

Universidade de Ribeirão Preto
Faculdade de Biotecnologia
Doutorado em Biotecnologia

VINÍCIUS ALVES SILVA

ANÁLISE GENÔMICA DE REPETIÇÕES EM TANDEM E MOTIVOS CpG
EM BACTÉRIAS DA FAMÍLIA *ANAPLASMATACEAE*

RIBEIRÃO PRETO - SP
2015

Ficha catalográfica preparada pelo Centro de Processamento Técnico da Biblioteca
Central da UNAERP

- Universidade de Ribeirão Preto -

S586a Silva, Vinícius Alves, 1981-
Análise genômica de repetições em tandem e motivos CpG em
bactérias da família *anaplasmataceae* / Vinícius Alves Silva. - -
Ribeirão Preto, 2015.
85 f.: il. color.

Orientador: Prof. Prof. Dr. Mozart de Azevedo Marins.

Tese (doutorado) - Universidade de Ribeirão Preto, UNAERP,
Biotecnologia. Ribeirão Preto, 2015.

1. Biotecnologia. 2. *Anaplasmataceae*. 3. *Tandem repeats*.
4. Genômica - Análise. I. Título.

CDD

VINÍCIUS ALVES SILVA

ANÁLISE GENÔMICA DE REPETIÇÕES EM BACTÉRIAS DA FAMÍLIA
ANAPLASMATACEAE

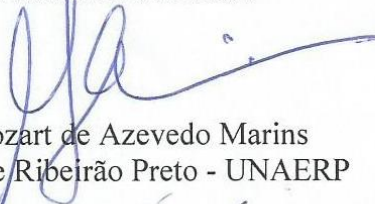
Tese de Doutorado apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade de Ribeirão Preto, para obtenção do título de Doutor em Biotecnologia.

Área de Concentração: Biotecnologia Aplicada à Saúde

Data da defesa: 18 de março de 2015

Resultado: Aprovado

BANCA EXAMINADORA



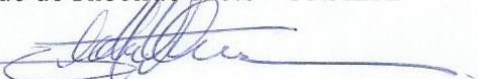
Prof. Dr. Mozart de Azevedo Marins
Universidade de Ribeirão Preto - UNAERP



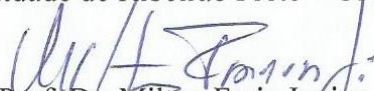
Prof. Dra. Jaqueline Carlos Funayama
Instituto Federal do Sul de Minas



Prof. Dra. Ana Lúcia Fachin Saltoratto
Universidade de Ribeirão Preto - UNAERP



Prof. Dr. Eduardo Carneiro Clímaco
Universidade de Ribeirão Preto - UNAERP



Prof. Dr. Milton Faria Junior
Universidade de Ribeirão Preto - UNAERP

RIBEIRÃO PRETO - SP
2015

ANÁLISE GENÔMICA DE REPETIÇÕES EM TANDEM E MOTIVOS CpG
EM BACTÉRIAS DA FAMÍLIA *ANAPLASMATACEAE*

Tese apresentada à Universidade de Ribeirão Preto, UNAERP, como requisito parcial para obtenção do título de Doutor em Biotecnologia.

Orientador: Prof. Dr. Mozart de Azevedo Marins

VINÍCIUS ALVES SILVA

ANÁLISE GENÔMICA DE REPETIÇÕES EM TANDEM E MOTIVOS CpG EM
BACTÉRIAS DA FAMÍLIA *ANAPLASMATACEAE*

Tese apresentada à Universidade de Ribeirão Preto, UNAERP, como requisito parcial para obtenção do título de Doutor em Biotecnologia.

Aprovação: __/__/__

Prof. Dr. Mozart de Azevedo Marins
Universidade de Ribeirão Preto
Orientador

Dedico este trabalho aos meus pais João e Maria das Graças pelo amor, motivação e apoio que sempre me deram no curso de toda a minha vida.

Obrigado, sem vocês isto não seria possível.

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Mozart de Azevedo Marins, pelas diretrizes de pesquisa, pelo incentivo, amizade e paciência durante a realização deste trabalho.

À DEUS pela saúde, força e proteção.

À Patrícia, pelo carinho, paciência e compreensão pelos momentos em que estive ausente.

À minha família que sempre me incentivou e me apoiou.

Ao Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - IFSULDEMINAS - pelo incentivo a mim concedido.

Aos professores, funcionários e colegas de pós-graduação (mestrado e doutorado) da Universidade de Ribeirão Preto - UNAERP.

À UNAERP.

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”

(Arthur Schopenhauer)

RESUMO

A identificação de casos de infecção de animais domésticos, selvagens, hospedeiros invertebrados e seres humanos por bactérias da família *Anaplasmataceae* cresce no Brasil e no mundo juntamente com a necessidade de criação de vacinas e métodos diagnósticos. Os carrapatos são os principais vetores e, após estas bactérias serem transmitidas a mamíferos, secretam proteínas efectoras a fim de subverter os processos celulares do hospedeiro e tornar a célula em um ambiente propício para replicação e desenvolvimento. Com os avanços de métodos e tecnologias de biologia molecular e pesquisas já realizadas, os genomas de várias espécies da família foram sequenciados e alguns genes associados à virulência. Mesmo diante de tais avanços, existem muitos genes com função desconhecida e questões que precisam ser melhor compreendidas, como a variabilidade genética e antigênica das bactérias, as interações patógeno-hospedeiro, os processos de proteção imunológica e antígenos que estimulem esses processos. Neste contexto, o objetivo do presente trabalho é realizar *in silico* as análises de Repetições adjacentes (*Tandem repeats*) e Motivos CpG e evidenciar alvos genômicos de bactérias da família *Anaplasmataceae* que possam estar relacionados com a variabilidade genética, interação patógeno-hospedeiro e ativação do sistema imunológico. Repetições adjacentes têm sido descritas como um importante mecanismo relacionado com a variabilidade e adaptação bacteriana, sendo também uma característica encontrada em genes relacionados com a patogenicidade. Motivos CpG têm sido eficientes adjuvantes na criação de vacinas que estimulam respostas imunes do hospedeiro. Para o trabalho, foram utilizados e desenvolvidos softwares buscando executar computacionalmente análises de repetições adjacentes e Motivos CpG em arquivos contendo dados do genoma das bactérias *Ehrlichia canis*, *Ehrlichia chaffeensis* e *Anaplasma phagocytophilum*. Os dados resultantes das análises foram armazenados em um banco de dados e correlacionados com informações existentes na literatura e com o resultado de ferramentas de busca em bancos de dados genômicos públicos. Espera-se que o trabalho gere resultados que auxiliem na caracterização de genes e que contribua, em um segundo momento, para o desenvolvimento de novas drogas antibacterianas e métodos diagnósticos.

PALAVRAS-CHAVE: *Anaplasmataceae*; *tandem repeats*; Motivos CpG.

ABSTRACT

The identification of cases of infection of domestic and wildlife animals, invertebrate hosts and humans by bacteria from Anaplasmataceae family grows in Brazil and in the world along with the need to develop vaccines and diagnostic methods. Ticks are the main vectors, and, after being transmitted to mammals, these bacteria secrete effector proteins to subvert the cellular processes of the host cell and render the cell in a conducive environment for replication and development. With the advances of molecular biology methods and technologies and research already conducted, the genome of several bacteria of the family were sequenced and some genes associated with virulence. Even with these advances, there are many genes with unknown function and issues that need to be better understood as the genetic and antigenic variability of bacteria, host-pathogen interactions, immune protection processes and antigens that stimulate these processes. In this context, the aim of this study is to perform *in silico* analyzes of Tandem Repeats and CpG Motifs to highlight genomic targets in bacteria Anaplasmataceae family that may be related to genetic variability, host-pathogen interaction and activation of the immune system through analyzes of Tandem Repeats and CpG motifs. Tandem Repeats has been described as an important mechanism related to bacterial adaptation and variation, and it is also a feature found in the pathogenicity-related genes. CpG motifs are efficient adjuvants in the creation of vaccines which stimulate host immune responses. For the study, softwares were used and developed to perform computationally analyzes of Tandem Repeats and CpG motifs in files containing the genome data of the bacteria *E. canis*, *E. chaffeensis* and *A. phagocytophilum*. The data resulting from this analysis were stored in a database and correlated with existing information in the literature and with the results of the genomic database search tools. It is expected that the work generate results that assist in the characterization of genes and contribute, in a second stage, to the development of new antibacterial drugs and diagnostic methods.

KEYWORDS: *Anaplasmataceae*; *tandem repeats*; *CpG Motifs*.

LISTA DE ILUSTRAÇÕES

Figura 1– Ilustração do ciclo de vida de <i>Ehrlichia spp.</i> e <i>Anaplasma spp.</i>	19
Figura 2 - Diferentes tamanhos dos <i>repeats</i> utilizados para classificação dos <i>TRs</i>	27
Figura 3 - Categorização do <i>TR</i> em relação à conservação dos <i>repeats</i>	28
Figura 4 - Mecanismos de expansão/contração de <i>TRs</i>	30
Figura 5 - Alteração na expressão genética mediada por <i>TRs</i> não-codificantes	31
Figura 6 - Possíveis posições de <i>TRs</i> não-codificantes na região de promotores.....	32
Figura 7 - Variações em <i>TRs</i> codificantes	33
Figura 8 - Variação no <i>TR</i> -codificante do gene <i>FLO1</i> altera a capacidade de adesão.	34
Figura 9 - Árvore filogenética baseada em sequências de aminoácidos do gene <i>TRP36</i> de isolados de <i>E. canis</i> de diferentes regiões do Brasil e do mundo.....	35
Figura 10 - Arranjo/regra para identificação de Motivos CpG no genoma bacteriano	39
Figura 11 - Forma de trabalho do sistema. Modelo MVC.....	44
Figura 12 - Tela principal do <i>CpG Highlight</i> . O usuário deve inicialmente carregar (botão "Load") o arquivo FASTA com o genoma da bactéria.	45
Figura 13- Tela com os Motivos CpG identificados e os botões para geração dos gráficos. .	46
Figura 14 - Tabelas com a relação entre os Motivos CpG e as regiões codificantes.	46
Figura 15- Fluxograma dos métodos aplicados na análise de <i>TRs</i>	48
Figura 16 - Modelo Entidade-Relacionamento do banco de dados.....	50
Figura 17 - Gráfico de Motivos CpG encontrados no genoma da bactéria <i>E. canis</i>	52
Figura 18 - Dispersão de Motivos CpG no genoma da bactéria <i>E. canis</i>	52
Figura 19 - Gráfico de Motivos CpG encontrados no genoma da bactéria <i>E. chaffeensis</i>	54
Figura 20 - Dispersão de Motivos CpG no genoma da bactéria <i>E. chaffeensis</i>	54
Figura 21 - Gráfico de Motivos CpG encontrados no genoma da bactéria <i>A. phagocytophilum</i>	56
Figura 22 - Dispersão de Motivos CpG no genoma da bactéria <i>A. phagocytophilum</i>	56
Figura 23 - Dispersão dos <i>TRs</i> ao longo do genoma das bactérias.....	60
Figura 24 - Dispersão dos <i>TRs</i> ao longo do genoma das bactérias. Parâmetros: VARSCORE > 1	60
Figura 25 – Sequência de DNA do genoma de <i>E. canis</i> (Linhagem Jake, GenBank: CP000107.1) correspondendo à região englobando o <i>TR</i> 468. As regiões dos <i>primers</i>	

REP468fwd-1 e REP468rev-1, englobando região de amplificação com 1050 pbs, está em negrito. As repetições estão sublinhadas.	66
Figura 26 – Foto representativa de eletroforese em gel de agarose de produtos de PCR com primers REP468fwd-1 e REP468rev-1 para amplificação de fragmento de 1050 pbs englobando o <i>TR</i> 468. Canaleta 1 – Marcador de tamanho de DNA 1 Kb (Ludwig); Canaletas 2-7 – Produtos de amplificação com DNA genômico de amostras de sangue canino confirmadas como positivas para infecção com <i>E. canis</i> . Canaleta 8: controle negativo (água ultra-pura).	66
Figura 27- Sequência de DNA do genoma de <i>E. canis</i> (Linhagem <i>Jake</i> , GenBank: CP000107.1) correspondendo à região englobando o <i>TR</i> 468 e a sequência de DNA do resultado de PCR da amplificação da região do <i>TR</i> 468 de uma amostra da cidade de Monte Negro, estado de Rondônia, Brasil.	67
Figura 28 - Padrão dos Motivos CpG identificados para os gêneros <i>Anaplasma</i> e <i>Ehrlichia</i> . 70	
Figura 29 - Esquema geral sobre a localização da sequencia de observação de <i>TRs</i> em uma região	75
Figura 30 - Esquema geral sobre a localização da sequencia de observação de <i>TRs</i> em uma região	75

LISTA DE TABELAS

Tabela 1- Classificação da ordem <i>Rickettsiales</i>	17
Tabela 2 - Sequência de nucleotídeos dos “ <i>primers</i> ” de PCR para amplificação da repetição <i>TR</i> -468	50
Tabela 3 - Comparação das propriedades genômicas das 3 bactérias.	51
Tabela 4 - Relação de Motivos CpG da bactéria <i>E. canis</i>	53
Tabela 5 - Relação de Motivos CpG da bactéria <i>E. chaffeensis</i>	55
Tabela 6 - Relação de Motivos CpG da bactéria <i>A. phagocytophilum</i>	57
Tabela 7- <i>TRs</i> em destaque encontrados no genoma da <i>E. canis</i>	58
Tabela 8 - <i>TRs</i> x <i>TRs</i> codificantes	59
Tabela 9 - <i>TRs</i> da bactéria <i>E. canis</i> com VARSCORE maior que 1.	61
Tabela 10 - <i>TRs</i> codificantes da bactéria <i>E. canis</i> com VARSCORE maior que 1.....	61
Tabela 11 - <i>TRs</i> codificantes da bactéria <i>E. chaffeensis</i> com VARSCORE maior que 1.	61
Tabela 12 - <i>TRs</i> codificantes da bactéria <i>A. phagocytophilum</i> com VARSCORE maior que 1.	62
Tabela 13 - <i>TRs</i> codificantes da bactéria <i>E. canis</i> com VARSCORE menor que 1 e TRFSCORE maior que 1000.....	62
Tabela 14 - <i>TRs</i> codificantes da bactéria <i>E. chaffeensis</i> com VARSCORE menor que 1 e TRFSCORE maior que 1000.....	63
Tabela 15 - <i>TRs</i> codificantes da bactéria <i>A. phagocytophilum</i> com VARSCORE menor que 1 e TRFSCORE maior que 1000.	63
Tabela 16 - <i>TRs</i> não-codificantes da bactéria <i>E. chaffeensis</i> com VARSCORE maior que 1...63	
Tabela 17 - <i>TRs</i> não-codificantes da bactéria <i>A. phagocytophilum</i> com VARSCORE maior que 1.	64
Tabela 18 - <i>TRs</i> não-codificantes da bactéria <i>E. canis</i> com VARSCORE menor que 1 e TRFSCORE maior que 1000.....	64
Tabela 19 - <i>TRs</i> não-codificantes da bactéria <i>E. chaffeensis</i> com VARSCORE menor que 1 e TRFSCORE maior que 1000.....	64
Tabela 20 - <i>TRs</i> não-codificantes da bactéria <i>A. phagocytophilum</i> com VARSCORE menor que 1 e TRFSCORE maior que 1000.	65
Tabela 21 - Motivos CpG destacados como potentes agonistas de TLR-9. Foram destacadas as referências, os hospedeiros e a posição no genoma das bactérias pesquisadas (em relação ao número de ocorrências no genoma).....	68

LISTA DE ABREVIATURAS E SIGLAS

AGC	Anaplasrose Granulocítica Canina
AGH	Anaplasrose Granulocítica Humana
ATP	Trifosfato de Adenosina
CDD	Banco de Dados de Domínio Conservado
DUF	<i>Protein of Unknown function</i>
EMC	Erliquiose Monocítica Canina
EMH	Erliquiose Monocítica Humana
NCBI	<i>National Center for Biotechnology Information</i>
NC	Nucleóide Condensado
ND	Nucleóide Descondensado
PCR	Reação em Cadeia da Polimerase
T4SS	Sistema de Secreção Tipo IV
TLR-9	Receptores <i>Toll-Like 9</i>
<i>TR</i>	<i>Tandem Repeat</i>
<i>TRs</i>	<i>Tandem Repeats</i>
TRFSCORE	índice do <i>TR</i> levando-se em consideração o tamanho, número de <i>repeats</i> e conservação dos <i>repeats</i>
VARSCORE	índice que indica o grau variabilidade do <i>TR</i> entre organismos de mesma espécie

SUMÁRIO

1 INTRODUÇÃO.....	15
2 REVISÃO DE LITERATURA	17
2.1 FAMÍLIA ANAPLASMATACEAE.....	17
2.1.1 CICLO DE VIDA E REPLICAÇÃO	18
2.1.2 MECANISMOS BACTERIANOS E A SUBVERSÃO DOS PROCESSOS CELULARES DO HOSPEDEIRO.....	18
2.1.2.1 SISTEMA DE SECREÇÃO TIPO IV.....	20
2.1.2.2 PROTEÍNAS DE SUPERFÍCIE	20
2.1.2.3 REGULAÇÃO DA EXPRESSÃO DE GENES DO HOSPEDEIRO.....	21
2.1.2.4 INIBIÇÃO DE APOPTOSE.....	21
2.1.2.5 SUBVERSÃO DA AUTOFAGIA	22
2.1.2.6 SEQUESTRO DE COLESTEROL	22
2.1.3 BACTÉRIAS ANALISADAS	22
2.1.3.1 EHRLICHIA CANIS.....	22
2.1.3.2 EHRLICHIA CHAFFEENSIS	24
2.1.3.3 ANAPLASMA PHAGOCYTOPHILUM	26
2.2 TANDEM REPEATS (TRS).....	27
2.2.1 IDENTIFICAÇÃO IN SILICO E LOCALIZAÇÃO DE TRS NO GENOMA	28
2.2.2 VARIABILIDADE DE TRS	29
2.2.3 VARIABILIDADE EM TRS NÃO-CODIFICANTES	31
2.2.4 VARIABILIDADE EM TRS CODIFICANTES	32
2.2.5 ALGUNS TRABALHOS ENVOLVENDO TRS E ANAPLASMATACEAE.....	34
2.3 MOTIVOS CPG	38
3 HIPÓTESE	41
4 OBJETIVOS GERAL E ESPECÍFICOS	42
5 MATERIAIS E MÉTODOS.....	43

5.1 MATERIAIS UTILIZADOS.....	43
5.2 MÉTODOS UTILIZADOS NA ANÁLISE DE MOTIVOS CPG.....	43
5.2.1 CPG HIGHLIGHT	43
5.2.1.1 ALGORITMOS	47
5.3 MÉTODOS UTILIZADOS NA ANÁLISE DE TRS	47
5.3.1 ANÁLISE LABORATORIAL DA VARIAÇÃO DE TRS	50
6 RESULTADOS	51
6.1 ANÁLISE DE MOTIVOS CPG.....	51
6.2 ANÁLISE DE TRS	57
6.2.1 TRS CODIFICANTES COM VARSCORE >1	61
6.2.2 TRS CODIFICANTES COM VARSCORE < 1 E COM TRFSCORE > 1000	62
6.2.3 TRS NÃO-CODIFICANTES COM VARSCORE >=1	63
6.2.4 TRS NÃO-CODIFICANTES COM VARSCORE < 1 E TRFSCORE>1000	64
6.2.5 TRS SIMILARES ENTRE AS TRÊS BACTÉRIAS.....	65
6.2.6 ANÁLISE IN VITRO DE VARIABILIDADE UTILIZANDO O TR 468 DE E. CANIS..	65
7 DISCUSSÃO.....	68
7.1 MOTIVOS CPG	68
7.1.1 MOTIVOS CPG DE DESTAQUE IDENTIFICADOS	69
7.1.2 GENES COM MOTIVOS CPG	70
7.2 TANDEM REPEATS.....	72
7.2.1 TRS CODIFICANTES COM VARSCORE MAIOR QUE 1	72
7.2.2 TRS CODIFICANTES COM VARSCORE < 1 E COM TRFSCORE > 1000	73
7.2.3 TRS NÃO-CODIFICANTES COM VARSCORE >=1	74
7.2.4 TRS NÃO-CODIFICANTES COM VARSCORE < 1 E TRFSCORE>1000	76
7.2.5 ANÁLISE IN VITRO DE VARIABILIDADE UTILIZANDO O TR 468 DE E. CANIS..	76
8 CONCLUSÃO.....	77
REFERÊNCIAS	78

1 INTRODUÇÃO

A família *Anaplasmataceae* é composta por gêneros de bactérias que se caracterizam por infectar animais domésticos, selvagens, hospedeiros invertebrados e seres humanos, causando infecção considerada uma zoonose emergente em diversas regiões. Os carrapatos são os principais vetores de transmissão e contribuem para a distribuição destas bactérias em zonas urbanas, rurais e silvestres. A identificação de indivíduos infectados no Brasil e no mundo tem aumentado e, por este motivo, diversas espécies da família têm tido destaque para a medicina veterinária e humana.

Com as técnicas de sequenciamento e o avanço de métodos e tecnologias de biologia molecular e bioinformática, alguns genes foram associados à virulência e a mecanismos de variabilidade entre amostras de uma mesma espécie. Mesmo diante de tais avanços e com o sequenciamento do genoma de diversos exemplares, existem vários genes com função desconhecida e questões que precisam ser melhor compreendidas para o desenvolvimento de vacinas e métodos de diagnósticos, como a variabilidade genética e antigênica das bactérias, as interações patógeno-hospedeiro, a proteção imunológica do hospedeiro e antígenos que estimulem os processos de proteção imune. Neste sentido, a utilização da bioinformática pode ajudar na compreensão dessas questões utilizando e correlacionando informações genômicas das bactérias com informações existentes na literatura para auxiliar a identificação de alvos a serem investigados visando o desenvolvimento de drogas, vacinas, marcadores moleculares, marcadores epidemiológicos e ferramentas de diagnóstico. Dentre as diversas técnicas de análise genômica existentes, as análises de repetições adjacentes e a de motivos CpG têm tido bastante destaque na literatura.

Repetições adjacentes são sequências de DNA que se repetem como múltiplas cópias adjacentes com a mesma orientação na cadeia de DNA. Diversos trabalhos têm associado as repetições adjacentes como importante característica de genes relacionados com a virulência, interação patógeno-hospedeiro e como mecanismo responsável por gerar diversidade entre as espécies.

Motivos CpG são dinucleotídeos CG não metilados juntamente com bases flangeadoras do genoma da bactéria utilizados para induzir respostas celulares do sistema imunológico de mamíferos hospedeiros, produzindo uma resposta imune quando o DNA da bactéria é identificado.

Neste trabalho, foram aplicadas as análises de Repetições Adjacentes e Motivos CpG nos dados genômicos de três espécies da família: *Ehrlichia canis* e *Ehrlichia chaffeensis* (gênero *Ehrlichia*) e *Anaplasma phagocytophilum* (gênero *Anaplasma*).

2 REVISÃO DE LITERATURA

2.1 FAMÍLIA ANAPLASMATACEAE

As bactérias Gram-negativas da família *Anaplasmataceae*, ordem *Rickettsiales*, reproduzem-se como parasitas obrigatórios no interior de vacúolos de células hematopoiéticas de mamíferos, especialmente macrófagos e monócitos (RIKIHISA, 1991; DUMLER et al., 2001; BOWMAN, 2011). A família compreende gêneros de bactérias capazes de infectar animais domésticos, selvagens, seres humanos e hospedeiros invertebrados, como artrópodes e helmintos (DUMLER et al., 2001). Os carrapatos são os principais vetores de transmissão e contribuem para a distribuição destas bactérias em zonas urbanas, rurais e silvestres (PAROLA et al., 2005, 2013; DANTAS-TORRES, 2008, 2010). Dentre as espécies da família, conforme Tabela 1 proposta por Dumler et al.(2001) baseada em semelhanças do gene RNA ribossômico 16S (16S rRNA), três possuem destaque para a medicina veterinária e humana. Estas espécies são a *E. canis*, agente etiológico da Erliquiose Monocítica Canina (EMC) (SKOTARCZAK, 2003), *E. chaffeensis*, associada à Erliquiose Monocítica Humana (EMH) (PADDOCK, 2003; RIKIHISA, 2010a) e *A. phagocytophilum*, que causa a Anaplasmoze Granulocítica Humana (AGH) e Canina (AGC) (RIKIHISA, 2010a, 2011). As citadas bactérias serão abordadas na análise genômica *in silico* realizada no trabalho.

Tabela 1- Classificação da ordem *Rickettsiales*.

Ordem	Família	Gênero	Espécie	
<i>Rickettsiales</i>	<i>Anaplasmataceae</i>	<i>Anaplasma</i>	<i>A. marginale</i>	
			<i>A. centrale</i>	
			<i>A. ovis</i>	
			<i>A. bovis</i>	
			<i>A. platys</i>	
			<i>A. phagocytophilum</i>	
		<i>Ehrlichia</i>	<i>E. ruminantium</i>	
			<i>E. canis</i>	
			<i>E. chaffeensis</i>	
			<i>E. muris</i>	
			<i>E. ewingii</i>	
			<i>Neorickettsia</i>	<i>N. helminthoeca</i>
				<i>N. risticii</i>
<i>N. sennetsu</i>				
<i>Wolbachia</i>	<i>W. pipientis</i>			

Fonte: Adaptado de DUMLER et al.(2001).

2.1.1 Ciclo de Vida e Replicação

O ciclo de vida das bactérias da família *Anaplasmataceae* envolve dois hospedeiros: carrapatos e mamíferos. Animais selvagens como ratos e veados são reservatórios primários destas bactérias, que são transmitidas para animais domésticos e seres humanos por carrapatos infectados (HAJDUSEK et al., 2013).

Uma vez transmitida a mamíferos, estas bactérias se replicam em granulócitos e macrófagos ou monócitos, subvertendo processos celulares e adquirindo compostos do hospedeiro para sua replicação e sobrevivência (RIKIHISA, 2010a, 2010b, 2011).

Segundo Zhang (2007) e Rikihisa (2010a), o desenvolvimento na célula de mamíferos envolve dois ciclos morfológicos: Nucleóide Condensado (NC) e Nucleóide Descondensado (ND), como pode ser visto na Figura 1.

Inicialmente, pequenas células na forma NC aderem à célula hospedeira e são internalizadas por fagocitose, formando-se um vacúolo. No fagossomo, NC passa por uma fase intermediária e transforma-se na forma ND, que se multiplica por fissão binária durante aproximadamente 48 horas. Ocorre novamente uma fase intermediária até as células maturarem para a forma NC (após 72 horas de infecção) e serem liberadas e internalizadas por outras células.

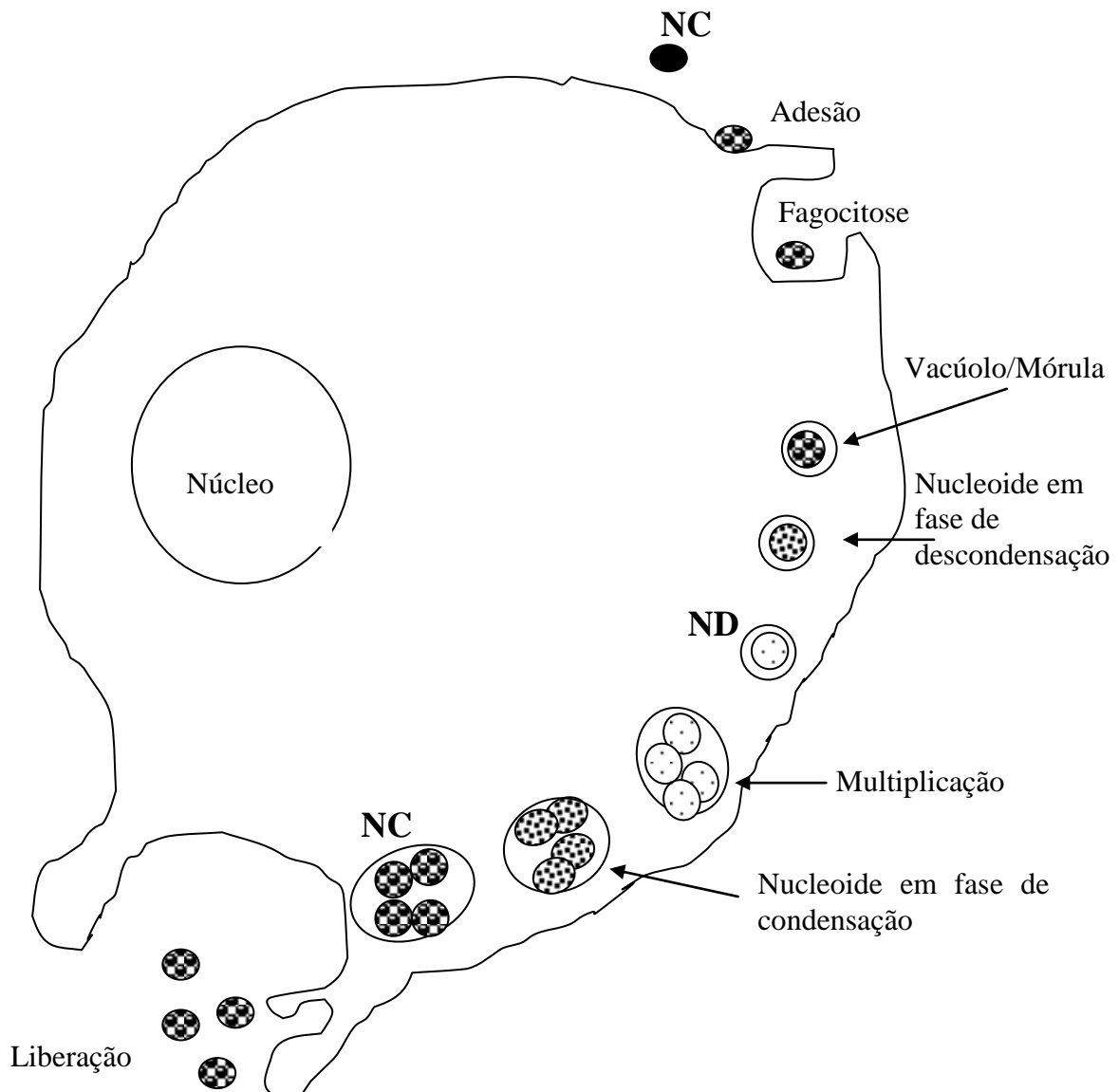
2.1.2 Mecanismos Bacterianos e a Subversão dos Processos Celulares do Hospedeiro

O processo de maturação de vacúolos celulares contendo microrganismos parasitas envolve normalmente sua fusão com lisossomos, os quais carregam enzimas digestivas que destroem os microrganismos contidos nos vacúolos (ZHANG, 2007). A sobrevivência de endobactérias depende de sua capacidade de resistirem ou evitarem a exposição à estas enzimas, havendo diferentes mecanismos adotados por bactérias para alcançarem este objetivo (ZHANG, 2007). A análise bioquímica e celular do ciclo de vida de bactérias da família *Anaplasmataceae*, como a *E. chaffeensis*, demonstra que ela evoluiu no sentido de inibir na célula hospedeira o processo de fusão dos lisossomos com os vacúolos (RIKIHISA, 2010a). Este processo de inibição envolve a participação de sistemas de secreção e proteínas efetoras que são secretadas para o interior da célula hospedeira onde, além do processo de inibição de fusão de lisossomos, interferem em diversos processos celulares como a ativação e supressão da atividade de quinases, regulação da atividade de receptores Toll, inibição de

apoptose, subversão da autofagia e regulação da expressão de genes envolvidos na defesa imune (RIKIHISA, 2011).

Diversos estudos têm apontado que domínios proteicos constituídos por repetições adjacentes são uma característica de muitas proteínas envolvidas nos processos descritos anteriormente (DOYLE, 2006; RIKIHISA, 2010a, 2011; MCBRIDE; WALKER, 2010; ZHANG, 2008; STUEN et al. 2013) , como por exemplo as proteínas efetoras p47, p120 e p200 de *E. chaffensis*, p36 de *E. canis* e proteínas de secreção de *A. phagocytophilum*. O estudo de tais padrões de repetição no genoma bacteriano pode facilitar a identificação e o entendimento de proteínas envolvidas nos processos de patogenicidade dessas bactérias.

Figura 1– Ilustração do ciclo de vida de *Ehrlichia spp.* e *Anaplasma spp.*



Fonte: Adaptado de Zhang (2007) e Rikihisa (2010a).

2.1.2.1 Sistema de secreção tipo IV

Com o objetivo de secretar proteínas, enzimas, DNA ou toxinas do citoplasma para o exterior da bactéria através da membrana citoplasmática, as bactérias utilizam pelo menos seis sistemas de secreção distintos, denominados sistemas de secreção do tipo I (T1SS) ao tipo VI (T6SS). Proteínas do sistema de secreção do tipo IV (T4SS) foram identificadas em genomas de bactérias da família *Anaplasmataceae* (RIKIHISA, 2010a). O T4SS, que é dependente de ATP, forma um canal composto por um complexo multiproteico, possibilitando o envio de proteínas efetoras e/ou de DNA para o interior da célula hospedeira para influenciar as atividades celulares com o objetivo de favorecer a sobrevivência intracelular bacteriana (RIKIHISA et. al., 2009).

2.1.2.2 Proteínas de superfície

As proteínas da superfície celular bacteriana fornecem uma importante interface para a interação patógeno-hospedeiro e desempenham um papel chave na patogênese e fisiologia das bactérias, sendo os principais alvos para o desenvolvimento de vacinas (GE & RIKIHISA, 2007).

Estudos sobre bactérias da família *Anaplasmataceae* revelaram que proteínas da membrana externa bacteriana, tais como MAP1, MAP2 e OMP-1, estimulam a resposta imune do hospedeiro e o protege contra a infecção. Ensaio bem sucedido com vacinas vivas atenuadas indicam que a imunidade induzida por proteínas imunorreativas é viável (MCBRIDE & WALKER, 2010). Essas proteínas têm sido estudadas e caracterizadas em *E. chaffeensis* e *E. canis* e *A. phagocytophilum*, sendo que uma característica comum entre elas envolve a presença de *tandem repeats*, repetições anquirina e a participação em processos de subversão da célula hospedeira (RIKIHISA 2010a, 2011; MCBRIDE;WALKER, 2010). Apesar de tais avanços, ainda há pouca informação sobre a interação de proteínas imunorreativas e o hospedeiro. Além disso, nem todas estas proteínas foram molecularmente caracterizadas (RIKIHISA, 2011). A caracterização das principais proteínas imunorreativas é um passo importante para combater a patogenicidade dessas bactérias. Outro desafio é compreender a diversidade genética e antigênica das proteínas, que apresentam variabilidade entre isolados de mesma espécie e de diferentes regiões (AGUIAR et al., 2013).

2.1.2.3 Regulação da expressão de genes do hospedeiro

Patógenos bacterianos podem alterar a expressão gênica do hospedeiro por meio de modificações nas histonas e remodelação da cromatina a fim de subverter respostas imunes e processos celulares e tornar a célula em um ambiente propício para a sua sobrevivência e desenvolvimento (STUEN et al., 2013). Proteínas efetoras de bactérias da família *Anaplasmataceae* são secretadas para a célula alvo, atuando na região promotora de genes de interesse. Um exemplo é a inibição da apoptose (explicada a seguir) por *A. phagocytophilum* e *E. chaffeensis* pela alteração da expressão de BFL1, um gene membro da família antiapoptótica (STUEN et al., 2013).

A. phagocytophilum modula a "explosão respiratória" ("*respiratory burst*") de granulócitos por meio da inibição da expressão do gene CYBB. A proteína AnkA se liga a regiões ricas em AT no promotor e diminui a sua transcrição (GARCIA-GARCIA et al., 2009; RENNOLL-BANKERT; DUMLER, 2012). De forma similar, a proteína gp200 de *E. canis* é secretada para o núcleo da célula, onde interage com a cromatina, potencialmente para modular a expressão do gene da célula hospedeira (MAVROMATIS et al., 2006).

Harrus et al. (2003) identificaram que a infecção por *E. canis* altera a expressão de proteínas do complexo de histocompatibilidade ou MHC (do inglês, *Major Histocompatibility Complex*), região mais densa de genes do genoma dos mamíferos que possui importante papel no sistema imunológico.

2.1.2.4 Inibição de apoptose

A apoptose, processo de morte celular programada, não é importante somente para a remoção de células danificadas ou não mais necessárias, mas também como um importante mecanismo imunológico inato celular para eliminar patógenos intracelulares.

A infecção por *A. phagocytophilum* e *E. chaffeensis* inibe a apoptose celular atuando na regulação de genes da célula do hospedeiro (RIKIHISA, 2010a). *A. phagocytophilum* inibe a apoptose de neutrófilos por até 48 a 96 horas, permitindo a reprodução das bactérias (GE; RIKISA, 2006). Dados de *microarray* de expressão de genes de neutrófilos humanos mostram que *A. phagocytophilum* evita a redução da expressão de BFL1, um membro da família antiapoptótica célula B de linfoma 2 (Bcl-2), ou seja, a infecção faz com que a expressão de genes da família Bcl-2 seja aumentada, evitando a atividade de caspases efetoras (RIKISA, 2010b, 2011).

2.1.2.5 Subversão da autofagia

A autofagia é um processo catabólico envolvendo a degradação e reciclagem de componentes celulares por meio da maquinaria lisossômica, controlando também a biomassa intracelular (RIKIHISA, 2011). Este processo desempenha também um importante papel no sistema imunológico inato das células eucarióticas ao sequestrar e digerir objetos intracelulares indesejáveis, como agentes patogênicos intracelulares (NIU et al., 2008). Pesquisas com *A. phagocytophilum* indicam que a bactéria subverte a autofagia para favorecer a infecção e, em um segundo momento, induz a autofagia celular para remodelar o citoplasma da célula hospedeira e disponibilizar espaço para o seu crescimento (NIU et al., 2008; RIKIHISA, 2011). Em *E. canis*, Niu et al. (2012) identificaram que o gene ECH_0825 é altamente expresso no início da infecção, sendo um dos genes que secretam proteínas para o citoplasma da célula hospedeira para atuar na inibição da autofagia.

2.1.2.6 Sequestro de colesterol

Diferentemente dos eucariotos, a maioria das bactérias não sintetiza, adquire ou necessita de colesterol. Bactérias da família *Anaplasmataceae* são dependentes de colesterol para a sobrevivência e crescimento, mas carecem de genes para a biossíntese ou modificação do colesterol (LIN; RIKISA, 2003). *A. phagocytophilum* e *E. chaffeensis* são capazes de incorporar na sua membrana uma quantidade substancial de colesterol do hospedeiro. Estudos mostram que a infecção por estas bactérias é facilitada em ratos com níveis elevados de colesterol no sangue e que a inibição das vias de absorção do colesterol bloqueia a infecção bacteriana (GE; RIKIHISA, 2007; RIKISA, 2010a).

2.1.3 Bactérias Analisadas

2.1.3.1 *Ehrlichia canis*

A bactéria *E. canis* reside como microcolônias dentro de vacúolos revestidos de membrana (mórulas), principalmente dentro de monócitos e macrófagos em hospedeiros mamíferos (McBRIDE et al., 1996). A espécie foi descrita pela primeira vez em 1935 na Argélia, mas atualmente é reconhecida por ter distribuição global e recebeu reconhecimento como um patógeno de importância veterinária principalmente depois de surtos em cães de

militares dos Estados Unidos no Vietnã, em 1963, pois resultou em cerca de 200 mortes (MAVROMATIS et al., 2006). No Brasil, a presença de *E. canis* foi documentada pela primeira vez no ano de 1973, em Belo Horizonte, no Estado de Minas Gerais. Atualmente há registros de mamíferos infectados nas cinco regiões do Brasil, apresentado inclusive diversidades no genoma de organismos das diferentes regiões (AGUIAR et al., 2013).

A bactéria é o agente etiológico primário da EMC e tem grande destaque devido sua alta incidência (OTRANTO et al., 2009a, 2009b; DANTAS-TORRES et al., 2008; GASSER et al., 2011). A EMC é uma doença caracterizada por três estágios bastante distintos: agudo, sub-clínico e crônico, sendo este último potencialmente fatal (SKOTARCZAK, 2003; HARRUS; WANER, 2011). Apesar da atenção dada à EMC pela veterinária, ela é considerada uma síndrome altamente variável pelas dificuldades significativas no diagnóstico diferencial com outras doenças, devido as suas diferentes fases e várias manifestações clínicas. Técnicas de diagnóstico tradicionais, incluindo hematologia, citologia, sorologia e isolamento são ferramentas importantes para a identificação da infecção, no entanto, um diagnóstico definitivo requer técnicas moleculares (HARRUS; WANER, 2011).

O genoma de *E. canis* consiste de um único cromossomo circular composto por 1.315.030 nucleotídeos, sendo a proporção G+C de 28,96%. Um total de 984 genes foram identificados no genoma de *E. canis*, incluindo genes de rRNA (5S, 16S e 23S), de forma que 73% do genoma é composto por regiões codificantes. Destes 984 genes, 925 codificam proteínas, sendo que 275 ainda possuem função desconhecida, conforme informações levantadas no sistema Entrez (*Global Query Cross-Database Search System*) mantido pelo NCBI.

Mavromatis et al. (2006) publicaram informações sobre o sequenciamento da *E. canis* str. Jake. O trabalho destaca que o genoma é menor do que outros membros da família, que possuem cerca de 1,5 Kb, mas tem um número semelhante de genes preditos.

Na análise realizada, foi identificado um pequeno grupo de proteínas (n = 12) com repetições em *tandem* e um outro grupo que codifica proteínas com domínio anquirina (n = 5).

Dos genes com *tandem repeats* (Ecaj_0017, Ecaj_0060, Ecaj_0062, Ecaj_0109, Ecaj_0220, Ecaj_0221, Ecaj_0387, Ecaj_0472, Ecaj_0529, Ecaj_0530, Ecaj_07, and Ecaj_0716), gp140 e gp36 (respectivamente Ecaj_0017 e Ecaj_0109), são glicosilados na região do *TR* e atuam indutivamente como adesinas, conforme semelhança com proteínas de *E. chaffeensis* e *E. ruminantium*. Ecaj_0062 é semelhante a proteínas da família WASP, que interagem com o citoesqueleto da célula e facilita a mobilidade bacteriana.

Das proteínas com domínio anquirina (Ecaj_0052, Ecaj_0221, Ecaj_0365, Ecaj_0387, e Ecaj_0627), gp200 (Ecaj_0365), é secretada para o núcleo, onde interage com a cromatina e potencialmente modula a expressão de genes da célula hospedeira. Esta função foi atribuída devido a semelhança da sequência de nucleotídeos com a proteína Anka de *A. phagocytophilum* (STUEN et al., 2013).

Na análise, não foram identificados genes envolvidos na produção de colesterol, que é necessário para o desenvolvimento da bactéria, sugerindo que o mesmo é adquirido do hospedeiro.

Uma caracterização molecular identificou um conjunto de proteínas imunorreativas que podem ser diferencialmente expressas, contribuindo para infecções persistentes nos hospedeiros naturais (DOYLE, 2006; MCBRIDE; WALKER,2010), sendo elas Ank200, gp36, gp140, gp200, família OMP-1 (25 genes) e MAP2. Uma característica dessas proteínas é a existência de *tandem repeats*, que podem ser os responsáveis pelas diferentes formas de expressão dos genes.

Outra importante análise foi a identificação de genes do sistema de secreção de tipo IV. Foram identificados em *E. canis* genes virB e virD. Proteínas vir desempenhar um papel chave nas interações célula hospedeiro-agente patogênico e estão associadas com a secreção de proteínas efectoras.

2.1.3.2 *Ehrlichia chaffeensis*

A *E. chaffeensis* é transmitida aos seres humanos por carrapatos da espécie *Amblyomma americanum* e é o agente etiológico da EMH, descoberta em 1986 (PADDOCK; CHILDS, 2003; WALKER, 2004; BOWMAN, 2011). Nos EUA a *E. chaffeensis* foi isolada pela primeira vez a partir amostras de sangue de um recruta militar, em Arkansas, em 1991. Após este isolamento, a vigilância americana identificou centenas de casos, inclusive fatais, de pacientes com doenças inexplicáveis após exposições de carrapatos, provavelmente erliquioses (PADDOCK; CHILDS, 2003). Estes resultados indicaram que a erliquiose tornava-se um problema de saúde pública generalizada, mas de magnitude indefinida.

Na América e Europa, a EMH é considerada um zoonose emergente e preocupa pela possibilidade de fatalidade quando não diagnosticada e tratada corretamente (PADDOCK et al. 2003). No Brasil, surgem a cada ano novos casos de EMH que chamam a atenção para a necessidade de vigilância em relação ao avanço desta doença (CALIC, 2004). Entretanto, embora a *E. chaffeensis* já tenha sido detectada no Brasil em amostras de sangue de veados

campeiros (MACHADO et. al, 2006) e em aves silvestres (MACHADO et. al, 2012), uma identificação precisa do agente infeccioso em humanos, relacionado ao quadro clínico, ainda não foi obtida (COSTA, 2006).

A evolução tecnológica dos exames para detecção e identificação, especialmente aqueles baseados na reação em cadeia da polimerase (PCR), tem sido importante para aprimorar o diagnóstico tanto para casos humanos como animais (SANTOS, 2009). Neste sentido, é possível que no Brasil, aliado à uma maior atenção dos pesquisadores e agentes de saúde, estes exames demonstrem que existe uma maior incidência de erliquioses humanas do que está relatado.

O genoma completo de *Ehrlichia chaffeensis* str. Arkansas foi sequenciado e possui um cromossomo circular com 1.176.248 nucleótidos, com um percentual G+C de 30,1%. Um total de 1159 genes foram identificados no genoma de *E. chaffeensis*, incluindo genes de rRNA (5S, 16S e 23S). Destes 1159 genes, 1105 são genes que codificam proteínas, sendo que 431 ainda possuem função desconhecida, conforme informações levantadas no ENTREZ. O tamanho médio dos genes é de 840 nucleotídeos e a área codificante ocupa 79.7% do genoma.

Hotopp et al. (2006) publicaram informações sobre o sequenciamento da *Ehrlichia chaffeensis* str. Arkansas. Os genes ausentes incluem aqueles que são necessários para a biossíntese de lipopolissacarídeo e peptidoglicano, tendência que pode também ser vista entre outros membros da família *Anaplasmataceae*.

A proteína 1 de membrana externa OMP1/superfamília P44 (também conhecida como MSP2) é composta de proteínas que são únicas à família *Anaplasmataceae*. As proteínas desta superfamília apresentam alta variação em diferentes amostras. Outra característica é a existência de TRs nesses genes, que geralmente excretam proteínas para o citoplasma e codificam as principais proteínas imunorreativas da bactéria, sendo que algumas já foram caracterizadas (Ank200, TRP120, TRP47, TRP32, família OMP-1 [22 genes] e MAP2). TRP120 foi também caracterizada em *E. canis* (ECH_0170) (YU et al., 2000). Polimorfismos em vários genes foram relatados, resultando em diferentes níveis de virulência, mas divergência genômica global é desconhecida (MIURA; RIKISA, 2007).

O genoma possui também uma baixa capacidade de codificação de genes envolvidos no metabolismo intermediário central, transporte e funções regulatórias, sugerindo a aquisição de compostos específicos do hospedeiro.

2.1.3.3 *Anaplasma phagocytophilum*

A. phagocytophilum ficou conhecida por causar doenças em ruminantes domésticos e cavalos ao longo de décadas. Mas, recentemente, a infecção foi detectada em outras espécies de mamíferos e em seres humanos, provocando a Anaplasmose Granulocítica Humana (AGH), uma importante doença imunopatológica transmitida por carrapatos Ixodídeos nos EUA, Europa e Ásia (STUEN et al. 2013).

A. phagocytophilum é uma das principais causas de erliquiose no mundo. Dados sorológicos recentes sugerem que até 15%-36% da população de áreas endêmicas (citadas anteriormente) foi infectada. Metade de todos os pacientes sintomáticos necessitam de hospitalização, e 5% -7 % necessitam de cuidados intensivos (NELSON et al., 2008).

No Brasil há relatos e estudos de detecção molecular de infecção em cães (SANTOS et al., 2011), aves selvagens e ovelhas (SANTOS et al., 2013). Estudos de soroprevalência encontraram evidências de infecção por *A. phagocytophilum* em pequenos ruminantes, eqüinos e cães no estado de Pernambuco, São Paulo e Paraná (SANTOS et al., 2011). Apesar desta evidência, não é possível considerar a AGH como uma doença endêmica no Brasil, uma vez que estudos são isolados e as amostras não são representativas.

A. phagocytophilum é um patógeno que evoluiu utilizando uma notável capacidade de colonizar e replicar, tanto em carrapatos quanto em células de mamíferos. No hospedeiro mamífero, a bactéria subverte a apoptose, a autofagia, inibe processos oxidativos e inflamatórios da célula para promover a sua persistência (SEVERO et al, 2012; RIKIHISA, 2011).

O genoma de *A. phagocytophilum* consiste de um único cromossomo circular composto por 1.471.282 nucleotídeos, sendo a proporção G+C de 41,6 %. Um total de 1412 genes foram identificados no genoma, incluindo genes de rRNA (5S, 16S e 23S). Destes 1412 genes, 1264 são genes que codificam proteínas, sendo que 552 ainda possuem função desconhecida, conforme informações levantadas no ENTREZ, o NCBI.

O trabalho de Hotopp et al. (2006) divulgou informações sobre o genoma de *A. phagocytophilum* str. HZ. As proteínas mais estudadas da bactéria incluem as de superfície, superfamília Omp-1/P44/Msp2. O genoma contém três genes omp-1, um MSP2, dois homólogos MSP2, um MSP4 e 113 p44. Essas proteínas são altamente imunogênicas nos pacientes infectados e têm sido o foco principal para o desenvolvimento de antígenos de diagnóstico diferencial e vacinas (MCBRIDE; WALKER, 2010). Características desses genes

incluem variabilidade genética, existência de TRS e secreção de proteínas para o citoplasma. Alguns possuem também domínio anquirina.

Em uma comparação com a *E. chaffeensis*, as duas espécies compartilham aproximadamente 500 genes, sendo que 470-580 genes são únicos para cada espécie. A maioria dos genes compartilhados são também compartilhados com outras bactérias da família, e a maior parte dos genes exclusivos codificam proteínas hipotéticas. Os genes específicos caracterizados de *A. phagocytophilum* incluem HGE-14 e HGE-2 e proteínas de membrana externa p44. Desses genes, HGE-14 e HGE-2 apresentam ser conservados e proteínas p44 possuem genes com um alto nível de variação de acordo com análises de isolados de duas regiões distintas dos Estados Unidos.

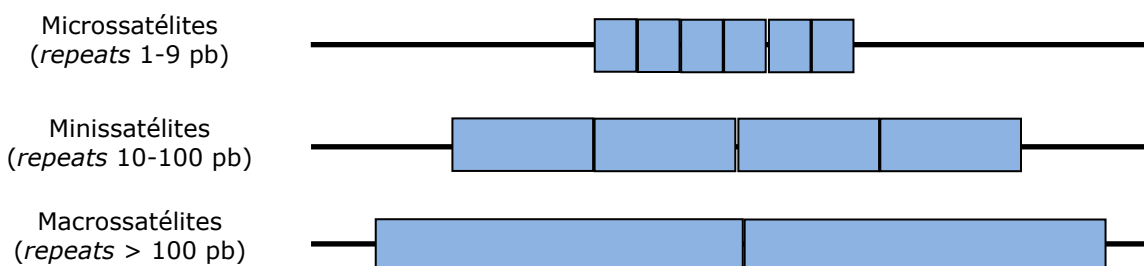
2.2 TANDEM REPEATS (TRs)

TRs são sequências de DNA que se repetem como múltiplas cópias adjacentes com a mesma orientação na cadeia de DNA, com elementos intervenientes ou não (BENSON, 1999). TRs foram descritos nos reinos Archaea, bactérias, eucariotos e em vírus (LEVDANSKY et al., 2008).

Em bactérias, TRs vêm sendo analisados por ser uma característica encontrada em genes relacionados à patogenicidade e também pela sua instabilidade, podendo expandir-se e/ou contrair-se resultando em uma modulação de genes específicos. Essa variabilidade genética permite uma rápida adaptação ao ambiente e um variado repertório para iludir o sistema imunológico do hospedeiro (ZHOU et al., 2013).

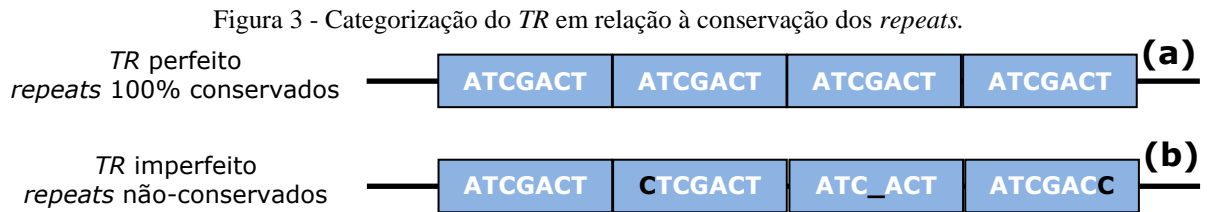
Estas repetições são classificadas em relação ao tamanho por microssatélites, minissatélites e macrossatélites, embora não haja um consenso geral sobre essas categorias. Os microssatélites são compostos por *repeats* de tamanho 1-9 pb, os minissatélites por *repeats* de 10-100 pb e os macrossatélites por *repeats* com mais de 100 pb (LEVDANSKY et al., 2008). A Figura 2 ilustra as três categorias descritas.

Figura 2 - Diferentes tamanhos dos *repeats* utilizados para classificação dos TRs



Fonte: Autor

Por exemplo, na Figura 3, o *repeat* ATCGACT, denominado também como sequência consenso, aparece quatro vezes no microsatélite em (a). Vale ressaltar que as repetições podem ser idênticas ou parcialmente degeneradas. No exemplo em (b), temos a troca de uma base nos *repeats* 2 e 4, e a ausência de G no *repeat* 3.



Fonte: Autor.

2.2.1 Identificação *in silico* e Localização de *TRs* no Genoma

Existem diversos programas disponíveis para a detecção de *TRs* em uma sequência de nucleotídeos, como eTandem (RICE et al., 2000), mpREPS (KOLPAKOV et al., 2003), T-Reks (KAJAVA, 2009) e Tandem Repeats Finder (TRF) (BENSON, 1999). Estes programas calculam um *score* linear para os *TRs* identificados baseado em 3 parâmetros: o comprimento de cada *repeat*, a conservação do *repeat* ao longo do *TR* e o número de unidades de *repeats* que o *TR* possui. Legendre et al. (2007) utilizou técnicas de aprendizado de máquina para desenvolver um modelo não-linear de algoritmo, denominado SERV, que produz um *score* capaz de prever a probabilidade de um *TR* variar em relação ao número de *repeats* entre diferentes estirpes da mesma espécie.

Além dos algoritmos de análise disponíveis, várias bases de dados de *TRs* anotados em procariotos têm sido criadas para facilitar a caracterização dos *TRs* identificados, como TRsDB (<http://minisatellites.u-psud.fr>), PSSRDb (<http://pssrdb.cdfd.org.in>) e MICAS (<http://180.149.48.108/micas/index.php>).

Após a identificação de *TRs*, uma importante classificação é em relação à localização de *TRs* no genoma: *TRs* codificantes (também designados como *TRs* intergênicos) e *TRs* não-codificantes (*TRs* intragênicos).

TRs não-codificantes podem ser encontrados em procariotos em regiões não transcritas. Em eucariotos, podem ser encontrados também em íntrons. *TRs* codificantes estão

localizados dentro da sequência de codificação de um gene, e podem ser transcritos em mRNA e traduzidos em uma proteína (LEVDANSKY et al., 2008).

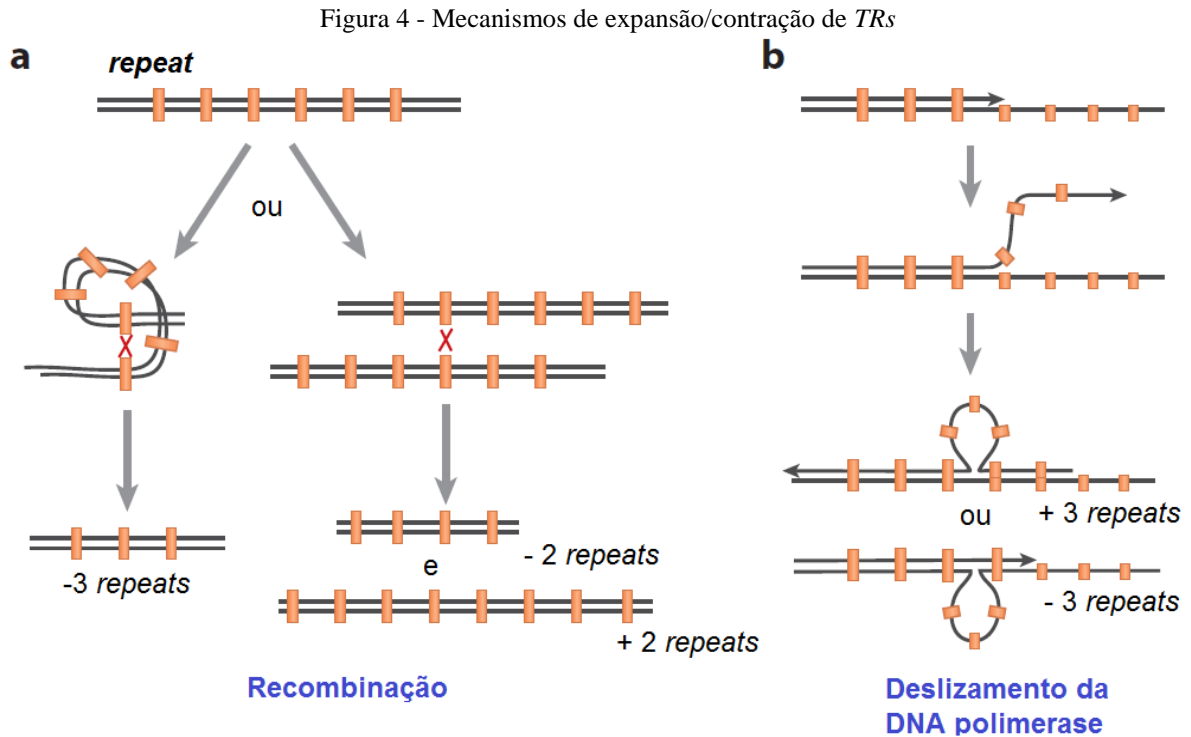
TRs em procariotos são encontrados em maior escala em regiões codificantes, pelo fato de que os genomas bacterianos são mais compactos do que o genoma de eucariotos. Por exemplo, em *Escherichia coli K-12*, 79,5% dos *TRs* são localizados em regiões codificantes, enquanto que no genoma do baiacu japonês (*Fugu rubripes*), apenas 11,6% dos *TRs* estão em regiões codificantes (VERSTREPEN et al., 2007).

Diversos estudos tem relacionado a presença de *TRs* à instabilidade genética e também como uma característica de genes relacionados a fatores de virulência, proteínas de superfície e proteínas imunorreativas (GEMAYEL, 2010, 2012; MCBRIDE; WALKER, 2010; RIKIHISA, 2011; ZHANG et al., 2008; ZHOU et al., 2013).

2.2.2 Variabilidade de *TRs*

A variabilidade de *TRs* tem sido atribuída a um dos principais mecanismos relacionados com a plasticidade genômica e diversidade entre organismos da mesma espécie, exercendo um importante papel na adaptação bacteriana (ZHOU et al., 2013). Regiões que contêm *TRs* são potencialmente hipermutáveis pela contração (supressão) ou expansão (inserção) das unidades de repetição que compõem o *TR* (GEMAYEL et al., 2012). Os polimorfismos (conhecidos como *VNTR* - *Variable Number of Tandem Repeat*) encontrados fornecem uma importante ferramenta para a epidemiologia bacteriana, como genotipagem de estirpes (RAMAZANZADEH; MCNERNEY, 2007).

A variabilidade fenotípica de *TRs* (contração/expansão) é o resultado de dois principais mecanismos genéticos: deslizamento da DNA polimerase durante o processo de replicação de DNA (*DNA slippage*) e recombinação (GEMAYEL et al., 2010), como mostra a Figura 4. No deslizamento da DNA polimerase (Figura 4 - b), a variabilidade ocorre devido a um mal pareamento (formação de uma "alça") entre a fita molde e a nova fita. Assim, quando a formação desta alça ocorre na nova fita, o reparo do mal pareamento resulta num aumento do número de repetições. Em contraste, quando a alça é formada na fita molde, isso resulta na diminuição do número de repetições. A recombinação ocorre durante a meiose quando acontece um alinhamento desigual entre as cromátides dos cromossomos homólogos (Figura 4 - a).



Fonte: Adaptado de Gemayel et al. (2010).

A recombinação tem sido vinculada ao rearranjo de *TRs* com *repeats* maiores, enquanto que o deslizamento da DNA polimerase é o mecanismo de variação subjacente dominante em *TRs* com *repeats* menores (GEMAYEL et al., 2010; ZHOU et al., 2013). No entanto, o mecanismo exato pelo qual cada tipo de *TR* varia ainda não é exato.

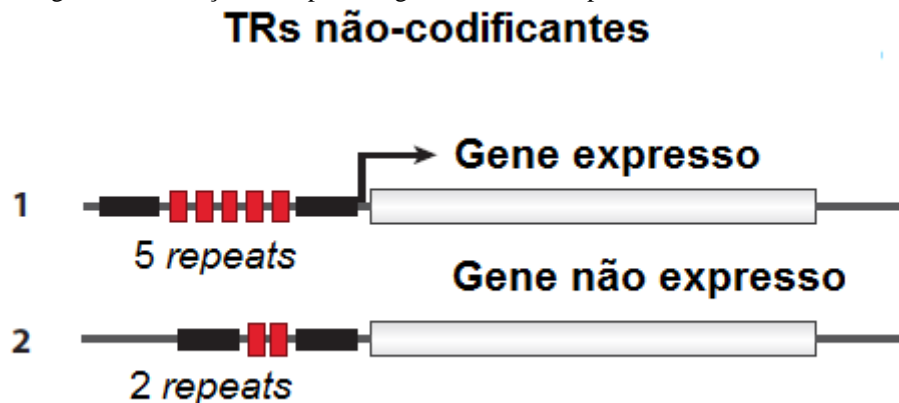
Embora *TRs* estejam propensos a serem instáveis, a frequência dessas alterações está relacionada a fatores intrínsecos (estruturais) e extrínsecos (ambientais). Em relação aos fatores intrínsecos, quanto mais longo e mais conservado for o *repeat*, maior será a frequência de mutação (GEMAYEL et al., 2012). Para Legendre et al.(2007), o número de *repeats* que compõem o *TR* influencia mais na instabilidade do que o tamanho e a conservação do *repeat*. Processos celulares, como aumento na taxa de transcrição também podem aumentar a instabilidade.

Exemplos clássicos incluem a contração de *repeats* CT após a infecção fúngica (SCHIMIDT; MITTER, 2004) em *TRs* usados para genotipagem de *Escherichia coli*, que demonstraram maior instabilidade em situações como aumento de temperatura e fome (ZHOU et al., 2013).

2.2.3 Variabilidade em *TRs* Não-codificantes

A variabilidade de *TRs* resulta em diferentes impactos de acordo com a localização no genoma. Várias evidências indicam que a variabilidade em *TRs* localizados em regiões regulatórias (não-codificantes) do genoma podem afetar a expressão dos genes (LI et al., 2012; JANSEN et al., 2012), como ilustra a Figura 5.

Figura 5 - Alteração na expressão genética mediada por *TRs* não-codificantes



Fonte: Adaptado de Gemayel et al. (2010)

Mais especificamente, *TRs* posicionados em torno da região de promotores e elementos regulatórios cis podem induzir a variação de fase, atuando como um importante mecanismo de diversificação da expressão e da função dos genes (GEMAYEL et al., 2012).

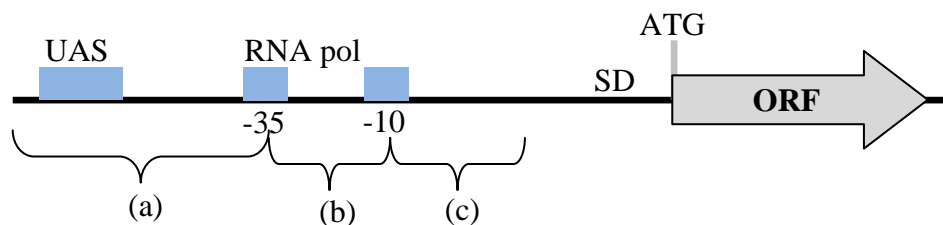
O posicionamento do *TR* na região não-codificante interfere diferentemente na transcrição do gene. A variação de *TRs* posicionados a mais de 35 sítios antes (upstream) do promotor (Figura 6 (a)) podem afetar a expressão do gene por alterar o número de sítios de ligação dos fatores de transcrição, afetando a iniciação da transcrição por modificar a ligação de proteínas regulatórias (ZHOU et al., 2013). Um exemplo deste evento foi descoberto em bactérias patogênicas. Em *Neisseria meningitidis*, alterações no número de *repeats* TAAA localizados no promotor do gene *nadA* (que codifica uma proteína que funciona como a invasina e adesina) induz a expressão variável da proteína de adesina (GEMAYEL et al., 2012).

A variação de *TRs* posicionados entre -35 e -10 sítios antes (*upstream*) do promotor (Figura 6 (b)) afeta a iniciação da transcrição por alterar a distância dos elementos promotores, comprometendo a ligação da RNA polimerase (MOXON et al., 1994). Van Ham et al. (1993) identificaram uma alteração na expressão do gene *hif* de *Haemophilus influenzae*,

um patógeno que adere a células epiteliais humanas com a ajuda de fímbrias de LKP. Essas fímbrias são codificadas pelo gene *hif* e são importantes para a infecção por *H. influenzae*. Variação na expressão destas fímbrias é mediada pela instabilidade de um *TR* com *repeats* dinucleotídeos posicionado entre os sítios -35 e -10 antes (*upstream*) do promotor. *TRs* identificados com 9, 10 ou 11 *repeats* resultam respectivamente em ausência de expressão, expressão elevada, e baixa expressão do gene *hif*.

Enquanto variações em *TRs* não-codificantes são mais comuns nas duas classes descritas anteriormente, alguns casos são mediados por *TRs* localizados entre o local de início da transcrição e o início do ORF (códon de início do gene) (Figura 6 (c)). Análise na expressão do gene *UspA1no* do patógeno *Moraxella catarrhalis*, revelou uma variação de expressão. O gene expressa uma adesina que auxilia na ligação com células epiteliais humanas, de forma que a instabilidade altera a sua capacidade de aderência. A evidência é que a variação nessa região afetou a eficiência na ligação de reguladores de transcrição e/ou a estabilidade do mRNA (ZHOU et al., 2013).

Figura 6 - Possíveis posições de *TRs* não-codificantes na região de promotores



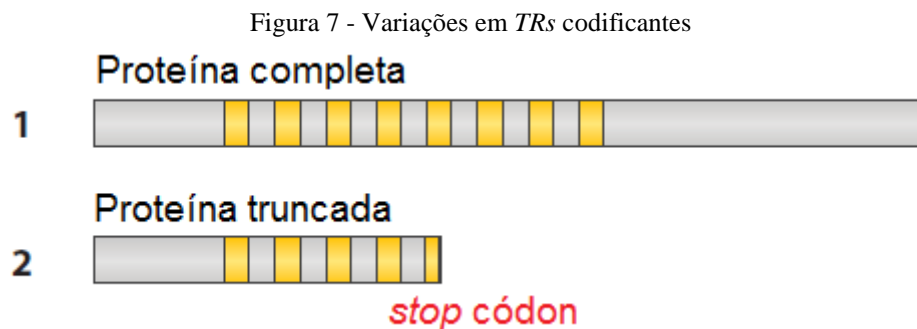
Fonte: Adaptado de Zhou et al. (2013).

2.2.4 Variabilidade em *TRs* Codificantes

TRs codificantes estão sendo apontados como um dos principais mecanismos moleculares responsáveis por gerar variabilidade em organismos da mesma espécie, desempenhando um papel fundamental na patogenicidade e adaptação bacteriana frente às mudanças que ocorrem no ambiente (ZHOU et al., 2013). *TRs* codificantes estão localizados dentro da sequência de codificação de um gene, e podem ser transcritos em mRNA e traduzidos em uma proteína (LEVDANSKY et al., 2008). Expansões ou contrações em *TRs* codificantes podem levar a um ganho ou perda da função do gene, levando a variações fenotípicas por alterar o número de *repeats* entre diferentes isolados, resultando na expansão ou contração do bloco de aminoácidos que formarão a proteína (VERSTREPEN et al., 2005).

Em agentes patogênicos procarióticos, esta capacidade de variabilidade permite alterar a morfologia de antígenos de superfície celular imunogênicos e adesinas, de forma a iludir o sistema imunológico do hospedeiro (VERSTREPEN et al., 2005). Em seres humanos, expansões de alguns genes estão diretamente relacionados com cerca de 30 doenças hereditárias humanas, como a síndrome X frágil, ataxia de Friedreich e distrofia miotônica (GEMAYEL et al., 2010).

Além de alterar funcionalmente a proteína, a variação em *TRs* codificantes podem também interromper a expressão do gene. No entanto, este mecanismo é dependente de uma característica do *TR*. Mais especificamente, se o tamanho do repeat do *TR* não é um múltiplo de três, rearranjos são capazes de induzir o aparecimento de um stop códon na cadeia de aminoácidos, gerando uma proteína não-funcional. *TRs* com repeat múltiplo de três está relacionado com alterações funcionais (GEMAYEL et al., 2010), como ilustra a Figura 7.



Fonte: adaptado de Gemayel et al. (2010).

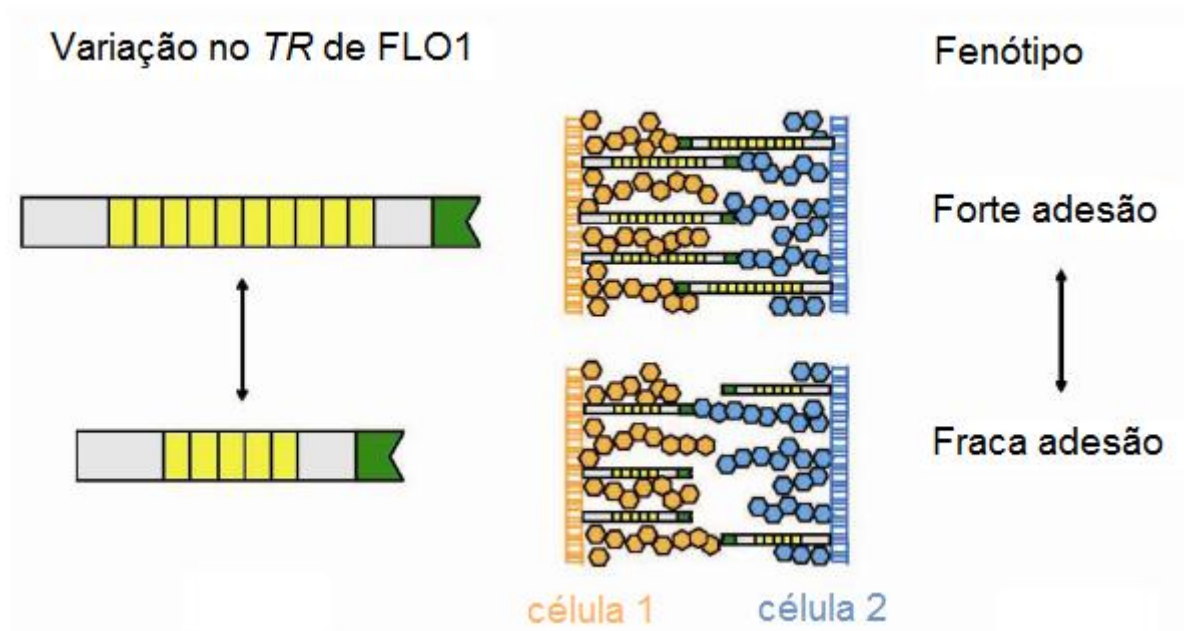
Na bactéria *Neisseria gonorrhoeae*, variações em *TRs* de genes de superfície celular da família p II permitem a evasão do sistema imunitário do hospedeiro. Mudanças nos *repeats* CTCTT de *TRs* codificantes de peptídeo sinal da membrana dos genes de superfície celular podem levar a tradução correta ou incorreta de proteínas (GEMAYEL et al., 2012).

Em relação à distribuição de *TRs* codificantes entre os genes, tem-se observado uma aglomeração em genes cujos produtos são expostos na superfície da célula ou envolvidos na biogênese de estruturas de superfície celular, como lipopolissacarídeos (LPS), adesinas, pili, fímbrias, cápsulas e também em genes envolvidos na patogenicidade e subversão de processos celulares do hospedeiro (MOXON et al., 1994; JORDAN et al., 2003; DOYLE et al., 2005, 2006; WAKEEL et al., 2009; RENNOLL-BANKERT; DUMLER, 2012; KURIAKOSE et al., 2012). Estes estudos comprovam que a variabilidade nesses genes contribui para a rápida

geração de diversidade nas estruturas de superfície, que podem servir como um mecanismo para escapar da resposta imune.

O trabalho de Verstrepen et al.(2005) destaca tal característica com a pesquisa sobre o gene FLO1 de *Saccharomyces cerevisiae*, que codifica uma adesina de superfície celular e facilita a adesão a uma célula ou superfície. O gene contém um *TR* variável em diferentes amostras com *repeats* de aproximadamente 100 nt. Testes fenotípicos mostram que a intensidade da aderência está diretamente relacionada ao comprimento do *TR*, como ilustra a Figura 8.

Figura 8 - Variação no *TR*-codificante do gene FLO1 altera a capacidade de adesão.



Fonte: Verstrepen et al.(2005).

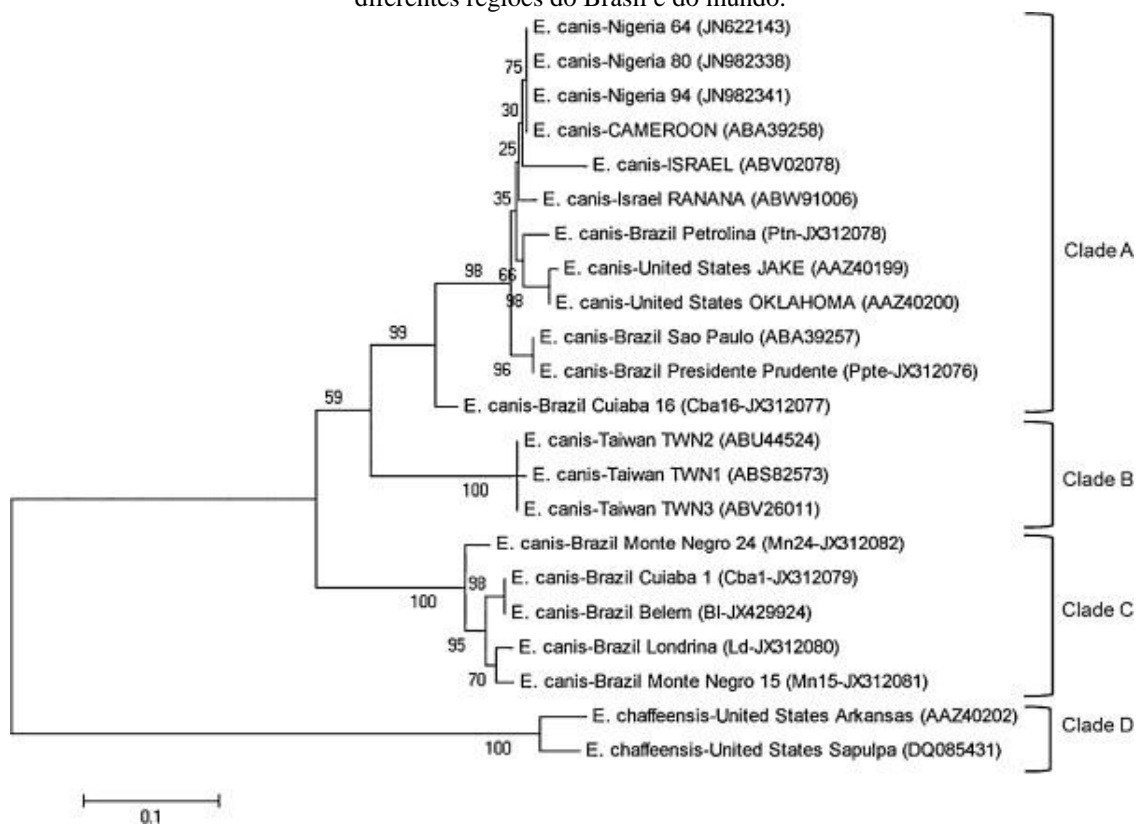
2.2.5 Alguns Trabalhos envolvendo *TRs* e *Anaplasmataceae*

Aguiar et al. (2013) desenvolveu um trabalho para identificar a diversidade genética de *E. canis* no Brasil levando-se em consideração a instabilidade do *TR* presente no gene TRP36 (também conhecido como gp36). O gene TRP36 é bastante utilizado em pesquisas sobre a diversidade genética de *E. canis* no mundo pela sua instabilidade, diferentemente dos genes 16S rRNA e as proteínas OMP-1, Dsb, TRP19 e TRP140, que demonstram um grande nível de conservação (DOYLE et al., 2005, 2006; ZHANG et al., 2008).

O gene TRP36 contém um importante epítipo na região do *TR*, sendo uma das principais proteínas imunorreativas associadas com as interações hospedeiro-patógeno, tais como a adesão e internalização, actina e nucleação e evasão da resposta imune.

O trabalho fez também uma comparação com as sequências do gene TRP36 identificadas em amostras de sangue de cães naturalmente infectados das cinco regiões geográficas do Brasil com sequências de isolados de diferentes localidades do mundo encontradas no GenBank. O resultado é mostrado na árvore filogenética da Figura 9.

Figura 9 - Árvore filogenética baseada em sequências de aminoácidos do gene TRP36 de isolados de *E. canis* de diferentes regiões do Brasil e do mundo.



Fonte: AGUIAR et al., 2013.

Foram identificados dois grandes genogrupos brasileiros distintos (A e C) e diversos subtipos com base na análise do gene TRP36. Uma interessante observação a ser feita em relação aos resultados do trabalho é que o gene TRP36 de amostras das bactérias de Brasil e Estados Unidos eram tidos como conservados (DOYLE et al., 2005, 2006; ZHANG et al., 2008). Os trabalhos comparavam as amostras São Paulo e a americana *Jake*, pertencentes ao genogrupo A.

Cruz et al. (2012) realizaram uma análise molecular e filogenética nos genes 16rRNA *groEL*, *dsb*, *gltA* e TRP36 de amostras retiradas do carrapato *Rhipicephalus microplus*

(conhecido popularmente como "carrapato do boi") em uma fazenda do estado de Minas Gerais com o objetivo de analisar a variabilidade de espécies *ehrlichia*. Árvores filogenéticas baseadas nos genes 16S rRNA , groEL , dsb e gltA mostraram uma *Ehrlichia spp.* isolada em um clado separado de qualquer outro relatado anteriormente. A análise molecular do gene TRP36, ortólogo de TRP47 de *E. chaffeensis*, mostrou uma nova unidade de repetição no TR do gene, quando comparados com os descritos na literatura. Conclui-se com os resultados que se tratava de uma nova espécie, denominada *E. mineirensis*, com novas propriedades antigênicas previstas para glicoproteína TRP36.

Wakeel et al. (2009) identificaram que a proteína TR47 (também denominada gp47) de *E. chaffeensis*, caracterizada em trabalhos anteriores por possui TRs em sua formação, ser diferencialmente expressa em diferentes isolados e ser efetora, interage com várias proteínas do hospedeiro associadas à sinalização, regulação transcricional e transporte de vesícula, como PCGF5, FYN, PTPN2 e CAP1. A pesquisa, que utilizou a técnica Imunofluorescência, destaca o importante papel da proteína na reprogramação dos processos celulares do hospedeiro, visando subverter os mecanismos de defesa e criar um ambiente hospitaleiro para a sobrevivência da bactéria.

O clássico trabalho de Doyle et al. (2005) caracterizou as glicoproteínas imunoreativas e ortólogas TRP36 e TRP47 de *E. canis* e *E. chaffeensis*. Os genes que codificam essas proteínas possuem TRs que contém epítomos reconhecidos pelo sistema imunológico do cão. Uma análise comparativa com isolados da América do Norte das duas bactérias identificou uma variação em relação ao número de *repeats* do TR. Para analisar a heterogeneidade global, os genes de amostras de diferentes continentes foram amplificados. Os resultados mostraram que os *repeats* são bem conservados, mas existem diferenças no número de *repeats* dos TRs. Desde então, diversos pesquisadores concentraram as análises relacionadas com a diversidade global de *E. canis* e *E. chaffeensis* nos genes TRP36 e TRP47.

Kuriakose et al. (2012) analisaram a ação de anticorpos contra as proteínas secretadas e imunorreativas TRP 32, TRP47 e TRP120 durante a infecção de *E. chaffeensis*, usando modelos *in vitro* e *in vivo*. As três proteínas possuem TRs em sua composição. A proteção foi demonstrada *in vitro* após tratamento profilático e administração terapêutica de anticorpos anti-TRP para os epítomos espécie-específicos. *In vivo*, a transferência passiva de soro com epítomos espécie-específicos TRP reduziu significativamente a carga de *Ehrlichia*, protegendo os camundongos contra a infecção letal. Os resultados demonstram que os anticorpos contra epítomos TRP32, TRP47 e TRP120 são protetores imunes durante a infecção de *E. chaffeensis* e proporcionam proteção semelhante aos anticorpos da família OMP. A pesquisa destaca que

estes epítomos devem ser incluídos no desenvolvimento de vacinas e terapias para doenças, porém, mais estudos precisam ser feitos para elucidar o mecanismo de proteção anti-TRP mediada por anticorpos durante a infecção por *E. chaffeensis*.

Zhang et al.(2008) realizaram um estudo visando identificar a diversidade genética e antigênica das proteínas imunorreativas gp36, gp200, gp140 e gp19 de estirpes de *E. canis* de três países de distintos continentes: Estados Unidos/Jake (EUA), Brasil/São Paulo (BR), Israel/611 (I-611) e Israel/Ranana (I-R). Dos genes estudados, dois possuem *TRs* sua formação (gp36 e gp140). Primeiramente foram analisadas as diversidades genéticas das quatro proteínas de cada uma das amostras por meio da amplificação via PCR e sequenciamento.

O gene gp36 foi o mais divergente. O *repeat* das amostras EUA, BR e I-R apresentam uma sequência idêntica de aminoácidos, mas com diferentes números de repetição, sendo EUA (12), BR(18) e I-R(10). O *TR* do gene da amostra I-611(11 *repeats*) apresentou diferenças nos nucleotídeos que compõe o *repeat*, resultando em 9 aminoácidos diferentes ao longo do *TR* da amostra. gp200 apresentou grande similaridade analisando a sequência de aminoácidos entre EUA e BR, sendo 99.6% de identidade (divergência de quatro aminoácidos) porém houve discrepância maior com a amostra de I-611 (94.3%).

gp19 e gp140 apresentaram sequencias idênticas para as estirpes EUA e BR e pequenas alterações nos *repeats* de I-611 e I-R, que resultaram em alterações nos aminoácidos da proteína.

Posteriormente, células infectadas com *E. canis* de cada isolado foram examinadas por *western immunoblotting* utilizando soro de cães naturalmente infectados de cada país, e as quatro proteínas imunorreativas foram identificadas em cada estirpe usando anti-soro, apresentando notáveis diferenças nas proteínas gp36 e gp140. A massa molecular da estirpe BR foi substancialmente maior que a estirpe EUA no gene gp36. Esta diferença é explicada pela maior quantidade de *repeats* do *TR* dos genes das estirpes BR (18 *repeats*) e EUA (12 *repeats*). Apesar da diferença de massa, não houve uma variação antigênica. Uma característica interessante foi a detecção de dois tipos de soro envolvendo isolados de Israel. O primeiro tipo reagiu fortemente com gp36 de estirpes de BR e EUA, e o segundo tipo (I-611) reagiu com uma proteína da estirpe IS cujo tamanho é similar com gp36 de EUA. Neste sentido, as evidências mostram que existem pelo menos duas estirpes de *E. canis* distintos circulantes em Israel. O artigo destaca a similaridade entre as estirpes EUA e BR, porém o trabalho utiliza uma estirpe do estado de São Paulo para a comparação. Estudos recentes

mostraram a diversidade genética entre estirpes de diferentes regiões do Brasil, levando-se em consideração o gene gp36 (AGUIAR et al., 2013).

Rennoll-Bankert e Dumler (2012) utilizaram a bactéria *A. phagocytophilum* para analisar a estratégia de patógenos bacterianos de alterar a expressão gênica do hospedeiro. A pesquisa destaca a ação da proteína efetora AnkA, que possui *TR* em sua formação e é capaz de promover modificações de histonas e remodelar a cromatina de genes do hospedeiro, a fim de subverter respostas de defesa imunes, permitindo a sobrevivência de bactérias. *A. phagocytophilum* modula a "explosão respiratória" ("*respiratory burst*") de granulócitos, diminuindo a transcrição de vários genes na oxidase do fagócito. AnkA é secretada para o núcleo das células infectadas e se liga a regiões ricas em AT do promotor do gene *CYBB*, diminuindo a expressão do gene.

2.3 MOTIVOS CpG

Um motivo que tem sido estudado em bactérias e utilizado como adjuvante na criação de vacinas envolve a análise de Motivos CpG (BODE et al., 2011). O genoma bacteriano possui uma frequência muito mais elevada de dinucleotídeos CpG do que o genoma de mamíferos (KRIEG et al., 1995; WAGNER, 2002). Além disso, os dinucleotídeos CpG encontrados no DNA de bactérias geralmente não são metilados, diferentemente do DNA dos mamíferos, que além de apresentar supressão de ocorrência de dinucleotídeos CpG, quando os mesmos aparecem no genoma, são geralmente metilados (KRIEG, 2002; CHAUNG, 2006). Essas duas características específicas constituem um sinal de perigo que pode ser reconhecido pelo sistema imunológico de mamíferos e, por conseguinte, uma resposta imune pode ser induzida quando estes sinais são encontrados.

O reconhecimento de Motivos CpG no hospedeiro requer o engajamento do receptor *Toll-like*. Os receptores *Toll-Like* (TLR) são proteínas transmembranas altamente conservadas que desempenham papel importante na detecção e reconhecimento de patógenos microbianos, bem como na sinalização para a produção de proteínas e citocinas pró-inflamatórias, que são essenciais para a ativação das respostas imunes. Existem vários tipos de receptores TLR, que são responsáveis por diferentes tipos de contaminação. O TLR 9 é o responsável pelo reconhecimento do DNA bacteriano e viral. Cada receptor *Toll-Like* tem sua própria via de sinalização intrínseca e induz respostas biológicas específicas contra micro-organismos (CHUANG et al., 2004).

Segundo Mutwiri (2003), Motivos CpG com capacidade de desencadear uma resposta imune no hospedeiro devem ser compostos por seis nucleotídeos e obedecer ao seguinte arranjo 5' → X1X2 CpG Y1Y2 → 3' no genoma bacteriano, sendo X1 uma purina, X2 uma purina ou timina, Y1 e Y2 são pirimidinas.

Figura 10 - Arranjo/regra para identificação de Motivos CpG no genoma bacteriano

...	X1	X2	C	G	Y1	Y2	...
	A G	A G T			T C	T C	

| operador lógico OU

Fonte: MUTWIRI, 2003.

A potencial aplicabilidade da utilização de Motivos CpG como forma de ativação de respostas imunes é comprovada em várias espécies.

O uso de oligonucleotídeos sintéticos de DNA contendo CpG não metilados juntamente com bases flanqueadoras (Motivo CpG) do DNA de uma determinada bactéria pode induzir respostas celulares que são reconhecidas pelo sistema imunológico de mamíferos, tornando o animal imune à bactéria (WANG, 2004). Baseados em resultados obtidos em espécies modelos, como ratos, ensaios clínicos estão em andamento em humanos para avaliar as terapias CpG para alergias, câncer e doenças infecciosas (CHAUNG, 2006; MUTWIRI et al., 2009; KLINMAN et al., 2009; KRIEG, 2012). Resultados *in vitro* e *in vivo* também confirmam a eficiência do tratamento em diversos animais domésticos como gatos, cavalos, aves, ovelhas e cães (MUTWIRI, 2003). No entanto, as aplicações potenciais para imunostimuladores de CpG em espécies de importância veterinária estão apenas começando a ser exploradas.

Vale ressaltar que os efeitos imunostimulantes de Motivos CpG são específicos de cada espécie e não existe uma regra geral, como número de ocorrência e localização no genoma, para seleção do Motivo que despertará a resposta imune.

Fundamentados pela técnica de utilização de Motivos CpG ao combate bacteriano, alguns softwares foram desenvolvidos para evidenciar alvos genômicos, com destaque para os sistemas AIMIE (MRÁZEK, 2008) e CpGCluster (HACKENBERG, 2006). O AIMIE tem como característica principal o foco na distribuição dos motivos CpG ao longo do genoma. Já o CpGCluster é um sistema que trabalha unicamente as aparições nucleotídicas CG e a distância entre os Motivos identificados ao longo do genoma. Ambas as ferramentas citadas

possuem código-fonte fechado, o que dificulta a utilização das mesmas para modificações e identificação dos seus algoritmos.

Essas ferramentas de bioinformática auxiliam na mineração de motivos CpG existentes no genoma bacteriano que possam ser utilizados como adjuvantes na criação de vacinas que estimulem o sistema humoral do hospedeiro (BODE et al., 2011), estratégia que tem se demonstrado viável para *Ehrlichia* spp e *Anaplasma* spp (MCBRIDE; WALKER, 2010).

3 HIPÓTESE

Como mostrado e discutido ao longo do capítulo 2, diversas pesquisas tem sido publicadas buscando ampliar a compreensão dos dados genômicos das três bactérias da família analisadas no trabalho. O desenvolvimento de vacinas e métodos de diagnóstico são necessários para estes patógenos humanos e veterinários, porém existem ainda algumas questões a serem compreendidas. Estas questões envolvem principalmente a variabilidade genética e antigênica das bactérias, a identificação de antígenos que estimulem a imunidade e a identificação de genes envolvidos nas interações moleculares entre patógeno-hospedeiro (MCBRIDE; WALKER, 2010).

A grande quantidade de genes ainda não caracterizados ou parcialmente caracterizados justificam a aplicação de métodos de bioinformática para correlacionar o grande volume de informações existentes em bancos de dados e na literatura.

TRs têm sido alvo de pesquisas e estão sendo considerados um importante mecanismo de variabilidade genética, permitindo uma rápida adaptação e variadas formas de iludir o sistema imunológico do hospedeiro. A presença de *TRs* é também uma característica de genes caracterizados como imunorreativos e envolvidos nos processos de subversão da célula hospedeira.

Motivos CpG são considerados eficientes coadjuvantes na criação de vacinas que estimulem respostas imunes do hospedeiro. Essa estratégia tem se mostrado viável em bactérias da família *Anaplasmataceae* (MCBRIDE; WALKER, 2010).

Neste contexto, a hipótese do presente trabalho é que as análises genômicas realizadas evidenciem alvos que possam ser utilizados para compreensão da patogenicidade, diversidade e distribuição geográfica, marcadores epidemiológicos, criação de drogas para o combate destas bactérias e ferramentas de diagnóstico. Vale ressaltar que, em um segundo momento, os resultados gerados serão explorados em laboratório.

4 OBJETIVOS GERAL E ESPECÍFICOS

O objetivo geral do trabalho é realizar *in silico* as análises de *TRs* e Motivos CpG e evidenciar alvos genômicos de bactérias da família *Anaplasmataceae* que possam estar relacionados com a variabilidade genética, interação patógeno-hospedeiro e ativação do sistema imunológico.

Os objetivos específicos são:

1. Identificar *TRs* no genoma de bactérias da família *Anaplasmataceae*;
2. Classificar os *TRs* identificados em *TRs* codificantes e não-codificantes e identificar genes que possuem *TRs*;
3. Identificar Motivos CpG no genoma de bactérias da família *Anaplasmataceae*;
4. Classificar Motivos CpG em codificantes e não-codificantes e identificar genes que possuem Motivos CpG;
5. Correlacionar alvos de destaque resultantes das análises com informações disponíveis na literatura e com ferramentas de busca em bancos de dados públicos;
6. Evidenciar nos resultados encontrados Motivos que poderão ser usados para ativação do sistema imunológico do hospedeiro.
7. Evidenciar genes com *TRs* que possam estar relacionados com a variabilidade em bactérias da família *Anaplasmataceae* e interação patógeno-hospedeiro;

5 MATERIAIS E MÉTODOS

5.1 MATERIAIS UTILIZADOS

Os materiais utilizados consistem em arquivos mantidos em bases de dados do *National Center for Biotechnology Information* (NCBI) contendo dados do genoma e dos genes das bactérias *E. canis* (CP000107.1), *E. chaffeensis* (NC_007799.1) e *A. phagocytophilum* (NC_007354.1).

5.2 MÉTODOS UTILIZADOS NA ANÁLISE DE MOTIVOS CPG

Para a análise de Motivos CpG que possam causar ação imunoestimulante em mamíferos ao ataque de bactérias da família *Anaplasmataceae*, foi desenvolvido o *software CpG Highlight 1.0*. O *software* consiste em uma nova ferramenta Java com código aberto e independente de plataforma que identifica Motivos CpG em genomas de bacterianos. *CpG Highlight* gera gráficos e tabelas destacando os Motivos CpG identificados, a distribuição dos motivos no genoma e os motivos presentes em regiões codificantes. Apesar da inspiração para o desenvolvimento do *software* ter surgido por meio de estudos envolvendo bactérias da família *Anaplasmataceae*, o *software* pode ser utilizado para analisar genomas bacterianos em geral.

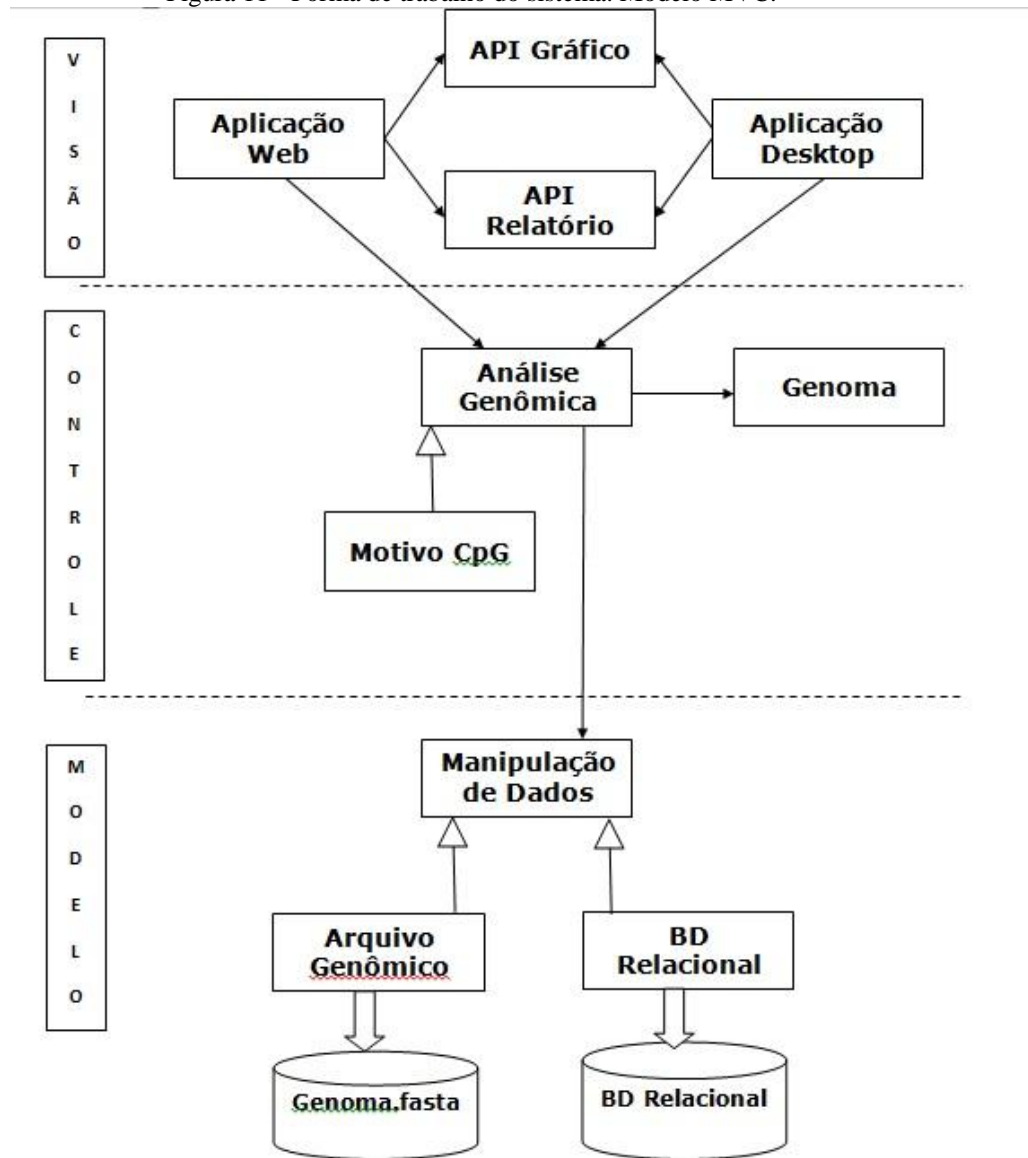
5.2.1 *CpG Highlight*

Com o objetivo de ser multiplataforma, *CpG Highlight* foi desenvolvido utilizando a linguagem Java e a biblioteca JFreeChart (GILBERT DAVID, 2008), uma biblioteca com código-fonte aberto para criação de gráficos de diversos tipos.

A implementação seguiu o modelo MVC (Modelo-Visão-Controle), um padrão de projeto de *software* que separa as classes que envolvem a apresentação dos dados (interação com o usuário), a lógica da aplicação (que englobam as técnicas de análise genômica abordadas) e classes de manipulação de arquivos. Desta forma, evoluções do *software* tornam-se mais ágeis, sendo possível alterar e disponibilizar novas formas de interface com o usuário, incluir novas regras de análise dos dados e permitir a interpretação de novos tipos de arquivos de entrada.

A partir do estudo envolvendo a análise de Motivos CpG, foi modelado o sistema proposto, conforme a Figura 11.

Figura 11 - Forma de trabalho do sistema. Modelo MVC.



Fonte: Autor.

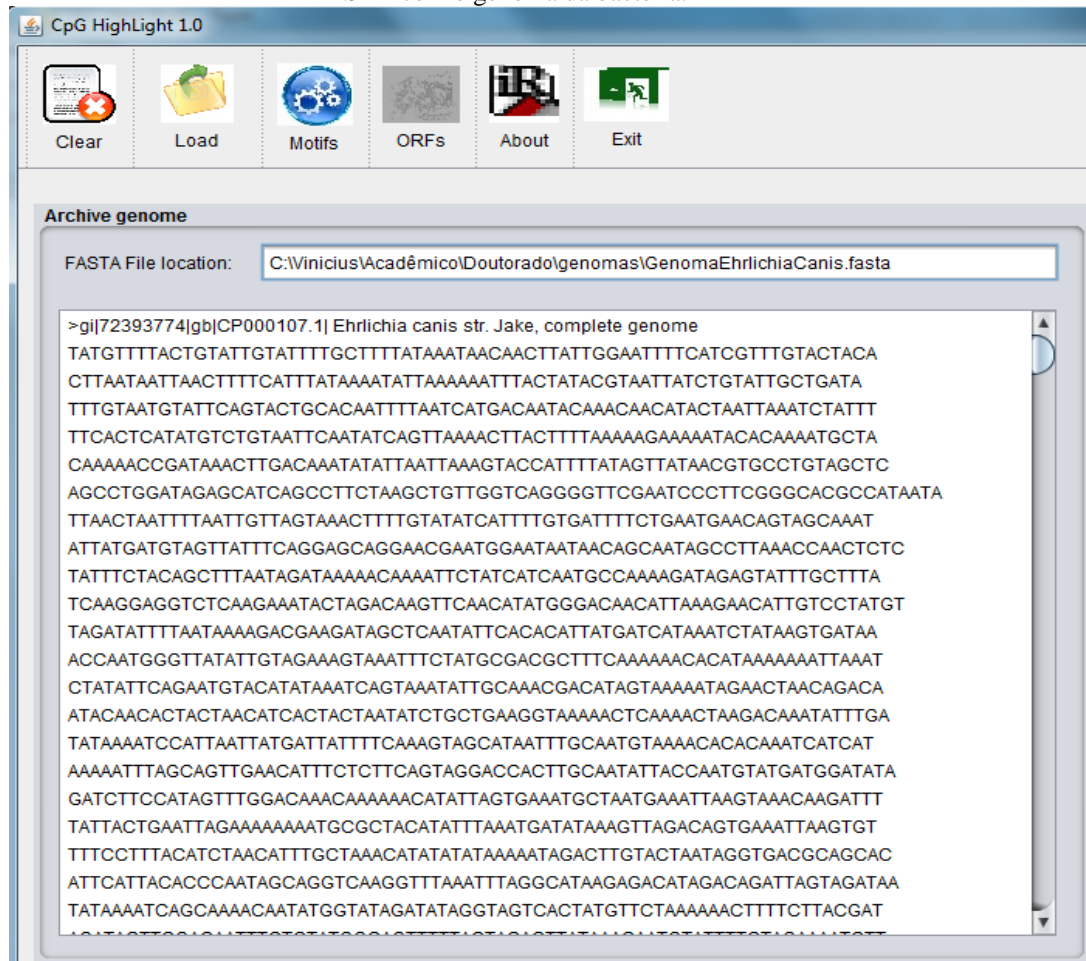
Ao ser executado, a tela inicial do *CpG Highlight* é apresentada, como mostra a **Figura 12**.

Inicialmente, o arquivo contendo o genoma completo da bactéria no formato FASTA deve ser carregado. Após o carregamento do arquivo, *CpG Highlight* identifica características do genoma como ilhas CG, o número de nucleotídeos e os Motivos CpG conforme arranjo proposto por Mutwiri (2003). É feita a contagem dos motivos identificados e a localização dos mesmos no genoma é armazenada. Com os dados coletados, o software disponibiliza a geração de gráficos que permitem visualizar o número de nucleotídeos, a proporção dos Motivos identificados e a distribuição dos Motivos ao longo do genoma utilizando um gráfico

de coordenadas cartesianas. A tela com a descrição dos motivos e os botões para geração dos gráficos pode ser vista na

Figura 13.

Figura 12 - Tela principal do *CpG Highlight*. O usuário deve inicialmente carregar (botão "Load") o arquivo FASTA com o genoma da bactéria.

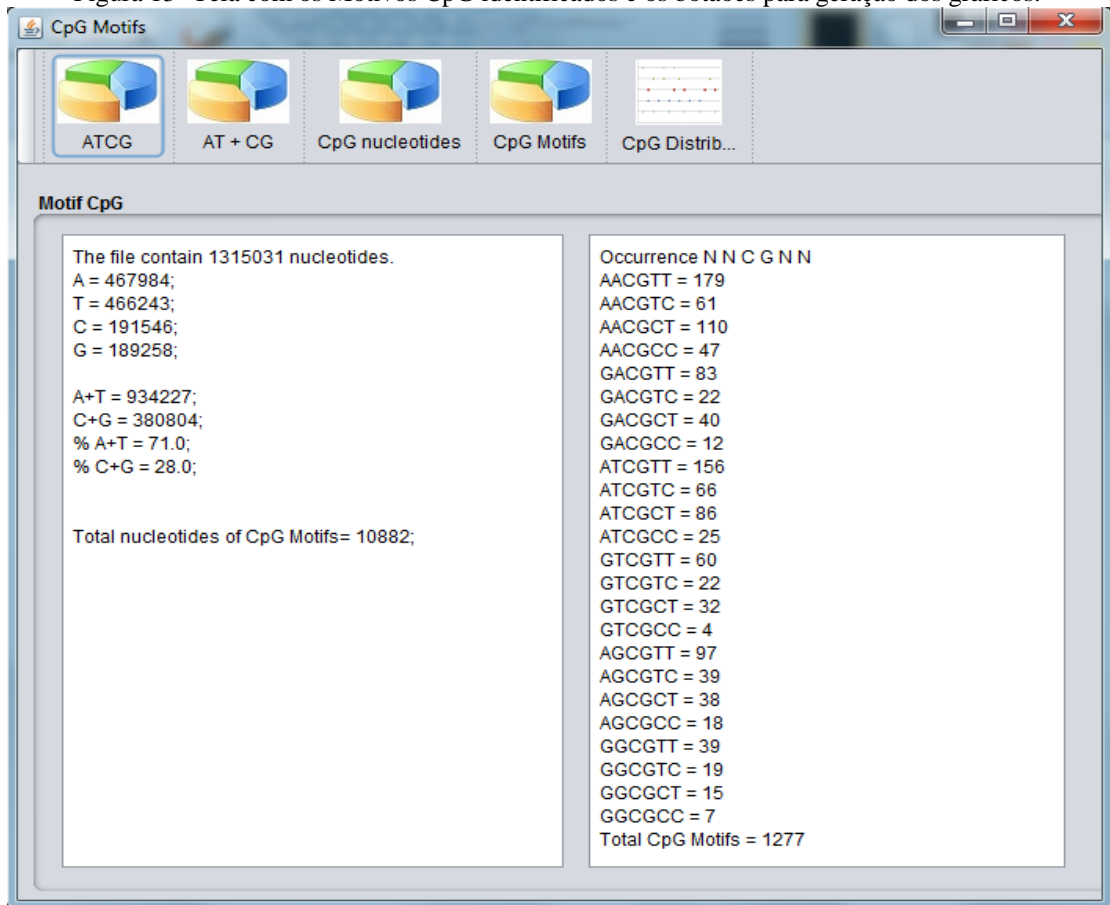


Fonte: Autor.

Depois da identificação dos Motivos CpG, o usuário pode carregar um arquivo contendo as regiões codificantes (arquivo com extensão .ffn) do genoma da bactéria. O software gera uma lista com os ORFs do arquivo e identifica a presença dos motivos presentes nos ORFs, criando uma relação entre eles e disponibilizando as tabelas da Figura 14.

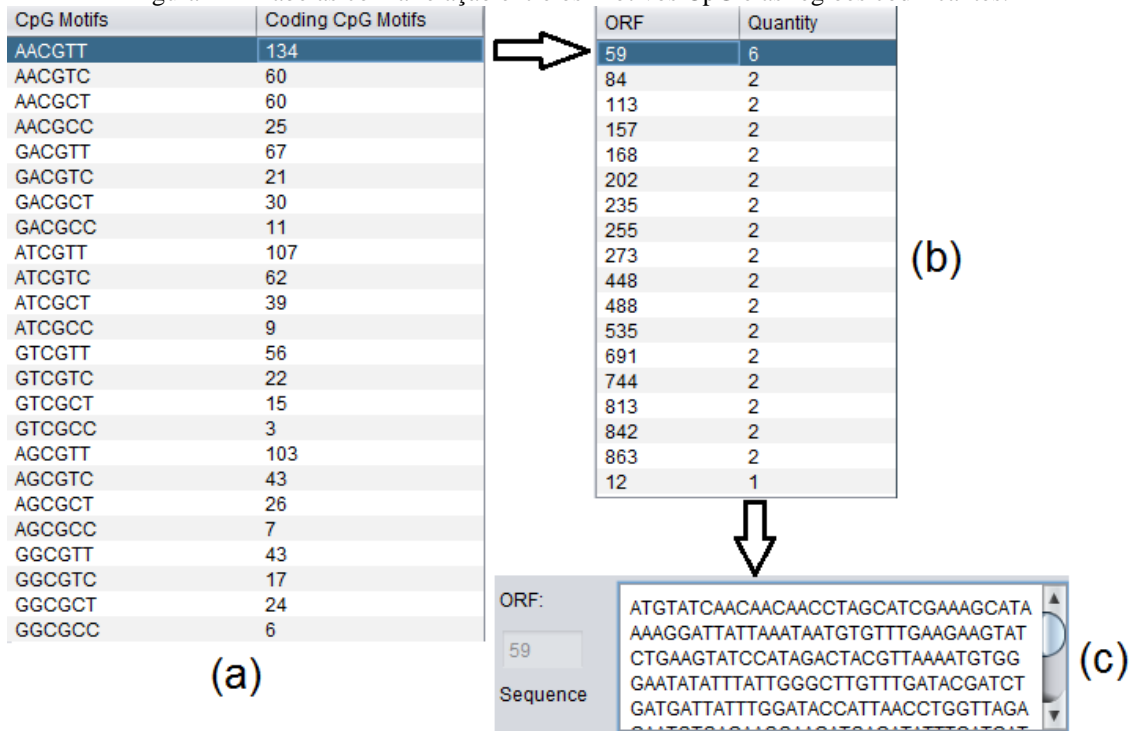
Primeiramente, o software exibe a tabela com a relação de Motivos CpG codificantes (a). Ao selecionar um motivo, como por exemplo AACGTT, que é o motivo que mais ocorre em regiões codificantes no exemplo, uma outra tabela (b) especifica em quais ORFs o motivo selecionado ocorre e o número de vezes que o mesmo aparece em cada ORF. Selecionando um ORF, a sequência de nucleotídeos é exibida (c).

Figura 13- Tela com os Motivos CpG identificados e os botões para geração dos gráficos.



Fonte: Autor.

Figura 14 - Tabelas com a relação entre os Motivos CpG e as regiões codificantes.



Fonte: Autor.

5.2.1.1 Algoritmos

O algoritmo de identificação de Motivos CpG baseia-se no arranjo de seis nucleotídeos proposto por Mutwiri (2003): $5' \rightarrow X1X2 \text{ CpG } Y1Y2 \rightarrow 3'$, em que : X1 é uma Purina, X2 é uma Purina ou T, Y1 e Y2 são Pirimidinas.

O algoritmo do *CpG Highlight* distribui os nucleotídeos do genoma em um vetor. O vetor é percorrido para identificar, quantificar e armazenar a localização dos Motivos CpG no genoma.

Sobre a análise de motivos codificantes, o software faz a leitura do arquivo com os ORFs gerando uma lista com cada um dos ORFs identificados. Cada item da lista é composto pelos dados dos ORFs, sendo um número sequencial de identificação, a sequência de nucleotídeos, e a posição de início e fim no genoma. O software percorre a lista identificando e vinculando os motivos CpG aos ORFs em que eles ocorrem.

Para a criação dos gráficos, o software cria um *dataset* com informações necessárias e aciona métodos do framework *JFreeChart*.

5.3 MÉTODOS UTILIZADOS NA ANÁLISE DE TRs

Como já discutido, existem diversos programas disponíveis para identificação de TRs em uma sequência de nucleotídeos. De um modo geral, estes programas calculam um *score* linear para os TRs identificados baseado no comprimento de cada *repeat*, na conservação do *repeat* ao longo do TR e no número de unidades de *repeats* que o TR possui. Legendre et al. (2007) propôs um modelo não-linear de algoritmo que produz um *score*, denominado VARSCORE, que pode prever a probabilidade de um TR variar em relação ao número de *repeats* entre diferentes estirpes da mesma espécie. A pontuação VARSCORE maior do que 1 indica uma alta probabilidade de que o TR pode variar entre diferentes estirpes ou isolados de uma espécie em particular (LEVDANSKY et al., 2008). Por esta característica, o *software* foi utilizado, visto que o fator variabilidade genética é uma ótica de análise abordada no presente trabalho.

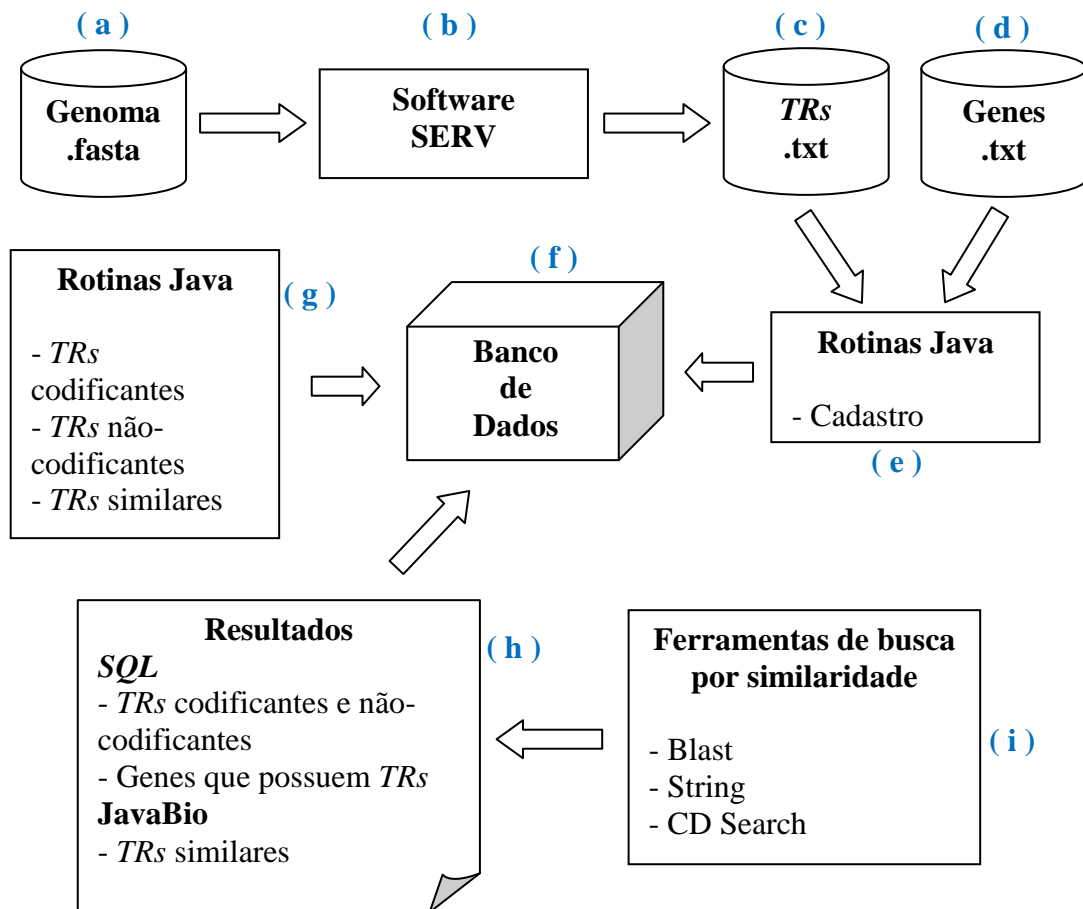
Com a escolha o *software* de identificação dos TRs e baseado nos objetivos do trabalho, foram elaboradas as seguintes perguntas para nortear os métodos para a análise:

1. Quantos e quais TRs estão em regiões codificantes (ORFs)?
2. Quais genes possuem TRs?
3. Quais TRs apresentam-se relevantes em relação à variabilidade e a fatores envolvidos na patogenia bacteriana?

4. *TRs* dos genes TRP36, GP36 e GP47 (genes muito utilizados na análise de variabilidade entre organismos da mesma espécie) estão presentes na análise realizada?
5. Quais *TRs* são comuns nas três bactérias pesquisadas? Esses *TRs* são relevantes?
6. Existem Motivos CpG dispostos em *tandem*?

Em seguida, foram realizadas as atividades do fluxograma apresentado na Figura 15 para os dados genômicos de cada uma das bactérias.

Figura 15- Fluxograma dos métodos aplicados na análise de *TRs*.



Fonte: Autor.

Inicialmente, foram submetidos ao *software* SERV o genoma completo de cada uma das três bactérias pesquisadas (a). SERV processou o genoma identificando os *TRs* e atribuindo como parâmetro diversos valores *score* (discutidos no capítulo 5), localização no genoma, sequência consenso e sequência de nucleotídeos que compõem cada um dos *TRs*. Em seguida, o resultado do *software* SERV foi salvo em um arquivo texto. Cada linha do arquivo corresponde a um *TR*, sendo que os parâmetros foram separados por tabulação (c). Em relação as áreas codificantes, foram obtidos arquivos texto com os ORFs e suas anotações diretamente do GenBank e do Entrez (d).

Para manipulação e correlação dos dados, foi desenvolvida uma aplicação Java utilizando a ferramenta NetBeans. Foram desenvolvidos métodos para fazer a leitura dos arquivos com *TRs* e genes e realizar a gravação dos dados (e) no banco de dados criado conforme o modelo entidade-relacionamento exibido na Figura 16 (f). Em seguida foram criados métodos para relacionar no banco de dados os *TRs* codificantes com seus respectivos genes e identificar *TRs* similares entre as 3 bactérias (h).

Para identificação de *TRs* similares entre as três bactérias, foram utilizados os métodos de alinhamento da biblioteca JavaBIO. Foi feito um alinhamento global entre os *TRs* e depois entre a sequência consenso dos *TRs* das três bactérias.

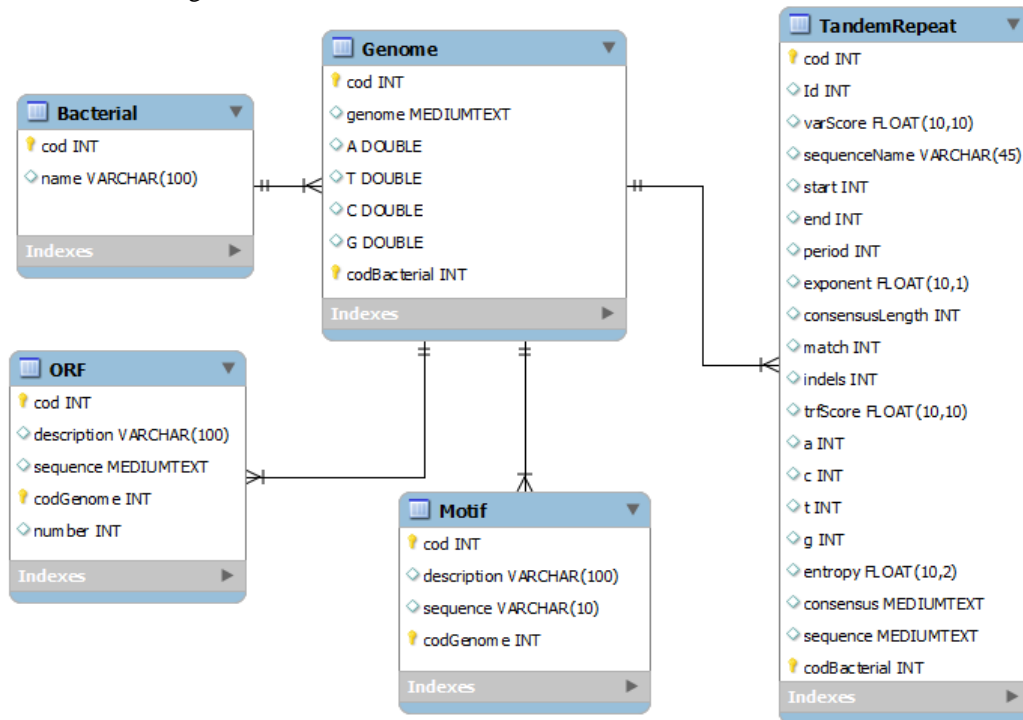
Para discussão dos resultados mais relevantes, foram utilizadas informações da literatura, bancos de dados públicos e *softwares* de busca por similaridade como: *Blast*, *String* e *CD-Search*.

O banco de dados do *String* proporciona uma perspectiva global de associações conhecidas e previstas, resultando em redes de genes que cobrem mais de 1100 organismos. Ele permite mineração automatizada de textos científicos para obter informações e fornece aos usuários informações estatísticas sobre qualquer enriquecimento funcional observado em suas redes (FRANCESCHINI et al, 2013).

A ferramenta de bioinformática *CD-Search* permite analisar genes, utilizando sua sequência de aminoácidos/nucleotídeos ou nome visando identificar padrões. Os padrões são sequências que determinam um domínio conservado. Cada domínio identificado é responsável por executar uma determinada função. O *CD-Search* utiliza o banco de dados de domínio conservado (CDD). O CDD trabalha com modelos de domínio importados de várias bases de dados externas como PFAM, SMART, COG, PRK e TIGRFAM (MARCHLER-BAUER et al., 2011).

Os programas *Blast* são os instrumentos de alinhamento mais populares entre os pesquisadores. Identificado um gene, os pesquisadores podem alinhar a sequência encontrada e compará-la com outras sequências. O *Blast* está disponível para ser usado na Internet. *Blast* utiliza o GenBank, que é um banco de dados referência de sequências, para obter os dados para suas análises. (WATSON et al., 2008).

Figura 16 - Modelo Entidade-Relacionamento do banco de dados.



Fonte: Autor.

5.3.1 Análise laboratorial da variação de TRs

Para confirmar se a região do TR 468 (TR com maior VARSCORE, discutido no próximo capítulo) encontrado no genoma da *E. canis* varia entre amostras de diferentes localizações geográficas conforme predito pelo *software* SERV, foram gerados *primers* para amplificar a região do TR 468 em amostras de sangue de cães infectados de diferentes regiões. Os *primers* estão descritos a seguir:

Tabela 2 - Sequência de nucleotídeos dos “primers” de PCR para amplificação da repetição TR-468

<i>Primer</i>	Sequência	Posição ^a
REP468fwd-2	GGAAACTTTAGCAATGAAAGTAATCCATG	690817 - 690845
REP468rev-1	GCCTGATACAATGCGTGATAAGAAGC	691841 – 691866

^a posição de acordo com sequência de DNA de *E. canis* (Linhagem Jake) GenBank no. CP000107.1.

6 RESULTADOS

Após a submissão dos arquivos contendo dados do genoma das três bactérias (detalhados na Tabela 3) ao software *CpG Highlight* e aos métodos implementados para análise de TRs (detalhados na Figura 15), foram encontrados os resultados descritos nos itens a seguir.

Tabela 3 - Comparação das propriedades genômicas das 3 bactérias.

Bactéria	N Bases do genoma	% G+C	N ORFs	N Bases codificantes	% G+C	% bases codificantes / total de bases
<i>E. canis</i>	1.315.030	35	925	948.282	30	72
<i>E. chaffeensis</i>	1.176.248	34	1105	936.276	31	79
<i>A. phagocytophilum</i>	1.471.282	29	1264	1.004.289	42	68

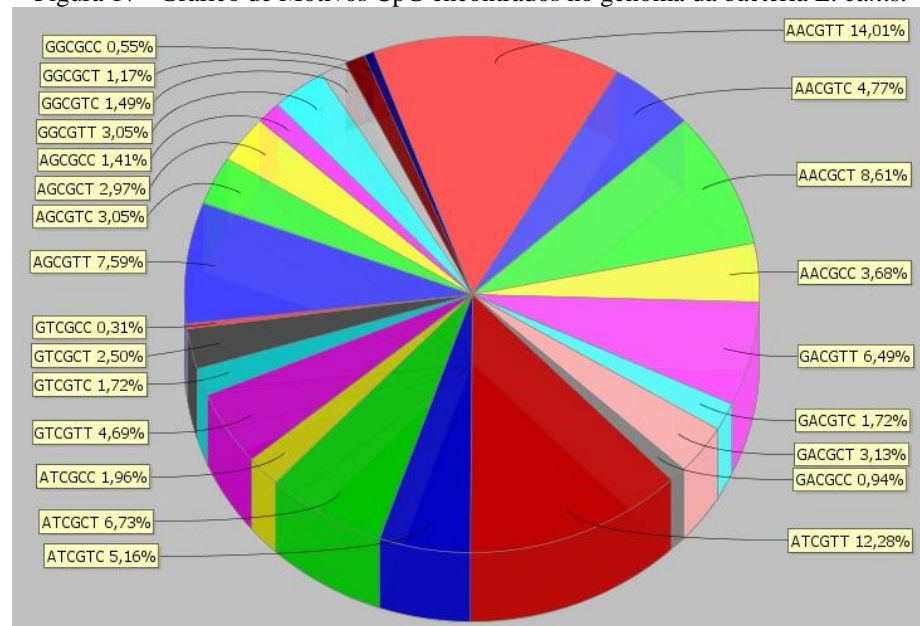
Fonte: Autor

6.1 ANÁLISE DE MOTIVOS CPG

Os resultados da análise de motivos CpG consistem em gráficos e tabelas gerados pelo software *CpG Highlight* que destacam os Motivos CpG identificados no genoma e também os Motivos que ocorrem em regiões codificantes. *CpG Highlight* foi desenvolvido visando a análise de Motivos CpG em bactérias da família *Anaplasmataceae*, porém, vale ressaltar que o *software* pode ser utilizado para análise de genomas bacterianos em geral.

Ao carregar o arquivo *fasta* com o genoma completo da bactéria, o software disponibiliza as funções para a criação de gráficos *pizza* sobre o número de nucleotídeos e sobre a quantificação dos motivos. O gráfico da Figura 17 mostra a proporção de cada um dos motivos identificados para a *E. canis*.

Figura 17 - Gráfico de Motivos CpG encontrados no genoma da bactéria *E. canis*.

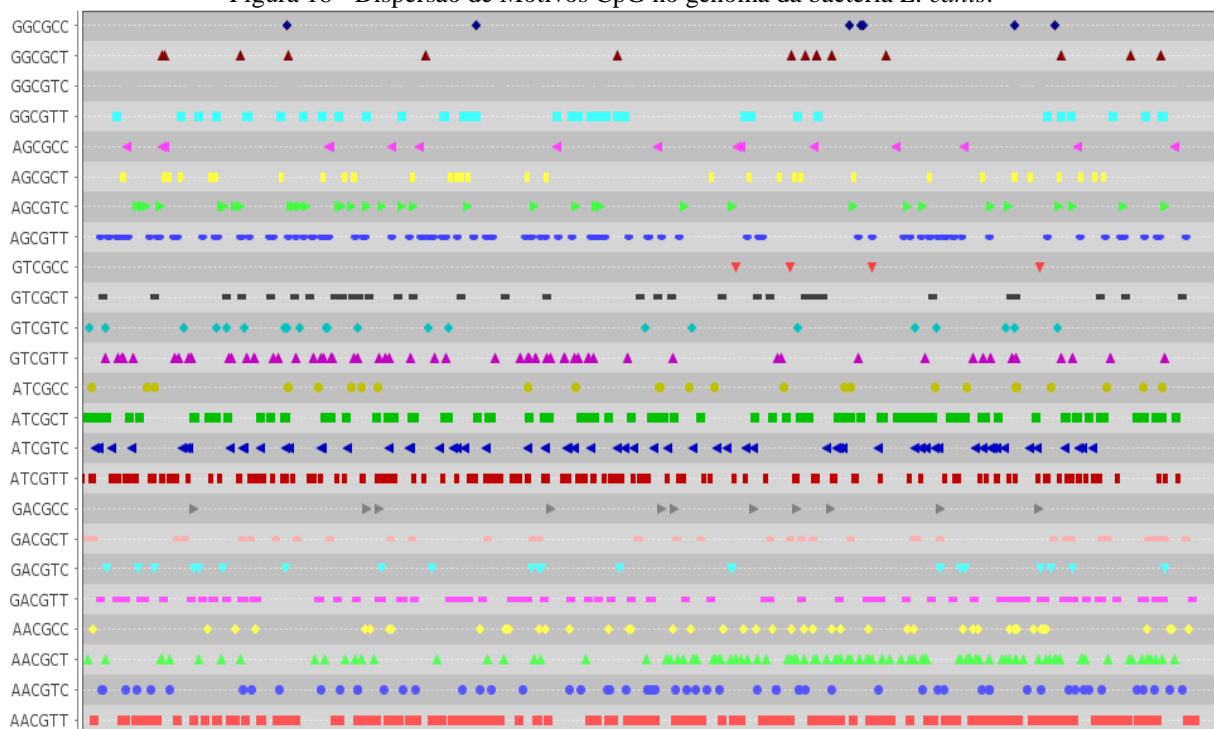


Fonte: Autor.

O software gera um gráfico de coordenadas artesanais que exibe a dispersão dos Motivos identificados ao longo do genoma da bactéria. A

Figura 18 mostra a dispersão dos motivos (eixo Y) no genoma da *E. canis*, da base 1 à base 1.315.030 (eixo X).

Figura 18 - Dispersão de Motivos CpG no genoma da bactéria *E. canis*.



Fonte: Autor.

A Tabela 4 mostra o número total de cada um dos 24 tipos de Motivos identificados e também a quantidade de Motivos CpG codificantes. Para cada um dos motivos, é possível identificar em quais genes eles ocorrem.

Tabela 4 - Relação de Motivos CpG da bactéria *E. canis*.

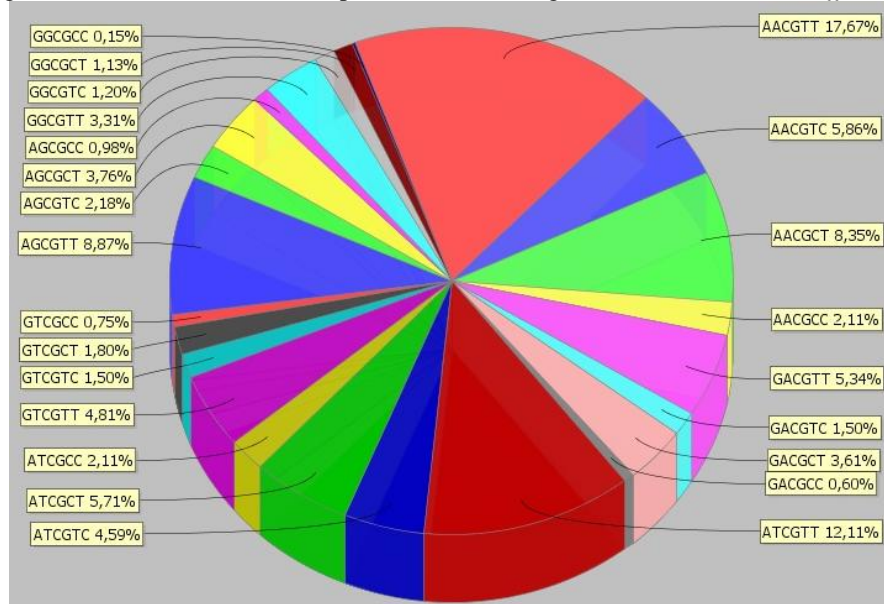
Motivos	Total	Codificantes	Não-codificantes	% Motivos codificantes
AACGTT	179	134	45	74,86
ATCGTT	156	120	36	76,92
AACGCT	110	80	30	72,73
AGCGTT	97	83	14	85,57
ATCGCT	86	71	15	82,56
GACGTT	83	74	9	89,16
ATCGTC	66	62	4	93,94
AACGTC	61	53	8	86,89
GTCGTT	60	51	9	85,00
AACGCC	47	34	13	72,34
GACGCT	40	37	3	92,50
AGCGTC	39	36	3	92,31
GGCGTT	39	34	5	87,18
AGCGCT	38	26	12	68,42
GTCGCT	32	26	6	81,25
ATCGCC	25	18	7	72,00
GACGTC	22	21	1	95,45
GTCGTC	22	19	3	86,36
GGCGTC	19	18	1	94,74
AGCGCC	18	17	1	94,44
GGCGCT	15	14	1	93,33
GACGCC	12	10	2	83,33
GGCGCC	7	6	1	85,71
GTCGCC	4	4	0	100,00
Total	1277	1048	229	82,07

Fonte: Autor.

Para a bactéria *E. chaffeensis*, foi repetido o procedimento utilizado com os dados do genoma e dos genes de *E. canis*. O gráfico da

Figura 19 exibe a proporção dos Motivos CpG identificados.

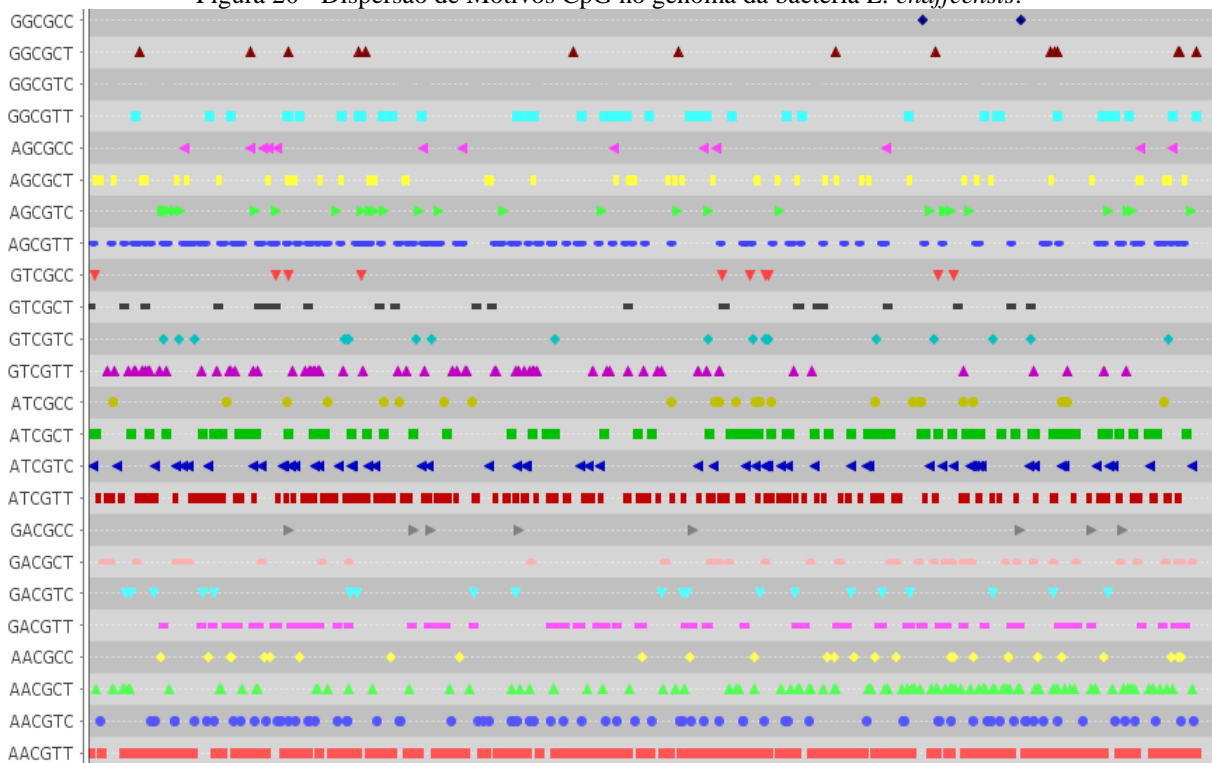
Figura 19 - Gráfico de Motivos CpG encontrados no genoma da bactéria *E. chaffeensis*.



Fonte: Autor.

Os motivos identificados estão dispersos no genoma da bactéria *E. chaffeensis* conforme gráfico da Figura 20.

Figura 20 - Dispersão de Motivos CpG no genoma da bactéria *E. chaffeensis*.



Fonte: Autor.

A Tabela 5 mostra o número total de cada um dos 24 tipos de Motivos identificados e também a relação de Motivos CpG codificantes x número total de motivos.

Tabela 5 - Relação de Motivos CpG da bactéria *E. chaffeensis*

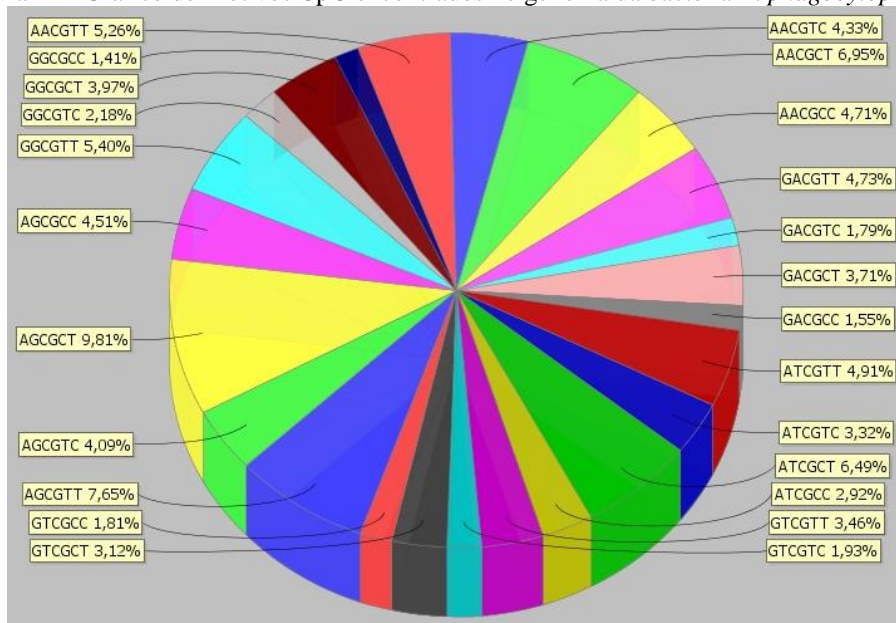
Motivos	Total	Codificantes	Não-codificantes	% Motivos codificantes
AACGTT	235	182	53	77,45
ATCGTT	161	133	28	82,61
AGCGTT	118	96	22	81,36
AACGCT	111	86	25	77,48
AACGTC	78	67	11	85,90
ATCGCT	76	60	16	78,95
GACGTT	71	62	9	87,32
GTCGTT	64	50	14	78,13
ATCGTC	61	52	9	85,25
AGCGCT	50	43	7	86,00
GACGCT	48	46	2	95,83
GGCGTT	44	38	6	86,36
AGCGTC	29	23	6	79,31
AACGCC	28	24	4	85,71
ATCGCC	28	24	4	85,71
GTCGCT	24	20	4	83,33
GACGTC	20	20	0	100,00
GTCGTC	20	19	1	95,00
GGCGTC	16	16	0	100,00
GGCGCT	15	13	2	86,67
AGCGCC	13	13	0	100,00
GTCGCC	10	6	4	60,00
GACGCC	8	7	1	87,50
GGCGCC	2	2	0	100,00
Total	1330	1102	228	82,86

Fonte: Autor.

Para a bactéria *A. phagocytophilum*, CpG Highlight retornou o gráfico de proporção dos Motivos CpG dos Motivos CpG conforme

Figura 21.

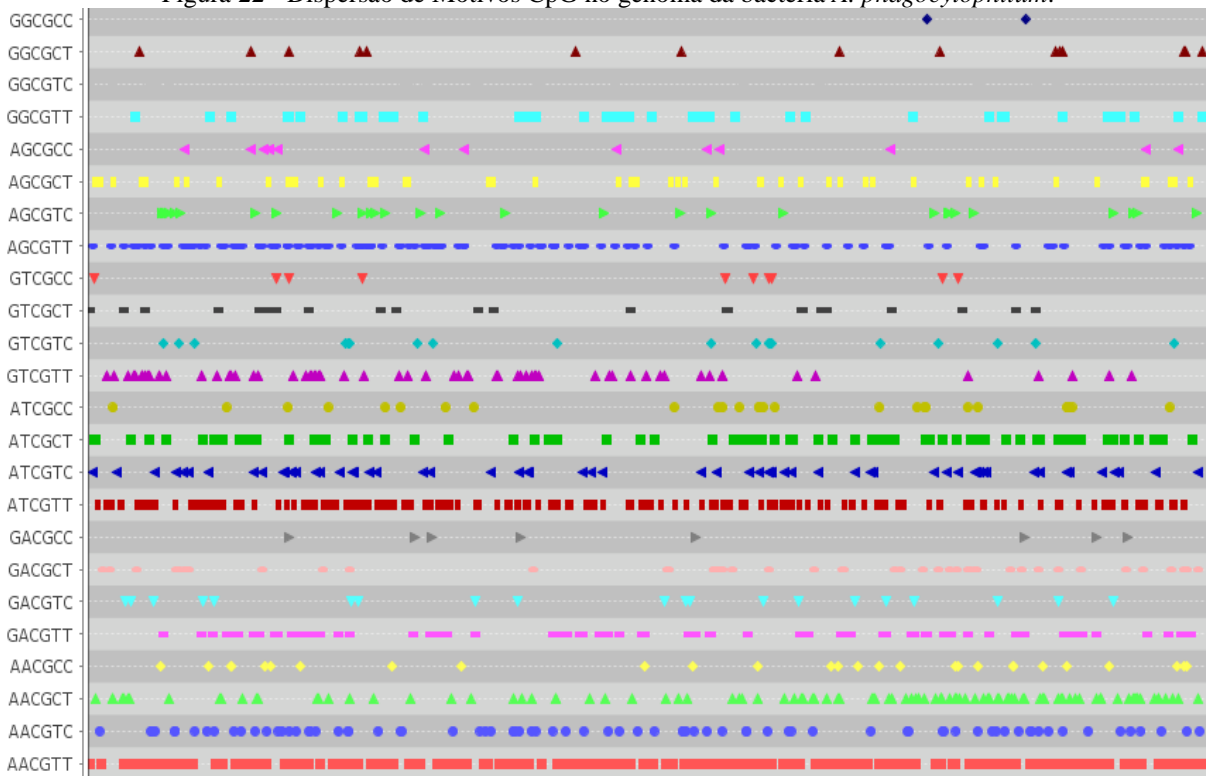
Figura 21 - Gráfico de Motivos CpG encontrados no genoma da bactéria *A. phagocytophilum*



Fonte: Autor.

Os motivos identificados estão dispersos no genoma da bactéria *A. phagocytophilum* como mostra o gráfico da Figura 22.

Figura 22 - Dispersão de Motivos CpG no genoma da bactéria *A. phagocytophilum*.



Fonte: Autor.

A Tabela 5 mostra o número total de cada um dos 24 tipos de Motivos identificados e também a quantidade de Motivos CpG codificantes com a respectiva porcentagem do número de motivos codificantes em relação ao número total de motivos encontrados.

Tabela 6 - Relação de Motivos CpG da bactéria *A. phagocytophilum*

Motivos	Total	Codificantes	Não-codificantes	% Motivos codificantes
AGCGCT	494	353	141	71,46
AGCGTT	385	234	151	60,78
AACGCT	350	231	119	66,00
ATCGCT	327	227	100	69,42
GGCGTT	272	175	97	64,34
AACGTT	265	176	89	66,42
ATCGTT	247	172	75	69,64
GACGTT	238	174	64	73,11
AACGCC	237	149	88	62,87
AGCGCC	227	159	68	70,04
AACGTC	218	153	65	70,18
AGCGTC	206	151	55	73,30
GGCGCT	200	144	56	72,00
GACGCT	187	138	49	73,80
GTCGTT	174	119	55	68,39
ATCGTC	167	140	27	83,83
GTCGCT	157	115	42	73,25
ATCGCC	147	111	36	75,51
GGCGTC	110	91	19	82,73
GTCGTC	97	65	32	67,01
GTCGCC	91	71	20	78,02
GACGTC	90	64	26	71,11
GACGCC	78	66	12	84,62
GGCGCC	71	55	16	77,46
Total	5035	3533	1502	70,17

Fonte: Autor.

6.2 ANÁLISE DE TRS

A Tabela 6 mostra o número de TRs encontrados em cada um dos genomas analisados pelo SERV (LEGENDRE, 2007). Cabe ressaltar que o *software* SERV utiliza internamente o *Tandem Repeats Finder* (TRF) (BENSON, 1999) para identificar os TRs. Após a identificação, SERV atribui valores a parâmetros de variabilidade aos TRs encontrados.

Tabela 6 - Número de TRs encontrados utilizando o SERV (LEGENBRE, 2007).

<i>TRs</i>		
<i>E. canis</i>	<i>E. chaffeensis</i>	<i>A. phagocytophilum</i>
901	740	245

Fonte: Autor.

A Tabela 7 mostra para fins de exemplificação o resultado do *software* SERV para os dois TRs mais significativos (em relação ao parâmetro VARSCORE) encontrados no genoma da bactéria *E. canis*. Os campos da tabela são:

- *Id*: número de identificação do *Tandem Repeat*;
- *VARSCORE*: índice que indica o grau variabilidade do *Tandem Repeat* entre organismos de mesma espécie;
- *Start*: posição de início do *Tandem Repeat* no genoma;
- *End*: posição de fim do *Tandem Repeat* no genoma;
- *TRFSCORE*: índice do *Tandem Repeat*, levando-se em consideração o tamanho, número de *repeats* e conservação dos *repeats*.
- *Consensus*: base comum do *repeat*, ou seja, bases que se repetem em *tandem* (de forma adjacente).
- *Sequence*: bases que compõem o TR.

Tabela 7- TRs em destaque encontrados no genoma da *E. canis*.

<i>Id</i>	<i>VAR score</i>	<i>Start</i>	<i>End</i>	<i>TRF score</i>	<i>Consensus</i>	<i>Sequence</i>
468	2.87	690884	691809	1693	TAGTATCTTGTAATT TACTTTTATCTTCAA CTTTTATTACAGGA AGTTTAGAGTCATC ACGCTTCTTATCTT	TAGTATCTTGTAATTTACTTTTAT CTTCAACTTTTATTACAGGAAGT TTAGAGTCATCACGCTTCTTATC TT TAGTATCTTGTAATTTACTTTTAT CTTCAACTTTTATTACAGGAAGT TTAGAGTCATCACGCTTCTTATC TT...
543	2.83	769335	769907	1071	GAACTTACACTTGA AGTATCACTATCATT ATTATACGGCTCTT CTTCTTT AGGTTGATAACTAT CTTCA	GAACTTACACTTGAAGTATCACT ATCATTATTATACGGCTCTTCTT CTTT AGGTTGATAACTATCTTCA GAACTTACACTTGAAGTATCACT ATCATTATTATACGGCTCTTCTT CTTT AGGTTGATAACTATCTTCA...

Fonte: Autor.

Para visualizar a dispersão dos *TRs* ao longo dos genomas foi desenvolvida uma funcionalidade que gera um gráfico e exibe a distribuição dos *TR* com diferentes valores para os parâmetros TRFSCORE e VARSCORE.

Os *TRs* destacados na imagem serão caracterizados posteriormente mediante consultas a base de dados e trabalhos correlatos publicados na literatura.

As figuras 23 e 24 mostram a dispersão de todos os *TRs* identificados ao longo do genoma das três bactérias.

As perguntas 1, 2 e 3 formuladas para direcionar o estudo e a análise de *TRs* questionavam sobre *TRs* não-codificantes, *TRs* codificantes, genes que possuem *TRs* e genes relevantes em relação à variabilidade e patogenicia. Em relação à variabilidade e patogenicia, foram considerados os parâmetros VARSCORE e TRFSCORE.

TRs codificantes foram relacionados aos genes que em eles ocorrem por meio de uma relação no banco de dados. Desta forma, os resultados apresentados nos itens sobre *TRs* foram adquiridos com a execução de consultas SQL ao banco da dados.

A Tabela 8 exibe o resultado dos *TRs* codificantes (intragênicos) identificados.

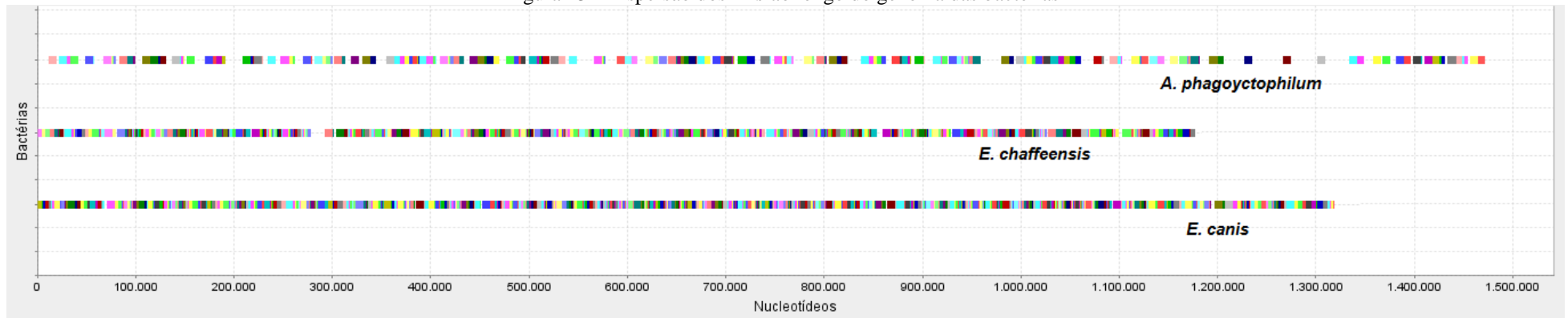
Tabela 8 - *TRs* x *TRs* codificantes

Bactérias	<i>TRs</i>		
	<i>TRs</i> identificados	<i>TRs</i> codificantes	% <i>TRs</i> codificantes
<i>E. canis</i>	901	384	42,62
<i>E. chaffeensis</i>	740	389	52,57
<i>A. phagocytophilum</i>	245	142	57,96

Fonte: Autor.

Os próximos itens mostram os resultados considerados mais relevantes da análise, levando-se em consideração os parâmetros VARSCORE e TRFSCORE.

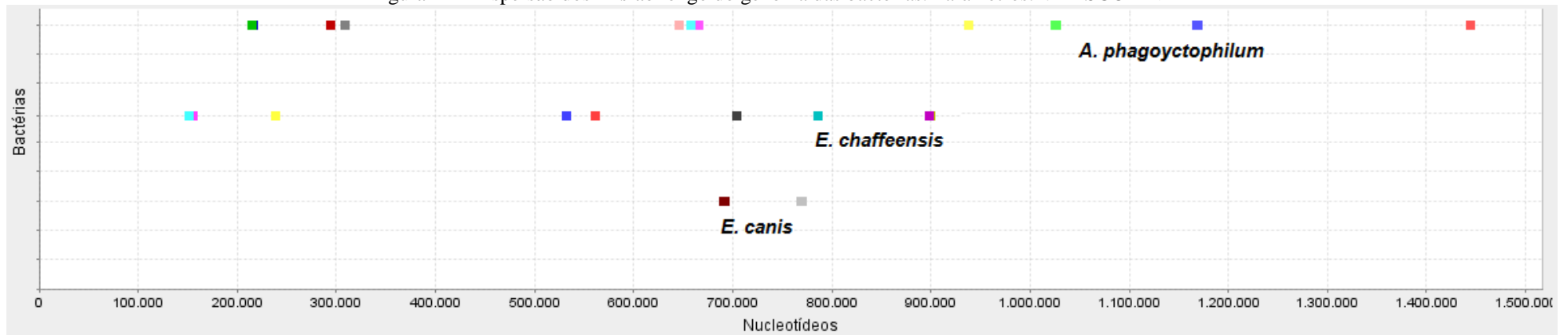
Figura 23 - Dispersão dos TRs ao longo do genoma das bactérias



Fonte:Autor.

O próximo gráfico exibe a dispersão dos TRs com parâmetro VARSCORE maior que 1.

Figura 24 - Dispersão dos TRs ao longo do genoma das bactérias. Parâmetros: VARSCORE > 1



Fonte:Autor.

6.2.1 TRs Codificantes com VARSCORE >1

Conforme exposto por Legendre et al (2007), TRs com VARSCORE maior que 1 são responsáveis por gerar variabilidade entre as bactérias. A Tabela 9 destaca que, para as bactérias *E. canis* e *E. chaffeensis*, uma alta porcentagem de TRs com VARSCORE maior que 1 encontra-se em regiões codificantes.

Tabela 9 - TRs da bactéria *E. canis* com VARSCORE maior que 1.

Bactéria	TRs VARSCORE >=1	TRs codificantes com VARSCORE >=1	%
<i>E. canis</i>	2	2	100
<i>E. chaffeensis</i>	10	9	90
<i>A. phagocytophilum</i>	11	4	36,4

Fonte:Autor.

As próximas tabelas mostram os TRs codificantes com VARSCORE >1 juntamente com seus respectivos genes e parâmetros.

Tabela 10 - TRs codificantes da bactéria *E. canis* com VARSCORE maior que 1.

Tandem Repeat			Gene	
TR ID	TRF SCORE	VAR SCORE	Gene	Descrição do Gene
468	1693	2,87	Ecaj_0472	hypothetical protein
543	1071	2,83	Ecaj_0530	TrbL/VirB6 plasmid conjugal transfer protein

Fonte:Autor.

Tabela 11 - TRs codificantes da bactéria *E. chaffeensis* com VARSCORE maior que 1.

Tandem Repeat			Gene	
TR ID	TRF SCORE	VAR SCORE	Gene	Descrição do Gene
90	791	3,72	ECH_0166	hypothetical protein
350	1298	2,52	ECH_0558	putative lipoprotein
461	216	1,96	ECH_0695	hypothetical protein
596	358	1,94	ECH_0878	hypothetical protein
591	226	1,46	ECH_0878	hypothetical protein
325	279	1,29	ECH_0531	hypothetical protein
140	158	1,11	ECH_0252	hypothetical protein
139	168	1,07	ECH_0252	hypothetical protein
85	225	1,03	ECH_0159	hypothetical protein

Fonte:Autor.

Tabela 12 - TRs codificantes da bactéria *A. phagocytophilum* com VARSCORE maior que 1.

Tandem Repeat			Gene	
TR ID	TRF SCORE	VAR SCORE	Gene	Descrição do Gene
197	950	3,42	APH_0964	triple helix repeat-containing collagen
33	699	3,02	APH_0217	hypothetical protein
239	176	1,13	APH_1379	hypothetical protein
32	507	1,11	APH_0215	hypothetical protein

Fonte:Autor.

6.2.2 TRs Codificantes com VARSCORE < 1 e com TRFSCORE > 1000

O *software* SERV possui uma limitação, pois para *repeats* com mais de 75 nucleotídeos, *repeats* de 2-10 nucleotídeos com mais de 90 unidades ou *repeats* de tamanho da unidade 11-20 nucleotídeos com mais de 80 unidades, o parâmetro VARSCORE não é calculado corretamente. O *software* emite uma mensagem indicando quando as dimensões do TR extrapolam sua faixa de operação. Para estes casos, recomenda-se o uso do parâmetro TRFSCORE como uma estimativa grosseira de variabilidade. O objetivo ao analisar TRs com alto valor TRFSCORE é, além de avaliar TRs que extrapolam as limitações do *software* SERV, evidenciar TRs significativos e seus respectivos genes, pois, como destacado no capítulo 2, a presença de TRs é uma característica de genes relacionados com a patogenicidade e mecanismos de subversão dos processos celulares do hospedeiro.

As tabelas desta subseção mostram os TRs codificantes com TRFSCORE maior que 1000 e VARSCORE menor que 1, juntamente com os respectivos genes em que os mesmos ocorrem.

Tabela 13 - TRs codificantes da bactéria *E. canis* com VARSCORE menor que 1 e TRFSCORE maior que 1000.

Tandem Repeat			Gene	
TR ID	TRF SCORE	VAR SCORE	Gene	Descrição do Gene
710	2894	-1,82	Ecaj_0716	hypothetical protein
9	2848	-1,43	Ecaj_0017	gp140
49	2725	-1,83	Ecaj_0060	hypothetical protein
539	2006	-1,83	Ecaj_0529	putative Type IV secretory pathway VirB6 components
47	1601	-1,83	Ecaj_0060	hypothetical protein
381	1496	-1,83	Ecaj_0387	ankyrin
46	1383	-1,83	Ecaj_0060	hypothetical protein
194	1287	-1,83	Ecaj_0220	hypothetical protein
706	1060	-1,82	Ecaj_0715	hypothetical protein

Fonte:Autor.

Tabela 14 - TRs codificantes da bactéria *E. chaffeensis* com VARSCORE menor que 1 e TRFSCORE maior que 1000.

Tandem Repeat			Gene	
TR ID	TRF SCORE	VAR SCORE	Gene	Descrição do Gene
287	2912	-1,83	ECH_0499	VirB6 family type IV secretion system protein
16	2044	-1,83	ECH_0039	120 kDa immunodominant surface protein
599	1608	-1,83	ECH_0878	hypothetical protein
605	1554	-1,83	ECH_0879	hypothetical protein
51	1387	-1,83	ECH_0108	hypothetical protein
301	1290	-1,83	ECH_0499	VirB6 family type IV secretion system protein
46	1110	-1,83	ECH_0106	hypothetical protein
646	1093	-1,83	ECH_0945	hypothetical protein

Fonte:Autor.

Tabela 15 - TRs codificantes da bactéria *A. phagocytophilum* com VARSCORE menor que 1 e TRFSCORE maior que 1000.

Tandem Repeat			Gene	
TR ID	TRF SCORE	VAR SCORE	Gene	Descrição do Gene
241	1906	-1,83	APH_1387	HGE-2 protein
93	1802	-1,75	APH_0455	HGE-14 protein
64	1642	-1,83	APH_0377	type IV secretion system VirB6 family protein
38	1498	-1,83	APH_0256	type I secretion system ATPase
67	1487	-1,83	APH_0377	type IV secretion system VirB6 family protein
60	1416	-1,83	APH_0377	type IV secretion system VirB6 family protein
142	1357	-1,83	APH_0720	hypothetical protein
4	1305	-1,83	APH_0032	hypothetical protein
3	1282	-1,06	APH_0032	hypothetical protein
105	1198	-1,83	APH_0546	hypothetical protein

Fonte:Autor.

6.2.3 TRs Não-Codificantes com VARSCORE ≥ 1

Segundo Levdansky et al. (2008), a variabilidade de TRs não codificantes podem regular a expressão de um gene, afetando a transcrição e tradução. A fim de verificar a existência de TRs não codificantes com alta probabilidade de variação, as tabelas a seguir mostram os TRs não-codificantes com VARSCORE maior que 1. A bactéria *E. canis* não possui TRs com estes valores de parâmetros.

Tabela 16 - TRs não-codificantes da bactéria *E. chaffeensis* com VARSCORE maior que 1.

TR ID	VARSCORE
522	1,18

Fonte:Autor.

Tabela 17 - TRs não-codificantes da bactéria *A. phagocytophilum* com VARSCORE maior que 1.

TR ID	VARSCORE
48	2,94
116	2,21
121	2,08
45	1,59
179	1,14
217	1,06
119	1,03

Fonte:Autor.

6.2.4 TRs Não-Codificantes com VARSCORE < 1 e TRFSCORE>1000

As próximas tabelas mostram os TRs não codificantes com VARSCORE menor que 1, mas com um alto valor TRFSCORE.

Tabela 18 - TRs não-codificantes da bactéria *E. canis* com VARSCORE menor que 1 e TRFSCORE maior que 1000.

ID	VARSCORE	TRFSCORE
710	-1,82	2894
9	-1,43	2848
49	-1,83	2725
27	-1,83	2077
539	-1,83	2006
174	-1,83	1958
173	-1,83	1917
26	-1,83	1888
314	-1,83	1789
891	-1,77	1786

Fonte:Autor.

Tabela 19 - TRs não-codificantes da bactéria *E. chaffeensis* com VARSCORE menor que 1 e TRFSCORE maior que 1000.

ID	VARSCORE	TRFSCORE
287	-1,83	2912
16	-1,83	2044
243	-1,83	1785
582	-1,83	1692
599	-1,83	1608
605	-1,83	1554
696	-1,83	1532
691	-1,83	1466
51	-1,83	1387
458	-1,83	1380

Fonte:Autor.

Tabela 20 - TRs não-codificantes da bactéria *A. phagocytophilum* com VARSCORE menor que 1 e TRFSCORE maior que 1000.

ID	VARSCORE	TRFSCORE
123	-1,83	2158
241	-1,83	1906
93	-1,75	1802
64	-1,83	1642
96	-1,83	1629
38	-1,83	1498
67	-1,83	1487
60	-1,83	1416
180	-1,83	1403
52	-1,83	1400

Fonte: Autor.

6.2.5 TRs Similares entre as três Bactérias

Utilizando a função de alinhamento global da biblioteca JAVABIO, foram feitos alinhamentos entre os TRs buscando identificar TRs similares entre as três bactérias, considerando um *score* de 70% de similaridade. Nenhum alinhamento realizado retornou um *score* maior ou igual a 70%. O processo foi repetido utilizando a sequência consenso dos TRs e, novamente, nenhum alinhamento retornou o *score* mínimo estipulado.

6.2.6 Análise *in vitro* de Variabilidade utilizando o TR 468 de *E. canis*

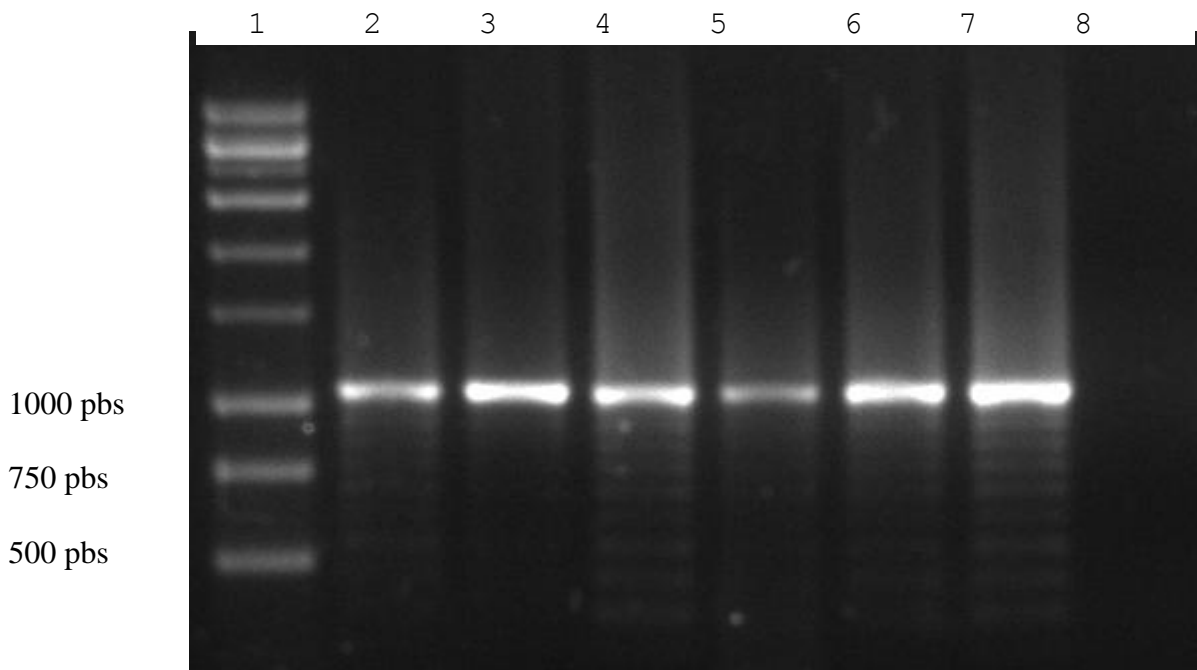
O TR 468 apresentou o maior valor VARSCORE conforme predito pelo software SERV. Desta forma, foi escolhido para uma análise de presença de polimorfismos de DNA que indicassem variabilidade entre diferentes isolados de *E. canis*. Foram realizadas reações de PCR com *primers* REP468fwd-1 e REP468rev-1 direcionados à amplificação de uma região de 1050 pbs do genoma de *E. canis* correspondendo à região englobando o TR 468 – Figura 25 – e DNA genômico de amostras de sangue canino de diferentes regiões geográficas do Brasil e também da cidade de Ribeirão Preto, previamente confirmadas como positivas para infecção com *E. canis* utilizando ensaios de PCR executados em nosso laboratório (SANTOS et al, 2009; OLIVEITA et al., 2009; CARDOSO et al., 2011) - Figura 26.

Figura 25 – Sequência de DNA do genoma de *E. canis* (Linhagem Jake, GenBank: CP000107.1) correspondendo à região englobando o *TR* 468. As regiões dos *primers* REP468fwd-1 e REP468rev-1, englobando região de amplificação com 1050 pbs, está em negrito. As repetições estão sublinhadas.

TATCATCCTCTATTTTT**GGAAC**TTTAGCAAT**GAAAGTAATCCATG**TTCTTTTTTATCACTTTG
 ATTATCATTGCTATAGTCACTAGAAATCTTGTAACCTACTCTTATCTTCAACTTTCTCTACTATT
 GGAAGTTTAGAGTCATCACGCTTCTTATCTTCAGTATCTTGTAATTTACTTTTTATCTTCAACTT
 TACTACAGGAAGTTTAGAGTCATCGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATC
TTCAACTTTTATTACAGGAAGTTTAGAGTCATCACGCTTCTTATCTTTAGTATCTTGTAATTTA
CTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACGCTTCTTATCTTTAGTATCTT
GTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACGCTTCTTATCTTT
AGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCGCGCTTC
 TTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAGTCAT
 CGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTT
 AGAGTCATCACGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACA
GGAAGTTTAGAGTCATCGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTT
TTATTACAGGAAGTTTAGAGTCATCACGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATC
TTCAACTTTTATTACAGGAAGTTTAGAGTCATCGCGCTTCTTATCTTTAGTATCTTGTAATTTA
CTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACGCTTCTTATCTTTAGTATCTT
 GTAATTTACTTTTTATCTTCAACTTTGACTATTGGAAGTTTAGAATCATCAACATTAGGTCCTGA
 AACTCCTACACCAACATG**CTTCTTATCACGCATTGTATCAGGC**ATATCATTTTCTATAGGAACA
 ATAAACTTAGCACCTTCCACAT

Fonte: Autor.

Figura 26 – Foto representativa de eletroforese em gel de agarose de produtos de PCR com primers REP468fwd-1 e REP468rev-1 para amplificação de fragmento de 1050 pbs englobando o *TR* 468. Canaleta 1 – Marcador de tamanho de DNA 1 Kb (Ludwig); Canaletas 2-7 – Produtos de amplificação com DNA genômico de amostras de sangue canino confirmadas como positivas para infecção com *E. canis*. Canaleta 8: controle negativo (água ultra-pura).



Fonte: Autor.

Os produtos de PCR foram sequenciados e comparados entre si. O sequencia do resultado do PCR realizado com uma amostra da cidade de Monte Negro, estado de Rondônia, apresentou

polimorfismos em relação as demais. A Figura 27 mostra a sequência amplificada no nosso experimento e a sequência equivalente da linhagem *E. canis jake*.

Figura 27- Sequência de DNA do genoma de *E. canis* (Linhagem *Jake*, GenBank: CP000107.1) correspondendo à região englobando o *TR* 468 e a sequência de DNA do resultado de PCR da amplificação da região do *TR* 468 de uma amostra da cidade de Monte Negro, estado de Rondônia, Brasil.

Fonte: Autor.

>Linhagem *E. canis Jake*

```
GGAAACTTTAGCAATGAAAGTAATCCATGTTCTTTTTTATCACTTTGATTATCATTGCTATAGTCAC
TAGAATCTTGTAACCTACTCTTATCTTCAACTTTCTCTACTATTGGAAGTTTAGAGTCATCACGCTTCTT
ATCTTCAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAGTCATCGCGCTTC
TTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACGCT
TCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACG
CTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCA
CGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCAT
CGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAGTC
ATCGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAG
TCATCACGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAG
AGTCATCGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTT
AGAGTCATCACGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTT
```

>Linhagem *E. canis Monte Negro*

```
GGAAACTTTAGCAATGAAAGTAATCCATGTTCTTTTTTTTTAACTTTGATTATCATTGCTATAGTCAC
TAGAATCTTGTAACCTACTCTTATCTTCAACTTTCTCTACTATTGGAAGTTTAGAGTCATCACGCTTCTT
ATCTTCAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAGTCATCGCGCTTC
TTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACGCT
TCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCACG
CTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCATCA
CGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTATTACAGGAAGTTTAGAGTCAT
CGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAGTC
ATCGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAGAG
TCATCACGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACAGGAAGTTTAG
AGTCATCGCGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACCAGGAAGTT
TAGAGTCATCACGCTTCTTATCTTTAGTATCTTGTAATTTACTTTTTATCTTCAACTTTTACTACCAGGAAGTT
```

A figura destaca a ocorrência de dois tipos de *repeats* que formam o *TR*, destacados como *repeat* verde e *repeat* azul:

Repeat verde: AGTATCTTGTAATTTACTTTTTATCTTCAACTTTTA^TTACAGGAAGTTTA

Repeat azul: AGTATCTTGTAATTTACTTTTTATCTTCAACTTTTA^CTACAGGAAGTTTA

O *repeat* verde apresenta a base T na posição 36, já o *repeat* azul possui a base C na mesma posição.

Em relação ao *TR* como um todo, a linhagem *Jake* possui 3 *repeats* azuis e 6 *repeats* verdes, enquanto a linhagem Monte Negro possui 4 *repeats* azuis e 4 *repeats* verdes.

7 DISCUSSÃO

Os dados resultantes da execução do software *CpG Highlight* evidenciaram os Motivos CpG identificados com maior frequência nos genomas pesquisados, os Motivos CpG codificantes e também os genes que esses Motivos ocorrem. O banco de dados alimentado com informações do software SERV e com as relações criadas pelas rotinas java (TRs codificantes e os genes em que eles ocorrem) permite destacar TRs instáveis (parâmetro VARSCORE), TRs significativos e instáveis (parâmetro TRFSORE), TRs codificantes, TRs não codificantes, TRs similares e genes que possuem TRs. Neste capítulo, os dados evidenciados serão discutidos e melhor caracterizados utilizando trabalhos existentes na literatura e as ferramentas de busca por similaridade descritas anteriormente: String, Blast, CD-Search.

7.1 MOTIVOS CpG

Muitos Motivos CpG têm sido utilizados em pesquisas com o objetivo de analisar seu potencial como agonista de TLR-9 em diferentes hospedeiros. A Tabela 21 mostra os Motivos CpG mais destacados em pesquisas como potentes agonistas de TLR-9.

Tabela 21 - Motivos CpG destacados como potentes agonistas de TLR-9. Foram destacadas as referências, os hospedeiros e a posição no genoma das bactérias pesquisadas (em relação ao número de ocorrências no genoma).

Motivos CpG	Hospedeiro	Posição (ocorrências)			Referência
		<i>E. canis</i>	<i>E. chaf.</i>	<i>A. phag.</i>	
AACGTT	Ratos, humanos, cães, bovinos, peixes	1	1	6	PASQUINI et al. (1999); UDEN e RAZ (1999); SHODA et al. (2001); KANDIMALLA et al. (2003); REN et al. (2010);
AACGCT	Murinos	3	4	3	KRIEG et al.(1995);
AGCGCT	Cães, ratos	14	10	1	REN et al. (2010); UDEN e RAZ (1999);
ATCGTC	Suíños	7	9	3	KAMSTRUP et al. (2001);
GACGTC	Humanos	17	17	22	UDEN e RAZ (1999);
GACGTT	Humanos, bovinos, coelhos, ratos	6	6	22	UDEN e RAZ (1999); SHODA et al. (2001); KRIEG et al.(2002); MUTWIRI et al. (2003); KRIEG et al.(2002); REN et al. (2010);
GTCGTT	Humanos, cães, gatos, suínos, equinos, caprinos	9	8	15	HARTMANN E KRIEG (2000); KRIEG et al.(2002); MUTWIRI et al. (2003);

7.1.1 Motivos CpG de destaque identificados

Em relação aos resultados da busca por Motivos CpG, os gráficos mostram uma grande similaridade entre o genoma das bactérias *E. canis* e *E. chaffeensis*. Em ambos os genomas, o motivo AACGTT teve o maior número de ocorrências. O motivo AACGTT foi caracterizado em trabalhos anteriores por estimular o sistema imunológico de ratos e humanos, aumentando a produção de citocinas pró-inflamatórias e a proliferação de células B (SATO et al., 1996; UDEN e RAZ, 1999; KANDIMALLA et al., 2003).

Com objetivo de desenvolver novos Motivos CpG que possam ser utilizados como adjuvante de vacina anti-rábica em cães, Ren et al. (2010) realizaram testes com os Motivos GACGTT, GTCGTT, AACGTT, ATCGTC e ATCGAT, observando a capacidade de induzir a proliferação de esplenócitos caninos. O resultado mostrou que, com exceção dos motivos GACGTT e ATCGAT, todos foram capazes de estimular a proliferação esplenócitos caninos, porém o motivo AACGTT obteve melhores resultados, sendo sugerido pelos autores como um potente motivo a ser utilizado como adjuvante de vacinas em cães. Ren et al. (2010) também comentam que existem poucos trabalhos envolvendo motivos CpG e cães.

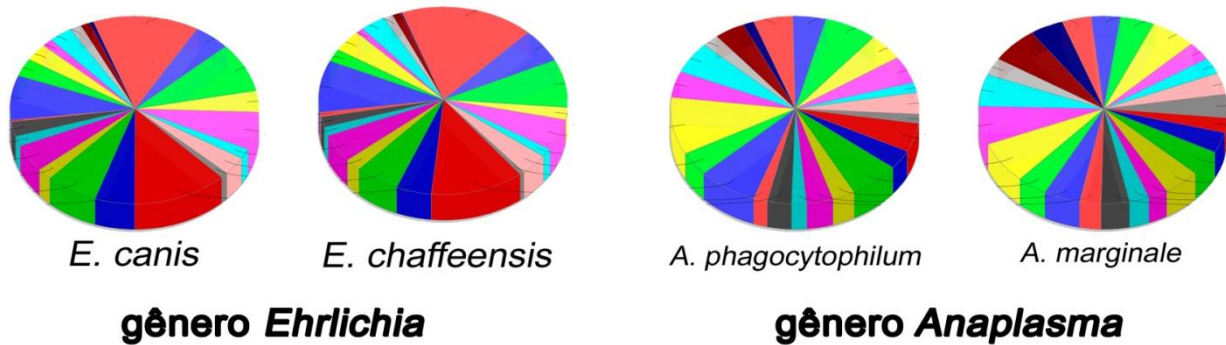
Nossos resultados mostram a alta concentração de motivos AACGTT no genoma da bactéria *E. canis*, principal agente etiológico da Erliquiose Monocítica Canina, justificando a necessidade de pesquisas envolvendo o motivo AACGTT como adjuvante vacinal.

O motivo ATCGTT foi o segundo motivo com mais ocorrências no genoma de *E. canis* e *E. chaffeensis*. Chuang et al. realizaram estudos com ODNs compostos pelos motivos ATCGTT, GACGTT, AACGTT e GTCGTT para avaliar a ativação de respostas imunes mediadas por TLR-9 em coelhos. Resultados mostraram que os motivos AACGTT e GTCGTT são potentes agonistas de TLR-9 para coelhos. Os motivos ATCGTT, GACGTT não ativaram uma resposta.

O genoma da bactéria *A. phagocytophilum* apresenta um número maior de ocorrências de Motivos CpG (5.035), tendo o Motivo AGCGCT aparecendo mais vezes. O motivo AGCGCT foi também caracterizado por estimular o sistema imunológico de ratos (KRIEG et al., 1995; UDEN e RAZ, 1999; REN et al., 2010). AGCGCT (maior ocorrência em *A. phagocytophilum*) também foi relacionado com a ativação de células NK (*Natural Killer*) em cães e ratos por Uden e Raz (1999), porém, não foram encontradas pesquisas recentes utilizando o Motivo.

Em nossas análises utilizando o *CpG HigLight* em genomas das bactérias *Ehrlichia canis*, *Ehrlichia chaffeensis*, *Anaplasma phagocytophilum* e *Anaplasma marginale*, observamos um interessante padrão nos gráficos de proporção, que pode ser visto na Figura 28.

Figura 28 - Padrão dos Motivos CpG identificados para os gêneros *Anaplasma* e *Ehrlichia*.



Fonte: Autor.

É perceptível a semelhança dos gráficos entre as bactérias de mesmo gênero: *Ehrlichia* (*E. canis* e *E. chaffeensis*) e *Anaplasma* (*A. phagocytophilum* e *A. marginale*). Tal característica pode ser interessante uma vez que a detecção de motivos envolve o reconhecimento de padrões moleculares do DNA da bactéria pelo sistema imunológico do hospedeiro. Esta observação permitiu a elaboração de algumas hipóteses que serão analisadas em laboratório em novos trabalhos, como: a utilização de motivos como adjuvantes vacinais pode ser realizada por gêneros? A combinação de motivos mais frequentes no genoma gera um motivo com maior atividade imunoestimulatória? A combinação de motivos de maior ocorrência de diferentes gêneros poderia ser utilizado como adjuvante para toda a família?

Apesar da utilização de motivos CpG como adjuvantes vacinais ser bastante pesquisada, existem ainda muitas questões não esclarecidas. Recentes pesquisas tentam resolver tais questões para definir melhor a identificação, categorização e utilização de motivos CpG agonistas (KRIEG et al., 2012).

7.1.2 Genes com Motivos CpG

Outra importante análise que pode ser feita com os resultados do *CpG Highlight* é a identificação de Motivos CpG concentrados em uma determinada região codificante. Em pesquisas de geneterapia, há evidências de que a presença de Motivos CpG em vetores de expressão de proteínas imunorreativas influenciam na ativação de resposta imunológica (Pasquini et al., 1999; KRIEG, 1999; SCHEULE, 2000; LESINA et al., 2010). Por esta razão, genes com maior ocorrência de motivos em sua formação foram destacados.

Arulkanthan et al. (1999) avaliaram o potencial do gene de membrana MSP1a da bactéria *Anaplasma marginale* para a vacinação contra a anaplasnose bovina. Os resultados demonstraram que a vacina com o plasmídeo com a sequência do gene MSP1a induziu a soroconversão de antígenos específicos em gado. Uma observação feita foi a presença de sete motivos CpG (cinco

GACGTC, um GACGTT, e um AACGTT) na sequência do gene MSP1a. Os autores acreditam que a presença dos Motivos influenciaram a resposta imune e registraram a necessidade de pesquisas mais específicas envolvendo a bactéria e Motivos CpG.

Em relação à ocorrência de Motivos CpG em regiões codificantes da *E. canis*, AACGTT aparece 6 vezes no gene ECAJ_0060 (*hypothetical protein*). Dentre os genes com mais motivos em sua composição, aparecem os genes ECAJ_0060 (14 motivos), ECAJ_0387 ankyrin (14 motivos) e ECAJ_0365 gp200 (9 motivos).

O gene ECAJ_0060 contém repetições anquirina, de acordo com a ferramenta String. O gene possui também três TRs com elevado valor TRFSCORE em sua formação: TR 46 (2725), TR 47 (1601) e TR 49 (1383). Utilizando a ferramenta Blast para localizar sequências similares no *GenBank*, observou-se que o gene possui semelhanças com genes de eucariotos. Repetições anquirina e semelhança com genes eucariotos são características de genes que codificam proteínas efetoras. O gene ECAJ_0387, caracterizado por possuir repetições anquirina, possui grande semelhança com o gene ECAJ_0060, conforme ferramenta String. O gene possui o TR 381 (TRFSCORE 1496) em sua formação. ECAJ_0365 gp200 é uma glicoproteína de membrana e é secretada para o núcleo, onde interage com a cromatina e potencialmente modula a expressão de genes da célula hospedeira (STUEN et al., 2013). O gene possui repetições anquirina e o TR 338 em sua formação. O TR 338 possui baixos valores TRFSCORE (49) e VARSORE (-0.8), porém Zhang et al.(2008) identificaram instabilidade do TR analisando cepas de diferentes continentes.

Em *E. Chaffeensis*, o Motivo CpG GACGCT aparece 8 vezes no gene ECH_0653 ankyrin *repeat-containing protein*. Dentre os genes com mais motivos em sua composição, aparecem os genes ECH_0653 ankyrin *repeat-containing protein* (27 Motivos), ECH_0945 *hypothetical protein* (10 Motivos).

O gene ECH_0653 ankyrin *repeat-containing protein* possui 4 TRs com valores TRFSORE entre 637 a 920 em sua composição. ECH_0945 possui o TR 646 (TRFSORE 1096) em sua formação. De acordo com as ferramentas String e BlastP, o gene possui similaridade somente com proteínas também hipotéticas, não sendo possível induzir uma função.

A bactéria *A. phagocytophilum* tem Motivos CpG espalhados pelo genoma e não concentrados em genes. O motivo AACGTC aparece 4 vezes no gene APH_1224 valyl-tRNA synthetase, que catalisa a fixação de Valina para o tRNA, conforme ferramenta *String*.

Dentre os genes com mais Motivos CpG em sua composição, aparecem os genes APH_0720 *hypothetical protein* (33 Motivos), que atua na replicação e reparação do DNA, APH_0709 ankyrin *repeat-containing protein* (27 Motivos) e APH_0049 *hypothetical protein* (24 Motivos).

APH_0720 possui o TR 142 (VARSORE 1357) em sua formação. De acordo com as ferramentas String e BlastP, o gene possui similaridade somente com proteínas também hipotéticas,

não sendo possível induzir uma função. APH_0709 possui o *TR* 133 (VARSCORE 833). O gene possui repetições anquirina em sua formação e semelhança com o gene APH_0377, um gene da família VirB do Sistema de Secreção tipo IV. APH_0049 possui o *TR* 8 (VARSCORE 989) em sua formação. O gene possui similaridade com proteínas de várias espécies do gênero *Anaplasma* e com gene da família VirB10 do Sistema de Secreção tipo IV.

7.2 TANDEM REPEATS

Foram analisados 4 grupos distintos de *TRs*, separados com base na localização no genoma (intragênico e intergênico) e nos valores de TRFSCORE e VARSCORE. Optou-se por não discutir resultados envolvendo genes hipotéticos que não retornaram informações nas pesquisas realizadas.

7.2.1 *TRs* Codificantes com VARSCORE maior que 1

No genoma da *E. canis*, foram encontrados 2 *TRs* com VARSCORE maior que 1, ambos em regiões codificantes, sendo eles o *TR* 468 (2,87), presente no gene Ecaj_0472 e o *TR* 543 (2,83), encontrado no gene Ecaj_0530.

O gene Ecaj_0472 consiste em uma proteína hipotética, mas por meio de consultas em Bancos de Dados genômicos utilizando CD-Seach, o gene codifica DNA topoisomerase 2. As topoisomerases são enzimas que controlam os enrolamentos e emaranhamentos do DNA, sendo essenciais durante a transcrição e a replicação. Inibidores de topoisomerase estão entre os medicamentos anticancerígenos e antibacterianos mais eficazes e mais usados, sendo um importante foco em pesquisas para o desenvolvimento de novas drogas (POMMIER, 2013).

O gene Ecaj_0530 codifica a *TrbL/VirB6 plasmid conjugal transfer protein*, classificado como uma proteína do Sistema de Secreção Tipo IV (T4SS).

Diversos trabalhos publicados na literatura utilizam o gene TRP36 (também chamado de gp36) para verificar a variabilidade genômica da bactéria *E. canis* em diferentes regiões (DOYLE et al., 2005; AGUIAR et al., 2013). O gene TRP36 possui o *TR* 103, com VARSCORE 0.25. Porém, existem postagens no GenBank do gene (código de acesso DQ085427.1) da *E. canis* com VARSCORE 1.49, sendo que a variação dos dois encontra-se justamente na região em que ocorre o *TR* 103.

Para os *TRs* codificantes da bactéria *E. chaffeensis* com VARSCORE maior que 1, destaque para os *TRs* com código identificador 90 (VARSCORE 3,72) e 350 (VARSCORE 2,52), respectivamente presentes nos genes ECH_0166 e ECH_0558. Ambos os genes consistem de proteínas hipotéticas, porém, consultas utilizando CD-Seach retornaram que o gene ECH_0166

consiste do gene ribonuclease E, que catalisa a degradação do RNA, e o gene ECH_0558 codifica uma lipoproteína, responsável pelo transporte de lipídeos em meios líquidos.

O gene ECH_0695 *hypothetical protein*, que contém o TR 461 (1,96), possui o padrão DUF - *Domain of Unknown function* - em sua sequência, conforme CD-Search. DUF consiste em um domínio encontrado em procariotos. Esse domínio é encontrado em diversas proteínas e é apontado como uma característica de proteínas efetoras, porém, sua função é desconhecida. Os genes ECH_0878 *hypothetical protein*, que possui padrão anquirina conforme String, e ECH_0252 (*hypothetical protein*) possuem uma interessante particularidade. Ambos possuem dois TRs com VARSCORE maior que um em sua composição (respectivamente TRs 591, 596 e 139, 140).

Trabalhos existentes na literatura relacionados com a diversidade genética de *E. chaffeensis* utilizam os genes p28 (LONG et al., 2002; MIURA e RIKIHISA, 2007), gp47 (DOYLE et al., 2006) e 120 kDa (CHEN et al., 1997) para analisar a variabilidade genômica da bactéria em diferentes regiões geográficas. O gene p28 é o ECH_1143 e, embora seja frequentemente citado em experimentos relacionados com a variabilidade bacteriana, não possui TRs em sua formação. O gene gp47 é o ECH_0166. ECH_0166 codifica uma proteína de superfície imunorreativa e possui o TR 90, com VARSCORE 3.72, sendo justamente o TR com maior VARSCORE encontrado para a bactéria. O gene ECH_0039, que codifica a proteína *120 kDa immunodominant surface protein*, contém em sua composição o TR 16. O TR 16 possui alto valor TRFSCORE (2044), fator este que deve estar relacionado com a sua variabilidade.

Para a bactéria *A. phagocytophilum*, destaque para o TR 197 (3,42), presente no gene APH_0964. APH_0964 codifica *triple helix repeat-containing collagen*. Recentes trabalhos mostraram a ligação entre a superexpressão de genes deste tipo a doenças envolvendo artérias lesionadas e vários tipos de cânceres em humanos (PYAGAY et al., 2005; DUMUS et al., 2006; KE et al., 2014). No entanto, recentemente, Duarte et al. (2014) publicaram os resultados de uma pesquisa que desconsideraram a hipótese da superexpressão do gene ser um marcador para doenças ou células cancerígenas. Nesta pesquisa, eles verificaram que indivíduos saudáveis ruivos apresentam uma expressão o gene cerca de até 400 vezes superior a qualquer outro indivíduo.

7.2.2 TRs Codificantes com VARSCORE < 1 e com TRFSCORE > 1000

Além dos TRs codificantes identificados com alto VARSCORE, foram identificados TRs com VARSCORE menor que 1, mas com alto valor do parâmetro TRFSCORE.

Para *E. canis*, o gene Ecaj_0716 possui o TR 710 (2848). O gene possui similaridade com proteínas hipotéticas das bactérias *E. chaffeensis* e *E. muris*. O gene Ecaj_0017, que contém o TR 9 (2848), codifica a glicoproteína de superfície gp140, caracterizada como proteína imunorreativa e

diferencialmente expressa em diversos trabalhos (DOYLE, 2006; ZHANG et al., 2008; MCBRIDE; WALKER, 2010).

O gene Ecaj_0060 possui três TRs significativos, sendo TRs 49, 46 e 47. O gene, descrito na discussão sobre Motivos CpG por ser o gene de *E. canis* com mais motivos em sua composição, contém repetições anquirina e semelhanças com gene hipotético do protozoário *Neospora caninum*.

De acordo com as ferramentas CD-Seach e String, os genes O gene Ecaj_0060, Ecaj_0529 (TR 539 (2006)), Ecaj_0220 (TR 194 (1287)) e Ecaj_0715 (TR 706(1060)) possuem repetições anquirina e pertencem a família VirB6 - Sistema de Secreção Tipo IV.

Sobre a bactéria *E. chaffeensis*, o gene ECH_0499 da família VirbB6 (T4SS) possui os TRs 287 (VARSCORE 2912) e 301 (VARSCORE 1290) em sua composição. O gene ECH_0039 possui o TR 16 (VARSCORE 2044) e codifica a proteína imunorreativa efetora p120, descrita em trabalhos relacionados com a diversidade genética da bactéria e por estar relacionada com a regulação da expressão de genes do hospedeiro (MCBRIDE; WALKER, 2010; ZHANG, 2008). O gene ECH_0106 (TR 46 (1110)) possui similaridade com o gene p28, presente em trabalhos existentes na literatura relacionados com a diversidade genética de *E. chaffeensis* (LONG et al., 2002).

No genoma da *A. phagocytophilum*, foram identificados o TR 241, gene APH_1387 - HGE-2 protein e o TR 93, gene APH_0455 - HGE-14 protein. Proteínas HGE são específicas da *A. phagocytophilum*. Elas são excretadas pelo Sistema de Secreção Tipo IV para o interior da célula hospedeira, mas sua função ainda é desconhecida.

APH_0377 e APH_0256 (TR 38) codificam proteínas do Sistema de Secreção, sendo que APH_0377 possui 3 TRs em sua formação (TRs 60, 64 e 67). O gene APH_0032 contém dois TRs em sua composição (TRs 3 e 4) e possui similaridade com o gene APH_0138 ligA, que codifica a ligase do DNA. APH_0546 (TR 105) mostrou semelhança funcional segundo a ferramenta String com a APH_0928, que possui repetições anquirina, e com APH_0377, que pertence ao sistema de secreção tipo IV.

Huang et al.(2010) e Huang et al.(2011) identificaram que os genes APH_1387 e APH_0032 são expressos durante o desenvolvimento intracelular de *A. phagocytophilum*. As proteínas desses genes são translocadas para o interior da célula hospedeira pelo T4SS, localizando-se em vacúolos ligados à membrana. Os trabalhos sugerem que esta identificação indica que os genes desempenham um papel importante na sobrevivência intracelular.

7.2.3 TRs Não-Codificantes com VARSCORE ≥ 1

Um gene, em bactérias, é composto basicamente por uma região codificadora (CDS), por uma região promotora (“upstream” ao gene) e por outras regiões, também importantes, como o

operador que inibe ou ativa a transcrição em diferentes taxas e o terminador, seqüência “downstream” ao final do gene que sinaliza à RNA polimerase a parada ou pausa na transcrição (GRIFFITHS, 1998).

Segundo XIONG (2006), as regiões promotoras em procariotos, em geral, ficam a aproximadamente –200 bases “upstream” à região codificadora da proteína. Como não se sabe exatamente onde está situado o início do sítio de transcrição do gene, definiu-se neste trabalho que as sequencias de observação das regiões intergênicas a se trabalhar seriam com até -500 b “upstream” da posição +1 da ORF.

Figura 29 - Esquema geral sobre a localização da sequencia de observação de TRs em uma região intergênica “upstream” de uma ORF qualquer da fita principal do DNA.

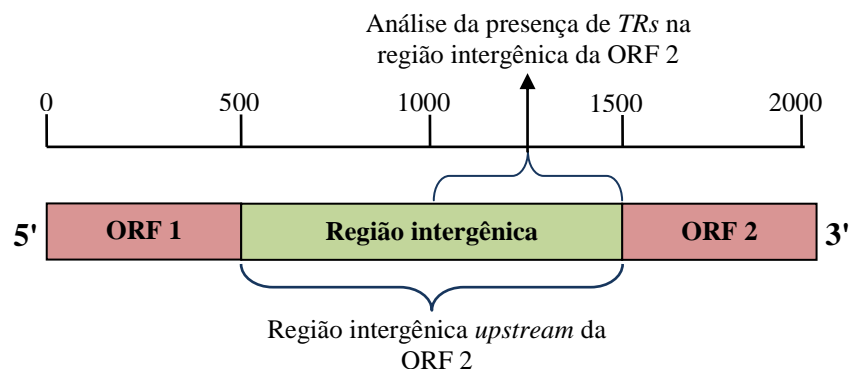
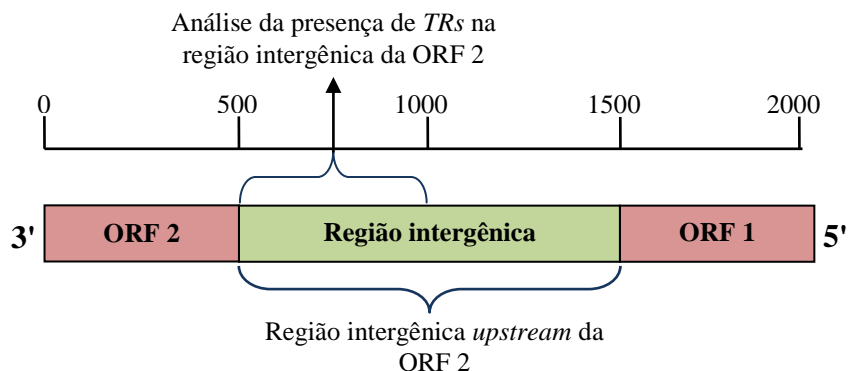


Figura 30 - Esquema geral sobre a localização da sequencia de observação de TRs em uma região intergênica “upstream” de uma ORF qualquer da fita secundária do DNA.



Para a bactéria *E. canis*, todos os TRs com VARSCORE maior que 1 estão em regiões codificantes.

Em *E. chaffeensis*, nenhum TR com VARSCORE maior que 1 está próximo a algum gene.

Para a *A. phagocytophilum*, o TR 116 (2,21) está localizado a 450 nt upstream do início do gene APH_0617 *hypotetical protein*. Segundo CD-Search, o gene possui o padrão DUF. APH_0617 está presente em todas as estirpes com genoma completo disponíveis no GenBank (*Dog2*, *JM*, *HZ2* e *HZ*).

O TR 119 (1,03) está a 308 nt do gene APH_0633 *hypotetical protein*. APH_0633 é similar a APH_0636 *hypotetical protein*, mas ambas não possuem precedentes nos bancos de dados analisados.

7.2.4 TRs Não-Codificantes com VARSCORE < 1 e TRFSCORE > 1000

O TRs 9 e 49 estão próximos de genes em *E. canis*. O TR 9 encontra-se a 460 nt de Ecaj_0016 *4Fe-4S ferredoxin iron-sulfur binding domain-containing* (proteínas ferro-enxofre). As proteínas ferro-enxofre atuam em diversos processos celulares, tal como expressão genética, a reparação de DNA, modificação do RNA, metabolismo central e respiração (ROCHE et al., 2012).

O TR 49 está a 411 nt de Ecaj_0061 *branched-chain alpha-keto acid dehydrogenase subunit E2*. Ecaj_0061 codifica a subunidade 2 do complexo multienzimático desidrogenase dos alfa-cetoácidos de cadeia ramificada.

Em *E. chaffeensis*, o TR 582 está a 391 de ECH_0871 *putative twitching motility protein PilT*. ECH_0871 codifica proteína associada a sistemas de adesão de bactérias e a fimbrias do tipo IV. O TR 599 está a 478 nt de ECH_0877 *ankyrin repeat-containing protein*. O TR 458 está a 69 nt de ECH_0685 *hypothetical protein*. ECH_0685 é similar a proteína ECH_0684, que possui repetições anquirina em sua formação.

Para *A. phagocytophilum*, o TR 123 (2158) está a 257 nt de APH_0653 *hypotetical protein*. APH_0653 está presente também nos genomas dos isolados HZ2, Dog2 e JM, mas não retornou nenhum resultado em relação a sua funcionalidade. TR 38 está a 484 nt de APH_0257 *quinone oxidoreductase*, que codifica a enzima quinona redutase.

7.2.5 Análise *in vitro* de Variabilidade utilizando o TR 468 de *E. canis*

O resultado do sequenciamento realizado com as amostras de ensaios de PCR realizados para amplificação da região do TR 468 mostrou diferenças entre a amostra da cidade de Monte Negro/ Rondônia, em relação as demais. É possível que a região apresente maior discrepância ao ser comparada com amostras de diferentes continentes.

Outro importante fator a ser destacado é que o TR 468 encontra-se em uma região codificante do genoma, no gene Ecaj_0472. Segundo resultados de consultas realizadas utilizando a ferramenta CD-Search, o gene codifica DNA topoisomerase 2, um alvo que tem sido foco para o desenvolvimento de novas drogas (POMMIER, 2013).

Variações em regiões codificantes resultam em diferenças no bloco de aminoácidos, podendo alterar também funcionalidade da proteína.

8 CONCLUSÃO

O presente trabalho implementou computacionalmente as análises de *TRs* e Motivos CpG com intuito de evidenciar alvos nos dados genômicos de bactérias da família *Anaplasmataceae* que possam estar relacionados com a variabilidade genética, interação patógeno-hospedeiro e ativação do sistema imunológico.

Para a análise de Motivos CpG, optou-se por desenvolver o *software CpG HighLight*, uma nova ferramenta Java com código aberto e independente de plataforma que pode ser utilizada por pesquisadores em genomas bacterianos.

Para análise de *TRs*, utilizou-se o software SERV para identificar os *TRs* nos genomas e categorizá-los em relação à sua instabilidade e significância. Os resultados do software SERV foram armazenados em um banco de dados, sendo criados métodos Java para correlacionar as informações.

Os resultados dos métodos computacionais gerados pelo *CpG HighLight* destacaram os Motivos CpG identificados nos genomas e os genes que possuem Motivos em sua composição. Em relação aos *TRs*, foram evidenciados *TRs* codificantes, não-codificantes, instáveis e significativos (parâmetros VARSCORE e TRFSCORE), além dos genes que possuem *TRs*.

Os resultados evidenciados foram discutidos utilizando informações da literatura e ferramentas de buscas em banco de dados genômicos públicos.

O conjunto de dados gerados poderão ser analisados em laboratório para compreensão das estratégias de patogenicidade das bactérias e elaboração de novas drogas e métodos de diagnósticos.

REFERÊNCIAS

- AGUIAR D.M. et al. Genetic diversity of Ehrlichia canis in Brazil. **Veterinary Microbiology**, v. 164, n. 3-4, p. 315–321, 2013.
- BENSON G. Tandem repeats finder: a program to analyze DNA sequences. **Nucleic Acids Research**, v. 27, n. 2, p. 573–580, 1999.
- BODE C. et al. CpG DNA as a vaccine adjuvant. **Expert Review Vaccines**, v. 10, n.4, p. 499-511, 2011.
- BOWMAN, D. D. Introduction to the alpha-proteobacteria: Wolbachia and Bartonella, Rickettsia, Brucella, Ehrlichia, and Anaplasma. **Top Companion Animal Medicine**, v. 26, n. 4, p. 173-177, 2011.
- CALIC, S. B. et al. Human ehrlichioses in Brazil: first suspect cases. **The Brazilian Journal of Infectious Diseases**, v. 8, n.3, p. 259-262, 2004.
- CARDOZO, G. P. et al. A glass bead protocol for recovery of host cell free Ehrlichia canis and quantification by Sybr-green real-time PCR. **Biocell**, v. 35, n. 1, p. 35-36, 2011.
- CHAUNG HC. CpG oligodeoxynucleotides as DNA adjuvants in vertebrates and their applications in immunotherapy. **International Immunopharmacology**, v. 6, n. 10, p. 1586–1596, 2006.
- CHUANG TH et al. Development of CpG-Oligodeoxynucleotides for Effective Activation of Rabbit TLR9 Mediated Immune Responses. **Plos Biology**. v. 9, n. 9, p. e108808, 2014.
- CHEN S. M. et al. Genetic and antigenic diversity of Ehrlichia chaffeensis: comparative analysis of a novel human strain from Oklahoma and previously isolated strains. **Journal Infectious Disease**. v. 175, n. 4, p. 856–863, 1997.
- COSTA P.S. et al. More about human monocytotropic ehrlichiosis in Brazil: serological evidence of nine new cases. **Brazilian Journal of Infectious Disease**, v. 10, n. 1, p. 7-10, 2006.
- DANTAS-TORRES, F. The brown dog tick, Rhipicephalus sanguineus (Latreille, 1806) (Acari: Ixodidae): from taxonomy to control. **Veterinary Parasitology**, v. 152, n. 3-4, p. 173-185, 2008.
- DANTAS-TORRES, F. Biology and ecology of the brown dog tick, Rhipicephalus sanguineus. **Parasites & Vectors**, v. 3, n. 1, p. 26-37, 2010.
- DOYLE, C.K. et al. A molecular characterization of *E. canis* gp36 and *E. chaffeensis* gp47 Tandem Repeats among isolates from different geographic localization. **Annals of New York Academy of Science**, v. 1063, p. 433-435, 2005.
- DOYLE, C.K. et al., Differentially Expressed and Secreted Major Immunoreactive Protein Orthologs of Ehrlichia canis and *E. chaffeensis* Elicit Early Antibody Responses to Epitopes on Glycosylated tandem repeats. **Infection and Immunity**, v. 74, n. 1, p. 711-720, 2006.
- DUARTE C. W. et al. Elevated Plasma Levels of the Pituitary Hormone Cthrc1 in Individuals with Red Hair but Not in Patients with Solid Tumors. **PLoS ONE**, v. 9, n. 6, p. e100449, 2014.

- DUMLER, J. S. et al. Reorganization of genera in the families Rickettsiaceae and *Anaplasmataceae* in the order Rickettsiales: unification of some species of Ehrlichia with Anaplasma, Cowdria with Ehrlichia and Ehrlichia with Neorickettsia, descriptions of six new species combinations and designation of Ehrlichia equi and 'HGE agent' as subjective synonyms of Ehrlichia phagocytophila. **International Journal of Systematic and Evolutionary Microbiology**, v.51, n., 6, p. 2145-2165, 2001.
- DUMUS T. et al. Expression analysis of the novel gene collagen triple helix repeat containing-1 (Cthrc1), **Gene Expression Patterns**, v.6, n.8, p. 935-940, 2006.
- GARCIA-GARCIA, J.C. et al. Silencing of host cell CYBB gene expression by the nuclear effector AnkA of the intracellular pathogen Anaplasma phagocytophilum. **Infection and Immunity**, v. 77, n. 6, p. 2385-2391, 2009.
- GEMAYEL R. et al. Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequence. **Annual Review of Genetics**, v. 44, n. 1, p. 445-477, 2010.
- GEMAYEL R. et al. Beyond Junk-Variable Tandem Repeats as Facilitators of Rapid Evolution of Regulatory and Coding Sequences. **Genes**, v. 3, n. 1, p. 461-480, 2012.
- GASSER, R. B. et al. Major prospects for exploring canine vector borne diseases and novel intervention methods using 'omic technologies. **Parasites & Vectors** , v. 4, n. 53, 2011.
- GE Y.; RIKIHISA Y. Anaplasma phagocytophilum delays spontaneous human neutrophil apoptosis by modulation of multiple apoptotic pathways. **Cellular Microbiology**, v. 8, n. 1, p. 1406–1416, 2006.
- GE Y., RIKIHISA Y. Surface-Exposed Proteins of Ehrlichia chaffeensis. **Infection and Immunity**, v. 75, n. 1, p. 3833-3841, 2007.
- HACKENBERG M. et al. CpGcluster: a distance-based algorithm for CpG-island detection, **BMC Bioinformatics**, v. 7, n.1, 2006.
- HAJDUSEK O. Interaction of the tick immune system with transmitted pathogens. **Frontiers in Cellular and Infection Microbiology**. v. 3, n. 26, 2013.
- HARRUS S. et al. Down-regulation of MHC class II receptors of DH82 cells, following infection with Ehrlichia canis. **Veterinary Immunology and Immunopathology**, v. 96, n. 3-4, p. 239-43, 2003.
- HARRUS S.; WANER T. Diagnosis of canine monocytotropic ehrlichiosis (Ehrlichia canis): An overview. **Veterinary Journal**, v. 187, n. 3, p. 292-296, 2011.
- HARTMANN G., KRIEG A. M. Mechanism and function of a newly identified CpG DNA motif in human primary B cells. **The Journal of Immunology**, v. 164, n. 2, p. 944–953, 2000.
- HERINGA, J. et. al. Tracking *repeats* using significance and transitivity. **Bioinformatics**, v. 20, n.1, p. 311–317, 2004.
- HOTOPP D. J. C. et al. Comparative genomics of emerging human ehrlichiosis agents. **PLoS Genetics**, v. 2, n. 12, p. e213, 2006.

HUANG B. et al. *Anaplasma phagocytophilum* APH_0032 is expressed late during infection and localizes to the pathogen-occupied vacuolar membrane. **Microbial Pathogenesis**, v. 49, n.5, p. 273-284, 2010.

HUANG B. et al. *Anaplasma phagocytophilum* APH_1387 Is Expressed throughout Bacterial Intracellular Development and Localizes to the Pathogen-Occupied Vacuolar Membrane. **Infection Immunity**, v. 78, n. 5, p. 1864-1873, 2010.

JANSEN A; GEMAYEL R.; VERSTREPEN K. J. Unstable microsatellite *repeats* facilitate rapid evolution of coding and regulatory sequences. **Genome Dynamics**, v.7, n. 1, p. 108-125, 2012.

JORDAN P. et al. Diversity in coding tandem repeats in related *Neisseria* spp. **BMC Microbiology**, v. 3, n. 1, p. 23-37, 2003.

KAJAVA, A.V. T-REKS: identification of *tandem repeats* in sequences with a K-meanS based algorithm. **Bioinformatics**, v. 25, n.1, p. 2632-2638, 2009.

KAMSTRUP S. et al. Response of porcine peripheral blood mononuclear cells to CpG-containing oligodeonucleotides. **Veterinary Microbiology**, v. 78, n. 4, p. 353-362, 2001.

KANDIMALLA, E. R. et al. Toll-like receptor 9: modulation of recognition and cytokine induction by novel synthetic CpG DNAs. **Biochemical Society Transactions**, v. 31, n. 1, p. 654-658, 2003.

KE Z. et al. Overexpression of collagen triple helix repeat containing 1 (CTHRC1) is associated with tumour aggressiveness and poor prognosis in human non-small cell lung cancer. **Oncotarget**, v. 5, n. 19, p. 9410-9424, 2014.

KLINMAN D. M. CpG oligonucleotides as adjuvants for vaccines targeting infectious diseases. **Advanced Drug Delivery Reviews**, v. 61, n. 3, p. 248–255, 2009.

KOLPAKOV, R., BANA G., KUCHEROV G. Mreps : efficient and flexible detection of *tandem repeats* in DNA. **Nucleic Acids Research**, v. 31, n. 13, p. 3672-3678, 2003.

KRIEG A. M. et al. CpG motifs in bacterial DNA trigger direct B-cell activation. **Nature**. v. 374, n. 6522, p. 546-549, 1995.

KRIEG A. M. CpG motifs in bacterial DNA and their immune effects. **Annual Review Immunology**. v. 20, n. 1, p. 709-760, 2002.

KRIEG A. M. CpG Still Rocks! Update on a accidental drug. **Nucleic Acid Therapeutics**, v. 22, n. 2, p. 77-89, 2012.

KURIAKOSE J. A. et al. Molecular basis of antibody mediated immunity against *Ehrlichia chaffeensis* involves species-specific linear epitopes in tandem repeat proteins. **Microbes and infection**, v. 14, n. 12, p. 1054–1063, 2012.

LABRUNA, M. B. et al. A preliminary investigation of *Ehrlichia* species in ticks, humans, dogs, and capybaras from Brazil. **Veterinary Parasitology**, v. 143, n. 2, p. 189-195, 2007.

LEGENDRE M., et. al. Sequence-based estimation of minisatellite and microsatellite *repeat* variability. **Genome Research**, v. 17, n. 1, p. 1787-1796, 2007.

- LEVDANSKY E. et al. Coding fungal *tandem repeats* as generators of fungal diversity. **Fungal Biology**, v. 22, n.1, p. 85-96, 2008.
- LIN, M.; RIKIHISA, Y. Ehrlichia chaffeensis and Anaplasma phagocytophilum lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. **Infection and Immunity**, v. 71, n. 9, p. 5324-5331, 2003.
- LONG S. W. et al. Antigenic variation of Ehrlichia chaffeensis resulting from differential expression of the 28-Kilodalton protein gene family. **Infection and Immunity**, v. 70, n. 4, p. 1824-1831, 2002.
- MACHADO R. Z. et al. Migratory and carnivorous birds in Brazil: reservoirs for Anaplasma and Ehrlichia species?. **Vector-Borne Zoonotic Diseases**, v.12, n. 8, p. 705-708, 2012.
- MACHADO R. Z. et al. Detection of Ehrlichia chaffeensis in Brazilian marshdeer (Blastocerus dichotomus). **Veterinary Parasitology**, v. 139, n. 1-3, p. 262-266, 2006.
- MARINS M., et al. Molecular evaluation of the incidence of Ehrlichia canis, Anaplasma platys and Babesia spp. in dogs from Ribeirão Preto, Brazil. **The Veterinary Journal**, v. 179, n. 1, p. 145-148, 2007.
- MAVROMATIS et al., The Genome of the Obligately Intracellular Bacterium Ehrlichia canis Reveals Themes of Complex Membrane Structure and Immune Evasion Strategies. **Journal of Bacteriology**, v. 188, n. 11, p. 4015-4023, 2006.
- McBRIDE, J. W. et al. Pcr detection of acute Ehrlichia canis infection in dogs. **Journal of Veterinary Investigation**, v. 8, n. 6, p. 441-447, 1996.
- McBRIDE J. W., WALKER D. H. Progress and obstacles in vaccine development for the ehrlichioses. **Expert Review of Vaccines**, v. 9, n. 9, p. 1071-1082, 2010.
- MIURA K., RIKIHISA Y. Virulence potential of Ehrlichia chaffeensis strains of distinct genome sequences. **Infection Immunity**, v. 75, n. 7, 3604–3613, 2007.
- MOXON E. R. et al. Adaptive evolution of highly mutable loci in pathogenic bacteria. **Current Biology**, v. 4, p. 24–33, 1994.
- MRÁZEK J. et al. AIMIE: a web-based environment for detection and interpretation of significant sequence motifs in prokaryotic genomes. **Bioinformatics**, v. 24, n. 8, p. 1041–1048, 2008.
- MUTWIRI G. et al. Biological activity of immunostimulatory CpG DNA motifs in domestic animals. **Veterinary immunology and immunopathology**, v. 91, p. 89-103, 2003.
- MUTWIRI G. et al. Approaches to enhancing immune responses stimulated by CpG oligodeoxynucleotides. **Advanced Drug Delivery Reviews**, v. 61, n. 51, p. 226–232, 2009.
- MUTWIRI G. TLR9 agonists: Immune mechanisms and therapeutic potential in domestic animals. **Veterinary Immunology and Immunopathology**, v. 148, n. 1-2, p. 85– 89, 2012.
- NELSON, C. et al. Whole genome transcription profiling of Anaplasma phagocytophilum in human and tick host cells by tiling array analysis. **BMC Genomics**, v. 9, n. 2, p. 364, 2008.

NIU H. et al. Subversion of cellular autophagy by *Anaplasma phagocytophilum*. **Cellular Microbiology**, v. 10, n. 3, p. 593-605, 2008.

NIU H et a. Ehrlichia type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. **Cellular Microbiology**, v.14, p. 1037-1050, 2012.

OLIVEIRA, L. P., et al. Molecular analysis of the rRNA genes of *Babesia* spp and *Ehrlichia canis* detected in dogs from RibeirAo Preto, Brazil. **Brazilian Journal of Microbiology**, v. 40, n. 2, p. 238-240, 2009.

OTRANTO, D. et al. Managing canine vector-borne diseases of zoonotic concern: part one. **Trends in Parasitology**, v. 25, n.4, p. 157-163, 2009.

OTRANTO, D. et al. Managing canine vector-borne diseases of zoonotic concern: part two. **Trends in Parasitology**, v. 25, n. 5, p. 228-235, 2009b.

PADDOCK, C.D.; CHILDS J. E. . Ehrlichia chaffeensis: a Prototypical Emerging Pathogen. **Clinical Microbiology Reviews**, v.16, n. 1, p. 37-64, 2003.

PAROLA, P. et al. Tick- and flea-borne rickettsial emerging zoonoses. **Veterinary Research**, v. 36, n. 3, p. 469-492, 2005.

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, 2013.

PEREZ, M., et al., Human Infection with *Ehrlichia Canis* Accompanied by Clinical Signs in Venezuela. **Annals of New York Academic Sciences**, v. 1078, p. 110-117, 2006.

POMMER Y. Drugging Topoisomerases: Lessons and Challenges. **ACS Chemical Biology**, v. 8, n.1, p. 82-95, 2013.

PYAGAY P. et al. Collagen triple helix repeat containing 1, a novel secreted protein in injured and diseased arteries, inhibits collagen expression and promotes cell migration. **Circulation Research**, v. 96, n. 2, p. 261-268, 2005.

RAMAZANZADEH R.; MCNERNEY R. Variable Number of Tandem Repeats (VNTR) and its Application in Bacterial Epidemiology. **Pakistan Journal of Biological Sciences**, v. 10, n. 16, p. 2612-2621, 2007.

RANKIN R. et al. CpG Motif Identification for Veterinary and Laboratory Species Demonstrates That Sequence Recognition Is Highly Conserved. **Antisense & Nucleic Acid Drug Development**, v. 11, n. 5, p. 333-340, 2001.

RECHE, Pedro. A. Prediction of methylated CpGs in DNA sequences using a support vector machine. **FEBS Letters**, v. 579, n. 20, p. 4302-4308, 2005.

RENNOLL-BANKERT, K. E.; DUMLER J. S. Lessons from *Anaplasma phagocytophilum*: chromatin remodeling by bacterial effectors. **Infectious Disorders Drug Targets** , v. 12, n. 5, p. 380-387. 2012.

- RICE P et al. EMBOSS: the European molecular biology open software suite. **Trends in Genetics**, v. 16, n. 6, p. 276–277, 2000.
- RIKIHISA, Y. The tribe Ehrlichieae and ehrlichial diseases. **Clinical Microbiology Reviews**, v. 4, n. 3, p. 286-308, 1991.
- RIKIHISA Y. et al. Type IV secretion system of *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. **Annual New York Academic Science**, v. 1166, p. 106-111, 2009.
- RIKIHISA Y. *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. **Nature Reviews Microbiology**, v. 13, n. 8, p. 328-339, 2010a.
- RIKIHISA Y. Molecular Events Involved in Cellular Invasion by *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum*. **Veterinary Parasitology**, v. 167, n. 1, p. 155-166, 2010b.
- RIKIHISA Y. Mechanisms of obligatory intracellular infection with *Anaplasma phagocytophilum*. **Clinical Microbiology Reviews**, v. 24, n. 3, p. 469–489. 2011.
- ROCHE B. et al. Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. **Biochimica et Biophysica Acta (BBA) - Bioenergetics**, v. 1827, n. 3, p. 455-469, 2013.
- SANTOS, F. et al. Molecular evaluation of the incidence of *Ehrlichia canis*, *Anaplasma platys* and *Babesia* spp. in dogs from Ribeirao Preto, Brazil. **The Veterinary Journal**, v. 179, n.1, p. 145-148, 2009.
- SANTOS, H. A. et. al. Detection of *Anaplasma phagocytophilum* in Brazilian dogs by real-time polymerase chain reaction. **Journal of Veterinary Diagnostic Investigation**, v. 23, n. 4, p. 770–774, 2011.
- SANTOS, H. A. et. al. Molecular epidemiology of the emerging zoonosis agent *Anaplasma phagocytophilum* (Foggie, 1949) in dogs and ixodid ticks in Brazil. **Parasites & Vectors**, v. 6, p. 348, 2013.
- SATO Y. et al. Immunostimulatory DNA sequences necessary for effective intradermal gene immunization. **Science**, v. 273, p. 352-354, 1996.
- SCHMIDT, A.L.; MITTER, V. Microsatellite mutation directed by an external stimulus. **Mutation Research Fundamental Molecular Mechanism Mutagenesis**, v. 568, p. 233-243, 2004
- SEVERO M. S. et al. *Anaplasma phagocytophilum*: deceptively simple or simply deceptive?. **Future Microbiology**, v. 7, p. 719–731, 2012.
- SHODA L. K. M. et al. Immunostimulatory CpG-modified plasmid DNA enhances IL-12, TNF- α , and NO production by bovine macrophages. **Journal of Leukocyte Biology**, v. 70, p. 103- 112, 2001.
- SKOTARCZAK, B. Canine ehrlichiosis. **Annals of Agriculture Environmental Medicine**, v.10, n. 2, p. 137-141, 2003.
- STUEN S et al. *Anaplasma phagocytophilum*-a widespread multi-host pathogen with highly adaptive strategies. **Frontier of Cellular and Infection Microbiology**, v. 3, n. 31, 2013.

- UDEN J. V.; RAZ E. Immunostimulatory DNA and applications to allergic disease. **Journal of Allergy and Clinical Immunology**. v. 104, n. 5, p. 902–910, 1999.
- VAN HAM S. M. et al. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. **Cell**, v. 73, n. 6, p. 1187–1196, 1993.
- VERSTREPEN K. J. et al. Intragenic tandem *repeats* generate functional variability. **Nature Genetics**. v. 37, n. 1, p. 986–990, 2005.
- YU X. J., McBride JW, Walker DH. Genetic diversity of the 28-kilodalton outer membrane protein gene in human isolates of *Ehrlichia chaffeensis*. **Journal of Clinical Microbiology**, v. 37, p. 1137–1143, 1999.
- YU X. J. et al. Molecular Cloning and Characterization of the 120-Kilodalton Protein Gene of *Ehrlichia canis* and Application of the Recombinant 120-Kilodalton Protein for Serodiagnosis of Canine Ehrlichiosis. **Journal of Clinical Microbiology**, p. 369–374, 2000.
- WAGNER H., Interactions between bacterial CpG-DNA and TLR9 bridge innate and adaptive immunity. **Current Opinion in Microbiology**, v. 5, n. 2, p. 62–69, 2002.
- WAKEEL A. et al. An *Ehrlichia chaffeensis* Tandem Repeat Protein Interacts with Multiple Host Targets Involved in Cell Signaling, Transcriptional Regulation, and Vesicle Trafficking. **Infection and Immunity**, v. 77, n. 5, p. 1734–1745, 2009.
- WALKER, D. H. et al. *Ehrlichia chaffeensis*: a prevalent, life-threatening, emerging pathogen. **Transactions of the American Clinical and Climatological Association**, v. 115, n. 1, p. 382-374, 2004.
- WANG Yong, et al. Cytosine Methylation Is Not the Major Factor Inducing CpG Dinucleotide Deficiency in Bacterial Genomes. **Journal of Molecular Evolution**, v. 58, n. 6, p. 692-700, 2004.
- WATSON James. D., et. al. **DNA Recombinante: Genes e Genomas**. 3 ed. Porto Alegre: Artmed, 2009.
- ZHANG J. Z. et. al. The developmental cycle of *Ehrlichia chaffeensis* in vertebrate cells. **Cellular Microbiology**. v. 9, n. 1, p. 610-618, 2007.
- ZHANG X. et. al. Genetic and Antigenic Diversities of Major Immunoreactive Proteins in Globally Distributed *Ehrlichia canis* Strains. **Clinical and Vaccine Immunology**, v. 15, n. 7, p. 1080-1088, 2008.
- ZHOW K. et al. The role of variable DNA tandem repeats in bacterial adaptation. **FEMS Microbiology Reviews**. v. 38, n. 1, p. 119-141, 2013.