador da motilidade baseada em actina, podendo interagir com as proteínas do hospedeiro para promover a sua sobrevivência (HAGLUND et al., 2010). A proteína Sca4 colocaliza com a proteína vinculina no local celular de aderência, ligando-se e ativando através de dois sítios conservados em todas as *Rickettsia* spp. (PARK et al., 2011). Curiosamente, Ku70, uma subunidade de DNA dependente nuclear proteína quinase localizada no citoplasma, bem como na membrana plasmática, serve como um receptor para OmpB, desempenhando um papel importante na internalização de *R. conorii* (MARTINEZ et al., 2005). Evidências em apoio a este achado sugerem que a ubiquitinação de Ku70 da célula hospedeira ocorre durante a infecção por *Rickettsia* recrutando ubiquitina ligase c-cbl para os locais de entrada (CHAN et al., 2009).

A polimerização da actina das células hospedeiras possui um importante papel na internalização de espécies de *Rickettsia*, envolvendo os mecanismos de sinalização de proteína GTPase Cdc42, fosfoinositídeo 3-quinase, c-Src e possivelmente outras proteínas tirosina quinases para ativar o complexo Arp2/3 (MARTINEZ, 2004; RYDKINA; TURPIN; SAHNI, 2008), havendo a participação de proteína clatrina e endocitose dependente de caveolina-2 (CHAN et al., 2009). As espécies de *Rickettsia* provavelmente utilizam proteínas membranolíticas, hemolisina C e fosfolipase D, codificadas pelos genes *tlyC* e *pIdA*, respectivamente, para romper as membranas fagossômicas e obter acesso ao citosol do hospedeiro (RADULOVIC et al., 1999; RENESTO et al., 2003; WHITWORTH et al., 2005). Uma fosfolipase A2 também foi predita no envolvido da entrada de espécies de *Rickettsia* nas células hospedeiras (SILVERMAN et al., 1992).

Para a espécie *R. parkeri* foram identificadas 21 proteínas nucleares GTPases Rho Rac1 e Rac2, além do fator promotor da nucleação do complexo WAVE e do complexo Arp2/3 (PETCHAMPAI et al., 2015). O complexo Arp2/3 atua durante a invasão de *R. parkeri* em células de mamíferos, enquanto o complexo WAVE e as proteínas GTPases da família Rho dependeram particularmente do tipo celular, no entanto, Rac1 e outra GTPase Cdc42 variam entre os tipos celulares (REED; SERIO; WELCH, 2012) (Figura 31).

Figura 31 Modelo de vias hospedeiras ativadas durante a invasão de *Rickettsia* e *Orientia* em células humanas. Na invasão da espécie *Rickettsia parkeri* em células hospedeiras de humanos a proteína Rho GTPase coopera com as proteínas Cdc42 e Rac1 na polimerização de actina, desempenhando papéis paralelos.



Fonte Reed et al. (2012).

Na Figura 31, verifica-se que a invasão celular por *R. parkeri* ocorreria por meio de uma cooperação entre estas proteínas para promoverem a polimerização de actina (REED; SERIO; WELCH, 2012). Da mesma forma, na caracterização funcional do complexo Arp2/3 do carrapato da espécie *Dermacento variabilis* identificou uma molécula relativamente conservada que está envolvida na invasão de *R. montanensis* (PETCHAMPAI et al., 2014b). Os estudos comparando os meios de invasão de espécies de *Rickettsia* do Grupo das Febres Maculosas em diferentes células demonstraram a possibilidade de mecanismos conservados que medeiam funções relacionadas com GTPase Rho e do complexo Arp2/3 (PETCHAMPAI et al., 2015). Segundo estes autores, a invasão de *Rickettsia* em células de artrópodes pode ocorrer através de uma via primária levando à nucleação de actina, enquanto a invasão em células endoteliais de mamíferos ocorre de forma redundante por caminhos que convergem para o recrutamento de actina com a ativação do complexo Arp2/3.

A motilidade intracelular baseada em movimentos de actina, em *Rickettsia* spp. do Grupo das Febres Maculosas ocorre por meio da formação de uma cauda de actina polar, o que não apenas facilita a sua movimentação, mas também desempenha um papel fundamental no seu estabelecimento, disseminação intercelular e patogênese (HEINZEN, 2003). Por outro lado, as espécies de *Rickettsia* do Grupo do Tifo não apresentam esta motilidade baseada em actina,

como é o caso da espécie *R. prowazekii*, ou apresentam os padrões de motilidade exibidos pela espécie *R. typhi* (HEINZEN, 2003; JENG et al., 2004). Para esta espécie, a disseminação de célula para célula é significativamente reduzida, devido possuírem comprimentos mais curtos de cauda de actina, no entanto, com taxa de movimentação muito semelhante quando comparada com a espécie *R. rickettsi* (HEINZEN, 2003).

A proteína da síndrome de Wiskott-Aldrich (WASP) denominada RickA é um fator promotor de nucleação capaz de ativar o complexo Arp2/3 e induzir polimerização de actina sendo o suficiente para direcionar a motilidade *in vitro* (GOUIN et al., 2004; JENG et al., 2004). Como esperado, a proteína RickA está presente em espécie de *Rickettsia* do Grupo das Febres Maculosa, mas não foi encontrado em *R. prowazekii* (GOUIN et al., 2004). A exceção entre as espécies deste grupo é a espécie *R. peacockii* que carece de motilidade baseada em actina, sendo uma espécie não patogênica para humanos. A motilidade baseada em actina em *R. typhi*, embora errático em relação aos movimentos lineares de espécie de *Rickettsia* spp. do Grupo das Febres Maculosas, na ausência de RickA sugere a possibilidade de existirem mecanismos redundantes de polimerização de actina em espécies de *Rickettsia* patogênicas (BALRAJ et al., 2008). Esta hipótese é ainda substanciada pela presença e expressão de RickA em *Rickettsia raoultii*, uma espécie recentemente isolada de carrapatos do gênero *Dermacentor*. A espécie *R. raoultii* codifica uma proteína de 565 aminoácidos altamente homóloga a RickA com outras espécies de *Rickettsia*, apresentando um defeito na motilidade baseada em actina durante a infecção *in vitro* de células L929 (BALRAJ et al., 2008).

A espécie *R. bellii*, pertencente ao Grupo Ancestral, codifica uma proteína ortóloga de RickA, formando caudas de actina para o movimento e colonização do núcleo da célula hospedeira. Ao entrar no núcleo desprovido de um complexo Arp2/3, que é necessário para a formação da cauda de actina, a bactéria aparentemente fica presa, resultando na abolição da motilidade baseada em actina. Presumivelmente, *R. bellii* utiliza RickA para entrar no núcleo da célula do hospedeiro, aonde se multiplica até que a membrana nuclear seja rompida, liberando as bactérias (OGATA et al., 2006). A presença de motilidade baseada em actina para espécies de *Rickettsia* representa uma importante característica do estilo de vida para o gênero *Rickettsia*, com a exceção das espécies do Grupo do Tifo que perdeu esta proteína durante a evolução reductiva dos seus genomas (WELCH; REED; HAGLUND, 2012).

Além da adesão e invasão, os resultados obtidos sugerem que as espécies de *Rickettsia* poderiam atuar na regulação de proteínas envolvidas com a formação de vesículas endossomais

durante a fagocitose e exocitose, participando do tráfego no transporte entre as organelas intracelulares. No intestino do carrapato da espécie *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcrito para a sequência da proteína da subunidade sigma do complexo de revestimento de vesícula Ap3, sendo regulada de forma negativa, havendo a sequência da proteína Sec23 e da sequência para a proteína de clatrina, sendo reguladas de forma positiva, no intestino.

No intestino do carrapato da espécie *A. sculptum* infectado pela mesma bactéria, observou-se a ocorrência da expressão de transcrito para a sequências das proteínas subunidades beta1, sigma, alfa e delta do complexo de revestimento de vesícula Ap3, subunidade menor do complexo adaptador de clatrina e de componente de sequência do complexo de revestimento de vesícula Ap2, fator de ribosilação que se liga a GTPase Ras/Arf sendo reguladas de forma negativa. No intestino do carrapato da espécie *A. sculptum* infectado por *R. amblyommatidis*, observou-se a ocorrência da expressão de transcrito para a sequência da proteína da subunidade delta do complexo de revestimento de vesícula Ap3 sendo regulada de forma negativa.

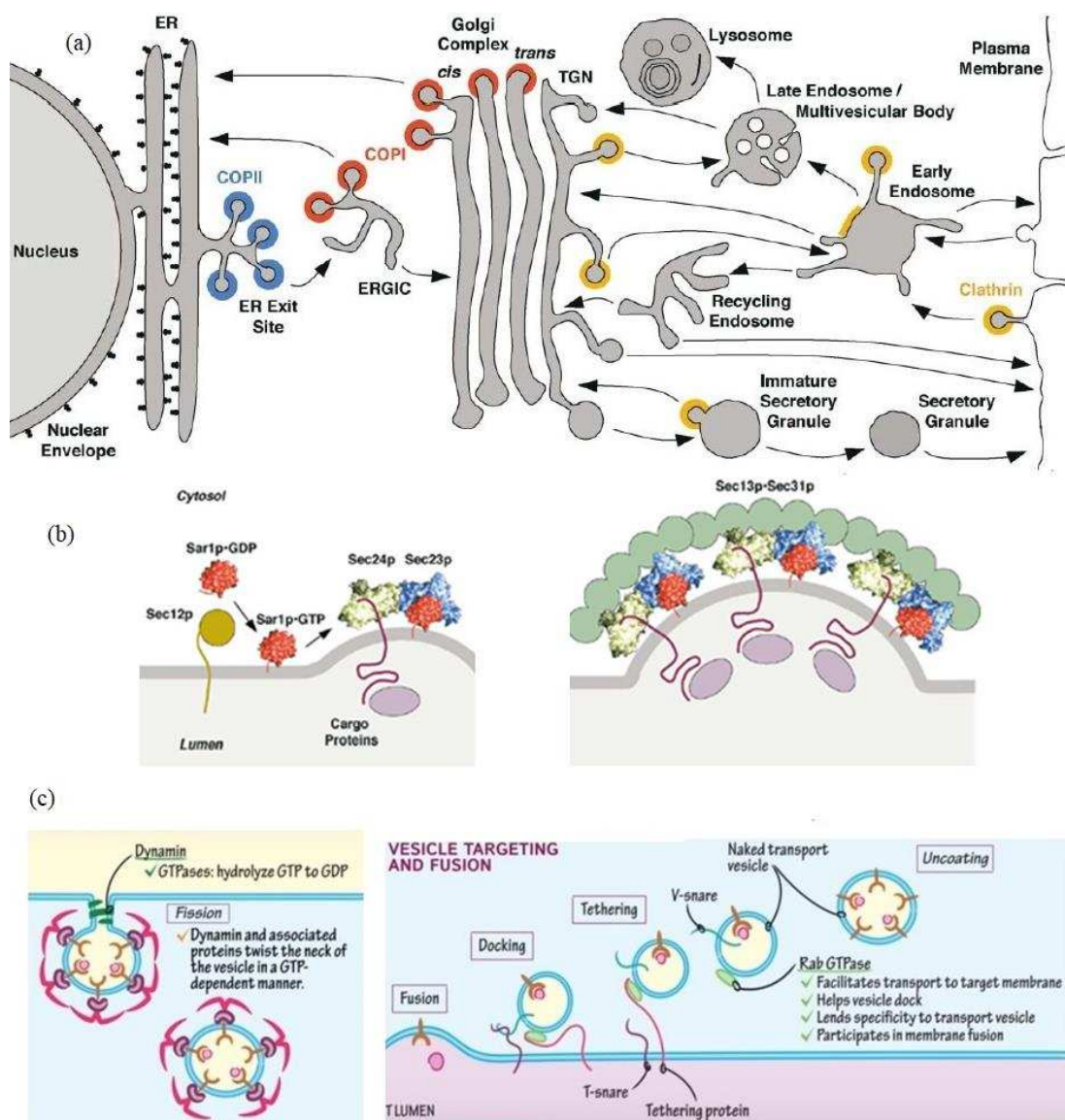
Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para as sequências de proteínas da subunidades Beta-1, sigma e delta do complexo de vesícula Ap3, uma sequência do complexo adaptador de clatrina e uma sequência para o fator 3 de ribosilação de ADP que se liga a GTPase Ras/Arf, Sec23 e uma sequência de Ap2 sendo reguladas de forma positiva no carrapato *A. aureolatum* em comparação com as mesmas sequências de transcritos, reguladas de forma negativa no carrapato *A. sculptum*. Com exceção do transcrito para a sequência do complexo de revestimento de vesícula Ap2 sendo regulada de forma negativa no carrapato *A. aureolatum*, e da sequência da proteína sec23 apresentando regulação positiva no carrapato *A. sculptum*.

Nos intestinos do carrapato *A. sculptum* infectado por *R. rickettsii* versus *R. amblyommatidis*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da subunidade delta do complexo de revestimento de vesícula Ap3 sendo regulada de forma negativa.

Diversos estudos têm observado que a manutenção e estabilidade do retículo endoplasmático, do complexo de Golgi e da membrana celular são aspectos importantes para a compreensão das rotas e vias intracelulares envolvidas na manutenção do ciclo de vida de vários patógenos no ambiente celular. As proteínas transportadoras e de revestimentos de vesículas

Ap1, Ap2 e Ap3 atuam no brotamento de membranas no complexo de Golgi e nas membranas do retículo endoplasmático contando com a participação de clatras e de pequenas proteínas GTPases do tipo Ras (OSTERMANN et al., 1993). O complexo de revestimento de vesícula Ap3 é homólogo de Ap1 e Ap2, sendo encontrado nas membranas do complexo de Golgi e dos endossomos, necessitando de Sar1 (HARA-KUGE et al., 1994) e do fator de ribosilação de ADP que se liga a GTPase Ras/Arf (OSTERMANN et al., 1993). O complexo de revestimento de vesícula Ap2 necessita de uma pequena GTPase Sar1p e de um fator de troca de guanina (BARLOWE; SCHEKMAN, 1993; BONIFACINO; GLICK, 2004; BUDNIK; STEPHENS, 2009) e dos subcomplexos Sec23p.Sec24p e Sec13p.Sec31p (BARLOWE, 1994) (Figura 32)

Figura 32 Esquema representativo dos compartimentos e vias de secreção. Destaque (a), lisosomal, vacuolar e endossomos, indicando o sistema de transportes de proteínas de revestimento de vesículas do tipo Ap1 (vermelho), Ap2 (azul) e Clatras (amarelo). Destaque (b), subcomplexos Sec23p.Sec24p e Sec13p.Sec31p. Destaque (c), o complexo de revestimento de vesícula Ap2 necessita de uma pequena GTPase Sar1p.



Fonte: Destaque (a), Bonifácio & Glick. (2004). Destaque (b), Barlowe et al. (1994). Destaque (c), Budnik & Stephens (2009).

Na Figura 32, verifica-se que a função da exportação do retículo endoplasmático (ER) por meio do complexo de revestimento de vesícula Ap2 e da endocitose da membrana plasmática por meio de clatrina são plenamente conhecidas sendo menos bem compreendidas, as funções exatas do complexo de revestimento de vesícula Ap1 durante a formação do retículo endoplasmático incorporado ao complexo de Golgi e de clatrina durante a formação do trans-Golgi (TGN), endossomos precoces e grânulos imaturos de secreção (BONIFACINO; GLICK, 2004). Segundo estes autores, os complexos de revestimento de clatrina são heterogêneos, contendo diferentes proteínas adaptadoras e acessórias em diferentes membranas. As clatrininas são restritas os locais pós-Golgi, incluindo a membrana plasmática e os caminhos entre a rede trans-Golgi e os endossomos (GASPER; SOT; WITTINGHOFER, 2010). Segundo estes autores, nos endossomos a complexidade das funções das proteínas GTPases Ras é aumentada com o elevado número de fatores relacionados a reguladores de troca de guanina e ativadores de GTPase (GAPs) como Di-Ras1 e Di-Ras2, o que as torna incapazes de reagir com fatores de troca de guanina da família Ras Cdc25.

De certa forma, ambos os carrapatos *Amblyomma* spp. e as espécies de *Rickettsia*, neste estudo poderiam utilizar desta maquinaria para controlar o transporte intracelular. Além da adesão, invasão e a fagocitose induzida, os nossos resultados sugerem, que as espécies de *Rickettsia* poderiam atuar na regulação de proteínas envolvidas com a formação de vesículas endossomais durante os processos de fagocitose e exocitose, participando do tráfego no transporte entre as organelas intracelulares, além de interagir com outras proteínas GTPases que garantem a estabilidade e a manutenção do ambiente celular.

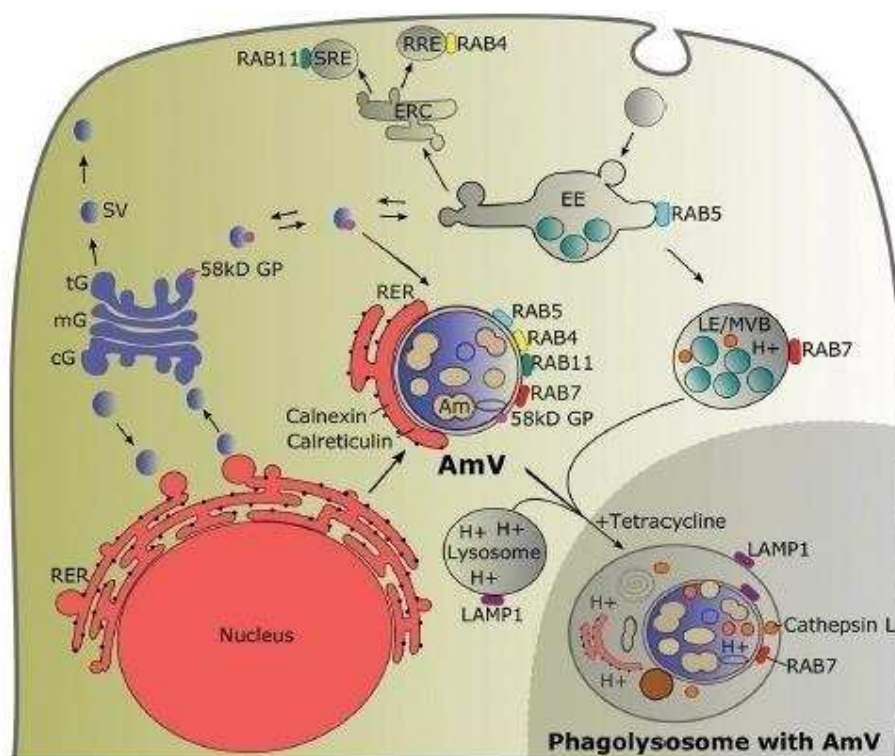
Para o carrapato da espécie *A. aureolatum* infectado por *R. rickettsii*, não se observou a ocorrência da expressão de transcritos para sequência da proteína GTPase Rab no seu intestino. Para o carrapato da espécie *A. sculptum* infectado pela mesma bactéria, observou-se a ocorrência da expressão de transcrito para sequência da proteína Rab1A sendo regulada de forma negativa no intestino. Para o carrapato da espécie *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para sequência das proteínas GTPs Rab20, Rab32, sendo reguladas de forma negativa no intestino.

Entre os intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, não se observou a ocorrência da coexpressão de transcritos para sequências de proteínas GTPases Rab.

Entre os intestinos do carrapato da espécie *A. sculptum* infectados por *R. rickettsii* versus *R. amblyommatis*, observou-se a ocorrência da coexpressão de transcritos para sequências de proteínas GTPase Rab-1A, sendo reguladas de forma negativa.

As principais funções atribuídas das GTPases Rab são regular as vias exocíticas e endocíticas das proteínas para modular o transporte e o tráfico de proteínas intracelulares (LARA, 2005; OLKKONEN; SLENMARK, 1997; SOJKA et al., 2013), atuando no sistema endossômico de artrópodes (FISCHER; EUN; DOOLAN, 2006), aonde podem ser manipuladas por bactérias para a sobrevivência de patógenos (BRUMELL; SCIDMORE, 2007). As bactérias *Rickettsiales* possuem diferenças para o estabelecimento e a sua manutenção no interior de vesículas celulares dos carrapatos (MAGUNDA et al., 2016). Segundo estes autores, a bactéria da espécie *Anaplasma marginale* ativa proteínas GTPase Rab responsáveis pela formação inicial dos vacúolos endossomais, reciclagem do endossomo e o retardo do amadurecimento do endossomo, o que contribui para o pH próximo de neutro dentro destas vesículas no interior das células do carrapato da espécie *Dermacentor andersoni* (Figura 33).

Figura 33 Modelo que resume a interação do patógeno *A. marginale* (AmV) (vacúolo) em células de carrapato da espécie *D. andersoni* sob condições de pH neutro com GTPases Rab nas vias endocitária e secretora. A inserção na parte inferior direita mostra a maturação do patógeno em um fagolisossomo multiorganelo com LAMP1, Rab7 e catepsina L. cG, cis-Golgi; mG, medial-Golgi; tG, trans-Golgi; proteína V-ATPase GP 58kD, proteína Golgi 58K; SG, grânulo de secreção.



Fonte: Magunda et al. (2016).

Na Figura 33, verifica-se que a proteína Rab5 presente no endossomo precoce (EE), é responsável pela reciclagem das vesículas endossomais contendo as proteínas Rab4 e Rab11, que compõem os endossomos de reciclagem rápida (RRE) e lenta (SRE), no centro de reciclagem endossomal (ECR), respectivamente, enquanto a proteína Rab7, é presente no endossomo tardio (LE) corpo multivesicular (MVB), no retículo endoplasmático rugoso (RER) e marcadores de retículo endoplasmático de calnexinae calreticulina e a proteína 58K no complexo de Golgi (MAGUNDA et al., 2016). O recrutamento de Rab1, Rab4A e Rab11A depende do fator de troca de guanina, enquanto a ativação de Rab10 é independente do recrutamento de clatrina (HUANG et al., 2010).

O patógeno *Anaplasma phagocytophilum* regularia o tráfego de membrana, atuando na reciclagem de endossomos por meio da regulação de uma proteína GTPase Rab10, o que facilita a sobrevivência intracelular independente de clatrina (MAGUNDA et al., 2016). Segundo estes autores, na ausência desta bactéria, este nicho vacuolar é levado ao amadurecimento para se tornar um fagolisossom o ativo e proteico impossibilitando a permanência da patógeno. A bactéria *Ehrlichia chaffeensis*, que infecta monócitos e macrófagos, também evita o encaminhamento dos endossomos das células hospedeiras sequestrando um regulador de tráfego de endossomo, e uma pequena GTPase Rab5 ligada à membrana, ao translocar o fator de translocação 2 (Etf-2) (YAN et al., 2018). Segundo estes autores, o Etf-2 é um efetor predito do sistema de secreção do tipo IV de *Ehrlichia*, sendo secretado no citoplasma da célula hospedeira, sugerindo um mecanismo único pelo qual a proteína GTPase Rab5 é regulada e mantida durante as inclusões desta bactéria beneficiando sua sobrevivência e replicação.

O impedimento do tráfico de vesículas para a formação dos fagolisossomos é uma estratégia de sobrevivência fundamental da maioria dos patógenos intracelulares que se replicam dentro de células de hospedeiros eucariotos (MAGUNDA et al., 2016). Da mesma forma que a bactéria intracelular obrigatória *E. chaffeensis* evita o encaminhamento para os fagolisossomas das células hospedeiras, escapando dos endossomos, fagossomos e dos vacúolos para o citosol (YAN et al., 2018), estes fenômenos tem sido descritos para outras bactérias especialmente as pertencentes aos gêneros, *Orientia* (RIKIHISA; ITO, 1979) e *Rickettsia* (WHITWORTH et al., 2005).

De certa forma, ambos os carrapatos *Amblyomma* spp. e as espécies de *Rickettsia*, neste estudo poderiam utilizar desta maquinaria para controlar o transporte intracelular durante o tráfego de vesículas, escapar dos endossomos, garantir a estabilidade e a manutenção do

ambiente celular para a sobrevivência e replicação dentro das células dos hospedeiros. Os resultados obtidos sugerem a ocorrência de expressão de transcritos para a sequência da proteína GTPase dinamina que atua em um processo de endocitose independente de clatrina.

No intestino do carrapato da espécie *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para a sequência da proteína GTPase dinamina sendo regulada de forma negativa. No intestino do carrapato da espécie *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para duas sequências da proteína GTPase dinamina sendo uma regulada de forma negativa e a outra de forma positiva. No intestino do carrapato da espécie *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para a sequência da proteína GTPase dinamina sendo regulada de forma negativa.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para sequências de proteínas GTPases dinamina, sendo regulada de forma positiva no carrapato da espécie *A. sculptum* e de forma negativa no carrapato da espécie *A. aureolatum*, havendo a presença de transcritos para a proteína do adaptador do complexo de clatrina sendo regulada de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*. Nos intestinos do carrapato da espécie *A. sculptum* infectados por *R. rickettsii* versus *R. amblyommatis*, não se observou a ocorrência da coexpressão de transcritos para sequências de proteínas GTPase dinamina.

A dinamina é uma proteína GTPase relacionada com a cisão da membrana progenitora das vesículas nascentes, participando de vários tipos de endocitose durante a internalização de vários vírus pertencentes ao gênero Flavivírus (PERRY; WOBUS, 2010; RAHN et al., 2011). Segundo estes autores, a maioria destes vírus dependem do pH ácido intracelular e da ativação de dinamina independente de clatrina para a sua internalização em células de mamíferos. No entanto, este processo requer uma proteína denominada caveola dependente do recrutamento de dinamina em condições de pH ácido (ZHU et al., 2012). As vias endocíticas utilizadas por diferentes Flavivirus transmitidos por mosquitos e carrapatos diferem ligeiramente umas das outras (MAYOR; PAGANO, 2007), podendo ser mediadas por clatrina, caveola dependente de dinamina e endocitose independente de caveola (SCHELHAAS, 2010). Além destes vírus, os Myxovirus possuem um domínio característico que regulada as proteínas relacionadas à dinamina. As proteínas Mx são evolutivamente GTPases muito semelhantes as dinaminas,

sendo o domínio GTP a parte mais conservada destas proteínas (PITOSSO et al., 1993; VERHELST; HULPIAU; SAELENS, 2013).

De certa forma, ambos os carrapatos *Amblyomma* spp. e as espécies de *Rickettsia*, neste estudo poderiam utilizar desta maquinaria para controlar os processos de fagocitose e endossitose independente de clatrina, utilizando a GTPase dinamina que proporciona outras vias para o recrutamento de actina sendo predito para atuarem na invasão de espécies diferentes de Flavivirus que infectam em especial os carrapatos.

Ankirina e proteínas fosfatases putativas relacionadas com o recrutamento de actina

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para uma sequência de proteína contendo Ank, sendo regulada de forma positiva, para a sequência da actina acetiltransferase subunidade A (NAT1) e para uma sequência da cadeia leve da proteína miosina, presente no módulo de fosforilação, sendo reguladas de forma negativa. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a expressão de transcrito para uma sequência da proteína contendo Ank, sendo regulada de forma negativa, para uma sequência da cadeia pesada de miosina sendo regulada de forma positiva, e uma sequência da cadeia leve de miosina sendo regulada de forma negativa, presente no módulo de fosforilação. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, não se observou a ocorrência da expressão de transcritos para a sequência de proteína Ank, havendo transcritos para uma sequência da proteína da cadeia pesada de miosina e para uma sequência da cadeia leve de miosina, sendo reguladas de forma negativa, presente no módulo de fosforilação no intestino

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência de proteína contendo Ank sendo regulada de forma positiva no carrapato da espécie *A. aureolatum* e de forma negativa no carrapato da espécie *A. sculptum*, havendo uma sequência de proteína da actina acetiltransferase subunidade A (NAT1) sendo regulada de forma negativa no carrapato *A. sculptum* e de forma positiva no carrapato *A. aureolatum*. Ainda, nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para duas sequências proteína da cadeia leve de miosina sendo uma regulada de forma positiva em ambos

os carrapatos, enquanto a outra foi regulada de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*.

Nos intestinos do carrapato da espécie *A. sculptum* infectados pelas espécies *R. rickettsii* versus *R. amblyommatis*, observou-se a ocorrência da coexpressão de transcritos para a sequência da proteína da miosina sendo uma regulada de forma positiva e a outra de forma negativa, respectivamente, havendo a coexpressão de transcritos para duas sequências da cadeia leve de miosina sendo reguladas de forma negativa em ambos os órgãos na presença destas *Rickettsia* spp..

No *I. scapularis* a expressão e regulação dos componentes do citoesqueleto variam em resposta a infecção por *A. phagocytophilum* em um determinado órgão, no entanto codificam os principais componentes sugerindo a existência de uma evolução do citoesqueleto conservado em carrapatos (CABEZAS-CRUZ et al., 2017a). Segundo estes autores, a regulação do RNAm de actina e os níveis da proteína foram significativamente reduzidos tanto nas glândulas salivares, quanto no intestino. No entanto, a expressão de transcritos para as sequências de proteínas laminas e tropomiosinas demonstraram níveis aumentados em resposta a infecção.

Para compensar a diminuição dos níveis de infecção de *A. phagocytophilum* em células embrionárias de carrapato da espécie *I. scapularis* ISE6, este patógeno possui a capacidade de remodelar o arcabouço do citoesqueleto após o silenciamento de genes da tropomiosina, aumentando os níveis dessas proteínas relacionadas à actina no intestino do carrapato (CABEZAS-CRUZ et al., 2017a; SULTANA et al., 2010). No entanto, a α -tubulina é diferencialmente expressa no carrapato da espécie *D. variabilis* em resposta a infecção por *Rickettsia montanensis*, sendo regulada de forma negativa no intestino e de forma positiva nas glândulas salivares (MACALUSO et al., 2003). Diversos microrganismos, incluindo *Rickettsia*, entre outros, usam a polimerização de actina para se moverem e colonizarem as células dos hospedeiros recrutando as proteínas do citoesqueleto para a montagem de uma cauda de actina (GOUIN et al., 2004).

Estes patógenos manipulam o citoesqueleto de actina por meio de efetores ou fatores de virulência injetados nas células que visam vias de sinalização intracelular cruciais (BHAVSAR; GUTTMAN; FINLAY, 2007; MUNTER; WAY; FRISCHKNECHT, 2006). A bactéria *Yersinia* possui em seu genoma a proteína efetora YpkA com função de fosforilação de actinas globulares que ocasionam estresse e a desmontagem das fibras (NAVARRO et al., 2007). O

vírus Vaccinia utiliza de filamentos de actina para sua motilidade, baseando-se também nas vias de sinalização de tirosina quinase, controlando a dinâmica de nucleação de actina (FRISCHKNECHT et al., 1999). Em células de mamíferos, a proteína de superfície de *Rickettsia* RickA ativa o Complexo Arp2/3 para induzir a polimerização e a formação de filopodia de actina (MARTINEZ, 2004). A proteína Anka presente no genoma do patógeno *A. phagocytophilum* é fosforilada pela tirosina quinase Abl-1, a qual facilita sua infecção em células humanas (LIN et al., 2007). Em granulócitos humanos, a Anka se liga ao DNA nuclear (PARK et al., 2004) e em neutrófilos induzindo a fosforilação de tirosina, ajudando na sobrevivência intracelular deste patógeno (THOMAS; FIKRIG, 2007).

Ankirina, proteínas putativas ribossomais, histonas relacionadas no controle da expressão gênica e processos epigenéticos

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para duas sequências de proteínas contendo domínio Set (SMYD) presentes nos módulos de transporte e secreção, além da sequência da proteína da histona demetilase específica de lisina presente no módulo de oxidação, respectivamente, e as sequências de proteínas de arginina N-metiltransferase PRMT1 e O-metiltransferase presentes no módulo de metilação, sendo reguladas de forma positiva, havendo transcritos para as sequências das proteínas ribossomais L2, L11, L15, L26, S2 40S sendo reguladas de forma positiva, e transcrito para duas sequências de outras proteínas ribossomais sendo reguladas de forma negativa. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para a sequência de proteína histona demetilase específica de lisina sendo regulada de forma negativa, havendo transcritos para as sequências das proteínas ribossomais L2, L4 60S e S12 30S sendo reguladas de forma negativa. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, não se observou a ocorrência da expressão de transcritos para a sequência destas proteínas contendo domínio SMYD Set e MYND, histona demetilase e as sequências de proteínas de arginina N-metiltransferase PRMT1 e O-metiltransferase, havendo transcritos para as sequências das proteínas ribossomais L2, L11, L15, L17 e S12 30S sendo reguladas de forma negativa.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência de proteína contendo domínio

SMYD Set e MYND presente no módulo de secreção sendo regulada de forma positiva, em ambos os carrapatos, havendo transcritos para esta sequência de proteína presente no módulo de transporte sendo regulada de forma negativa no carrapato *A. sculptum* e de forma positiva no carrapato *A. aureolatum*. Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se também a ocorrência da coexpressão de transcritos para a sequências das proteínas ribossomais L17 sendo regulada de forma negativa em ambos os carrapatos, L2 regulada de forma negativa no carrapato *A. sculptum* e de forma positiva no carrapato *A. areolatum*, havendo transcritos para duas sequências de proteínas ribossomais processadoras sendo reguladas de forma negativa em ambos os carrapatos.

Ainda nos intestinos, observou-se a ocorrência da coexpressão de transcritos para as sequências das proteínas fatores de iniciação translacionais IF1A sendo regulados de forma positiva em ambos os carrapatos, IF2 regulado de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*, havendo transcritos presentes no módulo de secreção para duas sequências da proteína IF3 de ribosilação do ADP e IF3T/GF-Beta, sendo a primeira possuindo os transcritos regulados de forma negativa no carrapato *A. sculptum* e de forma positiva no carrapato *A. areolatum*, enquanto a segunda foi regulada de forma oposta para as espécies de *Amblyomma*. Além disto, observou-se a ocorrência da coexpressão de transcritos para a sequências das proteínas do fator de iniciação IF2 e IF3 de ribosilação do ADP sendo reguladas de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*. Nos intestinos do carrapato da espécie *A. sculptum* infectados pelas espécies *R. rickettsii* versus *R. amblyommatis*, observou-se a ocorrência da coexpressão de transcritos para a sequência das proteínas ribossomais L17 e S12 30S presentes no módulo de secreção sendo reguladas de forma negativa na presença de ambas as bactérias.

O genoma do carrapato *I. scapularis* foi bem caracterizado, constituindo-se portanto um bom modelo para o estudo das interações patógeno-carrapato (AYLLÓN et al., 2015; GULIANUSS et al., 2016), em especial o papel das histonas e enzimas modificadoras de histonas (HMEs) durante a infecção por *A. phagocytophilum* de *I. scapularis* (CABEZAS-CRUZ et al., 2016). A proteína anquirina AnkA de *A. phagocytophilum* secretada através do T4SS é um efetor que se liga à cromatina (LIN et al., 2007; RIKIHISA; LIN; NIU, 2010), recrutando histona desacetilase 1 HDAC (GARCIA-GARCIA et al., 2009) para modificar a expressão gênica do hospedeiro (DING; ATMAKURI; CHRISTIE, 2003), exercendo um controle

epigenético do hospedeiro e na metilação global do DNA (RENNOLL-BANKERT et al., 2015; SINCLAIR; YEGNASUBRAMANIAN; DUMLER, 2015).

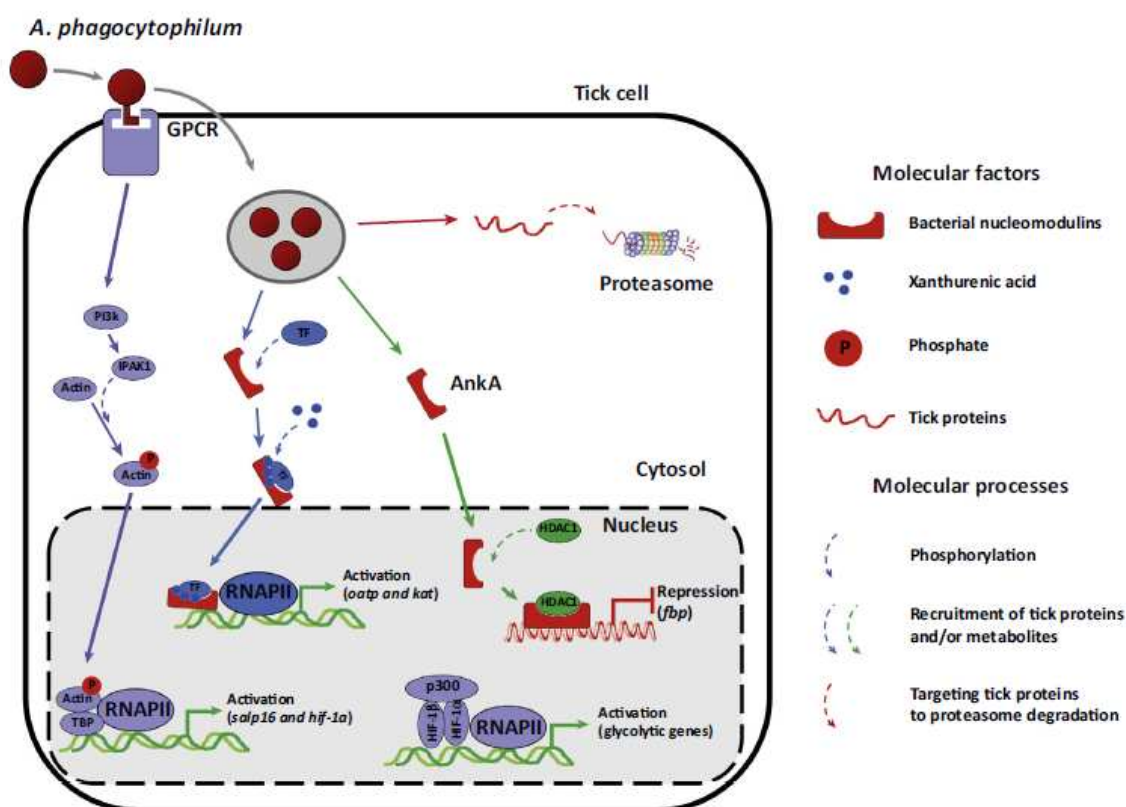
As proteínas contendo o domínio SET e MYND (SMYD) são uma classe especial de proteína lisina metiltransferases envolvidas na metilação de histonas H3H4 principalmente em resíduos específicos de lisina ou arginina (WOOD; SHILATIFARD, 2004). Dentro das famílias de histonas metilases, é observada no genoma do *I. scapularis* a ocorrência de sete famílias classificadas de acordo com seu domínio, incluindo SUV39, EZH, SET1, SET2, PRDM e SMYD (CABEZAS-CRUZ et al., 2016). A desacetilação ou metilação de histona é frequentemente observada em regiões promotoras de genes de defesa em células do hospedeiro infectadas por *A. phagocytophilum* (GARCIA-GARCIA et al., 2009), sugerindo um efeito amplo na transcrição de funções celulares (RENNOLL-BANKERT et al., 2015; SINCLAIR; RENNOLL-BANKERT; DUMLER, 2014). No carrapato da espécie *I. scapularis* foram identificados distintos membros do domínio Set, possuindo propriedades de sequência de aminoácidos específicas de queliceratos formando um cluster separado das sequências de mamífero e inseto (RENNOLL-BANKERT et al., 2015). Segundo estes autores, outros domínios, no entanto, foram semelhantes à estrutura do encontrado em humanos, com os domínios Anquirina, Pre-SET e SET.

A expressão gênica pode ser regulada por três mecanismos epigenéticos relacionados com a metilação do DNA, modificação de histonas e atividades de pequenos RNAs não codificantes (GÓMEZ-DÍAZ et al., 2012). Segundo estes autores, a metilação do DNA poderia silenciar a expressão gênica e remodelar a heterocromatina. Enquanto as modificações pós-traducionais em histonas, que incluem os principais componentes da cromatina, estariam relacionadas com a metilação, fosforilação, acetilação, ubiquitinação e sumoilação (ZHANG, 2001), podendo regular a transcrição, replicação, reparo do DNA e apoptose (NOWAK; CORCES, 2004).

O genoma do carrapato *I. scapularis* contém potencial para a expressão de diversas histonas, classificadas divididas em histona metilase, histona desmetilase, histona acetilase e histona desacetilase (CABEZAS-CRUZ et al., 2016). A acetilação destas proteínas regularia a estrutura da cromatina e a ligação de repressores e ativadores da transcrição gênica, enquanto a metilação regularia a organização da cromatina, dependendo de resíduos de lisina para a ativação ou repressão da expressão gênica (MOZZETTA et al., 2015). A ativação de vias de sinalização de carrapatos por *A. phagocytophilum* envolve o receptor acoplado à proteína G

(GPCR) que estimularia a proteína quinase ativada por Ixodes p21 (IPAK1) através da ativação de PI3K, resultando em sinalização mediada por actina (SULTANA et al., 2010). A ativação de IPAK1 fosforila a actina e acumulação desta proteína do citoesqueleto no núcleo medeia a ligação da RNA polimerase II (RNAPII) e proteína de ligação TATA box (TBP) a região do promotor para o gene *salp16*, induzindo sua expressão (CABEZAS-CRUZ et al., 2019) (Figura 34).

Figura 34 Mecanismos moleculares potencialmente implicados na modulação do metabolismo de carrapatos infectados por *A. phagocytophilum*. Destaque para a reprogramação transcricional e na modulação do proteoma.



Fonte: Cabezas-Cruz et al. (2019).

Na Figura 34, verifica-se a capacidade de *A. phagocytophilum* em regular de forma negativa a expressão gênica em neutrófilos de mamíferos está associada ao recrutamento da histona desacetilase 1 (HDAC1) hospedeira (CABEZAS-CRUZ et al., 2019) para os promotores de genes alvo pela proteína nucleomodulina contendo repetições de anquirina Anka (RENNOLL-BANKERT et al., 2015). Esta proteína secretada através do sistema de secreção do tipo IV T4SS, entra no núcleo do granulócito, se ligando a trechos de DNA rico em AT e recrutando a histona HDAC1 do hospedeiro para seus locais de ligação (CABEZAS-CRUZ et al., 2019). A infecção pode também induzir a ligação de HDAC1 e desacetilação da

histona H3, resultando na regulação negativa de muitos genes de defesa do hospedeiro, incluindo CYBB, que codifica componente NOX2 da NADPH oxidase (RENNOLL-BANKERT et al., 2015). Segundo estes autores, a regulação negativa do CYBB reduz a produção de ânion superóxido pela NADPH oxidase, o qual é essencial para a sobrevivência de *A. phagocytophilum* em neutrófilos. Homólogos de HDAC1 e H3 de mamíferos foram identificados no genoma da espécie de carrapato *I. scapularis* (CABEZAS-CRUZ et al., 2016).

A correlação negativa entre o aumento dos níveis de expressão de HDAC1 ocasiona a regulação da expressão do RNAm de proteína FBP, sugerindo que as histonas podem ser recrutadas por nucleomodulinas fagocitopênicas para atuar como repressores da expressão de genes ligados a glicose em glândulas salivares de carrapato (CABEZAS-CRUZ et al., 2017b). Segundo estes autores, a ativação da via fosfoinositido-3-quinase (PI3K), também foi sugerida para induzir a expressão de genes glicolíticos hif-1a, hif-1b e acetiltransferase p300 na infecção por *A. phagocytophilum*. A diminuição nos níveis de proteína FBP pode contribuir para a inibição da gliconeogênese que *A. phagocytophilum* induz em células de carrapato infectadas (VILLAR et al., 2015). Como consequência, o patógeno *A. phagocytophilum* altera os níveis de enzimas do metabolismo de aminoácidos (CABEZAS-CRUZ et al., 2017c), que, por sua vez, afetam os níveis intracelulares de vários aminoácidos e intermediários metabólicos (CABEZAS-CRUZ et al., 2019; TAANK et al., 2017; VILLAR et al., 2015).

Domínios PDZ, proteínas putativas relacionadas com a polaridade celular

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para a sequência de proteína contendo domínio PDZ e da sequência de proteína do disco largo (ionotrópico glutamato) sendo reguladas de forma positiva, presentes no módulo de secreção, havendo transcritos para a sequência da proteína Lin-12 sendo regulada de forma negativa presentes no módulo de fosforilação. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para uma sequência de proteína de domínio PDZ/DHR/GLGF, uma sequência da proteína sintenin PDZ, e uma sequência de proteína de receptor de ionotrópico glutamato sendo reguladas de forma negativa, havendo transcritos para uma sequência da proteína atrophin 1 aip 1 (guanilato), calmodulina e do P53 sendo reguladas de forma positiva presentes no módulo de fosforilação. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-

se a ocorrência da expressão de transcritos para uma sequência da proteína calmodulina sendo regulada de forma negativa, havendo transcritos para a sequência da proteína de receptor ionotrópico de glutamato e de uma proteína associada a apoptose sendo reguladas de forma positiva presentes no módulo de fosforilação.

No intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína sintenin sendo regulada de forma negativa no carrapato *A. sculptum* e de forma positiva para a espécie *A. aureolatum*, havendo transcritos para a sequência da proteína Atrophin aip 1 sendo regulada de forma positiva no carrapato *A. sculptum* e de forma negativa para a espécie *A. aureolatum*, presentes no módulo de secreção. Ainda nos intestinos dos carrapatos *Amblyomma* spp., observou-se a ocorrência da coexpressão de transcritos para as sequências das proteínas calmodulina e do fator de apoptose sendo reguladas de forma positiva para ambos os carrapatos, havendo transcritos para uma sequência da proteína Lin-12 sendo regulada de forma negativa para ambos os carrapatos presentes no módulo de fosforilação. Nos intestinos do carrapato da espécie *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para a sequência da proteína do receptor de ionotrópico glutamato presente no módulo de fosforilação sendo regulado de forma negativa na presença de *R. rickettsii* e de forma positiva na presença de *R. amblyommatis*.

As proteínas de domínios PDZ são módulos de interação de proteínas envolvido na manutenção da polaridade celular e na regulação da dinâmica sináptica (HUNG; SHENG, 2002; ROCHE et al., 2002). No carrapato, o vírus da encefalite regula dentro do núcleo da célula hospedeira a maquinaria de importação nuclear (MELIK et al., 2012). A replicação de flavivírus pode ocorrer por meio da ativação de uma proteína NS5 com ambas, as atividades de RNA metiltransferase e RNA-polimerase dependente de RNA (BROOKS et al., 2002). O silenciamento do gene responsável pela expressão desta proteína ocasiona modificações nas interações com as proteínas do hospedeiro dependentes de domínios PDZ, o que influencia na replicação deste patógeno transmitidos por carrapato (MELIK et al., 2012).

As proteínas virais de domínio PDZ interagem com outras proteínas quinase dependentes de cálcio como a calmodulina, a proteína receptora de glutamato e interleucina (MELIK et al., 2012). Nos insetos, o canal de cloreto é controlado por glutamato (Glu-Cl) que é um membro da superfamília “Cys-loop” de canais iônicos dependentes de ligantes, podendo ser inibido por Fipronil, um potente inseticida bloqueador dos canais de cloro controlados por

GABA ou pelo receptor de GABA (COLE; NICHOLSON; CASIDA, 1993; DONG et al., 2013). No camarão a proteína Pm-syntenin contendo domínios PDZ está ligada a um mecanismo que o protege contra infecções virais, indiretamente envolvida com uma via de sinalização que incluindo estes parceiros de ligação (GEIJSEN, 2001; GROOTJANS et al., 1997). Segundo estes autores, a proteína ligante de Pm-sintenina no hemócito de camarões infectados pelo vírus WSSV foi identificada como um adaptador de proteína interagindo com várias moléculas em uma via de transdução de sinal ligadas a proteínas de domínio PDZ.

A sintenina poderia regular a apoptose dependente de p53, sinalizando para a morte ou sobrevivência celular (LI et al., 2004). As proteínas Lin-7, Lin-2 e Lin-10 contêm domínios PDZ, os quais são conhecidos por mediar proteínas e interações de proteína com caudas C-terminais (SIMSKE et al., 1996). A proteína Lin-2 contém um domínio de ligação a calmodulina, um domínio PDZ, um domínio SH3 e um domínio guanilato quinase (HOSKINS et al., 1996). A proteína Lin-2 está relacionado a uma família de proteínas chamadas kitransmembrana de guanilato associada com proteínas membranas (MAGUKs), incluindo proteínas do disco largos (DlgA) (KIM, 1997).

Tpr, proteínas putativas, FKBP prolil cis trans isomerase, (PPIASE) e chaperonas moleculares relacionadas a fatores ambientais, estresse e uma gama de processos.

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência de expressão de transcritos para a sequência da proteína contendo Tpr sendo regulada de forma positiva, havendo transcritos para as sequências de proteínas de choque térmico Hsp40 sendo regulada de forma negativa. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência de expressão de transcritos para uma sequência da proteína contendo Tpr, havendo transcritos para uma sequência da proteína de choque térmico Hsp70 sendo reguladas de forma negativa. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para duas sequências da proteína contendo Tpr sendo uma regulada de forma negativa e a outra de forma positiva, havendo transcritos para uma sequência da proteína de choque térmico Hsp70 sendo regulada de forma positiva.

Nos intestinos das espécies de carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, não se observou a ocorrência da coexpressão de transcritos para sequências da proteína

contendo Hsp, havendo transcritos para uma sequência da proteína contendo Tpr sendo regulada de forma positiva em ambos ácaros. Nos intestinos do carrapato da espécie *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína contendo Tpr sendo regulada de forma negativa na presença da espécie *R. amblyommatis*, e de forma positiva na presença da espécie *R. rickettsii*, havendo transcritos para uma sequência da proteína de choque térmico Hsp90 sendo, regulada de forma positiva na presença de ambas espécies de *Rickettsia* estudadas.

As proteínas chaperonas de choque térmico Hsp90 e Hsp70 possuem a função de diferenciar sequências de proteínas dobradas de forma correta, daquelas que necessitam ser reparadas ou degradadas com padrões de ubiquitinação, o que envolveria a autofagia, sinalizando para a entrada na via de degradação lisossomal (CHEN; MALLAMPALLI, 2009). No entanto, estas proteínas também fornecem proteção contra a colonização bacteriana, bem como a inflamação (POCKLEY, 2003). As proteínas chaperonas Hsp90 e Hsp70 de classe II do sistema de secreção do tipo III (T3SS) são proteínas que contem motivos Tpr relacionadas à patogênese bacteriana (BÜTTNER et al., 2008; PALLEN; FRANCIS; FÄTTERER, 2003). Estas proteínas estão diretamente envolvidas com a translocação de fatores de virulência e adesão a células hospedeiras, atuando no bloqueio da maturação fagolisossômica (BRÄMS et al., 2006; CHAKRABORTY et al., 2008; EDQVIST et al., 2006).

Diversos estudos têm relatado que sua atuação contribui na patogênese de *Yersinia*, *Pseudomonas*, *Shigella*, *Mycobacterium*, *Porphyromonas* e *Francisella* (CERVENY et al., 2013). Em várias espécies de carrapatos Ixodídeos e Argasídeos as proteínas de choque térmico Hsp são conservadas (FRANCISCHETTI et al., 2005; LEWIS et al., 2015). Apesar do patógeno *Borrelia burgdorferi* possuir uma proteína que interage com proteínas de células de mamíferos contendo motivos Tpr, a interação é independente da região do motivo, sendo no entanto, relacionada com aspectos da invasão desta espiroqueta (GROSHONG et al., 2014). Segundo estes autores, para esta bactéria, a região da chaperona que compreende o motivo Tpr influencia na sua estabilidade proteica e infecciosidade, envolvendo diretamente uma interação proteína-proteína.

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcrito para a sequência da proteína FKBP peptidil-prolil-cis trans isomerase sendo regulada de forma positiva. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcrito para

esta proteína sendo regulada de forma negativa. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcrito para duas sequências desta proteína sendo uma regulada de forma negativa presente no módulo de secreção e a outra sendo regulada de forma positiva presente no módulo de transporte.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcrito para uma sequência da proteína FKBP sendo regulada de forma positiva em ambos os ácaros e presente no módulo de fosforilação. Nos intestinos do carrapato *A. sculptum* infectos por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para duas sequências desta proteína, havendo uma sendo regulada de forma negativa e a outra de forma positiva na presença de ambas as bactérias das espécies *R. rickettsii* e *R. amblyommatis*.

A peptidyl-prolyl cis-trans isomerase com atividade (PPIase) poderia catalisar a isomerização de ligações cis-trans peptídicas que precedem os resíduos de prolina, envolvendo uma ampla gama de processos, como divisão celular, regulação transcricional, tráfego de proteínas e RNA *splicing* (FERNANDES; FERREIRA; BOTELHO, 2018). Nas glândulas salivares do carrapato da espécie *I. scapularis*, estudos verificaram a ocorrência da expressão de enzimas com características imunogênicas como a ciclofilina A conhecida como isomerase FKBP peptidil-prolil cis-trans (FISCHER; SCHMID, 1990). FKBP pode ser imunossuprimida por ciclosporina A reduzindo a ativação de células T de defesa (LIU et al., 1991), podendo regular a concentração de Ca^{2+} em plaquetas ativadas (ELVERS et al., 2012). Várias espécies de patógenos, incluindo *Theileria parva* e *Theileria annulata* infectam causam alterações em leucócitos ocasionando um comportamento comum atribuído à vários tipos de células cancerígenas como o processo de imortalização, hiperproliferação e disseminação (TRETINA et al., 2015).

A proteína PPIase de *Theileria* é dependente de fosforilação, atuando na transformação de células hospedeiras, via Janus Kinase transdutor de sinal e ativador da transcrição (JAK/STAT), sendo é vital para este processo e para sua proliferação (MARSOLIER et al., 2015). Segundo estes autores, no entanto, em linfócitos B bovinos infectados em cultura com *Theileria annulata*, a via JAK/STAT é desencadeada principalmente em função da via de degradação reduzida pela proteína ubiquitina ligase bovina FBW7. Recentemente, um estudo relatou que a ciclofilina A do carrapato *Haemaphysalis longicornis* possui um domínio PPIase conservado, sendo expresso em múltiplos órgãos em todos os estágios de desenvolvimento

(BOLDBAATAR et al., 2008). Além disso, houve um relato sobre o gene putativo de imunofilina em carrapatos *Rhipicephalus (Boophilus) microplus* que tendo alta homologia com o gene da ciclofilina A, com o silenciamento gênico aumentando significativamente a taxa de infecção de *Babesia bovis* na progênie larval (BASTOS et al., 2009).

Proteínas histidinas quinases e motivos ricos em leucinas (LRRs), WD, F-box e Ubiquitina

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para quatro sequências da proteína contendo domínio de repetição WD sendo três reguladas de forma positiva e uma de forma negativa, havendo transcritos para uma sequência da proteína lumican sendo regulada de forma negativa. Ainda, no intestino deste carrapato, observou-se a ocorrência de transcritos para uma sequência da proteína Ubiquitina, sendo regulada de forma positiva.

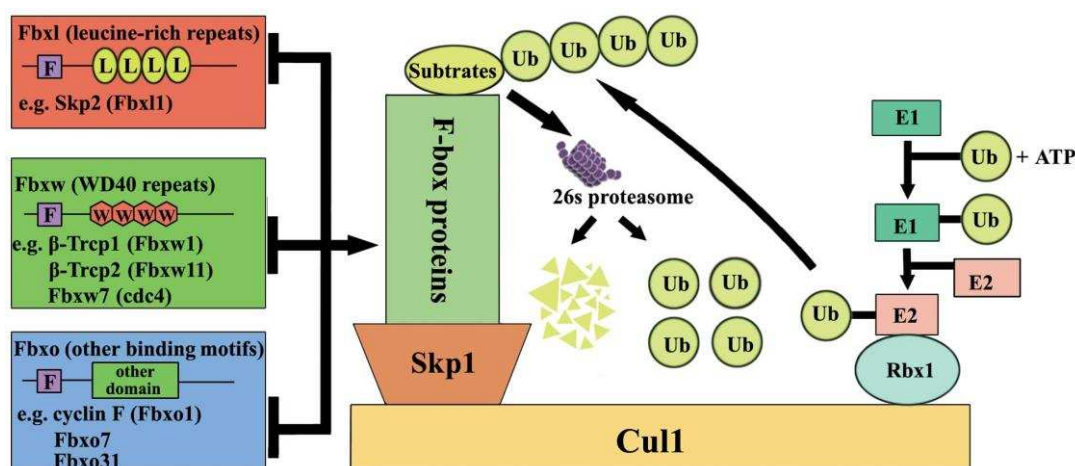
No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, não se observou a ocorrência da expressão de transcritos para sequência da proteína contendo domínio de repetição WD, havendo transcritos para uma sequência da proteína lumican sendo regulada de forma negativa. Não houve a ocorrência da expressão de transcritos para a sequência da proteína ubiquitina. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para uma sequência da proteína contendo domínio de repetição WD sendo regulada de forma negativa, não havendo transcritos para a sequência da proteína lumican. Ainda, no intestino deste carrapato, observou-se a ocorrência da expressão para uma sequência da proteína ubiquitina sendo regulada de forma positiva.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína contendo domínio de repetição WD sendo regulada de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*, presentes no módulo de secreção. Nos intestinos do carrapato *A. sculptum* infectos por espécies de *Rickettsia*, não se observou-se a ocorrência da coexpressão de transcritos para estas sequências na presença de ambas as espécies *R. rickettsii* e *R. amblyommatis*.

As proteínas F-box são os componentes de recrutamento de substrato dos complexos E3 da proteína de ligação a ubiquitina e determinam a especificidade do alvo através do

reconhecimento e da ligação de proteínas alvo para a ubiquitinação e depois para a degradação, apresentando na região C-terminal um domínio de interação de proteínas variável, como repetições Trp-Asp (WD40) e repetições ricas em leucina (LRR), bem como motivos desconhecidos, que são responsáveis pela identificação de substratos específicos (GONG; HUO, 2015) (Figura 35).

Figura 35 A ubiquitina é ativada pela enzima ativadora da ubiquitina (E1), transferida para a enzima conjugadora da ubiquitina (E2) e finalmente transferida para os substratos pelas ligases da ubiquitina (E3). As proteínas da ubiquitina são reconhecidas e degradadas pelo proteossoma 26s. O complexo E3 ligase consiste na variação de F-box de Skp1, Rbx1, e uma proteína Cullin.



Fonte: Gong et al. (2015).

Na Figura 35, verifica-se que em humanos, as proteínas F-Box da família FBXL são aquelas contendo motivos repetidos ricos em leucina, sendo as proteínas da família FBXW aquelas que pertencem à WD40, enquanto outras proteases F-box contêm os domínios de interações entre proteínas existentes e denominadas proteínas da família FBXO (GONG; HUO, 2015; NAKAYAMA; NAKAYAMA, 2006). As proteínas contendo motivos LRRs possuem diversas funções relacionadas às interações proteína-proteína, exercendo a função de receptores hormonais, inibidores enzimáticos, adesão celular e sinalização (KOBE, 2001), incluindo receptores de defesa Toll-likes (TLR) (BELL et al., 2003).

No mosquito *Anopheles gambiae*, as proteínas contendo tioester da família LRRs foram avaliadas quanto ao potencial antagonista para a colonização dos parasitas da malária, limitando a infecção por *Plasmodium* pela ativação de um sistema semelhante ao sistema de defesa complemento (FRAITURE et al., 2009). No carrapato da espécie *I. scapularis*, foram anotadas pelo menos 22 potenciais proteínas LRRs (SMITH; PAL, 2014). No carrapato *H. longicornis*, a proteína recombinante LRR possui uma atividade inibitória no crescimento *in vitro* de

Babesia gibsoni, estabelecendo uma tolerância mútua entre a função anti-babesial e o mecanismo de evasão do sistema imune inato do carrapato (FLORIN-CHRISTENSEN, 2009).

Segundo estes autores, TLRs reconhecem padrões que desempenham um papel vital na indução de genes antimicrobianos em alguns casos, exercendo o controle de imunidade adaptativa de respostas para outros. As proteínas lumican e as proteínas ricas em repetições de leucina atuam com função hemocítica contra diversas bactérias (KAWAI; AKIRA, 2006). Os TLR interagem com diferentes proteínas adaptadoras para ativar as vias MAPK e NF- κ B de uma forma específica e dependente do tipo celular (KAWAI; AKIRA, 2006; SUMBAYEV; YASINSKA, 2006). No geral, o TLR4 reconhece lipopolissacarídeos, enquanto TLR2 interage com muitos produtos microbianos diferentes, incluindo peptidoglicano e lipoproteínas da parede celular (KAWAI; AKIRA, 2006).

As proteínas ubiquitinadas nem sempre são degradadas, com a marcação da ubiquitina podendo ter funções não proteolíticas, também determinando a localização subcelular e atuando como um precursor na modificação pós-translacional das histonas e no reparo do DNA (FRIEDBERG, 2006). A degradação de proteínas está associada à enzima ativadora de ubiquitina (E1), enzima conjugadora de ubiquitina (E2) e ubiquitina ligase (E3) requerendo a participação do proteassoma 26S (COLLINS; GOLDBERG, 2017; ZHAO et al., 2015). Estudo com uma ubiquitina E1 mutante condicional demonstrou a utilidade da técnica para distinguir entre o proteassoma 26S da proteína ubiquitina dependente do proteassoma 20S, e a proteólise pela proteassoma dependente de ubiquitina (SHRINGARPURE; DAVIES, 2002). Entre as diversas famílias de proteínas como dismutases superóxidos, a ubiquitina ligase E3 previamente caracterizada em *I. scapularis* apresenta a capacidade de restringir a colonização bacteriana (SEVERO et al., 2013).

No carrapato *Dermacentor marginatus*, a defensina tem demonstrando uma atividade borrelcida evidente para a remoção de *Borrelia* (CHRUDIMSKÁ et al., 2014). Além disso, há evidências abundantes de que a ubiquitinação sinaliza para as vias do TLR, JAK/STAT e imunodeficiência (IMD) durante o desafio imune contra infecções bacterianas, virais e fúngicas (COLLINS; BROWN, 2010). Segundo estes autores, o circuito imunológico com a ubiquitinação não é exclusivo de *Drosophila* e poderia regular estas vias em artrópodes vetores de doenças, tais como: *A. gambiae*, *Aedes aegypti*, *Culex quinquefasciatus* e *I. scapularis*.

ABC transportadores e outras proteínas putativa na produção de heme e desintoxicação celular

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para cinco sequências da proteína ABC transporte sendo reguladas em sua maioria de forma positiva, havendo transcritos para a sequência da proteína da Subunidade Va do citocromo C oxidase sendo reguladas de forma positiva, e transcritos da sequência da proteína do citocromo P450 sendo regulada de forma positiva. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para cinco sequências desta proteína sendo reguladas de forma positiva, havendo transcritos para a sequência da proteína da Subunidade Va do citocromo C oxidase e do citocromo P450 sendo reguladas de forma negativa. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para uma sequência da proteína ABC sendo regulada de forma negativa, havendo a expressão transcritos para a sequência da proteína da Subunidade Va do citocromo C oxidase sendo regulada de forma negativa.

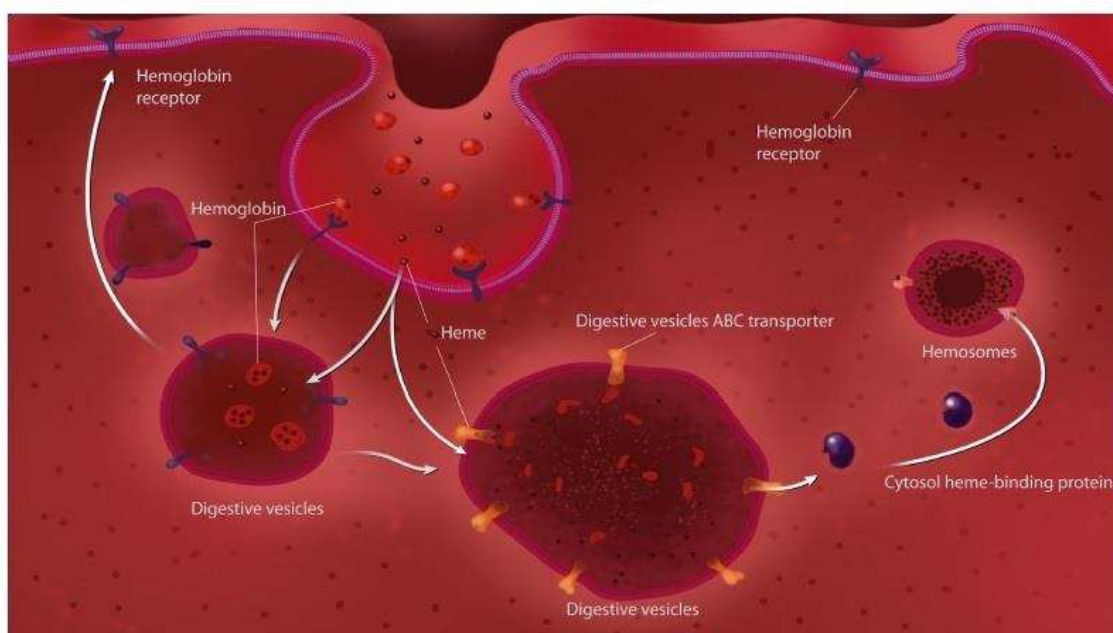
Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para cinco sequências de ABC transporte sendo regulada de forma positiva em ambos os carrapatos, exceto para a regulação de uma sequência na presença desta bactéria no intestino do carrapato *A. sculptum* sendo regulada de forma negativa, havendo a ocorrência de transcritos para a sequência da proteína do citocromo P450 sendo regulado de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*, presentes no módulo de secreção. Ainda nos intestinos dos carrapatos *Amblyomma* spp., observou-se a ocorrência da coexpressão de transcritos para a sequência da proteína do citocromo C sendo regulada de forma no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum* presentes no módulo de oxidação.

Nos intestinos do carrapato da espécie *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão para duas sequências da proteína ABC transporte presentes no módulo de secreção sendo ambas as sequências reguladas de forma negativa na presença de *R. amblyommatis*, havendo a regulação de forma negativa para uma destas sequências e uma regulação de forma positiva para a outra na presença de *R. rickettsii*, havendo transcritos para a sequência da proteína do citocromo C sendo regulados de forma negativa na presença de ambas as bactérias presentes no módulo de oxidação.

As proteínas ABC podem participar do transporte de drogas e metabolitos através da membrana celular, possuindo diversas estruturas e funções, que conferem uma “resistência multidrogas” fenotipicamente celular (HIGGINS, 2007). Atualmente, estas proteínas têm sido relacionadas no desenvolvimento de resistência adquirida por insetos à inseticidas (MOUNSEY et al., 2010; YOON et al., 2011), por helmintos às drogas anti-helmínticas (JAMES; DAVEY, 2009), no combate ao câncer em humanos à quimioterápicos (GOTTESMAN; FOJO; BATES, 2002) e por carrapatos das espécies *R. (B.) microplus* e *Ixodes ricinus* (MANGIA et al., 2018; POHL et al., 2014). A expressão de genes para sequências de proteínas transportadoras ABC ocorre dentro da membrana celular sendo responsável principalmente pela desintoxicação do citoplasma, podendo ser expressas nas membranas de organelas internas, como mitocôndrias e nos desmossomos (DEAN; HAMON; CHIMINI, 2001; LARA et al., 2015).

Os transportadores ABC têm sido principalmente relacionados à exportação de moléculas tóxicas, mas também demonstrou desempenhar um papel na importação de nutrientes e no transporte de outros substratos (POHL et al., 2014). Em carrapatos, as estirpes resistentes à acaricidas aumentam o transporte de moléculas heme e xenobióticas para as vesículas digestivas dos hemossomos (LARA, 2003; LARA et al., 2015). A expressão de ABC transporte e sua presença nos hemossomos no citosol das células digestivas é dedicada principalmente à desintoxicação de grupos heme (LARA et al., 2015) (Figura 36).

Figura 36 Modelo da via do trânsito de heme na célula digestora de carrapato da espécie *Rhipicephalus (B.) microplus*. Este modelo proposto integra dados sobre o papel de transportadores ABC e resultados de relatórios anteriores descrevendo a captação de hemoglobina, seguido pela exportação de heme da vesícula digestiva para a citosol e formação de heme agregados no hemossomo.



Fonte: Lara et al. (2015).

Na Figura 36, verifica-se o bombeamento de compostos tóxicos em uma via intracelular que não ocorre para o meio externo, tal como estudos têm apontado para outros transportadores ABC, como os que atuam nas membranas mitocondriais (LARA et al., 2015). No carrapato bovino *R. (B.) microplus* não ocorre a biossíntese do heme (BRAZ et al., 1999), o que exclui qualquer conexão desta rota de tráfego celular para a via metabólica do heme, indicando um mecanismo ancestral capaz de mediar o transporte de heme através de membranas, semelhante com o que ocorre em bactérias (BRAUN; HANTKE, 2011).

Recentemente, esta descoberta tem sido estendida para outras células eucarióticas, reforçando a hipótese de que, pelo menos, um sistema de transporte heme evoluiu em carrapatos (MANGIA et al., 2018; POHL et al., 2014). Segundo estes autores, a importância da proteína ABC quando expressas em mitocôndrias desempenha um papel central no processo de apoptose, aonde ocorre uma alta geração de espécies reativas de oxigênio desencadeadas pela liberação de citocromo C mitocondrial no citosol. A regulação de genes ABCTs mitocondriais junto com a expressão de isoformas enzimáticas do citocromo P450 pode estar correlacionada com os efeitos de drogas, além de outras moléculas xenobióticas e na geração de espécies reativas de oxigênio (ROS) que comprometem a função mitocondrial (MANGIA et al., 2018; POHL et al., 2014).

5.2.2 Resposta contra vírus

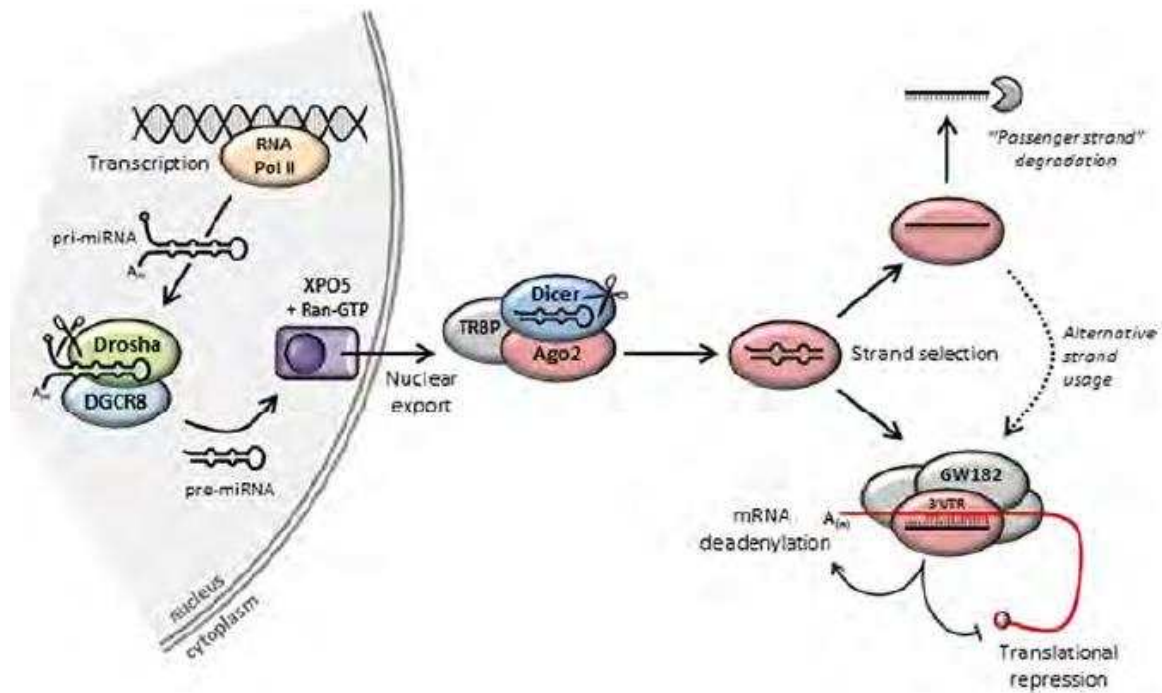
No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para a sequência da proteína DEAD box helicase de RNA dependente de ATP e da proteína Dicer-1 sendo reguladas de forma positiva. No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para duas sequências da proteína Dicer-1 sendo uma regulada de forma positiva e a outra de forma negativa. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatidis*, observou-se a ocorrência da expressão de transcritos para uma sequência da proteína helicase de DEAD box de RNA dependente de ATP sendo regulada de forma negativa.

Nos intestinos dos carrapatos *Amblyommas* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína do fator de reparo

de incisão de DNA ERCC-6 sendo regulada de forma positiva em ambos os intestinos destes ácaros, havendo transcritos para duas sequências da proteína Dicer-1 sendo reguladas de forma positiva no carrapato da espécie *A. aureolatum*, e de forma negativa para uma das sequências com a outra regulada de forma positiva no carrapato *A. sculptum*. Nos intestinos dos carrapatos *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína ERCC-6 sendo regulada de forma negativa na presença da espécie *R. amblyommatis* e de forma positiva na presença da espécie *R. rickettsii*.

A proteína Dicer (Dcr) é uma enzima do tipo RNase III responsável pelo processamento de RNA fita dupla (dsRNA) que atua na formação de RNAi (TIMMONS; FIRE, 1998) entre outros pequenos RNAs inibitórios (siRNAs) (HAMMOND et al., 2000). Os siRNAs são moléculas de RNA de fita dupla implicados no recrutamento de proteínas que formam o complexo ribonucleico silenciador de RNAi (RISC) (YANG; LU; ERICKSON, 2000). Os siRNAs são produzidos à partir do próprio gene alvo, como por exemplo, vírus de cadeia dupla, transposons ou transgene, enquanto os miRNAs, por outro lado, são produzidos à partir de regiões de transcrição auto complementares no genoma que não codificam proteínas (BARTEL, 2004). Os miRNAs são transcritos do genoma para criar miRNAs primários (pri-) com o primeiro ciclo de clivagem ocorrendo no núcleo, realizado pela enzima RNase III Drosha e pelo fator associado Pasha/DGCR8, resultando no miRNA precursor (pre-) (CLAYTON et al., 2018) (Figura 37).

Figura 37 Modelo de produção dos miRNAs pela RNA Pol II e sua passagem pela membrana. (A) Destaque para a exportação nuclear mediada pela Exportin 5 (XPO5) e Ran-GTP. (B) A segunda clivagem é realizada no citoplasma pela enzima Dicer RNase III como parte de um complexo proteico que produz o miRNA



Fonte: Clayton et al. (2018).

Na Figura 37, verifica-se que ocorre a seleção de uma das duas cadeias de miRNA que se liga a proteína Argonauta (Ago) e forma o RISC juntamente com outras proteínas, tais como GW182, alvo para o silenciamento promovendo a degradação de deadenilação e mRNA ou a repressão traducional (CLAYTON et al., 2018). Em insetos, ambos os siRNAs e miRNAs são produzidos pela ação de Dcr, possuindo dois homólogos diferentes, Dcr1 e Dcr2 (FRAGKLOUDIS et al., 2009), havendo as proteínas Ago2 e Dcr2 que são fatores-chave na via de RNAi antiviral (ALIYARI; DING, 2009; DONALD; KOHL; SCHNETTLER, 2012). No genoma do carrapato *I. scapularis*, pesquisas tem revelado diversas proteínas destas famílias por meio de homologia contra as proteínas Dcr e Ago de *Drosophila melanogaster* e *A. aegypti* (SCHNETTLER et al., 2014), as quais estão envolvidas no mecanismo de defesa de RNAi contra a infecção de arbovírus em mosquitos e em carrapatos (KURSCHEID et al., 2009).

Nestes ácaros, verifica-se uma significativa expansão genética para as proteínas da subfamília Ago com atividade antiviral em respostas a aplicação de RNAi de insetos em células do carrapato *I. scapularis* (SCHNETTLER et al., 2014). No entanto, os genomas de *Drosophila* e *Caenorhabditis elegans* podem codificar pelo menos 30 helicases ligadas a caixa DEAD,

enquanto nos genomas de bactérias como *Chlamydia* e *Borrelia*, transmitidas por carrapatos carecem de genes putativos para a proteína helicase DEAD box (KURSCHEID et al., 2009; SCHNETTLER et al., 2014). A helicase II do DNA dependente de ATP tem múltiplas funções no metabolismo do DNA, incluindo replicação, reparo e recombinação do DNA (COOPER; BOYLE; LOVETT, 2015).

A proteína HrpA helicase de caixa DEAD putativa em *B. burgdorferi*, possui um importante papel no centro de via regulatória global da expressão de gene envolvidos com o transporte em células de camundongos (STEIMER; KLOSTERMEIER, 2012). No entanto, o vírus da dengue transmitido por mosquitos das espécies *A. aegypti* e *Aedes albopictus*, possuem a capacidade genética limitada dependente de fatores celulares para completar seu ciclo de vida e facilitar a sua replicação (KHADKA et al., 2011), utilizando das proteínas RNA helicases da família DEAD box para modularem a estrutura do RNA, as quais são cruciais em muitos processos biológicos fundamentais (JANKOWSKY, 2011).

Estas proteínas são relacionadas a mediadores antivirais da imunidade inata, ou fatores essenciais do hospedeiro para a replicação viral (STEIMER; KLOSTERMEIER, 2012). A proteína DDX3 helicase de RNA colocalizada perto do núcleo da célula hospedeira é identificada como um cofator na replicação do RNA viral de Flavivírus durante a infecção pelo vírus do Nilo ocidental (ARIUMI et al., 2007), podendo detectar a presença de pequenas dsRNA com a participação de outras helicases que disparam a sinalização para o sistema de defesa (ZHANG et al., 2011). As proteínas helicases do hospedeiro sinalizariam para uma via antiviral, atuando num papel proviral durante a infecção pelo vírus da dengue (LI et al., 2015), de forma que a regulação de forma negativa da sua expressão poderia facilitar a infecção viral em *Aedes* spp. (FENG et al., 2017).

5.2.3 Oxidação

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para sequências das proteínas D-lactato desidrogenase e proteína mitocondrial peptidase alfa sendo reguladas de forma negativa, enquanto as proteínas oxidase S-2-dridroxi-ácido, citocromo C subunidade Va, dimetilglicina desidrogenase, histona desmetilase de lisina, FAD oxidoreductase e Titin foram reguladas de forma positiva.

No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para as sequências das proteínas da subunidade Va do citocromo C oxidase, NADH-ubiquinona redutase e tioredoxina H2, sendo reguladas de forma negativa, havendo a expressão de transcritos para as sequências das proteínas oxidase S-2-hidroxi-ácido, proteína putativa 1 e 2 do canal de cloro, subunidade alfa da peptidase processadora em mitocôndrias e dimetilglicina desidrogenase, reguladas de forma positiva.

No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para as sequências de proteína NADH-ubiquinona redutase, subunidade Va do citocromo C oxidase, subunidade beta peptidase, oxidase S-2-hidroxi-ácido, glicolato oxidase sendo reguladas de forma negativa, havendo entre estas a expressão de transcrito e regulação positiva para a proteína succinato desidrogenase.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para duas sequências da proteína oxidase S-2-hydroxy-ácido, uma sequência da proteína do canal de cloro 1 e 2, uma sequência de dimethylglycine desidrogenase, sendo reguladas de forma positiva em ambos os ácaros, havendo transcritos para uma sequência de NADH-ubiquinona redutase sendo regulada de forma negativa, e a coexpressão de transcritos para duas sequências das proteínas do citocromo C oxidase, Tioredoxin H2 e da subunidade alfa da peptidase processadora de mitocôndria sendo reguladas de forma negativa no carrapato da espécie *A. sculptum* e de forma positiva no carrapato da espécie *A. aureolatum*, com exceção da peptidase processadora de mitocôndria sendo regulada de forma negativa.

Nos intestinos do carrapato da espécie *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para três sequências da proteína oxidase S-2-hidroxi ácido sendo regulada de forma negativa na presença da espécie *R. amblyommatis* e de forma positiva na presença da espécie *R. rickettsii*, havendo transcritos coexpressos para duas sequências da proteína do citocromo C oxidase e transcritos para uma sequência da proteína redutase NADH-ubiquinona sendo reguladas de forma negativa e transcritos para uma sequência de peptidase processadora de mitocôndria sendo regulada de forma positiva na presença de ambas as espécies de *Rickettsia*.

Muitas espécies de patógenos fazem uso dos componentes do metabolismo do carrapato. Dentre estes, o centro modulador dos processos fisiológicos no ambiente celular, incluindo a

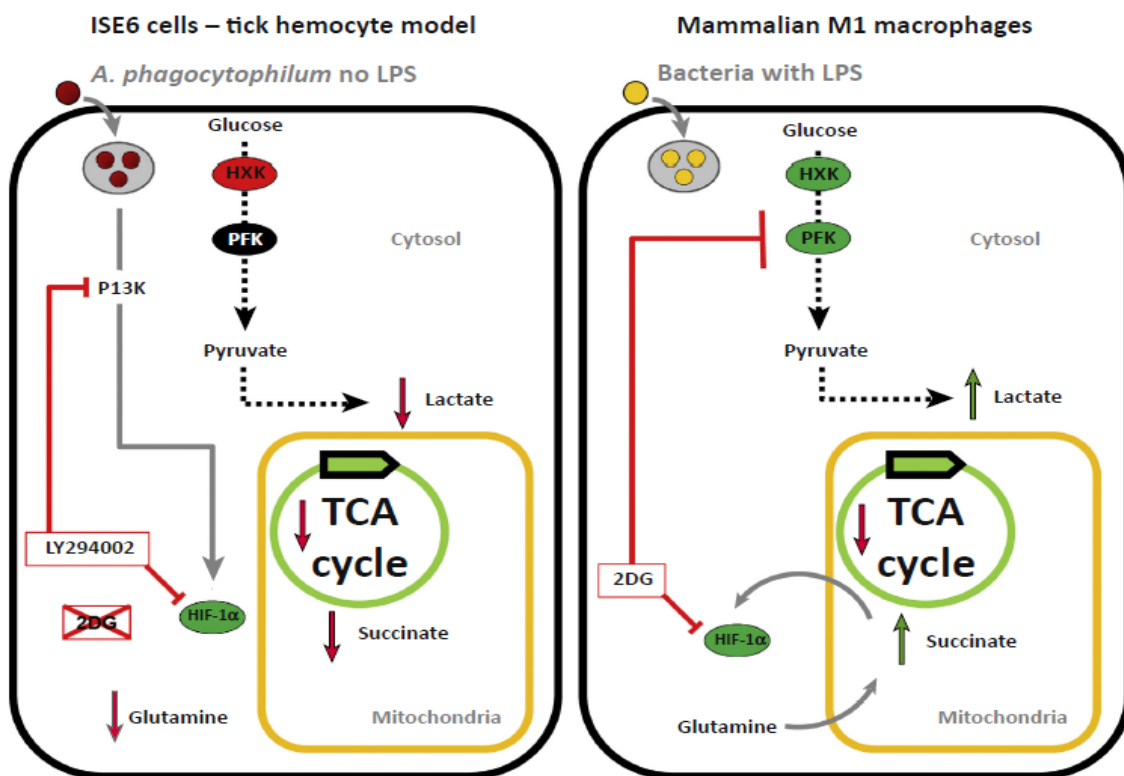
imunidade e a apoptose, dentre outras relações moleculares movidas pela coevolução e adaptação (CABEZAS-CRUZ et al., 2019). O patógeno *Ehrlichia ruminantium* tem a capacidade de detectar os níveis de ferro em células de mamíferos, respondendo com o aumento da expressão de genes de virulência, provenientes do T4SS, que participa na aquisição de ferro pelo patógeno (MOUMÈNE et al., 2018). No carrapato, a proteína de ligação à ferro em resposta a infecção por *A. phagocytophilum* seria hipoteticamente secretada por este sistema, contribuindo para superar a imunidade nutricional nestes ácaros (KAHLON et al., 2013).

A bactéria *B. burgdorferi*, por outro lado, não requer elevadas quantidades alta de ferro para sobreviver (Troxell; XU; YANG, 2012). A hemoglobina hospedeira é clivada a heme por três reações sequenciais envolvendo várias enzimas hemoglobinolíticas em carrapatos, podendo ser modulada por *A. phagocytophilum* por meios de enzimas hemoglobinolíticas (VILLAR et al., 2016). Estes autores sugerem que a infecção por *A. phagocytophilum* pode regular os níveis de heme de uma maneira específica do tecido com efeitos potenciais para o patógeno e a sobrevivência do vetor. Apesar dos macrófagos serem células imunes, divididos em inflamatórios (M1) e anti-inflamatórios (M2), estas células podem ser infectadas por vírus e bactérias intracelulares, as quais as utilizam para sua replicação, ocasionando diferenças nos perfis metabólicos em resposta à estímulos dos patógenos (PRICE; VANCE, 2014).

Segundo estes autores, os lipopolissacárido (LPS) estimulam a via transportadora de glicose nos macrófagos M1 durante o processo de glicólise aeróbica para aumentar o fluxo glicolítico e a produção de lactato. Em contraste, os macrófagos M2 têm um perfil metabólico comparável ao das células não estimuladas, com aumento dos processos do ciclo de ácido tricarboxílico (TCA) e atividade oxidativa. A imunidade celular nos carrapatos é mediada principalmente pelos hemócitos, os quais estão envolvidos na fagocitose de diferentes micróbios (HAJDUŠEK et al., 2013). O estudo do metabolismo de carrapatos em resposta à infecção por *A. phagocytophilum* baseado em células de *I. scapularis* linha ISE6 infectadas com fagócitos, sugerem que o perfil metabólico dos hemócitos em resposta à infecção do patógeno é muito diferente da resposta dos macrófagos inflamatórios M1 (CABEZAS-CRUZ et al., 2017b) (Figura 38).

Figura 38 Modelo do perfil metabólico de células ISE6 do carrapato *I. scapularis* infectadas por *A. phagocytophilum* (bactérias sem LPS) em hemócitos versus o modelo dos macrófagos M1 de mamíferos em resposta a outras bactérias intracelulares com LPS. Altos níveis de hexoquinase (HXK, verde) e fosfofrutoquinase (PFK, verde) estão associados com alta glicólise aeróbica e aumento da produção de lactato em macrófagos M1 de mamíferos. Os níveis e atividade de HXK do carrapato diminuíram (vermelho), e os níveis de PFK do carrapato

não se alteraram (negros) em resposta à infecção que foi associada com níveis mais baixos de lactato em células ISE6 infectadas.



Fonte Cabezas-Cruz et al. (2017b).

Na Figura 38, verifica-se que as células do carrapato *I. scapularis* infectadas *in vitro* por *A. phagocytophilum* não exibem a glicólise aeróbica observada em macrófagos de mamíferos M1 (CABEZAS-CRUZ et al., 2017b). Segundo estes autores, a ativação e estabilização do fator 1 alfa (HIF-1a) estimulado por hipóxia em células de carrapato ISE6 seria não dependente do acúmulo de succinato e da glicólise aeróbica. O inibidor da glicólise 2-desoxiglicose (2DG), que atua na inativação do HIF-1a em M1 macrófagos (NONNENMACHER; HILLER, 2018), não teria o mesmo efeito em células ISE6 (CABEZAS-CRUZ et al., 2017b). Segundo estes autores, em vez disto, a atividade do HIF-1a seria mediada pela ativação da via PI3K, podendo ser inibido por LY294002, o qual diminui sua atividade de ligação ao DNA, nestas células de carrapato. Uma propriedade metabólica compartilhada por macrófagos M1 e células ISE6 seria, no entanto, uma diminuição geral da atividade do TCA após infecção bacteriana destas (CABEZAS-CRUZ et al., 2017b; NONNENMACHER; HILLER, 2018).

A ausência de moléculas imunogênicas, como o LPS e o peptídeo diaminopimélico glicanos em *A. phagocytophilum* e outras bactérias transmitidas por *I. scapularis* foi

anteriormente sugerido para contribuir para a tolerância do vetor a patógenos transmitidos por carrapatos (SHAW et al., 2018). Curiosamente, os lipídios 1-palmitoil-2-oleoil-sn-glicerol-3-fosfoglicerol e 1-palmitoil-2-oleoil diacilglicerol, encontrados em *A. phagocytophilum*, ativam a via IMD em carrapatos (SHAW et al., 2017), podendo regular o metabolismo dos hemócitos do carrapato (CABEZAS-CRUZ et al., 2019). Segundo estes autores, isto demonstra que a tolerância à infecção pode estar associada à imunidade e resistência a patógenos transmitidos em carrapatos, sugerindo que a multiplicação bacteriana dentro das células hospedeiras pode afetar o vetor e, portanto, a sobrevivência do patógeno transmitido por carrapatos.

Notavelmente, pesquisas recentes mostram que o IMD está envolvido na regulação no metabolismo de glicose e lipídios em *Drosophila*, com a ativação constitutiva de IMD induzindo uma redução significativa na expressão de genes envolvidos na glicólise, gliconeogênese e ciclo do TCA em moscas (DAVOODI et al., 2019). O metabolismo lipídico do carrapato é explorado por diferentes microrganismos, os quais se utilizam dos ácidos graxos como importantes fontes de energia (BARISCH; SOLDATI, 2017). A bactéria *B. burgdorferi* que não é capaz de sintetizar ácidos graxos e colesterol, utiliza de todo o aparato da célula do hospedeiro vertebrado e, ou, da farinha de sangue do carrapato para fornecimento de lipídios (HOXMEIER et al., 2017). Similarmente, o colesterol do hospedeiro e outros lipídios são requeridos por *A. phagocytophilum* para infecção e multiplicação em células humanas (TRUCHAN; SEIDMAN; CARLYON, 2013).

A proteína redutase NADH-ubiquinona é uma enzima metabólica conservada localizada na membrana mitocondrial interna podendo catalisar a transferência de elétrons de NADH para a proteína coenzima Q (MERCADO-CURIEL et al., 2014). No intestino médio de carrapato da espécie *D. andersoni* o aumento da expressão de sequência da proteína TC22382 após a alimentação consiste nas necessidades de energia após a ingestão de componentes do sangue, diurese e balanço hídrico (MERCADO-CURIEL et al., 2011). Segundo estes autores, ocorre o aumento da expressão desta proteína NADH-ubiquinona após infecção por *B. bovis* e diminuição nos níveis de mRNA com a infecção por *A. marginale*.

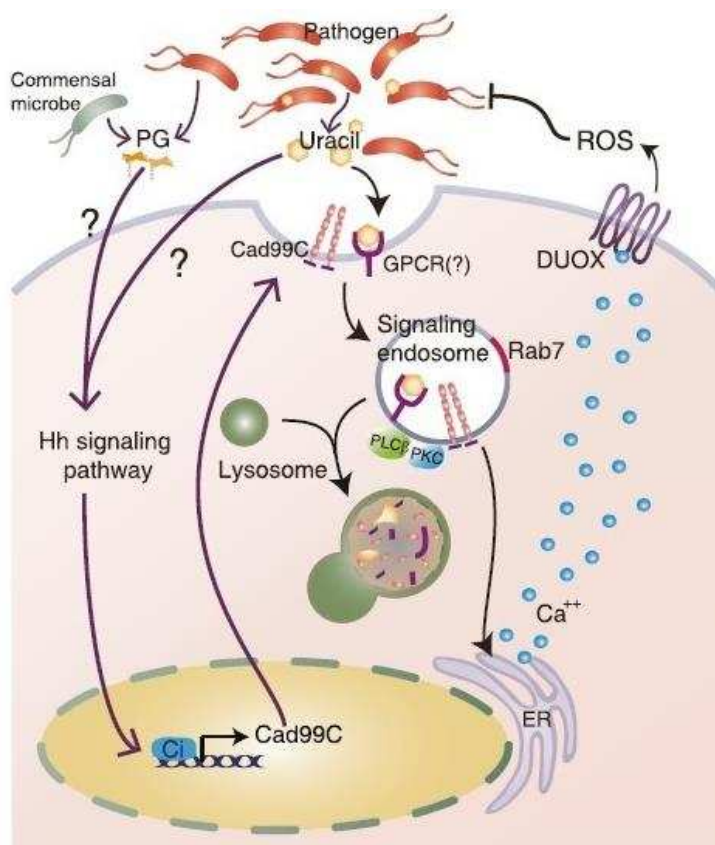
A proteína tioredoxina redutase (TrxR) pertence à família da piridina nucleotídeo-dissulfeto oxiredutase, incluindo também glutatona redutase, lipoamida desidrogenase e íon mercurica redutase (VLAMIS-GARDIKAS; HOLMGREN, 2002). Segundo estes autores, os membros desta família de flavoproteínas se ligariam a dissulfureto ativador de redox e a subunidade do FAD. A TrxR reduz a tioredoxina transferindo um elétron do NADPH ligado

para o sítio ativo da tiorredoxina, permitindo assim que a tiorredoxina reduza suas ligações dissulfeto (ADAMSON et al., 2013). O papel funcional da TrxR tem sido implicado em várias vias metabólicas, incluindo o controle do crescimento do organismo, suas funções imunológicas e anti-apoptose através do sistema tiorredoxina-tiorredoxina redutase (SALZ et al., 1994).

O sialotranscriptoma do carrapato *Amblyomma maculatum* revelou a presença de vários antioxidantes e selenoproteínas, incluindo a proteína TrxR, contribuintes para o sistema antioxidante do carrapato (KARIM; SINGH; RIBEIRO, 2011). As proteínas tirosinas poderiam atuar na mitigação do estresse oxidativo, na manutenção da colonização de patógenos (ADAMSON et al., 2013) e na homeostasia da comunidade bacteriana que colonizam estes ácaros (BUDACHETRI; KARIM, 2015). Os micróbios no intestino de carrapatos facilitam a colonização de agentes infecciosos, como *B. burgdorferi*, evidenciado pela redução na colonização de *Borrelia* após perturbação da microbiota intestinal (NARASIMHAN et al., 2014). O processamento da farinha do sangue envolve uma interação molecular complexa entre o vetor e a microbiota que reside dentro das células vetoriais (BUDACHETRI; KARIM, 2015).

Em moscas, ocorrem dois tipos de resposta imune relacionadas com a detecção e preservação de associações de bactérias benéficas, eliminando ao mesmo tempo aquelas prejudiciais e ativando uma dual oxidase (DUOX) e peroxidases que promovem uma eficaz homeostase para manutenção de colônias de microrganismos benéficos no intestino (LEE; HASE, 2014) (Figura 39).

Figura 39 Modelo de uracil ou peptidoglicano (PG) liberado pelos patógenos para induzir a via de sinalização Hh através de um mecanismo desconhecido, para manter altos nível de Cad99C na membrana apical. Moléculas PG de bactérias comensais também pode desempenhar um papel na manutenção do nível basal apropriado de Cad99C sob uma condição convencional antes da infecção do patógeno. Quando o uracil é reconhecido por um receptor acoplado à proteína G (GPCR), os endossomos de sinalização dependentes de Cad99C/PLCb/PKC facilitam a amplificação do sinal, induzindo a mobilização de Ca^{2+} dependente de PLCb para a geração de ROS dependente de DUOX.



Fonte Lee et al. (2015).

Na Figura 39, verifica-se que a proteína Dual oxidase (DUOX) transmembrana é um membro proeminente da família de proteína oxidase NADPH que se mostrou como fonte na geração de espécies reativas de oxigênio (ROS) (LEE et al., 2015), desempenhando um papel importante na manutenção da imunidade da mucosa (LEE; HASE, 2014). Esta proteína, originalmente descoberta em *C. elegans*, é implicada na geração de ROS que age na manutenção da microbiota intestinal em *A. gambiae* (VIJAY-KUMAR et al., 2010). Segundo estes autores, em mosquitos a DUOX durante a alimentação, juntamente com uma peroxidase heme, catalisa uma reticulação de tirosinas, formando uma barreira molecular e acelular no espaço luminal ao longo da camada epitelial do intestino denominada de rede de ditirosina (DTN).

Esta rede diminui a permeabilidade do intestino com vários elicitores imunológicos, protegendo-o da microbiota intestinal contra espécies comensais e patogênicas (DIAS et al., 2013). Segundo estes autores, nos carrapatos, algumas enzimas antioxidantes desempenham um papel importante na colonização de patógenos que manipulam a resposta imune, e dentre estes a expressão da DUOX durante a alimentação precoce do carrapato coincidindo com a replicação de bactérias no intestino apresentando uma rede distinta de ditirosina. O intestino dos carrapatos contém proteínas envolvidas na produção de ROS, o que também pode ajudar a controlar o crescimento de microrganismos (ANDERSON; SONENSHINE; VALENZUELA, 2008).

No genoma do *I. scapularis* foi anotado uma única DUOX além de várias proteínas peroxidase. A DUOX pertence à família da nicotinamida adenina dinucleotídeo fosfato (NADPH) e oxidase NOX (GEISZT et al., 2003), desempenhando um papel essencial na imunidade e homeostase da mucosa intestinal do carrapato (BAE; CHOI; LEE, 2010; YANG et al., 2014). Estes autores sugerem que peroxidases, assim como a DUOX, desempenham um papel importante na formação da DTN, influenciando na persistência dos patógenos da doença de Lyme nos carrapatos, com a participação de NADPH oxidase.

5.2.4 Proteólise

No intestino do carrapato *A. aureolatum* infectado por *R. rickettsii*, observou-se a ocorrência da expressão de transcritos para as sequências de proteínas ATPase transcricional do retículo endoplasmático, subunidade alfa 7 do proteassoma 26S, subunidade alfa 6A da proteína reguladora do proteassoma 26S e subunidade reguladora beta2 do proteassoma 20S sendo reguladas de forma negativa, enquanto as proteínas reguladas de forma positiva foram a subunidade reguladora beta do proteassoma 20S, fator de modulação TATA, fator de biogênese peroxissomos protease, subunidade reguladora tipo alfa PSMA7 / PRE6 do proteassoma 20S, protease dependente de ATP PIM1 / LON e proteína sensível ao nethylmaleimide.

No intestino do carrapato *A. sculptum* infectado por esta mesma bactéria, observou-se a ocorrência da expressão de transcritos para a sequência das proteínas ATPase, subunidade alfa 7 do proteassoma 26S, subunidade alfa 6A da proteína reguladora do proteassoma 26S, subunidade reguladora beta do proteassoma 20S, subunidade reguladora tipo alfa PSMA7 / PRE6 do proteassoma 20S e ATPaseRPT6 do complexo regulador do proteassoma 26S sendo

reguladas de forma negativa, havendo transcritos para as sequências de proteínas do fator de biogênese peroxissômica, ATPase transcricional do retículo endoplasmático, subunidade reguladora beta do proteassoma 20S, protease tipo Clp dependente de ATP e protease dependente de ATP PIM1/LON sendo reguladas de forma positiva. No intestino do carrapato *A. sculptum* infectado por *R. amblyommatis*, observou-se a ocorrência da expressão de transcritos para duas sequências da proteína da subunidade reguladora beta do proteassoma 20S sendo uma regulada de forma negativa e a outra de forma positiva.

Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína ATPase transcricional do retículo endoplasmático sendo regulada de forma negativa no carrapato *A. aureolatum* e de forma positiva no carrapato *A. sculptum*, uma sequência da proteína protease sendo regulada de forma negativa no carrapato *A. sculptum* e de forma positiva no carrapato *A. aureolatum*, havendo transcritos para a sequência da proteína protease dependente de ATP PIM1/LON e fator de biogênese peroxissômica sendo regulados de forma positiva em ambos os carrapatos. Ainda nos intestinos dos *Amblyomma* spp., observou-se a coexpressão de transcritos para a sequência das proteínas ATPaseRPT6 do complexo regulador do proteassoma 26S sendo regulada de forma positiva no carrapato *A. aureolatum* e de forma negativa no carrapato *A. sculptum*, havendo transcritos para a subunidade alfa 7 do proteassoma 26S sendo regulados de forma negativa em ambos os carrapatos.

Além disto, observou-se a ocorrência da coexpressão de transcritos para as sequências das proteínas da subunidade reguladora alfa do proteassoma 20S sendo regulados de forma positiva em ambos os carrapatos, havendo transcritos para a subunidade reguladora beta do proteassoma 20S sendo regulados de forma negativa no carrapato *A. sculptum* e de forma positiva no carrapato *A. aureolatum*. Nos intestinos do carrapato da espécie *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para as sequências das proteínas da subunidade reguladora alfa do proteassoma 20S sendo regulados de forma positiva na presença de ambas bactérias, havendo transcritos para a subunidade reguladora beta do proteassoma 20S sendo regulados de forma negativa na presença de ambas bactérias presente no módulo de proteólise.

Um dos principais caminhos para a degradação de proteínas na célula é o sistema de proteassoma dividido em duas categorias principais, sendo uma dependente de ubiquitina pelo proteassoma 26S e a outra independente de ubiquitina pelo proteassoma 20S (LÜLLMANN-

RAUCH, 2007). O terceiro, é um ativador de ligação do proteassoma do núcleo 20S ao 26S, sendo o proteassoma 19S responsável pela degradação de ubiquitina dependente de proteínas de atividade de ATPase AAA (LASKER et al., 2012). Além de reconhecer substratos ubiquitinados e transportados para o proteassoma 20S, o anel de ATPase ativa chaperonas com a hidrólise de ATP para o dobramento final de substratos durante a sua translocação para a câmara proteolítica deste proteassoma (RECHSTEINER; HILL, 2005).

Os proteassomas ao reduzirem a velocidade de degradação de material estranho como vírus ou bactérias, expõem por mais tempo o ataque microbiano e aumentam a capacidade de sinalizarem para apresentação de antígenos (KISSELEV et al., 1999). O aumento robusto das proteínas chaperonas em resposta ao estresse oxidativo é rigidamente regulado por mecanismos moleculares que incluem a ubiquitinação e não a oxidação (KÄSTLE et al., 2012). O proteassoma 20S é relativamente estável após a exposição H₂O₂, sendo responsável pela degradação de proteínas oxidadas (BAUGH; VIKTOROVA; PILIPENKO, 2009) como a proteína estimulante de apoptose p53 (ASHER et al., 2005), enquanto o proteassoma 26S é inativado (REINHECKEL et al., 1998). Segundo estes autores o proteassoma 20S, além de gerenciar a homeostase proteica durante o estresse oxidativo, também degrada muitas proteínas envolvidas na sinalização e regulação de processos celulares.

A proteína canônica Lon de *E. coli* é conservada e exerce atividade de protease e degradação intracelular de proteínas anormais (GOLDBERG et al., 1994), podendo se ligar ao DNA (ZEHNBAUER et al., 1981) e regular a atividade de chaperonas de choque térmico (GOFF; CASSON; GOLDBERG, 1984). A proteína Lon é uma serina protease e desempenha um papel importante na homeostase ao segmentar proteínas anormais e proteínas reguladoras instáveis (GOTTESMAN; HALPERN; TRISLER, 1981). A presença das proteínas Lon1 e Lon2 em *B. burgdorferi*, sendo reguladas por proteínas no sangue, sugeriu que Lon2 é envolvida no controle da homeostase, enquanto Lon1 desempenha um papel na adaptação da bactéria durante a transição do patógeno para o hospedeiro mamífero durante o período de hematofagia em decorrência da picada do carrapato (COLEMAN et al., 2009).

Várias proteínas derivadas de *D. variabilis* foram associadas à resposta do carrapato à infecção por *Rickettsia* (MACALUSO et al., 2001; MULENGA et al., 2003). Entre as moléculas putativamente identificadas e diferencialmente transcritas em resposta à infecção de *Rickettsia* tem-se a proteína ATPase vacuolar (V-ATPase) (PETCHAMPAI et al., 2014a). Apesar da variabilidade de subunidades em seus domínios, esta enzima utiliza a hidrólise de

ATP para acidificar compartimentos de endocitose mediada por receptor, tráfego intracelular e degradação de proteínas (WIECZOREK et al., 2009). Vários patógenos bacterianos requerem as proteínas V-ATPases nas células hospedeiras para facilitarem a vida intracelular, como, por exemplo, em *Listeria monocytogenes*, que depende desta proteína para a fuga fagossômica (PORTNOY et al., 1992), e na invasão de células de carrapatos por *A. marginale* (KOCAN et al., 2009). A acidificação de vacúolos não impede a infecção por *Rickettsia* do Grupo das Febres Maculosas, utilizando de proteínas V-ATPases em fossos membranares revestidos por clatrina durante a invasão de células hospedeiras (MUNDERLOH, 1995; WELCH; REED; HAGLUND, 2012).

5.2.5 Fosforilação

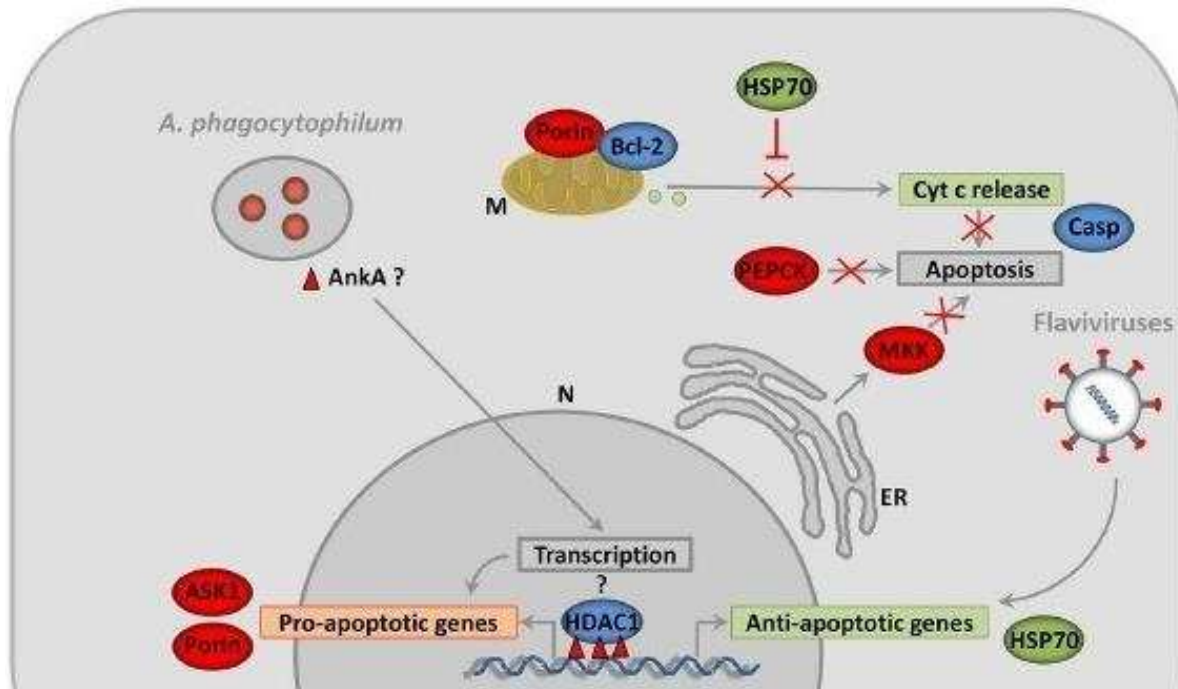
Nos intestinos dos carrapatos *Amblyomma* spp. infectados por *R. rickettsii*, observou-se a ocorrência da coexpressão de transcritos para uma sequência da proteína Tirosina quinase, Fosfatase 2B, ativadores de Mitogen quinase MKK4, Lisozima, DnaJ (Hsp40), proteína quinase de fator de morte, Ciclin T dependente quinase CDK9 e calmodulin sendo reguladas de forma positiva para ambos os carrapatos.

Nos intestinos do carrapato da espécie *A. sculptum* infectados por espécies de *Rickettsia*, observou-se a ocorrência da coexpressão de transcritos para as sequências das proteínas da subunidade reguladora alfa do proteassoma 20S sendo regulados de forma positiva na presença de ambas bactérias, havendo transcritos para a subunidade reguladora beta do proteassoma 20S sendo regulados de forma negativa na presença de ambas bactérias presente no módulo de proteólise.

A capacidade de *A. phagocytophilum* de regular negativamente a expressão gênica em neutrófilos foi associada ao recrutamento de HDAC1 para os promotores de genes alvo pela proteína contendo motivos de repetição de anquirina AnkA (RENNOLL-BANKERT et al., 2015). Este patógeno ao infectar as glândulas salivares, inibe a morte celular programada, diminuindo a expressão dos genes pro-apoptóticos como ASK1 e Porin, havendo a regulação negativa de Porin (AYLLÓN et al., 2015) associada à inibição de liberação mitocondrial do citocromo C (Cyt c) (DE LA FUENTE et al., 2017) (Figura 40).

Figura 40 Modelo de diferentes mecanismos para a inibição de apoptose de células hospedeiras por patógenos. N, núcleo; M, mitocôndria; ER, retículo endoplasmático; Cyt c, Citocromo c; ASK1, quinase 1 reguladora de sinal

de apoptose; MKK, proteína quinase ativada por mitógeno; HDAC1, histona desacetilase 1; AnkA, Ankyrin A; PEPCK, Fosfoenolpiruvato Carboxiquinase; FOXO, caixa de garfo O; Hid, Involução da cabeça defeituosa; JNK, junho amino-terminal quinases; Casp caspases.



Fonte - De la Fuente et al. (2017).

Na Figura 40, verifica-se diferentes comportamentos entre as moléculas e processos representados em verde sendo regulados de forma positiva, tanto quanto negativa para moléculas em resposta à infecção em vermelho, variando a resposta à infecção com as moléculas em azul (DE LA FUENTE et al., 2017). A bactéria *A. phagocytophilum* induz o estresse no ER em células de carrapatos ocasionando a redução dos níveis de MKK que inibe a apoptose (VILLAR et al., 2015). Segundo estes autores, outro mecanismo interessante de *A. phagocytophilum* para inibir a apoptose é a manipulação do metabolismo da glicose, reduzindo a níveis de PEPCK. No entanto, a infecção de células de carrapatos com flavivírus resulta na regulação positiva de genes como Hsp70 que inibem a apoptose (MANSFIELD et al., 2017).

5.3 Evolução de proteínas de *Rickettsia* secretadas, excretadas e ou de superfície

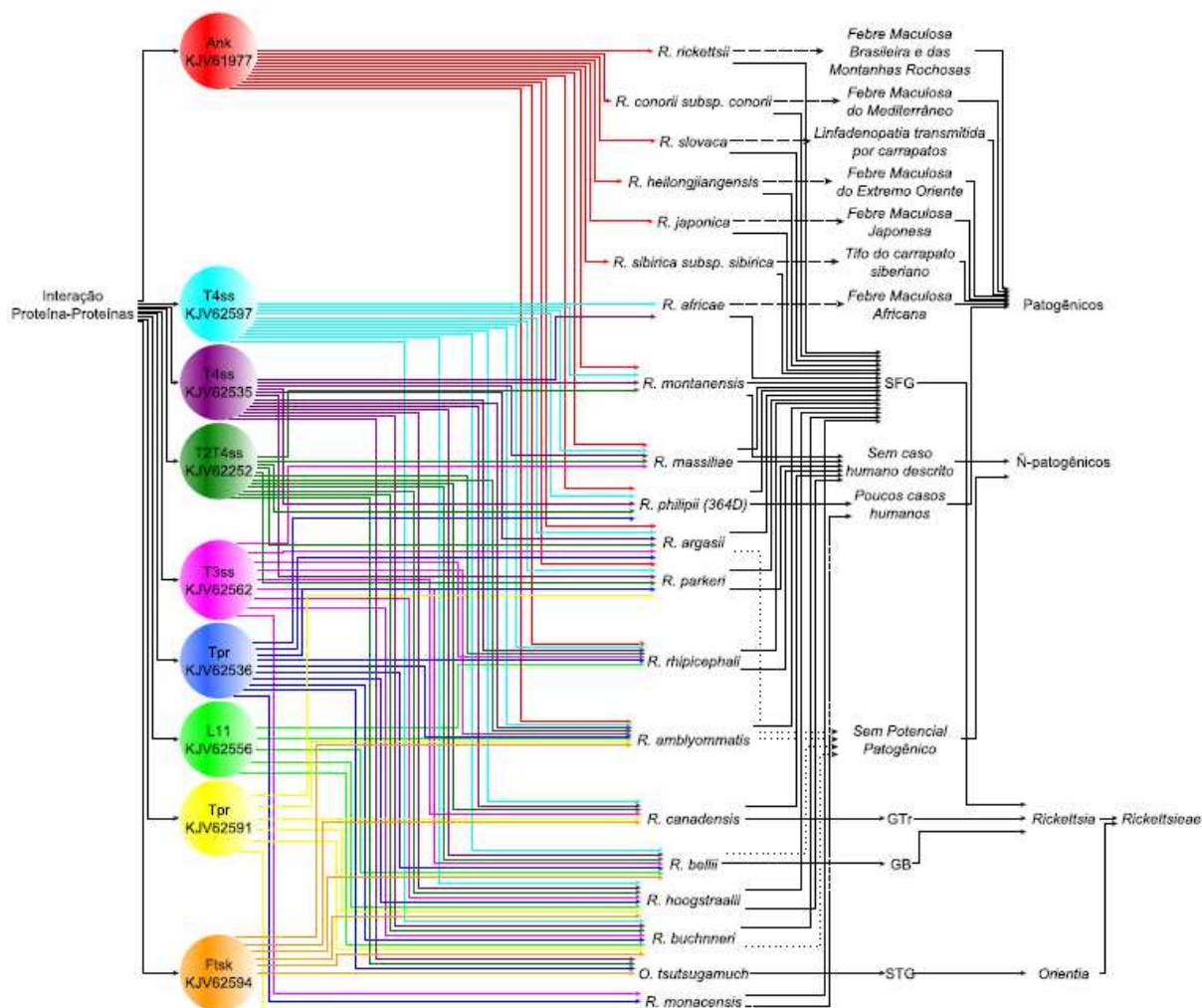
As anotações de genes são notoriamente difíceis em *Rickettsia* por causa do processo de degradação de gene, produzindo numerosos exemplos de um gene em uma espécie que é interrompida em vários fragmentos curtos em outras espécies. Por meio da análise de similaridade, obteve-se as seqüências de aminoácidos de proteínas similares as 91 proteínas de

membrana da espécie *R. amblyommatidis* que interagem com as proteínas do carrapato (Dados suplementares S1P) (Tabela 38). Diante deste quadro, obteve-se dentre as proteínas de *R. amblyommatidis* que interagiram com as proteínas do carrapato *I. scapularis* presentes no interatoma, algumas árvores filogenéticas (Dados não publicados).

Dentre os resultados das árvores filogenéticas utilizando de 91 sequências de proteínas do interatoma de membrana de *R. amblyommatidis* contra as proteínas provenientes de 75 genomas de bactérias, observou-se que 13 (15,85%) das árvores de sequências de proteínas de membrana possuem sequências similares e ortólogas para espécies de *Rickettsia* que infectam carrapatos, não havendo entre estas a presença de sequências de proteínas de espécies que infectam pulgas e, ou piolhos, ou qualquer outro inseto, podendo ocorrer a presença de proteína da espécie *Orientia tsugamush* que infecta ácaros, possuindo um elo evolutivo com *Rickettsia*, sugerindo haver um *status* de coevolução para estas sequências que interagiram e são moduladoras de determinadas vias moleculares entre *Rickettsia/Carrapatos*.

As árvores filogenéticas de proteínas de membrana de *Rickettsia* presentes no interatoma que continham apenas sequências de aminoácidos de proteínas pertencentes de *Rickettsia* capazes de infectar os carrapatos foram as seguintes: sequências de proteínas contendo domínios de repetição de tetratricopeptídeo (Tpr) IDs KJV62536, KJV62591.1 e da sequência de proteína chaperona do sistema de secreção do tipo III (T3SS) ID KJV62562.1; as sequências de proteínas responsáveis pela formação do sistema de secreção do tipo IV (T4SS) KJV62253.1, KJV62505.1, KJV62535.1, KJV62563.1, KJV62597.1 e do tipo II/IV (T2/T4SS) KJV62252.1; a sequência de proteína contendo motivos de repetição de anquiria (Ank) KJV61977.1; e as sequências de proteínas ribossomais metiltransferases (L11) KJV62584.1 e KJV62556.1 (Figura 41).

Figura 41 Fluxograma de proteínas ortólogas de membrana para as espécies de *Rickettsia* que infectam somente carrapatos. Destaque para as diferenças de patogenicidade e grupos de classificação taxonômica entre as espécies e a presença das proteínas, incluindo espécime do gênero *Orientia*, sendo este um Grupo Ancestral irmão do gênero *Rickettsia*.



Fonte: Autor deste trabalho.

Na Figura 41, verifica-se o fluxograma para as proteínas de membrana contendo ortólogos entre as espécies de *Rickettsia* que infectam carrapatos Ixodídeos e Argasídeos, além de *Orientia* que infecta ácaros. As árvores filogenéticas contendo sequências de ortólogos estariam entre as proteínas mais conservadas da espécie *R. amblyommatis* e espécies de *Rickettsia* que infectam os carrapatos. Visto aqui, também como ácaros há dois fatos importantes entre as relações ecológicas e evolutivas para estas espécies relacionadas, aonde no primeiro, os carrapatos e ácaros estariam agrupados na família Ácari e o outro abrange a evolução de *Rickettsia* e *Orientia* que pertencerem a família Rickettsieae.

Isto demonstra claramente indícios de tal fato, que ocorra e ou., esteja caminhando paralelamente em uma relação coevolutiva entre estes organismos, os quais possuindo em comum um ancestral, que no passado poderia ter adquirido também o ancestral entre estes dois gêneros. No entanto, é sensato colocar que outras bactérias pertencentes a Rickettsieae que não pertencem a *Rickettsia* e ou *Orientia*, também são transmitidas somente por carrapatos. Neste sentido, para as nossas árvores que possuem sequências de proteínas ortólogas somente de espécies que infectam os carrapatos, as proteínas dos organismos de *Orientia* são os últimos que possuiriam parentesco direto com estas sequências.

Na grande maioria das árvores filogenéticas, foi observado o agrupamento entre as proteínas que pertencem as espécies *R. belli* e *Rickettsia canadenses*, do Grupo Ancestral que infectam os carrapatos. De certa forma, algumas das proteínas destes organismos estão presentes entre as que possuem ortólogos para as espécies que infectam somente os carrapatos, demonstrando que estas proteínas poderiam ter seguido uma linha evolutiva em carrapatos. O que reforçaria a hipótese de coevolução entre estas proteínas e os carrapato seria o fato deste grupo ser seguido do Grupo do Tifo, apresentando separadamente as proteínas para as espécies *R. prowazekii* e *R. typhi*, que infectam ácaros e especialmente piolhos. Estas proteínas estariam muito distante do *statu* coevolutivo entre *Rickettsia* e carrapato. No entanto, a sua ausência nestas árvores de proteínas ortólogas e entre as espécies que infectam os carrapatos e, ou, os seus agrupamentos de forma correta no tempo evolutivo em outras árvores de proteínas de membrana. Isto demonstrou claramente que estas proteínas percorreram caminhos evolutivos diferentes, o que poderia ser também abrangido num contexto coevolutivo para os artrópodes hematófagos como o ambiente, aonde estes organismos poderiam ter evoluído.

Além destes grupos bem definidos, averiguamos a presença de proteínas do grupo de Transição com as espécies e, além das espécies. Este grupo apresenta bactérias que infectam ácaros *R. akari*; Ixodideo *R. australis*; Argasideo *R. hoogstraalii*; e pulgas *R. felis* e *R. asemboensis*, havendo uma mistura entre espécies de *Rickettsia* pertencentes ao Grupo das Febres Maculosas e Grupo do Tifo, sendo possível averiguar para as sequências de proteínas uma divergência entre a formação deste grupo, demonstrando haver de fato uma transição entre os grupos das espécies que infectam os carrapatos dos insetos. Dentre as árvores de proteínas ortólogas, somente sequências para a espécie *R. hoogstraalii* dentre estas foram observadas.

As demais sequências de proteínas abrangeram a maior diversidade e riqueza de espécies de *Rickettsia* que infectam os carrapatos pertencentes ao Grupo das Febres Maculosas.

Dentre as espécies que pertencem a este grupo mais recente evolutivamente em relação aos outros, observou-se uma separação de árvores filogenéticas entre as espécies patogênicas e não patogênicas. Neste estudo observamos que após a formação do clado das proteínas do Grupo de Transição a ocorrência do clado entre as proteínas pertencentes as bactérias que infectam especificamente o *I. scapularis* com a presença de *R. endonssibionte I. scapularis* e *R. bucheneri*, acompanhadas de *R. endonssibionte I. pacificus* além da *R. monacensis* isolada em *I. ricinus*, aonde a maioria não é patogênica, ou com poucos casos humanos relatados, sugerindo haver dentro deste gênero de carrapato um ambiente aonde estas espécies de *Rickettsia* poderiam ter evoluído, pertencentes a uma classificação antiga antes denominada Grupo da *Rickettsia helvética*. Dentre as árvores de proteínas ortólogas, praticamente todas estas espécies de *Rickettsia* tiveram proteínas representadas nestes cladogramas.

Prosseguindo, observamos a ocorrência das proteínas das espécies *R. montana*, *R. rhipicephalis*, *R. raoult*, *R. amblyommatis* e *R. massilea*. Esta última anteriormente dava nome ao Grupo da *Rickettsia massilea*. Dentre as árvores de proteínas ortólogas, somente a espécie *R. raoult* não tiveram proteínas representadas nestes cladogramas. No entanto, na ponta dos cladogramas observamos o restante das proteínas das espécies incluindo *R. japonica*, *R. argasi*, *R. heilongjiangensis*, *R. slova*, *R. sibirica*, *R. africae*, *R. peacocki*, *R. conorii*, *R. parkeri* e *R. rickettsii*, sendo estas a mais patogênica do Grupo das Febres Maculosas. Dentre as árvores de proteínas ortólogas, praticamente todas estas espécies de *Rickettsia* tiveram proteínas representadas nestes cladogramas. No entanto houve árvores de proteínas ortólogas para estas espécies que infectam somente os carrapatos que se divergiram entre as espécies patogênicas e não patogênicas.

Para finalizarmos, algumas árvores apresentaram em seus clados as proteínas mais conservadas filogeneticamente entre outras bactérias *Rickettsiales* e *Anaplasma* com a presença separada de clado único para *Orientia tsugamichi*, e *NeoRickettsia risticii*, além de *Anaplasma phagophilum*, *Ehrlichia chafeensis*, *Wolbachia pipiensis* e *Wolbachia endosymbiotic*.

As árvores filogenéticas para estas proteínas sugeriram que as pressões seletivas do hospedeiro contribuíram para o estado de evolução destas sequências. Isto até mesmo, para as proteínas não ortólogas de *Rickettsia* que infectam os carrapatos, mas também para outras bactérias mais distantes que também evoluíram dentro de artrópodes. Diante disso, seria possível generalizar os resultados obtidos para diversas espécies de carrapatos hospedeiros de

Rickettsia que contribuíram possivelmente para a evolução de proteínas de membrana de Rickettsieae, abrangendo até as proteínas de *Orientia* seu primeiro parente próximo, com as quais poderiam ativar e participar de vias funcionais e metabólicas essenciais nas células destes hospedeiros da família Ácari.

Diversos autores familiarizados com o estudo da relação parasito hospedeiro entre estes organismos, relataram a ocorrência de *Rickettsia* infectando uma rica diversidade de carrapatos hospedeiros (PAROLA et al., 2013). Fazendo um gancho com a riqueza e diversidade de hospedeiros carrapatos, dentre as espécies de *Rickettsia* que os infectam e apresentaram árvores filogenéticas contendo sequências de proteínas ortólogas entre as espécies que infectam estes ácaros, têm-se:

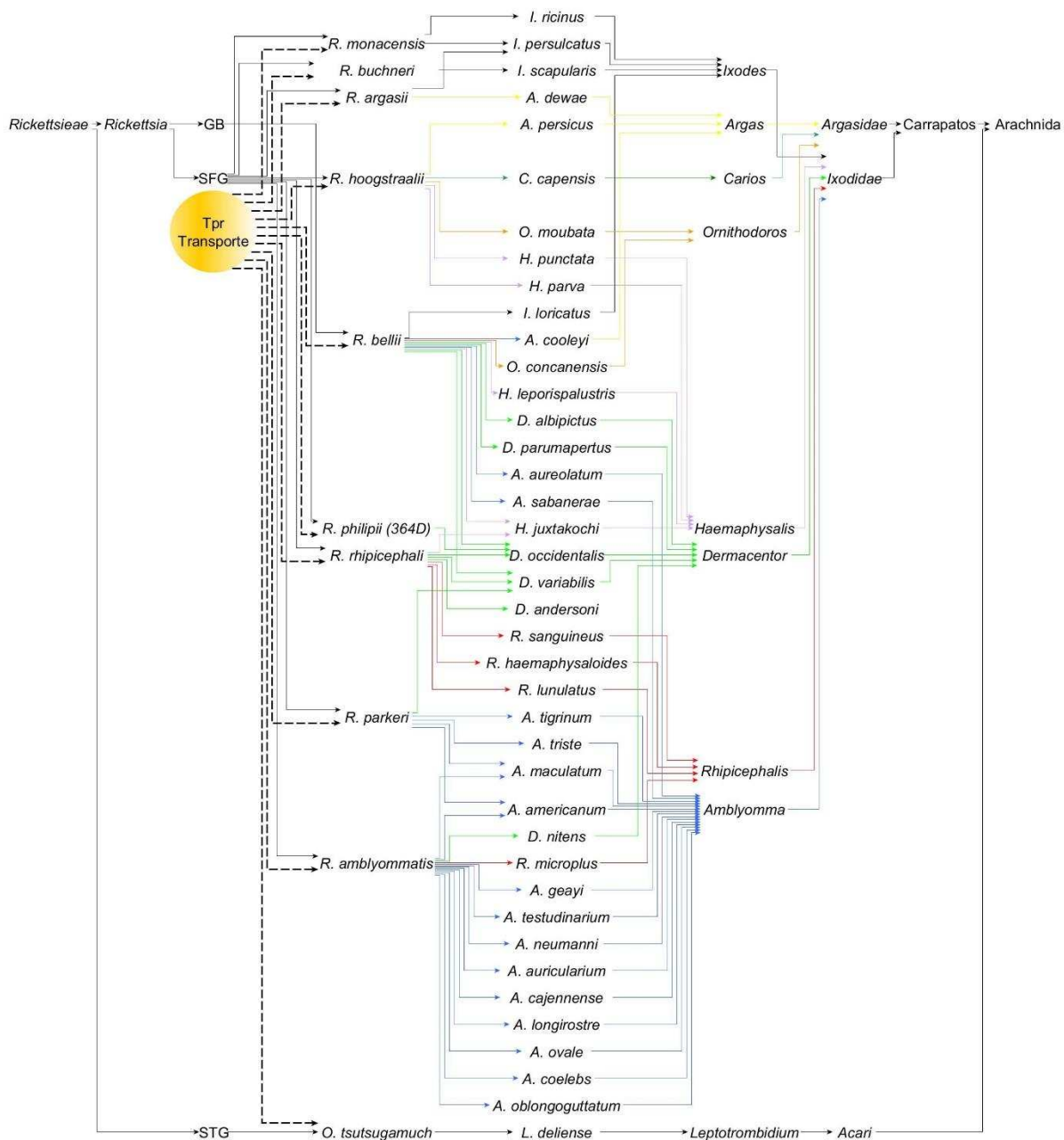
- A espécie *R. hoogstraalli* infecta os carrapatos das espécies *Carios capensis* (Japão Ásia), *Argas persicus*, *Ornithodoros moubata* (África subsariana) pertencentes aos Argasídeos, além de carrapato das espécies *Haemaphysalis punctata* e *Haemaphysalis sulcata* (Europa) pertencentes ao Ixodídeos.
- A espécie *R. argasii* infecta o carrapato da espécie *Argas dewae* (Austrália) (Argasídeo).
- A espécie *R. amblyommatis*, infecta diversas espécies de carrapatos *Amblyomma* spp., dentre elas, as espécies *Amblyomma americanum*, *Amblyomma cajennense*, *Amblyomma oblongoguttatum*, *Amblyomma ovale*, *Rhipicephalus microplus*, *Rhipicephalus sanguineus*, *Dermacentor nitens*, e *Amblyomma maculatum* (EUA, Costa Rica, e Panamá); sendo encontrada de forma mais comum no carrapato da espécie *A. americanum* (EUA, América do Norte e Central), além das espécies *Amblyomma cajennense*, *Amblyomma coelebs*, *Amblyomma neumanni*, *Amblyomma longirostre*, *Amblyomma geayi*, *Amblyomma auricularium* (Brasil, Argentina e Guiana Francesa).
- A espécie *R. monacensis*, infecta os carrapatos das espécies *Ixodes ricinus* (Europa), *Ixodes persulcatus* (Turquia), além de outras espécies de *Ixodes* spp. (Tunísia, Algeria, e Marrocos Norte da África).
- A espécie *R. parkeri*, infecta os carrapatos das espécies *Amblyomma maculatum*, *Amblyomma americanum*, *Dermacentor variabilis* (Sul dos EUA, América Central e Norte), além das espécies *Amblyomma triste*, *Amblyomma tigrinum* (Uruguai, Argentina, Brasil e Bolívia na América do Sul).
- A espécie *R. canadenses*, infecta os carrapatos das espécies *Haemaphysalis leporispalustris* (Canada e EUA, América do Norte e Central).

- As espécies REIS e *R. buchnerii* infectam o carrapato da espécie *I. scapularis* (América Central e do Norte).
- A espécie *R. rhipicephali*, infecta os carrapatos das espécies *Rhipicephalus haemaphysaloides* (Taiwan Ásia), *Rhipicephalus sanguineus*, *Rhipicephalus lunulatus*, (África Sub-Saharan), *Rhipicephalus sanguineus*, *Dermacentor occidentalis*, *Dermacentor variabilis*, *Dermacentor andersoni* (EUA América do Norte e Central), *Haemaphysalis juxtakochi* (Brasil Floresta Amazônica e Atlântica América do Sul).
- A espécie *R. philipe*, infecta o carrapato da espécie *Dermacentor occidentalis* (Califórnia, América do Norte e Central).
- A espécie *R. belli*, infecta os carrapatos das espécies *Dermacentor marginatus*, *Dermacentor reticulatus*, *Ixodes ricinus* SENLAT (old TIBOLA/DEBONEL – Europa) *Dermacentor silvarum*, *Dermacentor reticulatus*, *Dermacentor marginatus*, *Dermacentor nuttalli*, *Dermacentor niveus* (Norte da Ásia), *Haemaphysalis ornithophila*, *Haemaphysalis shimoga*, *Haemaphysalis lagrangei*, *Amblyomma testudinarium* (Sul da Ásia), havendo espécies de *Dermacentor* spp. (Marrocos North África).
- A espécie *R. argasi* infecta o carrapato da espécie *Argas dewae* (Argasídeo).
- A espécie *R. sibirica* subsp. *Sibirica*, infecta os carrapatos das espécies *Dermacentor nuttalli*, *Dermacentor marginatus*, *Dermacentor reticulatus*, *Dermacentor silvarum*, *Dermacentor sinicus*, *Haemaphysalis yeni*, *Haemaphysalis concinna*, e *Ixodes persulcatus* (Sibéria Rússia, China, e Mongólia na Ásia).
- A espécie *R. japonica*, infecta *Haemaphysalis flava*, *Haemaphysalis hystricis*, *Haemaphysalis longicornis*, *Haemaphysalis cornigera*, *Haemaphysalis formosensis*, *I. ovatus*, *D. taiwanensis* (Japão e Korea do Sul na Ásia).
- A espécie *R. conorii*, infecta *Rhipicephalus sanguineus* (Mediterrâneo Europa, Norte da África). A espécie *R. montanna*, infecta carrapatos das espécies *Dermacentor andersoni* (EUA e Canada) e *Dermacentor variabilis* (América do Norte e Central).
- A espécie *R. heilongjiangensis*, infecta os carrapatos das espécies *Haemaphysalis concinna*, *Haemaphysalis japonica douglasi*, *Haemaphysalis flava*, *Dermacentor silvarum* (Rússia, China, Korea do Sul, Japão na Ásia).
- A espécie *R. massiliae*. infecta os carrapatos das espécies *Rhipicephalus sanguineus* (Arizona e Califórnia Norte e Centro dos EUA).

- A espécie *R. slova*, infecta os carrapatos das espécies *Dermacentor marginatus*, *Dermacentor reticulatus* SENLAT (TIBOLA/DEBONEL Europa), havendo espécie de *Dermacentor* spp. (Ásia; Rússia, China) e *Dermacentor marginatus* SENELAT (TIBOLA DEBONEL – Algeria, Marrocos no Norte da África).
- A espécie *R. rickettsii*, infecta os carrapatos das espécies *Dermacentor andersoni*, *Dermacentor variabilis*, *Dermacentor occidentalis*, *Dermacentor nitens*, *Rhipicephalus sanguineus*, *Amblyomma cajennense*, *Amblyomma americanum* *Amblyomma imitator*, *Haemaphysalis leporispalustris* (Canada, EUA, México, Costa Rica, Panamá América do Norte e Central), havendo as espécies *Amblyomma cajennense*, *Amblyomma aureolatum*, *Rhipicephalus sanguineus* (Argentina, Brasil, e Colômbia na América do Sul).
- A espécie *R. africae*, infecta os carrapatos das espécies *Amblyomma variegatum*, *Amblyomma hebraeum*, *Amblyomma compressum*, *A. lepidum*, *Rhipicephalus annulatus*, *Rhipicephalus evertsi*, *Rhipicephalus decoloratus*, *Rhipicephalus sanguineus*, *Rhipicephalus geigy*, *Hyalomma impeltatum* (África Sub-Saharan), havendo a espécie *Amblyomma variegatum* trazido da África em 1800 aproximadamente (Guadalupe, St. Kitts, Nevis, Dominica, Ilhas Virgens EUA, Montserrat, St. Lucia, Martinica, e Antígua; América do Norte e Central, além da espécie *Amblyomma loculosum* (Ilhas Africanas do Pacífico), *Hyalomma aegyptium* (Ásia) (Turquia) *Hyalomma dromedarii* (Norte da África sub-Saharan Algeria e Egito).

A proteína Tpr ID KJV625361.1 apresentou uma árvore filogenética contendo sequências de proteínas similares e ortólogas para espécies de *Rickettsia* que infectam somente os carrapatos, (Figura 19) (Tabela 39). Para as espécies de *Rickettsia* que compartilharam estas sequências projetamos a possível interação entre as espécies de carrapatos hospedeiros, que poderia ocorrer com a participação direta destas proteínas (Figura 42).

Figura 42 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo motivos repetitivos de tetraco-peptídeos (Tpr) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62536.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura.



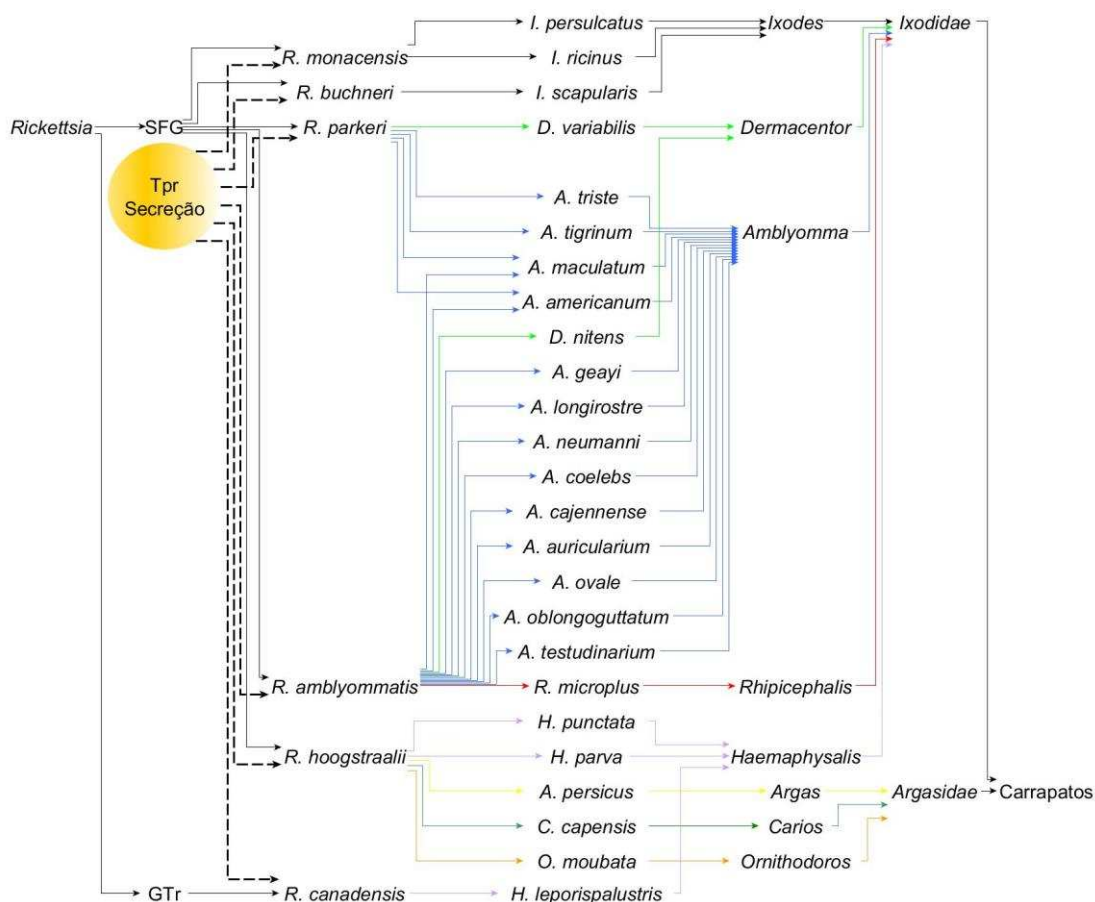
Fonte: Autor deste trabalho.

Na Figura 42, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação de Tpr e das proteínas dos carrapatos, com as quais poderiam interagir e regular a expressão de transcritos, podendo participar na

ativação de proteínas do hospedeiro da família GTPase envolvidas com o transporte intracelular.

A proteína Tpr ID KJV62591.1 apresentou uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 20) e (Tabela 40). Para as espécies de *Rickettsia* que compartilharam estas sequências projetamos a possível interação entre as espécies de carrapatos hospedeiros que poderia ocorrer com a participação direta destas proteínas (Figura 43).

Figura 43 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo motivos repetitivos de tetratricopeptídeos (Tpr) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Graphics. Destaque para a sequência de proteína ID KJV62591.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura.

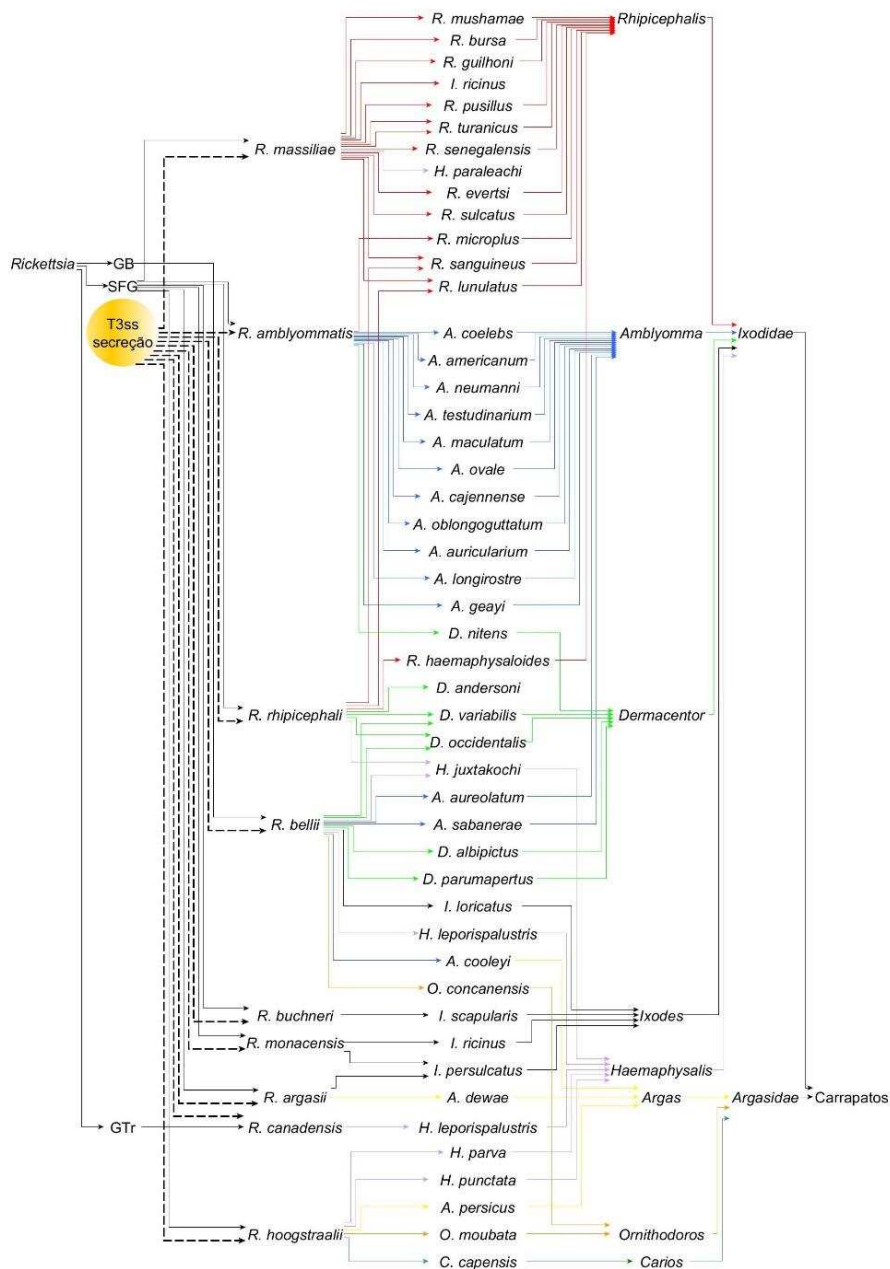


Fonte: Autor deste trabalho.

Na Figura 43, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro da família GTPase envolvidas com o transporte intracelular.

A proteína T3SS KJV62562.1 apresentou uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 21) (Tabela 41). Para as espécies de *Rickettsia* que compartilharam estas sequências projetamos a possível interação entre as espécies de carrapatos hospedeiros que poderia ocorrer com a participação direta destas proteínas (Figura 44).

Figura 44 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas T3SS, podendo conter motivos repetitivos de tetratricopeptídeos (Tpr) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Graphics. Destaque para a sequência de proteína ID KJV62562.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura.



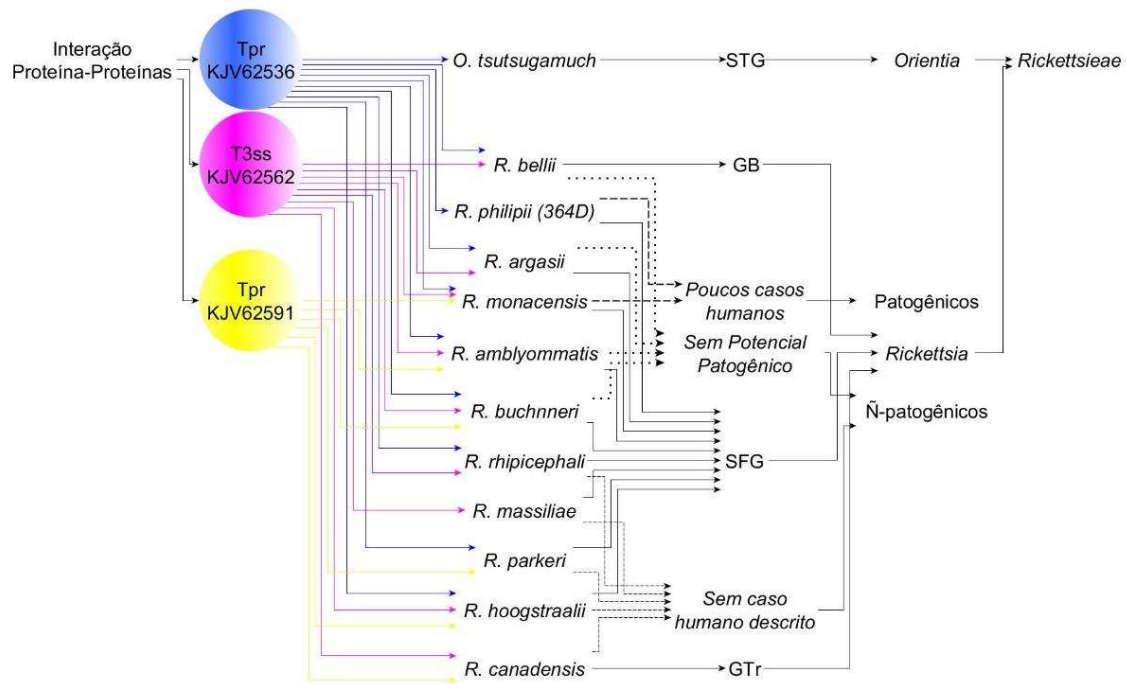
Fonte: Autor deste trabalho.

Na Figura 44, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro da família GTPase envolvidas com o transporte intracelular.

A proteína Tpr interagiram com distintas proteínas relacionadas a GTPase, em especial a Rho, sendo relacionadas com processos de invasão celular e fagocitose, além de outras proteínas GTPs que atuam na funcionalidade do transporte intracelular e que dizem respeito, aos processos de translação e alongamento na tradução de proteínas, transporte de proteínas, regulação, formação e organização de vesículas mediando o transporte e a exocitose de proteínas e outras moléculas; regulação da localização, estabilização e organização de proteínas endomembranas em função de pequenas GTPase, como Rab e Ras, podendo estas sinalizarem a tradução mediando eventos de invaginação com a formação de sinapse entre as vesículas endocitárias, levando a fusão das membranas.

No entanto, as sequências de aminoácidos das proteínas IDs KJV62536.1, KJV62562.1 e KJV62591.1 tiveram suas proteínas similares agrupadas resumidamente com poucas diferenças entre as espécies de *Rickettsia* que compartilharam estas três sequências de proteínas. Para as três houve a predominância de espécies não patogênicas com espécies apresentando baixo nível ou nenhum potencial patogênico e sem caso humano descrito (Figura 45).

Figura 45 Fluxograma representativo para as espécies de *Rickettsia* do grupo ancestral Belli, do grupo Canadense, do grupo das Febres Maculosas transmitidas por carrapatos com pouco potencial patogênico e, ou, espécimes endossimbiontes que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtidos com o uso do aplicativo *Yed Graphics*. Destaque para as proteínas Tpr ID KJV62536 no círculo (azul) e ID KJV62591 no círculo (amarelo) e para a proteína T3SS ID KJV62562 no círculo (rosa) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via de sinalização conservada para transporte e secreção via GTPase.



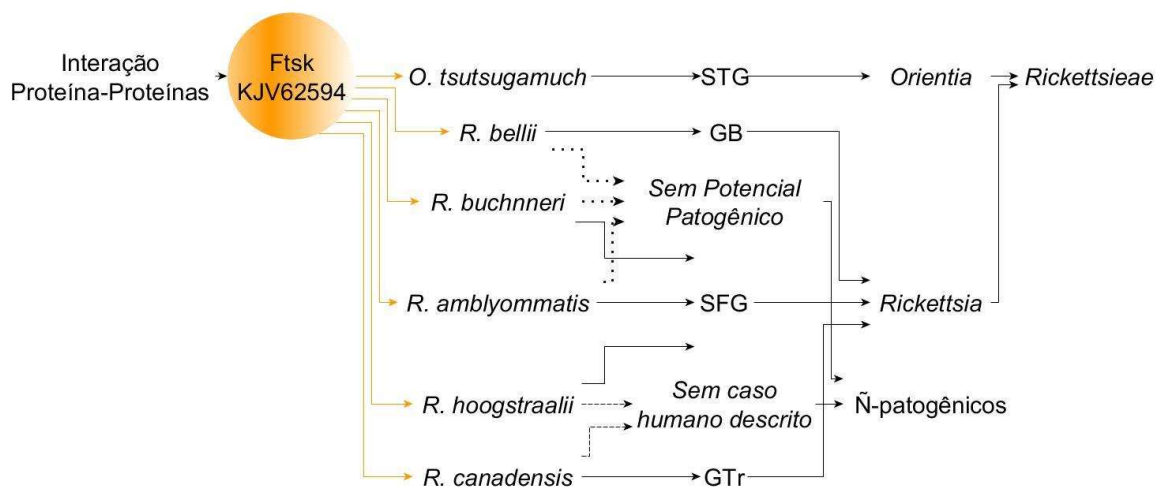
Fonte: Autor deste trabalho.

Na Figura 45, verifica-se que apenas as espécies *R. amblyommatis*, *R. buchneri*, *R. hoogstraalii* e *R. monacensis* apresentaram as duas proteínas Tpr e uma T3SS se destacando como aquelas que não possuem potencial patogênico, havendo entre as demais espécies do Grupo das Febres Maculosas as bactérias mais patogênicas de *Rickettsia* para a árvore filogenética desta proteína.

A proteína Ftsk ID KJV62594.1 apresentou uma árvore filogenética contendo sequências de proteínas similares podendo ser ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 22) (Tabela 42). Para as espécies de *Rickettsia* que compartilharam estas sequências projetamos a possível interação entre as espécies de carrapatos hospedeiros que poderia ocorrer com a participação direta destas proteínas (Figura 46).

Figura 46 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas recombinantes podendo ser responsáveis pelos movimentos de translocação e conjugação do DNA FtskSpoIIIIE, com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62594.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura.

amblyommatis uma espécie não patogênica que possui esta proteína essencial para uma via conservada de proteínas recombinantes conjugadora da família SpoIIIIE.



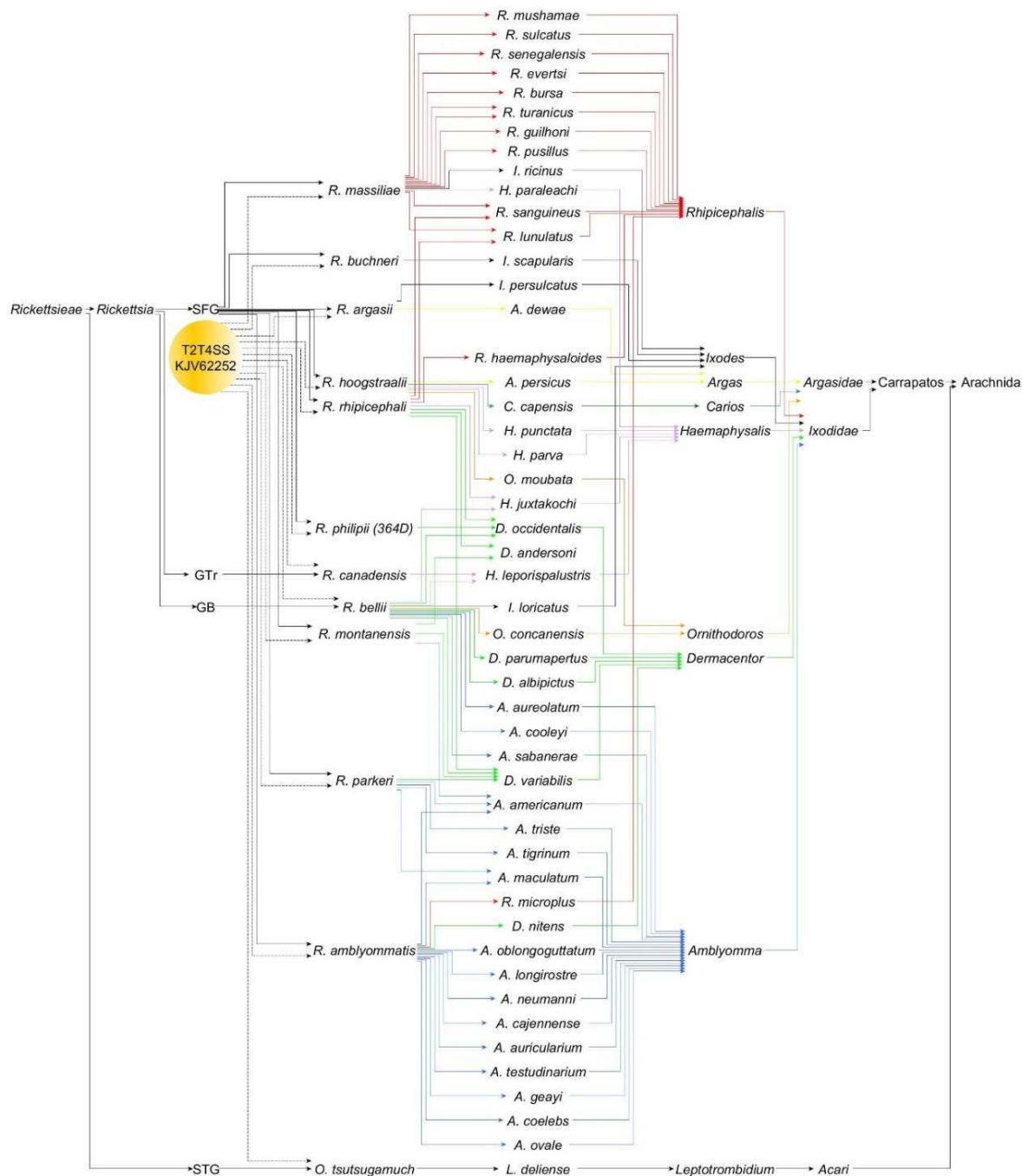
Fonte: Autor deste trabalho.

Na Figura 47, verifica-se que apenas a espécie *O. tsutsugamuch* apresentou a proteína Ftsk se destacando como aquela entre as espécies de *Rickettsia* que possuem potencial patogênico, havendo apenas espécies sem casos humanos descritos ou sem potencial patogênico do Grupo das Febres Maculosas, Grupo de Transição e Grupo Belli.

A proteína T2T4SS ID KJV62252.1 apresentou uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 23) (Tabela 43). Para as espécies de *Rickettsia* que compartilharam estas sequências de proteínas similares, projetamos a interação entre espécies de carrapatos podendo ocorrer a participação direta destas moléculas (Figura 48).

Figura 48 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas do sistema de secreção do tipo II/IV, com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Graphics. Destaque para a sequência de proteína ID KJV62252.1 pertencente a espécie *R.*

amblyommatis Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura.

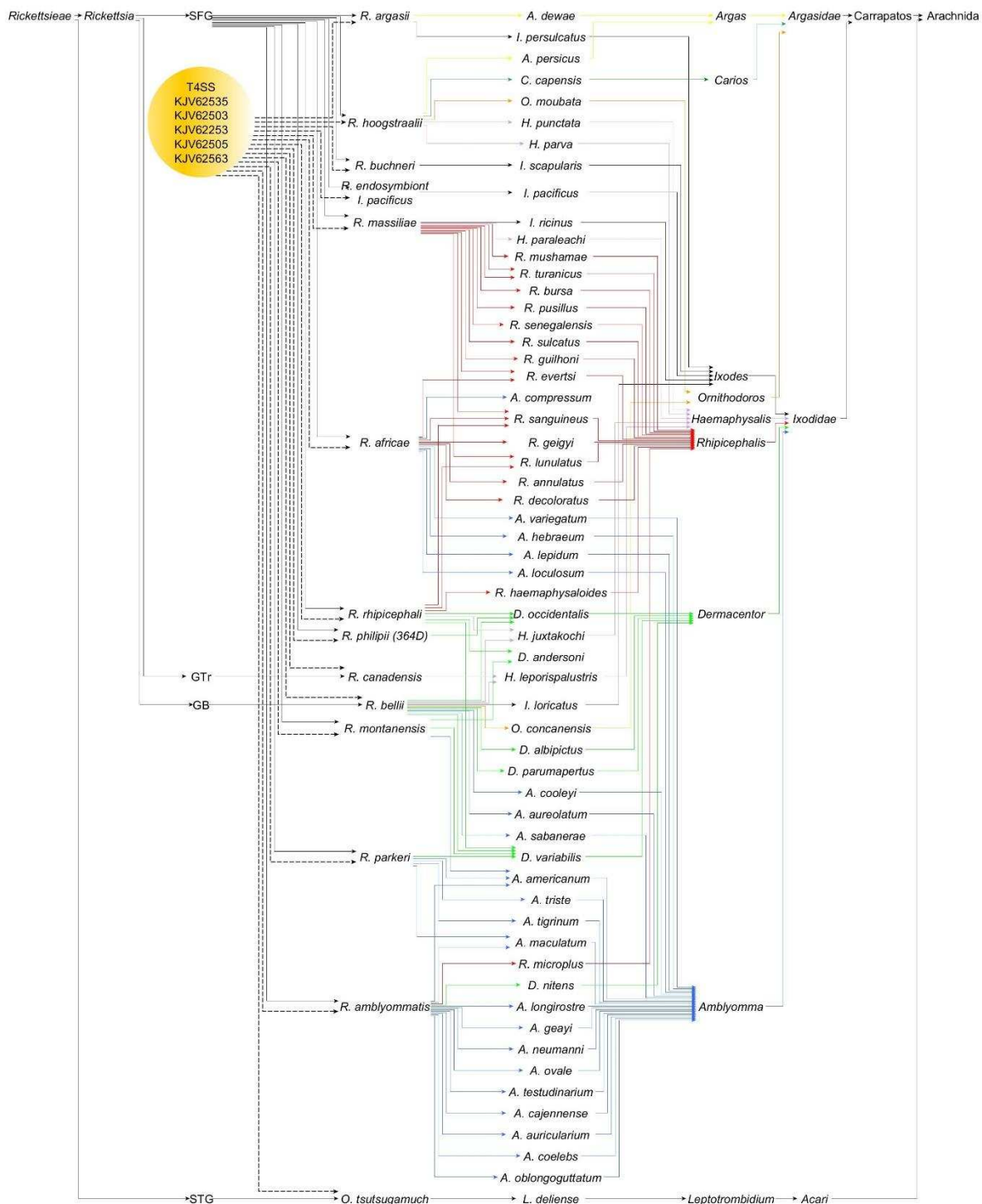


Fonte: Autor deste trabalho.

Na Figura 48, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro da família GTPase envolvidas com o transporte intracelular.

As proteínas T4SS ID KJV62253.1, KJV62505.1, KJV62503.1, KJV62535.1 e KJV62563.1 apresentaram uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 24-28) e (Tabela 44-48). Para as espécies de *Rickettsia* que compartilharam estas sequências de proteínas similares, projetamos a interação entre espécies de carrapatos podendo ocorrer a participação direta destas (Figura 49).

Figura 49 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo proteínas do sistema de secreção do tipo IV (T4SS) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para as sequências de proteínas ID KJV62253.1, KJV62505.1, KJV62503.1, KJV62535.1 e KJV62563.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diversas espécies de carrapatos conforme relatados na literatura.

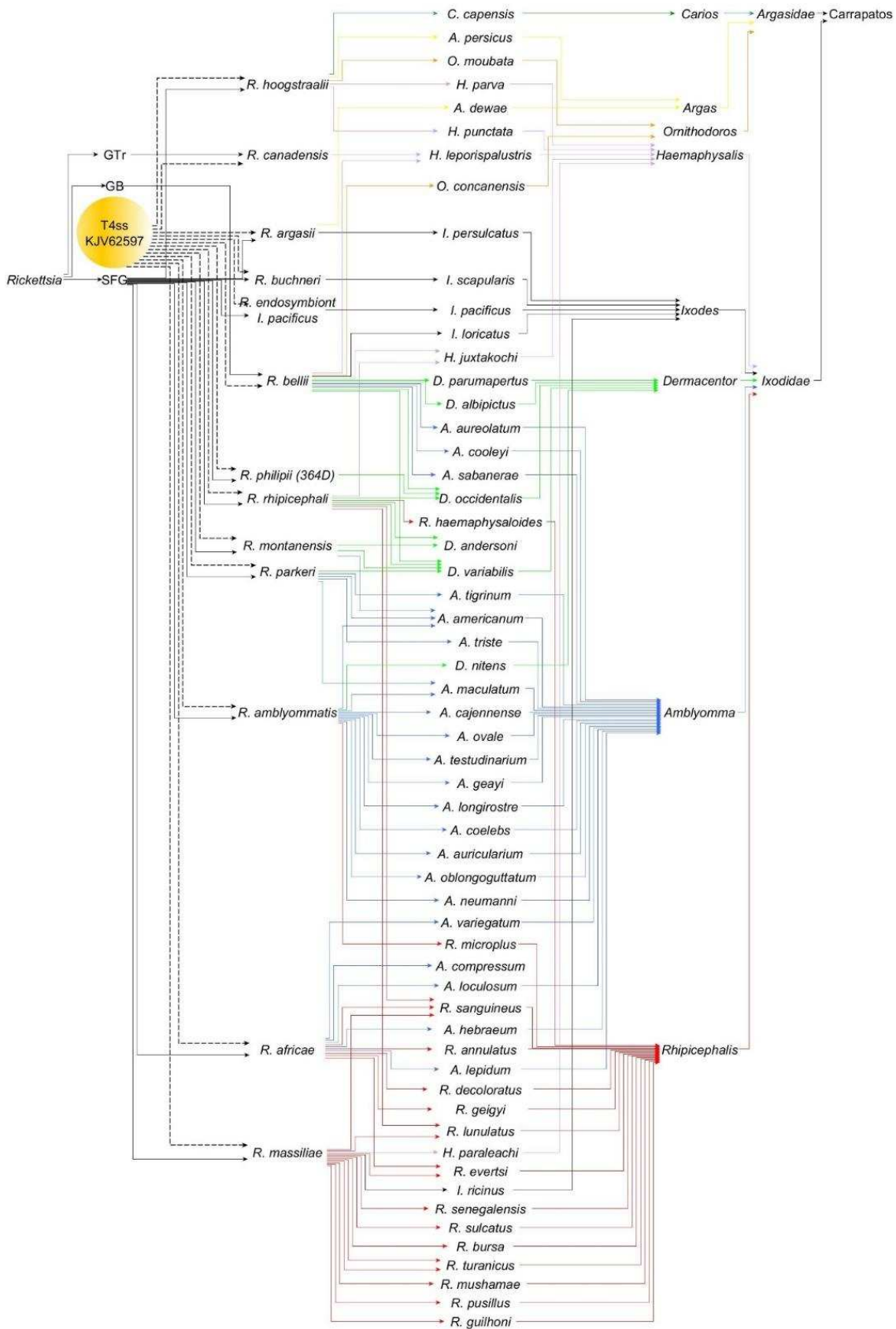


Fonte: Autor deste trabalho.

Na Figura 49, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro da família GTPase envolvidas com o transporte intracelular.

A proteína T4SS ID KJV62597.1 apresentou uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 27) e Tabela (47). Dentre as espécies de *Rickettsia* que compartilharam estas sequências de proteínas similares, projetamos a interação entre espécies de carrapatos podendo ocorrer a participação direta destas (Figura 50).

Figura 50 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas do sistema de secreção do tipo IV, com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV62597.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diferentes espécies de carrapatos, conforme relatados na literatura

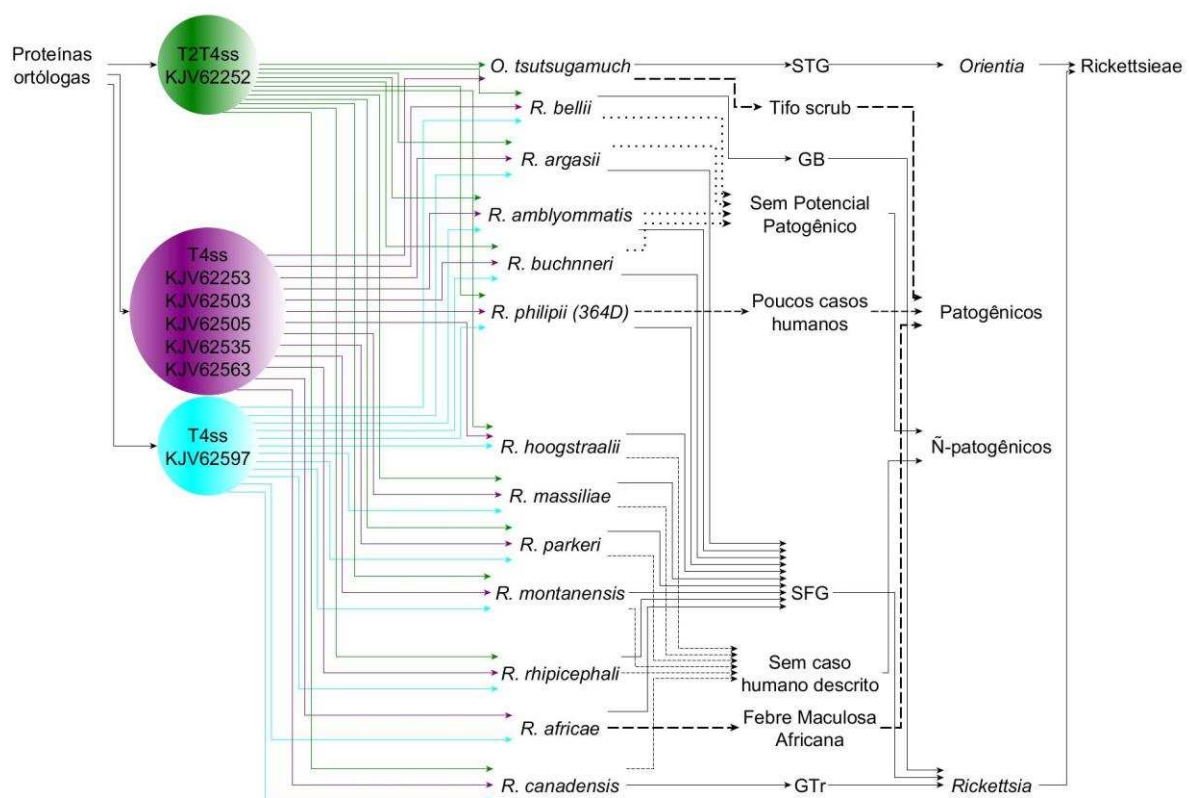


Fonte: Autor deste trabalho.

Na Figura 50, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro da família GTPase envolvidas com o transporte intracelular.

No entanto, a sequências de aminoácidos da proteína dos sistemas de secreção do T2T4SS ID KJV62252.1 e do T4SS ID KJV62253, KJV62505.1, KJV62503.1, KJV62535.1, KJV62563.1 e ID KJV62597.1 e suas proteínas similares foram agrupadas resumidamente entre as espécies de *Rickettsia* com predominância não patogênicas com espécies apresentando baixo nível ou nenhum potencial patogênico e sem caso humano descrito (Figura 51).

Figura 51 Fluxograma representativo para as espécies de *Rickettsia* do grupo ancestral Belli, do grupo Canadense, do grupo das Febres Maculosas transmitidas por carrapatos com pouco potencial patogênico e, ou, espécimes endossimbióticas que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtidos com o uso do aplicativo *Yed Grafics*. Destaque para as proteínas T2T4SS ID KJV62252.1 no círculo (verde) e T4SS ID KJV62253.1, KJV62505.1, KJV62503.1, KJV62535.1, KJV62563.1 no círculo (roxo) e ID KJV62597.1 no círculo (azul) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via conservada de sinalização para transporte e secreção.



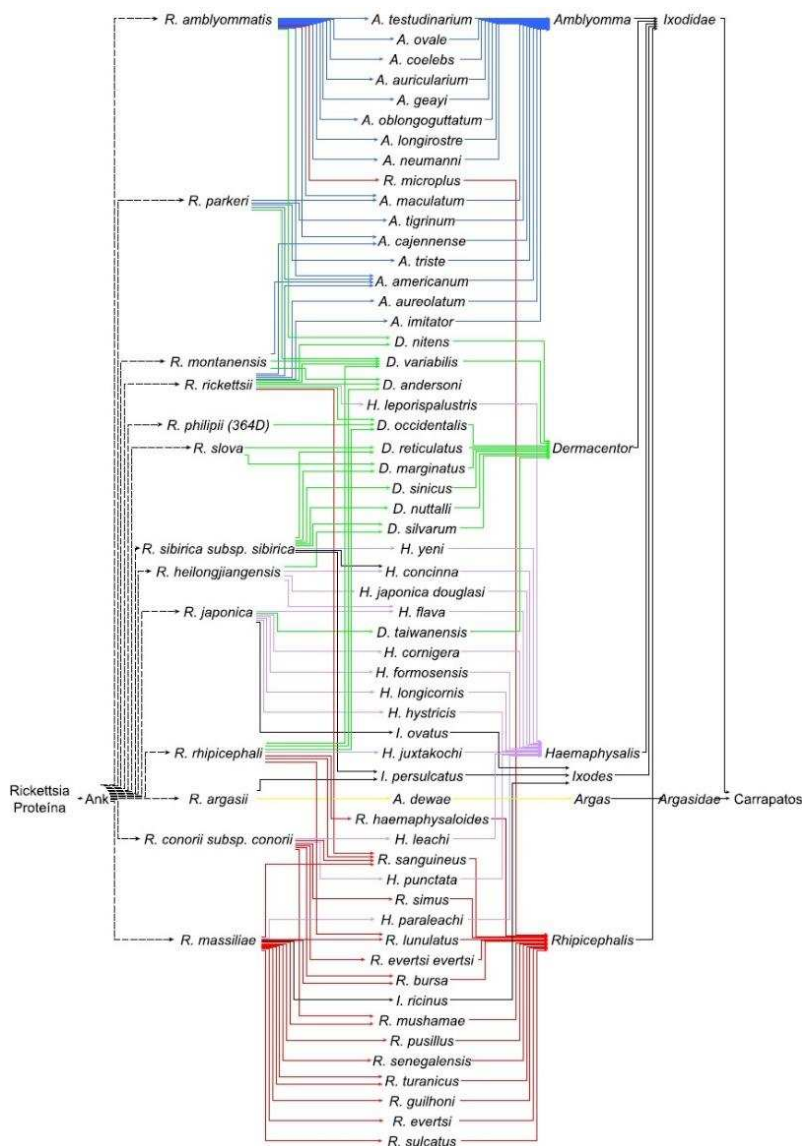
Fonte: Autor deste trabalho.

Na Figura 51, verifica-se que apenas as espécies *O. tsutsugamuch* do Grupo do Tifo Scrub não apresentou ortólogo para a sequência da proteína T4SS ID KJV62597.1 e a espécie *R. africae* do Grupo das Febres Maculosas não apresentou ortólogo para a sequência da proteína

T2T4SS KJV62252.1. Dentre as demais espécies estas se destacaram devido ao seu potencial patogênico, havendo espécies do Grupo das Febres Maculosas não patogênicas e com poucos casos humanos, espécie do Grupo de Transição e do Grupo Beli sem potencial patogênico.

A proteína Ank ID KJV61977.1 apresentou uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 29) (Tabela 49). Para as espécies de *Rickettsia* que compartilharam estas sequências projetamos a possível interação entre as espécies de carrapatos hospedeiros que poderia ocorrer com a participação direta destas proteínas (Figura 52).

Figura 52 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas contendo domínios repetidos de anquirina (Ank) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Grafics. Destaque para a sequência de proteína ID KJV61977.1 pertencente a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diversas espécies de carrapatos conforme relatados na literatura.



Fonte: Autor deste trabalho.

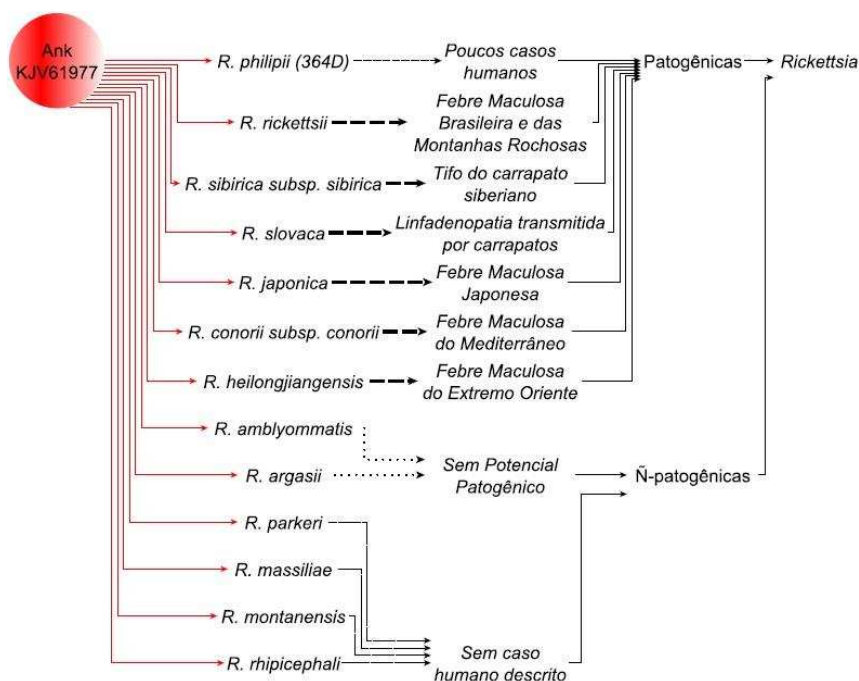
Na Figura 52, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro envolvidas na fosforilação de aminoácidos de proteínas. A sequência da proteína Ank ID KJV61977.1, entre outras proteínas previstas para o módulo de fosforilação foi capaz de interagir com sequências de proteínas relacionadas com o metabolismo de proteínas e aminoácidos por meio da fosforilação.

Nos níveis de expressão e regulação das sequências de transcritos para os carrapatos *Amblyomma* spp. infectados por *Rickettsia* relacionados com as proteínas do *I. scapularis* presentes no interatoma de proteínas de membrana e que interagem com a proteína Ank, verificou-se a sua atuação podendo regular processos de modificações pós-translacionais de proteínas, regulação positiva do nucleoplasma e transporte nuclear, regulação de fatores de transcrição, morfogênese, processos intermediários da organização de citoesqueleto e de morte celular.

No entanto, as sequências de aminoácidos das proteínas Ank ID KJV619771.1 e suas proteínas similares foram agrupadas resumidamente entre as espécies de *Rickettsia* com predominância das espécies mais patogênicas dentro do Grupo das Febres Maculosas, havendo espécies com nenhum potencial patogênico e sem caso humano descrito (Figura 53).

Figura 53 Fluxograma representativo para as espécies de *Rickettsia* do grupo das Febres Maculosas com potencial com potencial patogênico e, ou, espécimes endossimbiontes, transmitidas por carrapatos que não ocasionam riquetsioses e se distribuem por diversas parte do mundo, obtido com o uso do aplicativo *Yed Graphics*. Destaque para as proteínas contendo domínios repetitivos de anquirina (Ank) no círculo (vermelho) pertencentes a *R.*

amblyommatis uma espécie não patogênica que possui estas proteínas essenciais para uma via conservada de sinalização para fosforilação e metabolismo de aminoácidos, apoptose e epigenia.

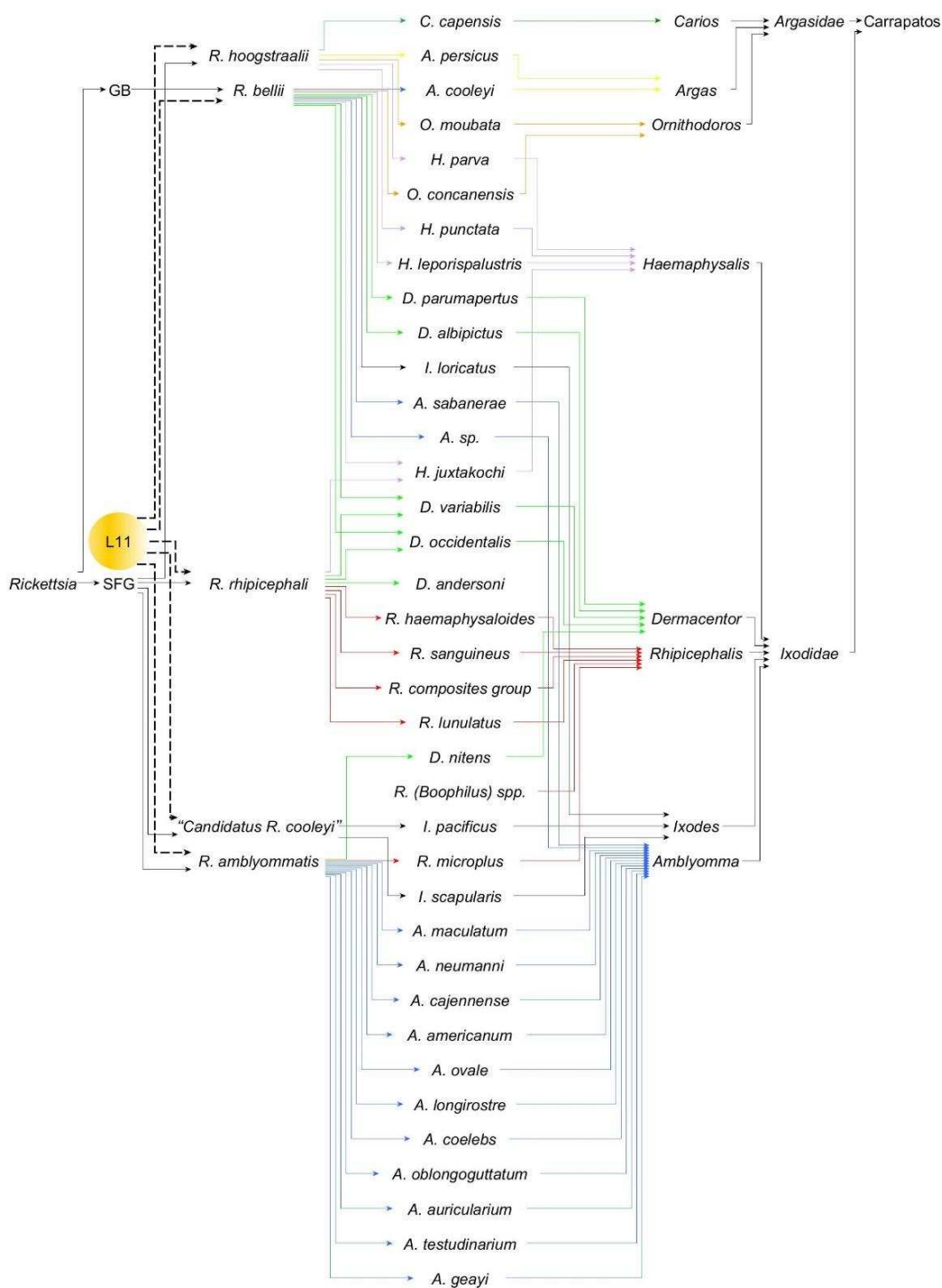


Fonte: Autor deste trabalho.

Na Figura 53, verifica-se que apenas as espécies *R. amblyommatis* e *R. argasii* apresentaram a proteína Ank se destacando como aquelas que não possuem potencial patogênico, havendo entre as demais espécies do Grupo das Febres Maculosas as bactérias mais patogênicas de *Rickettsia* para a árvore filogenética desta proteína.

As proteínas L11 IDs KJV62584.1 e KJV62556.1 apresentaram uma árvore filogenética contendo sequências de proteínas similares ortólogas somente de espécies de *Rickettsia* que infectam os carrapatos (Figura 30) (Tabela 50). Para as espécies de *Rickettsia* que compartilharam estas sequências projetamos a possível interação entre as espécies de carrapatos hospedeiros que poderia ocorrer com a participação direta destas proteínas (Figura 54).

Figura 54 Fluxograma de projeção da interação entre as espécies de *Rickettsia* que possuem as sequências da família de proteínas ribossomais metiltransferase (L11) com seus carrapatos hospedeiros, obtidos com o uso do aplicativo Yed Graphics. Destaque para as sequências de proteínas IDs KJV62584.1 e KJV62556.1 pertencentes a espécie *R. amblyommatis* Ac/Pa e as prováveis espécies de *Rickettsia* com proteínas similares podendo infectar diversas espécies de carrapatos, conforme relatados na literatura.



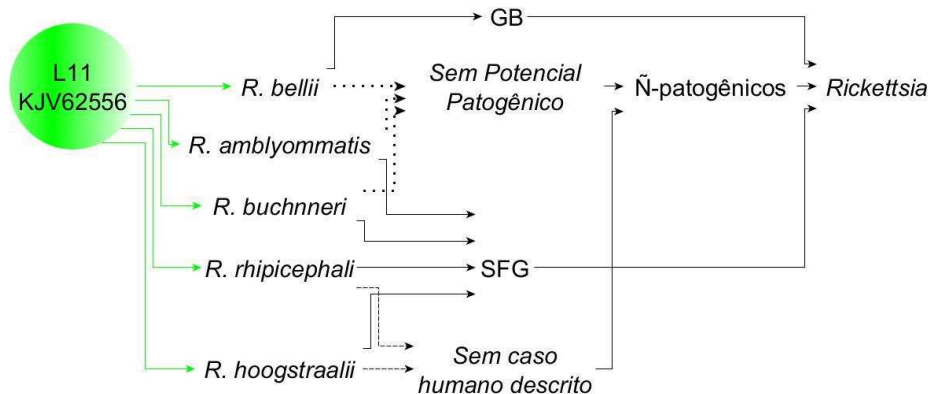
Fonte: Autor deste trabalho.

Na Figura 54, verifica-se o fluxograma de projeção para a interação entre espécies de *Rickettsia* e carrapatos, sendo possível a participação destas proteínas na ativação de proteínas do hospedeiro envolvidas na metilação do DNA e histonas. As sequências de proteínas (L11) ribossomais metiltransferases IDs KJV62584.1 e KJV62556.1 foram as únicas sequências preditas para o modulo de metilação.

Nos níveis de expressão e regulação das sequências de transcritos para espécies de carrapatos *Amblyomma* spp. infectados por *Rickettsia* spp. relacionados com as proteínas do *I. scapularis* presentes no interatoma de proteínas de membrana e, que interagem com as proteínas L11, verificou-se a sua atuação podendo participar de vias moleculares sendo capazes de interagir com as sequências de proteínas do carrapato relacionadas com a regulação de DNA, metilação de histonas H3-K9, alquilação de DNA e aminoácidos, com a formação do pile de conjugação.

No entanto, as sequências de aminoácidos das proteínas L11 ID KJV62584.1 e KJV62556.1 com suas proteínas similares foram agrupadas resumidamente entre as espécies de *Rickettsia* com predominância das espécies não patogênicas dentro do Grupo das Febres Maculosas e do Grupo Bellii, havendo espécies com nenhum potencial patogênico e sem caso humano descrito (Figura 55).

Figura 55 Fluxograma representativo para as espécies de *Rickettsia* do grupo Belli e do grupo das Febres Maculosas sem potencial patogênico e, ou, espécimes endossimbióticas, sem nenhum caso humano descrito, podendo ser transmitidas por carrapatos em diversas partes do mundo, obtidos com o uso do aplicativo *Yed Grafics*. Destaque para as proteínas ribossomais metiltransferase (L11) no círculo (verde) pertencentes a *R. amblyommatis* uma espécie não patogênica que possui estas proteínas essenciais para uma via de sinalização conservada para a metilação do DNA.



Fonte: Autor deste trabalho.

Na Figura 55, verifica-se que apenas a espécie *R. belli* pertencente ao Grupo Belli possui ortólogos para a sequências da proteína L11 entre as espécies do Grupo das Febres Maculosas sem casos humanos e com pouco potencial patogênico dentre as relacionadas deste grupo.

Diferentemente, das árvores filogenéticas para as sequências de proteínas de membrana da espécie *R. amblyommatidis* relacionadas acima e, apresentando sequências similares e ortólogas para espécies de *Rickettsia* que infectam os carrapatos, podendo participar de processos envolvendo vias de secreção celular, transporte intracelular, fosforilação de aminoácidos e metilação, as sequências da proteína NADH-quinona oxidoreductase subunidade B ID KJV61958.1; ATP sintase F1 beta subunit ID KJV61174.1 e citocromo B ID KJV61778.1 (Dados não publicados), apresentaram árvores filogenéticas enraizadas demonstrando parentesco com espécies de bactérias que compõem os gêneros *Anaplasma*, *Ehrlichia* e *Wolbachia*, dentro de *Anasplamaceas* além de *Orientia* e todos os genomas das 26 espécies de *Rickettsia*.

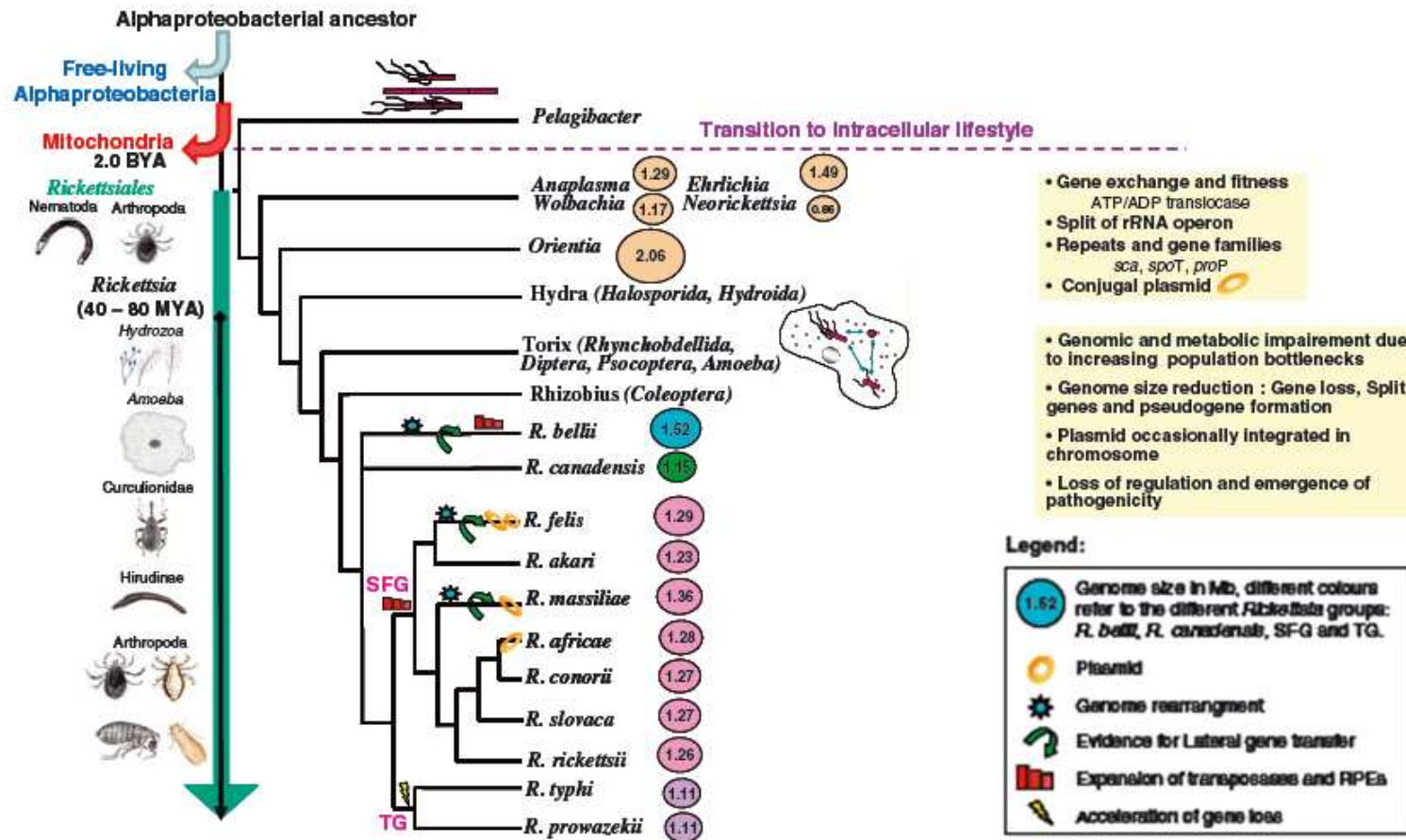
Estas proteínas estão relacionadas com processos de respiração celular, produção de ATP e energia, atuando na cadeia de transportes de elétrons, com oxidação reductiva e fosforilação oxidativa, homeostase com processos metabólicos oxidativos, além de respostas de defesa contra vírus etc. Além dessas proteínas, foram enraizadas as árvores filogenéticas para as sequências de proteínas succinato desidrogenase flavoproteína ID KJV61553.1 e succinato desidrogenase e fumarato reductase ferro sulfúrio ID KJV61447.1 (Dados não publicados), presentes no módulo de proteólise, as quais estão relacionadas com os processos celulares de modificação e catabolismo de proteínas dependente de ubiquitina.

Duplicação gênica, transferência horizontal de genes e rearranjos genômicos na evolução de *Rickettsia*

Diversos estudos sugerem que os carrapatos possuem vias metabólicas específicas que interagem com bactérias endossimbiontes, as quais dependeriam do ambiente da célula do hospedeiro que fornecem os substratos necessários e compostos químicos essenciais para a sua sobrevivência devido a evolução reductiva (ANDERSSON; KURLAND, 1998; DARBY et al., 2007; GOTTLIEB; LALZAR; KLASSON, 2015). Muitas das espécies específicas de genes são de tamanho pequeno e codificam proteínas hipotéticas com funções moleculares desconhecidas e sem similaridade com os genes de outras espécies. No entanto, muito tem sido esclarecido

sobre as taxas e padrões desta perda de genes (DAILY et al., 2015; FRASER et al., 1997; HEPBURN; SIMPSON, 2008) e seus efeitos remontando o passado de *Rickettsia* (MERHEJ; RAOULT, 2011) (Figura 56).

Figura 56 A simbiose de invertebrados e a associação com artrópodos apareceram precocemente em membros de Rickettsiales, como *Anaplasma*, *Ehrlichia*, *Wolbachia*, *NeoRickettsia* e *Orientia*. Destaque para as espécies de *Rickettsia* que se relacionaram com protistas unicelulares, como espécies de *Hydra* surgindo anteriormente do que a espécie *R. bellii*. SFG, Grupo das Febres Maculosas; TG, Grupo do Tifo; PSE, elementos palindrômicos de *Rickettsia*.



Na Figura 56, verifica-se que a transição de *Rickettsia* para um estilo de vida intracelular obrigatório dos organismos apresentados em hospedeiros avaliados tem sido associada à perda de genes, duplicação gênica e diversificação de funções (MERHEJ; RAOULT, 2011). Segundo estes autores, no entanto, plasmídeos conjugais podem ter estado presentes em ancestrais de *Rickettsia* e subsequentemente em outras linhagens com seus os genomas evoluindo à partir de duplicação gênica, transferência horizontal de genes e rearranjos genômicos como a redução do genoma, havendo a presença de elementos conjugativos em alguns desses genomas correlacionado com o aumento no número de transposons, o que ocasiona uma quebra geral na sintenia do genoma de *Rickettsia*.

Atualmente, sabe-se que os sistemas de secreção do tipo IV (T4SS) permitem inserir proteínas efetoras na célula hospedeira e modular as funções de várias vias e proteínas com um objetivo de subverter e explorar as funções moleculares do hospedeiro eucarioto (SAHNI et al., 2013). Segundo estes autores, os T4SS desempenham um papel na transferência de DNA para outras bactérias e para células hospedeiras, podendo adquirir e, ou, inserir o material genético no ambiente extracelular, e na inserção de toxinas, fatores de virulência além de outros efetores no citoplasma do hospedeiro.

Estes motivos estruturais possuem características de proteínas eucarióticas, atuando diretamente na interação proteína-proteína nos sistemas moleculares em que atuam, podendo ser encontradas no genoma de *R. bellii*, que contém genes homólogos de bactérias intracelulares de amebas, como *Legionell pneumophila* e *Protochlamydia amoebophila* (OGATA et al., 2006). A codificação de genes VirB para proteínas que formam o T4SS está associada ao parasitismo intracelular, sendo encontradas em todos os genomas de *Rickettsiales* atuando na conjugação genética entre as bactérias (OGATA, 2005). Estes genes possuem homólogos em *P. amoebophila*, o que pode ter sido adquirido de um protista hospedeiro (LAWLEY et al., 2003; OGATA et al., 2006), ou pode ter sido originado em um clado de bactérias intracelulares parasitas do ancestral eucarioto de Chlamydiales, e posteriormente transferidos entre os demais gêneros (GREUB; RAOULT, 2003; SCHMITZ-ESSER et al., 2004).

A presença de sequências curtas de VirB1, VirB2 e VirB7 do T4SS foram identificadas em 13 genomas de *Rickettsia* do Grupo das Febres Maculosas, sendo o gene VirD4 conservado com a função de ligação ao efector de moléculas do T4SS, enquanto o gene VirB5, que é necessário para a montagem do pilus foi reduzido nos genomas de *Rickettsia* em função do estilo de vida intracelular. No entanto, no genoma de *R. typhi* pertencente ao Grupo do Tifo, o

domínio C-terminal do gene *VirB9b* necessário para interações com as lipoproteínas do gene *VirB7*, é altamente truncado, e não pode auxiliar na montagem do canal do T4SS (GILLESPIE et al., 2009).

A transferência horizontal de genes (HGT) é um mecanismo pelo qual as bactérias adquirem genes de outras bactérias relacionadas, levando à inovação e a evolução de seus genomas, aumentando sua adaptabilidade para se estabelecer em novos nichos (OCHMAN; LAWRENCE; GROISMAN, 2000). O genoma da espécie *R. endosymbiontic I. scapularis* (REIS) transporta mais de 650 transposons, elementos gênicos móveis e amplificados de *Rickettsiales*, oriundos de genomas de outras bactérias (GILLESPIE et al., 2012). Nos genomas das espécies *R. felis*, *R. bellii* e *R. endosymbiontic I. scapularis* foram anotadas diversas sequências de proteínas contendo Ank, Tpr e repetições ricas em leucina LRR, estando presentes também nos genomas dos hospedeiros eucariotos, ou tendo sido adquiridas à partir do genoma ancestral de *Rickettsia*, durante o processo de coevolução (SAHNI et al., 2013).

As proteínas contendo repetições de Ank e Tpr foram implicadas na patogênese de vários patógenos intracelulares obrigatórios (FENN; BLAXTER, 2006; PARK et al., 2004) implicadas em interações patógeno-hospedeiro (CATUREGLI et al., 2000; CAZALET et al., 2004). Segundo estes autores, a proteína Tpr foi implicada na capacidade de *L. pneumophila* estabelecer a infecção ou manipular o seu hospedeiro ativando eventos de tráfego celular. A proteína Ank de *Anaplasma* spp., *Wolbachia* spp. e *Ehrlichia* spp. seria translocada para o citoplasma e para o núcleo da célula hospedeira, podendo interferir por meio da sinalização ao interagir com o citoesqueleto, alterando a transcrição gênica e ligando-se à cromatina do hospedeiro (FENN; BLAXTER, 2006; PARK et al., 2004). A análise global filogenômica entre as proteínas de algumas espécies de *Rickettsia* revelou um alto número de genes que codificam Ank e Tpr, estando presentes em quase 4% dos genomas de *R. bellii* e *R. felis* (OGATA, 2005; OGATA et al., 2006).

Análise do transcriptoma genômico de RNAm isolado da espécie *R. conorii* demonstrou que aproximadamente 15% dos genes, em sua maioria pertencentes ao reparo do DNA e à osmose por vias de estresse, cultivadas em células Vero e a maioria desses genes, é exclusiva do Grupo das Febres Maculosas (RENESTO et al., 2008). O perfil de transcrição global da espécie *R. rickettsii* sugeriu que, exceto por mudanças súbitas e mudanças na temperatura variando entre 37 à 42°C promove o desencadeamento da regulação diferencial de 56 genes (ELLISON et al., 2009). No entanto, a exposição de *R. prowazekii* em condições de temperatura

variando de 37 à 42°C promove o aumento em 23 vezes do número de genes expressos, dos quais 57% são caracterizados como genes induzidos por choque térmico (AUDIA; PATTON; WINKLER, 2008).

As espécies de *Rickettsia* codificam poucos reguladores transcricionais quando comparados com micróbios de vida livre, sugerindo que os genes adquiridos durante os estágios iniciais da evolução deste gênero são responsáveis pelas funções que facilitam a invasão e adaptação a diferentes hospedeiros (FUXELIUS et al., 2008). Estes autores sugerem que ao encontrar um ambiente hospedeiro adequado, a maioria dos reguladores de resposta transcricional não essenciais foram deliberadamente perdidos, sendo a aquisição de genes e a perda de genes o principal mecanismo de interação de adaptação entre bactérias e seus hospedeiros, tanto no estilo de vida patogênico, quanto endossimbiótico de *Rickettsiales* e outras bactérias.

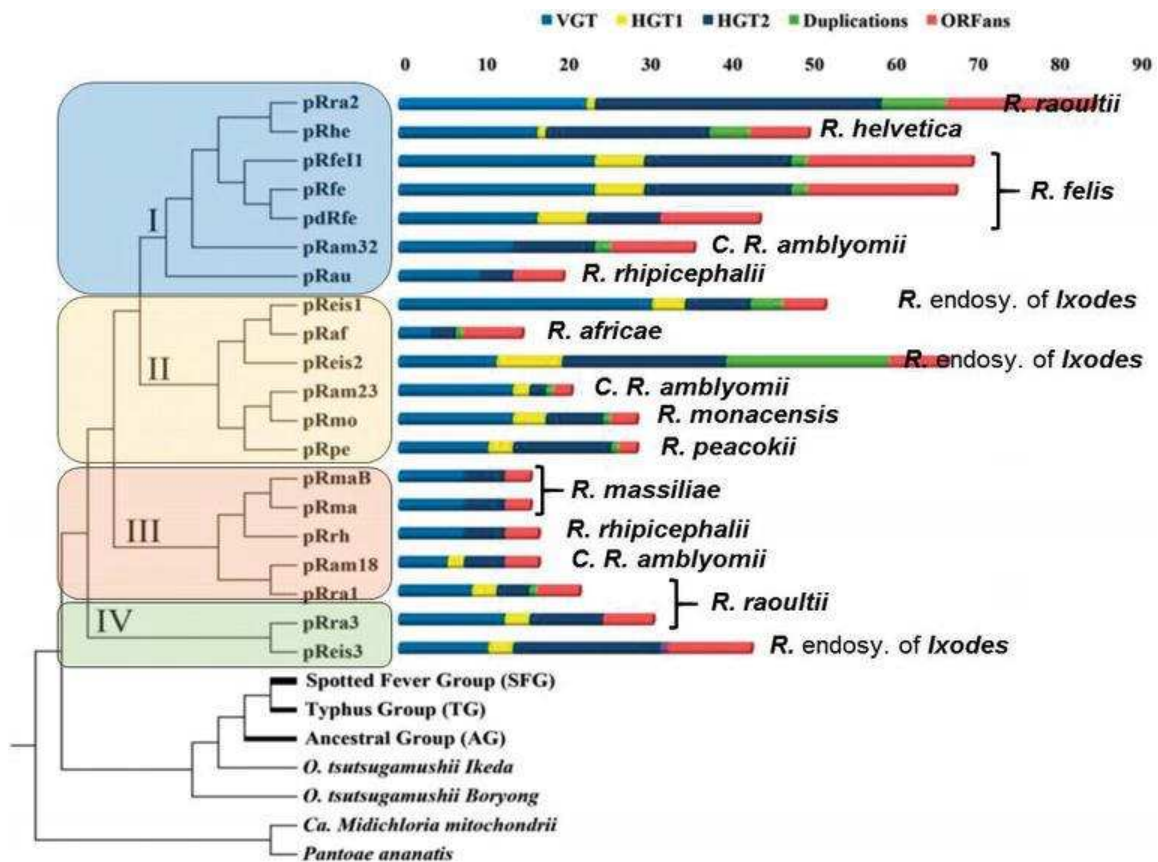
Os plasmídeos abrangem regiões genéticas muito grandes, 100 kilobases (kb), incluindo vários conjuntos de genes e sua integração frequente próximo do loci de RNAt sugere que muitos deles foram introduzidos em genomas bacterianos através de eventos de transferência mediados por fagos (FOURNIER et al., 2009). Segundo estes autores, na patogenicidade, os plasmídeos são chamados de “ilhas patogênicas” e em endossimbiontes “ilhas simbióticas” e o dogma que os plasmídeos não estão presentes em *Rickettsiaceae* foi refutado, com a eletroforese em gel de campo pulsado e análises de *Southern blot* de DNAs em diferentes espécies, o que sugere que podem estar disseminados no gênero.

A existência de plasmídeos, apesar da pressão exercida pela evolução do genoma redutivo, sugere um papel importante na biologia de *Rickettsia*, havendo entre os plasmídeos das 26 espécies atualmente reconhecidas deste gênero, a verificação que 11 espécies continham um a quatro plasmídeo com um tamanho variando de 12 a 83 kb, possuindo entre 15 a 85 genes (FOURNIER et al., 2009). Foi verificado entre eles que o pRICO, último ancestral comum dos atuais plasmídeos no gênero *Rickettsia*, era herdado verticalmente principalmente dos cromossomos de *Rickettsia/Orientia*, divergindo verticalmente em um ou vários plasmídeos nas espécies que compõem estes gêneros de bactérias (EL KARKOURI et al., 2016). Segundo estes autores, dos 747 genes codificadores de proteínas, 65% são de genes inteiros, com os 35% restantes parcialmente degradados, com os níveis de degradação variando entre os plasmídeos, relatando entre 16 a 40% em plasmídeos maiores (tamanho >47 kpb) e entre 44 a 59% em plasmídeos menores.

Foi observado que os plasmídeos são perdidos durante a passagem em série a longo prazo em células cultivadas, o que dificultaria a realização de estudos de ancestralidade para elucidar um único ou múltiplos ancestrais (EL KARKOURI et al., 2016). Segundo estes autores, os plasmídeos agruparam-se em quatro grupos putativos (I–IV), como a seguir: Grupo I Incluiu quatro grandes e três pequenos plasmídeos de cinco espécies: pRra2 em *R. raoultii*, pRhe em *R. helvetica*, pRfe, pRfeI1 e pDRfe em *R. felis*, pRam32 em Candidatus *R. amblyomii* e pRau em *R. rhipicephalii*; Grupo II Agrupou dois grandes e quatro pequenos plasmídeos pertencentes a cinco espécies: pReis1 e pReis2 em *R. endosimbionte I. scapularis*, pRaf em *R. africae*, pRam23 em Candidatus *R. amblyomii*, pRmo em *R. monacensis* e pRpe em *R. peacockii*; Grupo III Contém cinco pequenos plasmídeos de quatro espécies: pRam18 em Candidatus *R. amblyomii*, pRrh em *R. rhipicephalii*, pRra1 em *R. raoultii* e pRma e pRmaB em *R. massiliae*; e Grupo IV Reuniu um grande e um pequeno plasmídeo de duas espécies: pReis3 em *R. endosimbionte de I. scapularis* e pRra3 em *R. raoultii*.

No entanto, no nível inter-espécies, os plasmídeos pertencentes do mesmo grupo de *Rickettsia* mostraram conservações de sequência variável (GEORGIADES et al., 2011) (Figura 57).

Figura 57 Eventos evolutivos que moldaram os plasmídeos de *Rickettsia* e *Orientia*. Os genes utilizados foram hsp, dnaA, sca12, transposase 1 e 2, fosfolipase, traD, repetição de leucina, helix-turn-helix e timidilato quinase.



Fonte: Georgiades et al. (2011).

Na Figura 57, verifica-se que os plasmídeos de *Rickettsia* são um espelho da história evolutiva desse grupo bacteriano que envolve processos redutivos, eventos de duplicação e aquisição horizontal de genes necessários para se adaptar a um estilo de vida intracelular em eucariotos, sendo necessário determinar sua distribuição, evolução e o seu papel na adaptação e virulência no hospedeiro (GEORGIADES et al., 2011). O mecanismo de perda de genes tem sido uma estratégia bastante difundida na evolução dos genomas de *Rickettsiales*, e foi descoberto na espécie *R. endosimbionte* de *I. scapularis*, sendo a proliferação de elementos genéticos móveis, em particular, um elemento integrador RAGE (*Rickettsiales Amplified Genetic Element*) presente no cromossomo e nos plasmídeos (BALDRIDGE et al., 2008; GEORGIADES et al., 2011).

A espécie *R. endosimbionte* de *I. scapularis* codifica nove RAGEs conservados que incluem o T4SS semelhantes aos encontrados em outras espécies de *Rickettsia* compreendendo 35% do genoma total, tornando este microrganismo um dos elementos móveis bacterianos mais

plásticos e repetitivos (GEORGIADES et al., 2011). A presença de RAGE em *R. endosimbionte* de *I. scapularis* fornece a evidência mais robusta de que genes de *Rickettsia* conservados e associados ao estilo de vida intracelular foram adquiridos por via de mobilidade de elementos gênicos, podendo, através de contínuas invasões genômicas, fornecerem *insights* sobre a origem dos mecanismos de patogenicidade de *Rickettsia* (GILLESPIE et al., 2012). Segundo estes autores, os RAGEs são a fusão de genes da família Tra que codificam a proteína de transferência conjugal, podendo ser encontrados entre os genes traA e traD.

As espécies de *Rickettsia* compartilham 1,027 genes que provavelmente foram transferidos verticalmente de "proto-*Rickettsia*", onde a espécie *R. bellii* manteve todos estes, enquanto outras espécies, como *R. prowazekii* e *R. typhi* (128 genes perdidos), perderam grande parte destes (GEORGIADES et al., 2011). Segundo estes autores, a perda diferencial de genes contribuiu para a criação de novas espécies de *Rickettsia* podendo a perda de genes reguladores terem promovido o aumento de virulência em algumas espécies que infectam os carrapatos e vertebrados mamíferos. A evolução redutiva e a perda de genes envolvendo presumivelmente o *operon* Tra (FOURNIER et al., 2009), moldaram o conteúdo de alguns genomas de *Rickettsiales*, e a transferência horizontal de genes desempenharam um papel importante na evolução do genoma destas bactérias baseado na perda de genes e eventos de transferência horizontal de genes em *Rickettsia* spp., *Anaplasma* spp. e *Orientia* spp. mostrou que ocorreram vários eventos (GEORGIADES et al., 2011).

Segundo estes autores, o ganho genético é um evento conhecido que ocorreu durante a evolução das espécies de *Rickettsia*, havendo o fato de que muitos, dos genes presentes nos genomas de *Orientia*, vieram de transferência horizontal de genes ancestralmente e incluem transposases e proteínas contendo repetição de anquirina Ank que parecem ter sido adquiridas de vírus e protistas. No entanto, os genes doados pela *Orientia* para *Rickettsia* foram obtidos por *Firmicutes* spp., *Bacteroidetes* spp. e *Gamma proteobacteria* spp. com a análise do genoma completo de *O. tsutsugamushi* revelando a presença de histidina quinase Hk do T4SS, SpoT, Tra e Tpr (OGAWA et al., 2009). Segundo estes autores, as proteínas Hk atuam como transdutores de sinal em resposta a mudanças no ambiente, as proteínas da família SpoT atuam na resposta à falta de energia, enquanto as proteínas da família Tra participam da transferência de genes entre *Rickettsia* e outras bactérias.

Os resultados deste trabalho evidenciaram a presença em parte estas proteínas possuindo sequências ortólogas para *Rickettsia* que infectam carrapatos sendo as que mais divergiram em

relação as espécies que infectam insetos. A maioria dos ortólogos desempenham funções vitais na biologia ou no desenvolvimento celular geral sendo fixados em populações ou espécies por seleção negativa e mesmo assim, ocorrem perdas genéticas, ortólogos adquirem funções novas ou adicionais e domínios são excedidos para uma nova função (GABALDÓN; KOONIN, 2013; KANNAN; ROGOZIN; KOONIN, 2014). Segundo estes autores, os ortólogos geralmente possuem estrutura molecular, função, mecanismo de ação semelhante, resíduos conservados envolvidos na interação molecular e arquitetura de domínio entre espécies ou linhagens e podem ser rastreados até a última linhagem ancestral comum onde essa função se originou, sendo vistos como a base da universalidade dos processos biológicos celulares gerais, como transcrição, tradução, localização celular, secreção, transporte, metabolismo e a capacidade de anotar genes por homologia, conforme encapsulado na conjectura ortológica.

Neste sentido, havendo as associações de carrapato-patógeno que evoluíram para formar “*relações epigenéticas íntimas*” em que o patógeno induz reprogramação transcricional em carrapatos infectados, favorecendo a propagação de patógenos, mas também selecionando as adaptações ecológicas mais adequadas neste ácaro, as alterações fenotípicas e genéticas poderiam ser transmitidas para a próxima geração de carrapatos que evoluiriam para maximizar a circulação de patógenos na comunidade de hospedeiros. A ocorrência de evidências que os patógenos transmitidos por carrapatos induzem reprogramação transcricional em carrapatos infectados e células de vertebrados por meio de proteínas efetoras, nucleomodulinas secretadas para controlar a transcrição do hospedeiro e outros processos nucleares interagindo com a maquinaria epigenética do hospedeiro para induzir transcrições de reprogramação além de proporcionar o aumento da aptidão do carrapato.

6 CONCLUSÃO

Este trabalho constitui a primeira tentativa de explorar redes de interação de proteínas entre *Rickettsia* spp. e carrapatos utilizando métodos *in silico* para a construção do interatoma putativo entre as espécies estudadas. Obtivemos uma lista de proteínas putativas essenciais, havendo evidências experimentais e correlatos na literatura. Dentre as proteínas que mais divergiram em relação as espécies de *Rickettsia* que infectam carrapatos em relação as que infectam insetos, observamos as sequências das proteínas KJV62556.1 e KJV62584.1 (Ribossomal L11 *methyltransferase family protein*), KJV61977.1 (*Ankyrin repeat family protein*), KJV62536,1 e KJV62591,1 (*Tetratricopeptide repeat protein*), KJV62562,1 (*Type III secretion low calcium response chaperone LcrH/SycD*) e KJV62253.1 (*Type IV secretory system Conjugative DNA transfer family protein*). Estas proteínas constituem um importante grupo de proteínas de *Rickettsia* que infectam carrapatos Ixodídeos e Argasídeos a serem exploradas como fontes de novos alvos e vias, uma vez que são importantes para a sobrevivência do parasita, enquanto não possuem similaridade com outras sequências de proteínas de espécies de *Rickettsia* que infectam outros hospedeiros. Mais estudos experimentais são necessários para identificarmos inibidores específicos. Esses resultados ajudarão nos futuros esforços de descoberta de vias para o controle de *Rickettsia*. De importância crítica é necessário identificarmos e compreendermos quais fatores que distinguem as espécies patogênicas de não patogênicas, considerando os seus hospedeiros naturais e aqueles que por aspectos ecológicos e moleculares lhes permitem invadir e colonizar hospedeiros invertebrados e vertebrados, e quanto isto envolve a manipulação diretamente das vias celulares do carrapato.

REFERÊNCIAS BIBLIOGRÁFICA

ADAMSON, S. W. et al. Knockdown of selenocysteine-specific elongation factor in *Amblyomma maculatum* alters the pathogen burden of *Rickettsia parkeri* with epigenetic control by the Sin3 histone deacetylase corepressor complex. **PLoS ONE**, v. 8, n. 11, p. e82012, 25 nov. 2013.

AKOPIAN, D. et al. SecYEG activates GTPases to drive the completion of cotranslational protein targeting. **The Journal of Cell Biology**, v. 200, n. 4, p. 397–405, 18 fev. 2013.

ALIYARI, R.; DING, S.-W. RNA-based viral immunity initiated by the Dicer family of host immune receptors. **Immunological Reviews**, v. 227, n. 1, p. 176–188, jan. 2009.

ALTENHOFF, A. M. et al. Resolving the ortholog conjecture: Orthologs tend to be weakly, but significantly, more similar in function than paralogs. **PLoS Computational Biology**, v. 8, n. 5, p. e1002514, 17 maio 2012.

ALTENHOFF, A. M.; GLOVER, N. M.; DESSIMOZ, C. Inferring orthology and paralogy. In: **Evolutionary genomics**. [s.l.] Springer, 2019. p. 149–175.

ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–3402, 1997.

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**, v. 9, n. 1, p. 552, 2008.

ANDERSSON, S. G. E. et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. **Nature**, v. 396, n. 6707, p. 133–140, nov. 1998.

ANDERSSON, S. G. E.; KURLAND, C. G. Reductive evolution of resident genomes. **Trends in Microbiology**, v. 6, n. 7, p. 263–268, jul. 1998.

ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data** Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, , 2010.

ARAGÃO, H.; FONSECA, F. DA. Notas de ixodologia: VIII. Lista e chave para os

representantes da fauna ixodológica brasileira: notas de ixolodologia. **Memórias do Instituto Oswaldo Cruz**, v. 59, n. 2, p. 115–129, 1961.

ARIUMI, Y. et al. DDX3 DEAD-Box RNA helicase is required for hepatitis c virus RNA replication. **Journal of Virology**, v. 81, n. 24, p. 13922–13926, 15 dez. 2007.

ARNOLD, K. et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. **Bioinformatics**, v. 22, n. 2, p. 195–201, 15 jan. 2006.

ASHER, G. et al. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. **Genes & development**, v. 19, n. 3, p. 316–321, 2005.

AUDIA, J. P.; PATTON, M. C.; WINKLER, H. H. DNA Microarray Analysis of the Heat Shock Transcriptome of the Obligate Intracytoplasmic Pathogen *Rickettsia prowazekii*. **Applied and Environmental Microbiology**, v. 74, n. 24, p. 7809–7812, 15 dez. 2008.

AYLLÓN, N. et al. Systems biology of tissue-specific response to *Anaplasma phagocytophilum* reveals differentiated apoptosis in the tick vector *Ixodes scapularis*. **PLOS Genetics**, v. 11, n. 3, p. e1005120, 27 mar. 2015.

BADER, G. D. BIND: the Biomolecular Interaction Network Database. **Nucleic Acids Research**, v. 31, n. 1, p. 248–250, 1 jan. 2003.

BAE, Y. S.; CHOI, M. K.; LEE, W.-J. Dual oxidase in mucosal immunity and host-microbe homeostasis. **Trends in Immunology**, v. 31, n. 7, p. 278–287, jul. 2010.

BAIROCH, A. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Research**, v. 28, n. 1, p. 45–48, 1 jan. 2000.

BALDRIDGE, G. D. et al. Plasmids of the pRM/pRF Family Occur in Diverse *Rickettsia* Species. **Applied and Environmental Microbiology**, v. 74, n. 3, p. 645–652, 1 fev. 2008.

BALRAJ, P. et al. RickA expression is not sufficient to promote actin-based motility of *Rickettsia raoultii*. **PLoS ONE**, v. 3, n. 7, p. e2582, 9 jul. 2008.

BARISCH, C.; SOLDATI, T. Breaking fat! How mycobacteria and other intracellular pathogens manipulate host lipid droplets. **Biochimie**, v. 141, p. 54–61, out. 2017.

BARLOWE, C. COPII: A membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. **Cell**, v. 77, n. 6, p. 895–907, 17 jun. 1994.

BARLOWE, C.; SCHEKMAN, R. SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER. **Nature**, v. 365, n. 6444, p. 347–349, 1993.

BARROS-BATTESTI, D. M.; ARZUA, M.; BECHARA, G. H. **Carrapatos de importância médico-veterinária da região neotropical: um guia ilustrado para identificação de espécies**. [s.l.] ICTTD-3/Instituto Butantan, 2006.

BARTEL, D. P. MicroRNAs. **Cell**, v. 116, n. 2, p. 281–297, jan. 2004.

BASTOS, R. G. et al. Silencing of a putative immunophilin gene in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* increases the infection rate of *Babesia bovis* in larval progeny. **Parasites and Vectors**, v. 2, n. 1, p. 1–11, 2009.

BAUGH, J. M.; VIKTOROVA, E. G.; PILIPENKO, E. V. Proteasomes can degrade a significant proportion of cellular proteins independent of ubiquitination. **Journal of Molecular Biology**, v. 386, n. 3, p. 814–827, 2009.

BEATI, L. et al. *Rickettsia aeschlimannii* sp. nov., a new spotted fever group *Rickettsia* associated with *Hyalomma marginatum* ticks. **International Journal of Systematic Bacteriology**, v. 47, n. 2, p. 548–554, 1 abr. 1997.

BEATI, L. et al. Phylogeography and demographic history of *Amblyomma variegatum* (Fabricius) (Acari: Ixodidae), the tropical bont tick. **Vector-Borne and Zoonotic Diseases**, v. 12, n. 6, p. 514–525, jun. 2012.

BECHAH, Y. et al. Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. **Genome Research**, v. 20, n. 5, p. 655–663, 1 maio 2010.

BEDFORD, G. A. H. *Nuttalliella namaqua*, a new genus and species of Tick. **Parasitology**, v. 23, n. 2, p. 230–232, 6 abr. 1931.

BELL, E. J.; PICKENS, E. G. A toxic substance associated with the *Rickettsia* spp. of the spotted fever group. **The Journal of Immunology**, v. 70, n. 5, p. 461–472, 1953.

BELL, J. K. et al. Leucine-rich repeats and pathogen recognition in Toll-like receptors. **Trends in Immunology**, v. 24, n. 10, p. 528–533, out. 2003.

BERNSTEIN, F. C. et al. The protein data bank: A computer-based archival file for macromolecular structures. **Archives of Biochemistry and Biophysics**, v. 185, n. 2, p. 584–591, jan. 1978.

BERRADA, Z. L. et al. *Rickettsia rickettsii* (Rickettsiales: Rickettsiaceae) in *Amblyomma americanum* (Acari: Ixodidae) From Kansas. **Journal of Medical Entomology**, v. 48, n. 2, p. 461–467, 1 mar. 2011.

BHAVSAR, A. P.; GUTTMAN, J. A.; FINLAY, B. B. Manipulation of host-cell pathways by bacterial pathogens. **Nature**, v. 449, n. 7164, p. 827–834, 17 out. 2007.

BISSINGER, B. W. et al. Synganglion transcriptome and developmental global gene expression in adult females of the American dog tick, *Dermacentor variabilis* (Acari: Ixodidae). **Insect Molecular Biology**, v. 20, n. 4, p. 465–491, ago. 2011.

BLANC, G. et al. Reductive genome evolution from the mother of *Rickettsia*. **PLoS Genetics**, v. 3, n. 1, p. e14, 2007.

BLATCH, G. L.; LÄSSLE, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. **BioEssays**, v. 21, n. 11, p. 932–939, 11 out. 1999.

BOKOCH, G. M. Regulation of innate immunity by Rho GTPases. **Trends in Cell Biology**, v. 15, n. 3, p. 163–171, mar. 2005.

BOLDBAATAR, D. et al. Identification of two forms of cyclophilin from the hard tick *Haemaphysalis longicornis*. **Process Biochemistry**, v. 43, n. 6, p. 615–625, jun. 2008.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BONIFACINO, J. S.; GLICK, B. S. The mechanisms of vesicle budding and fusion. **Cell**, v. 116, n. 2, p. 153–166, jan. 2004.

BRÄMS, J. E. et al. Tetratricopeptide repeats are essential for PcrH chaperone function in *Pseudomonas aeruginosa* type III secretion. **FEMS Microbiology Letters**, v. 256, n. 1, p.

57–66, mar. 2006.

BRAUN, V.; HANTKE, K. Recent insights into iron import by bacteria. **Current Opinion in Chemical Biology**, v. 15, n. 2, p. 328–334, abr. 2011.

BRAY, N. L. et al. Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology**, v. 34, n. 5, p. 525–527, 4 maio 2016.

BRAZ, G. R. C. et al. A missing metabolic pathway in the cattle tick *Boophilus microplus*. **Current Biology**, v. 9, n. 13, p. 703–706, jul. 1999.

BROOKS, A. J. et al. The interdomain region of dengue NS5 protein that binds to the viral helicase NS3 contains independently functional importin β 1 and importin α/β -recognized nuclear localization signals. **Journal of Biological Chemistry**, v. 277, n. 39, p. 36399–36407, 27 set. 2002.

BRUMELL, J. H.; SCIDMORE, M. A. Manipulation of Rab GTPase function by intracellular bacterial pathogens. **Microbiology and Molecular Biology Reviews**, v. 71, n. 4, p. 636–652, 1 dez. 2007.

BUDACHETRI, K.; KARIM, S. An insight into the functional role of thioredoxin reductase, a selenoprotein, in maintaining normal native microbiota in the Gulf Coast tick (*Amblyomma maculatum*). **Insect Molecular Biology**, v. 24, n. 5, p. 570–581, out. 2015.

BUDNIK, A.; STEPHENS, D. J. ER exit sites - Localization and control of COPII vesicle formation. **FEBS Letters**, v. 583, n. 23, p. 3796–3803, 3 dez. 2009.

BURGDORFER, W.; BRINTON, L. P. Mechanisms of transovarial infection of spotted fever Rickettsiae in ticks. **Annals of the New York Academy of Sciences**, v. 266, n. 1 Pathobiology, p. 61–72, nov. 1975.

BURGDORFER, W.; VARMA, M. G. R. Trans-stadial and transovarial development of disease agents in arthropods. **Annual review of entomology**, v. 12, n. 1, p. 347–376, 1967a.

BURGDORFER, W.; VARMA, M. G. R. Trans-stadial and transovarial development of disease agents in arthropods. **Annual Review of Entomology**, v. 12, n. 1, p. 347–376, jan. 1967b.

BURGER, T. D. et al. Phylogenetic analysis of ticks (Acari: Ixodida) using mitochondrial genomes and nuclear rRNA genes indicates that the genus *Amblyomma* is polyphyletic. **Molecular Phylogenetics and Evolution**, v. 64, n. 1, p. 45–55, jul. 2012.

BURKMAN, E. J. **Genetic structure of *Amblyomma cajennense* (Acari : Ixodidae) populations based on mitochondrial gene sequences.** [s.l: s.n.].

BÜTTNER, C. R. et al. Structure of the *Yersinia enterocolitica* Type III secretion translocator chaperone SycD. **Journal of Molecular Biology**, v. 375, n. 4, p. 997–1012, jan. 2008.

CABEZAS-CRUZ, A. et al. *Anaplasma phagocytophilum* increases the levels of histone modifying enzymes to inhibit cell apoptosis and facilitate pathogen infection in the tick vector *Ixodes scapularis*. **Epigenetics**, v. 11, n. 4, p. 303–319, 2 abr. 2016.

CABEZAS-CRUZ, A. et al. Remodeling of tick cytoskeleton in response to infection with *Anaplasma phagocytophilum*. **Frontiers in Bioscience - Landmark**, v. 22, n. 11, p. 1830–1844, 2017a.

CABEZAS-CRUZ, A. et al. *Anaplasma phagocytophilum* infection subverts carbohydrate metabolic pathways in the tick vector, *Ixodes scapularis*. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. FEB, p. 1–17, 7 fev. 2017b.

CABEZAS-CRUZ, A. et al. *Ixodes scapularis* tick cells control *Anaplasma phagocytophilum* infection by increasing the synthesis of phosphoenolpyruvate from tyrosine. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. AUG, p. 1–16, 17 ago. 2017c.

CABEZAS-CRUZ, A. et al. Tick–pathogen interactions: The metabolic perspective. **Trends in Parasitology**, v. 35, n. 4, p. 316–328, abr. 2019.

CAMON, E. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. **Nucleic Acids Research**, v. 32, n. 90001, p. 262D – 266, 1 jan. 2004.

CARDWELL, M. M.; MARTINEZ, J. J. The Sca2 autotransporter protein from *Rickettsia conorii* is sufficient to mediate adherence to and invasion of cultured mammalian cells. **Infection and Immunity**, v. 77, n. 12, p. 5272–5280, 1 dez. 2009.

CASCALES, E.; CHRISTIE, P. J. The versatile bacterial type IV secretion systems. **Nature Reviews Microbiology**, v. 1, n. 2, p. 137–149, nov. 2003.

CATUREGLI, P. et al. AnkA: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. **Infection and Immunity**, v. 68, n. 9, p. 5277–5283, 1 set. 2000.

CAZALET, C. et al. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. **Nature Genetics**, v. 36, n. 11, p. 1165–1173, 3 nov. 2004.

CERVENY, L. et al. Tetratricopeptide repeat motifs in the world of bacterial pathogens: Role in virulence mechanisms. **Infection and Immunity**, v. 81, n. 3, p. 629–635, mar. 2013.

CHAKRABORTY, S. et al. Type IV Pili in *Francisella tularensis*: Roles of pilF and pilT in fiber assembly, host cell adherence, and virulence. **Infection and Immunity**, v. 76, n. 7, p. 2852–2861, 1 jul. 2008.

CHAN, Y. G. Y. et al. Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner. **Cellular Microbiology**, v. 11, n. 4, p. 629–644, abr. 2009.

CHAN, Y. G. Y.; RILEY, S. P.; MARTINEZ, J. J. Adherence to and invasion of host cells by spotted fever group *Rickettsia* species. **Frontiers in Microbiology**, v. 1, n. DEC, p. 1–10, 2010.

CHEN, B. B.; MALLAMPALLI, R. K. Masking of a nuclear signal motif by monoubiquitination leads to mislocalization and degradation of the regulatory enzyme cytidylyltransferase. **Molecular and Cellular Biology**, v. 29, n. 11, p. 3062–3075, 1 jun. 2009.

CHEN, R.; LI, L.; WENG, Z. ZDOCK: An initial-stage protein-docking algorithm. **Proteins: Structure, Function, and Genetics**, v. 52, n. 1, p. 80–87, 1 jul. 2003.

CHMELARĀ, J. et al. Sialomes and mialomes: A systems-biology view of tick tissues and tick–host interactions. **Trends in Parasitology**, v. 32, n. 3, p. 242–254, mar. 2016.

CHO, N.-H. et al. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host cell interaction genes. **Proceedings of the**

National Academy of Sciences, v. 104, n. 19, p. 7981–7986, 8 maio 2007.

CHOCHLAKIS, D. et al. Spotted fever group Rickettsiae in ticks in Cyprus. **Microbial Ecology**, v. 63, n. 2, p. 314–323, 11 fev. 2012.

CHRUDIMSKÁ, T. et al. Defensin from the ornate sheep tick *Dermacentor marginatus* and its effect on Lyme borreliosis spirochetes. **Developmental & Comparative Immunology**, v. 46, n. 2, p. 165–170, out. 2014.

CHUA, H. N.; SUNG, W. K.; WONG, L. Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. **Bioinformatics**, v. 22, n. 13, p. 1623–1630, 2006.

CLAYTON, S. A. et al. The role of microRNAs in glucocorticoid action. **Journal of Biological Chemistry**, v. 293, n. 6, p. 1865–1874, 9 fev. 2018.

COLE, L. M.; NICHOLSON, R. A.; CASIDA, J. E. Action of phenylpyrazole insecticides at the GABA-gated chloride channel. **Pesticide Biochemistry and Physiology**, v. 46, n. 1, p. 47–54, 1993.

COLEMAN, J. L. et al. Evidence that two ATP-dependent (Lon) proteases in *Borrelia burgdorferi* serve different functions. **PLoS Pathogens**, v. 5, n. 11, p. e1000676, 26 nov. 2009.

COLLINS, C. A.; BROWN, E. J. Cytosol as battleground: ubiquitin as a weapon for both host and pathogen. **Trends in Cell Biology**, v. 20, n. 4, p. 205–213, abr. 2010.

COLLINS, G. A.; GOLDBERG, A. L. The logic of the 26S proteasome. **Cell**, v. 169, n. 5, p. 792–806, maio 2017.

COMEAU, S. R. et al. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. **Bioinformatics**, v. 20, n. 1, p. 45–50, 1 jan. 2004.

COOPER, D. L.; BOYLE, D. C.; LOVETT, S. T. Genetic analysis of *Escherichia coli* RadA: functional motifs and genetic interactions. **Molecular Microbiology**, v. 95, n. 5, p. 769–779, mar. 2015.

COX, H. R. Cultivation of Rickettsiae of the Rocky Mountain spotted fever, typhus and Q fever groups in the embryonic tissues of developing chicks. **Science**, v. 94, n. 2444, p. 399–

403, 1941.

DAILY, J. et al. A work stealing based approach for enabling scalable optimal sequence homology detection. **Journal of Parallel and Distributed Computing**, v. 79–80, p. 132–142, maio 2015.

DARBY, A. C. et al. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. **Trends in Genetics**, v. 23, n. 10, p. 511–520, out. 2007.

DAVOODI, S. et al. The immune deficiency pathway regulates metabolic homeostasis in *Drosophila*. **The Journal of Immunology**, v. 202, n. 9, p. 2747–2759, 1 maio 2019.

DE JUAN, D.; PAZOS, F.; VALENCIA, A. Emerging methods in protein co-evolution. **Nature Reviews Genetics**, v. 14, n. 4, p. 249–261, 2013.

DE LA FUENTE, J. et al. Tick-pathogen interactions and vector competence: identification of molecular drivers for tick-borne diseases. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. APR, p. 1–13, 7 abr. 2017.

DEAN, M.; HAMON, Y.; CHIMINI, G. The human ATP-binding cassette (ABC) transporter superfamily. **Journal of Lipid Research**, v. 42, n. 7, p. 1007–1017, 2001.

DIAS, F. A. et al. Ovarian dual oxidase (Duox) activity is essential for insect eggshell hardening and waterproofing. **Journal of Biological Chemistry**, v. 288, n. 49, p. 35058–35067, 6 dez. 2013.

DING, Z.; ATMAKURI, K.; CHRISTIE, P. J. The outs and ins of bacterial type IV secretion substrates. **Trends in Microbiology**, v. 11, n. 11, p. 527–535, nov. 2003.

DONALD, C. L.; KOHL, A.; SCHNETTLER, E. New insights into control of arbovirus replication and spread by insect RNA interference pathways. **Insects**, v. 3, n. 2, p. 511–531, 29 maio 2012.

DONG, Y. et al. Cloning and polymorphism analysis of glutamate-gated chloride channel gene of *Laodelphax striatellus* (Hemiptera: Delphacidae). **Florida Entomologist**, p. 1168–1174, 2013.

DONGEN, S. M. VAN. **Graph clustering by flow simulation**, 2000.

DREHER-LESNICK, S. M. et al. Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. **BMC Microbiology**, v. 8, n. 1, p. 61, 2008.

DUAN, C. et al. Genomic and comparative genomic analyses of *Rickettsia heilongjiangensis* provide insight into its evolution and pathogenesis. **Infection, Genetics and Evolution**, v. 26, p. 274–282, ago. 2014.

DUBSKA, L. et al. Tick-borne zoonotic pathogens in ticks feeding on the common nightingale including a novel strain of *Rickettsia* spp. **Ticks and Tick-borne Diseases**, v. 3, n. 4, p. 265–268, set. 2012.

DURDEN, L. A. Taxonomy, host associations, life cycles and vectorial importance of ticks parasitizing small mammals. In: **Micromammals and Macroparasites**. Tokyo: Springer Japan, 2006. p. 91–102.

DZUL-ROSADO, K. et al. *Rickettsia rickettsii* isolation from naturally infected *Amblyomma parvum* ticks by centrifugation in a 24-well culture plate technique. **Open veterinary journal**, v. 3, n. 2, p. 101–5, 2013.

EDDY, S. R. Hidden markov models. **Current Opinion in Structural Biology**, v. 6, n. 3, p. 361–365, jun. 1996.

EDQVIST, P. J. et al. Tetratricopeptide repeats in the type III secretion chaperone, LcrH: their role in substrate binding and secretion. **Molecular Microbiology**, v. 59, n. 1, p. 31–44, jan. 2006.

EGEA, P. F. et al. Substrate twinning activates the signal recognition particle and its receptor. **Nature**, v. 427, n. 6971, p. 215–221, jan. 2004.

EGEKWU, N. et al. Transcriptome of the female synganglion of the black-legged tick *Ixodes scapularis* (Acari: Ixodidae) with comparison between Illumina and 454 systems. **PLoS ONE**, v. 9, n. 7, 2014.

EL KARKOURI, K. et al. Origin and Evolution of Rickettsial Plasmids. **PLOS ONE**, v. 11, n. 2, p. e0147492, 11 fev. 2016.

ELLISON, D. W. et al. Limited transcriptional responses of *Rickettsia rickettsii* exposed

to environmental stimuli. **PloS one**, v. 4, n. 5, p. e5612, 2009.

ELROD-ERICKSON, M.; MISHRA, S.; SCHNEIDER, D. Interactions between the cellular and humoral immune responses in *Drosophila*. **Current Biology**, v. 10, n. 13, p. 781–784, 2000.

ELVERS, M. et al. Intracellular cyclophilin A is an important Ca²⁺ regulator in platelets and critically involved in arterial thrombus formation. **Blood**, v. 120, n. 6, p. 1317–1326, 9 ago. 2012.

EREMEEVA, M. E. et al. Isolation and identification of *Rickettsia massiliae* from *Rhipicephalus sanguineus* ticks Collected in Arizona. **Applied and Environmental Microbiology**, v. 72, n. 8, p. 5569–5577, 1 ago. 2006a.

EREMEEVA, M. E. et al. Molecular typing of novel *Rickettsia rickettsii* isolates from Arizona. **Annals of the New York Academy of Sciences**, v. 1078, n. 1, p. 573–577, 1 out. 2006b.

ESCALANTE, A. A.; AYALA, F. J. Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. **Proceedings of the National Academy of Sciences**, v. 92, n. 13, p. 5793–5797, 20 jun. 1995.

ESTEVEZ, E. et al. Analysis of the salivary gland transcriptome of unfed and partially fed *Amblyomma sculptum* ticks and descriptive proteome of the saliva. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. NOV, p. 1–17, 21 nov. 2017.

ETIENNE-MANNEVILLE, S.; HALL, A. Rho GTPases in cell biology. **Nature**, v. 420, n. 6916, p. 629–635, dez. 2002.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, v. 39, n. 4, p. 783–791, 1985.

FENG, CHEN, AARON. J, MACKAY, CHRISTIAN. J, STOECKERT JR, DAVID. S, R. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. **Nucleic Acids Research**, v. 34, n. 90001, p. D363–D368, 1 jan. 2006.

FENG, T. et al. DEAD-Box helicase DDX25 is a negative regulator of Type I interferon pathway and facilitates RNA virus infection. **Frontiers in Cellular and Infection**

Microbiology, v. 7, n. AUG, p. 1–11, 4 ago. 2017.

FENN, K.; BLAXTER, M. *Wolbachia* genomes: revealing the biology of parasitism and mutualism. **Trends in Parasitology**, v. 22, n. 2, p. 60–65, fev. 2006.

FERNANDES, R.; FERREIRA, S.; BOTELHO, M. C. Commentary: *Theileria* parasites secrete a prolyl isomerase to maintain host leukocyte transformation. **Frontiers in Medicine**, v. 5, n. APR, p. 1–2, 2018.

FINN, R. D. Pfam: clans, web tools and services. **Nucleic Acids Research**, v. 34, n. 90001, p. D247–D251, 1 jan. 2006.

FINN, R. D.; MARSHALL, M.; BATEMAN, A. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. **Bioinformatics**, v. 21, n. 3, p. 410–412, 1 fev. 2005.

FISCHER, G.; SCHMID, F. X. The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. **Biochemistry**, v. 29, n. 9, p. 2205–2212, 6 mar. 1990.

FISCHER, J. A.; EUN, S. H.; DOOLAN, B. T. Endocytosis, endosome trafficking, and the regulation of *Drosophila* development. **Annual Review of Cell and Developmental Biology**, v. 22, n. 1, p. 181–206, nov. 2006.

FITCH, W. M.; MARKOWITZ, E. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. **Biochemical Genetics**, v. 4, n. 5, p. 579–593, out. 1970.

FLÓREZ, A. F. et al. Protein network prediction and topological analysis in *Leishmania major* as a tool for drug target selection. **BMC Bioinformatics**, v. 11, n. 1, p. 484, 2010.

FLORIN-CHRISTENSEN, M. Piroplasmids and ticks: a long-lasting intimate relationship. **Frontiers in Bioscience**, v. Volume, n. 14, p. 3064, 2009.

FOCIA, P. J. Heterodimeric GTPase core of the SRP targeting complex. **Science**, v. 303, n. 5656, p. 373–377, 16 jan. 2004.

FOURNIER, P.-E. et al. Analysis of the *Rickettsia africae* genome reveals that virulence

acquisition in *Rickettsia* species may be explained by genome reduction. **BMC Genomics**, v. 10, n. 1, p. 166, 2009.

FRAGKOUDIS, R. et al. Advances in dissecting mosquito innate immune responses to arbovirus infection. **Journal of General Virology**, v. 90, n. 9, p. 2061–2072, 1 set. 2009.

FRAITURE, M. et al. Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. **Cell Host & Microbe**, v. 5, n. 3, p. 273–284, mar. 2009.

FRANCISCHETTI, I. M. B. et al. The transcriptome of the salivary glands of the female western black-legged tick *Ixodes pacificus* (Acari: Ixodidae). **Insect Biochemistry and Molecular Biology**, v. 35, n. 10, p. 1142–1161, out. 2005.

FRASER, C. M. et al. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. **Nature**, v. 390, n. 6660, p. 580–586, dez. 1997.

FREY, B. J.; DUECK, D. Clustering by passing messages between data points. **Science**, v. 315, n. 5814, p. 972–976, 16 fev. 2007.

FRIEDBERG, E. C. Reversible monoubiquitination of PCNA: A novel slant on regulating translesion DNA synthesis. **Molecular Cell**, v. 22, n. 2, p. 150–152, 2006.

FRISCHKNECHT, F. et al. Actin-based motility of vaccinia virus mimics receptor tyrosine kinase signalling. **Nature**, v. 401, n. 6756, p. 926–929, out. 1999.

FRYXELL, K. J. The coevolution of gene family trees. **Trends in Genetics**, v. 12, n. 9, p. 364–369, set. 1996.

FU, Y.-X.; LI, W.-H. Maximum likelihood estimation of population parameters. **Genetics**, v. 134, n. 4, p. 1261–1270, 1993.

FUXELIUS, H.-H. et al. Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. **Genome Biology**, v. 9, n. 2, p. R42, 2008.

GABALDÓN, T.; KOONIN, E. V. Functional and evolutionary implications of gene orthology. **Nature Reviews Genetics**, v. 14, n. 5, p. 360–366, 4 maio 2013.

GALLETTI, M. F. B. M. et al. Virulence genes of *Rickettsia rickettsii* are differentially

modulated by either temperature upshift or blood-feeding in tick midgut and salivary glands. **Parasites and Vectors**, v. 9, n. 1, 2016.

GALPERIN, M. Y.; KOONIN, E. V. Who's your neighbor? New computational approaches for functional genomics. **Nature biotechnology**, v. 18, n. 6, p. 609, 2000.

GARCIA-GARCIA, J. C. et al. Epigenetic silencing of host cell defense genes enhances intracellular survival of the rickettsial pathogen *Anaplasma phagocytophilum*. **PLoS Pathogens**, v. 5, n. 6, p. e1000488, 19 jun. 2009.

GASPER, R.; SOT, B.; WITTINGHOFER, A. GTPase activity of Di-Ras proteins is stimulated by Rap1GAP proteins. **Small GTPases**, v. 1, n. 3, p. 133–141, 27 nov. 2010.

GEIJSEN, N. Cytokine-specific transcriptional regulation through an IL-5Ralpha interacting protein. **Science**, v. 293, n. 5532, p. 1136–1138, 10 ago. 2001.

GEISZT, M. et al. Dual oxidases represent novel hydrogen peroxide sources supporting mucosal surface host defense. **The FASEB Journal**, v. 17, n. 11, p. 1502–1504, ago. 2003.

GEORGIADES, K. et al. Gene gain and loss events in *Rickettsia* and *Orientia* species. **Biology Direct**, v. 6, n. 1, p. 6, 2011.

GILLESPIE, J. J. et al. *Rickettsia* phylogenomics: Unwinding the intricacies of obligate intracellular life. **PLoS ONE**, v. 3, n. 4, p. e2018, 16 abr. 2008.

GILLESPIE, J. J. et al. An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. **PLoS ONE**, v. 4, n. 3, p. e4833, 12 mar. 2009.

GILLESPIE, J. J. et al. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. **Journal of Bacteriology**, v. 194, n. 2, p. 376–394, 2012.

GILLESPIE, J. J. et al. Secretome of obligate intracellular *Rickettsia*. **FEMS Microbiology Reviews**, v. 39, n. 1, p. n/a-n/a, set. 2014.

GIMÉNEZ, D. F. Staining *Rickettsiae* in yolk-sac cultures. **Stain Technology**, v. 39, n. 3, p. 135–140, 12 jan. 1964.

GÖBEL, U. et al. Correlated mutations and residue contacts in proteins. **Proteins:**

Structure, Function, and Genetics, v. 18, n. 4, p. 309–317, abr. 1994.

GOFF, S. A.; CASSON, L. P.; GOLDBERG, A. L. Heat shock regulatory gene htpR influences rates of protein degradation and expression of the lon gene in *Escherichia coli*. **Proceedings of the National Academy of Sciences**, v. 81, n. 21, p. 6647–6651, 1 nov. 1984.

GOH, C.-S. et al. Co-evolution of proteins with their interaction partners 1 Edited by B. Honig. **Journal of Molecular Biology**, v. 299, n. 2, p. 283–293, 2 jun. 2000.

GOLDBERG, A. L. et al. ATP-dependent protease La (Lon) from *Escherichia coli*. In: **Methods in Enzymology**. [s.l: s.n.]. v. 244p. 350–375.

GOMES, G. G. **Condicionantes ambientais da infecção de carrapatos *Amblyomma* spp. (Acari: Ixodidae) por *Rickettsia* spp. (Rickettsiales: Rickettsiae)**. [s.l.] Universidade Federal de Viçosa, 2014.

GÓMEZ-DÍAZ, E. et al. Epigenetics of Host–Pathogen Interactions: The Road Ahead and the Road Behind. **PLoS Pathogens**, v. 8, n. 11, p. e1003007, 29 nov. 2012.

GONG, J.; HUO, J. New insights into the mechanism of F-box proteins in colorectal cancer (Review). **Oncology Reports**, v. 33, n. 5, p. 2113–2120, maio 2015.

GONG, S. et al. PSIBase: a database of protein structural interactome map (PSIMAP). **Bioinformatics**, v. 21, n. 10, p. 2541–2543, 15 maio 2005.

GOTTESMAN, M. M.; FOJO, T.; BATES, S. E. Multidrug resistance in cancer: role of ATP-dependent transporters. **Nature Reviews Cancer**, v. 2, n. 1, p. 48–58, jan. 2002.

GOTTESMAN, S.; HALPERN, E.; TRISLER, P. Role of sulA and sulB in filamentation by lon mutants of *Escherichia coli* K-12. **Journal of Bacteriology**, v. 148, n. 1, p. 265–273, 1981.

GOTTLIEB, Y.; LALZAR, I.; KLASSON, L. Distinctive genome reduction rates revealed by genomic analyses of two *Coxiella* - like endosymbionts in ticks. **Genome Biology and Evolution**, v. 7, n. 6, p. 1779–1796, jun. 2015.

GOUIN, E. et al. The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. **Nature**, v. 427, n. 6973, p. 457–461, jan. 2004.

GRABHERR, M. G. . et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature Biotechnology**, v. 29, n. 7, p. 644–652, 15 jul. 2011.

GREUB, G.; RAOULT, D. History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a Chlamydiales ancestor to plants 1 billion years ago. **Applied and Environmental Microbiology**, v. 69, n. 9, p. 5530–5535, 1 set. 2003.

GROOTJANS, J. J. et al. Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. **Proceedings of the National Academy of Sciences**, v. 94, n. 25, p. 13683–13688, 9 dez. 1997.

GROSHONG, A. M. et al. BB0238, a presumed tetratricopeptide repeat-containing protein, is required during *Borrelia burgdorferi* mammalian infection. **Infection and Immunity**, v. 82, n. 10, p. 4292–4306, out. 2014.

GUEDES, E. et al. Detection of *Rickettsia rickettsii* in the tick *Amblyomma cajennense* in a new Brazilian spotted fever-endemic area in the state of Minas Gerais. **Memórias do Instituto Oswaldo Cruz**, v. 100, n. 8, p. 841–845, dez. 2005.

GULIA-NUSS, M. et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. **Nature Communications**, v. 7, n. May 2015, 2016.

GUTIÉRREZ, R. et al. The effect of ecological and temporal factors on the composition of *Bartonella* infection in rodents and their fleas. **The ISME journal**, v. 8, n. 8, p. 1598, 2014.

HAGLUND, C. M. et al. *Rickettsia Sca2* is a bacterial formin-like mediator of actin-based motility. **Nature Cell Biology**, v. 12, n. 11, p. 1057–1063, 24 nov. 2010.

HAJDUŠEK, O. et al. Interaction of the tick immune system with transmitted pathogens. **Frontiers in Cellular and Infection Microbiology**, v. 3, n. JUL, p. 1–15, 2013.

HAMILTON, R.; BOOTS, M.; PATERSON, S. The effect of host heterogeneity and parasite intragenomic interactions on parasite population structure. **Proceedings of the Royal Society B: Biological Sciences**, v. 272, n. 1573, p. 1647–1653, 22 ago. 2005.

HAMMOND, S. M. et al. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. **Nature**, v. 404, n. 6775, p. 293–296, mar. 2000.

HARA-KUGE, S. et al. En bloc incorporation of coatomer subunits during the assembly of COP-coated vesicles. **The Journal of cell biology**, v. 124, n. 6, p. 883–892, 1994.

HARRINGTON, E. D.; JENSEN, L. J.; BORK, P. Predicting biological networks from genomic data. **FEBS Letters**, v. 582, n. 8, p. 1251–1258, 9 abr. 2008.

HAWLENA, H. et al. The arthropod, but not the vertebrate host or its environment, dictates bacterial community composition of fleas and ticks. **The ISME Journal**, v. 7, n. 1, p. 221–223, 28 jan. 2013.

HEINZEN, R. A. Rickettsial actin-based motility. **Annals of the New York Academy of Sciences**, v. 990, n. 1, p. 535–547, jun. 2003.

HEISE, S. R.; ELSHAHED, M. S.; LITTLE, S. E. Bacterial Diversity in *Amblyomma americanum* (Acari: Ixodidae) With a Focus on Members of the Genus *Rickettsia*. **Journal of Medical Entomology**, v. 47, n. 2, p. 258–268, 1 mar. 2010.

HENIKOFF, S.; HENIKOFF, J. G. Amino acid substitution matrices. In: **Advances in Protein Chemistry**. [s.l: s.n.]. v. 54p. 73–97.

HEPBURN, M. J.; SIMPSON, A. J. Tularemia: current diagnosis and treatment options. **Expert Review of Anti-infective Therapy**, v. 6, n. 2, p. 231–240, 10 abr. 2008.

HERMJAKOB, H. IntAct: an open source molecular interaction database. **Nucleic Acids Research**, v. 32, n. 90001, p. 452D – 455, 1 jan. 2004.

HIDALGO, M. et al. Rocky mountain spotted fever, Colombia. **Emerging infectious diseases**, v. 13, n. 7, p. 1058, 2007.

HIGGINS, C. F. Multiple molecular mechanisms for multidrug resistance transporters. **Nature**, v. 446, n. 7137, p. 749–757, abr. 2007.

HOLLAND, I. B.; SCHMITT, L.; YOUNG, J. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (Review). **Molecular Membrane Biology**, v. 22, n. 1–2, p. 29–39, 9 jan. 2005.

HOSKINS, R. et al. The *Caenorhabditis elegans* vulval induction gene *lin-2* encodes a member of the MAGUK family of cell junction proteins. **Development**, v. 122, n. 1, p. 97–111,

1996.

HOXMEIER, J. C. et al. Metabolomics of the tick-*Borrelia* interaction during the nymphal tick blood meal. **Scientific Reports**, v. 7, n. 1, p. 44394, 13 jun. 2017.

HUANG, B. et al. The *Anaplasma phagocytophilum*-occupied vacuole selectively recruits Rab-GTPases that are predominantly associated with recycling endosomes. **Cellular Microbiology**, v. 12, n. 9, p. 1292–1307, set. 2010.

HUANG, J. et al. Infection prevalences of common tick-borne pathogens in adult lone star ticks (*Amblyomma americanum*) and american dog ticks (*Dermacentor variabilis*) in Kentucky. **The American Journal of Tropical Medicine and Hygiene**, v. 85, n. 4, p. 718–723, 1 out. 2011.

HUBBARD, T. J. P. et al. SCOP: A structural classification of proteins database. **Nucleic Acids Research**, v. 25, n. 1, p. 236–239, 1 jan. 1997.

HUNG, A. Y.; SHENG, M. PDZ Domains: Structural modules for protein complex assembly. **Journal of Biological Chemistry**, v. 277, n. 8, p. 5699–5702, 22 fev. 2002.

HUYNEN, M. et al. Exploitation of gene context. **Current Opinion in Structural Biology**, v. 10, n. 3, p. 366–370, jun. 2000.

INOUE, N. et al. EB9 Characterization of phagocytic hemocytes in *Ornithodoros moubata*. **Medical Entomology and Zoology**, v. 52, n. Supplement, p. 142, 2001.

JACOBSON, R. L. Review of “What bugged the dinosaurs? Insects, disease and death in the cretaceous” by Poinar G. Jr. and Poinar R. **Parasites & Vectors**, v. 1, n. 1, p. 6, 2008.

JAMES, C. E.; DAVEY, M. W. Increased expression of ABC transport proteins is associated with ivermectin resistance in the model nematode *Caenorhabditis elegans*. **International Journal for Parasitology**, v. 39, n. 2, p. 213–220, jan. 2009.

JANKOWSKY, E. RNA helicases at work: binding and rearranging. **Trends in Biochemical Sciences**, v. 36, n. 1, p. 19–29, jan. 2011.

JENG, R. L. et al. A *Rickettsia* WASP-like protein activates the Arp2/3 complex and mediates actin-based motility. **Cellular Microbiology**, v. 6, n. 8, p. 761–769, 5 jul. 2004.

JONES, K. E. et al. Global trends in emerging infectious diseases. **Nature**, v. 451, n. 7181, p. 990–993, fev. 2008.

JONES, S.; THORNTON, J. M. Principles of protein-protein interactions. **Proceedings of the National Academy of Sciences**, v. 93, n. 1, p. 13–20, 9 jan. 1996.

JONGEJAN, F.; UILENBERG, G. The global importance of ticks. **Parasitology**, v. 129, n. S1, p. S3–S14, 19 out. 2004.

JOTHI, R. et al. Co-evolutionary analysis of domains in interacting proteins reveals insights into domain–domain interactions mediating protein-protein interactions. **Journal of Molecular Biology**, v. 362, n. 4, p. 861–875, set. 2006.

JUNCKER, A. S. et al. Prediction of lipoprotein signal peptides in gram-negative bacteria. n. 1994, p. 1652–1662, 2003.

KAHLON, A. et al. *Anaplasma phagocytophilum* Asp14 is an invasin that interacts with mammalian host cells via its c terminus to facilitate infection. **Infection and Immunity**, v. 81, n. 1, p. 65–79, jan. 2013.

KÄLL, L.; KROGH, A.; SONNHAMMER, E. L. L. A Combined Transmembrane Topology and Signal Peptide Prediction Method. **Journal of Molecular Biology**, v. 338, n. 5, p. 1027–1036, maio 2004.

KANG, Y.-J. et al. Extensive diversity of Rickettsiales bacteria in two species of ticks from China and the evolution of the Rickettsiales. **BMC Evolutionary Biology**, v. 14, n. 1, p. 167, 30 dez. 2014.

KANN, M. G. et al. Correlated evolution of interacting proteins: Looking behind the mirrortree. **Journal of Molecular Biology**, v. 385, n. 1, p. 91–98, jan. 2009.

KANNAN, S.; ROGOZIN, I. B.; KOONIN, E. V. MitoCOGs: Clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. **BMC Evolutionary Biology**, v. 14, n. 1, p. 1–16, 2014.

KARIM, S.; SINGH, P.; RIBEIRO, J. M. C. A deep insight into the sialotranscriptome of the Gulf Coast tick, *Amblyomma maculatum*. **PLoS ONE**, v. 6, n. 12, p. e28525, 21 dez. 2011.

KARPATY, S. E. et al. *Rickettsia amblyommatis* sp. nov., a spotted fever group *Rickettsia* associated with multiple species of *Amblyomma* ticks in North, Central and South America. **International Journal of Systematic and Evolutionary Microbiology**, v. 66, n. 12, p. 5236–5243, 1 dez. 2016.

KÄSTLE, M. et al. Chaperones, but not oxidized proteins, are ubiquitinated after oxidative stress. **Free Radical Biology and Medicine**, v. 53, n. 7, p. 1468–1477, out. 2012.

KAUR, S. J. et al. TolC-dependent secretion of an ankyrin repeat-containing protein of *Rickettsia typhi*. **Journal of Bacteriology**, v. 194, n. 18, p. 4920–4932, 15 set. 2012.

KAWAI, T.; AKIRA, S. TLR signaling. **Cell Death & Differentiation**, v. 13, n. 5, p. 816–825, 20 maio 2006.

KEIRANS, J. E.; CLIFFORD, C. M.; CORWIN, D. *Ixodes sigelos*, n. sp. (Acarina : Ixodidae), a parasite of rodents in Chile, with a method for preparing ticks for examination by scanning electron microscopy. **Acarologia**, v. 18, n. 2, p. 217–25, dez. 1976.

KERR, J. E.; CHRISTIE, P. J. Evidence for VirB4-mediated dislocation of membrane-integrated VirB2 pilin during biogenesis of the Agrobacterium VirB/VirD4 type IV secretion system. **Journal of Bacteriology**, v. 192, n. 19, p. 4923–4934, 2010.

KERRIEN, S. et al. The IntAct molecular interaction database in 2012. **Nucleic Acids Research**, v. 40, n. D1, p. D841–D846, 1 jan. 2012.

KHADKA, S. et al. A physical interaction network of dengue virus and human proteins. **Molecular & Cellular Proteomics**, v. 10, n. 12, p. M111.012187, dez. 2011.

KILLMASTER, L. F. et al. Isolation of a *Rickettsia slovaca*- like agent from *Dermacentor variabilis* ticks in vero cell culture. **Vector-Borne and Zoonotic Diseases**, v. 16, n. 1, p. 61–62, jan. 2016.

KIM, J.-G. et al. Predicting the interactome of *Xanthomonas oryzae* pathovar *oryzae* for target selection and DB service. **BMC Bioinformatics**, v. 9, n. 1, p. 41, 2008.

KIM, S. K. Polarized signaling: basolateral receptor localization in epithelial cells by PDZ-containing proteins. **Current Opinion in Cell Biology**, v. 9, n. 6, p. 853–859, dez. 1997.

KISSELEV, A. F. et al. The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. **Journal of Biological Chemistry**, v. 274, n. 6, p. 3363–3371, 5 fev. 1999.

KLAPPENBACH, J. A.; DUNBAR, J. M.; SCHMIDT, T. M. rRNA operon copy number reflects ecological strategies of bacteria. **Applied and Environmental Microbiology**, v. 66, n. 4, p. 1328–1333, 1 abr. 2000.

KLEINERMAN, G. et al. Molecular detection of *Rickettsia africae*, *Rickettsia aeschlimannii*, and *Rickettsia sibirica mongolitimonae* in camels and *Hyalomma* spp. ticks from Israel. **Vector-Borne and Zoonotic Diseases**, v. 13, n. 12, p. 851–856, dez. 2013.

KOBE, B. The leucine-rich repeat as a protein recognition motif. **Current Opinion in Structural Biology**, v. 11, n. 6, p. 725–732, 1 dez. 2001.

KOCAN, K. M. et al. Silencing of genes involved in *Anaplasma marginale*-tick interactions affects the pathogen developmental cycle in *Dermacentor variabilis*. **BMC Developmental Biology**, v. 9, n. 1, p. 42, 16 dez. 2009.

KONKEL, M. E.; TILLY, K. Temperature-regulated expression of bacterial virulence genes. **Microbes and Infection**, v. 2, n. 2, p. 157–166, fev. 2000.

KOONIN, E. V. Orthologs, paralogs, and evolutionary genomics. **Annual Review of Genetics**, v. 39, n. 1, p. 309–338, dez. 2005.

KOTSYFAKIS, M. et al. Tissue and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. **Scientific Reports**, v. 5, n. 1, p. 9103, 13 ago. 2015.

KOURANOV, A. The RCSB PDB information portal for structural genomics. **Nucleic Acids Research**, v. 34, n. 90001, p. D302–D305, 1 jan. 2006a.

KOURANOV, A. The RCSB PDB information portal for structural genomics. **Nucleic Acids Research**, v. 34, n. 90001, p. D302–D305, 1 jan. 2006b.

KROGH, A. et al. Hidden markov models in computational biology. **Journal of Molecular Biology**, v. 235, n. 5, p. 1501–1531, fev. 1994.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870–1874, jul. 2016.

KURSCHEID, S. et al. Evidence of a tick RNAi pathway by comparative genomics and reverse genetics screen of targets with known loss-of-function phenotypes in *Drosophila*. **BMC Molecular Biology**, v. 10, n. 1, p. 26, 2009.

KURTTI, T. J. et al. *Rickettsia buchneri* sp. nov., a rickettsial endosymbiont of the blacklegged tick *Ixodes scapularis*. **INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY**, v. 65, n. Pt 3, p. 965–970, 1 mar. 2015.

LA SCOLA, B.; RAOULT, D. Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. **Journal of clinical microbiology**, v. 35, n. 11, p. 2715–27, nov. 1997.

LABRUNA, M. B. et al. Isolation of *Rickettsia rhipicephali* and *Rickettsia bellii* from *Haemaphysalis juxtakochi* ticks in the state of Sao Paulo, Brazil. **Applied and Environmental Microbiology**, v. 73, n. 3, p. 869–873, 1 fev. 2007.

LABRUNA, M. B. et al. Allopatric speciation in ticks: genetic and reproductive divergence between geographic strains of *Rhipicephalus* (*Boophilus*) *microplus*. **BMC Evolutionary Biology**, v. 9, n. 1, p. 46, 2009.

LABRUNA, M. B. et al. Cross-mating experiments with geographically different populations of *Amblyomma cajennense* (Acari: Ixodidae). **Experimental and Applied Acarology**, v. 54, n. 1, p. 41–49, 1 maio 2011.

LARA, F. A. A new intracellular pathway of haem detoxification in the midgut of the cattle tick *Boophilus microplus*: aggregation inside a specialized organelle, the hemosome. **Journal of Experimental Biology**, v. 206, n. 10, p. 1707–1715, 15 maio 2003.

LARA, F. A. Tracing heme in a living cell: hemoglobin degradation and heme traffic in digest cells of the cattle tick *Boophilus microplus*. **Journal of Experimental Biology**, v. 208, n. 16, p. 3093–3101, 15 ago. 2005.

LARA, F. A. et al. ATP binding cassette transporter mediates both heme and pesticide

detoxification in tick midgut cells. **PLOS ONE**, v. 10, n. 8, p. e0134779, 10 ago. 2015.

LARKIN, M. A. et al. Clustal W and clustal X version 2.0. **Bioinformatics**, v. 23, n. 21, p. 2947–2948, 1 nov. 2007.

LASKER, K. et al. Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. **Proceedings of the National Academy of Sciences**, v. 109, n. 5, p. 1380–1387, 31 jan. 2012.

LAWLEY, T. . et al. F factor conjugation is a true type IV secretion system. **FEMS Microbiology Letters**, v. 224, n. 1, p. 1–15, jul. 2003.

LEE, K.-A. et al. Bacterial uracil modulates *Drosophila* DUOX-dependent gut immunity via hedgehog-induced signaling endosomes. **Cell Host & Microbe**, v. 17, n. 2, p. 191–204, fev. 2015.

LEE, W.-J.; HASE, K. Gut microbiota-generated metabolites in animal health and disease. **Nature Chemical Biology**, v. 10, n. 6, p. 416–424, 16 jun. 2014.

LESCOT, M. et al. The genome of *Borrelia recurrentis*, the agent of Deadly Louse-Borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. **PLoS Genetics**, v. 4, n. 9, p. e1000185, 12 set. 2008.

LEVIN, M. L. et al. Effects of *Rickettsia conorii* infection on the survival of *Rhipicephalus sanguineus* ticks. **Clinical Microbiology and Infection**, v. 15, p. 277–278, dez. 2009.

LEWIS, L. A. et al. Identification of 24h *Ixodes scapularis* immunogenic tick saliva proteins. **Ticks and Tick-borne Diseases**, v. 6, n. 3, p. 424–434, abr. 2015.

LI, LI.; CHRISTIAN, J. STOECKERT, JR., DAVID, S. R. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. **Genome Research**, v. 13, n. 9, p. 2178–2189, 1 set. 2003.

LI, A.-L. et al. A Novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. **Journal of Biological Chemistry**, v. 279, n. 47, p. 49251–49258, 19 nov. 2004.

LI, G. et al. DEAD-box RNA helicase DDX3X inhibits DENV replication via regulating type one interferon pathway. **Biochemical and Biophysical Research Communications**, v. 456, n. 1, p. 327–332, jan. 2015.

LI, H.; WALKER, D. H. rOmpA is a critical protein for the adhesion of *Rickettsia rickettsii* to host cells. **Microbial Pathogenesis**, v. 24, n. 5, p. 289–298, maio 1998.

LI, J.; MAHAJAN, A.; TSAI, M.-D. Ankyrin repeat: A unique motif mediating protein–protein interactions †. **Biochemistry**, v. 45, n. 51, p. 15168–15178, dez. 2006.

LIN, M. et al. *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. **Cellular Microbiology**, v. 9, n. 11, p. 2644–2657, nov. 2007.

LIU, J. et al. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. **Cell**, v. 66, n. 4, p. 807–815, 1991.

LIVELY, C. M. Host-parasite coevolution and sex. **BioScience**, v. 46, n. 2, p. 107–114, fev. 1996.

LO CONTE, L. SCOP: A structural classification of proteins database. **Nucleic Acids Research**, v. 28, n. 1, p. 257–259, 1 jan. 2000.

LÜLLMANN-RAUCH, R. History and morphology of the lysosome. **Lysosomes**, p. 1–16, 2007.

MACALUSO, K. R. et al. Infection and transovarial transmission of Rickettsiae in *Dermacentor variabilis* ticks acquired by artificial feeding. **Vector-Borne and Zoonotic Diseases**, v. 1, n. 1, p. 45–53, mar. 2001.

MACALUSO, K. R. et al. Differential expression of genes in uninfected and *Rickettsia*-infected *Dermacentor variabilis* ticks as assessed by differential-display PCR. **Infection and Immunity**, v. 71, n. 11, p. 6165–6170, 1 nov. 2003.

MAERE, S.; HEYMANS, K.; KUIPER, M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. **Bioinformatics**, v. 21, n. 16, p. 3448–3449, 15 ago. 2005.

MAGUNDA, F. et al. *Anaplasma marginale* actively modulates vacuolar maturation during intracellular infection of its tick vector, *Dermacentor andersoni*. **Applied and Environmental Microbiology**, v. 82, n. 15, p. 4715–4731, 1 ago. 2016.

MANGIA, C. et al. Exposure to amitraz, fipronil and permethrin affects cell viability and ABC transporter gene expression in an *Ixodes ricinus* cell line. **Parasites & Vectors**, v. 11, n. 1, p. 437, 31 dez. 2018.

MANS, B. J. et al. Comparative sialomics between hard and soft ticks: Implications for the evolution of blood-feeding behavior. **Insect Biochemistry and Molecular Biology**, v. 38, n. 1, p. 42–58, jan. 2008.

MANSFIELD, K. L. et al. Tick-borne pathogens induce differential expression of genes promoting cell survival and host resistance in *Ixodes ricinus* cells. **Parasites & Vectors**, v. 10, n. 1, p. 81, 15 dez. 2017.

MARCOTTE, E. M. Detecting protein function and protein-protein interactions from genome sequences. **Science**, v. 285, n. 5428, p. 751–753, 30 jul. 1999.

MARDIS, E. R. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**, v. 9, n. 1, p. 387–402, set. 2008.

MARIÉ, J.-L. et al. Rickettsiae in arthropods collected from red foxes (*Vulpes vulpes*) in France. **Comparative Immunology, Microbiology and Infectious Diseases**, v. 35, n. 1, p. 59–62, jan. 2012.

MARRELLI, M. T. et al. Taxonomic and phylogenetic relationships between neotropical species of ticks from genus *Amblyomma* (Acari: Ixodidae) inferred from second internal transcribed spacer sequences of rDNA. **Journal of Medical Entomology**, v. 44, n. 2, p. 222–228, 1 mar. 2007.

MARSOLIER, J. et al. *Theileria* parasites secrete a prolyl isomerase to maintain host leukocyte transformation. **Nature**, v. 520, n. 7547, p. 378, 2015.

MARTINEZ, J. J. Early signaling events involved in the entry of *Rickettsia conorii* into mammalian cells. **Journal of Cell Science**, v. 117, n. 21, p. 5097–5106, 1 out. 2004.

MARTINEZ, J. J. et al. Ku70, a component of DNA-dependent protein kinase, is a

mammalian receptor for *Rickettsia conorii*. **Cell**, v. 123, n. 6, p. 1013–1023, dez. 2005.

MARTINS, L. A. et al. The distinct transcriptional response of the midgut of *Amblyomma sculptum* and *Amblyomma aureolatum* ticks to *Rickettsia rickettsii* correlates to their differences in susceptibility to infection. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. APR, 28 abr. 2017.

MARTINS, T. F. et al. Geographical distribution of *Amblyomma cajennense* (sensu lato) ticks (Parasitiformes: Ixodidae) in Brazil, with description of the nymph of *A. cajennense* (sensu stricto). **Parasites & Vectors**, v. 9, n. 1, p. 186, 31 dez. 2016.

MASTROPAOLO, M. et al. Biological differences between two allopatric populations of *Amblyomma cajennense* (Acari: Ixodidae) in Argentina. **Experimental and Applied Acarology**, v. 53, n. 4, p. 371–375, 23 abr. 2011.

MATSUMOTO, K. et al. Experimental infection models of ticks of the *Rhipicephalus sanguineus* group with *Rickettsia conorii*. **Vector-Borne and Zoonotic Diseases**, v. 5, n. 4, p. 363–372, dez. 2005.

MAYOR, S.; PAGANO, R. E. Pathways of clathrin-independent endocytosis. **Nature Reviews Molecular Cell Biology**, v. 8, n. 8, p. 603–612, ago. 2007.

MAYS, S. E. et al. Prevalence of five tick-borne bacterial genera in adult *Ixodes scapularis* removed from white-tailed deer in western Tennessee. **Parasites & Vectors**, v. 7, n. 1, p. 473, 22 dez. 2014.

MAZIOLI, R. B. **Análise filogeográfica e filogenética do carrapato *Amblyomma cajennense* (FABRICIUS, 1787)**. [s.l.: s.n.].

MCCARTHY, F. M. et al. AgBase: A functional genomics resource for agriculture. **BMC Genomics**, v. 7, p. 1–13, 2006.

MCDADE, J. E.; NEWHOUSE, V. F. Natural history of *Rickettsia rickettsii*. **Annual Reviews in microbiology**, v. 40, n. 1, p. 287–309, out. 1986.

MCGEOCH, D. J. On the predictive recognition sequences of signal peptide. **Virus Research**, v. 3, n. 3, p. 271–286, 1985.

MEAD, D. G. et al. Absence of *Rickettsia rickettsii* and occurrence of other spotted fever group Rickettsiae in ticks from Tennessee. **The American Journal of Tropical Medicine and Hygiene**, v. 83, n. 3, p. 653–657, 1 set. 2010.

MEDIANNIKOV, O. et al. *Rickettsia raoultii* sp. nov., a spotted fever group *Rickettsia* associated with *Dermacentor* ticks in Europe and Russia. **INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY**, v. 58, n. 7, p. 1635–1639, 1 jul. 2008.

MEJIA, J. S.; BISHOP, J. V; TITUS, R. G. Is it possible to develop pan-arthropod vaccines? **Trends in Parasitology**, v. 22, n. 8, p. 367–370, ago. 2006.

MELIK, W. et al. Two PDZ binding motifs within NS5 have roles in tick-borne encephalitis virus replication. **Virus Research**, v. 169, n. 1, p. 54–62, out. 2012.

MELO, A. L. T. et al. *Rickettsia parkeri* infecting free-living *Amblyomma triste* ticks in the Brazilian Pantanal. **Ticks and Tick-borne Diseases**, v. 6, n. 3, p. 237–241, abr. 2015.

MENDES, T. A. O. et al. Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. **Molecular Biology and Evolution**, v. 30, n. 4, p. 951–963, 2013.

MERCADO-CURIEL, R. F. et al. Temporal characterisation of the organ-specific *Rhipicephalus microplus* transcriptional response to *Anaplasma marginale* infection. **International Journal for Parasitology**, v. 41, n. 8, p. 851–860, jul. 2011.

MERCADO-CURIEL, R. F. et al. Identification of *Rhipicephalus microplus* Genes That Modulate the Infection Rate of the *Rickettsia Anaplasma marginale*. **PLoS ONE**, v. 9, n. 3, p. e91062, 7 mar. 2014.

MERHEJ, V. et al. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. **Biology Direct**, v. 4, n. 1, p. 13, 2009.

MERHEJ, V.; RAOULT, D. Rickettsial evolution in the light of comparative genomics. **Biological Reviews**, v. 86, n. 2, p. 379–405, maio 2011.

MERING, C. V. STRING: a database of predicted functional associations between proteins. **Nucleic Acids Research**, v. 31, n. 1, p. 258–261, 1 jan. 2003.

MISHRA, S. Computational prediction of protein-protein complexes. **BMC Research Notes**, v. 5, n. 1, p. 495, 9 dez. 2012.

MORAN, N. A. Microbial minimalism. **Cell**, v. 108, n. 5, p. 583–586, mar. 2002.

MORAN, N. A.; WERNEGREEN, J. J. Lifestyle evolution in symbiotic bacteria: insights from genomics. **Trends in Ecology & Evolution**, v. 15, n. 8, p. 321–326, ago. 2000.

MOREIRA, H. N. S. **Anotação e montagem de transcriptomas de intestino médio e ovários do carrapato *Amblyomma Sculptum* , antes e após a infecção por *Rickettsia Amblyommii***. [s.l: s.n.].

MORENO, C. X. et al. Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. **Environmental Microbiology**, v. 8, n. 5, p. 761–772, maio 2006.

MOUMÈNE, A. et al. Iron starvation conditions upregulate *Ehrlichia ruminantium* type IV secretion system, tr1 transcription factor and map1 genes family through the master regulatory protein ErxR. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. JAN, p. 1–11, 2018.

MOUNSEY, K. E. et al. Increased transcription of glutathione S-transferases in acaricide exposed scabies mites. **Parasites & Vectors**, v. 3, n. 1, p. 43, 2010.

MOZZETTA, C. et al. Sound of silence: the properties and functions of repressive Lys methyltransferases. **Nature Reviews Molecular Cell Biology**, v. 16, n. 8, p. 499–513, 23 ago. 2015.

MULENGA, A. et al. Dynamics of *Rickettsia*-tick interactions: identification and characterization of differentially expressed mRNAs in uninfected and infected *Dermacentor variabilis*. **Insect Molecular Biology**, v. 12, n. 2, p. 185–193, abr. 2003.

MUNDERLOH, U. G. Cellular and molecular interrelationships between ticks and prokaryotic tick-borne pathogens. **Annual Review of Entomology**, v. 40, n. 1, p. 221–243, 1995.

MUNTER, S.; WAY, M.; FRISCHKNECHT, F. Signaling during pathogen infection. **Science Signaling**, v. 2006, n. 335, p. re5–re5, 16 maio 2006.

NAKAI, K.; KANEHISA, M. Expert system for predicting protein localization sites in gram-negative bacteria. **Proteins: Structure, Function, and Genetics**, v. 110, p. 95–110, 1991.

NAKAI, K.; KANEHISA, M. Refinement of the prediction methods of signal peptides for the genome analyses of *Saccharomyces cerevisiae* and *Bacillus subtilis*. **Genome Informatics**, v. 7, p. 72–81, 1996.

NAKAYAMA, K. I.; NAKAYAMA, K. Ubiquitin ligases: cell-cycle control and cancer. **Nature Reviews Cancer**, v. 6, n. 5, p. 369–381, maio 2006.

NARASIMHAN, S. et al. Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the lyme disease spirochete. **Cell Host & Microbe**, v. 15, n. 1, p. 58–71, jan. 2014.

NAVA, S. et al. Reassessment of the taxonomic status of *Amblyomma cajennense* (Fabricius, 1787) with the description of three new species, *Amblyomma tonelliae* n. sp., *Amblyomma interandinum* n. sp. and *Amblyomma patinoi* n. sp., and reinstatement. **Ticks and Tick-borne Diseases**, v. 5, n. 3, p. 252–276, abr. 2014.

NAVARRO, L. et al. Identification of a molecular target for the *Yersinia* protein kinase A. **Molecular Cell**, v. 26, n. 4, p. 465–477, maio 2007.

NEHER, E. How frequent are correlated changes in families of protein sequences? **Proceedings of the National Academy of Sciences**, v. 91, n. 1, p. 98–102, 4 jan. 1994.

NIKOH, N. et al. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. **Proceedings of the National Academy of Sciences**, v. 111, n. 28, p. 10257–10262, 15 jul. 2014.

NONNENMACHER, Y.; HILLER, K. Biochemistry of proinflammatory macrophage activation. **Cellular and Molecular Life Sciences**, v. 75, n. 12, p. 2093–2109, 3 jun. 2018.

NOOREN, I. M. A. New embo member's review: Diversity of protein-protein interactions. **The EMBO Journal**, v. 22, n. 14, p. 3486–3492, 15 jul. 2003.

NOWAK, S. J.; CORCES, V. G. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. **Trends in Genetics**, v. 20,

n. 4, p. 214–220, abr. 2004.

NUNES, E. D. C. et al. *Rickettsia amblyommii* infecting *Amblyomma sculptum* in endemic Spotted fever area from Southeastern Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 8, p. 1058–1061, 11 dez. 2015.

NUÑEZ, P. A.; SORIA, M.; FARBER, M. D. The Twin-Arginine translocation pathway in α -Proteobacteria is functionally preserved irrespective of genomic and regulatory divergence. **PLoS ONE**, v. 7, n. 3, p. e33605, 15 mar. 2012.

OCHMAN, H.; LAWRENCE, J. G.; GROISMAN, E. A. Lateral gene transfer and the nature of bacterial innovation. **Nature**, v. 405, n. 6784, p. 299–304, maio 2000.

OGATA, H. *Rickettsia felis*, from culture to genome sequencing. **Annals of the New York Academy of Sciences**, v. 1063, n. 1, p. 26–34, 1 dez. 2005.

OGATA, H. et al. Genome sequence of *Rickettsia bellii* illuminates the role of Amoebae in gene exchanges between intracellular pathogens. **PLoS Genetics**, v. 2, n. 5, p. e76, 2006.

OGAWA, M. et al. Shotgun proteomics of *Orientia tsutsugamushi*. **Clinical Microbiology and Infection**, v. 15, n. SUPPL. 2, p. 239–240, dez. 2009.

OLKKONEN, V. M.; SLENMARK, H. Role of Rab GTPases in membrane traffic. In: **Genome Biology 2001 2:5**. v. 176p. 1–85.

OSTERMANN, J. et al. Stepwise assembly of functionally active transport vesicles. **Cell**, v. 75, n. 5, p. 1015–1025, dez. 1993.

OVERBEEK, R. et al. The use of gene clusters to infer functional coupling. **Proceedings of the National Academy of Sciences**, v. 96, n. 6, p. 2896–2901, 16 mar. 1999.

PALLEN, M. J.; FRANCIS, M. S.; FÄTTERER, K. Tetratricopeptide-like repeats in type-III-secretion chaperones and regulators. **FEMS Microbiology Letters**, v. 223, n. 1, p. 53–60, jun. 2003.

PALOMAR, A. M. et al. Role of birds in dispersal of etiologic agents of tick-borne zoonoses, Spain, 2009. **Emerging Infectious Diseases**, v. 18, n. 7, p. 1188–1191, jul. 2012.

PARK, D. et al. Comparative interactomics analysis of protein family interaction

networks using PSIMAP (protein structural interactome map). **Bioinformatics**, v. 21, n. 15, p. 3234–3240, 1 ago. 2005.

PARK, H. et al. The *Rickettsia* surface cell antigen 4 applies mimicry to bind to and activate vinculin. **Journal of Biological Chemistry**, v. 286, n. 40, p. 35096–35103, 7 out. 2011.

PARK, J. et al. *Anaplasma phagocytophilum* AnkaA binds to granulocyte DNA and nuclear proteins. **Cellular Microbiology**, v. 6, n. 8, p. 743–751, 4 maio 2004.

PAROLA, P. et al. *Rickettsia slovaca* and *Rickettsia raoultii* in tick - borne rickettsioses. **Emerging Infectious Diseases**, v. 15, n. 7, p. 1105–1108, jul. 2009.

PAROLA, P. et al. Update on tick-borne rickettsioses around the world: A geographic approach. **Clinical Microbiology Reviews**, v. 26, n. 4, p. 657–702, 1 out. 2013.

PAROLA, P.; PADDOCK, C. D.; RAOULT, D. Tick-borne rickettsioses around the world: Emerging diseases challenging old concepts. **Clinical Microbiology Reviews**, v. 18, n. 4, p. 719–756, 1 out. 2005a.

PAROLA, P.; PADDOCK, C. D.; RAOULT, D. Tick-borne rickettsioses around the world: Emerging diseases challenging old concepts. **Clinical Microbiology Reviews**, v. 18, n. 4, p. 719–756, 1 out. 2005b.

PAROLA, P.; PADDOCK, C. D.; RAOULT, D. Tick-Borne rickettsioses around the world: Emerging diseases challenging old concepts. **Clinical Microbiology Reviews**, v. 18, n. 4, p. 719–756, 1 out. 2005c.

PAZOS, F. et al. Correlated mutations contain information about protein-protein interaction. **Journal of Molecular Biology**, v. 271, n. 4, p. 511–523, ago. 1997.

PAZOS, F.; VALENCIA, A. Similarity of phylogenetic trees as indicator of protein-protein interaction. **Protein Engineering, Design and Selection**, v. 14, n. 9, p. 609–614, set. 2001.

PELLEGRINI, M. et al. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 8, p. 4285–4288, 1999.

PERLMAN, S. J.; HUNTER, M. S.; ZCHORI-FEIN, E. The emerging diversity of *Rickettsia*. **Proceedings of the Royal Society B: Biological Sciences**, v. 273, n. 1598, p. 2097–2106, 7 set. 2006.

PERRY, J. W.; WOBUS, C. E. Endocytosis of murine norovirus 1 into murine macrophages is dependent on dynamin II and cholesterol. **Journal of Virology**, v. 84, n. 12, p. 6163–6176, 15 jun. 2010.

PETCHAMPAI, N. et al. Molecular and functional characterization of vacuolar-ATPase from the American dog tick *Dermacentor variabilis*. **Insect Molecular Biology**, v. 23, n. 1, p. 42–51, fev. 2014a.

PETCHAMPAI, N. et al. Novel identification of *Dermacentor variabilis* Arp2/3 complex and its role in rickettsial infection of the arthropod vector. **PLoS ONE**, v. 9, n. 4, p. e93768, 14 abr. 2014b.

PETCHAMPAI, N. et al. Identification of host proteins involved in rickettsial invasion of tick cells. **Infection and Immunity**, v. 83, n. 3, p. 1048–1055, 2015.

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

PINTER, A.; LABRUNA, M. B. Isolation of *Rickettsia rickettsii* and *Rickettsia bellii* in cell culture from the tick *Amblyomma aureolatum* in Brazil. **Annals of the New York Academy of Sciences**, v. 1078, n. 1, p. 523–529, 1 out. 2006.

PITOSSI, F. et al. A functional GTP-binding motif is necessary for antiviral activity of Mx proteins. **Journal of virology**, v. 67, n. 11, p. 6726–32, nov. 1993.

POCKLEY, A. G. Heat shock proteins as regulators of the immune response. **The Lancet**, v. 362, n. 9382, p. 469–476, ago. 2003.

POHL, P. C. et al. In vitro establishment of ivermectin-resistant *Rhipicephalus microplus* cell line and the contribution of ABC transporters on the resistance mechanism. **Veterinary Parasitology**, v. 204, n. 3–4, p. 316–322, ago. 2014.

POINAR, G. Evolutionary history of terrestrial pathogens and endoparasites as revealed in fossils and subfossils. **Advances in Biology**, v. 2014, p. 1–29, 2014.

POINAR, G.; BROWN, A. E. A new genus of hard ticks in Cretaceous Burmese amber (Acari: Ixodida: Ixodidae). **Systematic Parasitology**, v. 54, n. 3, p. 199–205, mar. 2003.

POINAR, G. O.; BUCKLEY, R. *Compluriscutula vetulum* (Acari: Ixodida: Ixodidae), a new genus and species of hard tick from lower Cretaceous Burmese amber. **Proceedings of the Entomological Society of Washington**, v. 110, n. 2, p. 445–450, abr. 2008.

POLICASTRO, P. F. et al. *Rickettsia rickettsii* growth and temperature-inducible protein expression in embryonic tick cell lines. **Journal of Medical Microbiology**, v. 46, n. 10, p. 839–845, 1 out. 1997.

PORTNOY, D. A. et al. Molecular determinants of *Listeria monocytogenes* pathogenesis. **Infection and Immunity**, v. 60, n. 4, p. 1263–1267, 1992.

PRICE, J. V.; VANCE, R. E. The macrophage paradox. **Immunity**, v. 41, n. 5, p. 685–693, nov. 2014.

RADULOVIC, S. et al. Identification and molecular analysis of the gene encoding *Rickettsia typhi* hemolysin. **Infection and Immunity**, v. 67, n. 11, p. 6104–6108, 1999.

RAHN, E. et al. Entry pathways of herpes simplex virus type 1 into human keratinocytes are dynamin- and cholesterol-dependent. **PLoS ONE**, v. 6, n. 10, 2011.

RAMAN, K.; YETURU, K.; CHANDRA, N. targetTB: A target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. **BMC Systems Biology**, v. 2, n. 1, p. 109, 2008.

RANDOLPH, S. E. Tick ecology: processes and patterns behind the epidemiological risk posed by ixodid ticks as vectors. **Parasitology**, v. 129, n. S1, p. S37–S65, 2004a.

RANDOLPH, S. E. Tick ecology: processes and patterns behind the epidemiological risk posed by ixodid ticks as vectors. **Parasitology**, v. 129, n. S1, p. S37–S65, 19 out. 2004b.

RAOULT, D.; ROUX, V. Rickettsioses as paradigms of new or emerging infectious diseases. **Clinical Microbiology Reviews**, v. 10, n. 4, p. 694–719, out. 1997.

RECHSTEINER, M.; HILL, C. P. Mobilizing the proteolytic machine: cell biological roles of proteasome activators and inhibitors. **Trends in Cell Biology**, v. 15, n. 1, p. 27–33, jan.

2005.

REED, S. C. O.; SERIO, A. W.; WELCH, M. D. *Rickettsia parkeri* invasion of diverse host cells involves an Arp2/3 complex, WAVE complex and Rho-family GTPase-dependent pathway. **Cellular Microbiology**, v. 14, n. 4, p. 529–545, abr. 2012.

REINHECKEL, T. et al. Comparative resistance of the 20S and 26S proteasome to oxidative stress. **Biochemical Journal**, v. 335, n. 3, p. 637–642, 1 nov. 1998.

RENESTO, P. et al. Identification and characterization of a phospholipase D–superfamily gene in Rickettsiae. **The Journal of Infectious Diseases**, v. 188, n. 9, p. 1276–1283, nov. 2003.

RENESTO, P. et al. *Rickettsia conorii* transcriptional response within inoculation eschar. **PLoS ONE**, v. 3, n. 11, p. e3681, 10 nov. 2008.

RENNOLL-BANKERT, K. E. et al. Chromatin-bound bacterial effector ankyrin A recruits histone deacetylase 1 and modifies host gene expression. **Cellular Microbiology**, v. 17, n. 11, p. 1640–1652, nov. 2015.

RIBEIRO, J. M. C. et al. An annotated catalog of salivary gland transcripts from Ixodes scapularis ticks. **Insect Biochemistry and Molecular Biology**, v. 36, n. 2, p. 111–129, fev. 2006.

RICHARDSON, D. H. S. The symbiotic habit. **Symbiosis**, v. 51, n. 2, p. 197–198, 3 set. 2010.

RIKIHISA, Y.; ITO, S. Intracellular localization of *Rickettsia tsutsugamushi* in polymorphonuclear leukocytes. **Journal of Experimental Medicine**, v. 150, n. 3, p. 703–708, 1979.

RIKIHISA, Y.; LIN, M.; NIU, H. Microreview: Type IV secretion in the obligatory intracellular bacterium *Anaplasma phagocytophilum*. **Cellular Microbiology**, v. 12, n. 9, p. 1213–1221, set. 2010.

RILEY, S. P. et al. The *Rickettsia conorii* autotransporter protein sca1 promotes adherence to nonphagocytic mammalian cells. **Infection and Immunity**, v. 78, n. 5, p. 1895–1904, 2010.

RIVES, A. W.; GALITSKI, T. Modular organization of cellular networks. **Proceedings of the National Academy of Sciences**, v. 100, n. 3, p. 1128–1133, 4 fev. 2003.

ROCHE, J. P. et al. Regulation of synaptic plasticity and synaptic vesicle dynamics by the PDZ protein scribble. **The Journal of Neuroscience**, v. 22, n. 15, p. 6471–6479, 1 ago. 2002.

ROTHBERG, J. M.; LEAMON, J. H. The development and impact of 454 sequencing. **Nature Biotechnology**, v. 26, n. 10, p. 1117–1124, 9 out. 2008.

RYDKINA, E.; TURPIN, L. C.; SAHNI, S. K. Activation of p38 mitogen-activated protein kinase module facilitates in vitro host cell invasion by *Rickettsia rickettsii*. **Journal of Medical Microbiology**, v. 57, n. 9, p. 1172–1175, 1 set. 2008.

SAHNI, S. K. et al. Recent molecular insights into rickettsial pathogenesis and immunity. **Future Microbiology**, v. 8, n. 10, p. 1265–1288, out. 2013.

SAITO, R. et al. A travel guide to cytoscape plugins. **Nature Methods**, v. 9, n. 11, p. 1069–1076, 6 nov. 2012.

SALWINSKI, L. The database of interacting proteins: 2004 update. **Nucleic Acids Research**, v. 32, n. 90001, p. 449D – 451, 1 jan. 2004.

SALZ, H. K. et al. The *Drosophila* maternal effect locus deadhead encodes a thioredoxin homolog required for female meiosis and early embryonic development. **Genetics**, v. 136, n. 3, p. 1075–1086, 1994.

SAYLER, K. A. et al. Cultivation of *Rickettsia amblyommii* in tick cells, prevalence in Florida lone star ticks (*Amblyomma americanum*). **Parasites & Vectors**, v. 7, n. 1, p. 270, 2014.

SCHELHAAS, M. Come in and take your coat off - how host cells provide endocytosis for virus entry. **Cellular Microbiology**, v. 12, n. 10, p. 1378–1388, out. 2010.

SCHMITZ-ESSER, S. et al. ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. **Journal of Bacteriology**, v. 186, n. 3, p. 683–691, 1 fev. 2004.

SCHNETTLER, E. et al. Induction and suppression of tick cell antiviral RNAi responses

by tick-borne flaviviruses. **Nucleic Acids Research**, v. 42, n. 14, p. 9436–9446, 18 ago. 2014.

SCHWARZ, A. et al. De novo *Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. **The FASEB Journal**, v. 27, n. 12, p. 4745–4756, dez. 2013.

SCHWARZ, A. et al. A Systems Level Analysis Reveals Transcriptomic and Proteomic Complexity in *Ixodes Ricinus* Midgut and Salivary Glands During Early Attachment and Feeding. **Molecular & Cellular Proteomics**, v. 13, n. 10, p. 2725–2735, out. 2014.

SEARS, K. T. et al. Surface proteome analysis and characterization of surface cell antigen (Sca) or autotransporter family of *Rickettsia typhi*. **PLoS Pathogens**, v. 8, n. 8, p. e1002856, 9 ago. 2012.

SENTAUSA, E. et al. Genome sequence of “*Rickettsia sibirica* subsp. *mongolitimonae*”. **Journal of Bacteriology**, v. 194, n. 9, p. 2389–2390, 1 maio 2012a.

SENTAUSA, E. et al. Genome sequence of *Rickettsia conorii* subsp. *caspia*, the agent of Astrakhan Fever. **Journal of Bacteriology**, v. 194, n. 17, p. 4763–4764, 1 set. 2012b.

SEVERO, M. S. et al. The E3 Ubiquitin Ligase XIAP restricts *Anaplasma phagocytophilum* colonization of *Ixodes scapularis* ticks. **The Journal of Infectious Diseases**, v. 208, n. 11, p. 1830–1840, 1 dez. 2013.

SHARAN, R.; ULITSKY, I.; SHAMIR, R. Network-based prediction of protein function. **Molecular Systems Biology**, v. 3, n. 1, p. 88, 13 jan. 2007.

SHAW, D. K. et al. Infection-derived lipids elicit an immune deficiency circuit in arthropods. **Nature Communications**, v. 8, n. 1, p. 14401, 14 abr. 2017.

SHAW, D. K. et al. Vector immunity and evolutionary ecology: The harmonious dissonance. **Trends in Immunology**, v. 39, n. 11, p. 862–873, nov. 2018.

SHRINGARPURE, R.; DAVIES, K. J. A. Protein turnover by the proteasome in aging and disease. **Free Radical Biology and Medicine**, v. 32, n. 11, p. 1084–1089, jun. 2002.

SIDERI, M. et al. Innate immunity in insects: surface-associated dopa decarboxylase-dependent pathways regulate phagocytosis, nodulation and melanization in medfly haemocytes.

Immunology, v. 123, n. 4, p. 528–537, abr. 2008.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 1, p. 539, 11 jan. 2011.

SILVERMAN, D. J. et al. Penetration of host cells by *Rickettsia rickettsii* appears to be mediated by a phospholipase of rickettsial origin. **Infection and Immunity**, v. 60, n. 7, p. 2733–2740, 1992.

SIMSER, J. A. et al. *Rickettsia monacensis* sp. nov., a spotted fever group *Rickettsia*, from ticks (*Ixodes ricinus*) collected in a european City Park. **Applied and Environmental Microbiology**, v. 68, n. 9, p. 4559–4566, 1 set. 2002.

SIMSKE, J. S. et al. LET-23 receptor localization by the cell junction protein LIN-7 during *Caenorhabditis elegans* vulval induction. **Cell**, v. 85, n. 2, p. 195–204, abr. 1996.

SINCLAIR, S. H. G.; YEGNASUBRAMANIAN, S.; DUMLER, J. S. Global DNA methylation changes and differential gene expression in *Anaplasma phagocytophilum*-infected human neutrophils. **Clinical Epigenetics**, v. 7, n. 1, p. 77, 29 dez. 2015.

SINCLAIR, S. H.; RENNOLL-BANKERT, K. E.; DUMLER, J. S. Effector bottleneck: microbial reprogramming of parasitized host cell transcription by epigenetic remodeling of chromatin structure. **Frontiers in Genetics**, v. 5, n. AUG, p. 1–10, 14 ago. 2014.

SMITH, A. A.; PAL, U. Immunity-related genes in *Ixodes scapularis* perspectives from genome information. **Frontiers in Cellular and Infection Microbiology**, v. 4, n. AUG, p. 1–12, 22 ago. 2014.

SNEL, B.; BORK, P.; HUYNEN, M. Genome evolution. **Trends in Genetics**, v. 16, n. 1, p. 9–11, jan. 2000.

SNIPEN, L.; USSERY, D. W. Standard operating procedure for computing pangenome trees. **Standards in Genomic Sciences**, v. 2, n. 1, p. 135–141, 28 jan. 2010.

SOCOLOVSKI, C. et al. Experimental infection of *Rhipicephalus sanguineus* with *Rickettsia conorii conorii*. **Clinical Microbiology and Infection**, v. 15, p. 324–325, dez. 2009a.

SOCOLOVSKI, C. et al. Transovarial and trans-stadial transmission of *Rickettsia africae* in *Amblyomma variegatum* ticks. **Clinical Microbiology and Infection**, v. 15, n. SUPPL. 2, p. 317–318, dez. 2009b.

SOCOLOVSKI, C. et al. Transmission of *Rickettsia conorii conorii* in naturally infected *Rhipicephalus sanguineus*. **Clinical Microbiology and Infection**, v. 15, n. SUPPL. 2, p. 319–321, dez. 2009c.

SODING, J. Protein homology detection by HMM-HMM comparison. **Bioinformatics**, v. 21, n. 7, p. 951–960, 1 abr. 2005.

SOJKA, D. et al. New insights into the machinery of blood digestion by ticks. **Trends in Parasitology**, v. 29, n. 6, p. 276–285, jun. 2013.

SONENSHINE, D. E. et al. First Transcriptome of the Testis-Vas Deferens-Male Accessory Gland and Proteome of the Spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). **PLoS ONE**, v. 6, n. 9, p. e24711, 16 set. 2011.

SONNHAMMER, E. Pfam: multiple sequence alignments and HMM-profiles of protein domains. **Nucleic Acids Research**, v. 26, n. 1, p. 320–322, 1 jan. 1998.

SONNHAMMER, E. L. L.; KROGH, A. A hidden Markov model for predicting transmembrane helices in protein sequences. 1998.

STARK, C. BioGRID: a general repository for interaction datasets. **Nucleic Acids Research**, v. 34, n. 9, p. D535–D539, 1 jan. 2006.

STEIMER, L.; KLOSTERMEIER, D. RNA helicases in infection and disease. **RNA Biology**, v. 9, n. 6, p. 751–771, 27 jun. 2012.

STROMDAHL, E. Y. et al. Infrequency of *Rickettsia rickettsii* in *Dermacentor variabilis* removed from humans, with comments on the role of other human-biting ticks associated with spotted fever group Rickettsiae in the United States. **Vector-Borne and Zoonotic Diseases**, v. 11, n. 7, p. 969–977, jul. 2011.

SULTANA, H. et al. *Anaplasma phagocytophilum* induces actin phosphorylation to selectively regulate gene transcription in *Ixodes scapularis* ticks. **The Journal of Experimental Medicine**, v. 207, n. 8, p. 1727–1743, 2 ago. 2010.

SUMBAYEV, V. V.; YASINSKA, I. M. Role of MAP kinase-dependent apoptotic pathway in innate immune responses and viral infection. **Scandinavian journal of immunology**, v. 63, n. 6, p. 391–400, 2006.

SZKLARCZYK, D. et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. **Nucleic Acids Research**, v. 39, n. Database, p. D561–D568, 1 jan. 2011.

TAANK, V. et al. Human rickettsial pathogen modulates arthropod organic anion transporting polypeptide and tryptophan pathway for its survival in ticks. **Scientific Reports**, v. 7, n. 1, p. 13256, 16 dez. 2017.

TATUSOV, R. L. The COG database: new developments in phylogenetic classification of proteins from complete genomes. **Nucleic Acids Research**, v. 29, n. 1, p. 22–28, 1 jan. 2001.

TEICHMANN, S. A. et al. Fast assignment of protein structures to sequences using the intermediate sequence library PDB-ISL. **Bioinformatics**, v. 16, n. 2, p. 117–124, 1 fev. 2000.

TELFER, S. et al. Species interactions in a parasite community drive infection risk in a wildlife population. **Science**, v. 330, n. 6001, p. 243–246, 8 out. 2010.

TETTELIN, H. et al. Comparative genomics: the bacterial pan-genome. **Current Opinion in Microbiology**, v. 11, n. 5, p. 472–477, out. 2008.

THOMAS, V.; FIKRIG, E. *Anaplasma phagocytophilum* specifically induces tyrosine phosphorylation of ROCK1 during infection. **Cellular Microbiology**, v. 9, n. 7, p. 1730–1737, jul. 2007.

THOMPSON, J. N. **The coevolutionary process**. [s.l.] University of Chicago Press, 1994.

THOMPSON, J. N. The evolution of species interactions. **Science**, v. 284, n. 5423, p. 2116–2118, 25 jun. 1999.

TIMMONS, L.; FIRE, A. Specific interference by ingested dsRNA. **Nature**, v. 395, n. October, p. 1998, 1998.

TITZ, B.; SCHLESNER, M.; UETZ, P. What do we learn from high-throughput protein

interaction data? **Expert Review of Proteomics**, v. 1, n. 1, p. 111–121, 2004.

TRESS, M. et al. Scoring docking models with evolutionary information. **Proteins: Structure, Function, and Bioinformatics**, v. 60, n. 2, p. 275–280, 24 jun. 2005.

TRETINA, K. et al. *Theileria*-transformed bovine leukocytes have cancer hallmarks. **Trends in Parasitology**, v. 31, n. 7, p. 306–314, jul. 2015.

TROUT FRYXELL, R. T. et al. Molecular Detection of Rickettsia Species Within Ticks (Acari: Ixodidae) Collected from Arkansas United States. **Journal of Medical Entomology**, v. 52, n. 3, p. 500–508, 1 maio 2015a.

TROUT FRYXELL, R. T. et al. Molecular detection of *Rickettsia* species within ticks (Acari: Ixodidae) collected from Arkansas United States. **Journal of Medical Entomology**, v. 52, n. 3, p. 500–508, 1 maio 2015b.

TROXELL, B.; XU, H.; YANG, X. F. *Borrelia burgdorferi*, a pathogen that lacks iron, encodes manganese-dependent superoxide dismutase essential for resistance to streptonigrin. **Journal of Biological Chemistry**, v. 287, n. 23, p. 19284–19293, 1 jun. 2012.

TRUCHAN, H. K.; SEIDMAN, D.; CARLYON, J. A. Breaking in and grabbing a meal: *Anaplasma phagocytophilum* cellular invasion, nutrient acquisition, and promising tools for their study. **Microbes and Infection**, v. 15, n. 14–15, p. 1017–1025, dez. 2013.

VARELA-STOKES, A. S. et al. *Rickettsia parkeri* in *Amblyomma maculatum* ticks, North Carolina, USA, 2009-2010. **Emerging Infectious Diseases**, v. 17, n. 12, p. 2350–2353, dez. 2011.

VERHELST, J.; HULPIAU, P.; SAELENS, X. Mx proteins: Antiviral gatekeepers that restrain the uninvited. **Microbiology and Molecular Biology Reviews**, v. 77, n. 4, p. 551–566, 1 dez. 2013.

VIJAY-KUMAR, M. et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-Like receptor 5. **Science**, v. 328, n. 5975, p. 228–231, 9 abr. 2010.

VILLAR, M. et al. Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by *Anaplasma phagocytophilum* infection in tick cells. **Molecular & Cellular Proteomics**, v. 14, n. 12, p. 3154–3172, 2015.

VILLAR, M. et al. The intracellular bacterium *Anaplasma phagocytophilum* selectively manipulates the levels of vertebrate host proteins in the tick vector *Ixodes scapularis*. **Parasites & Vectors**, v. 9, n. 1, p. 467, 25 dez. 2016.

VLAMIS-GARDIKAS, A.; HOLMGREN, A. Thioredoxin and glutaredoxin isoforms. In: **Methods in Enzymology**. [s.l: s.n.]. v. 347p. 286–296.

WALKER, D. H. Rocky Mountain spotted fever: a disease in need of microbiological concern. **Clinical Microbiology Reviews**, v. 2, n. 3, p. 227–240, jul. 1989.

WALKER, T. S. Rickettsial interactions with human endothelial cells in vitro: Adherence and entry. **Infection and Immunity**, v. 44, n. 2, p. 205–210, 1984.

WALKER, T. S.; WINKLER, H. H. Penetration of cultured mouse fibroblasts (L cells) by *Rickettsia prowazeki*. **Infection and Immunity**, v. 22, n. 1, p. 200–208, 1978.

WASS, M. N.; DAVID, A.; STERNBERG, M. J. Challenges for the prediction of macromolecular interactions. **Current Opinion in Structural Biology**, v. 21, n. 3, p. 382–390, jun. 2011.

WELCH, M. D.; REED, S. C. O.; HAGLUND, C. M. Establishing intracellular infection: escape from the phagosome and intracellular colonization (Rickettsiaceae). In: **Intracellular Pathogens II: Rickettsiales**. [s.l.] American Society of Microbiology, 2012. p. 154–174.

WELCH, R. A. et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. **Proceedings of the National Academy of Sciences**, v. 99, n. 26, p. 17020–17024, 24 dez. 2002.

WHITWORTH, T. et al. Expression of the *Rickettsia prowazekii* pld or tlyC Gene in *Salmonella enterica* serovar Typhimurium mediates phagosomal escape. **Infection and Immunity**, v. 73, n. 10, p. 6668–6673, 1 out. 2005.

WIECZOREK, H. et al. Vacuolar-type proton pumps in insect epithelia. **Journal of Experimental Biology**, v. 212, n. 11, p. 1611–1619, 1 jun. 2009.

WIKSWO, M. E. et al. Detection and identification of spotted fever group Rickettsiae in *Dermacentor* species from Southern California. **Journal of Medical Entomology**, v. 45, n.

3, p. 509–516, 2008.

WOOD, A.; SHILATIFARD, A. Posttranslational modifications of histones by methylation. In: **Advances in Protein Chemistry**. [s.l.: s.n.]. v. 67p. 201–222.

WRIGHT, C. L. et al. Experimental vertical transmission of *Rickettsia parkeri* in the Gulf Coast tick, *Amblyomma maculatum*. **Ticks and Tick-borne Diseases**, v. 6, n. 5, p. 568–573, jul. 2015a.

WRIGHT, C. L. et al. *Rickettsia parkeri* transmission to *Amblyomma americanum* by cofeeding with *Amblyomma maculatum* (Acari: Ixodidae) and potential for spillover. **Journal of Medical Entomology**, v. 52, n. 5, p. 1090–1095, set. 2015b.

XENARIOS, I. DIP: The database of interacting proteins: 2001 update. **Nucleic Acids Research**, v. 29, n. 1, p. 239–241, 1 jan. 2001.

YAN, Q. et al. Ehrlichia type IV secretion system effector Etf-2 binds to active RAB5 and delays endosome maturation. **Proceedings of the National Academy of Sciences**, v. 115, n. 38, p. E8977–E8986, 18 set. 2018.

YANAI, I.; DERTI, A.; DELISI, C. Genes linked by fusion events are generally of the same functional category: A systematic analysis of 30 microbial genomes. **Proceedings of the National Academy of Sciences**, v. 98, n. 14, p. 7940–7945, 3 jul. 2001.

YANG, D.; LU, H.; ERICKSON, J. W. Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in *Drosophila* embryos. **Current Biology**, v. 10, n. 19, p. 1191–1200, out. 2000.

YANG, X. et al. A dityrosine network mediated by dual oxidase and peroxidase influences the persistence of Lyme disease pathogens within the vector. **Journal of Biological Chemistry**, v. 289, n. 18, p. 12813–12822, 2 maio 2014.

YEANG, C.-H.; HAUSSLER, D. Detecting coevolution in and among protein domains. **PLoS Computational Biology**, v. 3, n. 11, p. e211, 2007.

YIN, H. et al. Antigenic diversity in maxadilan, a salivary protein from the sand fly vector of american visceral leishmaniasis. **The American Journal of Tropical Medicine and Hygiene**, v. 70, n. 3, p. 286–293, 1 mar. 2004.

YOON, K. S. et al. Brief exposures of human body lice to sublethal amounts of ivermectin over-transcribes detoxification genes involved in tolerance. **Insect Molecular Biology**, v. 20, n. 6, p. 687–699, dez. 2011.

YSSOUF, A. et al. First molecular detection of *Rickettsia africae* in ticks from the Union of the Comoros. **Parasites & Vectors**, v. 7, n. 1, p. 444, 2014.

ZANZONI, A. et al. MINT: A molecular interaction database. **FEBS Letters**, v. 513, n. 1, p. 135–140, 20 fev. 2002.

ZEHNBauer, B. A. et al. Identification and purification of the Lon+ (capR+) gene product, a DNA-binding protein. **Proceedings of the National Academy of Sciences**, v. 78, n. 4, p. 2043–2047, 1 abr. 1981.

ZEMTSOVA, G. et al. Co-feeding as a route for transmission of *Rickettsia conorii israelensis* between *Rhipicephalus sanguineus* ticks. **Experimental and Applied Acarology**, v. 52, n. 4, p. 383–392, 30 dez. 2010.

ZGURSKAYA, H. I. et al. Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria. **Frontiers in Microbiology**, v. 2, n. SEP, p. 1–13, 2011.

ZHANG, Y. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. **Genes & Development**, v. 15, n. 18, p. 2343–2360, 15 set. 2001.

ZHANG, Z. et al. DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. **Immunity**, v. 34, n. 6, p. 866–878, jun. 2011.

ZHAO, J. et al. mTOR inhibition activates overall protein degradation by the ubiquitin proteasome system as well as by autophagy. **Proceedings of the National Academy of Sciences**, v. 112, n. 52, p. 15790–15797, 29 dez. 2015.

ZHU, Y.-Z. et al. Japanese encephalitis virus enters rat neuroblastoma cells via a pH-dependent, dynamin and caveola-mediated endocytosis pathway. **Journal of Virology**, v. 86, n. 24, p. 13407–13422, 15 dez. 2012.

DADOS SUPLEMENTARES

- As análises de similaridade contra as proteínas do banco de dados do Peimap, Psimap, Pfam e String para as proteínas de *R. amblyommatis* (Dados suplementares S1a-S4a) como para o *I. scapularis* (Dados suplementares S1b-S4b);
- Resultados preliminares obtidos para os pares de proteínas que interagem encontrados nos bancos de dados do Peimap, Psimap, Pfam e String (Dados suplementares S1c-S4c);
- Resultados preliminares obtidos para os pares de proteínas destes organismos que interagem, representadas pelos IDs de identificação e com os valores de *score* final e normalizado, por meio das análises de similaridade contra o banco de dados do Peimap, Psimap, Pfam e String (Dados suplementares S1d-S4d);
- Resultados obtidos para os pares de proteínas que interagem, sendo representadas pelos IDs das proteínas e com os valores de *score* combinado para cada par de interação e a distribuição binomial da frequência de sua ocorrência em cada banco de dados, tratando-se do interatoma interespecífico geral (Dados suplementares S1e);
- Resultados preliminares obtidos para as proteínas de membrana preditas com o uso do aplicativo Signalp 4.1, Tmhmm 2.0c, LipoP 1.0a, Psortb 3.0.4 e Phobius 1.01 (Dados suplementares S1f-S5f);
- Resultados obtidos dos pares de proteínas de membrana da bactéria que interagem com as proteínas do carrapato, sendo representadas pelos IDs de identificação e o nome da proteína para cada par de interação (Dados suplementares S1g);
- Resultados para os pares de proteínas de membrana que interagem no interatoma e possui ortólogos para a *R. amblyommatis* e outras espécies de *Rickettsia* que infectam os carrapatos, sendo representadas pelos IDs de identificação e seus grupos de ortólogos preditos com o uso do aplicativo OrthoMCL 2.0 (Dados suplementares S1h);
- Resultados preliminares obtidos para averiguar as anotações de proteína e anotações de ontologia gênica no Uniprot, descrevendo as funções moleculares e os processos biológicos de proteínas destas interfaces de interação, sendo o arquivo Excel contendo as entradas, suas correspondências e as anotações de ontologia gênica GO para as correspondências (Dados suplementares S1i); o arquivo Align.html contendo os alinhamentos (Dados suplementares S2i); e o arquivo Sliminput.txt contendo um resumo dos resultados para as anotações GO (Dados suplementares S3i);
- Resultados de ontologia gênica obtidos por meio do uso e anotação da rede consenso de interação para as proteínas de membrana da espécie *R. amblyommatis* que interagem com as sequências de proteínas do carrapato *I. scapularis*, obtidos com o uso do pacote BINGO 3.0.3. Estes resultados sugerem a presença de diferentes módulos funcionais significativos com destaque para os módulos 0, 4, 13, 14 e 15 obtidos por meio do interatoma de proteínas de membrana (Dados suplementares S1j);
- Resultados preliminares da análise de qualidade inicial para as sequências gênicas de alta qualidade e confiabilidade do sequenciamento com o uso do aplicativo FastQC 0.11.5 (Dados suplementares S1k-S5k);
- Resultados preliminares da análise de trimagem com a retirada de possíveis adaptadores das *reads* fazendo o uso dos aplicativos Trimmomatic 0.36 e novamente do aplicativo FastQC 0.11.5 para averiguarmos a qualidade do sequenciamento com os dados trimados (Dados suplementares S11-S51);
- Resultados da análise da montagem “*de novo*” utilizando os arquivos das *reads* trimadas (Dados não publicados), obtidas por meio dos aplicativos Trimmomatic 0.36, avaliadas quanto a qualidade com o uso do aplicativo FastQC 0.11.5 (Dados suplementares S1m-S5m) para a obtenção das sequências de RNA dos transcritos (Dados não publicados), com o uso do aplicativo Trinity 2.5.1;
- Resultados da análise de similaridade com o uso do TBLASTn para as sequências de proteínas do carrapato presentes nos módulos funcionais do interatoma de proteínas de membrana contra as sequências de transcritos de RNA dos

transcriptomas montados para carrapatos *Amblyomma* spp. infectados e não infectados por *Rickettsia* spp. em diferentes órgãos (Dados suplementares S1n);

- Resultados dos IDs de identificação das sequências de RNA dos transcritos nos transcriptomas de carrapatos *Amblyomma* spp. infectados ou não por *Rickettsia*, com destaque para os valores de *fold change* e a média calculada para a definirmos, se a expressão destes transcritos é regulada de forma positiva ou negativa, podendo ainda, não haver níveis satisfatórios para esta mensuração (Dados suplementares S1o);
- Resultados da análise de similaridade com o uso do BLASTp para os pares de proteínas de membrana que interagem no interatoma para a *R. amblyommatis* contra as sequências de proteínas provenientes de genomas de outras bactérias e espécies de *Rickettsia*, sendo representadas pelos ID das proteínas, o ID da proteína similar e os valores de confiança e identidade (Dados suplementares S1p).

ANEXOS

Anexo 1 – Resultados dos alinhamentos das sequências gênicas das proteínas ortólogas presentes nos clusters 1816 e 1050 com o uso do CLUSTAL. Destaque (a) para as sequências gênicas obtidas para as bactérias do gênero *Rickettsia*, que infectam os carrapatos versus as que infectam os insetos, obtidas anteriormente com o uso do TBLASTn. Destaque (b) para as sequências de proteínas obtidas para as bactérias do gênero *Rickettsia*, que infectam os carrapatos versus as que infectam os insetos, obtidas anteriormente com o uso do BLASTp.

Cluster.1816.TBLASTn

```

1  CLUSTAL O(1.2.4) multiple sequence alignment
2
3
4  amb_vs_pro      ----- 0 36  amb_vs_pro      AARATGATTATTCAAGAGCTAGCAACTATAAAGAATAT----- 159
5  amb_vs_fel      ----- 0 37  amb_vs_fel      GACCCATTTTTAAGCTCAGCTTATAGTAATAAAGCTTTTGCCTTAAAGCAACTTTGTAGA 105
6  amb_vs_amb      ATGCAAAACATCGAAAAATTAGCATTTCATATTTATCTTATCATTAAATATTTTCAGTAGT 60 38  amb_vs_amb      AAACCTGACTACGAAGATGCTTATAACTATAAAGGAATGGTCTTAGCTAAGTTAGGAAAA 297
7  amb_vs_buc      ATGCAAAACATCGAAAAATTAGCATTTCATATTTATCTTATCATTAAATATTTTCAGTAGT 60 39  amb_vs_buc      AAACCTGACTACGAAGATGCTTATAACTATAAAGGAATGGTCTTAGCTAAGTTAGGAAAA 297
8  amb_vs_end      ATGCAAAACATCGAAAAATTAGCATTTCATATTTATCTTATCATTAAATATTTTCAGTAGT 60 40  amb_vs_end      AAACCTGACTACGAAGATGCTTATAACTATAAAGGAATGGTCTTAGCTAAGTTAGGAAAA 297
9  amb_vs_mon      ----- 0 41  amb_vs_mon      ----- 0
10
11
12  amb_vs_pro      ----- 0 44  amb_vs_pro      ---AATGTAGATATAGAAAACTTTATGATTCAAAAGAAATCAATCAATCATTGAAAGT 216
13  amb_vs_fel      ----- 0 45  amb_vs_fel      ATTGAAGAAGCTTTAGAAGTTTTAGATACAGCTATAAAAAATCAATCCCAAATATCAGGT 165
14  amb_vs_amb      GCTTTTGCAGAAAACCTTGATCAATCCTACTAATGCAAAAACAACAAATCGAGCAACAA 120 46  amb_vs_amb      TATCAAGAAGCAATACAAATATTATAATTTAGCAATCAAAATATGGTCTAGTTTACAGAA 357
15  amb_vs_buc      GCTTTTGCAGAAAACCTTGATCAATCCTACTAATGCAAAAACAACAAATCGAGCAACAA 120 47  amb_vs_buc      TATCAAGAAGCAATACAAATATTATAATTTAGCAATCAAAATATGGTCTAGTTTACAGAA 357
16  amb_vs_end      GCTTTTGCAGAAAACCTTGATCAATCCTACTAATGCAAAAACAACAAATCGAGCAACAA 120 48  amb_vs_end      TATCAAGAAGCAATACAAATATTATAATTTAGCAATCAAAATATGGTCTAGTTTACAGAA 357
17  amb_vs_mon      ----- 0 49  amb_vs_mon      ----- 0
18
19
20  amb_vs_pro      AATAAATCTAAAAATTTAAATAAAATAATGGAAAAATATCAAAATAATAATTTTCAAAC 60 52  amb_vs_pro      AGCATTGGCGCTAAAAATAGTTTCTAA-----GCTTATAATAAGCAAAAATAATA 265
21  amb_vs_fel      ----- 0 53  amb_vs_fel      TATTATGATAATAGAGSTATASTTTTAAAGTATTTTAGTCTCTATGAGGAAGCAATAAG 225
22  amb_vs_amb      AGCAAAATTCAGGATTCAAATATTTTAGCTGAAGAATATTTAAATATTT---GGTAAATCT 177 54  amb_vs_amb      GCCTATAATAATAAAGGTAATTCCTTTAAATAAATTAGGTAGATATCAAGAAGCAATAAAA 417
23  amb_vs_buc      AGCAAAATTCAGGATTCAAATATTTTAGCTGAAGAATATTTAAATATTT---GGTAAATCT 177 55  amb_vs_buc      GCCTATAATAATAAAGGTAATTCCTTTAAATAAATTAGGTAGATATCAAGAAGCAATAAAA 417
24  amb_vs_end      AGCAAAATTCAGGATTCAAATATTTTAGCTGAAGAATATTTAAATATTT---GGTAAATCT 177 56  amb_vs_end      GCCTATAATAATAAAGGTAATTCCTTTAAATAAATTAGGTAGATATCAAGAAGCAATAAAA 417
25  amb_vs_mon      ----- 0 57  amb_vs_mon      ----- 0
26
27
28  amb_vs_pro      TTATTGAATTATAAATATCGTACAAAATACAATTTACATTTTATTCCTAGTATAGGTTTT 120 60  amb_vs_pro      TAATA---TTAACACTAAGTGCTCATGGTAATATGAACGTCACCTCAATAATAAAAAT 321
29  amb_vs_fel      -----AAGTATGAAAACGGCGATAAAGTATTTTGATAAAGTTTTAAATCAA 45 61  amb_vs_fel      AGTTATGATCAAGCCCTTAAAGCTTGACTCGCTGGATGTTCTAACTTTTATTAATAAAGGA 285
30  amb_vs_amb      TTTTATAAATTAGGAAAATATCAAGAAGCAATAAAAAATTTTGATTTAGCTATTAATAT 237 62  amb_vs_amb      AATTTTGATCTTGCACTAAAATAATAATTTGCATTTTGCGGTAGCATATAATAATAAAGGT 477
31  amb_vs_buc      TTTTATAAATTAGGAAAATATCAAGAAGCAATAAAAAATTTTGATTTAGCTATTAATAT 237 63  amb_vs_buc      AATTTTGATCTTGCACTAAAATAATAATTTGCATTTTGCGGTAGCATATAATAATAAAGGT 477
32  amb_vs_end      TTTTATAAATTAGGAAAATATCAAGAAGCAATAAAAAATTTTGATTTAGCTATTAATAT 237 64  amb_vs_end      AATTTTGATCTTGCACTAAAATAATAATTTGCATTTTGCGGTAGCATATAATAATAAAGGT 477
33  amb_vs_mon      ----- 0 65  amb_vs_mon      -----AAGGTT 6
34
35

```

67				99			
68	amb_vs_pro	ACAAAAGTTAAT-----GCAAAAGCAACTTTTAAAAACA----	356	100	amb_vs_pro	ATGT-----	449
69	amb_vs_fel	TGGACTCTAAATAATTGGGGAAAATATAGAAAAGCACTAGGTGTTTGGATAAAGCTATT	345	101	amb_vs_fel	CTTATTGATATTAATAAAGCAATTGAAATTAATCCAAATGAGATTGAAATTTTATCATAAA	585
70	amb_vs_amb	AATTCCTTTAAGTAATTTAGGAAAAGCATCAAGAAGCAATAGAAAATTTGTGATCCTTGCCCTA	537	102	amb_vs_amb	ATTGAAACCTGTGATATAGCTATAAAAATATCAGCCCCATTGTGTAGAGCATACTATAAC	777
71	amb_vs_buc	AATTCCTTTAAGTAATTTAGGAAAAGCATCAAGAAGCAATAGAAAATTTGTGATCCTTGCCCTA	537	103	amb_vs_buc	ATTGAAACCTGTGATATAGCTATAAAAATATCAGCCCCATTGTGTAGAGCATACTATAAC	777
72	amb_vs_end	AATTCCTTTAAGTAATTTAGGAAAAGCATCAAGAAGCAATAGAAAATTTGTGATCCTTGCCCTA	537	104	amb_vs_end	ATTGAAACCTGTGATATAGCTATAAAAATATCAGCCCCATTGTGTAGAGCATACTATAAC	777
73	amb_vs_mon	AGTGCAATTAATGACTTAGGAGAATACCAAAATGCAATAGAAAATTTTGATCTGCTCTT	66	105	amb_vs_mon	ATTGAAACCTGTAATATAGCCATTAATATCAGCCCCATTGTGTAGAGCATACTATAAC	306
74				106		*	
75				107			
76	amb_vs_pro	-----AACATTACAAGAACTATCATATACCAGAAACCACTAAACTTGGATA	404	108	amb_vs_pro	-----G	450
77	amb_vs_fel	GAACTTGATCCTACTGATATAAAGCTTACGCTAATAAAGCAATATCTTTGAAAAAATTG	405	109	amb_vs_fel	AAAGGCTGGATTTATCAGAGCTTTATAAATATGATGAGGCAATGAAATGTTGTGATAAA	645
78	amb_vs_amb	AAATATAACCCCAAGATATCCAGAAGCATATTATAAATAAAGGAATATCCTTAATGTCTTTA	597	110	amb_vs_amb	AAGGCTATCTCTTTAAACATATTAAAAAATATCATGAAGCAATAGAGAATTGTGATTTA	837
79	amb_vs_buc	AAATATAACCCCAAGATATCCAGAAGCATATTATAAATAAAGGAATATCCTTAATGTCTTTA	597	111	amb_vs_buc	AAGGCTATCTCTTTAAACATATTAAAAAATATCATGAAGCAATAGAGAATTGTGATTTA	837
80	amb_vs_end	AAATATAACCCCAAGATATCCAGAAGCATATTATAAATAAAGGAATATCCTTAATGTCTTTA	597	112	amb_vs_end	AAGGCTATCTCTTTAAACATATTAAAAAATATCATGAAGCAATAGAGAATTGTGATTTA	837
81	amb_vs_mon	AAATATAACCCCAAGATATCCAGAAGCATATTATAAATAAAGGAATATCCTTAATGTCTTTA	126	113	amb_vs_mon	AAGGCTATCTCTTTAAACATATTAAAAAATATCATGAAGCAATAGAGAATTGTGATTTA	366
82		* * * * *		114			
83				115			
84	amb_vs_pro	-----	404	116	amb_vs_pro	CTACTTGAGTATAAT-----	465
85	amb_vs_fel	GGTAAATTTGATGAATGGTCCATATGTATAAATAAAGCTTTGGAAATAAATCCCAATTAT	465	117	amb_vs_fel	ATTATTGAGATTGATTCCTCTAGAACCTTACGGTTATTTTAAATAAAGGATCAATGTTACGC	705
86	amb_vs_amb	TGGCAACATCAAGAAGCAATAAAAAATTTTGATTTAGCTATTAAATATAAATCAGATTAT	657	118	amb_vs_amb	GCAATCAAAATATGATCCTACTTATACAGCATCATACTTAGAAAAAGGAATAGCTCTTAGA	897
87	amb_vs_buc	TGGCAACATCAAGAAGCAATAAAAAATTTTGATTTAGCTATTAAATATAAATCAGATTAT	657	119	amb_vs_buc	GCAATCAAAATATGATCCTACTTATACAGCATCATACTTAGAAAAAGGAATAGCTCTTAGA	897
88	amb_vs_end	TGGCAACATCAAGAAGCAATAAAAAATTTTGATTTAGCTATTAAATATAAATCAGATTAT	657	120	amb_vs_end	GCAATCAAAATATGATCCTACTTATACAGCATCATACTTAGAAAAAGGAATAGCTCTTAGA	897
89	amb_vs_mon	TGGCAACATCAGGAGCGATAGAAAATCTGATTTAGCTATTAAATATAAATCTGATTAT	186	121	amb_vs_mon	GCAATCGAATATGATCCTAATTTATATAGCATCATACTTAGAAAAAGGAGTATCTTTTAGA	426
90				122		* * * * *	
91				123			
92	amb_vs_pro	-----TAACATCGGTAATAATACTTTATGAGTATAAAAAATATCA	445	124	amb_vs_pro	-----	465
93	amb_vs_fel	TCTGAACTATATTTTGATAGAGCTRAGGCTTAAATTTGATATAGGTGRAAAAAGAAAAGCA	525	125	amb_vs_fel	CTCAAATAAAAAATATGAGGAAGCAATTGAAGCTTCAATATGGCTATTAATCTTATGCCA	765
94	amb_vs_amb	GCAGATGCTTATAACAACAAGGCTGCAGCTTTAATAAACCCTTAGGAAAATATCAAGAAGCA	717	126	amb_vs_amb	AAATTAGGTAARATATCAAGAAGCAATTAAATAGCTATGATCTAGCTATTAAGTATAAATCT	957
95	amb_vs_buc	GCAGATGCTTATAACAACAAGGCTGCAGCTTTAATAAACCCTTAGGAAAATATCAAGAAGCA	717	127	amb_vs_buc	AAATTAGGTAARATATCAAGAAGCAATTAAATAGCTATGATCTAGCTATTAAGTATAAATCT	957
96	amb_vs_end	GCAGATGCTTATAACAACAAGGCTGCAGCTTTAATAAACCCTTAGGAAAATATCAAGAAGCA	717	128	amb_vs_end	AAATTAGGTAARATATCAAGAAGCAATTAAATAGCTATGATCTAGCTATTAAGTATAAATCT	957
97	amb_vs_mon	GCAGATGCTTATAACAACAAGGCTGCAGCTTTAATAAACCCTTAGGAAAATATCAAGAAGCA	246	129	amb_vs_mon	AAATTAGGTAARATATCAAGAAGCAATTAAAGGCTACGATCTAGCTATTAAGTATAAATCT	486
98		* * * * *		130			
99							
131				147			
132	amb_vs_pro	-----	465	148	amb_vs_pro	-----	465
133	amb_vs_fel	ACTGAAGCTGAATATATTTTGAAGGTCGATGTTTATATGAATAAAAAGATATAAA	825	149	amb_vs_fel	-----	885
134	amb_vs_amb	GATTATGCTGAAAGTTACTTAGAAAAGGAATTACATTTGATTTAGCATGGGAAAAGAAAA	1017	150	amb_vs_amb	ACAGAAATAATTAAGCTTTAAGAAAAGGTAATGATCTTATGGCTGACGAGTTTGAA	1137
135	amb_vs_buc	GATTATGCTGAAAGTTACTTAGAAAAGGAATTACATTTGATTTAGCATGGGAAAAGAAAA	1017	151	amb_vs_buc	ACAGAAATAATTAAGCTTTAAGAAAAGGTAATGATCTTATGGCTGACGAGTTTGAA	1137
136	amb_vs_end	GATTATGCTGAAAGTTACTTAGAAAAGGAATTACATTTGATTTAGCATGGGAAAAGAAAA	1017	152	amb_vs_end	ACAGAAATAATTAAGCTTTAAGAAAAGGTAATGATCTTATGGCTGACGAGTTTGAA	1137
137	amb_vs_mon	GATTATGCTGAAAGTTACTTAGAAAAGGATTTGTTTTAGTTAAATTTAGGTAACATATA	546	153	amb_vs_mon	GAAGCAATAATTAAGCTTTAAGAAAATGAGTAATGATCTTATGGCTGATGAGTTTGAA	666
138				154			
139				155			
140	amb_vs_pro	-----	465	156	amb_vs_pro	-----	465
141	amb_vs_fel	GAAGCAGTTAAGGAGTTAATAAAGCTATTAAGTTTGAGCCGGATATTTTCATCTTTTAC	885	157	amb_vs_fel	-----	885
142	amb_vs_amb	GAAGCTAAAAGTAAATTTTCACTTAGCTCTTAAATATAAACCTAATCTTATGACAGAAATAT	1077	158	amb_vs_amb	GAAAAATTAAGATAAATGCAAGATAAATTTG	1167
143	amb_vs_buc	GAAGCTAAAAGTAAATTTTCACTTAGCTCTTAAATATAAACCTAATCTTATGACAGAAATAT	1077	159	amb_vs_buc	GAAAAATTAAGATAAATGCAAGATAAATTTG	1167
144	amb_vs_end	GAAGCTAAAAGTAAATTTTCACTTAGCTCTTAAATATAAACCTAATCTTATGACAGAAATAT	1077	160	amb_vs_end	GAAAAATTAAGATAAATGCAAGATAAATTTG	1167
145	amb_vs_mon	GAAGCTAAGGAAAATTTTAAATTTAGCTATTAAATACAAAACCTAATCTTATAGAAGATAT	606	161	amb_vs_mon	GAAAAATTAAGATAAATAAAAATAAATTTG	696
146				162			
				163			

Cluster.1816.BLASTp.

1	CLUSTAL O(1.2.4) multiple sequence alignment						
2							
3							
4	amb_pro	-----	0	29	amb_pro	-----NKSKNLNKIIKGVQNNFQTLNLYKYRTKYNLHFPSIGFKYDYSRASNYK	51
5	amb_fel	-----	0	30	amb_fel	INPKISGYDNRGIVLSILGLYEAA---IKSYDQALKLDSLDVLTFF--INKGWTLNNWG	102
6	EER21691	MKIIEKKFMQNIIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	60	32	EER21691	YGSSFTTEAYNNKGNLSLNKLGRYQEA---IKNFDLALKYNLHFVAVY--NNKGNLSNLG	174
7	amb_amb	-----MQNIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	52	33	amb_amb	YGSSFTTEAYNNKGNLSLNKLGRYQEA---IKNFDLALKYNLHFVAVY--NNKGNLSNLG	166
8	KJV62536.1	-----MQNIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	52	34	KJV62536.1	YGSSFTTEAYNNKGNLSLNKLGRYQEA---IKNFDLALKYNLHFVAVY--NNKGNLSNLG	166
9	amb_end	-----MQNIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	52	35	amb_end	YGSSFTTEAYNNKGNLSLNKLGRYQEA---IKNFDLALKYNLHFVAVY--NNKGNLSNLG	166
10	amb_buc	-----MQNIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	52	36	amb_buc	YGSSFTTEAYNNKGNLSLNKLGRYQEA---IKNFDLALKYNLHFVAVY--NNKGNLSNLG	166
11	KDO02633	MKIIEKKFMQNIIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	60	37	KDO02633	YGSSFTTEAYNNKGNLSLNKLGRYQEA---IKNFDLALKYNLHFVAVY--NNKGNLSNLG	174
12	EER22348	-----MQNIEKLAIFIFILSLIFFSSAFAENLINPTNARQQQIEQQSKIQDSNILAAE	52	38	EER22348	YNPHCAEAYNNKGNVLTNTIGQYQEA---ITLYDIATINYNASFAEAY--NNKGNLSNLG	166
13	amb_mon	-----	0	39	amb_mon	-----KGSALNDLG	9
14	CEO17529	-----	0	40	CEO17529	-----	0
15				41			
16				42			
17	amb_pro	-----	0	43	amb_pro	EYNVDIENLMIQKKSNSQSFESSIGAKIVSKPIISK----NNIILTSAHGNIERHFNNK	106
18	amb_fel	-----KYETAIKYFDKVLNQDPFLSSAYSNKAFAKRLCRIEAELEVLDTAIAK	48	44	amb_fel	KYRKALSVDLRAIEIDPDTIKAYANKAISLEKLGKFDVMYHNKALEINPNYSELYFDR	162
19	EER21691	YLNIGKSFYKLGKYQEAIKNFDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIQYNNLAIK	120	45	EER21691	KHQEAIEIENCDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	234
20	amb_amb	YLNIGKSFYKLGKYQEAIKNFDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIQYNNLAIK	112	46	amb_amb	KHQEAIEIENCDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	226
21	KJV62536.1	YLNIGKSFYKLGKYQEAIKNFDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIQYNNLAIK	112	47	KJV62536.1	KHQEAIEIENCDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	226
22	amb_end	YLNIGKSFYKLGKYQEAIKNFDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIQYNNLAIK	112	48	amb_end	KHQEAIEIENCDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	226
23	amb_buc	YLNIGKSFYKLGKYQEAIKNFDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIQYNNLAIK	112	49	amb_buc	KHQEAIEIENCDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	226
24	KDO02633	YLNIGKSFYKLGKYQEAIKNFDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIQYNNLAIK	120	50	KDO02633	KHQEAIEIENCDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	234
25	EER22348	YFNIGKSLYKLGKYQEAIKNYDLAIKYPDYEDAYNYKGMVLARLKGKYQEAIRKNDLSIK	112	51	EER22348	KYQETIENFDLALKYNSRYPEAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	226
26	amb_mon	-----	0	52	amb_mon	EYQNAIENFDLALKYNFRYPYAYYNGKISLMSLWQHQAIAKNFDLAIKYSYADAYNNK	69
27	CEO17529	-----	0	53	CEO17529	-----MSLWQHQAIAKNFDLAIKYSYADAYNNK	30
28				54		: : . : : : : :	
				55			
55				74	amb_end	YTASYLEKGIARLKLKGYQEAINSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	346
56	amb_pro	NTKVNKATPFKKQTLQETIIIPKQPKL---GYNIG-----NNILMSIKNINVLLEYN--	155	75	amb_buc	YTASYLEKGIARLKLKGYQEAINSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	346
57	amb_fel	AKALIDIGEREKALIDINKAIEINPNEIEFYHKKGWILSELYKYDEAIECCDRIIEIDSL	222	76	KDO02633	YTASYLEKGIARLKLKGYQEAINSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	354
58	EER21691	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	294	77	EER22348	YTASYLEKGIARLKLKGYQEAIAKYNLDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	346
59	amb_amb	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	286	78	amb_mon	YTASYLEKGVSPFRKLGKYQEAIAKYSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	189
60	KJV62536.1	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	286	79	CEO17529	YTASYLEKGVSPFRKLGKYQEAIAKYSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	150
61	amb_end	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	286	80			
62	amb_buc	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	286	81			
63	KDO02633	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	294	82	amb_pro	-----	155
64	EER22348	GATLNLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	286	83	amb_fel	RAIKFEPDISSYY-----	295
65	amb_mon	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	129	84	EER21691	LALKYKPNLMTEYTEIIKALRKLGNLMDAFEFEKLRKIMQDNL	397
66	CEO17529	GAALINLKGKYQEAIEETCDIAIKYQPHCVAEAYYKKAISLNILNKKYHEAIEIENCDLAIKYDPT	90	85	amb_amb	LALKYKPNLMTEYTEIIKALRKLGNLMDAFEFEKLRKIMQDNL	389
67		: . : * :* : : : : : : : : : : : :		86	KJV62536.1	LALKYKPNLMTEYTEIIKALRKLGNLMDAFEFEKLRKIMQDNL	389
68				87	amb_end	LALKYKPNLMTEYTEIIKALRKLGNLMDAFEFEKLRKIMQDNL	389
69	amb_pro	-----	155	88	amb_buc	LALKYKPNLMTEYTEIIKALRKLGNLMDAFEFEKLRKIMQDNL	389
70	amb_fel	EPYGYFNGSMLRLNKKYEEAIEAFNMAINLMPTEAEALYVLRGRCLYELKRYKEAVKEFN	282	89	KDO02633	LALKYKPNLMTEYTEIIKALRKLGNLMDAFEFEKLRKIMQDNL	397
71	EER21691	YTASYLEKGIARLKLKGYQEAINSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	354	90	EER22348	LALKYKPNLMTEYRGIKALRKLGNLMDAFEFEKLRKIMQDNL	389
72	amb_amb	YTASYLEKGIARLKLKGYQEAINSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	346	91	amb_mon	LAIKYPNLIIEYEAIKALRKLGNLMDAFEFEKLRKIMQDNL	232
73	KJV62536.1	YTASYLEKGIARLKLKGYQEAINSYDLAIKYSYDAESYLEKGITLISMGRKEAKVNFH	346	92	CEO17529	LAIKYPNLIIEYEAIKALRKLGNLMDAFEFEKLRKIMQDNL	193
74				93			
				94			

Cluster.1050.TBLASTn.

1 CLUSTAL O(1.2.4) multiple sequence alignment

```

2
3
4 amb.2_vs_amb      ATGACAATTAATAATCAAATCAACCAAAAGATAATCACAAATATGCAAAATCTTACTAAA 60 31
5 amb.1_vs_amb      ATGACAATTAATAATCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 32
6 amb.1_vs_end      ATGACAATTAATAATCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 33
7 amb.2_vs_end      ATGACAATTAATAATCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 34
8 amb.1_vs_buc      ATGACAATTAATAATCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 35
9 amb.2_vs_buc      ATGACAATTAATAATCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 36
10 amb.1_vs_mon      ATGACAATTAATAACCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 37
11 amb.2_vs_mon      ATGACAATTAATAACCAAATCAACCAAAAG-----ATAATCATAATATTACTAAA 51 38
12 amb.1_vs_fel      ----- 0 39
13 amb.2_vs_fel      ----- 0 40
14 amb.1_vs_pro      ----- 0 41
15 amb.2_vs_pro      -----ACTAAC-----TATGCTAGGCAATTTATCGAGGAACAGCTTGAA 39 42
16
17
18 amb.2_vs_amb      ACTAATTTGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 120 43
19 amb.1_vs_amb      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCCGCAAGAA 111 44
20 amb.1_vs_end      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 111 45
21 amb.2_vs_end      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 111 46
22 amb.1_vs_buc      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 111 47
23 amb.2_vs_buc      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 111 48
24 amb.1_vs_mon      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 111 49
25 amb.2_vs_mon      ACAAAATGGAATCTTCAATTTATCTCCCATATCTGTCCACATTAATACCATATCAGCAAGAA 111 50
26 amb.1_vs_fel      ----- 0 51
27 amb.2_vs_fel      ----- 0 52
28 amb.1_vs_pro      ----- 0 53
29 amb.2_vs_pro      ACAAAATC-----GCAATTTAAAAGCTCAAATAGAAAAA 72 54
30
31

```

```

32 amb.2_vs_amb      GAAAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 180
33 amb.1_vs_amb      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
34 amb.1_vs_end      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
35 amb.2_vs_end      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
36 amb.1_vs_buc      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
37 amb.2_vs_buc      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
38 amb.1_vs_mon      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
39 amb.2_vs_mon      GAGAATATCATTAACTCGACGATCACAATTTATACTGATATAATTAAGTATAAAATTAGAA 171
40 amb.1_vs_fel      -----AATATTAGTCATGAATTG--- 18
41 amb.2_vs_fel      -----AATATTAGTCATGAATTG--- 18
42 amb.1_vs_pro      -----AATATTAGTCATGAATTA--- 18
43 amb.2_vs_pro      GTAGAAAAAAGCTAGCTCATCCAATACAAAGTTTTTTGCTAATATTAGTCATGAATTA--- 129
44
45
46 amb.2_vs_amb      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 233
47 amb.1_vs_amb      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
48 amb.1_vs_end      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
49 amb.2_vs_end      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
50 amb.1_vs_buc      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
51 amb.2_vs_buc      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
52 amb.1_vs_mon      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
53 amb.2_vs_mon      TTACRAGAAGCAAATAATAAATTGAAAGA---ATCTGAG---TTATTTAAAAGAGATGT 224
54 amb.1_vs_fel      ---CGTACACCACCTTAATGCTATAATAGGCTTTTCGGAAATTCCTTATGTCGGAAAGAGAT 75
55 amb.2_vs_fel      ---CGTACACCACCTTAATGCTATAATAGGCTTTTCGGAAATTCCTTATGTCGGAAAGAGAT 75
56 amb.1_vs_pro      ---CGTACGCCACTTAATGCTATAATAGGCTTTTCAGAGATTTTGTGATGTCGGAACGAGAT 75
57 amb.2_vs_pro      ---CGTACGCCACTTAATGCTATAATAGGCTTTTCAGAGATTTTGTGATGTCGGAACGAGAT 186
58
59

```

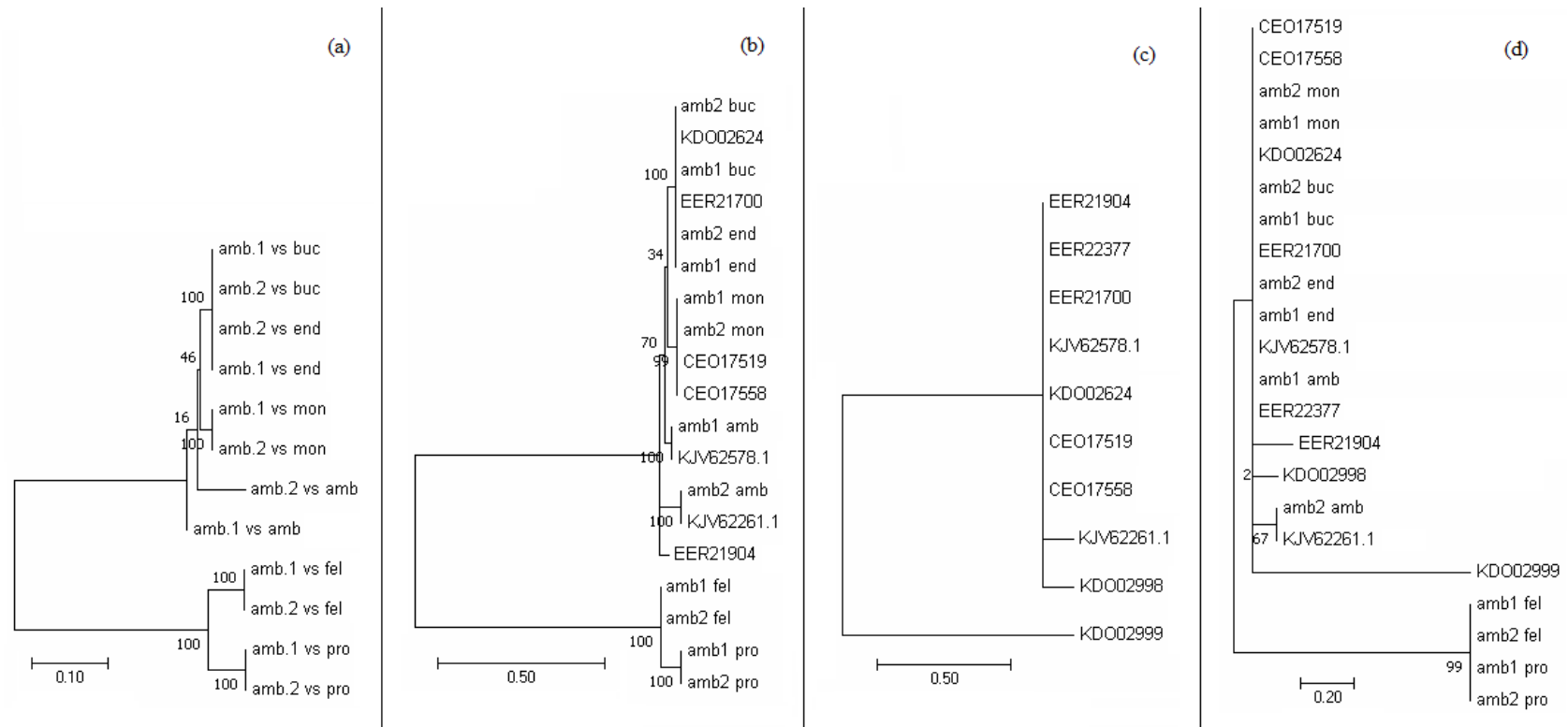

171				199							
172	amb.2_vs_amb	TATAGTGCATGATAGTGGCATAGGTCCTATCTAAAGAAATC-AGGCAACATCTTCATGAAA	772	200	amb.2_vs_amb	GTTTTGTAAAACGCTAATTTGGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	886				
173	amb.1_vs_amb	TATGGTGCATGATAGTGGTGTGGTTTGTCTCAAGAAAT-AGGCAACATCTTCATGAAA	763	201	amb.1_vs_amb	GTTTTGTAAAACGCTAATTTGGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
174	amb.1_vs_end	TATGGTGCATGATAGTGGTGTGGTTTGTCTAAAGAAATC-AGGCAACATCTTCATGAAA	763	202	amb.1_vs_end	GTTTTGTAAAACGCTAATTTAGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
175	amb.2_vs_end	TATGGTGCATGATAGTGGTGTGGTTTGTCTAAAGAAATC-AGGCAACATCTTCATGAAA	763	203	amb.2_vs_end	GTTTTGTAAAACGCTAATTTAGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
176	amb.1_vs_buc	TATGGTGCATGATAGTGGTGTGGTTTGTCTAAAGAAATC-AGGCAACATCTTCATGAAA	763	204	amb.1_vs_buc	GTTTTGTAAAACGCTAATTTAGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
177	amb.2_vs_buc	TATGGTGCATGATAGTGGTGTGGTTTGTCTAAAGAAATC-AGGCAACATCTTCATGAAA	763	205	amb.2_vs_buc	GTTTTGTAAAACGCTAATTTAGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
178	amb.1_vs_mon	TATGGTGCATGATAGTGGCATTGGTCTGTCTAAAGAAATC-AGGAAACATCTTCATGAAA	763	206	amb.1_vs_mon	GTTTTGTAAAACGCTAATTTGGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
179	amb.2_vs_mon	TATGGTGCATGATAGTGGCATTGGTCTGTCTAAAGAAATC-AGGAAACATCTTCATGAAA	763	207	amb.2_vs_mon	GTTTTGTAAAACGCTAATTTGGTGAATTTGGGTGGAGAGATTGAAGTAACCAAGCGAAAAAG	877				
180	amb.1_vs_fel	AAAAGTTATAGATACCGGCATCGGTATTGAGGAAAAAGATATCCCAAAAACCTTTATCGGC	497	208	amb.1_vs_fel	CGCTTACTAAAAAGCTGGTGAACCTTATGAACGGTAAATTTGATTTGCAAAGTAAGATAA	616				
181	amb.2_vs_fel	AAAAGTTATAGATACCGGCATCGGTATTGAGGAAAAAGATATCCCAAAAACCTTTATCGGC	497	209	amb.2_vs_fel	CGCTTACTAAAAAGCTGGTGAACCTTATGAACGGTAAATTTGATTTGCAAAGTAAGATAA	616				
182	amb.1_vs_pro	AAAAGTTATAGATACCGGCATCGGTATTGAGGAAAAAGATATCCCAAAAACCTTTATCGGC	497	210	amb.1_vs_pro	CTCTGACTAAAAAGCTAGTTGAACCTTATGAATGGTAAATTTGATTTGCAAAGTAAGATAA	616				
183	amb.2_vs_pro	AAAAGTTATAGATACCGGCATCGGTATTGAGGAAAAAGATATCCCAAAAACCTTTATCGGC	608	211	amb.2_vs_pro	CTCTGACTAAAAAGCTAGTTGAACCTTATGAATGGTAAATTTGATTTGCAAAGTAAGATAA	727				
184		* * * * *		212		* * * * *					
185				213							
186	amb.2_vs_amb	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	826	214	amb.2_vs_amb	GTAAAAGTACTACTATTGTCTGTAACATACCAGTTAAATTTACCTCTTCTTGACGATCTGA	946				
187	amb.1_vs_amb	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	215	amb.1_vs_amb	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTAA	937				
188	amb.1_vs_end	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	216	amb.1_vs_end	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTGA	937				
189	amb.2_vs_end	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	217	amb.2_vs_end	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTGA	937				
190	amb.1_vs_buc	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	218	amb.1_vs_buc	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTGA	937				
191	amb.2_vs_buc	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	219	amb.2_vs_buc	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTGA	937				
192	amb.1_vs_mon	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	220	amb.1_vs_mon	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTGA	937				
193	amb.2_vs_mon	AGATTGATAAAATTTGATGTTGC-----TATGCAATATAATAGACTCGAATCAGGATTAA	817	221	amb.2_vs_mon	GTAAAAGTACTACTATTGTCTGTAATATACCAATTAAGTTACCTCTTCTTGATGATCTGA	937				
194	amb.1_vs_fel	-GTTTGGTCAAATTTGATAGTGAGCTTAGTCGAAAAATATGAAGGTACCGGGCTTGGGCTGC	556	222	amb.1_vs_fel	ATGAAGGTACCACCATAACAATA-----	639				
195	amb.2_vs_fel	-GTTTGGTCAAATTTGATAGTGAGCTTAGTCGAAAAATATGAAGGTACCGGGCTTGGGCTGC	556	223	amb.2_vs_fel	ATGAAGGTACCACCATAACAATA-----	639				
196	amb.1_vs_pro	-ATTTGGTCAAATTTGATAGTGAGCTTAGTCGAAAAATATGAAGGTACTGGTCTTGGATTGC	556	224	amb.1_vs_pro	ATAAAAGTACTACAGTC-----	633				
197	amb.2_vs_pro	-ATTTGGTCAAATTTGATAGTGAGCTTAGTCGAAAAATATGAAGGTACTGGTCTTGGATTGC	667	225	amb.2_vs_pro	ATAAAAGTACTACAGTCACCAATA-----	750				
198		* * * * *		226		* * * * *					
199				227							
				228	amb.2_vs_amb	TTTATGATAAA	957	235	amb.2_vs_mon	TTTATGAT---	945
				229	amb.1_vs_amb	TTTATGATACA	948	236	amb.1_vs_fel	-----	639
				230	amb.1_vs_end	TTTATGATACA	948	237	amb.2_vs_fel	-----	639
				231	amb.2_vs_end	TTTATGAT---	945	238	amb.1_vs_pro	-----	633
				232	amb.1_vs_buc	TTTATGATACA	948	239	amb.2_vs_pro	-----	750
				233	amb.2_vs_buc	TTTATGAT---	945	240			
				234	amb.1_vs_mon	TTTATGATACA	948				

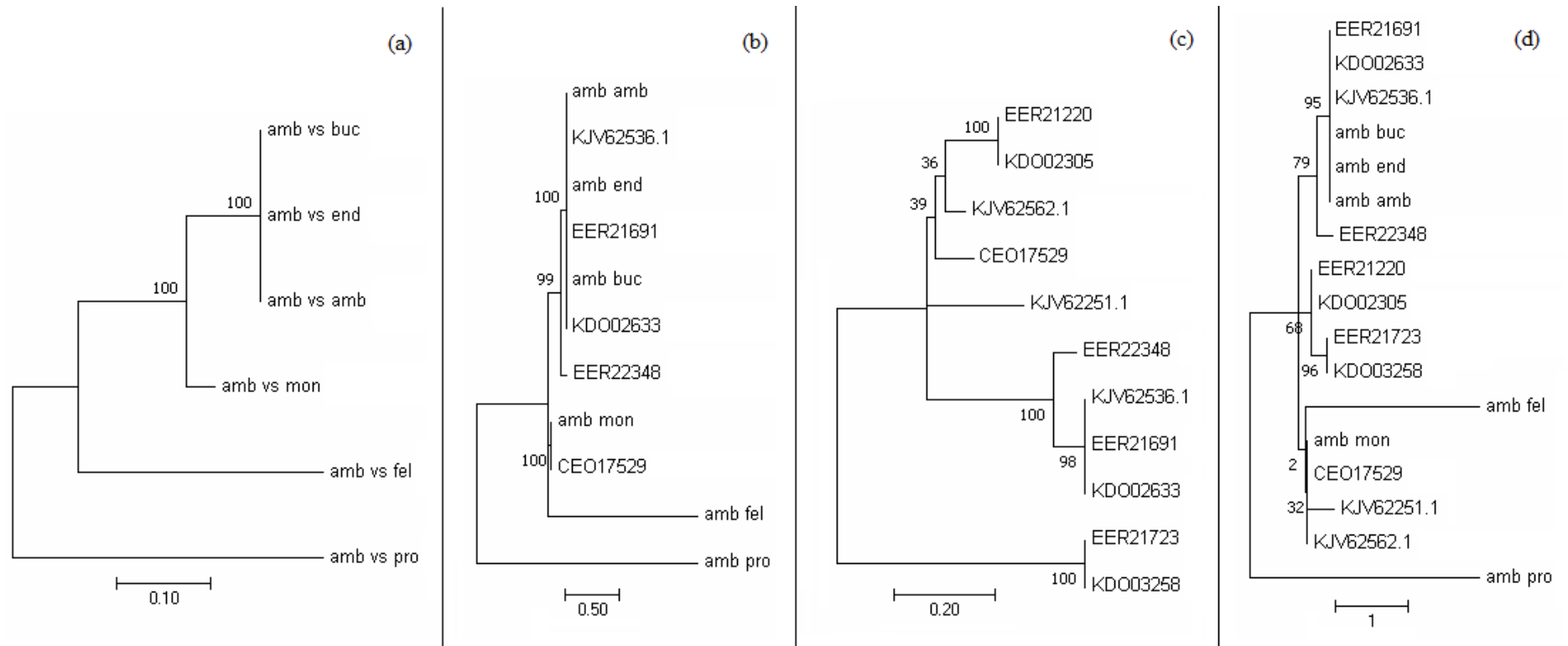
Cluster.1050.BLASTp.

1	CLUSTAL O(1.2.4) multiple sequence alignment				
2					
3					
4	amb1_fel	-----	0	44	
5	amb2_fel	-----	0	46	amb1_fel
6	amb1_pro	-----	0	47	amb2_fel
7	amb2_pro	-----TNYARQIIIEEQL	12	48	amb1_pro
8	amb1_amb	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	49	amb2_pro
9	KJV62578.1	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	50	amb1_amb
10	amb1_end	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	51	KJV62578.1
11	amb2_end	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	52	amb1_end
12	EER21700	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	53	amb2_end
13	amb1_buc	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	54	EER21700
14	amb2_buc	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	55	amb1_buc
15	KD002624	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	56	amb2_buc
16	CE017558	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	57	KD002624
17	amb1_mon	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	58	CE017558
18	amb2_mon	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	59	amb1_mon
19	CE017519	MTINNQNQPKDNHNI---TKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	56	60	amb2_mon
20	amb2_amb	MTINNQNQPKDNHNIANLTKTNWNLQLSPISVTLIPYQEEIINSKIITNYTDII-KYKL	59	61	CE017519
21	KJV62261.1	MTINNQNQPKDNHNIANLTKTNWNLQLSPISVTLIPYQEEIINSKIITNYTDII-KYKL	59	62	amb2_amb
22	EER21904	MTINNQNQPKDNHNIINLTKTNWNLQLSPISVTLIPYQEEIINSTIITNYTDII-KYKL	59	63	KJV62261.1
23				64	EER21904
24				65	
25				66	
26	amb1_fel	-----NISELHRTPLNAIIGFSEILMSERDAEKKNKYIKD	35	67	amb1_fel
27	amb2_fel	-----NISELHRTPLNAIIGFSEILMSERDAEKKNKYIKD	35	68	amb2_fel
28	amb1_pro	-----NISELHRTPLNAIIGFSEILMSERDEKCKRNYIKD	35	69	amb1_pro
29	amb2_pro	ETNRNLKAIQIEKVEKTSSTKFFANISELHRTPLNAIIGFSEILMSERDEKCKRNYIKD	72	70	amb2_pro
30	amb1_amb	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	71	amb1_amb
31	KJV62578.1	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	72	KJV62578.1
32	amb1_end	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	73	amb1_end
33	amb2_end	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	74	amb2_end
34	EER21700	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	75	EER21700
35	amb1_buc	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	76	amb1_buc
36	amb2_buc	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	77	amb2_buc
37	KD002624	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	78	KD002624
38	CE017558	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	79	CE017558
39	amb1_mon	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	80	amb1_mon
40	amb2_mon	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	81	amb2_mon
41	CE017519	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	112	82	CE017519
42	amb2_amb	E----LQEANNKLESELFKRDVVQNIHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	115	83	amb2_amb
43	KJV62261.1	E----LQEANNKLESELFKRDVVQNIHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	115	84	KJV62261.1
44	EER21904	E----LQEANNKLESELFKRDVVQNMVHGLKIPCNIGIFSLVTALYEIEDHPEKKRDYLAS	115	85	EER21904
45				86	

87					108		
88	amb1_fel	VYIKVIDTGIGIEEKDIPKTL	SAFGQIDSELSRKYEGTGLGLPLTKKLVELMNGKFDLQS	203	109	amb1_fel	KINEGTTITI----- 213
89	amb2_fel	VYIKVIDTGIGIEEKDIPKTL	SAFGQIDSELSRKYEGTGLGLPLTKKLVELMNGKFDLQS	203	110	amb2_fel	KINEGTTITI----- 213
90	amb1_pro	VYIKVIDTGIGIEEKDIPKTL	SAFGQIDSELSRKYEGTGLGLPLTKKLVELMNGKFDLQS	203	111	amb1_pro	KINKGTTV----- 211
91	amb2_pro	VYIKVIDTGIGIEEKDIPKTL	SAFGQIDSELSRKYEGTGLGLPLTKKLVELMNGKFDLQS	240	112	amb2_pro	KINKGTTVTI----- 250
92	amb1_amb	FQFMVHDSGVGLSQEIRQHL	HEKID--KFDVATQYHGLGSGLSFVKRLIGELNGEIEVTS	290	113	amb1_amb	EK GKSTTIVCNIPIKLPL
93	KJV62578.1	FQFMVHDSGVGLSQEIRQHL	HEKID--KFDVATQYHGLGSGLSFVKRLIGELNGEIEVTS	290	114	KJV62578.1	EK GKSTTIVCNIPIKLPL
94	amb1_end	FQFMVHDSGVGLSKEIRQHL	HEKID--KFDVAMQYNGLSGSLFVKRLISELNGEIEVTS	290	115	amb1_end	EK GKSTTIVCNIPIKLPL
95	amb2_end	FQFMVHDSGVGLSKEIRQHL	HEKID--KFDVAMQYNGLSGSLFVKRLISELNGEIEVTS	290	116	amb2_end	EK GKSTTIVCNIPIKLPL
96	EER21700	FQFMVHDSGVGLSKEIRQHL	HEKID--KFDVAMQYNGLSGSLFVKRLISELNGEIEVTS	290	117	EER21700	EK GKSTTIVCNIPIKLPL
97	amb1_buc	FQFMVHDSGVGLSKEIRQHL	HEKID--KFDVAMQYNGLSGSLFVKRLISELNGEIEVTS	290	118	amb1_buc	EK GKSTTIVCNIPIKLPL
98	amb2_buc	FQFMVHDSGVGLSKEIRQHL	HEKID--KFDVAMQYNGLSGSLFVKRLISELNGEIEVTS	290	119	amb2_buc	EK GKSTTIVCNIPIKLPL
99	KDO02624	FQFMVHDSGVGLSKEIRQHL	HEKID--KFDVAMQYNGLSGSLFVKRLISELNGEIEVTS	290	120	KDO02624	EK GKSTTIVCNIPIKLPL
100	CE017558	FQFMVHDSGIGLSKEIRKHL	HEKID--KFDVAIQYHGLGSGLSFVKRLIGELNGEIEVTS	290	121	CE017558	EK GKSTTIVCNIPIKLPL
101	amb1_mon	FQFMVHDSGIGLSKEIRKHL	HEKID--KFDVAIQYHGLGSGLSFVKRLIGELNGEIEVTS	290	122	amb1_mon	EK GKSTTIVCNIPIKLPL
102	amb2_mon	FQFMVHDSGIGLSKEIRKHL	HEKID--KFDVAIQYHGLGSGLSFVKRLIGELNGEIEVTS	290	123	amb2_mon	EK GKSTTIVCNIPIKLPL
103	CE017519	FQFMVHDSGIGLSKEIRKHL	HEKID--KFDVAIQYHGLGSGLSFVKRLIGELNGEIEVTS	290	124	CE017519	EK GKSTTIVCNIPIKLPL
104	amb2_amb	CQFIVHDSGIGLSKEIRQHL	HEKID--KFDVAMQYNRLESGLSFVKRLIGELGGEIEVTS	293	125	amb2_amb	EK GKSTTIVCNIPVKLPL
105	KJV62261.1	CQFIVHDSGIGLSKEIRQHL	HEKID--KFDVAMQYNRLESGLSFVKRLIGELGGEIEVTS	293	126	KJV62261.1	EK GKSTTIVCNIPVKLPL
106	EER21904	CQFMVHDSGIGLSKEIRQHL	HEKID--KFDVAMQYHGLGSGLSFVKR-----	280	127	EER21904	----- 280
107			: * * * * * : : : : : : : : : : : : : : * * : : *		128		
					129		

Anexo 2 – Resultados das análises filogenéticas, adotando das sequências obtidas por meio dos alinhamentos múltiplos de sequências gênicas e de aminoácidos das proteínas ortólogas presentes nos clusters 1816 e 1050. Destaque (a) para as sequências gênicas obtidas para as bactérias do gênero *Rickettsia*, que infectam os carrapatos versus as que infectam os insetos, obtidas anteriormente com o uso do TBLASTn. Destaque (b) para as sequências de proteínas obtidas para as bactérias do gênero *Rickettsia*, que infectam os carrapatos versus as que infectam os insetos, obtidas anteriormente com o uso do BLASTp.





Anexo 3 - Tabela para anotação dos níveis da expressão de transcritos e potenciais proteínas putativas para os módulos significativos do interatoma, fazendo o uso do Tblastn contra as sequências adquiridas por meio da montagem de *novo* para os dados de transcriptomas de intestino de carrapatos das espécies de *Amblyomma* infectados por espécies de *Rickettsia*.

Carrapato	<i>Rickettsia</i> sp.	Módulo	Proteína	Fold	Nome
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Fosforilação	ISCW007014-PA.1	5,153135409	Apoptosis stimulating of P53, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Fosforilação	ISCW002027-PA.1	-26,68320468	Myosin regulatory light chain, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Fosforilação	ISCW005413-PA.1	-26,68320468	Myosin regulatory light chain, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Fosforilação	ISCW001952-PA.1	2,390128659	Oxysterol-binding protein 1A, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Fosforilação	ISCW008598-PA.1	-4,119703507	PDZ Calmodulin, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Fosforilação	ISCW008263-PA.1	3,575716907	PDZ Ionotropic glutamate receptor, putative (Fragment)
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW007722-PA.1	1,335787358	#REF!
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW001280-PA.1	-3,33103356	(S)-2-hydroxy-acid oxidase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW004322-PA.1	-3,33103356	(S)-2-hydroxy-acid oxidase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW004674-PA.1	-3,33103356	(S)-2-hydroxy-acid oxidase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW005039-PA.1	-3,33103356	(S)-2-hydroxy-acid oxidase, putative (Fragment)
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW007947-PA.1	-1,162670625	Cell adhesion molecule, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW008682-PA.1	-4,479136884	Cytochrome C oxidase subunit Va
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW008740-PA.1	-5,421906103	Cytochrome C, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW007028-PA.1	-3,33103356	Glycolate oxidase, putative (Fragment)
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW009643-PA.1	1,168502067	Mitochondrial processing peptidase alpha subunit, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW003299-PA.1	-16,52261424	NADH-ubiquinone reductase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW001592-PA.1	-3,767308339	Processing peptidase beta subunit, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Oxidação	ISCW000555-PA.1	2,470112416	Succinate dehydrogenase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Proteólise	ISCW004008-PA.1	4,978307722	20S proteasome, regulatory subunit beta, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Proteólise	ISCW005987-PA.1	-7,532556795	20S proteasome, regulatory subunit beta, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Proteólise	ISCW006948-PA.1	1,899218378	20S proteasome, regulatory subunit beta, putative

<i>A.sculptum</i>	<i>R. amblyommatis</i>	Proteólise	ISCW007217-PA.1	-5,924312929	20S proteasome, regulatory subunit beta, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Proteólise	ISCW008949-PA.1	4,978307722	Proteasome, subunit beta, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Resposta	ISCW004825-PA.1	-4,377693768	DEAD box ATP-dependent RNA helicase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Resposta	ISCW006171-PA.1	-1,252322277	DEAD RNA Helicase DNA excision repair protein ERCC-6, putative (Fragment)
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW014417-PA.1	4,037032831	143-3 isoform A isoform 2 family protein
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW009589-PA.1	-6,230465636	ABC transporter, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW000200-PA.1	3,711917521	Adenylate kinase isoenzyme, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW001951-PA.1	-5,804879692	ATPase inhibitor, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW004820-PA.1	4,525266898	DNA-directed RNA polymerase subunit E', putative (Fragment)
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW000002-PA.1	-2,800907539	FKBP-type peptidyl-prolyl cis-trans isomerase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW014265-PA.1	5,001816912	Hsp90 protein, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW000999-PA.1	3,227248105	mRNA export protein, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW001340-PA.1	-3,4553449	Myosin heavy chain, skeletal muscle or cardiac muscle, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW005985-PA.1	-5,605508524	NADH:ubiquinone oxidoreductase, NDUFV1/51kDa subunit, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW004059-PA.1	5,533766122	Protein releasing factor, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW000476-PA.1	-8,091013282	Ribosomal protein L11, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW008173-PA.1	-11,39214099	Ribosomal protein L15, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW005425-PA.1	-6,355806091	Ribosomal protein L17, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW005878-PA.1	-25,14344318	Ribosomal protein L2, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW002308-PA.1	-2,850726313	Ribosomal protein S12, 30S putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW010407-PA.1	6,454272921	Ribosomal Ubiquitin/40S Ribosomal protein S27A fusion, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW013767-PA.1	-3,736631135	Thioredoxin peroxidase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Secreção	ISCW008293-PA.1	-4,295960411	WD e F-box domain protein, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Transporte	ISCW003422-PA.1	3,583155374	FKBP-type peptidyl-prolyl cis-trans isomerase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Transporte	ISCW004348-PA.1	-2,085273835	GTPase Rho, putative (Fragment)
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Transporte	ISCW009275-PA.1	-1,330433765	GTPase, putative
<i>A.sculptum</i>	<i>R. amblyommatis</i>	Transporte	ISCW007186-PA.1	-1,911318033	Tetratricopeptide repeat protein, putative (Fragment)

<i>A. sculptum</i>	<i>R. amblyommatis</i>	Transporte	ISCW000128-PA.1	2,972788482	Tetratricopeptide repeat protein, tpr, putative
<i>A. sculptum</i>	<i>R. amblyommatis</i>	Transporte	ISCW003700-PA.1	-1,524658648	Vesicle coat complex AP-3, delta subunit, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW001795-PA.1	6,548697376	Casein kinase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW007936-PA.1	6,548697376	Casein kinase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW002430-PA.1	5,64390047	FKBP-type peptidyl-prolyl cis-trans isomerase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW002994-PA.1	3,460088143	Hsp40 DnaJ, subfamily C, member 13, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW001645-PA.1	3,694395885	Lysozyme, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW005428-PA.1	3,218646275	Mitogen-activated protein kinase MKK4, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW001124-PA.1	6,982901871	Nonmuscle myosin essential light chain
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW008605-PA.1	4,61729772	Nonmuscle myosin essential light chain, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW007213-PA.1	-2,840345879	PDZ LIN-12 protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW004410-PA.1	2,957077606	Protein kinase C terminal domain-containing protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW003161-PA.1	3,393630114	Protein phosphatase 2B, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW000084-PA.1	2,570134735	Serine/threonine protein kinase, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW007829-PA.1	2,632553266	Tyrosine kinase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Metilação	ISCW004721-PA.1	4,684738324	O-methyltransferase, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Metilação	ISCW000176-PA.1	3,947486132	Protein arginine N-methyltransferase PRMT1, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW005039-PA.1	4,082606866	(S)-2-hydroxy-acid oxidase, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW002189-PA.1	5,450143109	Cytochrome C oxidase assembly protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW008682-PA.1	7,181586512	Cytochrome C oxidase subunit Va
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW001817-PA.1	2,907874006	Dimethylglycine dehydrogenase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW005219-PA.1	-2,180802074	D-lactate dehydrogenase 2, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW000019-PA.1	6,966833045	Fad oxidoreductase, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW004580-PA.1	3,208687242	Histone Lysine-specific histone demethylase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW009643-PA.1	-2,087962825	Mitochondrial processing peptidase alpha subunit, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Oxidação	ISCW006065-PA.1	3,625188183	Titin, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW007139-PA.1	1,376597953	20S proteasome, regulatory subunit alpha type PSMA7/PRE6, putative

<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW004008-PA.1	7,497113203	20S proteasome, regulatory subunit beta, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW006948-PA.1	5,280737615	20S proteasome, regulatory subunit beta, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW004146-PA.1	-2,246873949	26S protease regulatory subunit 6A, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW005463-PA.1	-2,57280647	26S proteasome alpha 7 subunit
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008741-PA.1	1,806943808	26S proteasome regulatory complex, ATPase RPT6, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW000683-PA.1	1,730318532	ATP-dependent protease PIM1/LON, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW007723-PA.1	5,161623247	Nethylmaleimide sensitive fusion family protein
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW000634-PA.1	4,39274531	Peroxisome biogenesis factor, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008144-PA.1	1,002425274	Protease, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW006222-PA.1	-1,433323418	Proteasome beta2 subunit, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008949-PA.1	7,497113203	Proteasome, subunit beta, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW007504-PA.1	3,458528168	TATA element modulatory factor, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008933-PA.1	-2,683397528	Transitional endoplasmic reticulum ATPase, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Resposta	ISCW004825-PA.1	5,372329317	DEAD box ATP-dependent RNA helicase, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Resposta	ISCW006171-PA.1	5,17410726	DEAD RNA Helicase DNA excision repair protein ERCC-6, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Resposta	ISCW000889-PA.1	2,579326393	Dicer-1, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW004310-PA.1	5,059494737	ABC transporter
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW009589-PA.1	7,657837096	ABC transporter, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW001681-PA.1	4,451591782	ABC transporter, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW006080-PA.1	2,774367384	ABC transporter, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW006081-PA.1	-3,476992064	ABC transporter, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW009274-PA.1	6,25975105	ADP-ribosylation factor 3
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW007976-PA.1	5,112542861	Ankyrin repeat containing protein
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW002574-PA.1	3,001404588	Cytochrome P450, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW004820-PA.1	4,083624328	DNA-directed RNA polymerase subunit E', putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW008140-PA.1	3,480605689	Grb2 associated binder (Gab), putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW010554-PA.1	7,67324752	NADH:ubiquinone oxidoreductase, NDUFS8/23 kDa subunit, putative (Fragment)

<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW003608-PA.1	4,010458547	PDZ Discs large protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW003477-PA.1	5,08639146	PDZ domain-containing protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW001178-PA.1	-2,507099347	Phospholipase C, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW005281-PA.1	4,43719156	PIN domain-containing RNA-binding protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW000929-PA.1	15,46125691	PlekhH1, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW009326-PA.1	5,229594305	Protein transport protein sec23, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW000556-PA.1	-3,876065191	Ribosomal processing protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW000476-PA.1	8,624627272	Ribosomal protein L11, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW008173-PA.1	8,449097771	Ribosomal protein L15, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW005878-PA.1	4,365068557	Ribosomal protein L2, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW002021-PA.1	9,64216315	Ribosomal protein L26, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW009206-PA.1	4,359501692	Ribosomal protein S2, 40S putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW004030-PA.1	-8,248464306	Ribosomal protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW010407-PA.1	8,401989381	Ribosomal Ubiquitin/40S Ribosomal protein S27A fusion, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW009641-PA.1	7,819526888	Sterol response element binding protein cleavage-activating protein, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW008220-PA.1	9,410973658	Translation initiation factor 1A
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW000011-PA.1	-3,020655094	Translation initiation factor 3 and TGF-beta interacting protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW022766-PA.1	4,256683528	Tumor rejection antigen (Gp96), putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW015782-PA.1	3,236012286	U5 snRNP-specific protein, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW000332-PA.1	3,194240306	WD-repeat protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW008108-PA.1	3,198066833	WD-repeat protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW009107-PA.1	3,874276438	WD-repeat protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Secreção	ISCW002911-PA.1	-3,813569803	WD-repeat protein, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW002199-PA.1	2,627671583	Ap-3 complex subunit sigma, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW006433-PA.1	3,409522883	Clathrin adaptor complex, small subunit, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW002525-PA.1	-2,930507784	GTPase Dynamin, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW004349-PA.1	-1,509872642	GTPase Rho, putative

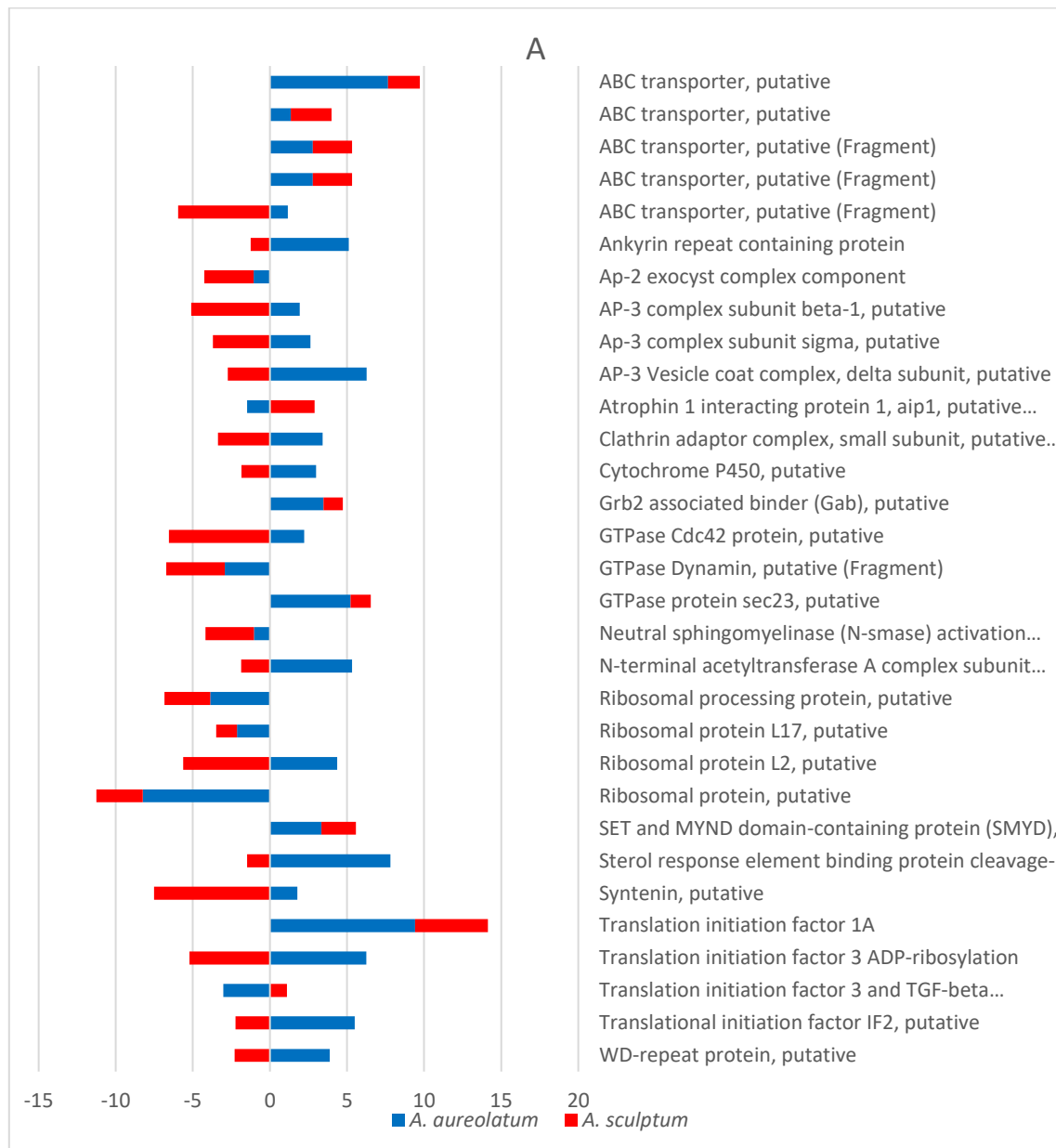
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW006741-PA.1	2,575626949	GTPase Rho, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW004348-PA.1	-1,796138584	GTPase Rho, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW007252-PA.1	3,393103731	GTP-binding protein, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW000156-PA.1	3,397472476	Guanine nucleotide-binding protein G, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW005645-PA.1	-3,84032269	Lumican, putative (Fragment)
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW000399-PA.1	5,332176431	N-terminal acetyltransferase A complex subunit NAT1, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW000409-PA.1	3,333904536	SET and MYND domain-containing protein (SMYD), putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW000128-PA.1	3,054784548	Tetratricopeptide repeat protein, tpr, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW008482-PA.1	5,495893244	Translational initiation factor IF2, putative
<i>A. aureolatum</i>	<i>R. rickettsii</i>	Transporte	ISCW003700-PA.1	6,27657569	Vesicle coat complex AP-3, delta subunit, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW004525-PA.1	-3,920551986	Ankyrin repeat containing protein
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW005225-PA.1	2,525706586	Apoptosis death- associated protein kinase, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW008769-PA.1	2,466805728	Cyclin T-dependent kinase CDK9, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW009238-PA.1	6,100417409	ETS domain-containing protein Elk-4, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW002764-PA.1	-2,891070487	Mitogen-activated protein kinase kinase MKK4, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW006910-PA.1	2,403683108	Mitogen-activated protein kinase kinase MKK4, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW002027-PA.1	-5,865822893	Myosin regulatory light chain, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW005413-PA.1	-5,865822893	Myosin regulatory light chain, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW007865-PA.1	-3,12734368	NADPH Dual oxidase 1, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW004102-PA.1	3,394702927	PDZ Calmodulin, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW008263-PA.1	-3,218589386	PDZ Ionotropic glutamate receptor, putative (Fragment)
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW003161-PA.1	4,597232545	Protein phosphatase 2B, putative (Fragment)
<i>A. sculptum</i>	<i>R. rickettsii</i>	Fosforilação	ISCW008686-PA.1	2,970964769	Serine/threonine protein kinase, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW001280-PA.1	2,625417375	(S)-2-hydroxy-acid oxidase, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW004674-PA.1	2,63465903	(S)-2-hydroxy-acid oxidase, putative
<i>A. sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW005039-PA.1	2,63465903	(S)-2-hydroxy-acid oxidase, putative (Fragment)
<i>A. sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW002944-PA.1	3,069925097	Chloride channel protein 1.2, putative

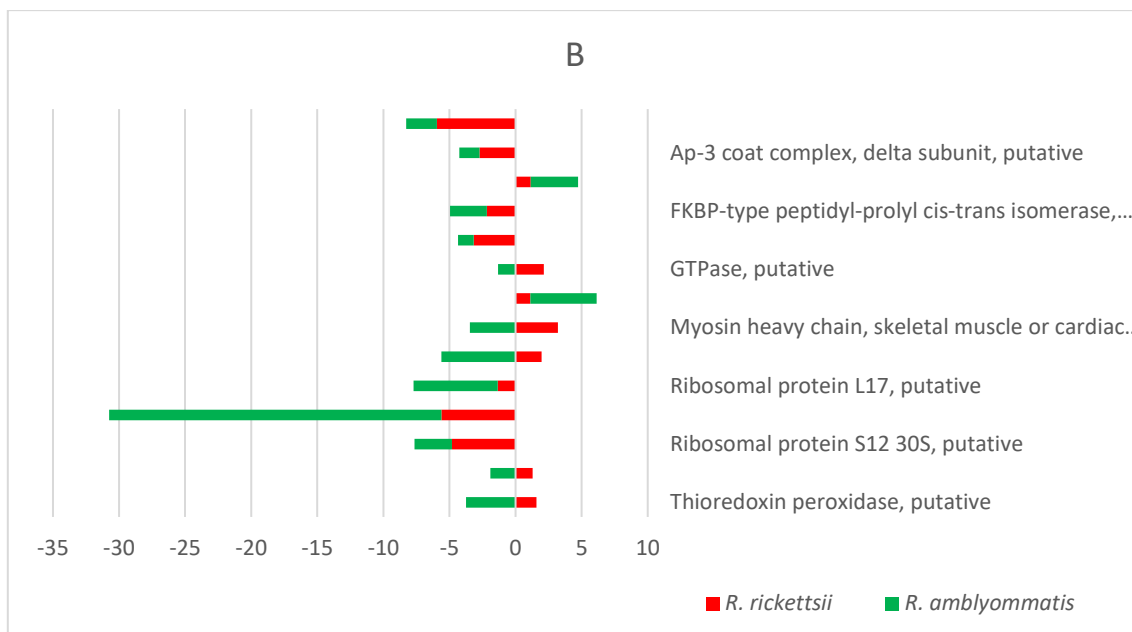
<i>A.sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW008682-PA.1	-3,926362469	Cytochrome C oxidase subunit Va
<i>A.sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW008740-PA.1	-2,424242727	Cytochrome C, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW001817-PA.1	5,055267702	Dimethylglycine dehydrogenase, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW009643-PA.1	4,337082471	Mitochondrial processing peptidase alpha subunit, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW003299-PA.1	-3,87119335	NADH-ubiquinone reductase, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Oxidação	ISCW001465-PA.1	-2,386763085	Thioredoxin H2 protein, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW007139-PA.1	-1,102394795	20S proteasome, regulatory subunit alpha type PSMA7/PRE6, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW005987-PA.1	-1,802927806	20S proteasome, regulatory subunit beta, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW006948-PA.1	1,322019697	20S proteasome, regulatory subunit beta, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW005463-PA.1	-1,716543449	26S proteasome alpha 7 subunit
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008741-PA.1	-2,290988466	26S proteasome regulatory complex, ATPase RPT6, putative (Fragment)
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW009180-PA.1	-5,024873731	ATPase, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW007470-PA.1	1,462033189	ATP-dependent Clp-type protease, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW000683-PA.1	3,152642214	ATP-dependent protease PIM1/LON, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW000634-PA.1	1,137799318	Peroxisome biogenesis factor, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008144-PA.1	-4,581327995	Protease, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Proteólise	ISCW008933-PA.1	1,227648532	Transitional endoplasmic reticulum ATPase, putative (Fragment)
<i>A.sculptum</i>	<i>R. rickettsii</i>	Resposta	ISCW000889-PA.1	3,558936528	Dicer-1, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Resposta	ISCW000890-PA.1	-2,491887939	Dicer-1, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW007915-PA.1	-5,961112843	ABC transporter, putative (Fragment)
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW006643-PA.1	-3,1860408	Ap-2 exocyst complex component
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW006433-PA.1	-3,365323833	Clathrin adaptor complex, small subunit, putative (Fragment)
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW015380-PA.1	-3,425245865	Elongation factor, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW000327-PA.1	-3,143862146	Neutral sphingomyelinase (N-smase) activation associated factor FAN, putative (Fragment)
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW005775-PA.1	-7,518784697	PDZ Syntenin, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW000600-PA.1	-3,208290818	PDZ/DHR/GLGF domaincontaining protein
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW005878-PA.1	-5,618188388	Ribosomal protein L2, putative

<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW007478-PA.1	-9,425546208	Ribosomal protein L4, 60S
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW002308-PA.1	-4,804750892	Ribosomal protein S12, 30S putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Secreção	ISCW007119-PA.1	-3,387136311	Signal recognition particle receptor, alpha subunit, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW007159-PA.1	-4,467617581	ADP-ribosylation factor 2
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW009274-PA.1	-5,218356922	ADP-ribosylation factor 3
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW001558-PA.1	-4,056815006	ADP-ribosylation factor, ARF, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW005629-PA.1	-5,101843978	AP-3 complex subunit beta-1, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW002199-PA.1	-4,817111922	Ap-3 complex subunit sigma, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW003614-PA.1	-6,545881944	Cdc42 protein, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW001040-PA.1	-3,789908355	FKBP-type peptidyl-prolyl cis-trans isomerase, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW006071-PA.1	-7,079785638	GTP binding protein Rab-1A
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW005274-PA.1	4,302403298	GTPase Dynamin, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW002525-PA.1	-4,130774024	GTPase Dynamin, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW003929-PA.1	-3,8893431	GTPase RAS-related protein, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW002009-PA.1	2,869648713	GTPase Rho, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW008890-PA.1	-4,898179765	Heat shock protein 70 (HSP70)-interacting protein, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW001027-PA.1	-3,283081847	Lumican, putative
<i>A.sculptum</i>	<i>R. rickettsii</i>	Transporte	ISCW003700-PA.1	-4,704684762	Vesicle coat complex AP-3, delta subunit, putative

Anexo 4 - Gráficos para a anotação dos níveis da expressão de transcritos e potenciais proteínas putativas coexpressas para os módulos significativos do interatoma, fazendo o uso do Tblastn contra as sequências adquiridas por meio da montagem de *novo* para dados de transcriptomas do intestino de espécies de carrapatos *Amblyomma* spp. infectados por *Rickettsia* spp..

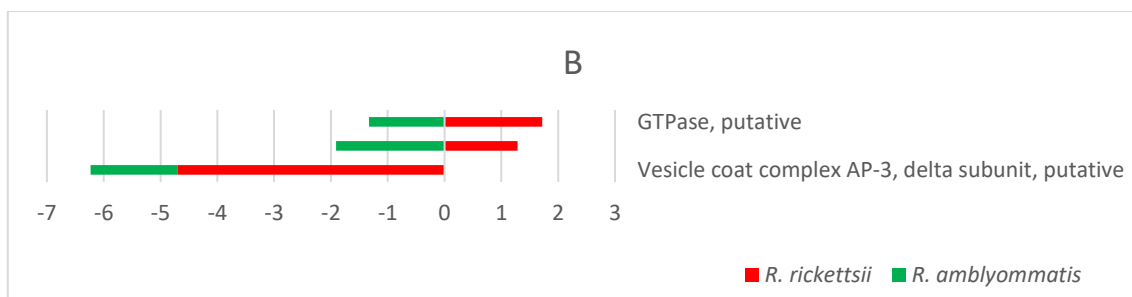
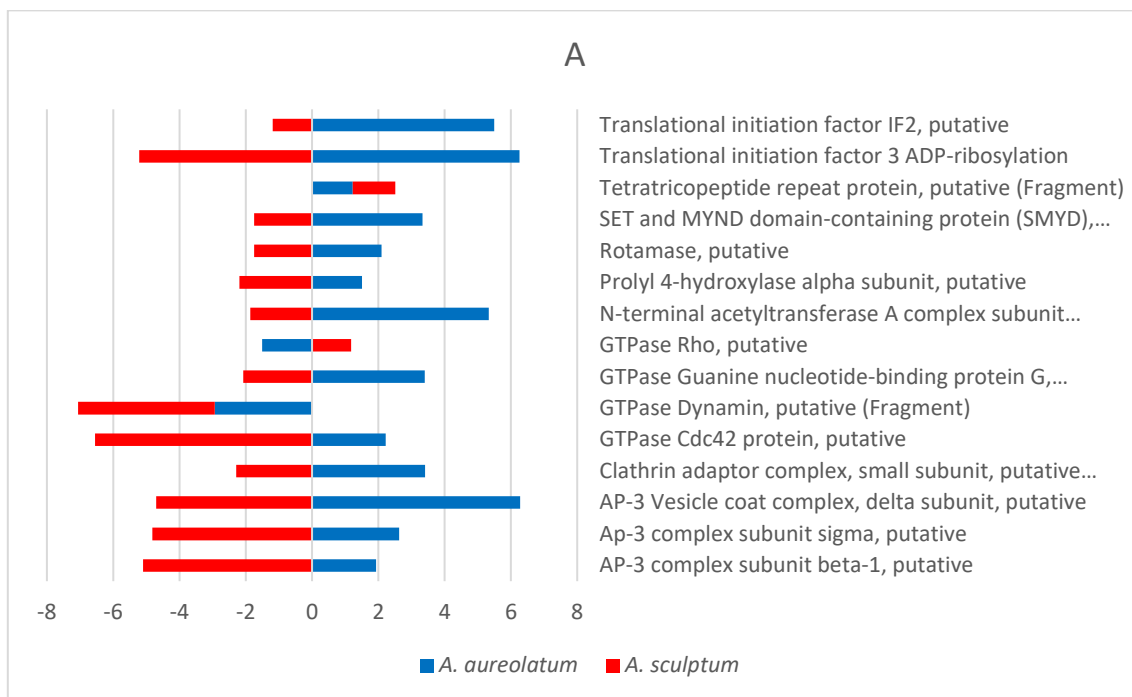
Secreção





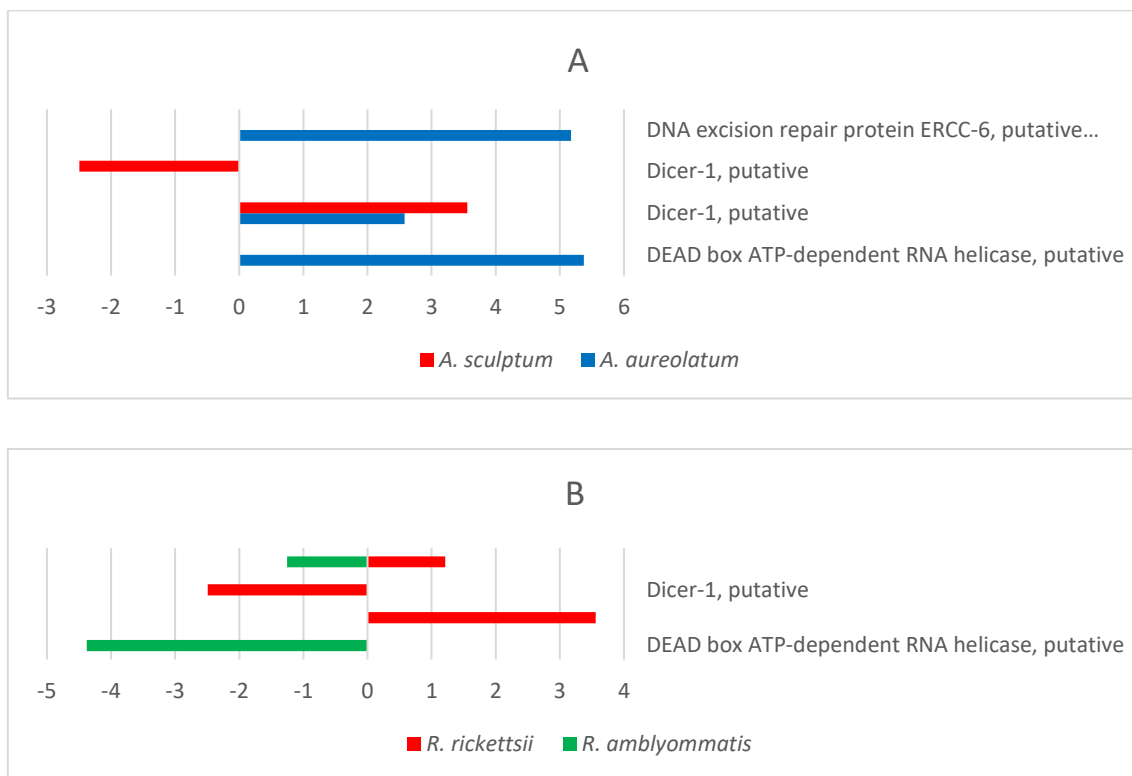
Gráficos destacando os níveis de regulação da expressão de transcritos coexpressos nos intestinos das espécies dos carrapatos, que interagiram com as proteínas relacionadas ao módulo funcional de secreção no interatoma de proteínas de membrana. (A) Níveis de transcritos coexpressos nos intestinos dos carrapatos das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*. (B) Níveis de transcritos coexpressos nos intestinos dos carrapatos das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*.

Transporte



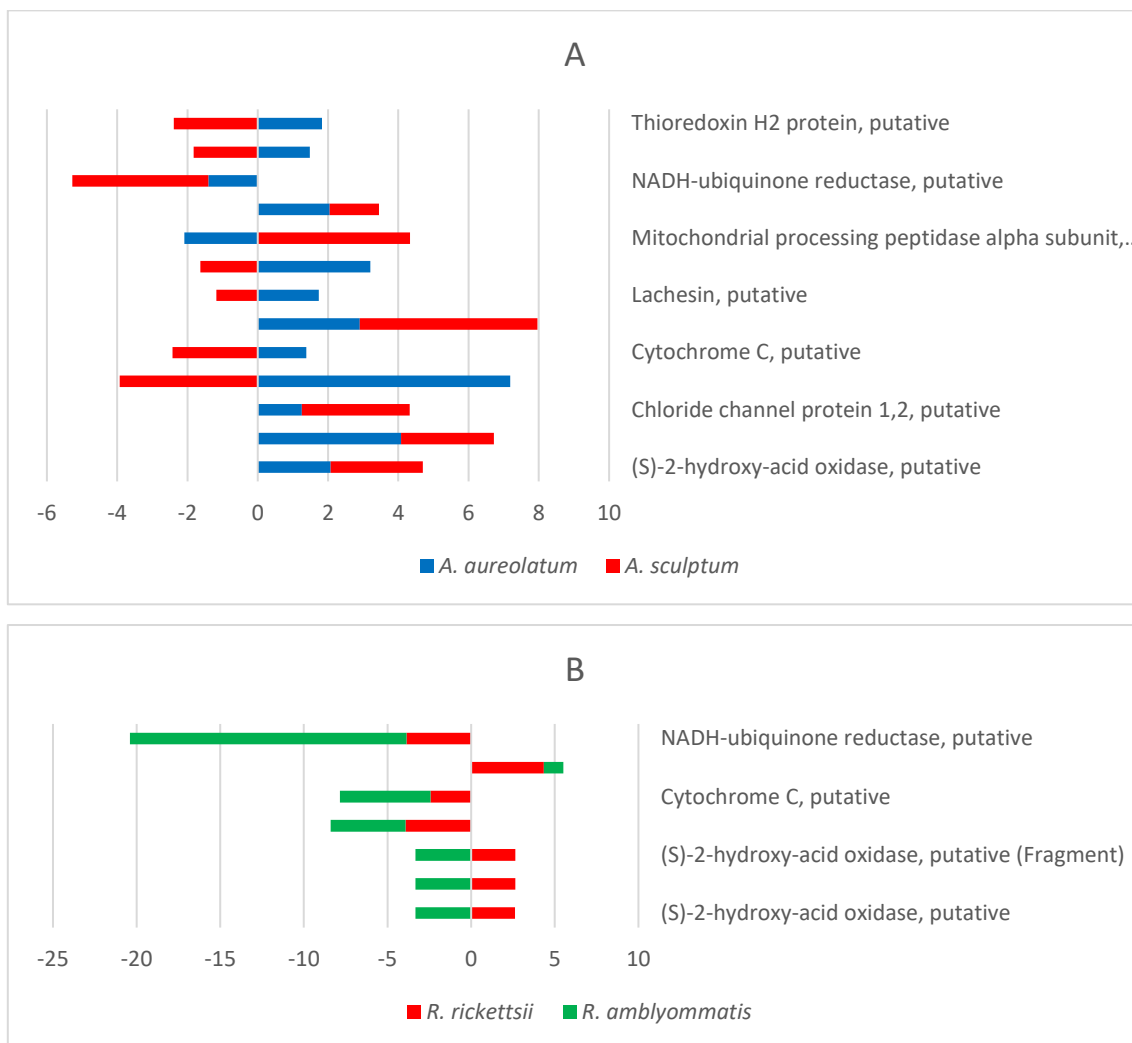
Gráficos destacando os níveis de regulação da expressão de transcritos coexpressos nos intestinos das espécies dos carrapatos, que interagiram com as proteínas relacionadas ao módulo funcional de transporte no interatoma de proteínas ortólogas de membrana. (A) Níveis de regulação da expressão de transcritos coexpressos nas amostras dos intestinos de carrapato da espécie *A. sculptum* infectados separadamente por *R. rickettsii* e *R. amblyommatis*. (B) Níveis de regulação da expressão de transcritos coexpressos nas amostras dos intestinos de carrapato da espécie *A. sculptum* infectados separadamente por *R. rickettsii* e *R. amblyommatis*.

Resposta contra vírus



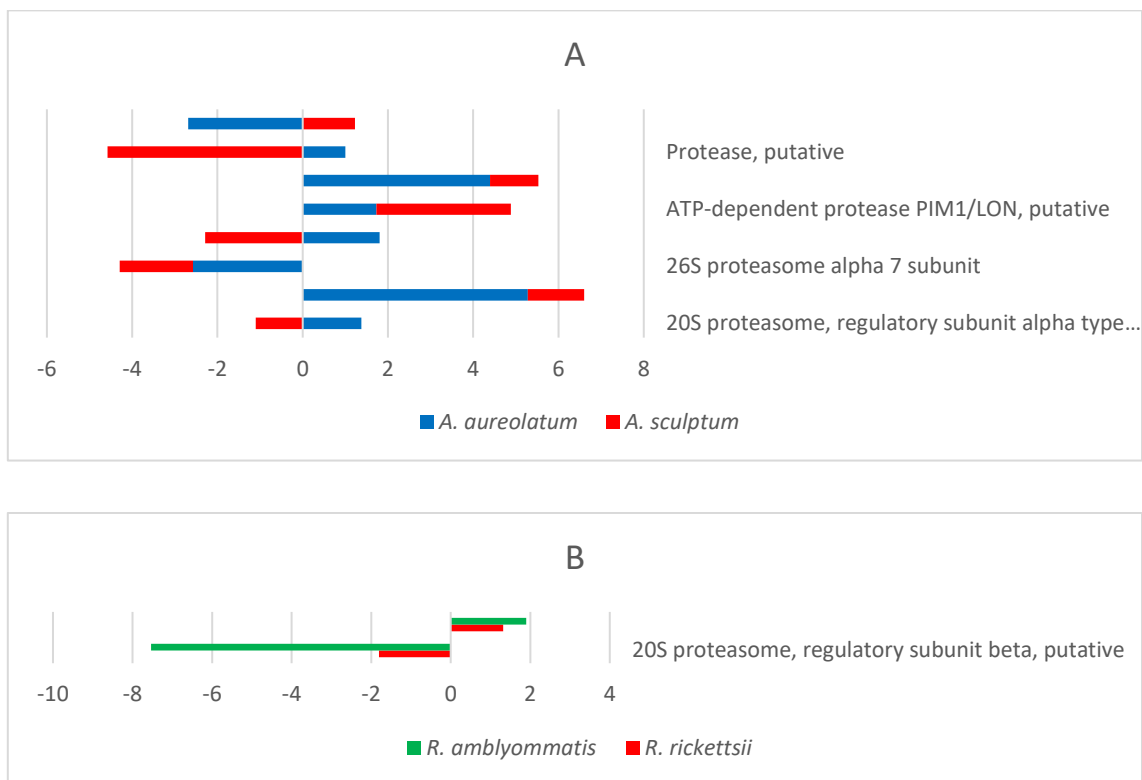
Gráficos destacando os níveis de regulação da expressão de transcritos coexpressos nos intestinos das espécies dos carrapatos, que interagiram com as proteínas relacionadas ao módulo funcional de resposta contra vírus no interatoma de proteínas de membrana. (A) Níveis de transcritos coexpressos nos intestinos dos carrapatos das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*. (B) Níveis de regulação da expressão de transcritos coexpressos nas amostras dos intestinos de carrapato da espécie *A. sculptum* infectados separadamente por *R. rickettsii* e *R. amblyommatis*.

Oxidação



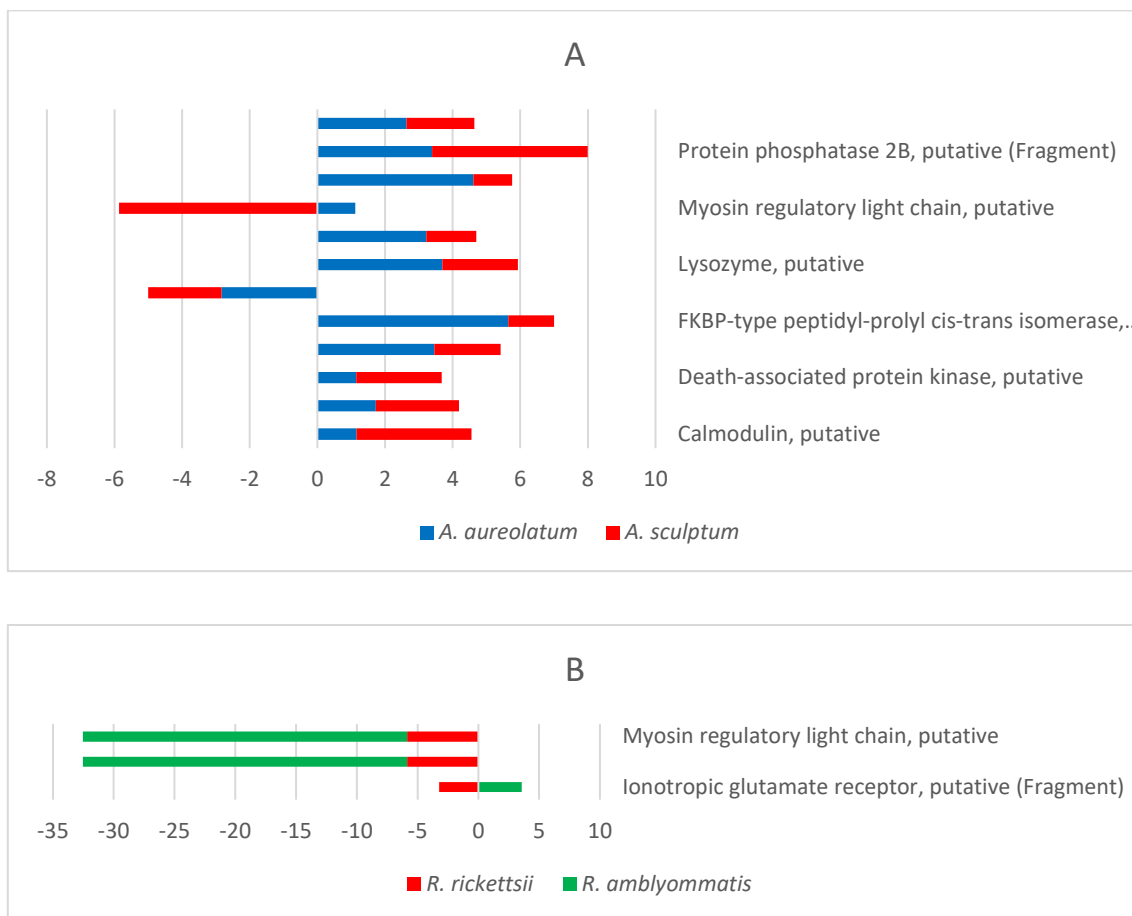
Gráficos destacando os níveis de regulação da expressão de transcritos coexpressos nos intestinos das espécies dos carrapatos, que interagiram com as proteínas relacionadas ao módulo funcional de oxidação no interatoma de proteínas de membrana. (A) Níveis de transcritos coexpressos nos intestinos dos carrapatos das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*. (B) Níveis de regulação da expressão de transcritos coexpressos nas amostras dos intestinos de carrapato da espécie *A. sculptum* infectados separadamente por *R. rickettsii* e *R. amblyommatis*.

Proteólise



Gráficos destacando os níveis de regulação da expressão de transcritos coexpressos nos intestinos das espécies dos carrapatos, que interagiram com as proteínas relacionadas ao módulo funcional de proteólise no interatoma de proteínas de membrana. (A) Níveis de transcritos coexpressos nos intestinos dos carrapatos das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*. (B) Níveis de regulação da expressão de transcritos coexpressos nas amostras dos intestinos de carrapato da espécie *A. sculptum* infectados separadamente por *R. rickettsii* e *R. amblyommatis*.

Fosforilação



Gráficos destacando os níveis de regulação de transcritos coexpressos nas espécies dos carrapatos que interagiram com as proteínas relacionadas ao módulo funcional de fosforilação no interatoma de proteínas de membrana. (A) Níveis de transcritos coexpressos nos intestinos dos carrapatos das espécies *A. aureolatum* e *A. sculptum* infectados por *R. rickettsii*. (B) Níveis de regulação da expressão de transcritos coexpressos nas amostras dos intestinos de carrapato da espécie *A. sculptum* infectados separadamente por *R. rickettsii* e *R. amblyommatis*.

Anexo 5 – Bactérias do gênero *Rickettsia* relatadas em carrapatos. Em destaque as estirpes isoladas e seus hospedeiros*: (i) Os grifos em amarelo - carrapatos infectados; (ii) e os grifos em verde - carrapatos confirmados como hospedeiros pelo isolamento destas bactérias.

<i>Rickettsia</i> spp.	Estirpe & Localização	Nº de acesso NCBI	Proteínas	Relato & Isolado*	Referências
<i>Rickettsia aeschlimannii</i>	MC16T, Marocos, Africa	CCER00000000.1		<i>Hyalomma marginatum</i> *	(BEATI et al., 1997)
				<i>Rhipicephalus appendiculatus</i>	(PAROLA; PADDOCK; RAOULT, 2005b)
				<i>Hyalomma excavatum</i>	(KLEINERMAN et al., 2013)
				<i>Ixodes ricinus</i>	(PAROLA et al., 2013)
				<i>Amblyomma variegatum</i>	(PAROLA et al., 2013)
<i>Rickettsia monacensis</i>	IrR/Munich	NZ_LN794217.1	1.447	<i>Ixodes ricinus</i> *	(SIMSER et al., 2002)
<i>Rickettsia sibirica</i>	subsp. mongolitimonae HA-91, China, 1991	NZ_AHZB00000000.1	1.175	<i>Hyalomma asiaticum</i> *	(SENTAUSA et al., 2012a)
	subsp. sibirica BJ-90, Beijing, China, 1990	NZ_AHIZ00000000.1	1.217	<i>Dermacentor sinicus</i> *	(SENTAUSA et al., 2012a)
				<i>Ixodes ricinus</i>	(PALOMAR et al., 2012)
<i>Rickettsia africae</i>	ESF-5	NC_012633.1	1.229	<i>Amblyomma variegatum</i> *	(SOCOLOVSKI et al., 2009b)
				<i>Rhipicephalus sanguineus</i>	(YSSOUF et al., 2014)
				<i>Hyalomma excavatum</i>	(KLEINERMAN et al., 2013)
				<i>Rhipicephalus (Boophilus) microplus</i>	(YSSOUF et al., 2014)
				<i>Rhipicephalus appendiculatus</i>	(YSSOUF et al., 2014)
<i>Rickettsia conorii</i>	subsp. conorii Malish 7, Africa do Sul	NC_003103.1	1.227	<i>Rhipicephalus sanguineus</i> *	(SOCOLOVSKI et al., 2009c)
	caspia A-167	NZ_AJUR00000000.1	1.132	<i>Rhipicephalus pumilio</i> *	(SENTAUSA et al., 2012b)
	subsp. israelensis ISTT CDC1	AJVP00000000.1	1.135	<i>Rhipicephalus sanguineus</i> *	(ZEMTSOVA et al., 2010)
	subsp. indica ITTR	NZ_AJHC00000000.1	1.135	<i>Rhipicephalus sanguineus</i> *	(PAROLA; PADDOCK; RAOULT, 2005c)
<i>Rickettsia helvetica</i>	C9P9	NC_016931.1	1.158	<i>Ixodes ricinus</i> *	(PAROLA; PADDOCK; RAOULT, 2005c)
<i>Rickettsia massiliae</i>	MTU5	NC_009900.1	1.118	<i>Rhipicephalus sanguineus</i> *	(CHOCHLAKIS et al., 2012)
	AZT80	NC_016931.1	1.195	<i>Rhipicephalus sanguineus</i> *	(EREMEEVA et al., 2006a)
				<i>Ixodes ricinus</i>	(MARIÉ et al., 2012)
<i>Rickettsia parkeri</i>	Grand Bay	NZ_LAOK00000000.1	1.223	<i>Amblyomma maculatum</i> *	(VARELA-STOKES et al., 2011)
	Tate's Hell	NZ_LAOO00000000.1	1.227	<i>Amblyomma maculatum</i> *	(VARELA-STOKES et al., 2011)
	AT#24	NZ_LAOL00000000.1	1.226	<i>Amblyomma triste</i> *	(MELO et al., 2015)
	"Portsmouth"	NC_017044.1	1.228	?	?
				<i>Dermacentor variabilis</i>	(HUANG et al., 2011)
				<i>Amblyomma americanum</i>	(WRIGHT et al., 2015b)
				<i>Amblyomma maculatum</i>	(WRIGHT et al., 2015a)
<i>Rickettsia rickettsii</i>	Iowa	NC_010263.3	1.386	<i>Dermacentor variabilis</i> *	(COX, 1941)
	Colombia	NC_016908.1	1.350	<i>Humano</i> *	(HIDALGO et al., 2007)
	Brazil	NC_016913.1	1.332	<i>Humano</i> *	(BELL; PICKENS, 1953)
	Sheila Smith, Missoula, MT	NC_009882.1	1.343	<i>Humano</i> *	(BELL; PICKENS, 1953)
	Hino, Oklahoma, 1964	NC_016914.1	1.335	<i>Humano</i> *	un-ref
	Hauke, Clermont, OH, 1986	NC_016911.1	1.340	<i>Humano</i> *	un-ref

	Arizona	NC_016909.1	1.343	<i>Rhipicephalus sanguineus</i> * <i>Amblyomma cajennense</i> <i>Amblyomma americanum</i> <i>Amblyomma parvum</i>	(EREMEEVA et al., 2006b) (GUEDES et al., 2005) (BERRADA et al., 2011) (DZUL-ROSADO et al., 2013)
"Candidatus Rickettsia amblyommii"	Darkwater, EUA	NZ_LAOH00000000.1	1.949	<i>Amblyomma americanum</i> *	un-ref
	Ac/Pa, Panama	NZ_LANR00000000.1	1.821	<i>Amblyomma cajennense</i> *	un-ref
	Ac/37, Rondônia, Brasil	NZ_CP012420.1	2.866	<i>Amblyomma cajennense</i> *	un-ref
	GAT-30v	NC_017028.1	1.378	?	?
				<i>Amblyomma sculptum</i> <i>Amblyomma americanum</i> <i>Amblyomma maculatum</i> <i>Rhipicephalus (Boophilus) microplus</i> <i>Rhipicephalus sanguineus</i> <i>Dermacentor variabilis</i> <i>Ixodes scapularis</i>	(NUNES et al., 2015) (SAYLER et al., 2014) (PAROLA et al., 2013) (PAROLA et al., 2013) (PAROLA et al., 2013) (PAROLA et al., 2013) (TROUT FRYXELL et al., 2015b)
<i>Rickettsia bellii</i>	OSU 95-389, Ohio, EUA	NC_009883.1	1.376	<i>Dermacentor variabilis</i> *	un-ref
	RML Mogi, Brasil	LAOJ00000000	1.671	<i>Amblyomma aureolatum</i> * <i>Ixodes ricinus</i>	(PINTER; LABRUNA, 2006) (DUBSKA et al., 2012)
<i>Rickettsia montanensis</i>	OSU 85-930, Montana, EUA, 1961	NC_017043.1	1.217	<i>Dermacentor variabilis</i> * <i>Dermacentor spp.*</i> <i>Amblyomma americanum</i>	(STROMDAHL et al., 2011) (PAROLA et al., 2013) (MEAD et al., 2010)
<i>Rickettsia rhipicephali</i>	3-7-female6-CWPP	NC_017042.1	1.266	?	?
	HJ#5, Brasil, Mata Atlântica	NZ_CP013133.1	1.213	<i>Haemaphysalis juxtakochi</i> *	(LABRUNA et al., 2007)
	Ect, EUA	NZ_LAOC00000000	1,067	<i>Rhipicephalus sanguineus</i> * <i>Dermacentor variabilis</i>	un-ref (WIKSWO et al., 2008)
<i>Rickettsia buchneri</i>	ISO7	NZ_JFKF00000000.1	1.323	<i>Ixodes scapularis</i> *	(KURTTI et al., 2015)
	<i>Rickettsia endosymbiont Ixodes scapularis</i>	NZ_CM000770.1	1.875	<i>Ixodes scapularis</i> *	(KURTTI et al., 2015)
<i>Rickettsia raoultii</i>	Khabarovsk	NZ_CP010969.1		<i>Dermacentor silvarum</i> * <i>Ixodes ricinus</i>	(MEDIANNIKOV et al., 2008) (PAROLA et al., 2013)
<i>Rickettsia slovaca</i>	13-B	NC_016639.1	1.260	<i>Dermacentor variabilis</i> *	(KILLMASTER et al., 2016)
	D-CWPP		1.261	<i>Dermacentor marginatus</i> * ?	(PAROLA et al., 2009) ?