Universidade Federal do Pará

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Biotecnologia

Estudo de genômica comparativa de *Corynebacterium pseudotuberculosis* linhagem 226 (biovar *ovis*)

Larissa Maranhão Dias

Belém - Pará
2015.

Universidade Federal do Pará

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Biotecnologia

Estudo de genômica comparativa de *Corynebacterium pseudotuberculosis* linhagem 226 (biovar *ovis*)

Larissa Maranhão Dias

Dissertação submetida ao Programa de Pós Graduação em Biotecnologia da UFPA como requisito parcial para obtenção do grau de Mestre em Biotecnologia

Orientador: Profª Dra. Adriana Ribeiro Carneiro

Belém – Pará
2015.

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Dias, Larissa Maranhão, 1987-
Estudo de genômica comparativa de *corynebacterium pseudotuberculosis* linhagem 226 (biovar *ovis*) / Larissa Maranhão Dias. - 2015.

Orientadora: Adriana Ribeiro Carneiro.
Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Ciências Biológicas, Programa de Pós-Graduação em Biotecnologia, Belém, 2015.

1. Corynebacterium pseudotuberculosis Patogenicidade. 2. Genômica. 3. Caprino Doenças. 4. Linfadenite caseosa. I. Título.

CDD 22. ed. 579.373

AGRADECIMENTOS

Primeiramente a Deus, pois foi Ele que me amparou e não permitiu que desistisse nos momentos mais difíceis durante o desenvolvimento desse trabalho.

A Universidade Federal do Pará, juntamente com o Programa de Pós-Graduação em Biotecnologia pela oportunidade de aprendizado.

A CAPES pela concessão da bolsa.

A minha orientadora, Profª. Drª. Adriana Carneiro por aceitar este desafio junto comigo e por todo incentivo, paciência e auxílio na construção do meu conhecimento durante o período de execução desse trabalho.

Ao Profº. Drº. Rommel Ramos, por sua atenção e disponibilidade em me ajudar sempre que precisei.

Ao Profº. Drº. Arthur Silva pela oportunidade de trabalhar em seu grupo de pesquisa.

A Silvanira e Soraya pela disponibilidade e atenção em tirar minhas dúvidas. Ao Davi e Vagner por toda auxílio dentro do laboratório.

A todos os meus amigos do LGBS pelos momentos de descontração e apoio. Em especial, agradeço aos amigos Yuri, Amália, David, Sávio, Pablo, Kenny, Edson e Rafael por toda ajuda na execução deste trabalho.

As minhas queridas amigas conquistadas no LGBS, Jorianne, Jaqueline e Ana Lídia pela paciência, apoio e colaboração no decorrer desse trabalho.

Ao Allan que sempre esteve presente e me auxiliou em diversos momentos durante esse período.

As minhas amigas Taline, Cássia, Vânia e Jéssica pelo apoio e compreensão nos momentos de ausência.

A minha família, em especial meus pais João e Marisa, e minha irmã Nayara pelos ensinamentos, exemplos e incentivo, sem os quais não teria chegado até aqui.

A todos o meu muito OBRIGADO!!!!!!!!!!!!!!!!!!!

SUMÁRIO

LISTA DE ABREVIATURAS

ACT	*Artemis Comparition Tool*
BHI	Meio de cultura de Infusão coração-cérebro
BLAST	*Basic Local Alignment Search Tool*
CCD	*Charge-coupled device*
CDS	Sequência Codificante de proteína (*Coding Sequencing*)
DDNTP's	Dideoxinucleotídeos
DNA	Ácido dexorribonucléico
GC	Guanina e Citosina
Gb	Giga bases
GEI's	Ilhas Genômicas
GO	*Gene Ontology*
GOLD	*Genomes Online Database*
LC	Linfadenite Caseosa
LGCM	Laboratório de Genética Celular e Molecular
LGT	Transferência Lateral de Genes
LGBS	Laboratório de Genômica e Biologia de Sistemas
LU	Linfangite Ulcerativa
MB	Mega bases
NCBI	*Nacional Center for Biotechnology Information*
NGS	*Next Generation Sequencing*
N50	Média estatística para avaliar um conjunto de *contigs* gerados por montagem *de novo*.
OFR	*Open Read Frames*
PAI's	Ilhas de Patogenicidade
PGM	*Personal genome Machine*
pb	Pares de bases
pH	potencial hidrogeniônico
RAST	*Rapid Annotation using Subsystem Technoly*
RNA	Ácido ribonucléico
RPGP	Rede Paraense de Genômica e Proteômica
rRNA	RNA ribossomal
SNP	Polimorfismo de Nucleotídeo Único
THG	Transferência Horizontal de Genes
tRNA	RNA transportador

LISTA DE TABELAS

LISTA DE FIGURAS

RESUMO

Corynebacterium pseudotuberculosis é uma bactéria Gram-positiva, intracelular facultativa, não-esporulante, não-capsulada e sem mobilidade, contudo possui fímbria, e pode assumir formas cocóides e filamentosas (pleomórfica), além disto, apresenta crescimento ótimo à 37°C. Este patógeno apresenta dois biovares: *ovis* que geralmente acomete pequenos ruminantes, e causa a doença linfadenite caseosa, e biovar *equi*, mais comum em equinos, bovinos, camelídeos, e bubalinos causando a Linfangite ulcerativa. A infecção por esta bactéria pode levar a condenação das carcaças e redução de lã (em ovinos e caprinos), leite e carne destes animais, e consequentemente a perdas econômicas para a indústria agropecuária mundial. Atualmente, ainda não existe uma vacina eficaz para estas doenças. A fim de obter um maior entendimento biológico entre as espécies o presente trabalho tem como objetivo principal analisar, por meio da genômica comparativa a linhagem *C. pseudotuberculosis* 226 biotipo *ovis* isolada de um caprino na Califórnia com outras linhagens do biovarar *ovis* e *equi*. Na análise de sintenia entre as linhagens foi possível identificar que a linhagem 226 apresenta alta conservação da ordem gênica entre as linhagens do biótipo *ovis*. Através de análises filogenômicas foi possível identificar que as linhagens I19 e 267 apresentaram maior e menor proximidade filogenética com a linhagem 226. A linhagem 1/06-A foi a que apresentou maior proximidade filogenômica entre as linhagens do biovar *equi*, quando comparadas a linhagem 226. Foram preditas 8 ilhas de patogenicidade, estando presente na ilha 1 os genes relacionados a virulência de *C. pseudotuberculosis* mais bem descritos na literatura. Não houveram regiões novas relacionadas a genes de virulência entre nenhuma das linhagens. Foram identificados 248 genes ortólogos entre as linhagens I19, 267 e 226 e 282 genes ortólogos entre as linhagens 258,1/06-A e 226. Com base nesse estudo é possível inferir que as linhagens do biovar *ovis* possuem um repertório gênico pouco variado e que as linhagens do biovar *equi* apresentam uma quantidade menor de genes compartilhados com a linhagem 226, corroborando com a diversidade gênica entre os biovares.

Palavras Chave: *C. pseudotuberculosis*, caprino, patogenicidade, biovar *ovis* e *equi*.

ABSTRACT

Corynebacterium pseudotuberculosis is a gram-positive, facultative intracellular, non-sporulating and non-encapsulated bacterium, it is non-motile although it has fimbriae, and can assume coccoid or filamentous forms (pleomorphic). Its optimum growth temperature is 37°C. This pathogen has two biovars: *ovis*, which usually affects small ruminants and causes caseous lymphadenitis, and biovar *equi*, more common in equines, bovines, camelids and bubalines, causing ulcerative lymphangitis. Its infection can lead to carcass condemnation and reduction in wool production (in ovines and caprines), milk production and meat production and, consequently, economic losses for the agricultural industry worldwide. Currently there is no effective vaccine against those illnesses. To obtain a better understanding of these species biologically, the main objective of this work is to analyze, using comparative genomics, the strain *C. pseudotuberculosis* 226 biovar *ovis*, isolated from a caprine in California, comparing it to other strains from biovars *ovis* and *equi*. The synteny analysis revealed highly conserved gene order between strain 226 and other biovar *ovis* strains. Phylogenomic analyses showed that the strains I19 and 267 are, respectively, the closest and the more distant phylogenetically from strain 226. Among biovar *equi* strains, the one with the greater phylogenomic proximity to strain 226 was strain 1/06-A. Eight pathogenicity islands were predicted, with *C. pseudotuberculosis* best characterized virulence genes in literature being present in island 1. No new regions related to virulence genes could be found compared to other strains. 248 orthologous genes could be found between strains I19, 267 and 226, while 282 orthologous genes could be found between strains 258, 1/06-A and 226. Based in this study it is possible to assume that strains from biovar *ovis* have a little varied gene repertory and strains from biovar *equi* have less genes shared with strain 226, reinforcing the genetic diversity between these biovars.

Keywords: *C. pseudotuberculosis*, goat, pathogenic, biovar *ovis* and *equi*.

1. INTRODUÇÃO

1.1. TECNOLOGIAS DE SEQUENCIAMENTO

Em 1977, Frederick Sanger desenvolveu um método de decodificação de DNA denominado método de terminação de cadeia, fazendo uso de DDNTP's (dideoxinucleotídeos) marcados com isótopos radioativos impedindo que os mesmos se ligassem a cadeia. Esta técnica, inicialmente utilizava gel de poliacrilamida, e o resultado era obtido através de autoradiografia e leitura manual das bases nitrogenadas (Sanger & Nicklen, 1977).

Por volta de 1987, a empresa *Applied Biosystems*, atualmente *Life Technologies* (https://www.lifetechnologies.com/), lançou no mercado o primeiro sequenciador automático que faz uso da eletroforese em capilares, esta tecnologia reduziu o contato manual durante a preparação do sequenciamento, aumentou a acurácia e rapidez na obtenção dos resultados. Ao longo dos anos diversas atualizações foram realizadas nesta plataforma (Shendure & Ji, 2008; Zhang *et al*., 2011).

Este equipamento foi amplamente utilizado por três décadas, o que levou a decodificação de genes completos e mais tarde genomas inteiros, sendo a tecnologia utilizada no Projeto Genoma Humano (Venter *et al*., 2001; Schuster, 2008). Apesar disto, o mesmo apresentava algumas desvantagens, dentre elas: uso de clonagem gênica, elevado tempo de corrida e alto custo por base sequenciada (Richardson, 2010).

Assim em meados de 2005, surgem às plataformas de sequenciamento de segunda geração, também denominadas de NGS (*Next Generation Sequencing*). A disponibilidade destas tecnologias proporcionou avanços na área genômica e consequentemente revolucionou a biologia e os conhecimentos sobre sequenciamento de DNA (Schuster, 2008; Pareek *et al*., 2011; *Liu et al*., 2012; El-Metwally *et al*., 2013; Yadav *et al*., 2014).

Dentre as plataformas de segunda geração podemos citar: 454 da Roche; MiSeq e HiSeq da Illumina; e SOLiD 5500xl W. Estes equipamentos apresentam algumas características comuns, geração de grande quantidade de dados em uma única execução, a alta acurácia, e redução considerável do tempo e custo por sequenciamento (Henson *et al*., 2012; El-Metwally *et al*., 2013).

1.1.1. Ion Torrent PGM™ (Personal Genome Machine)

Em 2010, a empresa *Life Technologies* disponibilizou comercialmente o Ion Torrent PGM (*Personal Genome Machine*), uma plataforma que possui um método de sequenciamento inovador baseado em chips semicondutores capazes de detectar mudanças de pH resultantes da liberação de íons de hidrogênio (H^+) que ocorre quando a enzima DNA polimerase incorpora o nucleotídeo, conforme demonstrado na figura 1. Este equipamento foi o primeiro a não utilizar fluorescência ou quaisquer instrumentos óticos, como câmeras CCD (charge-coupled device), sendo uma tecnologia que inaugura a era do sequenciamento pós-luz, levando a redução do custo por base sequenciada (Rusk, 2010; Merriman *et al*., 2012; Bragg *et al*., 2013).

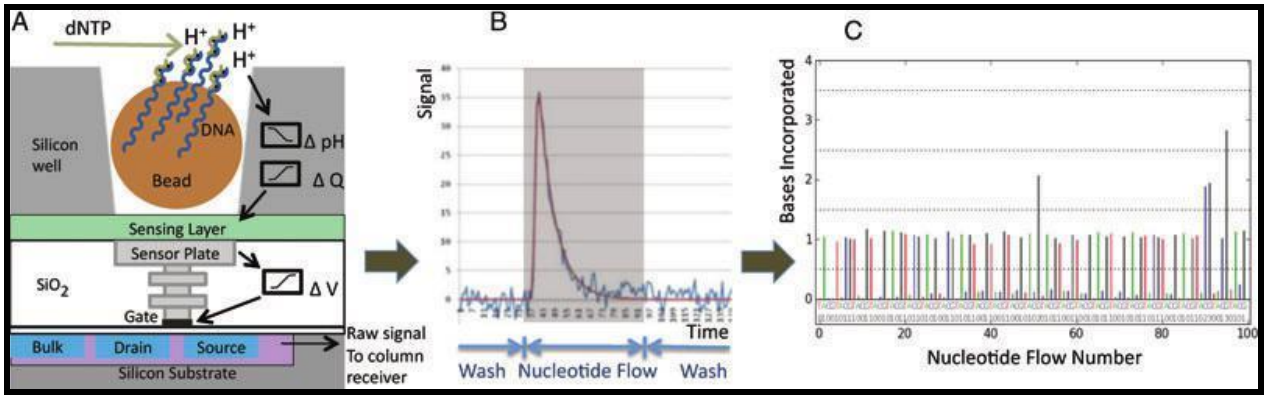


Figura 1. Processo de sequenciamento que ocorre no Ion Chip. Fonte: Merriman *et al*. 2012.

O sequenciamento no Ion Torrent (PGM) pode ser realizado em um dos três chips, os quais diferem de acordo com a quantidade de dados a ser gerada (*throughput*) (Figura 2). Atualmente, Ion Chip 314 pode produzir de 30 a 100 megabases (Mb), Ion Chip 316 de 300 Mb a 1 gigabases (Gb) e Ion Chip 318 600 Mb a 2 Gb (https://www.lifetechnologies.com/).

As leituras geradas por esta plataforma podem chegar ao comprimento de 400 pb (www.lifetechnologies.com) e é possível construir bibliotecas genômicas de fragmentos *(Silva et al*., 2012) e pareada, onde no ano de 2011 obteve-se o primeiro genoma completo de *Corynebacterium pseudotuberculosis* (Cp) 316 sequenciado e depositado ao NCBI (*National Center for Biotechnology Information*) com uso deste tipo de biblioteca (Ramos *et al*., 2012).

Neste ano a mesma empresa lançou a plataforma Ion Proton™, capaz de produzir com o chip PI cerca de 10 Gb (https://www.lifetechnologies.com/).

Figura 2. Chips semicondutores utilizados no Ion Torrent PGM. Fonte: https://www.lifetechnologies.com

O erro que pode ser observado com maior frequência nesta plataforma é o INDEL (Inserção/Deleção), o qual ocorre devido à química utilizada denominada "síntese", onde a detecção dos nucleotídeos ocorre a cada ciclo em que é realizada a introdução e remoção dos reagentes (Merriman & Rothberg, 2012). Devido à natureza deste método não há a detecção de múltiplos nucleotídeos em um dado ciclo, fazendo com que a leitura de regiões homopoliméricas não seja precisa, resultando no acréscimo ou remoção de uma base (Yeo *et al*., 2012).

O surgimento de novas tecnologias de sequenciamento possibilitou-se o sequenciamento de genomas completos em poucas horas (Henson *et al.,* 2012), o que contribuiu nos trabalhos nos três domínios da vida, como por exemplo, o domínio *Bacteria* que até o ano de 2015 apresentou 39.257 projetos relacionados (https://gold.jgi-psf.org/statistics). Esta nova tecnologia favorece o crescimento do número de *drafts* (com 15.635 projetos) e de genomas completos, atualmente há 2.031 projetos depositados nos bancos de dados públicos de acordo com a figura 3.

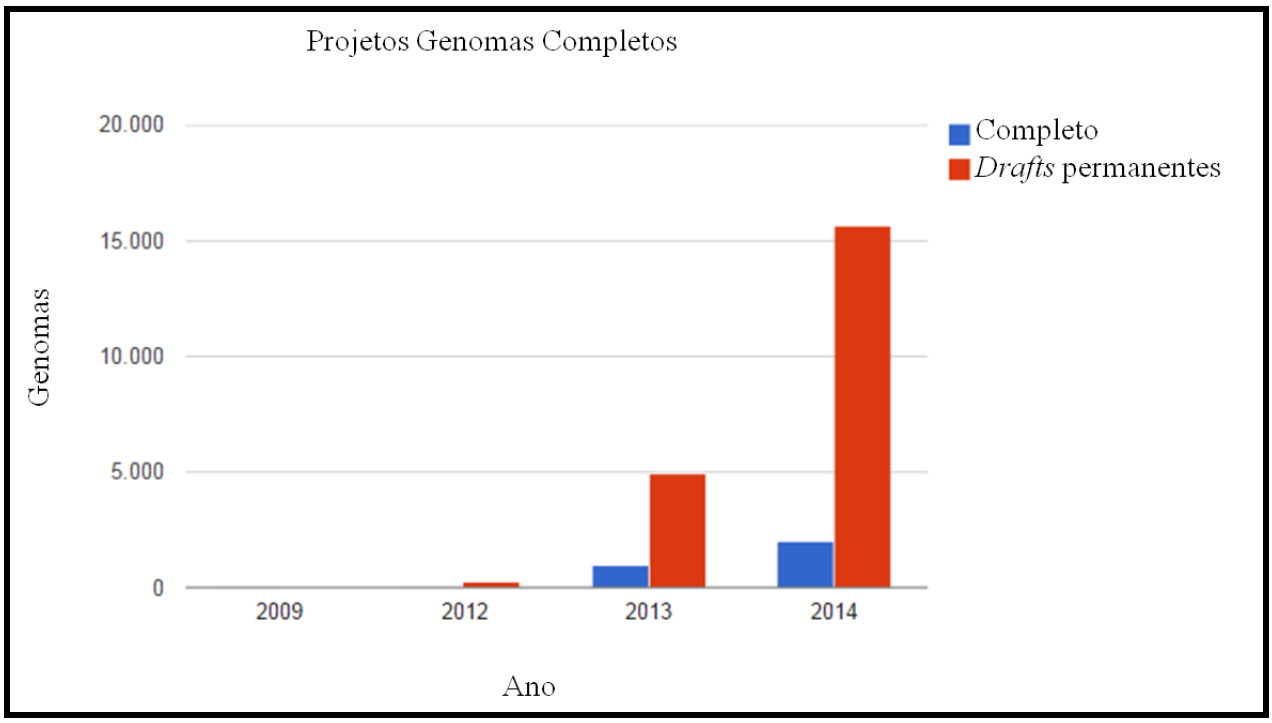


Figura 3. Quantidade total de genomas completos e *drafts* depositados em GOLD, onde: no eixo X representa o ano e no Y a quantidade de genomas. Fonte: GOLD.

A elevada produção de dados e o comprimento das leituras geradas por estas novas tecnologias de sequenciamento trouxe novos desafios para Bioinformática, pois novas ferramentas e *pipelines* computacionais precisaram ser desenvolvidas para realizar as etapas de montagem e fechamento de genomas, bem como tornou necessário estruturas computacionais mais robustas capazes de armazenar e processar estes dados (El-Metwally *et al*., 2013).

1.2. ESTRATÉGIAS DE MONTAGEM

A montagem é definida como um processo onde as leituras são agrupadas por identidade entre as bases, visando à reconstrução do genoma alvo, a mesma pode ser categorizada em duas metodologias: montagem por referência e *de novo* (Miller *et al*., 2010).

1.2.1. Montagem *de novo*

Na estratégia da montagem *de novo*, não são utilizados nenhum tipo de referência ou orientação (El-Metwally *et al*., 2013). Devido a sua complexidade ela pode ser subdividida em três grandes paradigmas que são: algoritmos gulosos, *overlap-layout-consensus* e grafo *de Bruijn* (Miller *et al*., 2010).

As ferramentas computacionais que fazem uso dos algoritmos gulosos iniciam o processo de montagem fazendo uma busca recursiva, testando todas as leituras umas contra as outras, buscando os melhores valores de sobreposições, este processo é executado até não se obter mais leituras. Quando o processo é finalizado as leituras são estendidas formando assim as sequências contíguas ou *contigs* (Nagarajan & Pop, 2013). Exemplos de ferramentas que usam este algoritmo são SSAKE (Warren *et al*., 2007), SHARCGS (Dohm *et al*., 2007) e SOAP*denovo* (Li *et al*., 2010).

A abordagem *overlap-layout-consesus* possui algumas características encontradas nos algoritmos gulosos, tendo em vista que o processo é a busca dos melhores valores de sobreposição fazendo alinhamentos das leituras umas contra as outras, e este resultado é utilizado para a construção do grafo, onde cada leitura representa um nó deste grafo e as sobreposições são as pontes que ligam estes nós (Pop & Salzberg, 2008). Dentre os softwares de montagem que utilizam esta abordagem pode-se citar: Mira 4.0.(Chevreux *et al*., 2004), SHORTY (Hossain *et al*., 2009) Readjoiner (Gonnella & Kurtz, 2012) e Fermi (Li, 2012).

O Grafo *de Bruijn* é considerado atualmente uma das estratégias mais utilizadas no desenvolvimento de *softwares* de montagem, isto é, devido ao seu alto desempenho, o qual proporciona o processamento de grande quantidade de dados. Nesta estratégia de montagem, o processo inicia com a quebra das leituras em subleituras de tamanho fixo denominado *k-mer*, após isto os *k-mers* são utilizados para a construção do grafo, sendo cada *k-mer* um nó do grafo e

as sobreposições entre eles, chamadas de *k-1*, são as pontes que ligam os nós, assim, não é necessário realizar o primeiro passo buscando os melhores valores de sobreposições e, as leituras propriamente ditas não são utilizadas para construção do grafo, resultando assim na redução da complexidade do grafo (Nagarajan & Pop, 2013). Exemplos de programas que utilizam esta abordagem são: Velvelt (Zerbino & Birney, 2008), ALLPATHS (Butler *et al.*, 2008) e ABySS (Simpson *et al*., 2009) e SPAdes (Bankevich *et al*., 2012).

1.3. ANOTAÇÃO FUNCIONAL

As tecnologias de sequenciamento de segunda geração proporcionaram o aumento no número de projetos e depósitos de genomas completos e *drafts* em banco de dados públicos, assim, várias linhagens de uma mesma bactéria podem ser sequenciadas, o que contribui para estudos mais detalhados na área da genômica. Entretanto, decifrar as sequências de DNA, identificar e anotar os genes e seus produtos proteicos faz-se necessário para agilizar o processo, bem como minimizar erros de nomenclaturas de produtos e genes presentes nos bancos de dados públicos através de uma revisão manual (Médigue & Moszer, 2007;Petty, 2010).

De maneira geral, anotação consiste em inferir conceitos biológicos às sequências de DNA, e tem por objetivo apontar as principais características da sequência genômica que pode ser classificada em dois níveis: automática (estática) ou manual (dinâmica) (Stein, 2001).

Na anotação automática são utilizados diversos programas que realizam predições de regiões abertas para leituras - *Open Read Frames* (ORF'S), e localização de regiões codificantes de proteínas- *Coding Sequencing* (CDS), que podem estar presentes em ORF'S preditas *in sílico*, além de RNA's ribossomais (rRNA) e transportadores (tRNA). Porém estes podem não identificar pequenos genes, além disto, *drafts* de genomas depositados podem conter altas taxas de erros de sequenciamento, acarretando em predições equivocadas de genes (Médigue & Moszer, 2007).

Com isso, faz-se necessário a anotação manual, a qual é dependente da etapa automática, onde os elementos estruturais são colocados em um contexto biológico, e são obtidas relações entre as proteínas, interações entre reguladores e/ou vias metabólicas. Esta etapa ainda pode contribuir para correção ou aumento da qualidade das anotações atribuídas (Médigue & Moszer, 2007).

Como exemplo de *softwares* que realizam anotação automática, temos: *FgenesB* (http://www.softberry.com/), Glimmer (http://www.genomics.jhu.edu/Glimmer/), GeneMark (http://exon.gatech.edu/genemark) e o sistema *web* RAST, uma ferramenta que faz a anotação automática dos principais elementos estruturais do genoma, o qual faz uso de subsistemas para implementar processos biológicos específicos como genes relacionados a parede celular, mecanismos de defesa entre outros (Aziz *et al*., 2008). Dentre os bancos de dados públicos utilizados para a curadoria da anotação manual podemos citar: UniProt (http://www.expasy.uniprot.org/), Pfam (http://www.sanger.ac.uk/Software/Pfam/) e InterPro (http://www.ebi.ac.uk/interpro/).

A partir dos dados biológicos obtidos da anotação automática e manual é possível compreender as informações funcionais, ou seja, pode-se incorporar elementos relacionados às funções e processos biológicos que permitem distribuir os genes em classes funcionais, estas análises podem ser obtidas via *Gene Ontology* (GO), um banco de dados de ontologia gênica capaz de realizar esta tarefa é o Blast2GO (Conesa *et al*., 2005). Assim, pode-se notar que o processo de anotação é um passo crucial para um melhor entendimento das interações e funcionamentos biológicos (Reed *et al*., 2006; Médigue & Moszer, 2007).

1.4. GENÔMICA COMPARATIVA

Os estudos de genômica comparativa iniciaram em meados de 1995 quando o grupo de Craig Venter publicou o sequenciamento de dois genomas bacterianos (Fleischmann *et al*., 2008; (Fraser *et al*., 2014). Desde 2005, a quantidade de dados genômicos depositados nos bancos de dados públicos vem crescendo rapidamente devido ao desenvolvimento das novas plataformas de sequenciamento, o que vem contribuindo para a realização de estudos na área da genômica comparativa (Field *et al*., 2006; Hu *et al*., 2011).

As pesquisas comparativas oferecem detalhes adicionais, que podem auxiliar na descoberta de genes alvos de interesse biotecnológico, biomédico, ambiental, etc. Através de análises pangenômicas, a qual consiste na comparação de várias linhagens de uma mesma espécie ou de diferentes espécies, é possível reconhecer diferenças e similaridades existentes entre os genomas e esclarecer quais sequências são capazes de divergir em mudanças fenotípicas nos

organismos e elucidar os mecanismos de virulência entre organismos patogênicos (Hu *et al*., 2011).

Um pangenoma é constituído pelo "genoma central", que configura os genes presentes entre todas as linhagens, isto é, os genes essenciais; o "genoma acessório" que compartilha genes entre duas ou mais cepas e inclui os genes que a bactéria necessita para sobreviver em um ambiente específico, além dos genes espécie-específicos pertencentes a uma única linhagem (Tettelin *et al*., 2005; Mira *et al*., 2010).

O genoma acessório pode estar presente em Ilhas Genômicas (GEIs), que são regiões do DNA que podem ter sido adquiridas por transferência horizontal (THG), contribuindo para a plasticidade, evolução, e a adaptação de microorganismos. Estas quando carregam genes de virulência são consideradas Ilhas de Patogenicidade (PAIs) (Moon, 2006; Juhas *et al*., 2009). Para realizar a predição destas ilhas diversos programas foram desenvolvidos como, por exemplo, o programa GIPSy, o qual identifica ilhas de patogenicidade por meio de análises de desvio de uso de códon, conteúdo GC e fatores de virulência entre uma bactéria de espécie patogênica e outra não patogênica (Soares *et al*., 2012). Além disto, estes ilhas podem ser comparadas através de programas como BRIG que, além disso, possibilita a geração de imagens que mostram comparações entre múltiplos de genomas procariotos (Alikhan *et al*., 2011).

Desta forma, nos estudos pangenômicos podem-se caracterizar o pangenoma de uma espécie como "aberto" ou "fechado". O pangenoma "aberto" indica a possibilidade do aumento de novos genes ao genoma *core* à medida que um novo genoma é inserido na análise, o mesmo não acontece no pangenoma "fechado", pois ele é caracterizado por apresentar pouca ou nenhuma aquisição de novos genes (Tettelin *et al*., 2005; Muzzi *et al*., 2007).

No estudo de Gao e colaboradores (2014), bactérias do gênero *Streptococcus* foram analisadas e revelaram que o genoma *core* aumenta de acordo com a inclusão de novos genomas. Neste estudo a cada novo genoma depositado para análise, 62 genes foram inclusos corroborando que este gênero possui um pangenoma aberto, ou seja, está em expansão (Gao *et al*., 2014).

Com o intuito de auxiliar nas análises de pangenoma diversos programas já foram desenvolvidos como PGAP que possibilita análises de variação genética e evolutivas. Esta ferramenta possibilita identificar o genoma *core* e dos *singletons* (genes únicos) (Zhao *et al*., 2012). Outra ferramenta que faz esse mesmo tipo de análise, porém de forma *online* é o *software* Panseq, com acréscimo de uma funcionalidade que é a identificação de regiões com

polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphisms*- SNPS) (Laing *et al*., 2010). O programa EDGAR fornece uma análise rápida de informações evolutivas, e ainda possibilita originar o gráfico de sintenia e o diagrama de Venn entre os genomas (Blom *et al*., 2009).

Com um número expressivo de depósitos de genomas deste microorganismo disponíveis em banco de dados de domínio público e seu interesse em diversas áreas, estudos de genômica comparativa desta bactéria já foram elucidados, como no trabalho de Soares e colaboradores 2013, através de estudos de pangenômica, relataram diferenças entre linhagens de *Corynebacterium pseudotuberculosis* e as que pertenciam ao biovar *ovis* apresentaram maior semelhança entre si, enquanto que as do biovar *equi* apresentaram maior diversidade gênica (Soares *et al*., 2013).

1.5. *Corynebacterium pseudotuberculosis*

Esta é classificada como um patógeno emergente

C. pseudotuberculosis pertence ao filo das Actinobacterias, onde estão inclusos os gêneros: *Mycobacterium*, *Nocardia*, *Rhodococcus* e *Corynebacterium*, que compõem o grupo CMNR, caracterizados por possuir um conteúdo de guaninas e citosinas (GC) entre 46 e 74%, de acordo com a espécie (Tauch & Sandbote, 2014). O gênero *Corynebacterium* possui 88 espécies validadas através de publicações, dentre estas, 53 são casos raros de infecções em humanos ou podem ser transmitidas através de zoonoses, as 35 restantes são encontradas no meio ambiente, água, alimentos, materiais sintéticos ou animais e aves (Tauch & Sandbote, 2014).

C. pseudotuberculosis é uma bactéria Gram-positiva, intracelular facultativa, não-esporulante, não-capsulada, sem mobilidade e possui fímbria. Pode assumir formas cocóides e filamentosas (pleomórfica), é um patógeno aeróbio facultativo que contem na composição de sua parede celular ácidos micólicos, peptídeoglicano e arabinogalactano, é capaz de crescer a uma temperatura ótima de 37°C em meio de cultura *Brain Heart Infusion* (BHI) (Selim, 2001; Connor *et al*., 2000; Dorella *et al*., 2006).

É uma bactéria mundialmente disseminada e descrita em diversos países como: Inglaterra, Canadá, Austrália, Brasil, Nova Zelândia, África do Sul e Estados Unidos, sendo um patógeno de

interesse médico, veterinário e biotecnológico (Arsenault *et al*., 2003; Paton *et al*., 2003; Dorella *et al*., 2006; Trost *et al*., 2010).

Esta bactéria pode ser classificada em dois biovares: *ovis* e *equi*, ambos diferenciados por suas propriedades bioquímicas, onde o biovar *ovis* é caracterizado pela redução de nitrato negativa e o biovar *equi* nitrato positiva (Dorella *et al*., 2006). O biovar *ovis* frequentemente acomete ovinos, caprinos e suínos causando a doença Linfadenite Caseosa (LC), em humanos pode causar a Linfadenite subaguda a crônica (Yeruham *et al*., 2004 ;Fontaine & Baird, 2008; Trost *et al*., 2010; Oliveira *et al*., 2014). O biovar *equi* acomete bubalinos, equinos, camelídeos e bovinos, e pode causar a doença Linfangite Ulcerativa (LU) (Pratt *et al*., 2005).

Estas doenças causam perdas na economia agropecuária mundial, pois há diminuição da produção de carne, leite e lã destes animais, levando a condenação das carcaças por completo e tem como principal característica o desenvolvimento de lesões piogranulomatosas (Figura 4) (Arsenault *et al*., 2003).

Figura 4. Lesões piogranulomatosas ocasionadas por *C. pseudotuberculosis*. Fonte: Fontaine & Baird, 2008.

O primeiro contato de *C. pseudotuberculosis* no hospedeiro ocorre o processo de fagocitose por macrófagos, entretanto este processo é ineficaz, pois a bactéria tem a capacidade de resistir à fagocitose, permitindo desta forma a sua multiplicação dentro da célula hospedeira e posterior rompimento celular, possibilitando a infecção em outros macrófagos (Songer, 2005).

Desta forma, seu processo de patogênese é bem caracterizado, entretanto os determinantes de virulência ainda não são totalmente elucidados (Dorella *et al*., 2006).

A medicação desta doença pode ser pela administração de antibióticos, entretanto estes recursos têm custos elevados além de não possuírem total eficácia (Olson *et al*., 2002). Além disso, pode-se usar a drenagem e aspiração dos linfonodos superficiais como alternativa, entretanto este procedimento é inviabilizado quando existem linfonodos internos, ademais esta intervenção não leva a cura, pois o animal permanece infectado, causando ainda cicatrizes na pele do animal e possível contaminação da bactéria no ambiente (Arsenault *et al*., 2003; Alves & Pinheiro, 2003).

Algumas vacinas estão disponibilizadas comercialmente no mercado nacional e internacional, entretanto as vacinas habilitadas para o uso em caprinos não apresenta a mesma eficácia para ovinos (Williamson, 2001). No Brasil, uma vacina a partir da linhagem 1002 vem sendo comercializada pela empresa Baiana de Desenvolvimento Agrícola (http://www.ebda.ba.gov.br/). Trata-se de uma cepa atenuada, porém seu uso é restrito em animais sadios e somente para dois tipos de hospedeiros, ovinos e caprinos (Dorella *et al*., 2009).

Atualmente, este organismo possui 19 genomas completos depositados no banco de dados do NCBI. Dentre estes, 15 foram depositados através de parcerias entre o Laboratório de Genômica e Biologia de Sistemas (Universidade Federal do Pará) e Laboratório de Genética Celular e Molecular (Universidade Federal de Minas Gerais) com apoio da Rede Paraense de Genômica e Proteômica (RPGP).

1.6. DOENÇAS EM CAPRINOS

Em caprinos *C. pseudotuberculosis* causa uma infecção crônica que pode apresentar dois tipos de manifestações. A externa, também conhecida como cutânea ou superficial, pois causa abscessos nos linfonodos linfáticos superficiais ou nos tecidos subcutâneos (Figura 5), os abscessos podem se desenvolver por um longo período, tornando-se inchado e envolto em cápsulas fibrosas, acarretando em perdas do pelo e ruptura, possibilitando a exposição das lesões purulentas, sendo estas algumas das causas de transmissões cruzadas entre animais por meio do contato com as mesmas (Fontaine & Baird, 2008).

Figura 5. Abscessos externos causados por *C. pseudotubeculosis* em pequenos ruminantes. Fonte: Fontaine & Baird, 2008.

A segunda é a forma visceral, estas lesões não podem ser notadas externamente, e tem como sítio de infecção normalmente linfonodos internos em órgãos como: fígado, pulmões, rins e glândulas mamárias, e de forma menos frequente em órgãos como coração, medula espinhal, cérebro, testículos, úteros e as articulações (Figura 6A e 6B). As lesões também podem ser assintomáticas, o que pode dificultar o diagnóstico da doença (Peel *et al*., 1997; Fontaine & Baird, 2008; Seyffert *et al*., 2010).

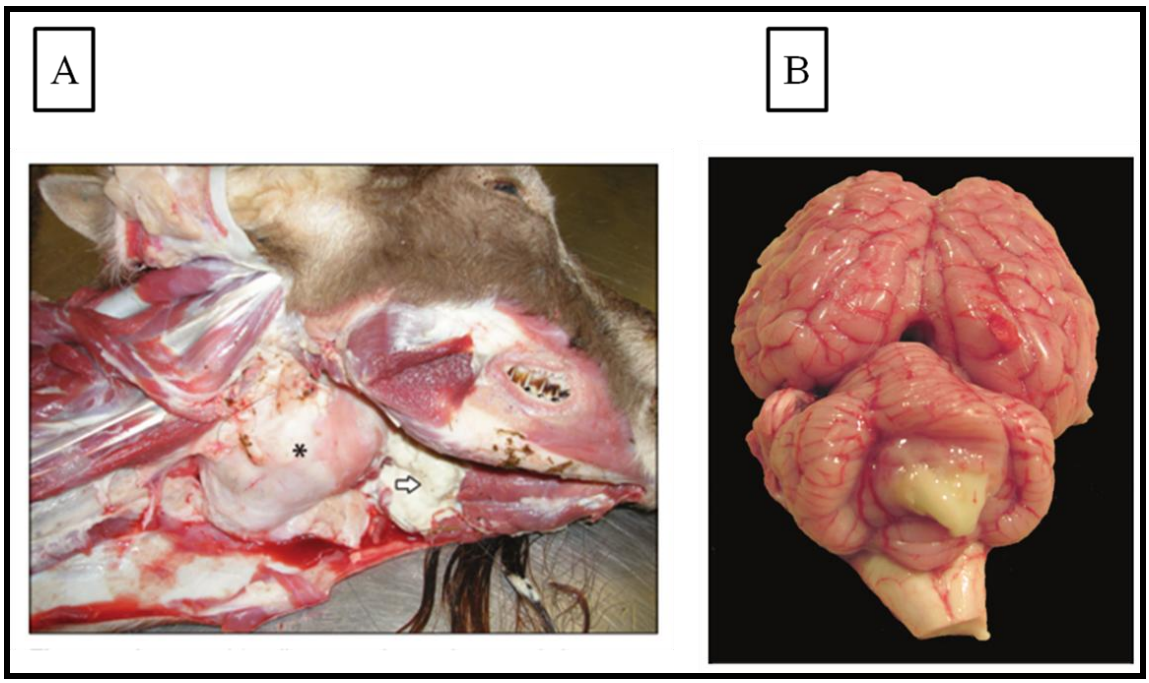


Figura 6. Abscessos internos causados por *C. pseudotuberculosis* em cabras. **A**, abscesso retrofaríngeo. **B**, abscesso cerebral. Fonte: Debien *et al*., 2013.

Esta bactéria é mundialmente distribuída e na Venezuela a criação de caprinos tem grande importância econômica, principalmente nas zonas áridas e semi-áridas, devido a falta de medidas sanitárias foi possível a disseminação desta doença, com o primeiro caso documentado de isolamento de *C. pseudotuberculosis* em cabras no ano de 1962, sendo estas importadas dos USA (Gallo & Morris, 1962;Chirino-Zárraga *et al*., 2006).

No América do Norte, especialmente no Canadá, as causas de morte em caprinos não estão bem documentadas, um estudo com 13 rebanhos de caprinos, teve 152 casos de cabras submetidas à necropsia, onde destes 54 foram de infecções causadas por esta bactéria que continham abscessos, tendo 3,9% de mortes causadas por LC (Debien *et al*., 2013).

Entre os meses de novembro de 2009 e agosto de 2010 um total de 466 cabras coreanas, a partir de 40 rebanhos teve soro coletado para a identificação de *C. pseudotuberculosis* nos abscessos presentes nos animais. Das 466 cabras, 267 (53,3%) deram soropositivas para *C. pseudotuberculosis* (Jung *et al*., 2015).

No Brasil a caprinocultura durante anos recentes apresentam uma tendência de crescimento (Brasileira de Pesquisa Agropecuária, 2008), com a região Nordeste do Brasil apresentando a maior produção de caprinos de ovinos (Brown *et al*., 1987). Em Rondón do Pará

(Estado do Pará), 237 vacas leiteiras com mastite bovina foram estudadas. Este tipo de doença está associada com diversos patógenos, dentre eles *Corynebacterium bovis*. Neste estudo, os percentuais de isolados de *Corynebacterium* spp. foi identificada em 8,3% dos animais com mastite clínica e 4,8% com mastite subclínica (Oliveira *et al*., 2011).

Diante deste contexto, faz-se necessário o estudo de várias linhagens deste organismo a fim de compreender a estrutura genômica, organização gênica e os eventos e/ou mecanismos que contribuem para a evolução desta bactéria. Este estudo poderá corroborar para um melhor esclarecimento entre as diferenças e similaridades entre o repertório gênico dos biovares *ovis* e *equi*. No presente trabalho uma linhagem de *C. pseudotuberculosis* do biovar *ovis* (226) foi comparada com outras duas linhagens do mesmo biovar e duas outras linhagens do biovar *equi*, tornando este estudo pioneiro.

2. OBJETIVOS

2.1. OBJETIVO GERAL

- ✓ Realizar análises comparativas da linhagem 226 de *C. pseudotuberculosis* biovar *ovis* isolada de caprino.

2.2. OBJETIVOS ESPECÍFICOS

- ✓ Realizar a montagem da linhagem 226 (*ovis*) de *C. pseudotuberculosis*;
- ✓ Realizar a anotação dos elementos estruturais como CDS, tRNA e rRNA e classificar as proteínas de acordo com o Gene Ontology;
- ✓ Realizar a análise filogenômica da linhagem 226 com as demais linhagens de *C. pseudotuberculosis*;
- ✓ Avaliar a ordem gênica entre *C. pseudotuberculosis* 226 e duas linhagens do biovar *ovis* e duas do biovar *equi*;
- ✓ Identificar os genes ortólogos e únicos entre *C. pseudotuberculosis* 226 e duas linhagens do biovar *ovis* e duas do biovar *equi*.
- ✓ Identificar as prováveis regiões de Ilhas de Patogenicidade em *C. pseudotuberculosis* 226.

3. MATERIAIS E MÉTODOS

3.1. OBTENÇÃO DA LINHAGEM 226

A *Corynebacterium pseudotuberculosis* linhagem 226 foi adquirida através da parceria entre a Universidade da Califórnia, USA, e o Laboratório de Genética Celular e Molecular (LGCM). Foi isolada a partir de abscessos de um caprino e identificada como nitrato redução negativa (*ovis*).

3.2. CRESCIMENTO BACTERIANO DA LINHAGEM 226 E EXTRAÇÃO DO DNA GENÔMICO

Esta linhagem foi acondicionada em meio BHI (Brain Heart Infusion) semi-sólido. Posteriormente, foi diluído 37 g meio BHI líquido (Himedia) em 1L de água MilliQ e acrescidos Tween® 80 a uma concentração de 0,05%. Para o inóculo foram utilizados 15 ml de meio com a amostra e colocadas sob agitação de 120 rotações por minuto (rpm) à 37°C durante 48 horas. Após o crescimento foram retirados 2 ml do inóculo e centrifugado a 130 rpm por 2 minutos para formação do *pellet,* o qual foi utilizado na etapa de extração de DNA. A extração de DNA foi de acordo com o do kit DNeasy® Blood & Tissue (Quiagen), utilizando o protocolo para cultura de células.

3.3. CONSTRUÇÃO DA BIBLIOTECA PARA O SEQUENCIAMENTO

A preparação da biblioteca de *C. pseudotuberculosis* linhagem 226 para o sequenciamento foi de acordo com o protocolo disponibilizado para *ION Xpress Plus gDNA Fragment Library Preparation* (*Life Tecnologies*), utilizando o kit *ION Xpress™ Plus fragment Library*. A fragmentação do DNA foi realizada pela enzima ION Shear Enzyme Mix II, gerando fragmentos na faixa de 400 pb. Posteriormente, os fragmentos foram purificados com *Agencourt AMPure XP reagent* (Beckman). Em seguida, a deposição foi realizada em chip de 318v2 de acordo com o

protocolo ION PGM™ Sequencing 400 kit e o sequenciamento por meio da plataforma Ion Torrent PGM™.

3.4. MONTAGEM DO GENOMA

3.4.1. Avaliação de qualidade e tratamento das leituras

A avaliação de qualidade dos dados brutos foi realizada utilizando o *software* FastQC (*http://www.bioinformatics.babraham.ac.uk/*). Este permite uma análise gráfica do valor de qualidade presente em cada base das leituras, auxiliando assim no processo de tomada de decisão para o tratamento dos dados, antes do processo de montagem, utilizando como entrada para análise um arquivo no formato fastq contendo as leituras oriundas do sequenciamento.

3.4.2. Montagem *de novo* e fechamento de *gaps*

O processo de montagem foi realizado com o montador Mira versão 4.0 (Chevreux *et al*., 2004), sendo este executado de acordo com os parâmetros para dados produzidos pela plataforma Ion Torrent PGM™, estes foram relacionados ao arquivo manifest. Para esta linhagem foram utilizado o parâmetro **readgroup**, sendo utilizado **fragment** que indica o tipo de biblioteca.

Os *contigs* gerados pelo programa Mira foram utilizados como entrada para o *software* Lasergene (www.dnastar.com) para a obtenção do *scaffold*, utilizando-se como referência à linhagem 42/02-A de *C. pseudotuberculosis*, selecionada nos resultados de alinhamentos por meio de *blast* local. Contudo, o *scaffold* gerado continha N´s (regiões de *gaps*), os quais foram fechados utilizando os *softwares* Artemis (Rutherford *et al*., 2000), para edição do *scaffold* e CLC *workbenk* (www.clcbio.com) na geração de mapeamentos contra os dados brutos.

3.5. ANOTAÇÃO FUNCIONAL

A anotação do genoma foi realizada em duas etapas: automática e manual. Neste trabalho, a etapa de anotação automática foi feita pelo sistema *web* RAST, utilizando-se o RAST clássico com as opções de correção de erros automáticos, correção de *frameshifts* e aviso caso exista

algum erro durante o processo de anotação. Pelo RAST foi possível também visualizar a distribuição de subsistemas por categoria, quantidade de regiões codificantes e quantidade de RNAs (Aziz *et al*., 2008).

A segunda etapa foi realizada pela ferramenta Artemis (Rutherford *et al*., 2000), que utiliza um arquivo de entrada no formato *embl*, gerado a partir da anotação automática. Neste programa, foi realizada a curadoria manual das CDS's, através de busca de similaridade em nível proteico utilizando o algoritmo Blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi), adotando o banco de dados não redundante (nr). Além disso, foram usados bancos de dados como: UniProt (http://www.uniprot.org/) para identificar sigla de genes, InterProscan (Zdobnov & Apweiler, 2001) e Pfam (http://pfam.xfam.org/) para identificar motivos e domínios proteicos conservados. Para a anotação funcional foi utilizada a base de dados do GO usando o programa Blast2GO (Conesa *et al*., 2005), onde foi possível identificar os principais processos biológicos e funções moleculares presentes em *C. pseudotuberculosis* linhagem 226, adotando-se o nível 3 de classificação.

Além disso, o mapa genômico foi gerado através do programa CG*View* (Stothard & Wishart, 2005), tendo como entrada o arquivo *embl* que contem as CDS's, rRNA's e tRNA's.

3.6. ANÁLISE COMPARATIVA

3.6.1. Análise de filogenômica e sintenia da ordem gênica

O programa Gegenees foi utilizado para a análise de filogenômica e identificação de similaridade entre as linhagens, este programa realiza alinhamentos através de blastn entre os genomas, onde cada dado entre dois genomas é representado por uma pontuação, sendo esta representada através de uma matriz (Agren *et al*., 2012).

A analise de conservação da ordem gênica foi realizada por meio do programa Gepard (Krumsiek *et al*., 2007), o qual utiliza como entrada arquivos no formato fasta, resultando em um gráfico de sintenia (*dotplot*), que permite identificar regiões que apresentaram alta conservação da ordem gênica. As regiões do gráfico onde observou-se quebra de sintenia foram analisadas pela ferramenta Artemis *Comparison Tool* (ACT) (Carver *et al*., 2008) a fim visualizar as regiões

novas entre os genomas, rearranjos, translocações, deleções e inserções, usando-se como arquivo de entrada o *embl* e o resultado de blast local entre os dois genomas comparados.

3.6.2. Identificação de genes Ortólogos

O *software* PGAP (Zhao *et al*., 2012) foi adotado para a identificação dos genes ortólogos e únicos, onde foram fornecidos três arquivos de entrada:.nuc, .pep e .function. Estes foram gerados utilizando o programa Artemis (Rutherford *et al*., 2000) a partir do arquivo *embl*. Para a execução do programa PGAP foram adotados os parâmetros de *e-value* 0.00001, identidade de 80 e 90%, e cobertura de 90%. Para esta etapa foram utilizados os genomas completos disponíveis na base de dados de domínio público, conforme a Tabela 1.

Tabela 1. Genomas completos depositados de *C. pseudotuberculosis* no NCBI. *Biovar não informado. Fonte: NCBI.

Linhagem	Biovar	País de Origem	Hospedeiro	Número de Acesso
C231	*ovis*	Brasil	Ovino	CP001829
1002	*ovis*	Brasil	Caprino	CP001809
I19	*ovis*	Israel	Bovino	CP002251
PAT10	*ovis*	Argentina	Ovino	CP002924
42/02-A	*ovis*	Austrália	Ovino	CP003062
3/99-5	*ovis*	Escócia	Ovino	CP003152
FRC41	*ovis*	França	Humano	CP002097
267	*ovis*	Estados Unidos	Camelídeo	CP003407
P54B96	*ovis*	África do Sul	Bovino	CP003385
VD57	*ovis*	Brasil	Caprino	CP009927
CIP 52.97	*equi*	Kénia	Equino	CP003061
1/06-A	*equi*	Califórnia	Equino	CP003082
316	*equi*	Califórnia	Equino	CP003077
31	*equi*	Egito	Bubalino	CP003421
258	*equi*	Bélgica	Equino	CP003540
Cp162	*equi*	Reino Unido	Camelídeo	CP003652
***48252**	-	Noruega	Humano	CP008922
***CS_10**	-	Noruega	Linhagem de laboratório	CP008923
***Ft_219367**	-	Noruega	Caprino	CP008924

3.6.3. Predição de Ilhas de Patogenicidade

Para a predição de ilhas de patogenicidade foi utilizado o programa GIPsy (Soares *et al*., 2012), onde foi possível identificar os genes presentes nestas regiões que possivelmente possam ter suas funções relacionadas à patogenicidade e virulência da bactéria.

4. RESULTADOS E DISCUSSÃO

4.1. MONTAGEM *DE NOVO* DO GENOMA 226

Diversos estudos demonstram a importância de realizar trimagem e filtros de qualidade antes de se iniciar o processo de montagem (Carneiro *et al*., 2012), entretanto, devido à alta qualidade dos dados produzidos pelo sequenciamento esta etapa não foi necessária, como podemos visualizar na figura 7, onde grande parte das leituras apresentaram valor de qualidade acima de *Phred* 20 (1 erro a cada 100 pb).

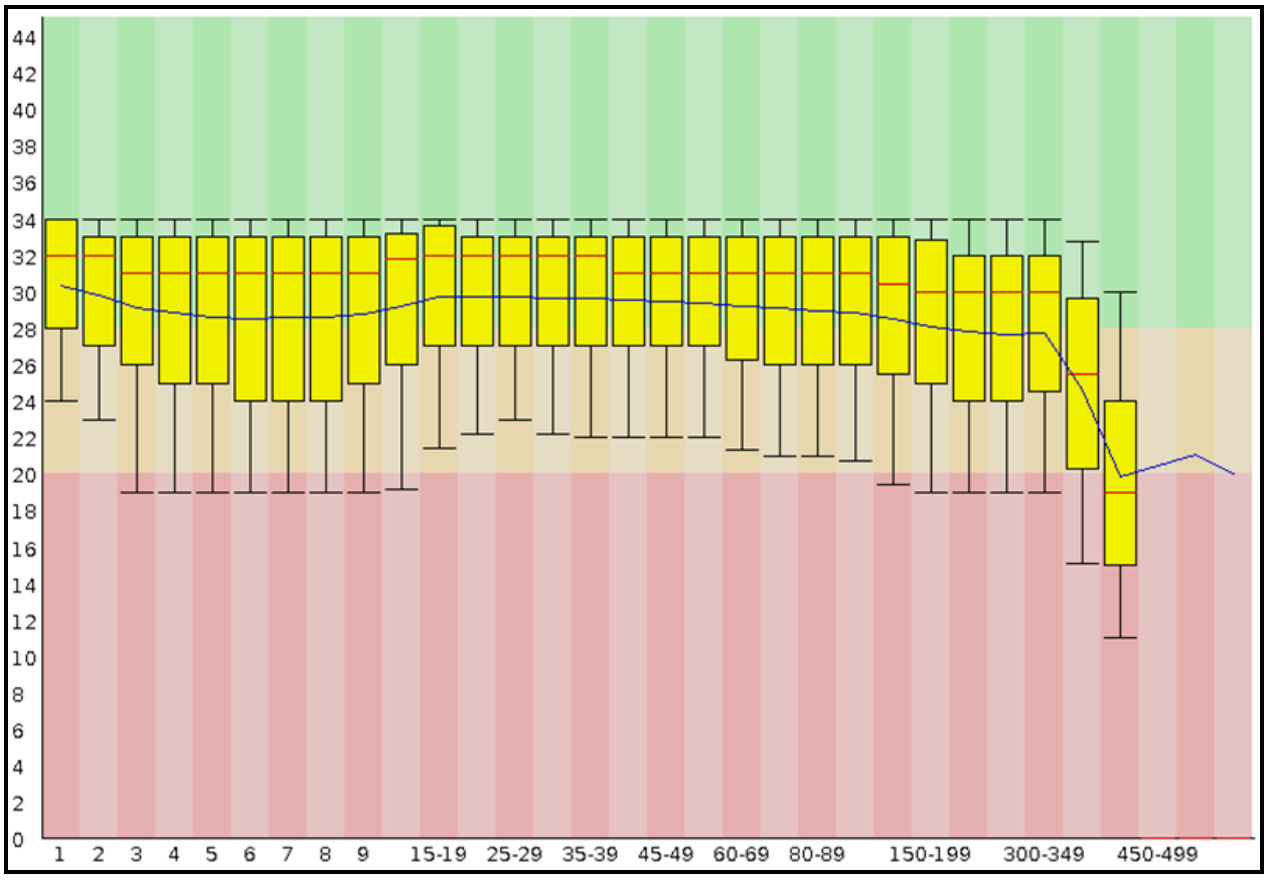


Figura 7. Avaliação de qualidade dos dados brutos gerada pelo programa FastQC; onde o eixo Y representa o valor de qualidade das bases, o Y o valor de qualidade *Phred*; a linha vermelha representa a mediana das leituras e a azul a média.

Após o sequenciamento foram obtidas 370.840.598 leituras e uma cobertura estimada de ~158x, sendo esta foi calculada com base na fórmula de cobertura estimada (Figura 8). Baseada no genoma de referência de 2,34 Mb da linhagem *C. pseudotuberculosis* 42/02-A, de acordo com o resultado de alinhamento local.

$$\text{Cobertura} = \frac{TL*QL}{TE}$$

Figura 8. Fórmula de cobertura estimada, onde: TL: tamanho de leituras; QL: quantidade de leituras e TE: tamanho estimado do genoma.

Após a montagem com o Mira foi obtido o resultado de acordo com a tabela 2.

Tabela 2. Resultado de montagem da linhagem 226 com o programa Mira.*N50: média estatística para avaliar um conjunto de *contigs* gerados por montagem *de novo*.

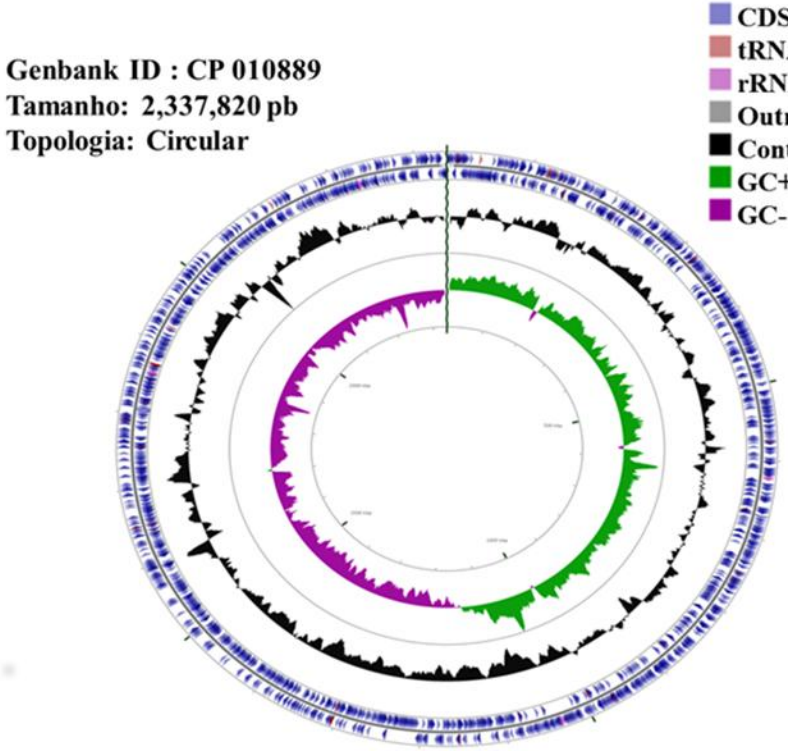
	Quantidade de bases	**Menor *contig***	**Maior *contig***	***N50**	**Total de *contigs***
Mira	2.358.218	357	824549	370.035	27

Após, os *contigs* foram submetidos ao programa Lasergene para geração do *scaffold* preliminar. Para a retirada de N's (regiões de *gaps*) (alinhando os dados brutos contra os *contigs*) e edição da fita genômica os programas Artemis e CLC foram utilizados. Dessa forma, o genoma de *C. pseudotuberculosis* linhagem 226 foi finalizada com 2.337.820 pb agrupadas em um único *contig*.

4.2. ANOTAÇÃO FUNCIONAL DO GENOMA DE *C. pseudotuberculosis* 226

A geração massiva de dados genômicos depositados em bancos de dados públicos, de maneira geral, contribuiu para diversos estudos na área científica. Entretanto, o aumento exponencial de genomas anotados de forma automática e sem revisão manual podem propagar erros relacionadas à nomenclatura de produtos e outros elementos estruturais do genoma (D'Afonseca *et al*., 2012; Stein, 2001).

A fim de minimizar a dispersão desses erros, após a predição automática utilizando a ferramenta *web* RAST (Aziz *et al*., 2008), uma edição manual dos elementos estruturais do genoma foi realizada no programa Artemis (Rutherford *et al*., 2000). Após esta revisão, o genoma de *C. pseudotuberculosis* 226 resultou em um total de 2.138 CDS, 4 *clusters* de rRNA (5S,16S e 23S), 49 de tRNA e 72 pseudogenes. As características estruturais, e o mapa genômico de *C. pseudotuberculosis* 226 podem ser visualizados na figura 9.



Características genômicas	*C.pseudotuberculosis 226*
GenBank ID	CP010889
Tamanho (pb)	2.337.820
% Conteúdo GC (genoma)	52,18%
% Codificante	84,7%
Genes	2.271
Proteínas	2.138
Operons de rRNA	4x(16S-23S-5S)
Número de tRNA	49
Pseudogenes	72
Transposases, Integrases e Proteínas de fagos	10

Figura 9. Mapa genômico gerado pelo programa CG*View* e características estruturais de *C. pseudotuberculosis* 266.

As 2.138 CDS preditas no genoma de *C. pseudotuberclosis* 226 foram classificadas em diversas funções e processos biológicos utilizando o terceiro nível de classificação baseadas no banco de dados de ontologia gênica (*Gene Ontology*).

Nos processos biológicos relacionados à resposta a estresses apresentou 66 genes envolvidos (Figura 10), sendo que estes podem estar associados à resistência da bactéria ao ambiente hostil do hospedeiro durante o processo infeccioso. Após a fagocitose, os macrófagos tornam o seu meio intracelular rapidamente ácido, este processo afeta de forma negativa o metabolismo do hospedeiro, além disso, outros estresses intracelulares também contribuem de maneira negativa, como estresse osmótico, térmico e nitrosativo (Pinto *et al.*, 2014). Como o gene *recF* que atua no reparo e replicação do DNA.

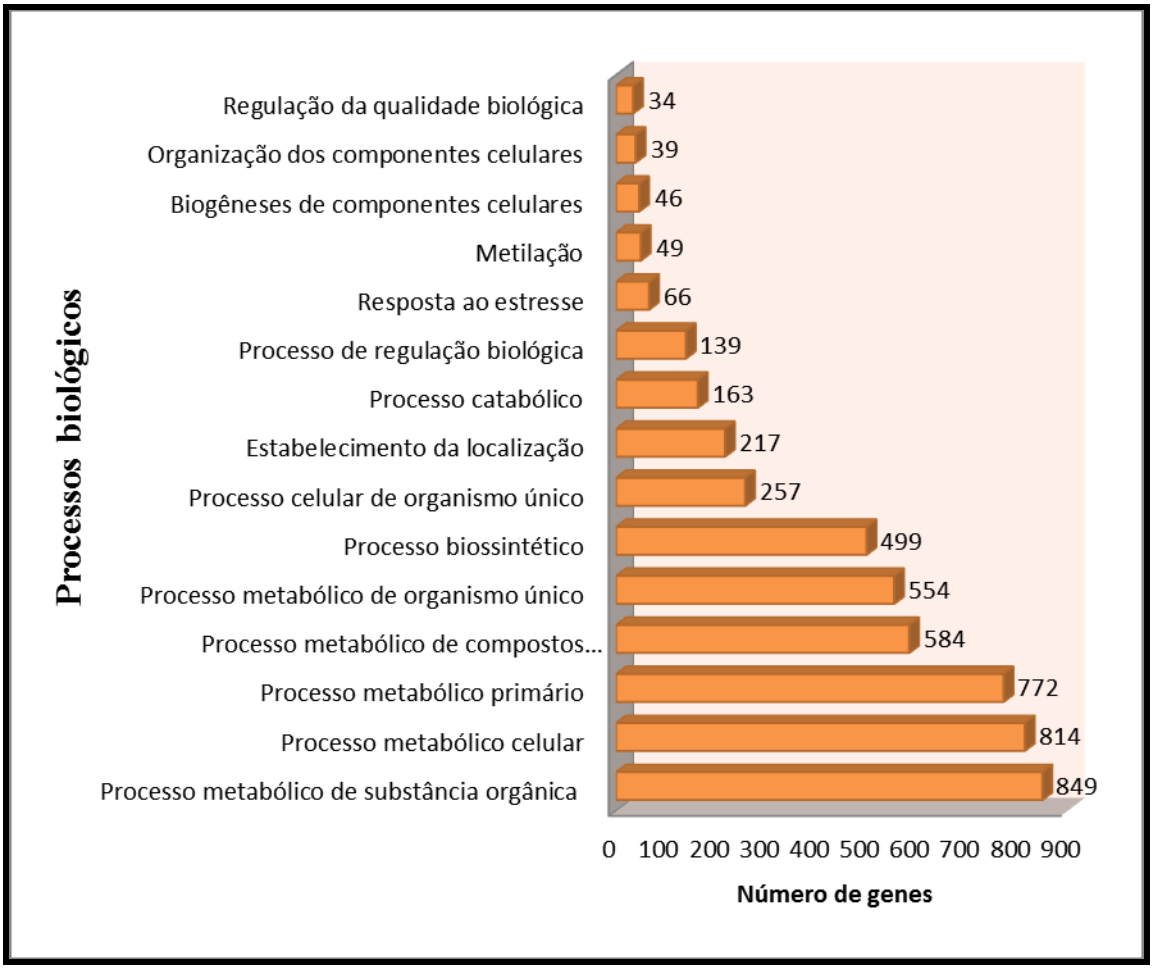


Figura 10. Classificação dos processos biológicos de *C. pseudotuberciulosis* 226 utilizando o terceiro nível baseado no *Gene Ontology*.

O gene *dtxR* (repressor da toxina diftérica), que atua na repressão da ligação de ferro que atua como um regulador global na expressão do gene *tox* (D'Afonseca *et al*., 2012), sendo este gene presente dentre as funções biológicas mais representativas com 549 genes relacionados a função de ligações de compostos orgânicos cíclicos (Figura 11).

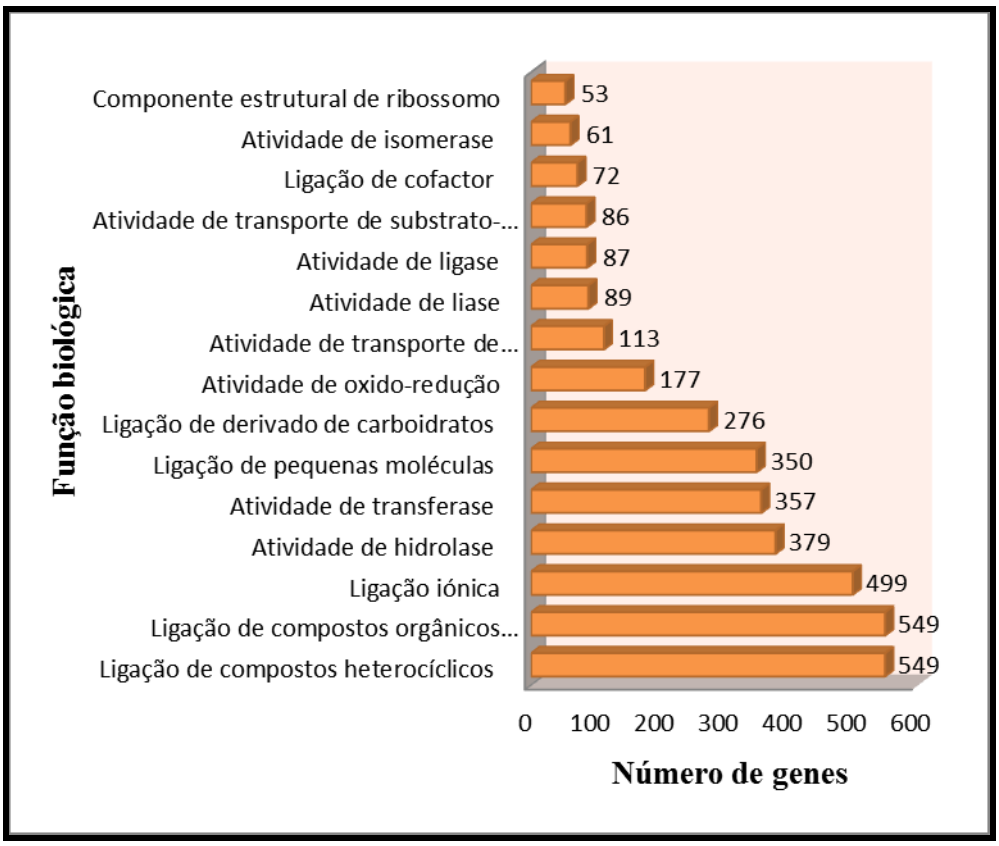


Figura 11. Classificação das funções biológicas de *C. pseudotuberculosis* 226 utilizando o terceiro nível baseado no *Gene Ontology*.

4.3. ANÁLISE FILOGENÔMICA E SINTENIA DA ORDEM GÊNICA

Através do programa Gegenees, comparou-se o genoma da Cp226 com outras 19 linhagens de biovar *ovis* e *equi* disponíveis no NCBI e o conteúdo variável entre estes genomas estão representados em percentuais de similaridade em uma matriz (*Heatmap*) (Figura 12). Além disto, os percentuais de similaridade foram usados para gerar a árvore filogenômica, onde pode-

se observar que todas as linhagens pertencentes a cada biovar formaram dois agrupamentos distintos (Figura 13).

As linhagens do biotipo *ovis* apresentam alta similaridade (acima de 90%) e ao compararmos a Cp226 com 10 linhagens deste biovar, a I19 apresentou maior similaridade (99,91%), a qual foi isolada a partir de um hospedeiro bovino em Israel (Silva *et al*., 2011), enquanto a que obteve menor similaridade foi a CpC231 (99,45%), entretanto, a Cp267 foi a escolhida para as outras análises deste trabalho devido ter a mesma localização geográfica de Cp226, pois foi isolada de uma lhama com abscessos na Califórnia, USA (Lopes *et al*., 2012). Outro resultado importante foi em relação as linhagens CS_10, Ft_219367 e 48252, que apesar de não terem sido informadas quanto ao tipo de biovar agruparam-se no mesmo clado das linhagens do biotipo *ovis*, sugerindo-se que estas sejam deste biovar.

Estes resultados corroboram com o estudo de Soares e colaboradores (2013), em que as linhagens do biotipo *ovis* apresentam um comportamento mais clonal, quando comparadas às linhagens do biotipo *equi* (Soares *et al*., 2013).

O genoma da Cp226 quando comparado com os das linhagens do biovar *equi,* observou-se que Cp162 e Cp1/06-A apresentam maior e menor similaridade com 93,65% e 93,37%, respectivamente. Entretanto, as análises de sintenia e de genes ortólogos selecionou-se a linhagem Cp258, que foi a segunda com maior valor de percentual de similaridade (93,63%), devido está linhagem ter sido resequenciada e reanotada de forma bem criteriosa, o que permitiu identificar genes responsáveis pela diferenciação entre o biovar *ovis* e *equi* de *C. pseudotuberculosis*, como os genes *nar,* responsáveis pela redução de nitrato redutase, os quais estão organizados em operon *narI, narJ* e *narH* e não tinham sido anotados nos genomas disponíveis na base de dados do NCBI.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
C. diphtheriae 241	100.0	99.98	3.22	3.22	3.36	3.23	3.21	3.21	3.2	3.21	3.21	3.21	3.21	3.21	3.2	3.2	3.2	3.21	3.21	3.21	3.2	3.19
C. diphtheriae HC01	99.98	100.0	3.24	3.26	3.38	3.27	3.25	3.26	3.26	3.26	3.26	3.26	3.26	3.26	3.25	3.26	3.27	3.27	3.27	3.27	3.26	3.25
C. pseudotuberculosis 31	3.14	3.15	100.0	98.17	98.73	98.25	97.51	95.54	94.92	94.98	95.09	95.08	95.08	95.1	95.09	95.13	95.07	95.09	95.11	95.1	94.69	94.85
C. pseudotuberculosis CIP 52.97	3.17	3.18	97.52	100.0	99.24	98.27	97.27	95.32	94.33	94.36	94.5	94.49	94.49	94.51	94.49	94.53	94.48	94.5	94.51	94.49	94.07	94.27
C. pseudotuberculosis 258	3.21	3.21	95.89	97.03	100.0	96.82	95.2	93.53	92.23	92.26	92.39	92.38	92.38	92.4	92.36	92.39	92.38	92.38	92.37	92.37	92.02	92.15
C. pseudotuberculosis 316	3.22	3.23	97.83	98.43	99.3	100.0	97.57	96.04	94.53	94.57	94.69	94.7	94.7	94.71	94.67	94.71	94.7	94.69	94.69	94.69	94.32	94.46
C. pseudotuberculosis 1/06-A	3.3	3.3	98.22	98.61	98.75	98.78	100.0	95.92	95.47	95.53	95.66	95.64	95.64	95.66	95.68	95.7	95.62	95.64	95.69	95.66	95.24	95.42
C. pseudotuberculosis 162	3.22	3.23	95.71	96.11	96.52	96.68	95.39	100.0	95.22	95.26	95.39	95.39	95.39	95.41	95.36	95.4	95.39	95.39	95.38	95.37	94.99	95.16
C. pseudotuberculosis PAT10	3.17	3.18	93.46	93.5	93.57	93.5	93.31	93.57	100.0	99.64	99.69	99.68	99.68	99.69	99.68	99.71	99.79	99.81	99.8	99.78	99.36	99.55
C. pseudotuberculosis 1002	3.17	3.17	93.53	93.54	93.6	93.54	93.37	93.61	99.65	100.0	99.92	99.81	99.81	99.82	99.78	99.82	99.71	99.73	99.73	99.71	99.32	99.48
C. pseudotuberculosis VD57	3.2	3.21	93.53	93.6	93.66	93.6	93.41	93.67	99.59	99.83	100.0	99.85	99.85	99.87	99.83	99.87	99.76	99.78	99.78	99.76	99.31	99.53
C. pseudotuberculosis 48252	3.18	3.18	93.54	93.58	93.64	93.58	93.37	93.66	99.59	99.72	99.85	100.0	99.99	99.93	99.82	99.87	99.76	99.78	99.77	99.75	99.3	99.52
C. pseudotuberculosis CS_10	3.13	3.14	93.53	93.55	93.59	93.56	93.4	93.63	99.58	99.71	99.84	99.99	100.0	99.92	99.82	99.86	99.75	99.76	99.77	99.75	99.3	99.53
C. pseudotuberculosis Ft_219367	3.11	3.12	93.56	93.59	93.68	93.62	93.42	93.69	99.62	99.74	99.86	99.92	99.92	100.0	99.85	99.88	99.78	99.79	99.8	99.78	99.33	99.54
C. pseudotuberculosis 3/99-5	3.2	3.21	93.53	93.59	93.59	93.57	93.44	93.63	99.57	99.67	99.82	99.81	99.82	99.82	100.0	99.9	99.75	99.75	99.85	99.77	99.31	99.52
C. pseudotuberculosis FRC41	3.17	3.17	93.57	93.62	93.65	93.6	93.46	93.66	99.6	99.72	99.86	99.86	99.86	99.86	99.9	100.0	99.76	99.79	99.84	99.8	99.3	99.56
C. pseudotuberculosis 226	3.23	3.24	93.51	93.55	93.63	93.57	93.37	93.65	99.69	99.61	99.76	99.75	99.75	99.76	99.75	99.76	100.0	99.9	99.85	99.86	99.45	99.66
C. pseudotuberculosis I19	3.19	3.2	93.54	93.59	93.65	93.58	93.4	93.66	99.72	99.64	99.79	99.77	99.77	99.79	99.77	99.8	99.91	100.0	99.88	99.89	99.43	99.66
C. pseudotuberculosis 42/02-A	3.2	3.2	93.56	93.6	93.62	93.59	93.45	93.67	99.71	99.64	99.78	99.77	99.77	99.79	99.85	99.86	99.86	99.87	100.0	99.89	99.46	99.63
C. pseudotuberculosis P54B96	3.22	3.22	93.54	93.58	93.63	93.59	93.42	93.65	99.69	99.62	99.76	99.74	99.74	99.76	99.78	99.81	99.88	99.89	99.89	100.0	99.41	99.64
C. pseudotuberculosis C231	3.25	3.25	93.53	93.55	93.66	93.59	93.4	93.66	99.69	99.64	99.73	99.73	99.73	99.74	99.73	99.73	99.87	99.85	99.87	99.82	100.0	99.57
C. pseudotuberculosis 267	3.17	3.18	93.29	93.32	93.4	93.35	93.16	93.42	99.44	99.37	99.52	99.51	99.51	99.52	99.52	99.56	99.67	99.65	99.63	99.64	99.18	100.0

Figura 12. *Heatmap* entre 20 linhagens de *C. pseudotuberculosis* e 2 de *C. diphtheriae*. Os números em vermelho indicam baixa similaridade e verde alta similaridade.

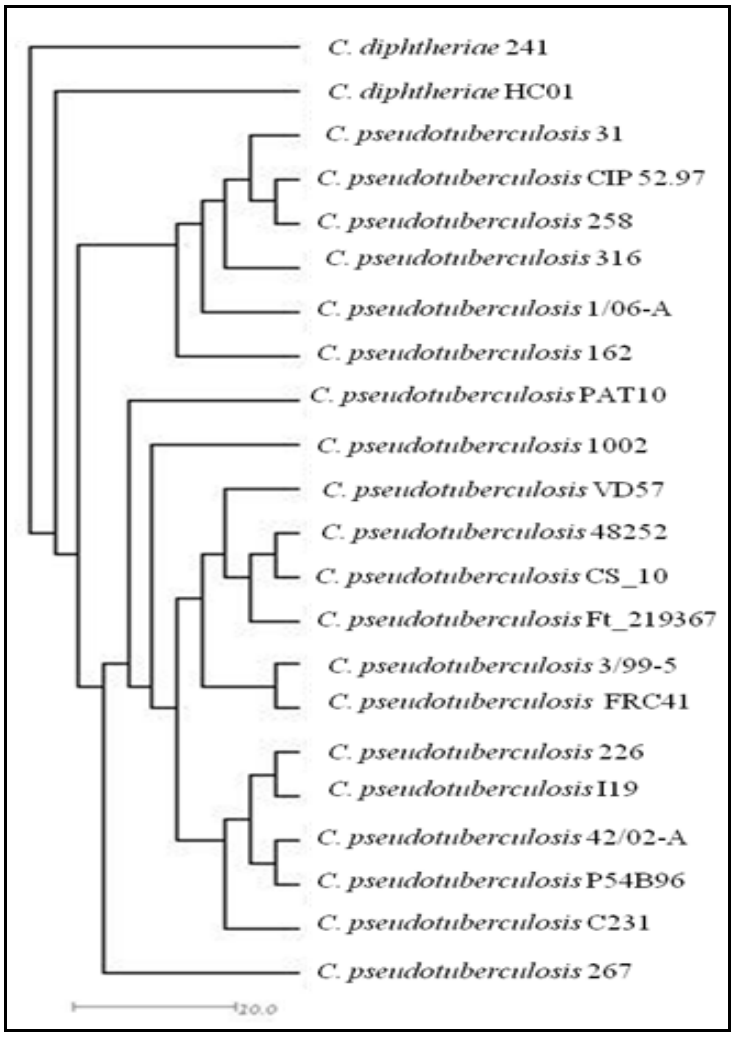


Figura 13. Árvore filogenômica entre as 20 linhagens de *C. pseudotuberculosis* e 2 de *C. diphtheriae*, gerada pelo programa SpliTree a partir do resultado do programa Gegenees.

As duas linhagens escolhidas do biovar *ovis* de acordo com o *heatmap* foram comparadas em relação à sintenia da ordem gênica e os resultados podem ser visualizados na Figura 13.

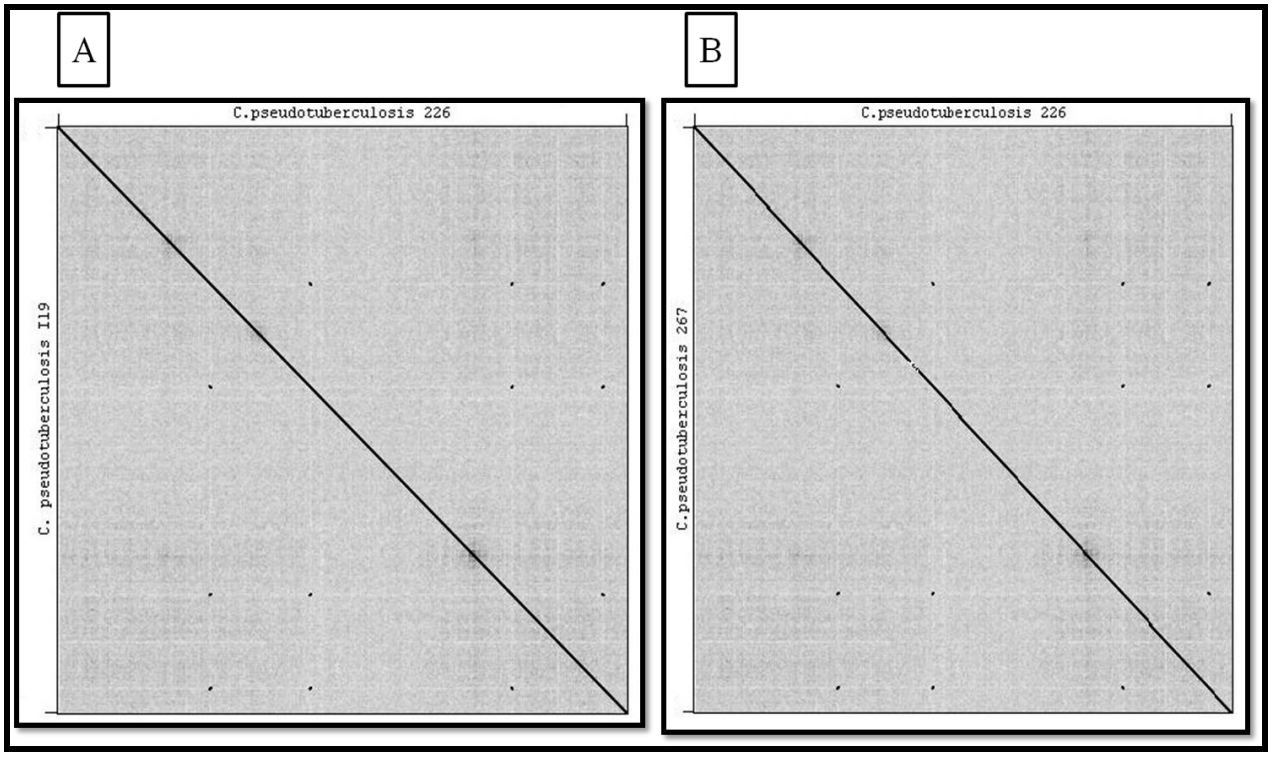


Figura 14. *Dotplot* de sintenia gerado pelo programa Gepard entre as linhagens do biovar *ovis* com Cp226. **A**, análise de sintenia entre a linhagem I19 e 226 de *C. pseudotuberculosis*; **B**, análise de sintenia entre as linhagens 267 e 226 de *C. pseudotuberculosis*.

Na análise de sintenia da linhagem 226 com as linhagens do biovar *equi* Cp1/06-A (Figura 14A) e Cp258 (Figura 14B), embora também apresentem a ordem gênica altamente conservada identificaram-se regiões de quebra, muito provavelmente devido à presença/ausência de genes que podem estar relacionadas à diferença entre os biotipos.

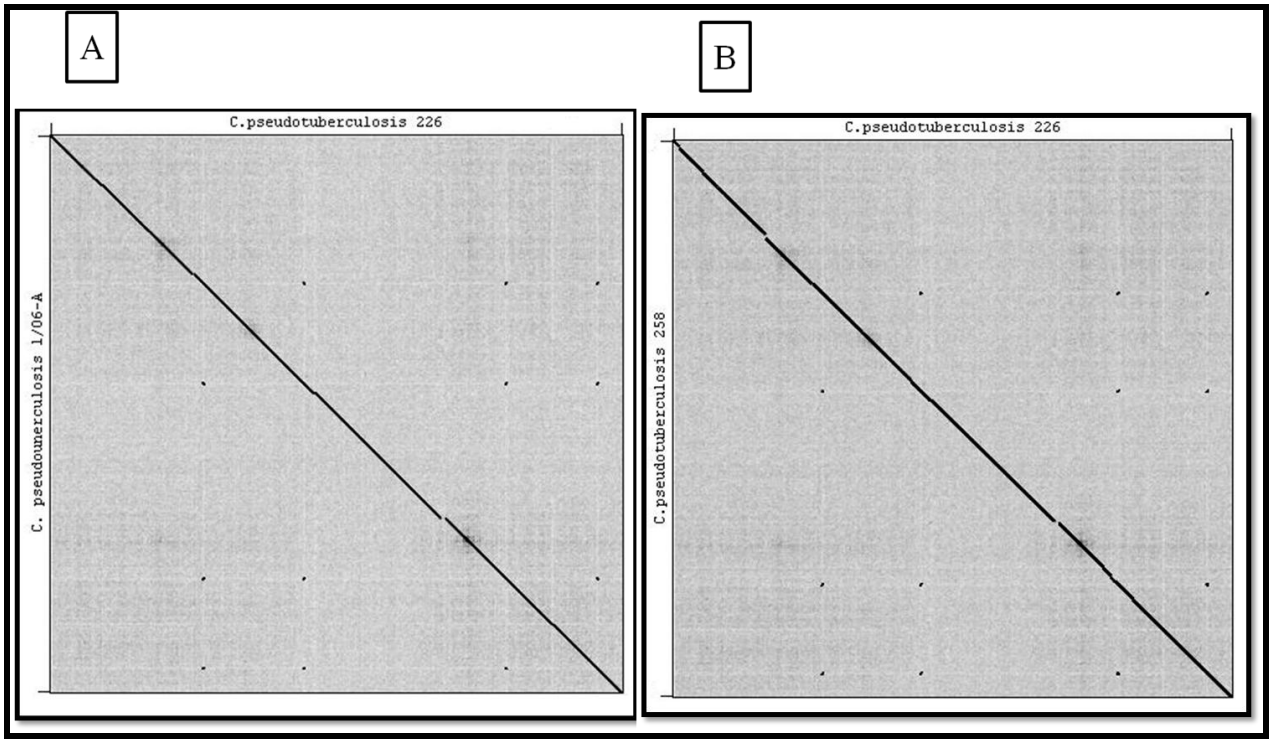


Figura 15.*Dotplot* de sintenia gerado pelo programa Gepard entre as linhagens do biovar *equi* com Cp226. **A**, análise de sintenia entre a linhagem 1/06-A e 226 de *C. pseudotuberculosis*; **B**, análise de sintenia entre as linhagens 258 e 226 de *C. pseudotuberculosis*.

Além disso, visando identificar as regiões do genoma que apresentam rearranjos ou inversões, o alinhamento múltiplo entre as sequências genômicas de todas as linhagens do biovar *ovis*, incluindo a Cp226 realizado no programa Mauve permitiu identificar que a linhagem 1002 de *C. pseudotuberculosis* é a única que possui rearranjo (Figura 15) e esta será re-sequenciada e re-montanda para correção desta região, visto que estes resultados corroboram com os dados de mapa ótico (Dados gerados no Laboratório de Genética Celular e Molecular-LGCM da Universidade Federal de Minas Gerais), uma tecnologia Argus da OpGen de alta resolução que ordena mapas de restrição do genoma inteiro (http://opgen.com).

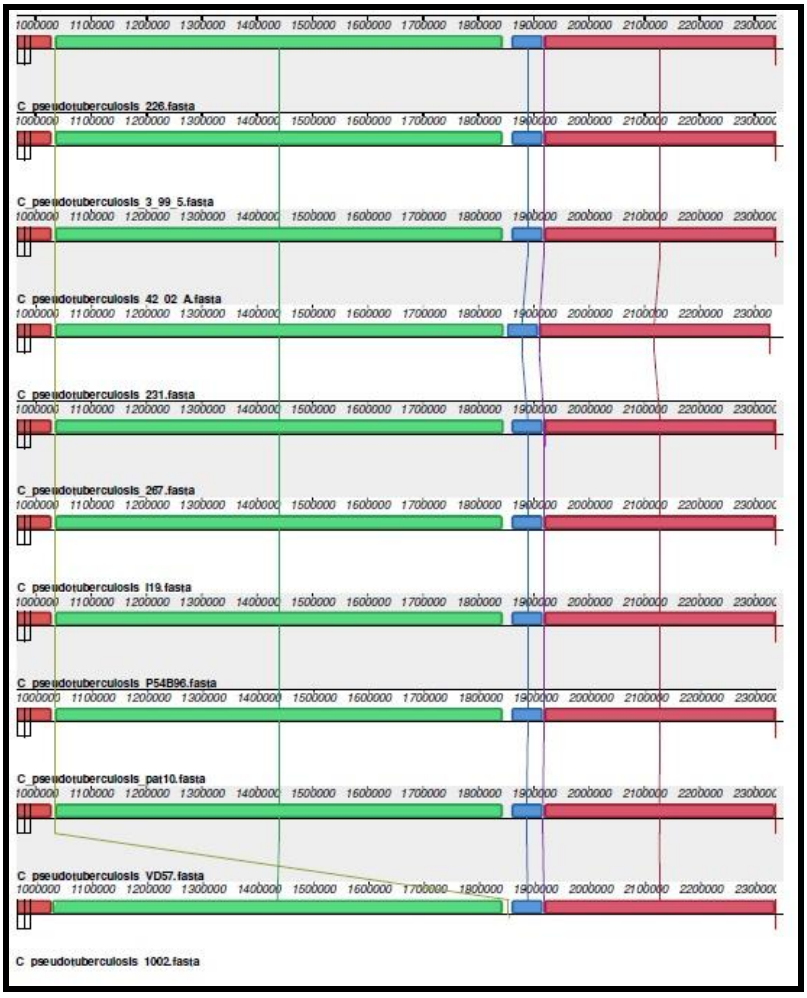


Figura 16. Alinhamento feito pelo programa Mauve entre as linhagens *ovis* e a linhagem 226, demonstrando uma região de bloco gênico com rearranjos na linhagem 1002.

Nos alinhamentos realizados entre os biovares *equi* e a linhagem 226 não foram observados rearranjos de blocos gênicos (Figura 16), com exceção da linhagem CIP52.97.

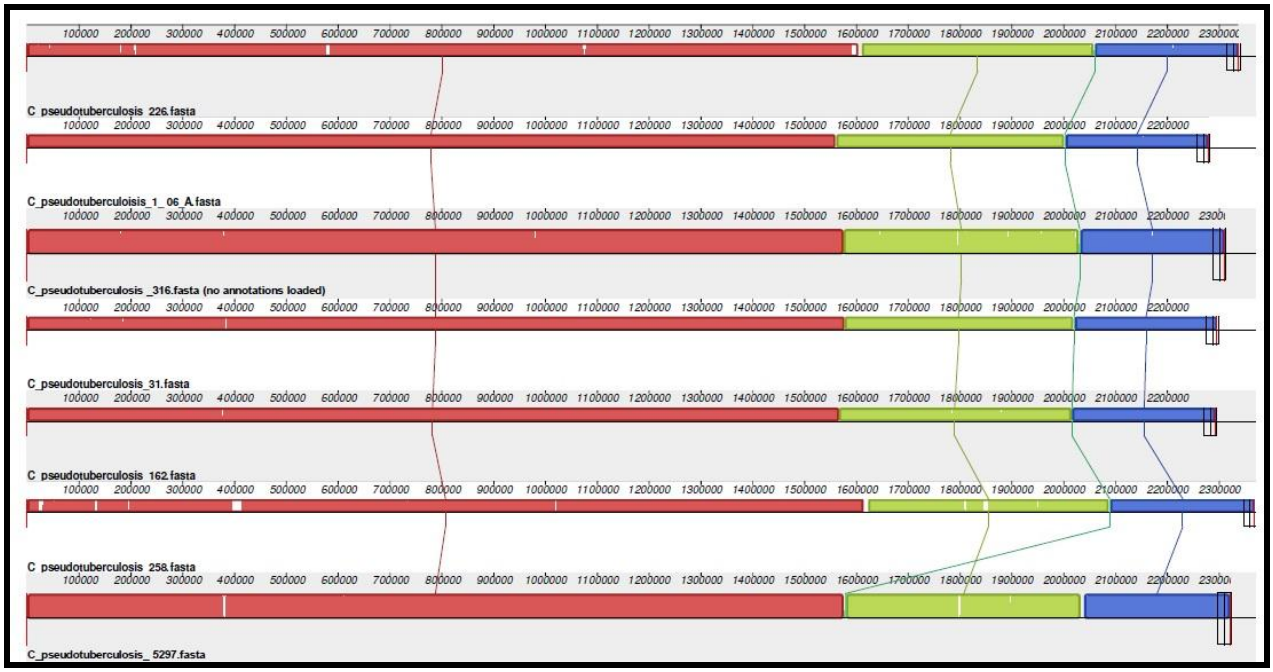


Figura 17. Alinhamento feito pelo programa Mauve entre os biovares e*qui* e a linhagem 226, demonstrando uma região de bloco gênico com rearranjos na linhagem CIP 52.97.

Na análise utilizando a ferramenta ACT, não foi identificada nenhuma região nova entre as linhagens do biotipo *ovis* quando comparadas à Cp226, entretanto, identificou uma inversão no gene *ilvB* (Figura 17), o qual em Cp226 aparece na *frame* de leitura -1 e na Cp267 o mesmo aparece na *frame* de leitura +3. Este gene codifica a acetocolase sintase em *E. coli* e faz parte de um operon que atua na regulação de promotores no processo de biossíntese de valina e isoleucina (Friden *et al*., 1982). Esta região necessita de uma nova análise para confirmar se este evento de inversão é verdadeiro ou se ocorreu um erro relacionado à montagem de Cp267.

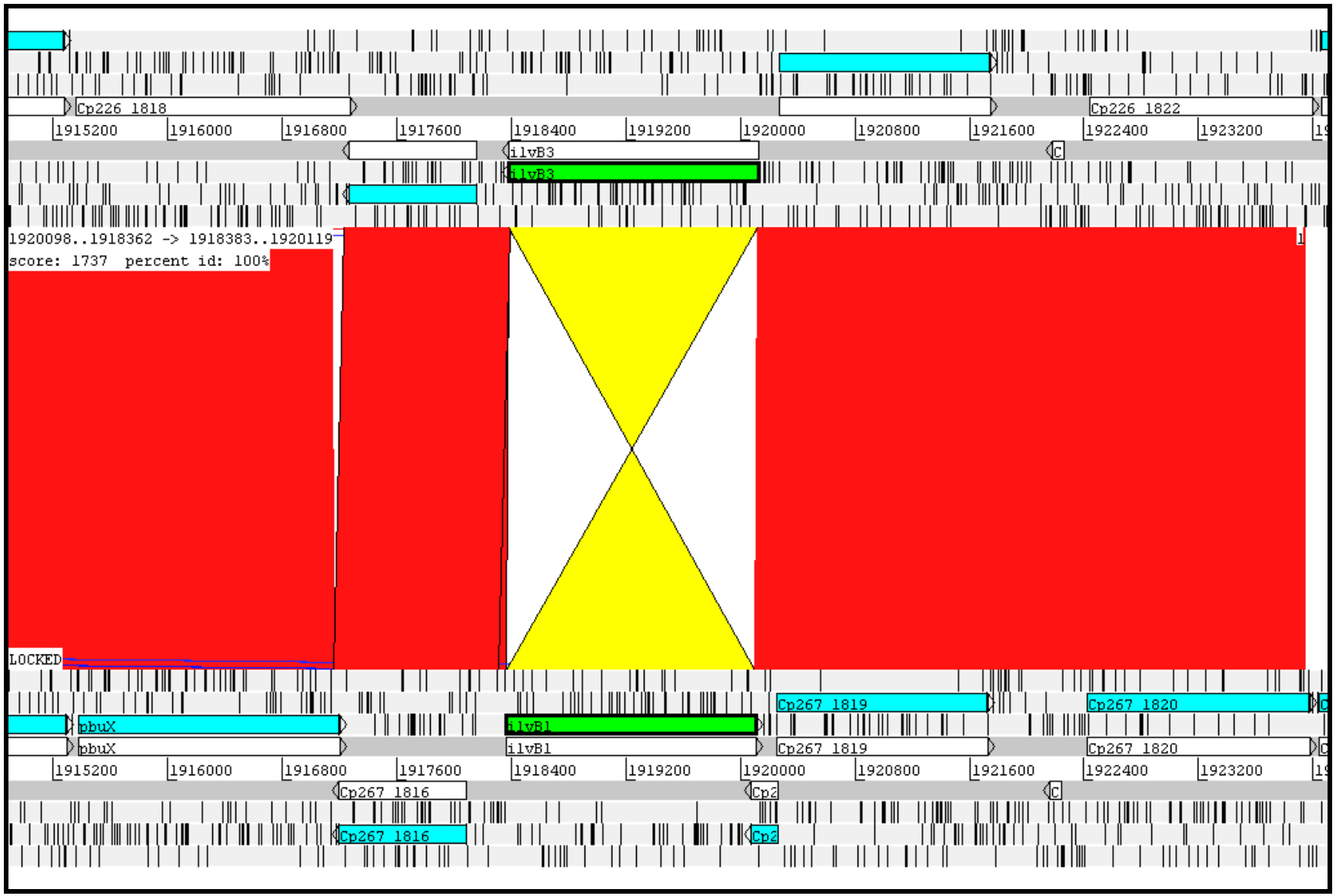


Figura 18. Imagem gerada pelo programa Artemis ACT; região de inversão do gene *ilvB* encontrada na linhagem 226 (-1) e Cp267(+3).

4.4. IDENTIFICAÇÃO DE GENES ORTÓLOGOS E ÚNICOS

Na identificação de genes ortólogos entre as linhagens Cp226, CpI19 e Cp267 pertencentes ao biovar *ovis* foram compartilhados 1749 genes, conforme ilustrado na figura 18.

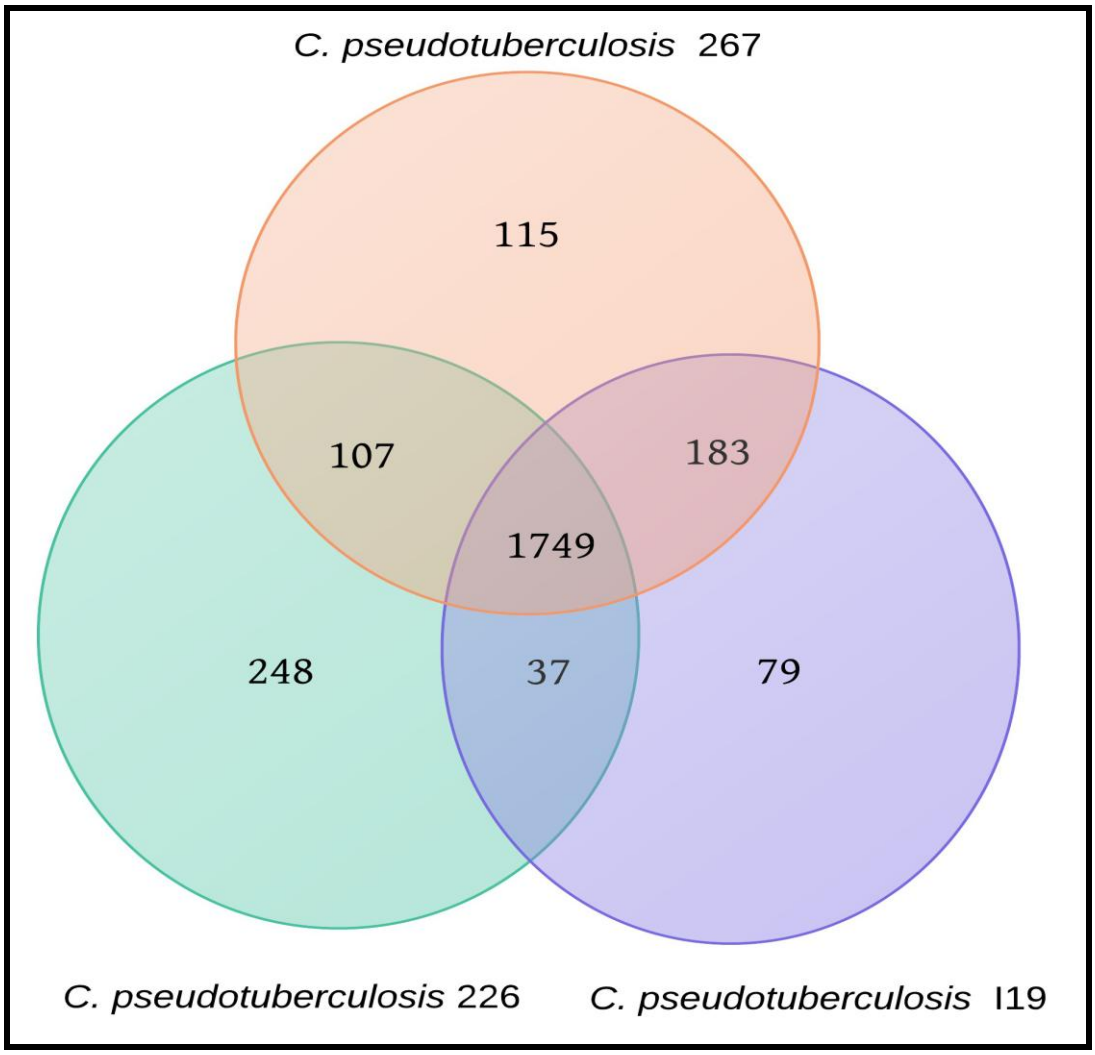


Figura 19. Gráfico de Venn entre as linhagens I19 e 267 (*ovis*) com a linhagem 226.

Destaca-se que 248 genes estão presentes somente em Cp226, dentre estes genes, foi identificado o gene *gnlfT1* que codifica a proteína UDP-transferase galactofuranose, que é um importante componente da parede celular de *C. pseudotuberculosis* e de espécies do grupo CMNR. Em ensaios de clonagem e caracterização deste gene em *Mycobacterium tuberculosis* foi possível revelar este como um possível alvo terapêutico em micobactérias (Rose *et al.*, 2006).

Entretanto seu papel em *C. pseudotuberculosis* ainda não foi descrito na literatura, necessitando de novos estudos que possam confirmar sua função na virulência deste patógeno.

Entre as linhagens do biovar *equi* Cp1/06-A e Cp258 quando comparadas a linhagem 226 foram compartilhados 1.592 genes (Figura 19). A linhagem 226 apresentou 282 genes únicos. Destes os genes Cp226_0206, Cp226_0207 e Cp226_0208 codificam produtos associados a famílias de elementos genéticos móveis, tais como transposons ou sequências de inserção que coficam a enzima transposase, a presença desses elementos únicos no genoma de *C. psedotuberculosis* 226 contribui para a aquisição de genes relacionados à virulência através de transferência horizontal de genes (Ruiz *et al*., 2011).

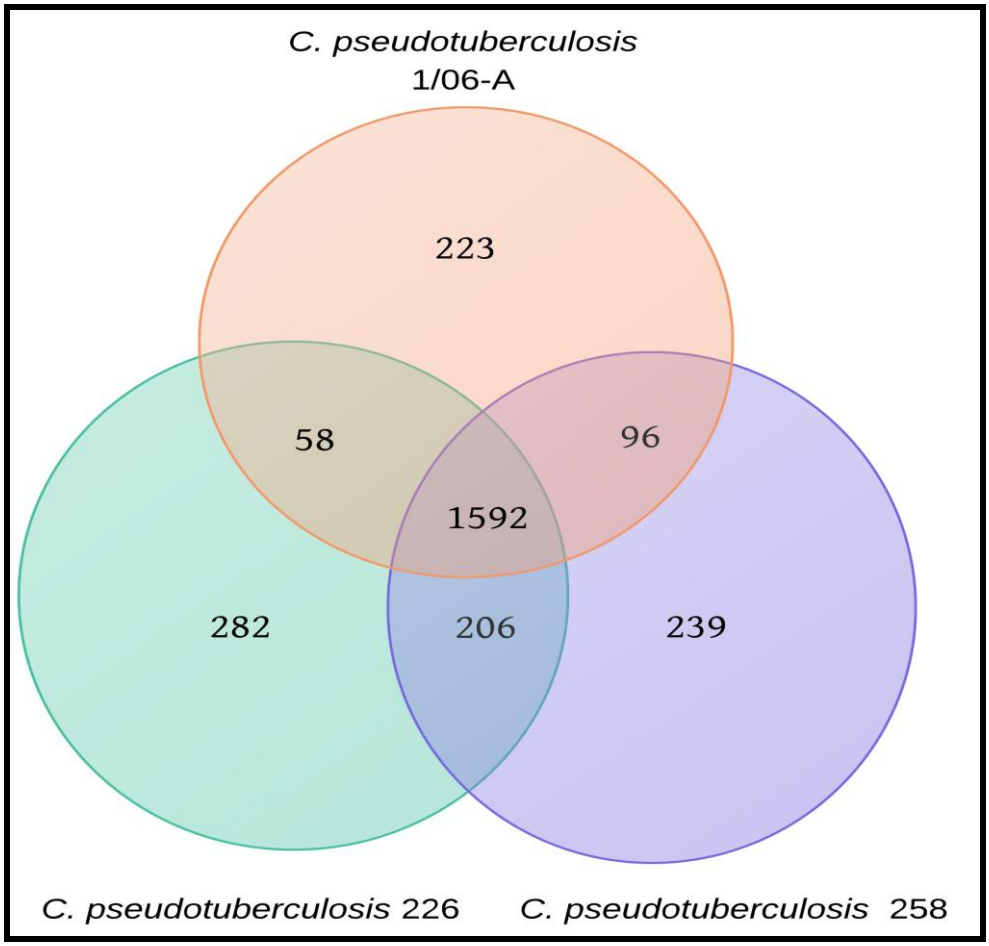


Figura 20. Gráfico de Venn entre as linhagens 1/06-A e 258 (*equi*) com a linhagem 226.

Diversos estudos demonstram as diferenças e similaridades entre os biovar *ovis* e *equi*. De acordo com Bolt, (2009), através de análises entre sequências conservadas de diferentes linhagens de *C. pseudotuberculosis* isoladas a partir de diferentes hospedeiros e localizações geográficas foi possível identificar que as linhagens do biotipo *ovis* apresentam mutações de ponto, em vez de recombinação. Este evento sugere a limitação quanto à evolução do repertório gênico do biovar *ovis*, enquanto que as cepas do biovar *equi* apresenta maior diversidade gênica. Desta forma, os resultados gerados em neste trabalho corroboram com os dados produzidos por (Bolt, 2009).

4.5. ANÁLISE DAS REGIÕES DE ILHAS DE PATOGENICIDADE

Em *C. pseudotuberculosis* 226 foram preditas 8 ilhas de patogenicidade, com um total de 102 genes (Figura 20). Na ilha 1 foram identificados 12 genes, dentre eles foi encontrado o *cas5*, gene associado a CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) e que confere resistência contra fagos e plasmídeos (Kunin *et al*., 2007).

Além disso, o gene *pld* que é codificador da proteína fosfolipase D também foi identificado nesta ilha. Ele é considerado um dos principais e bem descritos na literatura quanto ao seu papel na virulência de *C. pseudotuberculosis*, sua enzima catalisa a dissociação da esfingomielina e possibilita o aumento da permeabilidade vascular, isto permite a multiplicação e sobrevivência do patógeno nas células, ocasionando a difusão por macrófagos para os glânglios linfáticos (Hodgson *et al*., 1999;Baird & Fontaine, 2007).

Os genes *fagA* (proteína integral de membrana), *fagB* (transportador de enterobactina de ferro), *fagC* (ATP-proteína de ligação da membrana citoplasmática) e *fagD* (Sideróforos-proteína de ligação de ferro) também estão presentes flanqueando o gene *pld*. Este operon está relacionado à captação de ferro, e contribui na persistência na infecção de *C. pseudotuberculosis* em caprinos (Billington *et al*., 2002;Aquino de Sá *et al*., 2013). Estes genes relacionados à virulência de *C. pseudotuberculosis* situados em regiões de ilhas de patogenicidade corroboram com a aquisição dos mesmos através de transferência horizontal, que fornecem comprovações a respeito de seu estilo de vida e mecanismos de patogenicidade usadas pelo microorganismo durante o processo infeccioso (Ruiz *et al*., 2011).

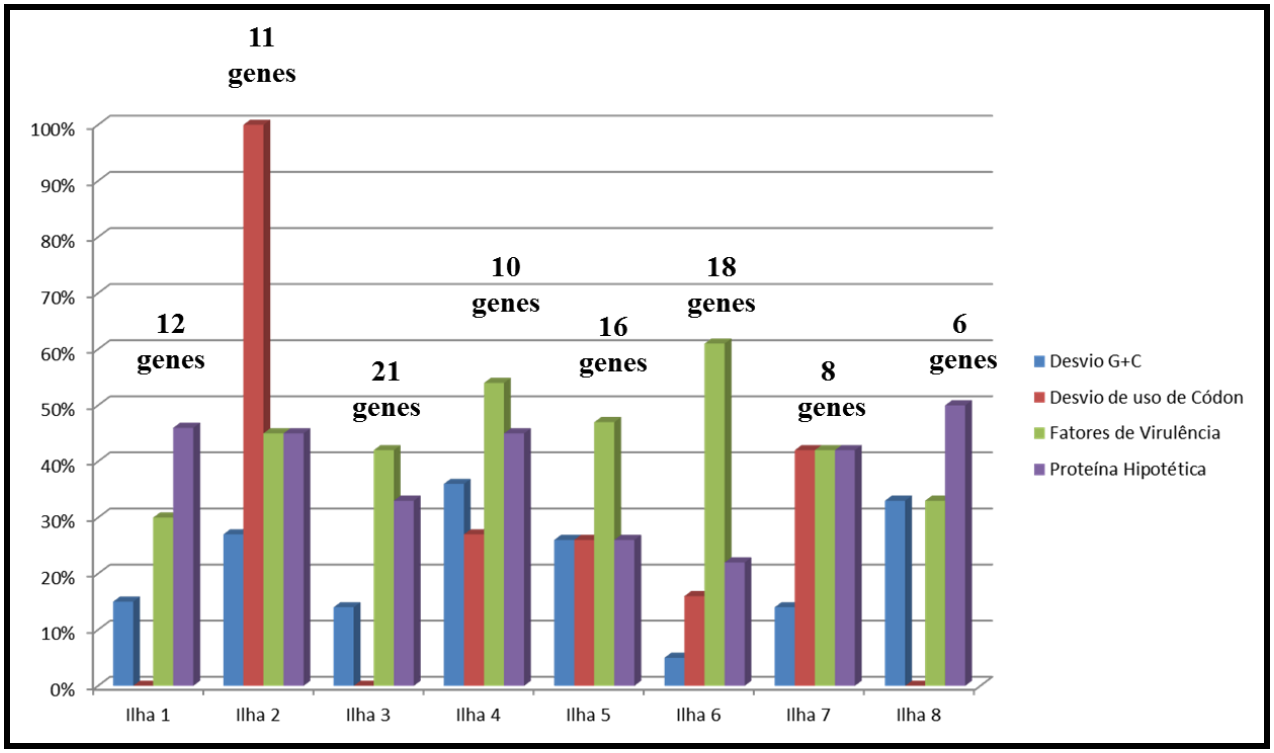


Figura 21. Ilhas de patogenicidade preditas em Cp226 pelo programa GIPSy; onde no eixo X está representado a porcentagem e no Y a as ilhas.

A ilha 2 (Apêndice) foi a única que apresentou maior força de predição com 27% de conteúdo G+C, 100% de desvio de uso de códon, 45% de fatores de virulência e 45% de proteínas hipotéticas. Nesta ilha foram identificados os genes *potA, potC, potD, glpt, mprA1, senX3A* e cinco proteínas hipotéticas. O operon *potA,C* e *D* tem função de transporte de poliaminas, estas são essenciais para o crescimento e sobrevivência das células (Stein & Finlay, 2006). Entretanto, estes genes não são descritos na literatura quanto a sua função em *C. pseudotuberculosis*, contudo devido a sua funcionalidade e presença nesta região infere-se que provavelmente estes estejam envolvidos nos mecanismos de sobrevivência da bactéria.

5. CONCLUSÃO

O presente trabalho apresenta dados pioneiros de um organismo de interesse médico, veterinário e biotecnológico e nos permite concluir que:

- ✓ O repertório gênico de *C. pseudotuberculosis* linhagem 226 apresenta 2138 CDS, 12 rRNA, 49 tRNA, 72 pseudogenes e conteúdo GC de 52,18%, mostrando que as características estruturais se mostraram similares entre as linhagens de *C. pseudotuberculosis* analisadas;
- ✓ Na análise filogenômica entre as linhagens foi possível identificar que a linhagem 226 possuiu maior proximidade filogenômica com a linhagem I19 e menor similaridade com a linhagem 267 do biovar *ovis*;
- ✓ No biovar *equi* observou-se maior similaridade filogenômica com a linhagem 162 e menor similaridade com a linhagem 1/06-A, demonstrando a diversidade gênica encontrada entre as linhagens deste biovar;
- ✓ A ordem gênica entre as linhagens do biovar *ovis* e *C. pseudotuberculosis* 226 apresentaram alta conservação gênica. E entre as linhagens do biovar *equi*, apesar de exibir algumas regiões de quebra a conservação da ordem gênica manteve-se bastante conservada;
- ✓ Na análise comparativa foi possível identificar 1749 genes compartilhados entre as linhagens do biovar *ovis* e 1592 entre as linhagens do biovar *equi*. A linhagem 226 apresentou 248 genes, únicos quando comparada com as linhagens do biovar *ovis* e 282 genes únicos com as do biovar *equi*, confirmando que as linhagens do biovar *ovis* são mais clonais e as do biovar *equi* apresentam maior diversidade gênica;
- ✓ Através destas análises foi possível inferir que as linhagens do biovar *ovis* apresentam pouca variação em seu repertório gênico independente de tipo de hospedeiro ou localização geográfica. Já as linhagens do biotipo *equi* possuem menor quantidade genes compartilhados com a linhagem alvo do estudo, corroborando com a diversidade gênica entre os biovares.
- ✓ A ilha 2 apresentou maior força de predição, entretanto os genes encontrados nesta região não são descritos na literatura quanto ao seu papel na virulência e/ou patogenicidade em *C. pseudotuberculosis*.

6. REFERÊNCIAS

Agren J, Sundström A, Håfström T, & Segerman B (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. PloS one, 7(6), e39107.

Alikhan N-F, Petty N K, Ben Zakour N L, & Beatson S a (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC genomics, 12(1), 402. BioMed Central Ltd.

Alves, F.S.F. e Pinheiro R R (2003). Linfadenite caseosa: recomendações e medidas profiláticas. Sociedade Nacional de Agricultura, ano 100.

Aquino de Sá M D C, Gouveia G V, Krewer C D C, Veschi J L A, de Mattos-Guaraldi A L, & da Costa M M (2013). Distribution of PLD and FagA, B, C and D genes in *Corynebacterium pseudotuberculosis* isolates from sheep and goats with caseus lymphadenitis. Genetics and molecular biology, 36(2), 265–8.

Arsenault J, Girard C, Dubreuil P, Daignault D, Galarneau J-R, Boisclair J, Simard C, & Bélanger D (2003). Prevalence of and carcass condemnation from maedi–visna, paratuberculosis and caseous lymphadenitis in culled sheep from Quebec, Canada. Preventive Veterinary Medicine, 59(1-2), 67–81.

Aziz R K, Bartels D, Best A a, DeJongh M, Disz T, Edwards R a, Formsma K, Gerdes S, Glass E M, Kubal M, Meyer F, Olsen G J, Olson R, Osterman A L, Overbeek R a, McNeil L K, Paarmann D, Paczian T, Parrello B, Pusch G D, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, & Zagnitko O (2008). The RAST Server: rapid annotations using subsystems technology. BMC genomics, 9, 75.

Baird G J, & Fontaine M C (2007). *Corynebacterium pseudotuberculosis* and its role in ovine caseous lymphadenitis. Journal of comparative pathology, 137(4), 179–210.

Bankevich A, Nurk S, Antipov D, Gurevich A a, Dvorkin M, Kulikov A S, Lesin V M, Nikolenko S I, Pham S, Prjibelski A D, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G, Alekseyev M a, & Pevzner P a (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology, 19(5), 455–77.

Billington S J, Esmay P a., Songer J G, & Jost B H (2002). Identification and role in virulence of putative iron acquisition genes from *Corynebacterium pseudotuberculosis*. FEMS Microbiology Letters, 208, 41–45.

Blom J, Albaum S P, Doppmeier D, Pühler A, Vorhölter F-J, Zakrzewski M, & Goesmann A (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC bioinformatics, 10, 154.

Bolt (2009). The population structure of the *Corynebacterium diphtheriae* group by.

Bragg L M, Stone G, Butler M K, Hugenholtz P, & Tyson G W (2013). Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS computational biology, 9(4), e1003031.

Brown C C, Olander H J, & Alves S F (1987). Synergistic hemolysis-inhibition titers associated with caseous lymphadenitis in a slaughterhouse survey of goats and sheep in Northeastern Brazil. Canadian journal of veterinary research = Revue canadienne de recherche veterinaire, 51, 46–49.

Butler J, Maccallum I, Kleber M, Shlyakhter I A, Belmonte M K, Lander E S, Nusbaum C, & Jaffe D B (2008). ALLPATHS : De novo assembly of whole-genome shotgun microreads, 810–820.

Carneiro A R, Ramos R T J, Barbosa H P M, Schneider M P C, Barh D, Azevedo V, & Silva A (2012). Quality of prokaryote genome assembly: indispensable issues of factors affecting prokaryote genome assembly quality. Gene, 505(2), 365–367. Elsevier B.V.

Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell B G, Parkhill J, & Rajandream M-A (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics (Oxford, England), 24(23), 2672–6.

Chevreux B, Pfisterer T, Drescher B, Driesel A J, Müller W E G, Wetter T, & Suhai S (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome research, 14(6), 1147–59.

Chirino-Zárraga C, Scaramelli A, & Rey-Valeirón C (2006). Bacteriological characterization of *Corynebacterium pseudotuberculosis* in Venezuelan goat flocks. Small Ruminant Research, 65(1-2), 170–175.

Conesa A, Götz S, García-Gómez J M, Terol J, Talón M, & Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics (Oxford, England), 21(18), 3674–6.

Connor K M, Quirie M M, Baird G, & Donachie W (2000). Characterization of United Kingdom isolates of *Corynebacterium pseudotuberculosis* using pulsed-field gel electrophoresis. Journal of clinical microbiology, 38(7), 2633–7.

D'Afonseca, Ali, Santos A, Pinto, Magalhães, Faria, Barbosa, Guimarães, Eslabão, Almeida, Abreu, Neto, Carneiro, Cerdeira L, Ramos R, Hirata-Jr, Mattos-Guaraldi A, Trost, Tauch, Silva, Schneider, Miyoshi, & Azevedo V (2012). Reannotation of the *Corynebacterium diphtheriae* NCTC13129 genome as a new approach to studying gene targets connected to virulence and pathogenicity in diphtheria. Open Access Bioinformatics.

Debien E, Hélie P, Buczinski S, Lebœuf A, Bélanger D, & Drolet R (2013). Proportional mortality: A study of 152 goats submitted for necropsy from 13 goat herds in Quebec, with a special focus on caseous lymphadenitis, 54(June), 581–587.

Dohm J C, Lottaz C, Borodina T, & Himmelbauer H (2007). SHARCGS , a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing, 1697–1706.

Dorella F A, Gustavo L, Achecoa C P, Liveirab S C O, Iyoshia A M, & Zevedoa V A (2006). Review article *Corynebacterium pseudotuberculosis* : microbiology , biochemical properties , pathogenesis and molecular studies of virulence, 37, 201–218.

Dorella F a, Pacheco L G, Seyffert N, Portela R W, Meyer R, Miyoshi A, & Azevedo V (2009, March). Antigens of *Corynebacterium pseudotuberculosis* and prospects for vaccine development. Expert review of vaccines.

El-Metwally S, Hamza T, Zakaria M, & Helmy M (2013). Next-Generation Sequence Assembly: Four Stages of Data Processing and Computational Challenges. (S Markel Ed)PLoS Computational Biology, 9(12), e1003345.

Field D, Wilson G, & van der Gast C (2006). How do we compare hundreds of bacterial genomes? Current opinion in microbiology, 9(5), 499–504.

Fleischmann R D, Adams M D, White O, Clayton R A, Ewen F, Kerlavage A R, Bult C J, Tomb J, Dougherty B A, Merrick J M, Mckenney K, Sutton G, Fitzhugh W, Fields C, Jeannie D, Scott J, Shirley R, Liu L, Glodek A, Kelley J M, Janice F, Phillips C A, Spriggs T, Hedblom E, & Cotton M D (2008). Whole-Genome Random Sequencing and Assembly of *Haemophilus Influenzae* Rd Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2887657.

Fontaine M C, & Baird G J (2008). Caseous lymphadenitis. Small Ruminant Research, 76(1-2), 42–48.

Fraser C, Jd G, White O, Ra C, Rd F, Cj B, & Ar K (2014). The minimal gene complement of *Mycoplasma genitalium* . PubMed Commons, 1–2.

Friden P, Newman T, & Freundlich M (1982). Nucleotide sequence of the ilvB promoter-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP. Proceedings of the National Academy of Sciences of the United States of America, 79(October), 6156–6160.

Gallo, P.; Morris H (1962). First observed cases of caseous lymphadenitis by *Corynebacterium pseudotuberculosis* in goats from Venezuela)., 1962.

Gao X-Y, Zhi X-Y, Li H-W, Klenk H-P, & Li W-J (2014). Comparative genomics of the bacterial genus *Streptococcus illuminates* evolutionary implications of species groups. PloS one, 9(6), e101229.

Gonnella G, & Kurtz S (2012). Readjoiner: a fast and memory efficient string graph-based sequence assembler. BMC bioinformatics, 13(1), 82. BMC Bioinformatics.

Henson J, Tischler G, & Ning Z (2012). Europe PMC Funders Group Next-generation sequencing and large genome assemblies, 13(8), 901–915.

Hodgson A L ., Carter K, Tachedjian M, Krywult J, Corner L A, McColl M, & Cameron A (1999). Efficacy of an ovine caseous lymphadenitis vaccine formulated using a genetically inactive form of the *Corynebacterium pseudotuberculosis* phospholipase D. Vaccine, 17(7-8), 802–808.

Hossain M S, Azimi N, & Skiena S (2009). Crystallizing short-read assemblies around seeds. BMC bioinformatics, 10 Suppl 1, S16.

Hu B, Xie G, Lo C-C, Starkenburg S R, & Chain P S G (2011). Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. Briefings in functional genomics, 10(6), 322–33.

Juhas M, van der Meer J R, Gaillard M, Harding R M, Hood D W, & Crook D W (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS microbiology reviews, 33(2), 376–93.

Jung B Y, Lee S-H, Kim H-Y, Byun J-W, Shin D-H, Kim D, & Kwak D (2015). Serology and clinical relevance of *Corynebacterium pseudotuberculosis* in native Korean goats (Capra hircus coreanae). Tropical Animal Health and Production.

Krumsiek J, Arnold R, & Rattei T (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics (Oxford, England), 23(8), 1026–8.

Kunin V, Sorek R, & Hugenholtz P (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome biology, 8(4), R61.

Laing C, Buchanan C, Taboada E N, Zhang Y, Kropinski A, Villegas A, Thomas J E, & Gannon V P J (2010). Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. BMC bioinformatics, 11, 461.

Li H (2012). Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. Bioinformatics (Oxford, England), 28(14), 1838–44.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, & Wang J (2010). De novo assembly of human genomes with massively parallel short read sequencing, 265–272.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, & Law M (2012). Comparison of next-generation sequencing systems. Journal of biomedicine & biotechnology, 2012, 251364.

Lopes T, Silva A, Thiago R, Carneiro A, Dorella F A, Rocha F S, dos Santos A R, Lima A R J, Guimarães L C, Barbosa E G V, Ribeiro D, Fiaux K K, Diniz C A A, de Abreu V A C, de Almeida S S, Hassan S S, Ali A, Bakhtiar S M, Aburjaile F F, Pinto A C, Soares S D C, Pereira U D P, Schneider M P C, Miyoshi A, Edman J, Spier S, & Azevedo V (2012). Complete genome sequence of *Corynebacterium pseudotuberculosis* strain Cp267, isolated from a llama. Journal of bacteriology, 194(13), 3567–8.

Médigue C, & Moszer I (2007). Annotation, comparison and databases for hundreds of bacterial genomes. Research in microbiology, 158(10), 724–36.

Merriman B, R&D Team I T, & Rothberg J M (2012). Progress in ion torrent semiconductor chip based sequencing. Electrophoresis, 33(23), 3397–417.

Merriman B, & Rothberg J M (2012). Progress in ion torrent semiconductor chip based sequencing. Electrophoresis, 33(23), 3397–417.

Miller J R, Koren S, & Sutton G (2010). NIH Public Access, 95(6), 315–327.

Mira A, Martín-cuadrado A B, Auria G D,& Rodríguez-valera F (2010). The bacterial pan-genome : a new paradigm in microbiology, 45–57.

Muzzi A, Masignani V, & Rappuoli R (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug discovery today, 12(11-12), 429–39.

Nagarajan N,& Pop M (2013). Sequence assembly demystified. Nature reviews. Genetics, 14(3), 157–67. Nature Publishing Group.

Oliveira C M C, Sousa M G S, Mendonça C L, Silveira J A S, Oaigen R P, Juliano S, Andrade T, Diomedes J, C A O C M, Sousa M G S, Silva N S, Mendonça C L, & Silveira J A S (2011). Prevalência e etiologia da mastite bovina na bacia leiteira de Rondon do Pará , estado do Pará 1, 31(2), 104–110.

Oliveira M, Barroco C, Mottola C, Santos R, Lemsaddek A, Tavares L, & Semedo-Lemsaddek T (2014). First report of *Corynebacterium pseudotuberculosis* from caseous lymphadenitis lesions in Black Alentejano pig (Sus scrofa domesticus). BMC veterinary research, 10(1), 218.

Olson M E, Ceri H, Morck D W, Buret A G, & Read R R (2002). Biofilm bacteria : formation and comparative susceptibility to antibiotics Résumé, 86–92.

Pareek C S, Smoczynski R, & Tretyn A (2011). Sequencing technologies and genome sequencing. Journal of applied genetics, 52(4), 413–35.

Paton M W, Walker S B, Rose I R, & Watt G F (2003). Prevalence of caseous lymphadenitis and usage of caseous lymphadenitis vaccines in sheep flocks. Australian veterinary journal, 81(1-2), 91–5.

Peel M M, Palmer G G, Stacpoole a M, & Kerr T G (1997). Human lymphadenitis due to *Corynebacterium pseudotuberculosis*: report of ten cases from Australia and review. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 24(2), 185–91.

Petty N K (2010). Genome annotation: man versus machine. Nature reviews. Microbiology, 8(11), 762. Nature Publishing Group.

Pinto A C, de Sá P H C G, Ramos R T J, Barbosa S, Barbosa H P M, Ribeiro A C, Silva W M, Rocha F S, Santana M P, de Paula Castro T L, Miyoshi A, Schneider M P C, Silva A, & Azevedo V (2014). Differential transcriptional profile of *Corynebacterium pseudotuberculosis* in response to abiotic stresses. BMC genomics, 15, 14.

Pop M, & Salzberg S L (2008). Bioinformatics challenges of new sequencing technology. Trends in genetics : TIG, 24(3), 142–9.

Pratt S M, Spier S J, Carroll S P, Vaughan B, Whitcomb M B, & Wilson W D (2005). Evaluation of clinical characteristics, diagnostic test results, and outcome in horses with internal infection caused by *Corynebacterium pseudotuberculosis*: 30 cases (1995-2003). Journal of the American Veterinary Medical Association, 227(3), 441–8.

Ramos R T J, Silva A, Carneiro A R, Pinto A C, Soares S D C, Santos A R, Almeida S S, Guimarães L C, Aburjaile F F, Barbosa E G V, Dorella F A, Rocha F S, Cerdeira L T, Barbosa M S, Tauch A, Edman J, Spier S, Miyoshi A, Schneider M P C, & Azevedo V (2012). Genome sequence of the *Corynebacterium pseudotuberculosis* Cp316 strain, isolated from the abscess of a Californian horse. Journal of bacteriology, 194(23), 6620–1.

Reed J L, Famili I, Thiele I, & Palsson B O (2006). Towards multidimensional genome annotation. Nature reviews. Genetics, 7(2), 130–41.

Richardson P (2010). Special Issue: Next Generation DNA Sequencing. Genes, 1(3), 385–387.

Rose N L, Completo G C, Lin S, Mcneil M, Palcic M M, & Lowary T L (2006). Expression , Purification , and Characterization of a Galactofuranosyltransferase Involved in *Mycobacterium tuberculosis* Arabinogalactan Biosynthesis, 6721–6729.

Ruiz J C, D'Afonseca V, Silva A, Ali A, Pinto A C, Santos A R, Rocha A a M C, Lopes D O, Dorella F a, Pacheco L G C, Costa M P, Turk M Z, Seyffert N, Moraes P M R O, Soares S C, Almeida S S, Castro T L P, Abreu V a C, Trost E, Baumbach J, et al. (2011). Evidence for reductive genome evolution and lateral acquisition of virulence functions in two *Corynebacterium pseudotuberculosis* strains. PloS one, 6(4), e18551.

Rusk N (2010). Torrents of sequence. Nature Methods, 8(1), 44–44. Nature Publishing Group.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M a, & Barrell B (2000). Artemis: sequence visualization and annotation. Bioinformatics (Oxford, England), 16(10), 944–5.

Sanger F, & Nicklen S (1977). DNA sequencing with chain-terminating, 74(12), 5463–5467.

Schuster S C (2008). Next-generation sequencing transforms today ' s biology, 5(1), 16–18.

Selim A.S (2001). Oedematous skin disease of buffalo in Egypt, 258, 241–258.

Seyffert N, Guimarães a S, Pacheco L G C, Portela R W, Bastos B L, Dorella F a, Heinemann M B, Lage a P, Gouveia a M G, Meyer R, Miyoshi a, & Azevedo V (2010). High seroprevalence of caseous lymphadenitis in Brazilian goat herds revealed by *Corynebacterium pseudotuberculosis* secreted proteins-based ELISA. Research in veterinary science, 88(1), 50–5. Elsevier Ltd.

Shendure J,& Ji H (2008). Next-generation DNA sequencing. Nature biotechnology, 26(10), 1135–45.

Silva A, Ramos R T J, Ribeiro Carneiro A, Cybelle Pinto A, de Castro Soares S, Rodrigues Santos A, Silva Almeida S, Guimarães L C, Figueira Aburjaile F, Vieira Barbosa E G, Alves Dorella F, Souza Rocha F, Souza Lopes T, Kawasaki R, Gomes Sá P, da Rocha Coimbra N A, Teixeira Cerdeira L, Silvanira Barbosa M, Cruz Schneider M P, Miyoshi A, Selim S A K, Moawad M S, & Azevedo V (2012). Complete genome sequence of *Corynebacterium pseudotuberculosis* Cp31, isolated from an Egyptian buffalo. Journal of bacteriology, 194(23), 6663–4.

Silva A, Schneider M P C, Cerdeira L, Barbosa M S, Ramos R T J, Carneiro A R, Santos R, Lima M, D'Afonseca V, Almeida S S, Santos A R, Soares S C, Pinto A C, Ali A, Dorella F a, Rocha F, de Abreu V A C, Trost E, Tauch A, Shpigel N, Miyoshi A, & Azevedo V (2011). Complete genome sequence of *Corynebacterium pseudotuberculosis* I19, a strain isolated from a cow in Israel with bovine mastitis. Journal of bacteriology, 193(1), 323–4.

Simpson J T, Wong K, Jackman S D, Schein J E, & Jones S J M (2009). ABySS : A parallel assembler for short read sequence data, 1117–1123.

Soares S C, Abreu V a C, Ramos R T J, Cerdeira L, Silva A, Baumbach J, Trost E, Tauch A, Hirata R, Mattos-Guaraldi A L, Miyoshi A, & Azevedo V (2012). PIPS: pathogenicity island prediction software. PloS one, 7(2), e30848.

Soares S C, Silva A, Trost E, Blom J, Ramos R, Carneiro A, Ali A, Santos A R, Pinto A C, Diniz C, Barbosa E G V, Dorella F a, Aburjaile F, Rocha F S, Nascimento K K F, Guimarães L C, Almeida S, Hassan S S, Bakhtiar S M, Pereira U P, Abreu V a C, Schneider M P C, Miyoshi A, Tauch A, & Azevedo V (2013). The pan-genome of the animal pathogen

Corynebacterium pseudotuberculosis reveals differences in genome plasticity between the biovar ovis and equi strains. PloS one, 8(1), e53818.

Songer (2005). The genus *Corynebacterium*. Veterinary microbiology, 2005.

Stein L (2001). Genome annotation: Fron sequence to biology, 2(July).

Stein M A, Leung K Y, Zwick M, Portillo F G, & Finlay B B (1996). Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. Molecular Microbiology, 20(1), 151–164.

Stothard P, & Wishart D S (2005). Circular genome visualization and exploration using CGView. Bioinformatics, 21(4), 537–539.

Tauch A, & Sandbote J (2014). The Family Corynebacteriaceae. (E Rosenberg, E F DeLong, S Lory, E Stackebrandt, and F Thompson Eds). Berlin, Heidelberg: Springer Berlin Heidelberg.

Tettelin H, Masignani V, Cieslewicz M J, Donati C, Medini D, Ward N L, Angiuoli S V, Crabtree J, Jones A L, Durkin A S, Deboy R T, Davidsen T M, Mora M, Scarselli M, Margarit y Ros I, Peterson J D, Hauser C R, Sundaram J P, Nelson W C, Madupu R, et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America, 102(39), 13950–5.

Trost E, Ott L, Schneider J, Schröder J, Jaenicke S, Goesmann A, Husemann P, Stoye J, Dorella F A, Rocha F S, Soares S D C, D'Afonseca V, Miyoshi A, Ruiz J, Silva A, Azevedo V, Burkovski A, Guiso N, Join-Lambert O F, Kayal S, & Tauch A (2010). The complete genome sequence of *Corynebacterium pseudotuberculosis* FRC41 isolated from a 12-year-old girl with necrotizing lymphadenitis reveals insights into gene-regulatory networks contributing to virulence. BMC genomics, 11(1), 728. BioMed Central Ltd.

Venter J C, Adams M D, Myers E W, Li P W, Mural R J, Sutton G G, Smith H O, Yandell M, Evans C a, Holt R a, Gocayne J D, Amanatides P, Ballew R M, Huson D H, Wortman J R, Zhang Q, Kodira C D, Zheng X H, Chen L, Skupski M, et al. (2001). The sequence of the human genome. Science (New York, N.Y.), 291(5507), 1304–51.

Warren R L, Sutton G G, Jones S J M, & Holt R a (2007). Assembling millions of short DNA sequences using SSAKE. Bioinformatics (Oxford, England), 23(4), 500–1.

Yadav N K, Shukla P, Omer A, Pareek S, & Singh R K (2014). Next generation sequencing: potential and application in drug discovery. TheScientificWorldJournal, 2014, 802437.

Yeo Z X, Chan M, Yap Y S, Ang P, Rozen S, & Lee A S G (2012). Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. PloS one, 7(9), e45798.

Yeruham I, Friedman S, Perl S, Elad D, Berkovich Y, & Kalgard Y (2004). A herd level analysis of a *Corynebacterium pseudotuberculosis* outbreak in a dairy cattle herd. Veterinary dermatology, 15(5), 315–20.

Zdobnov E M, & Apweiler R (2001). signature-recognition methods in InterPro, 17(9), 847–848.

Zerbino D R, & Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research, 18(5), 821–9.

Zhang J, Chiodini R, Badr A, & Zhang G (2011). The impact of next-generation sequencing on genomics. Journal of genetics and genomics Yi chuan xue bao, 38(3), 95–109. Elsevier Limited and Science Press.

Zhao Y, Wu J, Yang J, Sun S, Xiao J,& Yu J (2012). PGAP: pan-genomes analysis pipeline. Bioinformatics (Oxford, England), 28(3), 416–8.

APÊNDICIE

Ilhas	Gene	Produto	Força de predição
Ilha 1	Cp226_0022	Proteína hipotética	Fraca
	Cp226_0023	Proteína hipotética	
	Cp226_0024	Proteína hipotética	
	Cp226_0025	Proteína hipotética	
	Cp226_0026	Proteína hipotética	
	Cp226_0027	Proteína hipotética	
	cas5	Proteína associada ao CRISPR	
	Cp226_0029	Proteína de elemento de inserção	
	pld	Fosfolipase D	
	fagC	Proteína de ligação ao ATP da membrana citoplasmática	
	fagB	Transportador de ferro-enterobactina	
	fagA	Proteína integral de membrana	
	fagD	Proteína de ligação ferro-sideróforo	
Ilha 2	Cp226_0185	Proteína hipotética	Forte
	Cp226_0186	Proteína hipotética	
	Cp226_0187	Proteína hipotética	
	Cp226_0188	Proteína hipotética	
	Cp226_0189	Proteína hipotética	
	potA1	Proteína de ligação ao ATP do importador de Espermidina/ putrescina Pota	
	potC	Permease do sistema de transporte de espermidina / putrescina potC	
	potD	Proteína periplasmática de ligação Espermidina/ putrescina	
	glpT	Transportador de glicerol-3-fosfato	
	mprA1	Regulador de resposta	
	senX3A	Proteína que contém o domínio sensor	

		de histidina kinase	
Ilha 3	Cp226_0450	Proteína hipotética	Fraca
	Cp226_0451	Proteína que contêm o domínio sacarase-ferredoxina	
	metI	Permease do transportador ABC de metionina	
	metN	Proteína de ligação ao ATP do transportador ABC de metionina	
	metQ	Proteína de ligação ao substrato do transportador ABC de metionina	
	Cp226_0455	Proteína hipotética	
	metQ1	Proteína de ligação ao substrato do transportador ABC de metionina	
	Cp226_0457	Proteína hipotética	
	Cp226_0458	Proteína hipotética	
	Cp226_0459	Proteína hipotética	
	Cp226_0460	Proteína hipotética	
	Cp226_0461	Proteína de ligação ao substrato do transportador ABC de manganês	
	mntB3	Proteína de ligação ao ATP do transportador ABC de manganês	
	mntC	Proteína de membrana do transportador ABC de manganês	
	mntD3	Proteína de membrana do transportador ABC de manganês	
	mntR	Regulador transcricional da família DtxR	
	spoU1	tRNA (citidine(34)-2'-O)-metiltransferase	
	Cp226_0468	Proteína hipotética	
	metX	Homosserine O-acetiltransferase	
	htaB	Receptor de hemina na superfície da célula	
Ilha 4	*ciuA*	Proteína de ligação ao substrato do transportador ABC de ferro	Normal
	ciuB	Proteína que contém o domínio permease do transportador ABC de	

		ferro	
	ciuC	Permease do transportador ABC de ferro	
	ciuD	Proteína de ligação ao ATP do transportador ABC de ferro	
	ciuE	Proteína relacionada à biossíntese de sideróforos	
	ciuF	Proteína de efluxo	
	Cp226_1026	Proteína hipotética	
	Cp226_1027	Proteína hipotética	
	Cp226_1028	Proteína hipotética	
	Cp226_1029	Proteína hipotética	
	Cp226_1030	Proteína hipotética	
Ilha 5	*oppA5*	Proteína de ligação ao substrato do transportador ABC de peptídeos	Normal
	oppB4	Permease do sistema de transporte de oligopeptídeos oppB	
	oppC4	Permease do sistema de transporte de oligopeptídeos oppC	
	oppdF2	Proteína de ligação ao ATP do sistema de transporte de peptídeos OppD	
	Cp226_1616	Proteína hipotética	
	Cp226_1618	Enzima piridoxal fosfato da família YggS	
	Cp226_1619	Proteína hipotética	
	Cp226_1620	Proteína hipotética	
	Cp226_1621	Proteína hipotética	
	Cp226_1622	Binding-protein-dependent transport system inner membrane component	
	Cp226_1623	Proteínas bacterianas extracelulares de ligação a soluto, família 5 Middle	
	Cp226_1624	Proteína similar à proteínas secretadas	
	Cp226_1625	Transportador de efluxo de drogas do tipo MFS	
	Cp226_1626	Transportador ABC de oligopeptídeos	
	Cp226_1627	Proteína de ligação ao ATP do	

		transportador ABC	
	Cp226_1628	Proteína hipotética	
Ilha 6	Cp226_1979	Proteína ancorada à membrana	Normal
	Cp226_1980	Proteína ancorada à superfície da membrana	
	Cp226_1981	Fator de von Willebrand, tipo A (vWF)	
	Cp226_1982	Proteína hipotética	
	srtA2	Sortase A	
	Cp226_1984	Proteína hipotética	
	oppD3	Sistema de transporte ABC, subunidade de ligação ao ATP	
	oppcD1	Permease do sistema de transporte de oligopeptídeos	
	oppB5	Permease do sistema de transporte de oligopeptídeos	
	pitB	Fosfato permease	
	Cp226_1990	Proteína hipotética	
	Cp226_1991	Oxirredutase	
	fadF	Oxirredutase Ferro-Enxofre	
	Cp226_1993	Proteína hipotética	
	Cp226_1994	Proteína de regulação transcricional	
	phuC	Proteína yusV similar à permease de transporte de dicitrato de ferro(III)	
	fecD2	Permease do sistema de transporte de dicitrato de ferro(III)	
	Cp226_1997	Proteína que contém o domínio do componente permease do transportador ABC	
Ilha 7	Cp226_2009	Lisil-tRNA sintetase	Normal
	Cp226_2010	Hidrolase/aciltransferase	
	Cp226_2011	Proteína hipotética	
	Cp226_2012	Proteína hipotética	
	udgA	UDP-glicose desidrogenase	
	dcd	Desoxicitidina trifosfato deaminase	

	Cp226_2015	Succinato semialdeído desidrogenase (NAD)	
	Cp226_2016	Proteína hipotética	
Ilha 8	*pspA1*	Phage shock protein A	Fraca
	Cp226_2109	Proteína hipotética	
	Cp226_2110	Proteína hipotética	
	Cp226_2111	Proteína de ligação ao ATP do transportador ABC	
	Cp226_2112	Regulador transcricional da família GntR	
	Cp226_2113	Proteína hipotética	
	pspA1	Phage shock protein A	