



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE DESENVOLVIMENTO SUSTENTÁVEL DO SEMIÁRIDO  
UNIDADE ACADÊMICA DE ENGENHARIA DE BIOTECNOLOGIA  
CURSO DE GRADUAÇÃO EM ENGENHARIA DE BIOTECNOLOGIA E  
BIOPROCESSOS**

**YURI VINÍCIUS VERÍSSIMO DE LIMA**

**PREDIÇÃO E CARACTERIZAÇÃO DE ILHAS GENÔMICAS EM  
BACTÉRIAS DO GÊNERO *MYCOPLASMAS*, UTILIZANDO O MÉTODO DE  
CLUSTERIZAÇÃO *MEAN SHIFT***

**SUMÉ**

**2015**

**YURI VINÍCIUS VERÍSSIMO DE LIMA**

**PREDIÇÃO E CARACTERIZAÇÃO DE ILHAS GENÔMICAS EM  
BACTÉRIAS DO GÊNERO *MYCOPLASMAS*, UTILIZANDO O MÉTODO DE  
CLUSTERIZAÇÃO *MEAN SFHIT***

Monografia apresentada ao Curso de Graduação em Engenharia de Biotecnologia e Bioprocessos do Centro de Desenvolvimento Sustentável do Semiárido, da Universidade Federal de Campina Grande, como requisito para obtenção do título em Bacharel em Engenharia de Biotecnologia e Bioprocessos.

**ORIENTADOR: RAFAEL TRINDADE MAIA**

**SUMÉ**

**2015**

L732p

Lima, Yuri Vinícius Veríssimo de.

Predição e caracterização de ilhas genômicas em bactérias do gênero Mycoplasmas, utilizando o método de clusterização Mean Shift. / Yuri Vinícius Veríssimo de Lima. - Sumé - PB: [s.n], 2015.

69 f.

Orientador: Prof. Dr. Rafael Trindade Maia.

Monografia - Universidade Federal de Campina Grande; Centro de Desenvolvimento Sustentável do Semiárido; Curso de Graduação em Engenharia de Biotecnologia e Bioprocessos.

1. Microbiologia. 2. Bactérias - Genoma. 3. Mycoplasmas. I. Título.

CDU: 579 (043.3)

**YURI VINÍCIUS VERÍSSIMO DE LIMA**

**PREDIÇÃO E CARACTERIZAÇÃO DE ILHAS GENÔMICAS EM  
BACTÉRIAS DO GÊNERO *MYCOPLASMAS*, UTILIZANDO O  
MÉTODO DE CLUSTERIZAÇÃO *MEAN SFHIT*.**

**Monografia apresentada ao Curso  
de Engenharia de Biosistemas do  
Centro de Desenvolvimento  
Sustentável do Semiárido da  
Universidade Federal de Campina  
Grande, como requisito parcial para  
obtenção do título de Bacharel em  
Engenharia de Biosistemas.**

**BANCA EXAMINADORA:**

---

**Professor Dr. Rafael Trindade Maia.  
Orientadora – UAEDUC/CDSA/UFCG**

---

**Professora Dr<sup>a</sup> Thaís Gaudêncio do Rêgo.  
Examinadora I – UATEC/CDSA/UFCG**

---

**Professora Dr<sup>a</sup> Ana Verônica Silva do Nascimento.  
Examinador II – UATEC/CDSA/UFCG**

**SUMÉ - PB**

**2015**

*Á Deus por iluminar meu caminho e pela  
possibilidade de realização de um sonho.*

*Ofereço*

*Aos meus pais, Tony e Magnólia e á minha  
querida irmã Isabelle, pelo incentivo,  
confiança, apoio e amor incondicional.*

*Dedico*

“Ambição e conquista sem contribuição não têm significado”. (Neil Tolkien)

## RESUMO

Os custos de sequenciamento genético, nos últimos anos foram bastante reduzidos, induzindo um crescimento na quantidade de espécies de *Mycoplasmas* sequenciadas. A disponibilidade de inúmeros dados genômicos estruturais, trouxe consigo informações a serem exploradas, com destaque para as Ilhas Genômicas (IGs). As IGs são regiões do genoma de bactérias, que podem conter genes adquiridos por transferência horizontal de outros organismos, que pode conferir adaptações, como resistência a antibióticos e virulência. A detecção e caracterização destas regiões permitem que pesquisadores identifiquem os genes responsáveis pelas adaptações e desenvolvam novas vacinas e antibióticos. Existem atualmente, métodos para a detecção de IGs, mas requerem aperfeiçoamento. Neste trabalho se propõe um método alternativo para detecção de IGs em *Mycoplasmas*, utilizando o método de clusterização *Mean Shift*, e a caracterização das IGs, agrupando as proteínas por meio de uma Classificação Filogenética de Proteínas Codificadas em Genomas Completos (COGs).

Palavras-Chave: *Mycoplasmas*, Ilhas Genômicas, Clusterização, Mean-Shift, COGs.

## ABSTRACT

Costs with genetic sequencing have been greatly reduced in the past few years, leading to an increase in Mycoplasma species sequenced. The availability of numerous structural genomic data provided the exploitation of such information, especially for Genomic Islands (GIs). GIs are regions of the bacterial genome that may contain genes acquired by horizontal transfers from other organisms, which can confer adaptations, such as virulence and antibiotic resistance. Detection and characterization of these regions allows researchers to identify genes responsible for these adaptations and develop new vaccines and antibiotics. Current methods for GIs detection require improvement. This study proposes an alternative method for detection of Mycoplasma GIs, the Mean Shift clustering method, and characterization of these GIs, gathering the proteins by a phylogenetic classification of Proteins Encoded in Clusters of Orthologous Groups (COGs).

Keywords: Mycoplasmas, Genomic Islands, Clustering, Mean-Shift, COGs.



## LISTA DE FIGURAS

FIGURA 1 – DESCRIÇÃO IGs (Ilhas Genômicas)	18
FIGURA 2 – Segmento de um genoma representado no formato FASTA	37
FIGURA 3 – Lista de pontos interpretados para o genoma.	37
FIGURA 4 – Coleta do sequenciamento da <i>M. Californicum</i> em arquivo FASTA	41
FIGURA 5 – Sequenciamento da <i>M. Californicum</i> em arquivo FASTA	41
FIGURA 6 – Saída do algoritmo Mean Sfhit	42
FIGURA 7 – <i>Mycoplamas</i> e a posição da provável ilha genômica	43
FIGURA 8 – Nome das proteínas e as suas posições	43

## LISTA DE TABELAS

TABELA 1 – <i>Mycoplasmas</i> /código de acesso no NCBI.	38
TABELA 2 – <i>Mycoplasmas</i> Utilizadas/ Provável Ilha Genômica	45
TABELA 3 – Posições e Caracterização das Ilhas genômicas	46
TABELA 4 – Porcentagem das categorias do COGs	50

## LISTA DE ABREVIATURAS

DNA – Ácido Desoxirribonucleico

IG – Ilha Genômica

KB – Kilo Pares de Base

MB – Mega Pares de Base

NCBI – Centro Nacional de Informação Biotecnológica

RNA – Ácido Ribonucleico

THG – Transferência Horizontal de Genes

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	14
1.1 Contextualização .....	14
1.2 Objetivos.....	16
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	17
2.1 Ilha Genômica (IG).....	17
2.2 Transferência Horizontal de Genes (THG) .....	19
2.2.1 Ascensão da THG.....	21
2.2.2 Importância da THG.....	22
2.2.3 THG e Evolução Microbiana.....	23
2.3 Métodos para detecção de ilhas genômicas.....	24
2.3.1 Métodos baseados em comparação genômica .....	24
2.3.2 Métodos baseados em composição de sequência .....	24
2.4 Clusterização .....	25
2.5 <i>Mycoplasmas</i> .....	26
2.5.1 <i>Mycoplasma Agalactiae</i> .....	26
2.5.2 <i>Mycoplasma arthritidis</i> .....	27
2.5.3 <i>Mycoplasma bovis</i> .....	27
2.5.4 <i>Mycoplasma bovoculi</i> .....	27
2.5.5 <i>Mycoplasma californicum</i> .....	27
2.5.6 <i>Mycoplasma canadense</i> .....	28
2.5.7 <i>Mycoplasma Capricolum</i> .....	28
2.5.8 <i>Mycoplasma conjunctivae</i> .....	28
2.5.9 <i>Mycoplasma Crocodyli</i> .....	29
2.5.10 <i>Mycoplasma cynos</i> .....	29
2.5.11 <i>Mycoplasma díspar</i> .....	29
2.5.12 <i>Mycoplasma fermentans</i> .....	29

2.5.13 <i>Mycoplasma flocculare</i> .....	30
2.5.14 <i>Mycoplasma Gallisepticum</i> .....	30
2.5.15 <i>Mycoplasma genitalium</i> .....	30
2.5.16 <i>Mycoplasma Haemocanis</i> .....	31
2.5.17 <i>Mycoplasma haemofelis</i> .....	31
2.5.18 <i>Mycoplasma hominis</i> .....	31
2.5.19 <i>Mycoplasma hyopneumoniae</i> .....	32
2.5.20 <i>Mycoplasma hyorhinis</i> .....	32
2.5.21 <i>Mycoplasma Leachii</i> .....	32
2.5.22 <i>Mycoplasma meleagridis</i> .....	32
2.5.23 <i>Mycoplasma mobile</i> .....	33
2.5.24 <i>Mycoplasma mycoides</i> .....	33
2.5.25 <i>Mycoplasma ovis</i> .....	33
2.5.27 <i>Mycoplasma penetrans</i> .....	34
2.5.28 <i>Mycoplasma pneumoniae</i> .....	34
2.5.29 <i>Mycoplasma pulmonis</i> .....	34
2.5.30 <i>Mycoplasma putrefaciens</i> .....	34
2.5.32 <i>Mycoplasma synoviae</i> .....	35
2.5.33 <i>Mycoplasma wenyonii</i> .....	35
2.5.34 <i>Mycoplasma yeatsii</i> .....	35
<b>3 METODOLOGIA</b> .....	<b>36</b>
3.1 Parâmetros .....	36
3.1.1 Tamanho dos fragmentos .....	37
3.1.2 Coeficiente do desvio padrão .....	37
3.1.3 Quantidade de fragmentos artificiais.....	38
3.2 Coleta de Dados.....	38
3.3 Entrada dos Dados .....	41

3.4 Saída dos Dados .....	42
3.5 Coleta de Proteínas .....	43
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>45</b>
<b>5 CONSIDERAÇÕES FINAIS.....</b>	<b>54</b>
<b>REFERÊNCIAS .....</b>	<b>55</b>

# 1 INTRODUÇÃO

## 1.1 Contextualização

*Mycoplasmas* são organismos simples, com o conjunto mínimo de organelas necessárias para o crescimento e replicação. As espécies de *Mycoplasmas* são reconhecidas na natureza como patógenos de humanos, animais, aves, peixes, répteis, artrópodes e plantas. (KUMAR et al., 2014).

Em 1995, aconteceu o primeiro sequenciamento completo de um genoma da espécie, a *Mycoplasma genitalium* (FRASER et al., 1995), que se destaca, ao possuir um genoma de tamanho de 580kb, considerado o organismo com o menor genoma entre as espécies bacterianas.

O sequenciamento do genoma de *M. genitalium* levou ao surgimento da célula mínima e conceitos importantes do genoma. Com a diminuição dos custos de sequenciamento, a quantidade de *Mycoplasmas* sequenciadas só aumenta.

Com a disponibilidade de dados genômicos, foram possibilitados avanços significativos na compreensão da diversidade bacteriana. O alto volume de organismos sequenciados trouxe à tona inúmeras informações, principalmente no que se refere à estrutura genômica. Descobriu-se, por exemplo, em algumas delas, partes significantes de DNA de outras bactérias (HENTSCHEL;HACKER, 2001).

Este processo é conhecido por Transferência Horizontal de Genes (THG). A obtenção de genes de outros organismos é um fenômeno que gera uma mudança no genótipo de uma *Mycoplasma*, levando à criação de novas características ou até mesmo novas espécies (GOGARTEN et al., 2002). Estas *Mycoplasmas* possuem a vantagem de possuir além dos genes responsáveis por codificar funções essenciais, genes responsáveis por funções adaptativas (JUHAS et al., 2009). Os genes adquiridos por THG encontram-se em regiões denominadas Ilhas Genômicas (IGs).

IGs compreendem relativamente grandes regiões genômicas, por exemplo, 10-200 kb, que são adquiridas por meio da THG. Algumas IG relacionadas são conhecidas como patogênicas podendo conferir resistência à *Mycoplasma*, outras contêm genes que produzem metabólitos secundários, enquanto outras degradam produtos químicos. Apesar de muitas IGs conferirem às bactérias adaptações com efeitos prejudiciais aos humanos, algumas delas também podem ser responsáveis por adaptações com efeitos benéficos (CHE et al., 2014).

Desde a descoberta de IGS (BLUM et al.,1994), estudos extensos têm sido focados na sua detecção em diferentes genomas durante os últimos anos. Vários métodos para predição de IGS têm sido propostos, embora a precisão destes métodos não seja tão boa quanto desejável, é urgente desenvolver ferramentas computacionais para identificar todas as IGS.

Nesse trabalho identificamos e caracterizamos as IGS em organismos pertencentes ao gênero *Mycoplamas*. A caracterização das proteínas presentes nas IGS fornece uma maior base para a identificação e caracterização destas regiões que possuem genes diferentes do organismo e foram possivelmente adquiridas por transferência horizontal.



## 1.2 Objetivos

### Objetivo Geral

Identificação e caracterização de ilhas genômicas em bactérias do gênero *Mycoplasma* utilizando o método de clusterização *Mean Shift*.

### Objetivos específicos

Identificação das prováveis Ilhas Genômicas a partir do resultado do agrupamento do *Mean Shift*.

Caracterização das ilhas genômicas, pelo parâmetro proposto pelo COGs (Clusters de Grupos Ortólogos).

Comparação das ilhas de *Mycoplasmas* já descritas na literatura, com as aqui preditas.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Ilha Genômica (IG)

Estima-se que existam mais de 10 milhões de espécies diferentes ocupando as mais diversas regiões do planeta Terra (ALBERTS et al., 2015). Cada uma destas espécies possui as suas peculiaridades, diferenciando-se em aparência, tamanho e habitat. Todas estas características são determinadas pelo genoma de cada organismo (LEWIN, 2004).

Com o advento da técnica de sequenciamento, uma quantidade grande de informações acerca de sequências de DNA foi disponibilizada. Atualmente, é possível comparar genomas de microrganismos inteiros, com isso podemos identificar e comparar as semelhanças e as diferenças entre os organismos.

Uma descoberta que provém destas comparações é que os genomas microbianos podem abrigar uma quantidade significativa e variável de DNA de espécies diferentes da sua. Isto porque tem a capacidade de trocar material genético entre células não descendentes (mecanismos de transferência horizontal).

Desde que ocorreu o primeiro sequenciamento de um genoma, ficou claro que aquisição de um novo DNA é um mecanismo comum para a evolução bacteriana (OCHMAN et al., 2000), e que os genomas de todas as bactérias de vida livre apresentam um grande número de genes recém-adquiridos (NAKAMURA et al., 2004).

Um agente primário na mudança genômica, as Ilhas Genômicas (IGs), são formadas por um grupo de dezenas a centenas de genes cujos produtos podem cooperar para conferir funções complexas ao organismo (DOBRINDT et al., 2004).

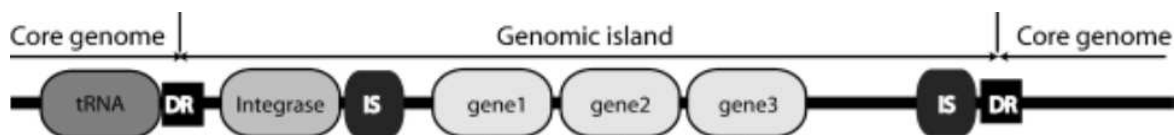
Ilhas genômicas (IGs) são, em essência, segmentos distintos de DNA, geralmente entre 10 e 200 kb de comprimento, que foram adquiridos por transferência horizontal de genes (THG) e são normalmente encontrados em alguns isolados de uma determinada espécie ou estirpe (HACKER; KAPER, 2000). As diferentes famílias de IGs, algumas evolutivamente muito antigas, foram reconhecidas com base na sequência e homologias funcionais. (BURRUS et al., 2002; JUHAS et al., 2007; VERNIKOS; PARKHILL, 2008).

As primeiras classes de IGs a serem descritas eram as ilhas de patogenicidade, assim chamadas, pois eram regiões cromossômicas instáveis com características e fenótipos variáveis associados à virulência (GROISMAN; OCHMAN, 1996).

Atualmente as IGs representam um grupo muito mais amplo, diversificado e de vários tamanhos, com uma grande abundância no genoma bacteriano (DOBRINDT et al., 2004). A capacidade de codificação das IGs não está limitada as funções de patogenicidade, mas possui vasta diversidade, incluindo inúmeras características, por exemplo, simbiose (SULLIVAN et al., 2002).

IGs são relativamente grandes segmentos de DNA, cujas características de nucleotídeos, diferem do resto do cromossomo. IGs, na maioria das vezes são inseridos em genes tRNA e ladeado por repetições diretas (do inglês, *Direct Repeats*, DR). IGs normalmente abrigam genes que estejam envolvidos na mobilidade genética, como integrases, transposases e sequências de inserção (do inglês, *Insertion Sequences*, IS), como mostra a Figura 1:

**Figura 1.** Descrição IGS



Fonte : JUHAS, 2009

IGs muitas vezes carregam genes que oferecem uma vantagem seletiva para as bactérias hospedeiras. De acordo com seu conteúdo genético, IGs podem ser descritos como patogenicidade, simbiose, metabólicas, de fitness ou ilhas de resistência (DOBRINDT et al., 2004; SCHMIDT; HENSEL, 2004).

As ilhas são descritas e caracterizadas por alguns fatores: (i) ilhas genômicas estão presentes nos genomas de muitas bactérias, mas ausentes nos genomas de grupos de descendentes estreitamente relacionados, (ii) em sua maioria possuem entre 10-200 kb; no entanto, existem algumas com tamanhos menores (1-10 kb) que são denominado de 'Ilhotas' genômicas, (iii) diferem em conteúdo GC e na sua utilização de códons, além de que podem ser reconhecidos pela quantidade de nucleotídeo, que geralmente diferem do resto do cromossomo, (iv) o locus do tRNA pode atuar como alvo para a integração de DNA estranho, (v) muitas vezes possuem genes ou pseudogenes que codificam para mobilidade genética, tais como integrases, transposases e elementos de inserção de sequência, (vi) além de serem frequentemente instáveis (HENTSCHEL; HACKER, 2001).

É amplamente reconhecido que a THG é facilitada por IGS, e desempenha um papel crucial na evolução das espécies bacterianas. Após a excisão a partir do genoma do hospedeiro, IGS podem desempenhar um papel importante na transferência de parte do cromossomo do hospedeiro para a bactéria receptora (HOCHHUT et al., 2000). Por exemplo, a presença de uma grande variedade de sistemas de secreção em muitas IGS sugere que estes são utilizados não apenas para a transferência de IGS, mas também para a transferência do DNA cromossômico do hospedeiro. IGS poderiam sofrer uma recombinação com o cromossomo do hospedeiro, com um impacto significativo sobre a evolução da bactéria hospedeira. Facilitando a evolução em “saltos quânticos”. A sua aquisição ou perda, pode alterar drasticamente o estilo de vida de uma bactéria (GROISMAN; OCHMAN, 1996; OCHMAN et al., 2000; VAN DER MEER; SENTCHILO, 2003; VERNIKOS et al., 2007).

Deve-se notar que muitos dos genes encontrados atualmente em IGS são novos e de função desconhecida, sem homólogos detectáveis em outras espécies, mas, no entanto, eles podem conferir vantagem seletiva para o organismo hospedeiro (HSIAO et al., 2005).

Vários trabalhos durante anos, demonstram a importância que IGS tem e tiveram em promover a THG, com a distribuição de uma gama de funções adaptativas para a bactéria hospedeira.

Várias linhas de pesquisas mostram evidências que sugerem a participação de IGS na evolução de vários grupos de espécies bacterianas, descobriram conjuntos conservados de genes no núcleo, que eram necessários para manutenção da bactéria. A contribuição de vários grupos de pesquisa ao redor do mundo tornou o processo de THG por meio de IGS, um dos campos mais investigados da microbiologia (JUHAS et al., 2009).

## 2.2 TRANSFERÊNCIA HORIZONTAL DE GENES (THG)

Em relação à estrutura celular, os seres vivos se dividem em dois grandes grupos: Eucariotos e Procariotos. Os Eucariotos possuem seu material genético em um compartimento limitado por uma membrana chamada núcleo e compreendem as plantas, fungos e animais. Já os Procariotos não possuem seu DNA limitado por nenhuma membrana (ALBERTS et al., 2015).

A transferência da informação genética nos eucariotos é feita basicamente de forma vertical, através da reprodução, porém nos organismos procariotos ela também pode ocorrer de outra forma (MADIGAN et al., 2015).

Nestes, a transferência de informações pode ocorrer tanto vertical, como horizontalmente. Na primeira forma, as informações genéticas são passadas da célula pai para a célula filha, durante a fase de divisão celular (LAWRENCE; ROTH, 1996).

Na transferência horizontal, ocorre a transmissão não genealógica de material genético de um organismo para outro, sendo uma fonte de novos genes e funções para o receptor desse material genético que foi transferido (GOLDENFELD; WOESE, 2007).

A transferência horizontal de genes (THG) é a maneira que organismos possuem de trocar informações genéticas com outros organismos, (BROWN, 2003). Em bactérias, essa troca de informação pode ser feita de três maneiras transdução (CALENDAR, 1988), transformação (LORENZ; WACKERNAGEL, 1994) ou conjugação (CLEWELL, 1993).

A transdução é um mecanismo de transferência gênica entre bactérias mediada por bacteriófagos (fagos), que são vírus especializados em infectar células bacterianas (TENOVER, 2006). No momento em que o material genético do bacteriófago é introduzido, o tipo de transdução varia dependendo da biologia do fago e do ambiente celular. Na transdução generalizada, todos os genes contidos dentro de uma célula infectada, sejam eles pertencentes ao DNA cromossômico, plasmidial ou até mesmo de outro vírus podem ser empacotados junto ao DNA do fago dentro do capsídeo proteico e transferidos a outras bactérias após a lise. Entretanto, na transdução especializada somente genes bacterianos específicos são transferidos, independente de sua localização, sendo considerado um processo mais eficiente (TORTORA et al., 2012).

Os bacteriófagos são considerados um importante meio pelo qual os microrganismos se adaptam a novos ambientes de forma mais permanente. Tal característica oferece vantagem seletiva as bactérias patogênicas e conseqüentemente maior impacto em saúde pública, principalmente em relação a transdução de genes de resistência a antimicrobianos (DE LA CRUZ & DAVIES, 2000).

Transformação é o processo pelo qual a bactéria é capaz de realizar recombinação genética por meio da absorção ativa de DNA extracelular proveniente de diversas fontes como cromossomos, plasmídeos e fagos, dispersos no ambiente. Para tanto é necessário que a bactéria esteja em estado de competência, cujas alterações

fisiológicas permitem a captação do DNA doador (BRIGULLA & WACKERNAGEL, 2010). Assim, se o DNA livre for composto por genes de resistência a antibióticos, a célula receptora provavelmente adquire a resistência correspondente, caso a integração ao material genético seja eficiente (FOLEY & LYNNE, 2008). Portanto, os requisitos necessários para que ocorra a transformação natural é a liberação e persistência do DNA doador no ambiente, a presença de uma célula receptora competente e por fim a capacidade de integração ao DNA receptor (KELLY et al., 2009).

A conjugação é um mecanismo de transferência gênica entre bactérias no qual contrariamente à transformação, existe a necessidade de um contato entre a célula doadora de material genético e a célula receptora. Esse processo é orientado por elementos genéticos móveis que transportam genes responsáveis pela codificação de funções para sua própria transferência a outras bactérias hospedeiras, além de outras atividades como resistência a antimicrobianos (THOMAS; NIELSEN, 2005; KELLY et al., 2009). Por isso, a conjugação frequentemente é associada à plasmídeos, pequenos elementos genéticos que se multiplicam independente do cromossomo da célula bacteriana (TORTORA et al., 2012). O processo é mediado por um pili conjugativo (pili sexual), uma estrutura proteica semelhante a uma haste que auxilia na união das células. Após o contato inicial entre as duas células bacterianas, as membranas celulares se fundem promovendo estabilidade. Uma cópia de fita simples do DNA plasmidial é transferido a partir de uma sessão denominada origem de transferência, para a célula receptora (BAUMAN, 2009). Pelo sistema de conjugação, é possível transferir entre duas bactérias não somente um plasmídeo completo, mas sequências extensas de quaisquer elementos genéticos, desde que estejam integrados fisicamente à sessão da origem de transferência (THOMAS; NIELSEN, 2005). A conjugação é considerada o principal processo de disseminação de genes de resistência a antibióticos entre as populações bacterianas (ROWE-MAGNUS; MAZEL, 1999).

### 2.2.1 Ascensão da THG

Os fundadores da filogenia molecular usaram a informação molecular de diferentes proteínas e genes para reconstruir as relações filogenéticas entre os organismos (ZUCKERKANDL; PAULNG, 1965). Logo perceberam, no entanto, que os genes diferentes poderiam levar as incongruências filogenéticas e filogenias conflituosas, principalmente em microrganismos, ao agrupar grupos de espécies ou

espécies que são divididas por outros marcadores morfológicos, fisiológicos ou moleculares (HILARIO; GOGARTEN, 1993; GUPTA; SINGH, 1994; GOLDING; GUPTA, 1995; WHITEHOUSE et al., 1998).

Por outro lado, os mecanismos para a transferência de material genético entre microrganismos são conhecidos desde o início na biologia molecular e na área da pesquisa da genética molecular (LEDERBERG; TATUM, 1946; ZINDER; LEDERBERG, 1952; STOCKER et al., 1953). O conceito de transferência horizontal de genes entre organismos surgiu no início da década de 1990 como uma explicação alternativa para os eventos de filogenia conflituosa (HILARIO; GOGARTEN, 1993). Desde então, novos e abundantes dados, reforçam essa ideia, especialmente com o surgimento da era genômica, o que permitiu a comparação de conjuntos completos de genes entre organismos.

### 2.2.2 Importância da THG

Desde o surgimento da THG, um número considerável de estudos tem sido publicados sobre os genes que foram adquiridos dessa forma, (GOGARTEN et al., 2002; LERAT et al., 2005), tanto em bactérias (SAUNDERS et al., 1999; OCHMAN et al., 2000) e archaea (DOOLITTLE; LOGSDON, 1998; FAGUY; DOOLITTLE, 1999), bem como em eucariotos (ANDERSON, 2005). Estes estudos mostram que a transferência genética pode ocorrer entre organismos, mas também entre os domínios em todas as direções possíveis: de bactérias para archaea (REST; MINDEEL, 2003), de archaea para as bactérias (GOPHNA et al., 2004), de archaea para eukarya (ANDERSON et al., 2003), das bactérias para eukarya (WAKTINS; GREY, 2006), a partir de eukarya para bactérias (GULJAMOW et al., 2007) e até mesmo entre eukarya (NEDELICU et al., 2008). No entanto, é na evolução bacteriana e de archaea que a THG tem sido aceita e documentada.

A identificação da THG, ainda é uma questão a resolver, diversos critérios tem sido propostos e discutidos que permita essa identificação com uma menor dificuldade, com o viés no uso de códons e na composição de bases que são diferentes em relação aos outros genes do genoma (RAGAN 2002; LAWRENCE; OCHMAN, 2002).

Critérios baseados na utilização de códons e na composição de base foram objetos de diversas críticas, já que o conjunto de genes, os que residem e os que foram transferidos, passam pelas mesmas tendências mutacionais, uma vez que fazem parte do

mesmo genoma, desta forma os critérios poderiam permitir a identificação dos eventos recentes de THG, mas podem ser ineficientes na detecção de eventos que ocorreram há muito tempo (KOSKI ET AL., 2001; KUO; OCHMAN, 2009).

### 2.2.3 THG e Evolução Microbiana

Uma grande quantidade de trabalho tem sido realizada nos últimos anos, para tentar lidar com a importância da THG na evolução de bactérias e archaeas (FAGUY; DOOLITTLE, 1999; NELSON et al., 1999; OCHMAN et al., 2000; OCHMAN, 2001; GOGARTEN et al., 2002; KURLAND et al., 2003; PHILIPPE; DOUADY, 2003; PÁL et al., 2005; CHOI; KIM, 2007; KOONIN; LOBO, 2008), que nos mostra que o mecanismo, desempenha um papel maior na evolução microbiana maior do que se pensava anteriormente.

Vários estudos sugerem que a THG pode ser efetivamente mais frequente para distâncias evolutivas curtas e intermediárias, mas raro entre organismos separados por grandes períodos de tempo evolutivo (OCHMAN et al., 2000; BRUGGER et al., 2002; NAKAMURA et al., 2004; GE et al., 2005; CHOI KIM, 2007; DAGAN et al., 2008).

Wagner et al. (2008) analisaram a evolução de 2091 sequências de inserção em 438 genomas de procariotos completamente sequenciados e encontraram apenas 30 casos de transferência entre clades distantemente relacionadas, 23 destes eventos pareciam ser antigos, enquanto apenas sete foram recentes.

Os casos de transferências de genes entre archaea e bactéria que foram descritos (REST; MINDELL, 2003; GOFNA et al., 2004), mostram que a THG pode afetar a evolução nos procariotos ao longo das diferentes épocas evolutivas.

Os estudos sugerem que, apesar do fato de que a THG pode ser mais frequente entre organismos relacionados intimamente, também pode ter lugar entre os organismos distantemente relacionados, contribuindo para a evolução microbiana.

A THG é uma força importante na modulação da evolução nos procariotos, apesar da troca de genes serem mais fáceis em organismos estreitamente relacionados, a THG ocorreu entre ambos os domínios.



## 2.3 Métodos para detecção de ilhas genômicas.

Existem dois métodos para a detecção de IGs em bactérias. Método baseado em comparação de genomas com uso de sequências genômicas de espécies filogeneticamente próximas, e outro método baseado em composição de sequência, que utiliza apenas o genoma da bactéria de interesse para encontrar IGs,

### 2.3.1 Métodos baseados em comparação genômica

Com o crescimento de genomas sequenciados, a abordagem do uso da análise genômica comparativa para detectar IGs se torna possível. Os genomas de espécies filogeneticamente próximas são utilizados para comparação. Os métodos baseados na comparação genômica comparam múltiplas sequências genômicas para prever IGs. O genoma investigado é alinhado com dois ou mais genomas relacionados. As regiões presentes em ambos os genomas provavelmente não têm origem horizontal, enquanto as regiões encontradas apenas no genoma da bactéria investigada são consideradas IGs.

Dentre os métodos de comparação se destacam, **MobilomeFinder** (OU et al., 2007), que requer a seleção manual dos genomas utilizados para comparação, **MOSAIC** (CHIAPELLO et al., 2005) o qual utiliza um conjunto de dados de referência, com um simples tratamento dos resultados de alinhamento que permite uma definição robusta da estrutura em mosaico e o **IslandPick** (LANGILLE et al., 2008) que realiza uma busca automática.

A principal limitação dessa abordagem é a dependência da disponibilidade de genomas relacionados que foram sequenciados, cerca da metade dos genomas não têm um número mínimo de espécies / linhagens relacionadas para que possa ser feita análise comparativa entre os genomas. Assim, a detecção de IGs em tais genomas pode não ser aplicável, no entanto com o crescente número de genomas sequenciados, essa dificuldade tende a diminuir.

### 2.3.2 Métodos baseados em composição de sequência

Esse método se baseia em que o genoma de uma espécie possui uma assinatura única de sequência, então IGs podem ser detectadas pela sua composição que difere do resto do genoma da bactéria (CHE et al., 2001).

A composição de bases e o padrão no uso de códon, são bem semelhantes entre os genes de um genoma, podemos diferenciar as sequências que foram adquiridas por transferência horizontal de genes, devido às diferenças que apresentam em relação ao resto do genoma (LAWRENCE; OCHMAN, 1998).

Uma simples medida de diferenciação é o conteúdo de GC. Nesta medida toma-se uma janela deslizante de tamanho usual entre 10 e 50 kb e calcula-se a frequência das bases G e C para cada janela, comparando-as com a frequência média do genoma. Janelas cujas frequências diferirem da frequência média esperada são consideradas IGs. (KARLIN, 2001).

A maioria dos métodos baseados em composição de sequência requer uma medida da composição da sequência para capturar a assinatura do genoma, os principais métodos baseados em composição de sequência são: **IslandPath** (HSIAO et al., 2003) que calcula a frequência de 24 dinucleotídeos, combinada com a presença de pelo menos um gene típico de THG para a predição de IGs em bactérias; **PAI-IDA** (TU; DING, 2003) utiliza análise discriminante para detectar regiões genômicas que desviam do resto do genoma, de acordo com três critérios de composição: conteúdo de GC, frequência de dinucleotídeos e uso de códons.

## 2.4 Clusterização

Bioinformática é a ciência de gerenciar, minerar, integrar e interpretar informações de dados biológicos em diversos níveis do organismo (WANG et al., 2005). O acúmulo de dados biológicos em bancos de dados em várias partes do mundo demanda o desenvolvimento de novas ferramentas para interpretar este grande volume de dados (WANG et al., 2005; ATTWOOD et al., 2011).

A Descoberta de Conhecimento em Banco de Dados (do Inglês *Knowledge Discovery in Databases*, KDD) surgiu da necessidade de extrair informações úteis de grandes conjuntos de dados e desempenha um papel importante na análise de dados biológicos e solução de problemas emergentes na área de bioinformática (WANG et al., 2005). A Mineração de Dados é a principal etapa no processo de KDD, que também envolve a aquisição das informações, pré-processamento, transformação e interpretação dos resultados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Mineração de dados consiste na aplicação de algoritmos para extrair padrões de interesse dos dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). A

Clusterização é um método de mineração de dados. Algoritmos de Clusterização são utilizados para separar elementos de um conjunto de dados, em grupos de objetos, de acordo com a sua semelhança. Estes grupos são conhecidos como *clusters* ou agrupamentos (JAIN; DUBES, 1988; JAIN; MURTY, 1999)

## 2.5 – *Mycoplasmas*

*Mycoplasmas* pertencem à classe dos *Mollicutes*. Eles estão entre os organismos de menor tamanho de genoma de 580 kb a 2.2 mb , além de serem capazes de autorreplicação em meio isento de células (LO, 1992), como bactérias, elas não possuem núcleo, mas contêm DNA e RNA, porém ao contrário da maioria das bactérias, eles não portam qualquer tipo de parede celular (MARTY, 1993), causam doenças graves em muitas espécies animais (assim como plantas), onde eles podem afetar sistemas de múltiplos órgãos ou causar doenças crônicas. *Mycoplasmas* são muitas vezes difíceis de detectar e de erradicar devido a mecanismos de evasão e modulação da resposta imune do hospedeiro (BASESMAN et al., 1996).

As proteínas de *Mycoplasma* são suficientemente semelhantes às proteínas animais, com isso, o sistema imune do corpo não reconhece à *Mycoplasma* como uma estrangeira ou pode fazer com que o organismo produza anticorpos que ataquem às células animais e acabam produzindo uma doença autoimune (TULLY et al., 1996).

### 2.5.1 *Mycoplasma Agalactiae*

A *M. agalactiae*, possui um genoma de tamanho de 1,05 mb, com um conteúdo GC de 29,7%, ela é responsável pela agalactia contagiosa em pequenos ruminantes domésticos, uma síndrome listada pela Organização Mundial de Saúde Animal e responsável por graves danos para a indústria de laticínios (CORRALES et al., 2007)

Possui uma difícil erradicação já que a *M. agalactiae* continua a infectar novas manadas por muitos anos após a infecção inicial. Embora as características da doença estejam bem documentadas, pouco se sabe sobre os determinantes de patogenicidade de *M. agalactiae* ou os mecanismos envolvidos na infecção e na sua resistência (DAMASSA et al., 1992).

### 2.5.2 *Mycoplasma arthritidis*

*M. arthritidis* é um agente patogênico natural de ratos, causando severa poliartrite, com lesões semelhantes às observadas na artrite reumatoide humana.

O organismo foi isolado de camundongos, o seu sequenciamento mostrou um genoma de tamanho de 1 mb, com um conteúdo GC de 30.7%. (WASHBURN, 2002).

### 2.5.3 *Mycoplasma bovis*

O *M. bovis* que foi identificado pela primeira vez como um agente causador da mastite em 1961 e reconhecido como um importante agente patogênico bovino causador de doenças respiratórias em 1976, possui um genoma de tamanho de 1,15 mb e um conteúdo GC de 29,3% (CASWELL et al., 2008).

A pneumonia causada pela *M. bovis* tornou-se mais comum com o desenvolvimento da indústria de carne, induzida pelos transportes de longa distância, com uma péssima condição de confinamento. Uma vez que este microrganismo é resistente a vários antibióticos, incluindo beta-lactâmicos, não possui nenhuma vacina eficaz disponível comercial, *M. bovis* tem causado uma perda econômica significativa nos Estados Unidos, Canadá e a maioria da Europa (WHITE et al., 2010).

### 2.5.4 *Mycoplasma bovoculi*

*M. bovoculi*, a primeira sequência completa do genoma desse organismo foi isolado em 1968, a partir de um olho bovino no Canadá, possui um genoma de tamanho de 0.95 mb e um conteúdo GC de 28.2% (LANGFORD et al., 1973).

É um agente etiológico da conjuntivite bovina e está também associado com o desenvolvimento da infecção mais grave, queratoconjuntivite bovina (POSTA et al., 2008).

### 2.5.5 *Mycoplasma californicum*

*M. californicum* é uma bactéria que causa a mastite, apresenta forte infectividade, e possui uma má resposta ao tratamento com antibióticos, gerando grandes perdas econômicas. Apesar de sua importância, pouca informação genética da *M. californicum* está disponível (HEWICKER-TRAYTWEIN et al., 2002).

O organismo foi isolado a primeira vez na Califórnia, depois de um surto de mastite (Jasper, 1981), mas devida a sua importância e poucas informações genéticas, foi isolada no Japão a partir do leite bovino, para que possa se fornecer alguma base para investigação dessa espécie no futuro, possui um genoma de tamanho 0.95 mb e um conteúdo GC de 30,8% (NOGUCHI et al., 2008).

### 2.5.6 *Mycoplasma canadense*

*M. canadense* é uma bactéria que causa a mastite bovina, a infecção gerada por essa bactéria resulta em uma diminuição imediata e acentuada na produção de leite, uma enorme quantidade de células inflamatórias infiltrando o leite de todos os quadrantes, além de que persiste por longos períodos (BALL, 1986).

Apesar da importância da sua importância, pouca informação genética sobre o organismo está disponível. A sequência inteira do seu genoma foi isolada a partir do leite bovino no Japão, a *M. canadense* possui um genoma de 0.85 mb, com um conteúdo GC de 24.3%. (NOGUCHI, 2008).

### 2.5.7 *Mycoplasma Capricolum*

*M. capricolum* possui um genoma de tamanho 1.2 mb com um conteúdo GC de 23.7%, é um agente etiológico da agalaxia, doença que afeta pequenos ruminantes, associada com inflamação crônica, artrite e mastite. Embora *M. capricolum* seja considerado um dos membros menos patogênicos, os surtos provocados por este agente pode ter um impacto significativo sobre as indústrias caprinas devido à perda de produção de leite e aumento da mortalidade (GÓMEZ-MARTÍN et al., 2013)

### 2.5.8 *Mycoplasma conjunctivae*

*M. conjunctivae*, contém um genoma de tamanho 1 mb com um conteúdo GC de 28,6%, é um agente etiológico da ceratoconjuntivite infecciosa, causa infecções oculares graves que levam à cegueira e perfuração da córnea, particularmente em *Alpine ibex* e *camurças* (BARILE, 1972). Tendo em vista as condições físico-químicas severas que protegem o olho de ser infectado por microrganismos patogênicos, *M. conjunctivae* tem uma função eficiente para a adesão, que não a deixa ser expulsa, pelo fluido lacrimal (HANG et al., 1995).

### 2.5.9 *Mycoplasma Crocodyli*

O organismo foi isolado a partir de articulações e pulmões de crocodilos. Esta bactéria que provoca poli artrite em crocodilos, seu sequenciamento revelou um genoma de tamanho de 1.10 mb e o seu conteúdo GC é de 27,6%. (KIRCHHOFF, 1997).

### 2.5.10 *Mycoplasma cynos*

*M. Cynos* contém um genoma de tamanho de 1.15 mb com um conteúdo GC de 25.7%, parasita do trato respiratório canino. Várias espécies de *Mycoplasma* foram isoladas das vias respiratórias de cães com doença clínica. A Doença Respiratória Infeciosa Canina (DRIC) é uma doença multifatorial, envolvendo ambos os agentes virais e bacterianos (CHALKER et al., 2004).

Chavla (2007), sugere em evidências recentes que *M. cynos*, em particular, é um importante agente etiológico da DRIC. Não se sabe sobre o papel que *M. cynos* desempenha em DRIC, se a bactéria é um agente patogênico ou um patógeno secundário oportunista.

### 2.5.11 *Mycoplasma díspar*

*M. díspar* é frequentemente isolada de animais com doenças respiratórias, mas também pode ser detectada em animais saudáveis, normalmente transmitida pelas secreções respiratórias dos animais infectados (TEGTMEIER et al., 1999). Esta espécie foi isolada em diferentes países, possui um genoma de tamanho de 1.25 mb, com um conteúdo GC de 29%.

### 2.5.12 *Mycoplasma fermentans*

O sequenciamento *M. fermentans*, mostrou um genoma de tamanho de 1.15mb com um conteúdo GC de 29.6%, além de estar associada com várias doenças humanas. Por exemplo, a presença de *M. fermentans* com a síndrome da imunodeficiência adquirida (SIDA) dos pacientes foi relatado em 1989 (LO et al., 1989).

Seya (2002), sugere que o potencial de *M. fermentans* para interagir com o sistema imunitário tem sido intensivamente investigado. Estudos têm demonstrado que as proteínas associadas a membranas de *M. fermentans* podem interagir com monócitos

e macrófagos, *M. fermentans* foi proposta como um cofator do vírus da imunodeficiência humana para a transmissão e progressão da virulência.

#### 2.5.13 *Mycoplasma flocculare*

O sequenciamento de *M. flocculare*, mostrou um genoma de tamanho de 0.95 mb com um conteúdo GC de 28,9% (SIQUEIRA et al., 2013.), geralmente considerada um colonizador da nasofaringe suína, além de ser um agente patogênico oportunista em coinfeccões provocadas por outras Mycoplasma. (OPRIESSNIG et al., 2011).

#### 2.5.14 *Mycoplasma Gallisepticum*

*M. gallisepticum* contém um genoma de tamanho 1,15mb com um conteúdo GC de 31.6%, agente patogênico do trato respiratório em aves, causando uma inflamação grave da traqueia, pulmões e nos sacos de ar, em especial na presença de uma coinfeccão (SATO et al., 1970). Este agente patogênico invade, sobrevive e multiplica-se dentro de uma variedade de células não fagocíticas, tais como glóbulos vermelhos de galinha, células HeLa, e fibroblastos de frango (WINNTER et al., 2000).

Além disso, *M. gallisepticum* é conhecido por colonizar diversos tecidos extrapulmonares, incluindo o sangue, coração, baço, fígado e cérebro (INDIKOVA et al., 2013).

#### 2.5.15 *Mycoplasma genitalium*

Identificada em 1980, *M. genitalium* é uma bactéria da classe das *Mollicutes* que coloniza o aparelho reprodutor masculino e feminino. Dado o pequeno tamanho do seu genoma 0.75mb e conteúdo GC de 31.7%, foi uma das primeiras bactérias a ser totalmente sequenciada e o primeiro genoma sintetizado quimicamente. (FRASER, 1995)

*M. genitalium* causa uma infecção transmitida sexualmente, é prevalente e emergente epidemiologicamente, está relacionada à doença inflamatória pélvica e cervicite (TAYLOR-ROBISON, 2011).

### 2.5.16 *Mycoplasma Haemocanis*

*M. Haemocanis* possui um tropismo para os glóbulos vermelhos hospedeiros vertebrados, é um agente patogênico do sangue que pode causar anemia em cães (HULME-MOIR et al., 2010.)

O DNA da bactéria foi isolado a partir do sangue de um cão infectado, foi totalmente sequenciado, mostrou um genoma um tamanho de 1.10 mb, com um conteúdo GC de 35.3% (NAÍLA, 2012).

### 2.5.17 *Mycoplasma haemofelis*

*M. haemofelis* é uma infecção comum em gatos que provoca a anemia hemolítica aguda, que destrói ou remove hemácias pela atuação de anticorpos antieritrocitários. Uma ampla variedade de sintomas clínicos, incluindo a anemia, pirexia, letargia, e esplenomegalia caracteriza a doença, que se não for tratada pode resultar em morte (FELDER et al., 2011).

O DNA da *M. haemofelis* foi purificado a partir do sangue retirado de um gato infectado em um momento de alta parasitemia, foi identificado um genoma de 1.35 mb, com um conteúdo GC de 38.85% ( PETERS et al., 2010).

### 2.5.18 *Mycoplasma hominis*

*M. hominis* é um habitante comum vaginal que está associado com a vaginose bacteriana. A bactéria é considerada inofensiva para a maior parte das mulheres que não estão grávidas, mas em gestantes pode causar infecções intra-amnióticas, causando uma ruptura prematura das membranas, gerando um nascimento prematuro (MASSARO et al., 2009).

Sequenciou o genoma completo, a partir do líquido amniótico e da placenta de um isolado de *M. hominis*, de gravidezes que resultaram em nascimentos prematuros, identificou um genoma com um tamanho de 0.85 mb, com um conteúdo GC de 26.9% (MATTHEW et al., 2015).



### 2.5.19 *Mycoplasma hyopneumoniae*

*M. hyopneumoniae* contém um genoma de tamanho 1.05 mb com um conteúdo GC de 28.5%, causa pneumonia enzoótica, que é uma pneumonia leve, que acomete os suínos. Este organismo altamente infeccioso tem uma distribuição mundial.

A infecção por *M. hyopneumoniae* resulta em lesões pulmonares graves e perdas de produção, um relativo controle foi conseguido através de programas de vacinação, mas a pneumonia enzoótica continua a ser um grande problema econômico na indústria de suínos (MAES, 2008).

### 2.5.20 *Mycoplasma hyorhinis*

*M. hyorhinis* é um patógeno suíno que causa uma doença crônica, após a sua colonização inicial nas superfícies mucosas do trato respiratório. *M. hyorhinis* também pode infectar células humanas. A sequência completa do genoma de *M. hyorhini* foi derivada de animais naturalmente infectados, possui um genoma de tamanho de 1 mb, com um conteúdo GC de 25.9%. (ELKIN et al., 2010.)

### 2.5.21 *Mycoplasma Leachii*

*Mycoplasma leachii* foi inicialmente isolado a partir de fluidos de bezerros do sul de Queensland, na Austrália, o seu sequenciamento gerou um genoma de tamanho de 1.20 mb, com um conteúdo GC de 23.5% e sua patogenicidade foi estabelecido por infecção experimental, foi relatada com pouca frequência como causa de poli artrite em bezerros e mastite em vacas; o patógeno também foi isolado a partir de fetos abortados pulmão bovino e de pequenos ruminantes anfitriões (MANSO-SILVÁN et al., 2009).

### 2.5.22 *Mycoplasma meleagridis*

*M. Meleagridis* é relatado como um patógeno específico de perus que provoca mortalidade embrionária, anormalidades esqueléticas e mau desempenho do crescimento na descendência infectada (Yamamoto, 1991).

No entanto, o isolamento ocasional de frangos que apresentem sinais de doenças respiratórias e/ou perdas de produção de ovos sugere que *M. meleagridis* também pode ser um agente patogênico de frangos. Para começar a responder a estas questões, foi

realizada a sequência completa do genoma de *M. Meleagridis*, com um genoma de tamanho de 1mb, e o seu conteúdo GC de 25,9%. (KHIARI et al., 2011).

### 2.5.23 *Mycoplasma mobile*

*M. mobile* é uma das *Mycoplasmas* em forma de frasco (1 µm x 0,3 µm), foi originalmente isolada a partir de um peixe (KIRCHHOFF; ROSENGARTEN 1984).

Acredita-se que seja um agente patogênico de mamíferos, cresce a cerca de 20 °C, um pouco menor do que a maioria das outras *Mycoplasmas* estudadas, cujo (ILOFF, 2002). No entanto, única característica mais marcante de *M. mobile* é a sua capacidade de deslizar (definida como translocação suave, enquanto em contato com uma superfície) em superfícies de vidro ou plástico em velocidades de até 7 mm/s (ROSENGARTEN; KIRCHHOFF 1987).

O sequenciamento do genoma da *M. mobile*, gerou um genoma de tamanho de 0.95 mb, com um conteúdo GC de 25% (JAFFE; CHURCH, 2004).

### 2.5.24 *Mycoplasma mycoides*

O sequenciamento do genoma *M. mycoides*, mostrou um genoma de tamanho de 1.19 mb e o seu conteúdo GC de 24%, os membros da espécie das *M. mycoides*, estão entre os mais virulentos das *Mycoplasmas*, causando doenças em todo o mundo, com perdas significativas de gado bovino e caprino ( RODRIGUEZ et al., 1983; MANSO-SILVAN, 2007).

### 2.5.25 *Mycoplasma ovis*

*M. ovis*, foi descrito pela primeira vez em 1934. O seu sequenciamento, mostrou um genoma de tamanho de 0.90 mb e um conteúdo GC de 31.7%. Este organismo pode causar uma infecção aguda em ovinos podendo resultar em síndrome de anemia, afetando a produção industrial. (NEIMARK et al., 2004).

### 2.5.26 *Mycoplasma parvum*

*M. parvum*, parasita suíno descrita pela primeira vez por Splitter (1950), é relatada por se acumular frequentemente em grandes números em um único glóbulo vermelho. Foi sequenciada a partir do sangue de um porco, que foi naturalmente

infectado, o seu genoma é de tamanho de 0.75 mb, e o seu conteúdo GC é de 27%. (NAÍLA et al., 2013).

#### 2.5.27 *Mycoplasma penetrans*

*M. penetrans*, infecta seres humanos na via urogenital e respiratória. Uma característica típica é a penetração nas células humanas, essa bactéria está associada principalmente com a infecção por HIV-1 que contribui para a deterioração do sistema imune, particularmente em adultos na Europa e nos EUA (FINCH, 1992).

O *M. penetrans* foi isolado a partir de um paciente HIV-negativo, com uma doença sistêmica associada à infecção por *M.penetrans*, o organismo possui um genoma de tamanho de 1.55 mb, com um conteúdo GC de 25.7 % ( YÃNEZ et al., 1999).

#### 2.5.28 *Mycoplasma pneumoniae*

O sequenciamento de *M. pneumoniae* mostrou um genoma de tamanho de 0.81 mb com um conteúdo GC de 40%. Patógeno comum do trato respiratório superior e inferior de seres bovinos de todo o mundo (KUHNER et al., 2009).

Embora a infecção seja suave, na maioria dos casos, os pacientes podem desenvolver ocasionalmente essa infecção para uma doença fatal (NARITA, 2010).

#### 2.5.29 *Mycoplasma pulmonis*

*M. Pulmonis*, agente etiológico da micoplasmose respiratória em ratos e camundongos. Considerado como o modelo mais útil para o estudo de infecções respiratórias derivadas de *Mycoplasmas*, incluindo a infecção causada por mielofibrose primária em seres humanos (CITTI, 1997).

A análise da sequência completa do genoma de *M.pulmonis*, revelou que ele fornece pistas para decifrar os respectivos papéis dos fatores de virulência, o seu genoma possui um tamanho de 1.15 mb, com um conteúdo GC de 26.6% (CHAMBAUD, 2011).

#### 2.5.30 *Mycoplasma putrefaciens*

*M. putrefaciens*, agente etiológico da agalactia contagiosa em caprinos, uma síndrome listada pela Organização Mundial de Saúde Animal (NICHOLAS, 2008), a

síndrome inclui mastite que pode ser acompanhada por conjuntivite e poli artrite. Foi identificada em 1976 (TULLY et al., 1976).

O sequenciamento dessa bactéria, mostrou um tamanho de 1.05 mb, com um conteúdo GC de 26.95%, além de possibilitar análises comparativas para obter informações o polimorfismo interespecies. (CALCUT, 2011).

#### 2.5.31 *Mycoplasma suis*

*M. Suis* é um parasita dos eritrócitos de suínos, a bactéria pode invade o eritrócito levando consigo uma alta virulência, fora a sua resistência aos antibióticos, além disso, evidências crescentes apontam que *M. suis* também é um agente zoonótico. O DNA foi isolado a partir de leitões infectados, onde possui um genoma de tamanho de 0.90 mb, com um conteúdo GC de 31.1%. (GROEBEL et al., 2009).

#### 2.5.32 *Mycoplasma synoviae*

*M. synoviae*, principal patógeno de aves de capoeira comerciais, gerando osteoartrite, sinusite e lesões no trato respiratório das aves, tem como principais mecanismos patogênicos, a citoaderência e a hemadsorção (KANG, 2002).

Um conjunto de uma sequência híbrida do *M. synoviae* completa foi sequenciada, gerando um genoma de tamanho de 0.95 mb, com um conteúdo GC de 28.3% (May et al., 2015).

#### 2.5.33 *Mycoplasma wenyonii*

*M. wenyonii* é um parasita de bovinos, gerando infecções que associam a perda de peso, infertilidade e diminuição da produção de leite (GENOVA, 2011). Possui uma baixa morbidez e os animais podem tornar-se portadores crônicos (FUJIHARA, 2011).

O genoma completo de *M. wenyonii* foi sequenciado, montado e anotado, para que se pudesse entender melhor a sua patogenicidade, metabolismo e evolução, o seu genoma possui um tamanho de 0.85 mb, com um conteúdo GC de 33.9%. (ANDREA et al., 2012).

#### 2.5.34 *Mycoplasma yeatsii*

*M. yeatsii*, é um patógeno de ruminantes, é costumeiramente isolado do canal auditivo dos caprinos, foi isolado a partir de uma cabra na Califórnia, possui um genoma de tamanho de 1.05 mb, com um conteúdo GC de 25.75%. (KENT et al., 2012).

### 3 METODOLOGIA

Neste trabalho, foi utilizado um método alternativo para detecção de IGs em genomas de *Mycoplasma*, através da utilização do algoritmo *Mean Shift*.

Fundamenta-se na hipótese de que é possível separar as sequências genômicas adquiridas horizontalmente das originais através de um algoritmo de clusterização, que é um método de mineração de dados, utilizado para separar elementos de um conjunto de dados, em um grupo, de acordo com sua semelhança, esses grupos são chamados de *clusters*, que possibilita a detecção das IGs. Os *clusters* são as próprias IGs, pois contém os fragmentos potencialmente adquiridos por transferência horizontal, já que partimos do princípio que os agrupamentos separados do genoma original contém genes de outros organismos, devido a sua composição distinta, coloca os fragmentos que contém os genes estranhos em *clusters* diferentes daqueles que contém os fragmentos nativos do genoma.

Fukunaga (1975) propôs o mean-shift inicialmente em 1975, para ser aplicado na área de reconhecimento de padrões, o método possui aplicações em diversas áreas, tais como processamento de imagens, visão computacional e mineração de dados.

O método utilizado foi desenvolvido pelo bacharel em ciências da computação Daniel Miranda de Brito do Centro de Informática da Universidade Federal da Paraíba, que propôs um método alternativo para detecção de IGs em bactérias com o algoritmo de clusterização, utilizando as bases ATCG como unidade de composição de sequência.

#### 3.1 Parâmetros

O método alternativo que utilizamos para detecção de IGs possui três parâmetros ajustáveis: o tamanho dos fragmentos (janelas), coeficiente do desvio padrão para seleção dos fragmentos artificiais e quantidade de fragmentos artificiais selecionados, os dois últimos parâmetros são utilizados para o ajuste automático do parâmetro *bandwidth* (parâmetro usado pelo *Mean Shift* para fazer a mineração de dados nos fragmentos, calculado automaticamente) ótimo do algoritmo de clusterização, que é necessário para que as IGs sejam encontradas.

### 3.1.1 Tamanho dos fragmentos

O primeiro parâmetro é o tamanho dos fragmentos (janela), que define a unidade básica de predição, do ponto de vista de algoritmos de agrupamento, o ponto no espaço d-dimensional (onde d é 4, pois consideram-se as bases A,T,C e G). Para exemplificar a escolha desse parâmetro, foi feita uma sequência de um genoma qualquer:

**Figura 2.** Segmento de um genoma representado no formato FASTA.

```
>Sequência
"TAAGTTATTATTAG"
```

Fonte: Elaborado pelo autor

Foi selecionada uma janela (fragmento) de tamanho 5, consideram-se os fragmentos "TAAGT", "TATTA" e "TTTAG" na predição de IGs. A Figura 3 mostra como são vistos como pontos no espaço 4-dimensional:

**Figura 3.** Lista de pontos interpretados para o genoma

P1:(2.0, 2.0, 0.0, 1.0)

P2:(2.0, 3.0, 0.0, 0.0)

P3:(1.0, 3.0, 0.0, 1.0)

Fonte: Elabora pelo autor

O método divide o genoma original em fragmentos de tamanhos iguais, cujo valor é fornecido pelo usuário. Para realizar os testes nos genomas Mycoplasma, foi utilizado o tamanho de janela 50kb (50000 nucleotídeos).

### 3.1.2 Coeficiente do desvio padrão

O segundo parâmetro define o quão distante estão os fragmentos artificiais dos fragmentos originais do genoma. Para isso, é calculada a média e desvio padrão de cada base (A, T, C e G) dos fragmentos originais do genoma. Com esses valores calculados, fragmentos candidatos à predição são recuperados de uma base com genomas de

diversas espécies. Um fragmento artificial é aceito se o seu somatório de bases for menor que média menos o desvio padrão ou maior que a média somando o valor desvio padrão dos fragmentos do genoma original.

### 3.1.3 Quantidade de fragmentos artificiais

O número de fragmentos artificiais aceitos dizem quantas vezes o processo de seleção deve ser feito, tendo os fragmentos artificiais selecionados, executa-se o algoritmo de clusterização *Mean Shift*. Usando parâmetros bandwidth decrescentes, é repetido o processo de agrupamento, até que todos os fragmentos artificiais sejam mantidos em agrupamentos isolados. Nesse ponto, temos a clusterização ótima, que nos dará informações sobre as IGs presentes no organismo. Os fragmentos artificiais são usados apenas para a definição do parâmetro ótimo e não são analisados para identificação de IGs.

São considerados IGs fragmentos que são mantidos em clusters separados daqueles que contêm os fragmentos nativos do genoma. Esses clusters separados são aqueles que contêm no máximo 200 kb em bases (definido a partir do tamanho da janela, caso seja 50kb são clusters com no máximo 4 elementos).

## 3.2 Coleta de Dados

Todas as *Mycoplasma* sequenciadas usadas no trabalho foram adquiridas no Centro Nacional de Informações sobre Biotecnologia ( do inglês, National Center for Biotetchnology Information - NCBI, que possibilita recursos que organizam informações sobre genomas, incluindo sequências, mapas, cromossomos e montagens..

**Tabela 1.** *Mycoplasmas*/código de acesso no NCBI

<i>Mycoplasma</i>	Código de Acesso (ID)
<i>Mycoplasma agalactiae</i>	1028
<i>Mycoplasma arthritis</i>	1021
<i>Mycoplasma bovis</i>	1150

---

<i>Mycoplasma bovoculi</i>	11292
<i>Mycoplasma californicum</i>	32001
<i>Mycoplasma canadense</i>	35947
<i>Mycoplasma capricolum</i>	521
<i>Mycoplasma conjunctivae</i>	2176
<i>Mycoplasma crocodyli</i>	1732
<i>Mycoplasma cynos</i>	15822
<i>Mycoplasma dispar</i>	36400
<i>Mycoplasma fermentans</i>	1234
<i>Mycoplasma flocculare</i>	11293
<i>Mycoplasma gallisepticum</i>	1113
<i>Mycoplasma genitalium</i>	474
<i>Mycoplasma haemocanis</i>	11914
<i>Mycoplasma haemofelis</i>	1663
<i>Mycoplasma hominis</i>	3075

---



---

<i>Mycoplasma hyopneumoniae</i>	190
<i>Mycoplasma hyorhinis</i>	1922
<i>Mycoplasma leachii</i>	1696
<i>Mycoplasma meleagridis</i>	17586
<i>Mycoplasma mobile</i>	1651
<i>Mycoplasma mycoides</i>	720
<i>Mycoplasma ovis</i>	23941
<i>Mycoplasma parvum</i>	22974
<i>Mycoplasma penetrans</i>	1037
<i>Mycoplasma pneumoniae</i>	1028
<i>Mycoplasma pulmonis</i>	1029
<i>Mycoplama putrefaciens</i>	2769
<i>Mycoplasma suis</i>	2628
<i>Mycoplasma synoviae</i>	1649
<i>Mycoplasma wenyonii</i>	11761

---

---

*Mycoplasma yeatsii*

10726

---

Fonte: NCBI - Centro Nacional de Informações sobre Biotecnologia  
Elaborado pelo autor

O primeiro passo foi selecionar as espécies de mycoplasma que já tiveram seus genomas sequenciados, para coletar a informação no formato FASTA que é o formato padrão usada na área de bioinformática para representar uma sequência de DNA.

**Figura 4.** Coleta do sequenciamento da *M. Californicum* em arquivo FASTA

**Mycoplasma californicum**  
 Representative genome: **Mycoplasma californicum**  
 Download sequences in FASTA format for genome, protein  
 Download genome annotation in GFF, GenBank or tabular format  
 BLAST against Mycoplasma californicum genome, protein

Fonte: NCBI - Centro Nacional de Informações sobre Biotecnologia

### 3.3 Entrada dos Dados

O programa de predição de IGs com *Mean Shift* recebe como entrada o genoma no formato FASTA. A extensão do arquivo utilizada é a *fna*, que contém o DNA completo. A Figura 5 mostra um segmento de um genoma de *M.californicum*, de tamanho 0.95 mb, representado neste formato.

**Figura 5.** Sequenciamento da *M. Californicum* em arquivo FASTA

#### **Mycoplasma californicum strain ST-6, complete genome**

NCBI Reference Sequence: NZ\_CP007521.1

[GenBank](#) [Graphics](#)

```
>gi|752677826|ref|NZ_CP007521.1| Mycoplasma californicum strain ST-6, complete genome
GTGGGGATGCGAAAATCAGAAAATGAAAAAATAATATTTTATCGCTTAATGAATCTTTTATAAGTCAA
TGAGTGTGAAATATCAGATAAAATGATGATGAAAACCTTTGTTCTTCCCTTAAAAATAGATAAAGTAGA
AAATAATGTTGTTTTTTTAGTTCTTCTTAAGCAAGATATCTTAAATGTTCTTAGTAGAACTCCTATC
CACGACAAAATCTTAAAGTTTTGGCTGATATCGTAGGTCAAAATGTTTCTTATAATAAATTGATAAAA
CACTTAAAAAAATCCCAGAAAAGATTAGAAAATAGAAAAACAACAAGAATTAAATAGAAAAAGTATATC
TAAAAATTTACGTAAAAAGACTAATGATTTTTCAAGAGAATATACTTTCGAAACTTATGTTGAGGGTAAA
TTTAACTCGGAAGCTTTACGAATGGCTCGAGCAGTCTTGAAGGAGAACGCGAATATAATCCTATATTTA
TGACTGCTAATTCAGGGCTGGGAAAACACATCTATTACACGCTTTAGGAAATGAATTACAAAAACAAA
TAGTGATGTTATCTATATAAATCCACCAGATTCAGTCTGAAATATCTATTCTTTTCAAGAAAAATAAC
CCTAAAAAATAAGCAAAGAGTTGATGAATTATTACAGGCCGATGTTTAAATGTTGATGATTTTCAAT
```

Fonte: NCBI - Centro Nacional de Informações sobre Biotecnologia

A primeira linha do arquivo inicia com o símbolo '>', seguido do nome e do identificador da sequência e, opcionalmente, por comentários. A partir da segunda linha, tem-se a sequência do genoma, geralmente com 70 caracteres por linha. Na representação do genoma, cada letra representa uma base nucleica. A letra 'A' representa a base Adenina, 'T' representa Timina, 'C', Citosina e 'G', Guanina, caso apareça a letra 'N', é a representação qualquer uma das bases.

### 3.4 Saída dos Dados

Após a conversão da representação do genoma, o agrupamento será feito pelo algoritmo *Mean Sfhit*, saída do algoritmo nos define as potencias IGs e as suas posições. A Figura 6 mostra a saída do algoritmo e a definição de uma provável ilha genômica e a sua posição.

**Figura 6.** Saída do Algoritmo *Mean Sfhit*

```
0,00 - 0,05Mb. Ponto: (17616.0, 17328.0, 7007.0, 8049.0).Cluster: 1
0,05 - 0,10Mb. Ponto: (18000.0, 17012.0, 6978.0, 8010.0).Cluster: 1
0,10 - 0,15Mb. Ponto: (17813.0, 16987.0, 7106.0, 8094.0).Cluster: 1
0,15 - 0,20Mb. Ponto: (17401.0, 17036.0, 7505.0, 8058.0).Cluster: 1
0,20 - 0,25Mb. Ponto: (16566.0, 17830.0, 7356.0, 8248.0).Cluster: 1
0,25 - 0,30Mb. Ponto: (17247.0, 17411.0, 7204.0, 8138.0).Cluster: 1
0,30 - 0,35Mb. Ponto: (19723.0, 15384.0, 7181.0, 7712.0).Cluster: 2
0,35 - 0,40Mb. Ponto: (16399.0, 18210.0, 7697.0, 7694.0).Cluster: 1
0,40 - 0,45Mb. Ponto: (16767.0, 17701.0, 7890.0, 7642.0).Cluster: 1
0,45 - 0,50Mb. Ponto: (18903.0, 16538.0, 7316.0, 7243.0).Cluster: 1
0,50 - 0,55Mb. Ponto: (17598.0, 16361.0, 8291.0, 7750.0).Cluster: 1
0,55 - 0,60Mb. Ponto: (14896.0, 18356.0, 8915.0, 7833.0).Cluster: 3
0,60 - 0,65Mb. Ponto: (17080.0, 17495.0, 8048.0, 7377.0).Cluster: 1
0,65 - 0,70Mb. Ponto: (17371.0, 17149.0, 8200.0, 7280.0).Cluster: 1
0,70 - 0,75Mb. Ponto: (17964.0, 16873.0, 8001.0, 7162.0).Cluster: 1
0,75 - 0,80Mb. Ponto: (13365.0, 12141.0, 11111.0, 13383.0).Cluster: 4
0,80 - 0,85Mb. Ponto: (15071.0, 15291.0, 9008.0, 10627.0).Cluster: 5
0,85 - 0,90Mb. Ponto: (12532.0, 12779.0, 12078.0, 12611.0).Cluster: 6
0,90 - 0,95Mb. Ponto: (10115.0, 9799.0, 15135.0, 14951.0).Cluster: 7
```

Provavel Ilha Genomica: 0,300 - 0,350Mb.

Provavel Ilha Genomica: 0,550 - 0,600Mb.

Numero de Clusters: 7

Bandwidth: 1600.0

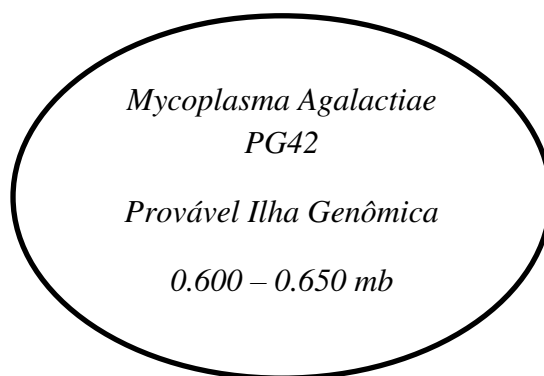


Fonte: Elaborado pelo autor

### 3.5 Coleta de Proteínas

Após a identificação das espécies de *Mycoplasmas* que apontam uma provável ilha genômica, analisamos todas as proteínas que integram a posição que destaca a IGS, pelo NCBI, que possui a contagem e o detalhe de proteínas de cada organismo sequenciado.

**Figura 7.** *Mycoplasma* e a posição da provável ilha genômica



Fonte: Elaborado pelo autor

A posição do arquivo de saída é dada em Mb, que corresponde a  $10^6 = 1000000$ , então, se a posição 0,60 corresponde a 600000 e 0,65 a 650000, então coletamos todas as proteínas que estiverem nesse intervalo. A Figura 8 mostra as posições da *Mycoplasma* e os nomes das proteínas em cada posição.

**Figura 8.** Nome das proteínas e as suas posições

POSIÇÕES										PROTEÍNAS
Name	Accession	Start	Stop	Strand	GeneID	Locus	Locus tag	Protein product	Length	Protein name
chr	<a href="#">NC_009497.1</a>	541306	542442	+	-	-	MAG_RS02355	<a href="#">WP_011949631.1</a>	378	hypothetical protein
chr	<a href="#">NC_009497.1</a>	542432	543322	+	-	-	MAG_RS02360	<a href="#">WP_011949632.1</a>	296	hypothetical protein
chr	<a href="#">NC_009497.1</a>	543309	544196	+	-	-	MAG_RS02365	<a href="#">WP_041308754.1</a>	295	hypothetical protein
chr	<a href="#">NC_009497.1</a>	544222	545793	-	-	-	MAG_RS02370	<a href="#">WP_011949634.1</a>	523	phosphomannomutase
chr	<a href="#">NC_009497.1</a>	545838	546797	+	-	-	MAG_RS02375	<a href="#">WP_011949635.1</a>	319	membrane protein
chr	<a href="#">NC_009497.1</a>	546817	547236	+	-	-	MAG_RS02380	<a href="#">WP_011949636.1</a>	139	antitermination factor NusB
chr	<a href="#">NC_009497.1</a>	547337	547939	-	-	-	MAG_RS02385	<a href="#">WP_011949637.1</a>	200	ACP phosphodiesterase
chr	<a href="#">NC_009497.1</a>	548206	548478	+	-	-	MAG_RS02390	<a href="#">WP_011949638.1</a>	90	phosphocarrier protein HPr
chr	<a href="#">NC_009497.1</a>	548505	549440	+	-	-	MAG_RS02395	<a href="#">WP_011949639.1</a>	311	peptide ABC transporter permease
chr	<a href="#">NC_009497.1</a>	549441	550559	+	-	-	MAG_RS02400	<a href="#">WP_011949640.1</a>	372	peptide ABC transporter permease

Fonte: NCBI - Centro Nacional de Informações sobre Biotecnologia

Ao coletarmos todas as proteínas que foram apontadas nas posições que possuem uma provável ilha genômica, agrupamos as proteínas a partir de uma categorização de genes, por meio da Classificação Filogenética de Proteínas Codificadas em Genomas Completos (Phylogenetic classification of proteins encoded in complete genomes – COGs) (TATUSOV, 2000).

Cada grupo contém proteínas ou grupos de parálogos, a partir de pelo menos três linhagens. As proteínas são classificadas pela sua função, distribuídas nessas categorias:

- Citoesqueleto
- Controle do ciclo celular e mitose
- Estrutura secundária
- Metabolismo
- Modificação Pós-Transducional
- Motilidade Celular
- Produção e conversão de energia
- Proteínas da Membrana
- Proteínas Hipotéticas
- Replicação e Reparação
- Tráfego intracelular e secreção
- Transcrição
- Transdução do sinal

Após toda a coleta de dados e categorização das proteínas, foram analisadas mediante a literatura as proteínas que desenvolvem resistência à bactéria, se está relacionada com virulência ou adaptação de vida a bactéria, gerando vantagem seletiva.

## 4 RESULTADOS E DISCUSSÃO

Utilizamos 34 genomas de *Mycoplasmas* que já foram sequenciadas, 16 delas apontaram uma provável ilha genômica (Tabela 2):

<i>Mycoplasma</i>	Ilha Genômica
<i>Mycoplasma agalactiae</i> PG2, <i>Mycoplasma arthritidis</i> , <i>Mycoplasma bovis</i> PG45, <i>Mycoplasma californicum</i> , <i>Mycoplasma canadense</i> , <i>Mycoplasma crocodyli</i> MP145, <i>Mycoplasma fermentans</i> JER, <i>Mycoplasma gallisepticum</i> CA06, <i>Mycoplasma haemocanis str. Illinois</i> , <i>Mycoplasma haemofelis</i> Ohio2, <i>Mycoplasma parvum str. Indiana</i> , <i>Mycoplasma pneumoniae</i> M129-B7, <i>Mycoplasma putrefaciens</i> Mput9231, <i>Mycoplasma suis str. Illinois</i> , <i>Mycoplasma synoviae</i> 53, <i>Mycoplasma yeatsii</i> HM274B	SIM
<i>Mycoplasma bovoculi</i> , <i>Mycoplasma capricolum</i> ATCC 27343, <i>Mycoplasma conjunctivae</i> , <i>Mycoplasma cynos</i> C142, <i>Mycoplasma dispar</i> , <i>Mycoplasma flocculare</i> ATCC27399, <i>Mycoplasma genitalium</i> G37, <i>Mycoplasma hominis</i> , <i>Mycoplasma hyopneumoniae</i> J, <i>Mycoplasma hyorhinae</i> SK76, <i>Mycoplasma leachii</i> PG50, <i>Mycoplasma meleagridis</i> strain, <i>Mycoplasma mobile</i> 163K, <i>Mycoplasma mycoides</i> SC str. PG1, <i>Mycoplasma ovis str. Michigan</i> , <i>Mycoplasma penetrans</i> Hf-2 Dna, <i>Mycoplasma pulmonis</i> UAB CTIP, <i>Mycoplasma wenyonii str. Massachusetts</i>	NÃO

Com o resultado de todas as *Mycoplasmas*, que possuem uma provável IGS. Foi feita uma categorização por meio de uma Classificação Filogenética de Proteínas Codificadas em Genomas (COGs). A tabela a seguir (Tabela 3), mostra as *Mycoplasmas* que possuem IGS, qual sua posição, outras IGS citadas na literatura e as proteínas presentes nessas regiões.

**Tabela 3.** Posições e Caracterização das Ilhas genômicas

<b>Genoma</b>	<b>Deteccão - Ilha Genômica (Mb) (Este Trabalho)</b>	<b>Deteccão – Ilha Genômica (Outros trabalhos)</b>	<b>Características</b>
<i>M.agalactiae</i> (PG2)	0.600 – 0.650	0.640 – 0.657 (SIRAND-PUGNET et al., 2007).	Proteínas da membrana, metabolismo, transdução do sinal, proteínas hipotéticas.
<i>M.arthritis</i> (158L3-1)	0.000 – 0.050	-	Controle do ciclo celular, metabolismo, proteínas da membrana, proteínas hipotéticas, transdução do sinal.
<i>M.bovis</i> (PG45)	0.700 – 0.750	0.700 – 0.740 (LI et al., 2011).	Metabolismo, proteínas da membrana, produção e conversão de energia, proteínas hipotéticas, motilidade celular.
<i>M.californicum</i> (ST-6)	0.300 – 0.350	-	Metabolismo, proteínas da membrana, produção e conversão de energia, replicação e reparação, proteínas hipotéticas.
	0.550 – 0.600	-	Proteínas hipotéticas, proteínas da membrana, metabolismo.
<i>M.canadense</i> (HAZ 360_1)	0.350 – 0.400	-	Proteínas da membrana, metabolismo, transdução do sinal, proteínas hipotéticas.
	0.550 – 0.650	-	Produção e conversão de energia, controle do ciclo celular, transcrição, replicação e reparação, modificação pós-transducional, proteínas hipotéticas, transdução do sinal, proteínas da

			membrana, metabolismo.
<i>M.crocodyli</i> (MP145)	0.400 – 0.450	-	Controle do ciclo celular, metabolismo, proteínas da membrana, replicação e reparação, motilidade celular, proteínas hipotéticas, transdução do sinal, estrutura secundária.
	0.800 - 0.900	-	Proteínas da membrana, produção e conversão de energia, controle do ciclo celular, metabolismo, replicação e reparação, motilidade celular, proteínas hipotéticas, transdução do sinal, tráfego intracelular e secreção.
<i>M.fermentans</i> (JER)	0.400 – 0.450	-	Produção e conversão de energia, controle do ciclo celular, replicação e reparação, parede de celular, motilidade celular, modificação pós-transducional, proteínas da membrana, metabolismo, proteínas hipotéticas, transdução do sinal, tráfego intracelular e secreção.
<i>M.gallisepticum</i> (CA06)	0.400 – 0.450	0.400 – 0.410 (SZCZEPANEK et al., 2010).	Transdução do sinal, metabolismo, replicação e reparação, proteínas hipotéticas, proteínas da membrana.
<i>M.haemocanis</i> (srt.Illinois)	0.200 – 0.250	-	Proteínas hipotéticas, proteínas da membrana, metabolismo, motilidade celular.
	0.400 – 0.650	-	Proteínas hipotéticas, transdução do sinal, metabolismo, proteínas da membrana, controle celular, tráfego intracelular e secreção.
	0.750 – 0.800	-	Proteínas hipotéticas, proteínas da membrana, replicação e reparação.



<i>M. haemofelis</i> ( <i>Ohio2</i> )	0.650 – 0.700	-	Proteínas hipotéticas.
	0.950 – 1.000	-	Proteínas hipotéticas, proteínas da membrana.
<i>M. parvum</i> ( <i>str. Indiana</i> )	0.000 – 0.050	-	Controle do ciclo celular, metabolismo, proteínas da membrana, transdução do sinal.
	0.100 – 0.200	-	Transdução do sinal, proteínas da membrana, metabolismo, proteínas hipotéticas.
	0.300 – 0.350	-	Controle do ciclo celular, metabolismo, proteínas da membrana, replicação e reparação, motilidade celular.
<i>M. pneumoniae</i> ( <i>M129-B7</i> )	0.550 – 0.600	0.600 – 0.650 (XIAO et al., 2015).	Proteínas da membrana, metabolismo, replicação e reparação, produção e conversão de energia, transdução do sinal, proteínas hipotéticas.
<i>M. putrefaciens</i> ( <i>Mput9231</i> )	0.050 – 0.100	-	Produção e conversão de energia, metabolismo, proteínas da membrana, motilidade celular, proteínas hipotéticas, transdução do sinal.
	0.300 – 0.350	-	Produção e conversão de energia, controle do ciclo celular, metabolismo, proteínas da membrana, transcrição, replicação e reparação, motilidade celular, modificação pós-transducional, proteínas hipotéticas, transdução do sinal, tráfego intracelular e secreção, citoesqueleto.
	0.550 – 0.650	-	Produção e conversão de energia, controle celular, metabolismo, proteínas da membrana, replicação e reparação,

	0.800 – 0.850	-	<p>motilidade celular, modificação pós-transducional, estrutura secundária, transdução do sinal, proteínas hipotéticas.</p> <p>Produção e conversão de energia, controle do ciclo celular, metabolismo, proteínas da membrana, replicação e reparação, modificação pós-transducional, proteínas hipotéticas, transdução do sinal, tráfego intracelular e secreção.</p>
<i>M.suis</i> (Illinois)	0.100 – 0.150	0.030 – 0.100 (OEHLERKING et al., 2011).	Proteínas hipotéticas
	0.500 – 0.550	0.400 - 0.500 (OEHLERKING et al., 2011).	Proteínas hipotéticas, proteínas da membrana e metabolismo.
<i>M.synoviae</i> (53)	0.200 – 0.300	0.220 – 0.225 (VASCONCELOS et al., 2005).	Motilidade celular, proteínas da membrana.
<i>M.yeatsii</i> (GM274B)	0.00 – 0.050	Sem posição especifica (Breton et al., 2012).	Proteínas da membrana, produção e conversão de energia, controle do ciclo celular, metabolismo, motilidade celular, modificação pós-transducional, proteínas hipotéticas, transdução do sinal, tráfego intracelular e secreção, citoesqueleto.
	0.400 – 0.450	Sem posição especifica (Breton et al., 2012).	Controle do ciclo celular, metabolismo, proteínas da membrana, replicação e reparação, proteínas hipotéticas, transdução do sinal, tráfego intracelular e secreção, transcrição, motilidade celular, citoesqueleto.

Fonte: NCBI – Centro Nacional de Informações de Biotecnologia.  
Elaborado pelo autor

Analisando a caracterização das regiões de IGs, pode-se notar que algumas categorias aparecem mais que outras, desse modo, iremos analisar as categorias que mais se destacaram e as proteínas que fazem parte, procurando por fatores de virulência, patogenicidade ou resistência, que podem indicar uma vantagem seletiva à *Mycoplasma*. Na tabela 4 a seguir, mostramos a porcentagem de cada categoria na caracterização:

**Tabela 4.** Porcentagem das categorias do COGs

<b>Categoria</b>	<b>Porcentagem (%)</b>
Proteínas da Membrana	16
Proteínas Hipotéticas	16
Metabolismo	14
Transdução do Sinal	11
Replicação e Reparação	8
Controle do Ciclo Celular	8
Motilidade Celular	7
Produção e Conversão de Energia	7
Tráfego Intracelular e Secreção	4
Modificação Pós-Transducional	4
Citoesqueleto	2
Transcrição	2
Estrutura Secundária	1

Fonte: Elaborado Pelo Autor

### Proteínas da Membrana

Essas proteínas servem como mediadores importantes da comunicação entre o citoplasma da célula e o ambiente extracelular. Elas também determinam se os metabolitos secundários, tais como antibióticos serão sintetizados, exportados ou importados (CHATER et al., 2010). Eles representam cerca de 20-30% do genoma de um organismo (MILLER et al., 2005).

Neste trabalho, verificou-se uma constância das proteínas membranares, nas regiões que indicam IGs, em todas as *Mycoplasmas*, essas proteínas estiveram presente, junto com as proteínas hipotéticas. Uma das proteínas que fazem parte dessa categoria, que

mais se destacaram, foi os transportadores ABC (do inglês, ATP-Binding Cassette Transporters), que são membros de uma superfamília de proteínas transmembranares.

Em procariotos funcionam como importadores, exportadores, transportam uma grande variedade de substratos através das membranas extra e intracelular, além de possuírem resistência a múltiplas drogas (ERKENS et al., 2012).

Os transportadores ABC, são essenciais para viabilidade celular, virulência e patogenicidade. O sistema de absorção de ferro, que é uma particularidade dessa proteína, chama atenção, já que a absorção de ferro é um grande obstáculo para o crescimento e a sobrevivência de bactérias patogênicas (RAKESHKUMAR et al., 2014).

Outro grupo bastante presente, foi as lipoproteínas, cerca de 30 lipoproteínas entre as regiões, foi às lipoproteínas que são um conjunto de proteínas, que desempenham um papel na interação entre a bactéria e a célula hospedeira, não só na adesão e co-agregação, mas com a estimulação e liberação de citocinas pró-inflamatórias e na translocação de fatores de virulência em células hospedeiras, realçando o seu potencial virulento com uma grande gama de mecanismos (KOVACS-SIMON et al., 2011; MUHLRADT et al., 1998).

## Transdução do Sinal

É o processo através do qual uma célula converte um tipo de sinal ou estímulo em outro. A maioria dos processos de transdução de sinal envolvem sequências ordenadas de reações bioquímicas dentro da célula, que são levadas a cabo por proteínas ativadas por mensageiros secundários, resultando numa via de transdução de sinal. As proteínas são componentes essenciais, como canais iônicos, reconhecimento, além de que são essências para o funcionamento e especificidade das vias de sinalização.

A caracterização de proteínas que fazem parte da transdução de sinal é de grande importância para entendermos melhor os seus mecanismos de respostas aos diferentes estímulos e da hierarquia intracelular das diferentes vias de sinalização (GALPERIN, 2004).

A característica mais importante das proteínas de transdução de sinal bacteriana é sua organização modular. A maioria dos domínios dessas proteínas pode ser facilmente reconhecida e, às vezes, associada a funções bioquímicas particulares (SCHELER, 2013).

Algumas proteínas sinalizadoras que estavam nas regiões com provável IGs, por exemplo, GTP-binding, que atua como interruptor molecular em vias de sinalização, regulando as funções de outras proteínas. (KRAUSS, 2008).

## Proteínas Hipotéticas

Apesar de quantidades cada vez maiores de dados biológicos, tais como os sequenciamentos e dados genômicos a partir de experiências de elevado rendimento, se tem uma deficiência funcional para muitas proteínas recentemente sequenciadas. (ENAULT et al., 2005). Uma grande parte dos genes dentro de um genoma é identificada como "desconhecida", ou "hipotética", limitando a nossa compreensão da virulência e patogenicidade destes organismos.

Cada proteína é um produto de gene que interage com o meio ambiente celular de alguma forma, com a função de promover o crescimento da célula. Isto sugere que este grande número de proteínas descaracterizadas nas *Mycoplasmas* podem desempenhar um papel crucial na sua sobrevivência e viabilidade ou podem contribuir para a sua aptidão em diferentes ambientes (PERSERNER, 2003).

Neste trabalho, se teve um grande número de proteínas hipotéticas, regiões inteiras preenchidas por proteínas hipotéticas. Muitas destas proteínas hipotéticas podem desempenhar um papel importante na adesão do agente patogênico, no sistema hospedeiro e na modulação imunitária, e também no seu estilo de vida intracelular.

Assim, uma das principais tarefas na era pós-genômica é a anotação do genoma, a fim de capitalizar o conhecimento adquirido através dos esforços na área de sequenciamento (MAZANDU-MULDER, 2012).

.

## Metabolismo Energético

Uma característica indispensável de qualquer organismo vivo é o seu metabolismo. O metabolismo é uma rede complexa de reações químicas catalisadas por enzimas, que são codificadas nos genes. Ele usa fontes de energia - nutrientes - para sintetizar moléculas que o organismo necessita, incluindo precursores da biomassa e vários produtos secundários, tais como moléculas de defesa e comunicação. (KUWAHARA et al., 2007).

A categorização das IGs neste trabalho apresentou um grande número de proteínas que atuam nessa rede, e que metabolizam amino ácidos, nucleotídeos, íons inorgânicos,

carboidratos, coenzimas e lipídios. As proteínas que metabolizam carboidrato foram as que mais se destacaram.

Todas as espécies de *Mycoplasma* exibem genes relacionados com o metabolismo energético, já que existem diferentes processos que elas utilizam para gerar energia na forma de ATP, permitindo que elas persistam em ambientes com condições extremas, conferindo uma vantagem seletiva para a bactéria. (KOUMANDOU et al., 2014).

A ATP sintase, um complexo de proteínas que fornecem energia através da síntese de ATP foi relatada 20 vezes nas regiões que apontam IGs. Constantemente caracterizada pelo metabolismo energético, é primordial para os processos bioenergéticos. A maioria das vias bioenergéticas usa uma cadeia de transporte de elétrons para gerar um gradiente de prótons através da membrana, e a energia liberada pelo fluxo de elétron para compensar este gradiente, é utilizada pela ATP sintase para gerar ATP.

Existe uma tese de que a THG, aconteceu para uma forma variante a ATP sintase relacionada a subunidade C, que é utilizada para translocação de Na<sup>+</sup>, supondo que essa variante confira uma vantagem seletiva para bactéria em ambientes altamente salinos. (DIBROVA et al., 2010).

O metabolismo de nucleotídeos também esteve presente nas regiões de IGs, ele é central para todos os sistemas biológicos, devido seu papel essencial na transferência de informação genética e energia, esse metabolismo têm sido descrito como um dos mais antigos na evolução. (CAETANO-ANOLLES et al., 2009). Muitos compostos intermediários associados a este módulo metabólico foram intimamente associados com a química pré-biótica e a origem da vida. (BECERRA, 1998).

## 5 CONSIDERAÇÕES FINAIS

A eficácia do método desenvolvido foi comprovada através dos testes de predição de IGs nos genomas das Mycoplasmas, identificamos 16 prováveis IGs em 34 Mycoplasmas, 7 delas descritas na literatura e 9 não documentadas, ao todo encontramos 29 posições que indicam uma provável IGS. Os resultados obtidos mostraram que o método se configura como uma abordagem possível para a predição de IGs em bactérias. A caracterização nos permitiu observar como as IGs se manifestam nos organismos e atribuem características vantajosas.

Algumas IGs detectadas não foram previamente documentadas, porém a investigação dos genes presentes nestas ilhas sugere que estas regiões contêm genes adquiridos horizontalmente. Investigação futura é necessária a fim de se comprovar biologicamente esta hipótese.

Como contribuição, a simplicidade do método pode ser considerada uma vantagem em relação a outros métodos existentes. O método possui uma abordagem simples para predição de IGs.

## REFERÊNCIAS

- ALBERTS, B.; et al. Molecular biology of the cell. **Garland Science**, Taylor and Francis Group. 2015.
- ANDERSSON, J. Lateral gene transfer in eukaryotes. **Cell. Mol. Life Sci**, v.62, p.1182–1197. 2005.
- ANDERSSON, J.; et al. Phylogenetic analysis of Diplomonad genes reveals frequent lateral gene transfer affecting eukaryotes. **Curr. Biol**, v.13, 94–104. 2003.
- ANDREA, P.; et al. Complete Genome Sequence of *Mycoplasma wenyonii* Strain Massachusetts. **J Bacteriol**, v.194, p.19. 2012.
- BALL, HJ. Experimental production of bovine and ovine mastitis with a *Mycoplasma canadense* isolate. **Vet. Rec**, v.118, p.72–73. 1986.
- BARILE, M.; et al. Isolation and characterization of *Mycoplasma conjunctivae* sp. n. from sheep and goats with keratoconjunctivitis. **Infect. Immun**, v.5, p.70-76. 1972.
- BASEGGIO, N.; et al. Size and genomic location of the pMGA multigene family of *Mycoplasma gallisepticum*. **Microbiology**, v.142, p.1429–1435. 1996.
- BASEMAN, J.; TULLY, J.; *Mycoplasmas*: Sophisticated, reemerging, and burdened by their notoriety. **Emerging Infectious Diseases**, v.3, p.31-32. 1997.
- BASEMAN, JB.; REDDY, SP.; DALLO, SF. Interplay between *mycoplasma* surface proteins, airway cells, and the protean manifestations of *mycoplasma*-mediated human infections. **Respir Crit Care Med**, v.154, p.137-144. 1996.
- NAKAMURA, Y.; et al. Biased biological functions of horizontally transferred genes in prokaryotic genomes. **Nat Genet**, v.36, n.7, p.760-6. 2004.
- BLUM, G.; et al. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. **Infect. Immun**, v.62, p.606–614. 1994.



BRETON, M.; et al. Distribution and diversity of mycoplasma plasmids: lessons from cryptic genetic elements. **BMC Microbiology**, v.12, p.257. 2012.

BRIGULLA, M.; WACKERNAGEL, W. Molecular aspects of gene transfer and foreign DNA acquisition in prokaryotes with regard to safety issues. **Applied Microbiology Biotechnology**, v.86, p.1027-41. 2010.

BRUGGER, K.; et al. Mobile elements in archaeal genomes. **FEMS Microbiol, Lett.** v.206, p.131–41. 2002.

BURRUS, V.; et al. **Mol Microbiol**, v.46, n.3, p.601-10. 2002.

CAETANO-ANOLLES, G.; et al. The origin and evolution of modern metabolism. **Int J Biochem Cell Biol**, v.41, n.2, p.285–297. 2009.

CALCUTT, J.; FOECKING, F. Genome Sequence of *Mycoplasma putrefaciens* Type Strain KS1. **J Bacteriol**, v.193, p.21. 2011.

CASWELL, JL.; ARCHAMBAULT, M.; *Mycoplasma bovis* pneumonia in cattle. **Animal Health Research Reviews**, v.8, p.161–86. 2008.

CHALKER, VJ.; et al. *Mycoplasmas* associated with canine infectious respiratory disease. **Microbiology Reading**, v.150, p.3491–97. 2004.

CHAMBAUD, I. The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. **Nucleic Acids Res**, v.29, p.2145-53. 2001.

CHAMBAUD, I.; WROBLEWSKI, H.; BLANCHARD, A. Interactions between mycoplasma lipoproteins and the host immune system. **Trends Microbiol.**, v.7, p.493–99. 1999.

CHATER KF.; et al. The complex extracellular biology of *Streptomyces*. **FEMS Microbiol Rev**, v.13, n.2, p.171–198. 2010.

CHE, DONGSHENG.; et al. EGID: an ensemble algorithm for improved genomic island detection in genomic sequences. **Bioinformatics**, v.7, p.311. 2011.

CHIAPELLO, H.; BOURGAI, I.; SOURIVONG, F. Systematic determination of the mosaic structure of bacterial genomes: species backbone versus strain-specific loops. **BMC Bioinformatics**, v.6, p.171. 2005.

CHOI, I.; KIM, S. Global extent of horizontal gene transfer. **Proc. Natl Acad. Sci**, v.104, p.4489–94. 2007.

CHRISTIE, P.; et al. Biogenesis, architecture, and function of bacterial type IV secretion systems. **Annu. Rev. Microbiol**, v.59, p.451–85. 2005.

CHVAL, S.; et al. Simultaneous canine distemper virus, canine adenovirus type 2, and *Mycoplasma cynos* infection in a dog with pneumonia. **Vet. Pathol**, v.44 ,p.508–512. 2007.

CITTI C. *Mycoplasma* genetic variation and its implication for pathogenesis. **Wien. Klin. Wochensh**, v.9, p.562–68. 1997.

CORRALES, JC.; et al. Contagious agalactia in small ruminants. **Small Ruminant**, v.68, p.154-66. 2007.

DAGAN, T.; ARTZY-RANDRUP, Y.; MARTIN, W. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. **Proc. Natl. Acad. Sci**, v.105, n.10, p.39–44. 2008.

DAMASSA, AJ.; WAKENELL, PS.; BROOKS, DL. *Mycoplasmas* of goats and sheep. **J Vet Diagn Invest**, v.4, p.101–113. 1992.

DE LA CRUZ, F.; DAVIES, J. Horizontal gene transfer and the origin of species: lessons from bacteria. **Trends in microbiology**, v. 8, n. 3, p. 128–133. 2000.

DIBROVA, DV.; GALPERIN, MY.; MULKIDJANIAN, AY. Characterization of the N-ATPase, a distinct, laterally transferred Na<sup>+</sup>-translocating form of the bacterial F-type membrane ATPase. **Bioinformatics**, v.26, n.12, p.1473-76. 2010.

DOBRINDT, U.; et al. Genomic islands in pathogenic and environmental microorganisms. **Nat. Rev. Microbiol**, v.2, n.5, p.414–24. 2004.

DOOLITTLE, W.; LOGSDON, J. Do Archaea have a mixed heritage? **Curr. Biol**, v.8, p.209–11. 1998.

ELKIND, E.; et al. *Mycoplasma hyorhinis* upregulates calpastatin and inhibits calpain-dependent proteolysis in SH-SY5Y neuroblastoma cells. **FEMS Microbiol**, v.304, p.62–68. 2010.

ENAULT, F.; SUHRE, K.; CLAVERIE, J.M. Phydbac “Gene Function Predictor”: a gene annotation tool based on genomic context analysis. **BMC Bioinformatics**, v.6, p.247. 2005.

ERKENS G.; et al. Energy coupling factor-type ABC transporters for vitamin uptake in prokaryotes. **Biochemistry**, v.51, p.4390–96. 2012.

FAGUY, D. M.; DOOLITTLE, W. F. Lessons from the *Aeropyrum pernix* genome. **Curr. Biol**, v.9, p.883–86. 1999.

FELDER, KM.; HOELZLE, K.; RITZMANN M. et al. Hemotrophic *Mycoplasmas* induce programmed cell death in red blood cells. **Cell Physiol Biochem**, v.27, p.557–564. 2011.

FINCH, L. *Mycoplasmas, Molecular Biology and Pathogenesis*. **American Society for Microbiology**, v.1, p. 523–45. 1992

FOLEY, S.; LYNNE, A. M. Food animal-associated *Salmonella* challenges, pathogenicity and antimicrobial resistance. **Journal of Animal Science**, v. 86, p. 173–87. 2008.

FRASER, CM.; GOCAYNE, JD.; WHITE, O. The minimal gene complement of *Mycoplasma genitalium*. **Science**, v.270, p.397–403. 1995.

FUJIHARA, Y. Prevalence of hemoplasma infection among cattle in the western part of Japan. **J. Vet. Med. Sci**, v.73, p.1653–55. 2011.

FUKUNAGA, K.; HOSTETLER, L. The estimation of the gradient of a density function, with applications in pattern recognition. **Information Theory, IEEE Transactions on**, v.21, p. 32-40. 1975.

GAILLARD, M.; et al. The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. **J. Bacteriol**, v.18, n.5, p.1999–2013. 2006.

GALPERIN, M.Y. Bacterial signal transduction network in a genomic perspective. **Environ Microbiol**, v.6, p.552-67. 2004.

GE, F.; WANG, L. S.; KIM, J. The cobweb of life revealed by genome scale estimates of horizontal gene transfer. **PLoS Biol**, v.3, p.316. 2005.

GENOVA, SG. Severe anemia associated with *Mycoplasma wenyonii* infection in a mature cow. **Vet. J**, v.52, p.1018–21. 2011.

GERHARD, K. Biochemistry of signal transduction and regulation. **Wiley-VCH**, v.235. 2008.

GOGARTEN, J.P.; DOOLITTLE, W.F.; LAWRENCE, J.G. Prokaryotic evolution in light of gene transfer. **Mol. Biol. Evol**, v.19 , p.2226–38. 2002.

GOLDENFELD, N.; WOESE, C. Biology's next revolution. **Nature**, v.445, p.369. 2007.

GOLDING, G. B.; GUPTA, R. S. Protein based phylogenies support a chimeric origin for eukaryotic genome. **Mol. Biol. Evol**, v.12, p.1–6. 1995.

GÓMEZ-MARTÍN, A.; et al. Contagious agalactia due to *Mycoplasma* spp. in small dairy ruminants: epidemiology and prospects for diagnosis and control. **Vet. J**, v.198, p.48–56. 2013.

GOPHNA, U.; CHARLEBOIS, R. L.; DOOLITTLE, W. F. Have archaeal genes contributed to bacterial virulence? **Trends Microbiol**, v.12, p.213–19. 2004.

GROEBEL, K.; HOELZLE, K.; WITTENBRINK, M.; et al. *Mycoplasma suis* invades porcine erythrocytes. **Infect. Immun**, v.77, p.576–84. 2009.

GROISMAN, EA. OCHMAN. Pathogenicity islands: bacterial evolution in quantum leaps. **Cell**, v.87, n.5, p.791-4. 1996.

GULJAMOW A.; et al. Horizontal gene transfer of two cytoskeletal elements from a Eukaryote to a Cyanobacterium. **Curr. Biol**, v.17, p.757–59. 2007.

GUELL, M.; et al. Transcriptome complexity in a genome-reduced bacterium. **Science**, v.509, p.68–71. 2009.

GUPTA R. S.; SINGH, B. Phylogenetic analysis of 70 kD heat-shock protein sequences suggest a chimeric origin for the eukaryotic cell nucleus. **Curr. Biol**, v.4, p.1104–14. 1994.

GUPTA R. S.; et al. Generating Symmetry in the Asymmetric ATP-binding Cassette (ABC) Transporter Pdr5 from *Saccharomyces cerevisiae*. **The Journal of Biological Chemistry**, v.289, n.22, p.15272-279. 2014.

HANG, Q.; F, YOUNG.; F, ROSS. Identification and characterization of a *Mycoplasma hyopneumoniae* adhesin. **Infect. Immun.** v.63, p.1013-19. 1995.

HE, DONGSHENG.; et al. An Accurate Genomic Island Prediction Method for Sequenced Bacterial and Archaeal Genomes. **Journal of Proteomics & Bioinformatics**, 2014.

HACKER, J. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extra intestinal Escherichia coli isolates. **Microb Pathog**, v.8, n.3, p.213-25. 1990.

HACKER, J.; et al. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. **Mol Microbiol**, v.23, n.6, p.1089-97.1997.

HENTSCHEL, U.; HACKER, J. Pathogenicity islands: the tip of the iceberg. **Microbes and infection**, v.3, n.7, p.545-48. 2001.

HEWICKER-TRAUTWEIN, M.; et al. Outbreak of pneumonia and arthritis in beef calves associated with *Mycoplasma bovis* and *Mycoplasma californicum*. **Vet. Rec**, v.151, p.699–703. 2002.

HILARIO, E.; GOGARTEN J. P. Horizontal gene transfer of ATPase genes—the tree of life becomes a net of life. **BioSystems**, v.31, p.111–19. 1993.

HSIAO, WILLIAM.; et al. IslandPath: aiding detection of genomic islands in prokaryotes, **Bioinformatics**, v.19, p.418-20. 2003.

HULME-MOIR, KL.; et al. Use of real-time polymerase chain reaction to monitor antibiotic therapy in a dog with naturally acquired *Mycoplasma haemocanis* infection. **J. Vet. Diagn. Invest**, v.22, p.582–87. 2010.

- JAFFE, J.; CHURCH, G. Proteogenomic mapping as a complementary method to perform genome annotation. **Proteomics**, v.4, p.59–77. 2004.
- JASPER, DE. *Mycoplasma californicum*, a new species from cows. **Int. J. Syst. Bacteriol.** v.31, p.339–45. 1981.
- JUHAS, M. Sequence and functional analyses of Haemophilus spp. genomic islands. **Genome Biol**, v.8, n.11, p.237. 2007.
- JUHAS, M.; et al. Genomic islands: tools of bacterial horizontal gene transfer and evolution. **FEMS microbiology reviews**, v. 33, n. 2, p.376-93. 2009.
- KANG, M. Virulence of recent isolates of *Mycoplasma synoviae* in turkeys. **Avian Dis.** v.46, p.102–10. 2002.
- KARLIN, S.; CAMPBELL, AM. Which bacterium is the ancestor of the animal mitochondrial genome? **Proc Natl Acad Sci**, v.91, p.12842-846. 1994.
- KARLIN, S. Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. **Trends in microbiology**, v.9, p.335-43. 2001.
- ILOFF, J. In Molecular biology and pathogenicity of *Mycoplasmas*. **Phylogeny and evolution**, v.1, p.31–44. 2002.
- INDIKOVA, I.; et al. Role of the GapA and CrmA cytoadhesins of *Mycoplasma gallisepticum* in promoting virulence and host colonization. **Infect Immun.** v.34, p.181-86. 2013.
- KELLY, B. G.; VESPERMANN, A.; BOLTON, D. J. The role of horizontal gene transfer in the evolution of selected foodborne bacterial pathogens. **Food and Chemical Toxicology**, v. 47, p. 951-68. 2009.
- KENT, BN.; FOECKING, MF.; CALCUTT, MJ. Development of a novel plasmid as a shuttle vector for heterologous gene expression in *Mycoplasma yeatsii*. **J Microbiol Methods**, v.91, p.121–27. 2012.
- KHIARI, A.; et al. Isolation of *Mycoplasma meleagridis* from chickens. **Avian Dis**, v.55, p.8–12. 2011.

KIRCHHOFF, H. *Mycoplasma crocodyli* sp. nov., a new species from crocodiles. **Int J Syst Bacteriol**, v.47, p.742-46. 1997.

KIRCHHOFF, H.; ROSENGARTEN, R. Isolation of a motile mycoplasma from fish. **J. Gen. Microbiol**, v.130, p.2439–445. 1984.

KOONIN, E. V.; MAKAROVA, A. S.; ARAVIND, L. Horizontal gene transfer in Prokaryotes. Quantification and classification. **Annu. Rev. Microbiol**, v.55, 709–42. 2001.

KOSKI L. B.; MORTON R. A.; GOLDING G. B; Codon bias and base composition are poor indicators of horizontal gene transfer. **Mol. Biol. Evol**, v.18, p.404–12. 2001.

KOUMANDOU, VL.; KOSSIDA, S. Evolution of the F<sub>0</sub>F<sub>1</sub> ATP Synthase Complex in Light of the Patchy Distribution of Different Bioenergetic Pathways across Prokaryotes. **PLoS Computational Biology**, v.10, n.9. 2014.

KOVACS-SIMON, A.; TITBALL, RW.; MICHELL, SL. Lipoproteins of Bacterial Pathogens . **Infection and Immunity**, v.79, n.2, p.548-61. 2011.

KUHNER, S.; et al. Proteome organization in a genome-reduced bacterium. **Science**, v.502, p.35–40. 2009.

KUMAR, A; SRIVASTAVA, C.; SINGH, V.; et al. Electrophoretic analysis of indian isolates of *Mycoplasma agalactiae* and *Mycoplasma bovis* by SDS-PAGE and Immunoblotting. **Vet Med Int**. 2014.

KUO, C.; OCHMAN, H. The fate of new bacterial genes. **FEMS Microb. Rev**, v.33, p.38–43. 2009.

KURLAND, C.; CANBACK, B.; BERG, O. Horizontal gene transfer: a critical view. **Proc. Natl. Acad. Sci**, v.100, p.9658–62. 2003.

KUWAHARA, H.; et al. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. **Current biology**, v.17, p.881–86. 2007.

LANGFORD, EV.; LEACH, RH. Characterization of a mycoplasma isolated from infectious bovine keratoconjunctivitis: *M. bovoculi* sp. **Can. J. Microbiol**, v.19, p.1435–44. 1973.

LANGILLE, MG.; HSIAO, WW.; BRINKMAN, FS. Evaluation of genomic island predictors using a comparative genomics approach. **BMC Bioinformatics**, v.9, p.329. 2008.

MILLER, JP. Large-scale identification of yeast integral membrane protein interactions. **Proc. Natl. Acad. Sci**, v.102, n.34, p.12123-8. 2005.

LAWRENCE, J. G.; OCHMAN, H. Reconciling the many faces of lateral gene transfer. **Trends Microbiol**, v.10, p.1–4. 2002.

LAWRENCE, G.; OCHMAN, H. Molecular archaeology of the *Escherichia coli* genome. **Proceedings of the National Academy of Sciences**, v.95, p.9413-17. 1998.

LEDERBERG, J.; TATUM E. L. Gene recombination in *Escherichia coli*. **Nature**, v.158, p.558. 1946.

LERAT, E.; et al. Evolutionary origins of genomic repertoires in Bacteria. **PLoS Biol**, v.3, p.130. 2005.

LEWIN, B. *Genes VIII*. New Jersey: **Pearson Prentice Hall**, 2004.

LI, Y; et al. The Complete Genome Sequence of *Mycoplasma bovis* Strain Hubei-1. **PLoS ONE**, v.6, p.6. 2011.

LO, S. *Mycoplasmas: Molecular Biology and Pathogenesis* *Mycoplasmas and AIDS*. **American Society for Microbiology**. v.2, p.525-45. 1992a.

LO, S.; et al. Virus-like infectious agent (VLIA) is a novel pathogenic mycoplasma: *Mycoplasma incognitus*. **Am. J. Trop. Med. Hyg**, v.41, p.586–600. 1989b.

MADIGAN, T. et al. *Brock Biology of microorganisms*. **Pearson**. 2015.

MAES, D. et al. Control of *Mycoplasma hyopneumoniae* infections in pigs. **Vet Microbiol**, v.126, p.297–309. 2008.

MANSO-SILVAN, L.; et al. Phylogeny of the *Mycoplasma mycoides* cluster based on analysis of five conserved protein-coding sequences and possible implications for the taxonomy of the group. **Int. J. Syst. Evol. Microbiol**, v.57, p.2247–58. 2007.



MANSO-SILVÁN, L.; et al. *Mycoplasma leachii* sp.nov. as a new species designation for *Mycoplasma* sp. bovine group 7 of Leach, and reclassification of *Mycoplasma mycoides* subsp. *Mycoides* LC as a serovar of *Mycoplasma mycoides* subsp. *capri*. **Int. J. Syst. Evol. Microbiol**, v.59, p.1353–58. 2009.

MARQUES, LM.; et al. Use of a Polymerase Chain Reaction for Detection of *Mycoplasma Dispar* in the Nasal Mucus of Calves. **Journal of Veterinary Diagnostic Investigation**, v.19, p.103-06. 2007.

MARTY, A. *Mycoplasma Infections*. Services University of the Health Sciences **Department of Pathology Syllabus**, v.4, p.91-4. 1993.

MASSARO, G.; et al. Interleukin-6 and *Mycoplasma hominis* as markers of preterm birth and related brain damage: our experience. **J. Matern Fetal Neonatal Med**. v.22, p.1063–67. 2009.

MATTHEW, J. et al. Identification of a gene in *Mycoplasma hominis* associated with preterm birth and microbial burden in intra-amniotic infection. **Am. J. Obstet. Gynecol**, v.212, p.799. 2015.

MAY, A. et al. Complete Genome Sequence of *Mycoplasma synoviae* Strain WVU 1853T. **Genome Announc**, v.3, p.563. 2015.

MAZANDU, GK.; MULDER, NJ. Function Prediction and Analysis of Mycobacterium tuberculosis Hypothetical Proteins. **International Journal of Molecular Sciences**, v.13 n.6, p.7283-7302. 2012.

HOCHHUT, B.; MARRERO, J.; WALDOR, MK. Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139. **J Bacteriol**, v.182, n.7, p.2043-7. 2000.

MUCH, P.; et al. *Mycoplasma gallisepticum*: Influence of cell invasiveness on the outcome of experimental infection in chickens. **FEMS Immunol Med Microbiol**, v.34, p.181–86. 2002.

MÜHLRADT, PF.; et al. Structure and specific activity of macrophage-stimulating lipopeptides from *Mycoplasma hyorhinis* . **Infect Immun**, v.66, n.10, p.4804–10. 1998.

NAÍLA C.; et al. Genome Sequence of *Mycoplasma parvum* (Formerly *Eperythrozoon parvum*), a Diminutive Hemoplasma of the Pig. **J. Bacteriol**, v.1, p.6. 2013.

NAÍLA, C.; et al. Complete genome sequence of *Mycoplasma haemocanis* strain illinois. **J. Bacteriol**. v.194, p.1605–06, 2012.

NAKAMURA, Y.; et al. Biased biological functions of horizontally transferred genes in prokaryotic genomes. **Nat. Genet**, v.36, p.760–66. 2004.

NARITA, M. Pathogenesis of extrapulmonary manifestations of *Mycoplasma pneumoniae* infection with special reference to pneumonia. **J Infect Chemother**, v.16, p.2–9. 2010.

NEDELCO A. M.; et al. Adaptative eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. **J. Evol. Biol**, v.21, p.1852–60. 2008.

NEIMARK, H.; HOFF, B.; GANTER, M. *Mycoplasma ovis* comb. nov. (formerly *Eperythrozoon ovis*), an eperythrocytic agent of haemolytic anaemia in sheep and goats. **Int. J. Syst. Evol. Microbiol.**, v.54, p.365–71. 2004.

NELSON K. E.; et al. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. **Nature**, v.399, 323–29.1999.

NICHOLAS, RA. Contagious agalactia, **In Manual of diagnostic tests and vaccines for terrestrial animals**, v.7, n.5, p.987-94. 2008.

NOGUCHI, H.; TANIGUCHI, T.; ITOH, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. **DNA Res**, v.15, p.387–96. 2008.

OCHMAN, H. Lateral and oblique gene transfer. **Curr. Opin. Genet. Dev**. v.11, 616–19. 2001.

OCHMAN H.; LAWRENCE J. G.; GROLSMAN, E. A. Lateral gene transfer and the nature of bacterial innovation. **Nature**, v.405, 299–304. 2000.

OEHLERKING, J. et al. Complete Genome Sequence of the Hemotrophic *Mycoplasma suis* Strain KI3806. **Journal of Bacteriology**, v.193, n.9, 2369-70. 2011.

OPRIESSNIG, T.; GIMÉNEZ-LIROLA, LG.; HALBUR, PG. Polymicrobial respiratory disease in pigs. **Anim Health Res. Rev**, v.12, p.133–48. 2015.

OU, HY.; HE, X.; HARRISON, EM. web-based tools for *n silico* and experimental discovery of bacterial genomic islands. **Nucleic Acids Research**, v.35, p.97-104. 2007.

PÁL, C.; PAPP, B.; LERCHER, M. J. Adaptative evolution of bacterial metabolic networks by horizontal gene transfer. **Nat. Genet**, v.37, p.1372–75. 2005a.

PÁL, C., PAPP, B., LERCHER, M. J. Horizontal gene transfer depends on gene content of the host. **Bioinformatics**, v.21, p.222–23. 2005b.

PERSENER, J. Bioinformatics and Functional Genomics. **John Wiley & Sons**. 2003.

PETERS, I.; HELPS, R.; GRUFFYDD-JONES, T.; et al. Antigen specificity of the humoral immune response to *Mycoplasma haemofelis* infection. **Clin. Vaccine Immunol**, v.17, p.1238–43. 2010.

PHILIPPE, H.; DOUADY, C.; Horizontal gene transfer and phylogenetics. **Curr. Opin. Microbiol**, v.6, p.498–505. 2003.

POLLACK, JD. *Mycoplasma* genes: a case for reflective annotation. **Trends Microbiol**, v.5, p.413-19. 1997.

POSTMA, GC.; CARFAGNINI, JC.; MINATEL, L. *Moraxella bovis* pathogenicity: an update. **Comp. Immunol. Microbiol. Infect. Dis**, v.31, p.449–58. 2008.

RAGAN, M. A. Reconciling the many faces of lateral gene transfer. **Trends. Microbiol**, v.10, p.4. 2002.

RAZIN, S.; YOGEV D.; NAOT, Y. Molecular biology and pathogenicity of *Mycoplasmas*. **Microbiol Mol Biol Rev: MMBR**, v.62, p.1094–156. 1998.

VERNIKOS, GS.; PARKHILL, J. Resolving the structural features of genomic islands: a machine learning approach. **Genome Res**, v.18, n.2, p.331-42. 2008.

REST, J. S.; MINDELL, D. P. Retroids in Archaea: phylogeny and lateral origin. **Mol. Biol. Evol**, v.2, p.1134–142. 2003.

RODRIGUEZ, JL.; et al. High mortality in goats associated with the isolation of a strain of *Mycoplasma mycoides*. **Zentralbl Veterinarmed**, v.42, p.587–93. 1995.

ROSENGARTEN, R.; et al. Host-pathogen interactions in mycoplasma pathogenesis: virulence and survival strategies of minimalist prokaryotes. **Int. J. Med. Microbiol**, v.290, p.15–25. 2000.

ROWE-MAGNUS, D. A.; MAZEL, D. Resistance gene capture. *Current Opinion in Microbiology*, v. 2, p. 483-88, 1999.

SAUNDERS, N. J.; HOOD, D. W.; MOXON, E. R. Bacterial evolution: Bacteria play pass the gene. **Curr. Biol**, v.9, p.180–83. 1999.

SCHELER, G. *Self-organization of signal transduction. F1000 research*, v.2, p.116. 2013.

SENTCHILO, V.; et al. Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. Strain B13. **Mol Microbiol**, v.49, p.93–104. 2003.

SEYA, T.; MATSUMOTO, M. A lipoprotein family from *Mycoplasma fermentans* confers host immune activation through Toll-like receptor 2. **Int J Biochem Cell Biol**, v.34, p.901–06. 2002.

SIQUEIRA, FM.; et al. New insights on the biology of swine respiratory tract *Mycoplasmas* from a comparative genome analysis. **BMC Genomics**, v.14, p.175, 2013.

SIRAND-PUGNET, P.; et al. Being Pathogenic, Plastic, and Sexual while Living with a Nearly Minimal Bacterial Genome. **Matic I, ed. PLoS Genetics**, v.3, n.5, p.75. 2007.

SPLITTER, EJ. *Eperythrozoon suis* n. sp. and *Eperythrozoon parvum* n. sp., 2 new blood parasites of swine. **Science**. v.111, p.513–14. 1950.

STOCKER, B.; ZINDER, N. D.; Lederberg J. Transduction of flagellar characters in *Salmonella*. **J. Gen. Microbiol**, v.9, p.410. 1953.

- SULLIVAN, JT.; et al. Comparative Sequence Analysis of the Symbiosis Island of *Mesorhizobium loti* Strain R7A. **Journal of Bacteriology**, v.184, n.11, p.3086-95. 2002.
- SUTTON, RH. *Eperythrozoon ovis*—a blood parasite. **Vet J**, v.18, p.156–64. 1969.
- TATUSOV, RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. **Nucleic Acids Res.** v.28, p.33-36. 2000.
- TAYLOR-ROBINSON, D.; JENSEN, JS. *Mycoplasma genitalium*: from chrysalis to multicolored butterfly. **Clin. Microbiol.** V.24, p.498–514. 2011.
- TEGTMEIER, C.; et al . Pathological and microbiological studies on pneumonic lungs from Danish calves. **J Med Vet Br**, v.46, p.693–700. 1999.
- TENOVER, F.C. Mechanisms of antimicrobial resistance in bacteria. **The American Journal of Medicine.** v.119, p.3-10. 2006.
- THOMAS, C. M.; NIELSEN, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. **Nature Reviews Microbiology**, v. 3, p. 711– 21. 2005.
- TORTORA, G. J.; FUNKE, B. R.; CASE, C. L. **In: Microbiologia**, 10.ed. 2012.
- TU, QIANG.; DING, DAFU. Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. **FEMS microbiology letters**, v. 221, p.269-75. 2003.
- TULLY, JG.; et al. Characterization of some caprine Mycoplasmas, with proposals for new species, *Mycoplasma capricolum* and *Mycoplasma putrefaciens*. **J. Gen. Microbiol**, v.85, p.102–20. 1974.
- VAN DER MEER, JR; SENTCHILO, V. Genomic islands and the evolution of catabolic pathways in bacteria. **Curr Opin Biotechnol**, v.14, p.248–54. 2003.
- VASCONCELOS, ATR.; et al. Swine and Poultry Pathogens: the Complete Genome Sequences of Two Strains of *Mycoplasma hyopneumoniae* and a Strain of *Mycoplasma synoviae*. **Journal of Bacteriology**, v.187, n.16, p.5568-77. 2005.

VERNIKOS, GS.; Thomson, NR.; Parkhill, J.; Genetic flux over time in the *Salmonella* lineage. **Genome Biol**, v.8, p.100. 2007.

WATKINS, R. F.; GRAY, M. W. The frequency of eubacterium-to-eukaryote lateral gene transfer shows significant cross-taxa variation within Amoebozoa. **J. Mol. Evol.** v.63, p.801–14. 2006.

WHITE, BJ.; et al. Mollicutes species and *Mycoplasma bovis* prevalence and association with health outcomes in beef feeder calves at arrival and initial treatment for bovine respiratory disease. **Can Vet J**, v.51, p.1016–18. 2010.

WHITEHOUSE D. B. et al. A phylogenetic approach to the identification of phosphoglucomutase genes. **Mol. Biol. Evol**, v.15,p. 456–62. 1998.

WINNER, F.; ROSENGARTEN, R.; CITTI C. In vitro cell invasion of *Mycoplasma gallisepticum*. **Infect Immun**, v.68, p.4238–44. 2000.

XIAO, L.; et al. Comparative genome analysis of *Mycoplasma pneumoniae*. **BMC Genomics**, v.16, n.1, p.610. 2015.

YAMAMOTO, R. *Mycoplasma meleagridis* infection. **Diseases of poultry**, v.9, p.212-33. 1991.

YÁÑEZ, A.; et al. *Mycoplasma penetrans* bacteremia and primary antiphospholipid syndrome. **Emerg. Infect Dis**, v.5, p.164–67. 1999.

YI-CHANG, LIU.; I-HSUAN, LIN.; WEI-JEN CHUNG. Proteomics Characterization of Cytoplasmic and Lipid-Associated Membrane Proteins of Human Pathogen *Mycoplasma fermentans* M64. **PLoS One**, v.7, p.4. 2012.

ZINDER, N. D.; Lederberg, J. Genetic exchange in *Salmonella*. **J. Bacteriol.** 64, 679–699. 1952.

ZUCKERKANDL, E.; PAULING, L. Molecules as documents of evolutionary history. **J. Theor. Biol.** v.8, p.357–66. 1965.