CRISTINA KRAEMER ZIMPEL

**Sequenciamento, anotação e análise do genoma completo de *Mycobacterium*
bovis cepa SP38**

São Paulo
2017

CRISTINA KRAEMER ZIMPEL

**Sequenciamento, anotação e análise do genoma completo de *Mycobacterium*
bovis cepa SP38**

Dissertação/Tese apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Mestre/Doutor em Ciências.

Departamento:

Medicina Veterinária Preventiva e Saúde Animal

Área de concentração:

Epidemiologia Experimental Aplicada às Zoonoses

Orientadora:

Profa. Dra. Ana Marcia de Sá Guimarães

De acordo: _____
Orientadora

São Paulo
2017

Obs: A versão original se encontra disponível na Biblioteca da FMVZ/USP

Autorizo a reprodução parcial ou total desta obra, para fins acadêmicos, desde que citada a fonte.

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO

(Biblioteca Virginie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo)

T. 3498
FMVZ

Zimpel, Cristina Kraemer
Sequenciamento, anotação e análise do genoma completo de *Mycobacterium bovis* cepa SP38. / Cristina Kraemer Zimpel. -- 2017.
81 p. : il.

Dissertação (Mestrado) - Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Departamento de Medicina Veterinária Preventiva e Saúde Animal, São Paulo, 2017.

Programa de Pós-Graduação: Epidemiologia Experimental Aplicada às Zoonoses.

Área de concentração: Epidemiologia Experimental Aplicada às Zoonoses.

Orientador: Profa. Dra. Ana Marcia de Sá Guimarães.

1. *Mycobacterium bovis*. 2. Sequenciamento de genoma. 3. Complexo *Mycobacterium tuberculosis*. 4. Genômica comparativa. I. Título.

São Paulo, 24th May 2016

CERTIFIED

We certify that the Research "Sequencing, annotation and genomic analyses of Mycobacterium bovis strain SP38", protocol number CEUAX 3468170315, under the responsibility Ana Marcia De Sá Guimarães, agree with Ethical Principles in Animal Research adopted by Ethic Committee in the Use of Animals of the School of Veterinary Medicine and Animal Science of University of São Paulo, and was approved in the meeting of day May 13, 2015.

Certificamos que o protocolo do Projeto de Pesquisa intitulado "Sequenciamento, anotação e análise do genoma completo de Mycobacterium bovis cepa SP38", protocolado sob o CEUAX nº 3468170315, sob a responsabilidade de Ana Marcia De Sá Guimarães, está de acordo com os princípios éticos de experimentação animal da Comissão de Ética no Uso de Animais da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, e foi aprovado na reunião de 13 de maio de 2015.

Profa. Dra. Denise Tabacchi Fantoni
Presidente da Comissão de Ética no Uso de Animais
Faculdade de Medicina Veterinária e Zootecnia da Universidade
de São Paulo

Roseli da Costa Gomes
Secretaria Executiva da Comissão de Ética no Uso de Animais
Faculdade de Medicina Veterinária e Zootecnia da Universidade
de São Paulo

FOLHA DE AVALIAÇÃO

Autor: ZIMPEL, Cristina Kraemer

Título: Sequenciamento, anotação e análise do genoma completo de *Mycobacterium bovis* cepa SP38.

Dissertação apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Data: ____/____/____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

AGRADECIMENTO ESPECIAL

Ao conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) número de processo 133126/2015-3 e ao Programa Jovem Talento do Ciência sem Fronteiras (CsF) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) de número A008_2013 pelo apoio financeiro que permitiu a realização desse estudo.

AGRADECIMENTOS

Agradeço à minha família, especialmente meus pais por sempre apoiarem minhas escolhas. Meus irmãos e minha cunhada Jaque, que apesar da distância, incentivam o meu estudo. Sem vocês eu não conseguiria estar realizando meus sonhos.

Às minhas amigas de infância, Maiara e Gabrielle, que sempre posso contar e são família que escolhi levar para a vida toda.

Com tantas mudanças, todas as pessoas que passaram pelo meu caminho deixam uma marca e muita saudade, os amigos de Santa Maria, cada um seguiu seu caminho, mas sempre lembrarei de vocês com carinho. Talita, Gabi Werle, Gabi Rocha, Andressa, e vários outros, guardo com carinho essa época. Aos meus amigos de Curitiba também me ajudaram muito. Mara, Júlia, Vivien, obrigada por todos os dias de alegria que vieram junto com a residência, com vocês aprendi tanto, profissionalmente e pessoalmente.

Em São Paulo fiz novos amigos que são muito importantes hoje, fazem com que esse desafio se torne incrivelmente maravilhoso. Aos amigos que fiz no VPS, tantas risadas e ao mesmo tempo nos preocupamos, em especial Juliana e Antonio. Aos funcionários do VPS que tornam esse lugar tão divertido. Ao Danival que sempre resolve nossas dúvidas e escuta nossos problemas. Agradeço ao professor Paulo que aceitou me receber de mala e cuia no Laboratório de Raiva. Ao LZB e ao prof. Marcos, que estão sempre dispostos a ajudar nos projetos e por tantos ensinamentos.

Obrigada muito a Celminha sempre será nossa segunda mãe longe de casa. Aos amigos do ICB, Maysa e Aline sempre prontas para o trabalho e diversão, agradeço muito por ter vocês como amigas. Iza, Camilinha, Naty e Regis, vocês são incríveis. Ao Lucas e ao Gui, amigos desde Purdue, os quais tantas histórias divertidas podemos contar. A todos vocês, sei que seremos amigos por muito tempo.

Aos animais que amo tanto e me alegram nessa escolha da profissão de veterinária, Fiinha, Memphis, Tina e Pi.

Em especial, agradeço a minha amiga e também orientadora, Ana Marcia. Obrigada por acreditar em mim, e por ser minha inspiração de cada dia. Espero um dia retribuir todos teus esforços em permitir a realização desse sonho.

Tantas pessoas que tem um papel importante nessa etapa, obrigada a todos por fazerem parte da minha vida e me ajudarem a ser uma pessoa e pesquisadora melhor.

“Impossible is just the possible that someone has not put enough effort
to make it come true”
- Miguel Nicolelis

RESUMO

ZIMPEL, C. K. **Sequenciamento, anotação e análise do genoma completo de *Mycobacterium bovis* cepa SP38**. [Sequencing, annotation and genomic analysis of *Mycobacterium bovis* strain SP38]. 2017. 81 p. Dissertação (Mestrado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2017.

A tuberculose é uma doença infectocontagiosa causada por bactérias do complexo *Mycobacterium tuberculosis* (MTBC) que afeta humanos e/ou animais. Membros desse complexo evoluíram clonalmente e possuem grande similaridade genômica, diferenciando-se por polimorfismos de nucleotídeo único (SNPs) e regiões de diferença (RDs). Dentre os patógenos da tuberculose em animais, *Mycobacterium bovis*, causador da tuberculose bovina, é o membro do MTBC de maior importância global. Desta maneira, o presente estudo tem por objetivo o sequenciamento, a anotação e a análise da estirpe brasileira SP38 de *M. bovis*, seguido da genômica comparativa desse com outros genomas de *M. bovis* depositados no GenBank. *Mycobacterium bovis* SP38 apresenta um genoma tradicional de micobactéria tuberculosa, sendo esse único, circular com 4.347.646 pb, alto conteúdo de GC (65,6%) e 4.216 genes, incluindo 154 pseudogenes, 3 genes de rRNA (RNA ribossomal), 45 de tRNA (RNA transportador), 2 de ncRNA (RNA não codificante), 1 tmRNA (RNA transferência-mensageiro) e 4.011 sequências de DNA codificante (CDSs) (NZ_CP015773.1). Dentre as CDSs, a maioria (2.805 - 69,93%) foi anotado com função e 1.206 (30,07%) como hipotéticos. Para a genômica comparativa, os 31 genomas completos ou em *drafts* de *M. bovis* depositados no GenBank, 32 genomas de *Mycobacterium bovis* BCG e 23 genomas de *Mycobacterium tuberculosis* foram selecionados. Análises *in silico* dos padrões de RD resultaram na exclusão de três genomas anotados equivocadamente como *M. bovis* virulentos. A análise de genes ortólogos sugere que *M. bovis* está sob processo de *decay* genômico. A quantificação de sítios polimórficos indica uma maior variabilidade genética em números totais (8.335 em *M. tuberculosis*, 3.448 em *M. bovis* virulentos, e 1.088 em *M. bovis* BCGs) e comparações par-a-par ($p \leq 0,05$) de *M. tuberculosis* em relação a *M. bovis* virulentos e BCGs, indicando uma maior pressão evolutiva sob *M. tuberculosis*, contrastando com o fato de que *M. bovis* é capaz de infectar um maior número de espécies hospedeiras que *M. tuberculosis*. A maioria desses sítios polimórficos estão localizados em CDSs hipotéticos (31,7% - 51,3%), sendo associados a família gênica PE/PPE, e apresentam uma proporção de mutações não sinônimas crescentes pela ordem *M. bovis* BCG, *M. bovis* virulentos e *M. tuberculosis* (48,90%, 51,92% e 59,52%, respectivamente). Essa menor proporção de mutações não sinônimas e a categorização funcional dissimilar entre CDSs

contendo sítios polimórficos, indica que *M. bovis* BCG está sujeito a diferentes pressões seletivas quando comparado a *M. bovis* virulentos e *M. tuberculosis*. Por fim, a análise filogenética baseada em sítios polimórficos indica agrupamentos filogenéticos de *M. bovis* suportados pela classificação dos Complexos Clonais (CCs) e não por hospedeiros de origem dos isolados, confirmando que sítios polimórficos podem ser utilizados para classificação filogenética de linhagens genéticas desta espécie bacteriana. Além do mais, 2/28 (7,14%) genomas de *M. bovis* não puderam ser classificados nos CCs atualmente descritos, sugerindo a existência de complexos ainda não determinados. Este estudo representa o primeiro genoma de uma estirpe nacional de *M. bovis* a ser completamente sequenciado e a primeira análise de genômica comparativa de genomas desta espécie bacteriana.

Palavras chave: *Mycobacterium bovis*. Sequenciamento de genoma. Genômica comparativa. Complexo *Mycobacterium tuberculosis*.

ABSTRACT

ZIMPEL, C. K. **Sequencing, annotation and genomic analysis of *Mycobacterium bovis* strain SP38**. [Sequenciamento, anotação e análise do genoma completo de *Mycobacterium bovis* cepa SP38]. 2017. 81 p. Dissertação (Mestrado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2017.

Tuberculosis is an infectious disease caused by bacteria of the *Mycobacterium tuberculosis* Complex (MTBC) that affects human beings and/or animals. Members of this complex clonally evolved and have high genomic similarity, differentiated by single nucleotide polymorphisms (SNPs) and regions of difference (RDs). Among the animal tuberculosis pathogens, *Mycobacterium bovis*, the causative agent of bovine tuberculosis, is the MTBC member of greatest global importance. Therefore, the aim of the present study is to sequence, assemble and annotate the genome of the Brazilian strain SP38 of *M. bovis*, followed by the comparative genomics with other *M. bovis* genomes available in GenBank. *Mycobacterium bovis* SP38 has a traditional mycobacteria genome. It has a single and circular chromosome with 4,347,646 bp, high GC content (65.6%), and 4,216 genes, including 154 pseudogenes, 3 rRNA genes (ribosomal RNA), 45 tRNA (transfer RNA), 2 ncRNA (non-coding RNA), 1 tmRNA (transfer-messenger RNA), and 4,011 coding DNA sequences (CDSs) (NZ_CP015773.1). The majority of CDSs (2,805 - 69,93%) was annotated with function and 1,206 (30,07%) are hypothetical. For the comparative genomics analyses, the 31 genomes (complete and drafts) of *M. bovis* available in GenBank, 32 *Mycobacterium bovis* BCG and, 23 of *Mycobacterium tuberculosis* were chosen. *In silico* analysis of the RDs patterns resulted in the exclusion of three genomes, mistakenly annotated as virulent *M. bovis*. Orthologous gene analysis suggests that strains of *M. bovis* are under genomic decay. The quantification of polymorphic sites indicates the greater variability in absolute numbers (8,335 in *M. tuberculosis*, 3,448 in virulent *M. bovis*, and 1,088 in *M. bovis* BCG) and in pairwise comparisons ($p \leq 0,05$) of *M. tuberculosis* compared to virulent *M. bovis* and *M. bovis* BCG, suggesting that *M. tuberculosis* is under high evolutionary pressure. This is in contrast to the fact that *M. bovis* is capable of infecting a higher number of host species than *M. tuberculosis*. Most of these polymorphic sites are located in hypothetical CDSs (31.7% - 52.3%), being associated with PE/PPE family, and demonstrating a nonsynonymous mutations proportion of the following increasing order: *M. bovis* BCG, virulent *M. bovis* and *M. tuberculosis* (48.90%, 51.92% and 59.52%, respectively). This lower proportion of nonsynonymous mutations and the dissimilar functional categorization of CDSs with polymorphic sites indicates that *M. bovis* BCG is subjected to different selective pressure

when compared to virulent *M. bovis* and *M. tuberculosis*. Finally, the phylogenetic analysis based on polymorphic sites indicates that the phylogenetic grouping of *M. bovis* is supported by Clonal Complexes (CCs), and not by the host of *M. bovis* isolates, confirming that polymorphic sites can be used for phylogenetic classification of genetic lineages of this bacterial species. Furthermore, 2/28 (7.14%) genomes of *M. bovis* could not be classified in the currently described CCs, suggesting the existence of complexes yet to be determined. This study represents the first genome of a Brazilian strain of *M. bovis* to be completely sequenced and the first comparative genomic analysis of the genomes of this bacterial species.

Key words: *Mycobacterium bovis*. Genome sequencing. Comparative genomic. *Mycobacterium tuberculosis* Complex.

LISTA DE FIGURAS

Figura 1 - Ilustração evolutiva dos micro-organismos pertencentes ao complexo <i>Mycobacterium tuberculosis</i>	23
Figura 2 - Mapa circular do genoma completo de <i>Mycobacterium bovis</i> estirpe SP38.....	38
Figura 3 - Gradiente de sintenia com genomas de <i>Mycobacterium bovis</i> , <i>M. bovis</i> BCG e <i>M. tuberculosis</i>	44
Figura 4 - Diagrama de Venn demonstrando compartilhamento de genes ortólogos entre <i>Mycobacterium bovis</i> , <i>M. bovis</i> BCG e <i>M. tuberculosis</i>	48
Figura 5 - Padrão de distribuição de sítios polimórficos com avaliação <i>pairwise</i> entre genomas de <i>Mycobacterium bovis</i> BCG, <i>M. bovis</i> e <i>M. tuberculosis</i>	50
Figura 6 - Diagrama de Venn demonstrando proteínas únicas e compartilhamento de proteínas ortólogas envolvidas em sítios polimórficos, entre genomas de <i>Mycobacterium bovis</i> BCG, <i>M. bovis</i> e <i>M. tuberculosis</i>	54
Figura 7 - Diagrama de Venn demonstrando proteínas únicas e compartilhamento de proteínas ortólogas com mutações não sinônimas, entre <i>M. bovis</i> BCG, <i>M. bovis</i> e <i>M. tuberculosis</i> , e sua categorização em COGs (<i>Cluster of Orthologous Groups</i>)	56
Figura 8 - Árvore filogenética obtida pelo método <i>Neighbor-joining</i> , utilizando SNPs (polimorfismo de nucleotídeo único) de membros de Complexo <i>Mycobacterium tuberculosis</i> (MTBC). Os genomas de <i>Mycobacterium bovis</i> são acompanhados da identificação espécie hospedeira afetada, local de isolamento, número SB do <i>spoligotyping</i> e agrupados conforme os Complexos Clonais	58

LISTA DE TABELAS

Tabela 1 - Principais Regiões de Diferença (RDs) utilizadas na diferenciação de membros do Complexo <i>Mycobacterium tuberculosis</i>	22
Tabela 2 - Regiões de diferença (RDs) avaliadas para determinação dos Complexos clonais Europeu 1, Africano 1 e Africano 2 de <i>Mycobacterium bovis</i>	34
Tabela 3 - Comparação das características gerais dos genomas completos de <i>Mycobacterium bovis</i> SP38, <i>M. bovis</i> AF2122/97 e <i>M. tuberculosis</i> H37Rv.....	39
Tabela 4 - Padrões binários, OCTcode e SB referentes aos espoligotipos dos genomas de <i>Mycobacterium bovis</i> depositados no GenBank.....	40
Tabela 5 - Resultados dos perfis que determinaram os Complexos Clonais dos genomas de <i>Mycobacterium bovis</i> depositados no Genbank.....	42
Tabela 6 - Genes ortólogos presentes nos 28 genomas <i>Mycobacterium bovis</i> categorizados pelos COGs (<i>Cluster of Orthologous</i>)	44
Tabela 7 - Proteínas que não apresentam homólogos e são encontradas apenas em um dos 28 genomas de <i>Mycobacterium bovis</i>	46
Tabela 8 - Número de grupos de proteínas homólogas encontrados unicamente nos genomas de <i>Mycobacterium tuberculosis</i>	48
Tabela 9 - Categorização e as determinações de mutações sinônimas e não sinônimas dos sítios polimórficos encontrados em <i>Mycobacterium bovis</i> de acordo com as funções dos COGs (<i>Cluster of Orthologous</i>)	50
Tabela 10 - Categorização e as determinações de mutações sinônimas e não sinônimas dos sítios polimórficos encontrados em <i>Mycobacterium tuberculosis</i> de acordo com as funções dos COGs (<i>Cluster of Orthologous</i>)	51
Tabela 11 - Categorização e as determinações de mutações sinônimas e não sinônimas dos sítios polimórficos encontrados em <i>Mycobacterium bovis</i> BCG de acordo com as funções dos COGs (<i>Cluster of Orthologous</i>)	53

LISTA DE ABREVIATURAS

CC - Complexo Clonal

CDS - sequência de DNA codificante

COG - *Cluster of Orthologous Groups*

Conteúdo de GC - guanina e citosina

CRISPR - *clustered regularly interspaced palindromic repeats*

CTAB - brometo de cetiltrimetilamônio

DR - repetições diretas

ENA - *European Nucleotide Archive*

HGT - transferência horizontal de genes

HPC - cloreto de hexadecilpiridínio

IGS - *Institute for Genome Sciences*

LSP- *large sequence polymorphisms*

MIRU - Unidades Repetitivas Micobacterianas

ML- máxima verossimilhança (*Maximum Likelihood*)

MTBC - complexo *Mycobacterium tuberculosis*

NaCl - cloreto de sódio

NCBI - *National Center for Biotechnology Information*

ncRNA - RNA não codificante

NJ - *Neighbor-Joining*

OIE - *World Organisation for Animal Health*

pb - pares de base

PE - prolina-glutamato

PCR - Reação em Cadeia da Polimerase

PGAP - *Prokaryotic Genome Annotation Pipeline*

PGF - família de genes parálogos

PKS - policetídeos sintase

PNCEBT - Programa Nacional de Controle e Erradicação da Brucelose e da Tuberculose Animal

PPE - prolina-prolina-glutamato

RDs - regiões de diferença

RNA - ácido ribonucleico

rRNA - ácido ribonucleico ribossomal

SNPs - polimorfismos de nucleotídeo único

tmRNA - ácido ribonucleico transferência-mensageiro

tRNA - ácido ribonucleico transportador

USP - Universidade de São Paulo

VNTR - *Variable Number Tandem Repeat*

VPS - Departamento de Medicina Veterinária Preventiva e Saúde Animal

SUMÁRIO

1	INTRODUÇÃO	18
2	REVISÃO DE LITERATURA	20
2.1	TUBERCULOSE	20
2.2	COMPLEXO <i>MYCOBACTERIUM TUBERCULOSIS</i>	21
2.3	<i>MYCOBACTERIUM BOVIS</i>	24
2.3.1	Avaliação da Epidemiologia Molecular de <i>M. bovis</i>	26
2.3.2	Genômica comparativa de <i>M. bovis</i>	27
3	OBJETIVOS	29
3.1	OBJETIVO GERAL	29
3.2	OBJETIVOS ESPECÍFICOS	29
4	MATERIAIS E MÉTODOS	30
4.1	ISOLAMENTO DA ESTIRPE DE <i>MYCOBACTERIUM BOVIS</i> E EXTRAÇÃO DO DNA	30
4.2	BIBLIOTECAS GENÔMICAS E SEQUENCIAMENTO	31
4.3	MONTAGEM E FECHAMENTO DO GENOMA	31
4.4	PREDIÇÃO DE GENES E ANOTAÇÃO	32
4.5	GENÔMICA COMPARATIVA	32
4.5.1	Escolha dos genomas	32
4.5.2	Marcadores epidemiológicos	33
4.5.3	Complexo Clonal	33
4.5.4	Parálogos, ortólogos e sintenia gênica	34
4.5.5	Identificação e análise de sítios polimórficos	35
4.5.6	Filogenômica	36
5	RESULTADOS E DISCUSSÃO	37
5.1	O GENOMA DE <i>M. BOVIS</i> SP38, UMA TRADICIONAL MICOBACTÉRIA TUBERCULOSA	37
5.2	GENÔMICA COMPARATIVA	39
5.2.1	Nem todos os genomas depositados no GenBank são <i>M. bovis</i> virulentos	39
5.2.2	O espoligotipo mais comum dos genomas de <i>M. bovis</i> virulentos depositados no GenBank é o SB0140	40
5.2.3	Complexo clonal Europeu 2 é encontrado no Brasil	42
5.2.4	Regiões sem sintenia gênica são observadas entre os genomas de <i>M. bovis</i> virulentos	43
5.2.5	Grupos de genes ortólogos	44
5.2.5.1	Genoma <i>core</i> de <i>M. bovis</i> apresenta um grande número de CDSs de função desconhecida	44
5.2.6	Pesquisa de <i>singletons</i> de <i>M. bovis</i> demonstra <i>decay</i> genômico	46

5.2.6.1	Grupos de genes ortólogos entre os genomas de <i>M. bovis</i> , <i>M. bovis</i> BCG e <i>M. tuberculosis</i> : não existem CDSs exclusivamente presente em todos os genomas de <i>M. bovis</i>	47
5.2.7	Sítios Polimórficos	49
5.2.7.1	Estirpes de <i>M. tuberculosis</i> possuem um maior número de sítios polimórficos quando comparado a estirpes de <i>M. bovis</i> virulentos e <i>M. bovis</i> BCGs	49
5.2.7.2	Anotação funcional dos sítios polimórficos: <i>M. bovis</i> BCG está sujeito a pressões evolutivas distintas daquelas de micobactérias virulentas.	50
5.2.7.3	Genes ortólogos envolvidos em sítios polimórficos: <i>M. bovis</i> e <i>M. tuberculosis</i> compartilham um maior número de genes sob pressão evolutiva entre si do que com <i>M. bovis</i> BCG	54
5.2.8	Filogenia inferida por SNPs agrupa linhagens de <i>M. bovis</i> conforme os Complexos Clonais	57
7	CONSIDERAÇÕES FINAIS	59
	REFERÊNCIAS	60
	APÊNDICES	77

1 INTRODUÇÃO

A tuberculose é uma doença infectocontagiosa que afeta milhões de pessoas e animais mundialmente e é causada por membros do complexo *Mycobacterium tuberculosis* (MTBC). O MTBC é representado pelos patógenos altamente adaptados aos seres humanos, *Mycobacterium tuberculosis* e *Mycobacterium africanum* linhagens 5 e 6, e também pelos patógenos adaptados a animais, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium pinnipedii*, *Mycobacterium microti*, *Mycobacterium orygis*, *Mycobacterium suricattae*, *Mycobacterium mungi*, “dassie bacillus” e “chimpanzee bacillus” (COSCOLLA; GAGNEUX, 2014; RODRIGUEZ-CAMPOS et al., 2014). Esses micro-organismos são caracterizados por serem álcool-ácido resistentes e por possuírem atividade intracelular, principalmente em macrófagos. São imóveis, não esporulados, de crescimento lento e possuem parede celular rica em lipídeos complexos (KIESER; RUBIN, 2014). Além disso, esses patógenos possuem grande similaridade genômica (> 99,95% de identidade nas regiões alinháveis) e evolução clonal, em que surgiram a partir de um ancestral de *M. tuberculosis*. As principais características que tornam essas espécies diferentes são as variações fenotípicas relacionadas aos níveis de patogenicidade e adaptabilidade às espécies hospedeiras e de pequenas variações genotípicas, como polimorfismos de nucleotídeo único (SNPs) e áreas de deleções, conhecidas como regiões de diferença (RDs) (GALAGAN, 2014).

Dentre as espécies do MTBC, a que constitui a maior ameaça à saúde animal é *M. bovis*. Esse patógeno é conhecido como o principal causador da tuberculose bovina, uma doença de notificação obrigatória conforme normas da OIE (*World Organisation for Animal Health*), mas também pode causar doença em outras espécies animais. A infecção nos bovinos ocorre principalmente pela inalação de aerossóis provenientes de hospedeiros contaminados, através da qual a micobactéria atinge o trato respiratório para residir e proliferar dentro dos macrófagos. Essa multiplicação intracelular pode levar à formação de lesões granulomatosas em diversos órgãos, mas principalmente no trato respiratório. Já em humanos, a contaminação por *M. bovis* se dá sobretudo pela ingestão de produtos de origem animal contaminados. A transmissão via aerossol animal-humano é considerada rara. A doença por *M. bovis* tende a não persistir na população humana, sendo dificilmente transmitida entre indivíduos, mesmo em casos de doença pulmonar (DE LA RUA-DOMENECH et al., 2006; MALAMA; MUMA; GODFROID, 2013).

No Brasil a prevalência da doença causada por *M. bovis* em bovinos e bubalinos é considerada baixa (~ 1,3% dos rebanhos nacionais) e o controle é regulamentado pelo Programa Nacional de Controle e Erradicação da Brucelose e da Tuberculose Animal (PNCEBT)

(BRASIL, 2006). Pouco se sabe sobre a real prevalência de *M. bovis* em humanos (DE KANTOR et al., 2008; SOBRAL et al., 2011; SILVA et al., 2013). Já em animais selvagens, os poucos relatos dessa doença são referentes à animais mantidos em cativeiro, e não se sabe se algum animal silvestre poderia atuar como reservatório e ter capacidade de potencializar a transmissão do patógeno entre rebanhos domésticos ou animais selvagens (ZIMPEL et al., 2017).

Enquanto que *M. tuberculosis* é altamente adaptado a humanos, *M. bovis* é capaz de infectar uma ampla gama de hospedeiros, com variável persistência populacional. Algumas espécies animais infectadas por *M. bovis* podem ser classificadas como reservatórios, enquanto outras como hospedeiros *dead-end*. Fatores determinantes dessa variável adaptabilidade hospedeira e virulência permanecem desconhecidos (BEHR; GORDON, 2015). Uma ferramenta para compreender diferenças genéticas que podem ser responsáveis por essas diferenças é o sequenciamento de genomas. O sequenciamento e anotação de genomas de *M. bovis*, assim como a comparação entre os diferentes genomas de espécies do MTBC, poderá ajudar a compreender melhor os mecanismos de atuação bacteriana tanto no ambiente como no hospedeiro, permitindo gerar dados genotípicos que possam ser experimentalmente validados no futuro quanto às suas expressões fenotípicas. Assim, o presente trabalho tem como objetivo sequenciar o genoma completo de uma estirpe nacional de *M. bovis* e realizar genômica comparativa entre diferentes genomas do MTBC disponíveis em bancos de dados públicos. O sequenciamento completo da estirpe de *M. bovis* SP38 permitirá a descrição completa da composição molecular de um isolado nacional pela primeira vez.

2 REVISÃO DE LITERATURA

2.1 TUBERCULOSE

A tuberculose é uma doença infecto-contagiosa causada por bactérias pertencentes ao complexo *Mycobacterium tuberculosis* (MTBC). Essa doença é caracterizada por comprometimento do trato respiratório, local onde a bactéria coloniza macrófagos e irá residir e se proliferar, podendo então levar à formação de lesões granulomatosas (forma pulmonar). Caso atinja vasos sanguíneos ou linfáticos, pode acometer diversos outros órgãos e até o sistema nervoso (forma extrapulmonar) (PAI et al., 2016). A tuberculose humana, geralmente relacionada com os patógenos adaptados aos humanos, é transmitida via aerossol proveniente de hospedeiros infectados e é considerada uma doença crônica, levando anos para se desenvolver (DE JONG; ANTONIO; GAGNEUX, 2010; PAI et al., 2016). Já na tuberculose causada por micobactérias animais do MTBC, a principal transmissão entre animais se dá por via aerossol e contato com secreções de espécies infectadas (MALAMA; MUMA; GODFROID, 2013), e para humanos é por consumo de leite não pasteurizado e carne *in natura* proveniente de animais contaminados e, em menor escala, por via aerossol ao contato com animais infectados (DE LA RUA-DOMENECH et al., 2006; MALAMA; MUMA; GODFROID, 2013). A transmissão de *M. bovis* entre humanos pode ocorrer, mas não é comum, mesmo em casos de doença pulmonar, como ocorre em na infecção por *Mycobacterium tuberculosis*, que é um patógeno altamente adaptado a seres humanos (KANTOR; LOBUE; THOEN, 2010).

Anualmente, aproximadamente 10 milhões de novos casos da doença em humanos são reportados por todo o mundo, e o Brasil ocupa a 18ª posição no *ranking* da doença, com aproximadamente 70 mil casos e quatro mil mortes por ano, conforme a Organização Mundial da Saúde (WHO, 2016). Apesar desses dados serem relacionados à *Mycobacterium tuberculosis* (tuberculose humana), sabe-se que *M. bovis* (tuberculose animal) também afeta seres humanos, mas ainda não é conhecida a real proporção de humanos afetados por essa bactéria, principalmente no Brasil. Estima-se que até 2% dos casos sejam decorrentes dessa espécie em países desenvolvidos (MIGNARD; PICHAT; CARRET, 2006). Já em regiões onde o controle da tuberculose bovina é ausente ou ineficaz, esses números são desconhecidos, e a importância dessa transmissão zoonótica é frequentemente negligenciada (OLEA-POPELKA et al., 2017).

A tuberculose animal, também é amplamente disseminada, sendo o bovino a espécie hospedeira mais acometida, resultando em grandes perdas econômicas para a pecuária (El-Sayed et al., 2016). Destaca-se que animais selvagens são sensíveis ao MTBC, com especial

atenção aos animais mantidos em cativeiro e aos hospedeiros de manutenção de *M. bovis* (MONTALI; MIKOTA; CHENG, 2001; DE LISLE et al., 2002; MICHEL et al., 2006; UNE; MORI, 2007; DE GARINE-WICHATITSKY et al., 2013; PALMER, 2013). No Brasil, apesar de um reservatório selvagem para *M. bovis* nunca ter sido identificado, casos de tuberculose por esse patógeno e *M. tuberculosis* em animais de zoológicos são frequentes e apresentam alta morbidade e/ou mortalidade (ZIMPEL et al., 2017). O PNCEBT tem conseguido controlar a doença em bovinos e bubalinos, sendo que a estimativa de prevalência é aproximadamente 1,3% em rebanhos comerciais (BRASIL, 2006). Porém, as principais fontes de infecções que levam aos surtos nacionais de tuberculose bovina e à manutenção da doença ainda são pouco conhecidas. Outro agravante para o controle da tuberculose bovina é a ocorrência dos reservatórios selvagens que podem disseminar a doença para outros animais e seres humanos (KANEENE et al., 2010; MILLER; SWEENEY, 2013; PALMER, 2013). Felizmente até o momento, reservatórios selvagens de manutenção de *M. bovis* não foram detectados no Brasil (RAMOS et al., 2014), mas a vigilância epidemiológica constante é de primordial importância, uma vez que pouco se sabe sobre a capacidade de transmissão, virulência e persistência das estirpes nacionais de *M. bovis* e seus determinantes genéticos.

2.2 COMPLEXO *MYCOBACTERIUM TUBERCULOSIS*

O MTBC é o causador da tuberculose em animais e humanos. Esse grupo de bactérias possui atividade intracelular, são imóveis, não esporulados, e de crescimento lento em cultura. Possuem uma parede celular rica em lipídeos complexos, como o ácido micólico, sendo então classificadas como álcool-ácido resistentes por coloração de Ziehl-Neelsen (KIESER; RUBIN, 2014). Esse complexo é composto pelos patógenos altamente adaptados a humanos e patógenos adaptados aos animais. *M. tuberculosis* linhagens antiga (1 - Filipinas e Oceano Índico) e modernas (2 - Leste Asiático; 3 - Leste Africano; 4 - Europa, Américas e África), e *Mycobacterium africanum* linhagens 5 e 6 (também caracterizados como *M. africanum* linhagens 1 e 2) fazem parte das micobactérias “humanas”, sendo o primeiro distribuído mundialmente e o principal causador da tuberculose neste hospedeiro e o segundo restrito à região Oeste da África (COSCOLLA; GAGNEUX, 2014; RODRIGUEZ-CAMPOS et al., 2014). Dentre as micobactérias “animais” destacam-se *M. bovis*, detectado principalmente em bovinos, mas encontrado em diversas outras espécies por todo o mundo, *Mycobacterium caprae* em ruminantes domésticos da Europa, *Mycobacterium pinnipedii* em mamíferos marinhos na Oceania, *Mycobacterium microti* em roedores na Grã-Bretanha e os relatados geralmente no continente Africano, sendo *Mycobacterium orygis* em antílopes, *Mycobacterium suricattae* em

suricatos, *Mycobacterium mungi* em mangustos, “dassie bacillus” infectando hírax e “chimpanzee bacillus” (SMITH et al., 2006a; ALEXANDER et al., 2010; COSCOLLA et al., 2013; COSCOLLA; GAGNEUX, 2014; GALAGAN, 2014; RODRIGUEZ-CAMPOS et al., 2014; DIPPENAAR et al., 2015). Dentre essas espécies, as principais estudadas e de maior importância para saúde pública e pecuária são *M. tuberculosis*, amplamente distribuído pelo mundo causando doença em seres humanos, mas podendo também afetar animais em eventos de *spillover*, e *M. bovis*, principal causador de tuberculose em animais, em especial os bovinos.

Outra característica importante dos membros do MTBC é a sua grande similaridade genética oriunda de uma evolução clonal (SMITH et al., 2006a; DOS VULTOS et al., 2008; GALAGAN, 2014). Neste contexto, o sequenciamento do primeiro genoma de *M. bovis* (estirpe AF2122/97) teve como objetivo avaliar essa relação entre os diferentes membros do MTBC, pois acreditava-se que o patógeno ancestral do complexo era *M. bovis* ou um patógeno relacionado a bovinos. Devido a presença de eventos de deleções genômicas em *M. bovis*, concluiu-se que o mesmo se originou de um ancestral de *M. tuberculosis* (GARNIER et al., 2003; GALAGAN, 2014). Esses eventos consistem em deleções de RDs (Tabela 1) e SNPs (RODRIGUEZ-CAMPOS et al., 2014), que permitem a diferenciação entre as espécies do MTBC (Figura 1).

Tabela 1 - Principais Regiões de Diferença (RDs) utilizadas na diferenciação de membros do Complexo *Mycobacterium tuberculosis*

RD	Tamanho (kb)	Genes (quantidade)	Ausência entre membros do MTBC e <i>M. canettii</i>
1	9,5	Família PE/PPE (2) ESX-1 (7)	<i>M. bovis</i> BCG
4	12,7	Proteínas hipotéticas (7) Proteínas de membrana (2) Proteínas conservadas (2) EpiA (1) GmdA (1) Açúcar transferase (1)	<i>M. bovis</i> , <i>M. bovis</i> BCG
9	2,0	Proteínas hipotéticas (1) Piridoxamine 5'-fosfato oxidase (1) Oxidoreductase (1) cobL (1)	<i>M. bovis</i> , <i>M. bovis</i> BCG, <i>M. africanum</i> , <i>M. microti</i> , <i>M. pinnipedii</i> e <i>M. caprae</i>
12	2,8	moaE1 (1) Proteína hipotética (2) Citocromo P450 Cyp141 (1)	<i>M. bovis</i> , <i>M. bovis</i> BCG, <i>M. caprae</i> e <i>M. canettii</i>

Fonte: adaptado de Brosch et al., (2002) e Warren et al., (2006).

Figura 1 - Ilustração evolutiva dos micro-organismos pertencentes ao Complexo *Mycobacterium tuberculosis*

Fonte: (RODRIGUEZ-CAMPOS et al., 2014).

Legenda: Observar escala evolutiva em que os micro-organismos do complexo *Mycobacterium tuberculosis* surgiram a partir do ancestral comum, sendo dos membros do MTBC o mais próximo do ancestral é *M. tuberculosis*. A evolução é baseada na perda de regiões de diferença (RDs) (caixas cinza), também por mutações de nucleotídeos (SNPs) (caixas brancas) e entre parênteses o padrão de perda dos espaçadores do espoligotipo.

Já foi sugerido que *Mycobacterium canettii* também pertenceria ao MTBC. Esse patógeno está relacionado com infecções em humanos, mas é restrito à região do Corno da África e nunca foi descrito afetando animais (ABOUBAKER OSMAN et al., 2015). Todavia, por possuir genoma maior e indícios de sofrer recombinações e transferência horizontal de genes (HGT), que não são descritos como características de membros do MTBC, essa espécie não é agrupada no complexo (SUPPLY et al., 2013). Assim, sugere-se que *M. tuberculosis* é o micro-organismo mais próximo do ancestral comum (GALAGAN, 2014; RODRIGUEZ-CAMPOS et al., 2014) que deu origem ao MTBC e ao *M. canettii*.

Essa particularidade de evolução clonal do MTBC reflete também na dificuldade de

compreensão genotípica dessas micobactérias, sendo que as sequências nucleotídicas dos genomas de membros do MTBC, incluindo o gene do RNA ribossomal 16S (16S rRNA), possuem mais de 99,9% de identidade (GARNIER et al., 2003). Com isso, esses micro-organismos poderiam ser classificados como ecotipos da mesma espécie (SMITH et al., 2006a). Porém, em decorrência de variações fenotípicas, com diferentes níveis de patogenicidade e adaptabilidade à diferentes hospedeiros (DUNN; NORTH, 1995; BROSCHE et al., 2002; MEDINA et al., 2006; RENWICK; WHITE; BENGIS, 2007; NEDELTCHEV et al., 2009), SNPs e RDs, esses patógenos têm sido classificados como diferentes espécies (BROSCH et al., 2002; GARNIER et al., 2003; MOSTOWY et al., 2005; PERIWAL et al., 2015).

Outro agravante da semelhança entre membros do MTBC é a identificação laboratorial do patógeno, uma vez que o diagnóstico clínico, o isolamento em cultura tradicional e a identificação por Ziehl Neelsen não diferenciam as espécies bacterianas (RODRIGUEZ-CAMPOS et al., 2014). Por isso, as RDs, além de permitir estudar as características evolutivas, são amplamente utilizadas para a diferenciação dos micro-organismos pertencentes ao MTBC e também *M. canettii*. Para tal, os padrões de presença e ausência dessas regiões foram utilizados por Warren et al., (2006) que, através de diferentes *primers* (oligo iniciadores) e um fluxograma, permitiu a realização do diagnóstico molecular de maneira eficaz e rápida. Esse protocolo só foi possível em decorrência de sequenciamentos de genomas de espécies do MTBC, que vem como uma ferramenta chave para ajudar a compreender a biologia desses micro-organismos, e também para se identificar novas formas de diagnóstico, controle e até tratamento da tuberculose (GOLBY et al., 2013).

2.3 MYCOBACTERIUM BOVIS

Mycobacterium bovis é considerada uma bactéria generalista, ou seja, uma espécie capaz de infectar múltiplas espécies hospedeiras (BÄUMLER; FANG, 2013). Ao contrário de *M. tuberculosis*, que não persiste em outras espécies que não o homem (i.e. uma espécie especialista), *M. bovis* pode se manter em diversos reservatórios animais, como é o caso dos bovinos, cervídeos de cauda branca (*Odocoileus virginianus*) nos EUA (Michigan), texugos (*Meles meles*) no Reino Unido, cusu (*Trichosurus vulpecula*) na Nova Zelândia, porcos (*Sus scrofa*) selvagens na península Ibérica e búfalos (*Syncerus caffer*) na África do Sul (COLEMAN; COOKE, 2001; DONNELLY et al., 2003; ARANAZ et al., 2004; RENWICK; WHITE; BENGIS, 2007; MILLER; SWEENEY, 2013). *Mycobacterium bovis* é também capaz de participar em eventos de *spillover*, cujos hospedeiros acometidos dificilmente transmitem o patógeno entre indivíduos da mesma espécie, como é no caso de humanos. As razões que levam

a essa variabilidade hospedeira de *M. bovis* ainda permanecem desconhecidas.

A identificação dos fatores de virulência de MTBC pode ser uma chave importante na categorização dos diferentes mecanismos de ação dentre esses micro-organismos. São diversos os fatores de virulência descritos (e.g. sistemas de secreção e de dois componentes, lipoproteínas, proteínas da parede celular) e muitos estão envolvidos principalmente com adaptação do bacilo no macrófago hospedeiro (FORRELLAD et al., 2013). Alguns estudos sugerem que mutações no sistema de dois componentes PhoP/PhoR poderiam explicar porquê as micobactérias animais do complexo perderam a habilidade de serem transmitidas entre humanos (GONZALO-ASENSIO et al., 2014). Para isso avaliou-se a baixa transmissibilidade como sinônimo de virulência, na tentativa de mostrar que *M. bovis* seria uma forma atenuada de *M. tuberculosis*, mas isso ainda é questionado (BEHR; GORDON, 2015). Mutações nesse mesmo sistema também não explicam porque *M. bovis* é capaz de infectar e ser transmitido entre diferentes espécies de animais selvagens, não bovinas, com maior sucesso que entre humanos (e.g. caprinos, ovinos, búfalos, cervídeos, bisões, cusus, texugos, porcos selvagens, dentre outros). Essas também não explicam a adaptabilidade de outras espécies de micobactérias animais, como o *M. caprae*, *M. microti*, *M. orygis* e *M. pinnipedii*, que são encontradas em espécies animais não bovinas, e que em alguns casos, como com o *M. pinnipedii*, foram encontradas causando graves infecções em humanos do Novo Mundo antes da introdução de *M. tuberculosis* nas Américas (BOS et al., 2014). Conclui-se então que a adaptação bacteriana aos diferentes hospedeiros animais é muito mais complexa do que proposto por Gonzalo-Asensio et al. (2014) e pode, inclusive, envolver fatores genéticos ainda desconhecidos das diferentes espécies hospedeiras (BEHR; GORDON, 2015).

Para se compreender as características únicas de *M. bovis*, o sequenciamento de genomas de diferentes estirpes poderá identificar distinções genéticas, quando comparado aos outros membros do MTBC, que podem estar relacionadas a essas variações fenotípicas. Até o momento, 5.285 genomas (completos ou *drafts*) pertencentes ao MTBC estão depositados no banco de dados do GenBank do NCBI (*National Center for Biotechnology Information*), incluindo 5.182 genomas de *M. tuberculosis*. Em relação às estirpes de *M. bovis*, 68 genomas encontram-se disponíveis (30 *M. bovis* virulentos, 37 *M. bovis* BCG e 1 *M. bovis* AN5). Dentre os genomas de *M. bovis* virulentos, apenas três estão completamente sequenciados: *M. bovis* AF2122/97 (GARNIER et al., 2003), *M. bovis* 1595 (KIM et al., 2015), ambos isolados de bovinos, *M. bovis* ATCC BAA-935 (proveniente de *M. bovis* AF2122/97) e *M. bovis* 30 (ZHU et al., 2016) (proveniente da ATCC-19210). Dentre os genomas “incompletos” (*drafts*), 17 são isolados de bovinos, 5 de porcos selvagens, 2 de humanos, 1 de ovino e 1 de macaco. Além

disso, 9 isolados são de origem nacional, os demais englobam Argentina (7), Espanha (3), Coréia do Sul (3), Uganda (2), Reino Unido (2), China (1), Uruguai (1), França (1) e Mali (1). Portanto, ao contrário de *M. tuberculosis*, dados genômicos sobre *M. bovis* virulentos são mais escassos e pouco explorados.

2.3.1 Avaliação da Epidemiologia Molecular de *M. bovis*

Como forma de avaliação da distribuição epidemiológica das estirpes dessa espécie, as metodologias amplamente utilizadas são espoligotipagem (*spolityping*) e número variável de repetições em sequência (VNTR - *Variable Number Tandem Repeat*) e Unidades Repetitivas Micobacterianas (MIRU) (MIRU-VNTR). Ambas são realizadas através da técnica da PCR (Reação em Cadeia da Polimerase), geralmente de maneira complementar e comumente utilizadas em investigações de surtos da tuberculose (ZHANG et al., 2010). A espoligotipagem é baseada no *locus* do CRISPR (*clustered regularly interspaced palindromic repeats*), que possui repetições diretas (DR) com sequências conservadas de 36 pb (pares de base) intercaladas por espaçadores (*spacers*) que variam de 35 a 41 pb, totalizando 43 espaçadores (ZHANG et al., 2010). Essas regiões são amplificadas pela PCR e os produtos hibridizados em membranas de nylon contendo sondas complementares a esses espaçadores. Com o advento do sequenciamento de genomas, a espoligotipagem pode ser alternativamente realizada com as *reads* do sequenciamento genômico, com genomas completos ou *drafts* (genoma incompleto) através dos programas SpolPred e SpoTyping (COLL et al., 2012; XIA; TEO; ONG, 2016). O uso apenas da técnica de espoligotipagem não é aconselhada para epidemiologia molecular, pois eventos homoplásicos fazem com que o mesmo padrão de espoligotipo possa ser encontrado em diferentes linhagens de *M. bovis* (SMITH, 2012). Por sua vez, o MIRU-VNTR, padronizado primeiramente para *M. tuberculosis*, analisa distintos *loci* ao longo do genoma e é baseado em 24 regiões que resultam em diferentes tamanhos em decorrência do padrão de repetições em cada *locus* (SUPPLY et al., 2006). Assim, a implementação dessas duas técnicas conjuntas torna a análise epidemiológica dos isolados mais confiável (SUPPLY et al., 2006).

Outro método de genotipagem de micobactérias é a pesquisa de fragmento de inserção IS6110. Esse é muito utilizado para *M. tuberculosis*, pois essa espécie possui diversas cópias do fragmento de inserção. Por sua vez, *M. bovis* possui apenas 1 ou 2 cópias do IS6110, impossibilitando o uso desta técnica para diferenciação das estirpes (ALLIX et al., 2006; SMITH et al., 2006a; RAMOS et al., 2014).

Recentemente foi proposto a separação das estirpes de *M. bovis* em diferentes linhagens (SMITH, 2012), permitindo a diferenciação desse patógeno em quatro categorias agrupadas

pelos Complexos Clonais: Europeu 1, Europeu 2, Africano 1 e Africano 2. As caracterizações são referentes a presença ou ausência de RDs (RdEu1, RdAF1 e RdAf2) pesquisados com diferentes *primers* por meio da PCR, a identificação de SNPs (gene *guaA*) por meio do sequenciamento gênico e aos diferentes espoligotipos. Esses complexos são distribuídos conforme as distintas regiões geográficas do mundo, sendo Europeu 1 a linhagem mais encontrado nas Ilhas Britânicas, África do Sul, Oceania, Novo Mundo (com exceção do Brasil) e Coréia do Sul (SMITH et al., 2011). Neste estudo, foram utilizados 36 isolados de *M. bovis* do Brasil e apenas 6 (16%) apresentaram as características de Europeu 1. O Complexo Clonal Europeu 2 foi logo caracterizado e identificado como o mais prevalente na Península Ibérica, mas isolados brasileiros não foram avaliados (RODRIGUEZ-CAMPOS et al., 2012). O complexo Africano 1 é encontrado na África Subsaariana (MÜLLER et al., 2009) e o Africano 2 envolve países do Leste Africano (BERG et al., 2011). Assim, não existem estudos compreensivos sobre a distribuição de Complexos Clonais de *M. bovis* no Brasil, e apenas o Europeu 1 foi detectado no território nacional. A pesquisa desses complexos pode ser utilizada para determinar os padrões genotípicos nas diversas regiões do mundo e estabelecer um padrão de linhagens das estirpes de *M. bovis* (SMITH, 2012).

A separação filogenética por diferentes linhagens baseada em SNPs tem sido comumente utilizada para identificar as 7 linhagens descritas de *M. tuberculosis*, em que é possível visualizar as estirpes agrupadas conforme os padrões de mutações (NAMOUCHI et al., 2012; COSCOLLA; GAGNEUX, 2014; PHELAN et al., 2016). No entanto, devido aos poucos estudos de genômica comparativa entre *M. bovis*, essa metodologia ainda não foi implementada para essa espécie.

2.3.2 Genômica comparativa de *M. bovis*

Estudos de genômica comparativa intra-espécie do MTBC são comumente realizados, tanto para avaliação de diferentes linhagens e relações filogenéticas quanto para identificação e caracterização de mutações, deleções, genes de resistência a antimicrobianos, genes de virulência, genes relacionados ao metabolismo microbiano, dentre outros. Essas pesquisas são geralmente realizadas utilizando genomas de *M. tuberculosis* (KATO-MAEDA et al., 2001; ILINA et al., 2013; COSCOLLA; GAGNEUX, 2014; LIU et al., 2014; PHELAN et al., 2016), e as comparações envolvendo *M. bovis* são normalmente resultantes de estudos que compreendem os genomas do MTBC (COLE, 2002; GARCIA PELAYO et al., 2009; JOSHI et al., 2012; MCGUIRE et al., 2012; RUE-ALBRECHT et al., 2014; DE LA FUENTE et al., 2015; PERIWAL et al., 2015).

Ao contrário das diferentes estirpes de *M. tuberculosis*, não houveram trabalhos detalhados de genômica comparativa abrangendo os diferentes genomas de *M. bovis* virulentos. Diversos trabalhos descrevem a utilização de *reads* provenientes do sequenciamento dessa espécie como ferramenta de investigação epidemiológica de surtos de tuberculose de isolados bovinos ou de animais selvagens (BIEK et al., 2012; GLASER et al., 2016; CRISPELL et al., 2017), analisando SNPs e modelos matemáticos de cadeia de transmissão para responder as diversas perguntas epidemiológicas. Outro estudo utilizou *reads* de quatro isolados de *M. bovis* do Reino Unido e seus transcriptomas para estudar padrões de expressão gênica de *M. bovis* (GOLBY et al., 2013). Apenas dois estudos envolveram genomas completos ou *drafts* de *M. bovis* virulentos de maneira comparativa. O primeiro sequenciou três genomas de *M. bovis* e um de *M. caprae*, representativos de diferentes prevalências, graus de lesão e hospedeiros, correlacionando a genômica comparativa com seus fenótipos, o que permitiu observar a maior ocorrência de SNPs em estirpes que induziam maior grau de lesão (DE LA FUENTE et al., 2015). Por sua vez, o segundo consistiu de uma análise filogenética e evolutiva de isolados de *M. bovis* e concluiu que, além de indícios de recombinações em determinados sítios, é possível datar o início da diversificação das linhagens nos EUA (há aproximadamente cinco mil anos), e que as linhagens de evolução mais rápida poderiam estar relacionadas à manutenção da bactéria nos diferentes hospedeiros, podendo levar a novos surtos (PATANÉ et al., 2017). Assim a execução da genômica comparativa compreensiva de genomas completos ou *drafts* de *M. bovis* virulentos, e em comparação com diferentes genomas do MTBC, ainda não foi realizada, e poderá fornecer mais dados para a compreensão da biologia desses patógeno no que se refere a mecanismos de atuação frente aos hospedeiros e ambiente.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Sequenciar e anotar o genoma completo da estirpe SP38 de *Mycobacterium bovis* e compará-lo a outros genomas de *M. bovis* depositados em bancos de dados públicos.

3.2 OBJETIVOS ESPECÍFICOS

- Sequenciar o genoma de *M. bovis* estirpe SP38 a partir de DNA extraído de um primo isolamento;
- Montar, fechar, anotar e depositar o genoma em um banco de dados público;
- Realizar genômica comparativa de *M. bovis* SP38 com genomas do MTBC através de software específicos.

4 MATERIAIS E MÉTODOS

4.1 ISOLAMENTO DA ESTIRPE DE *MYCOBACTERIUM BOVIS* E EXTRAÇÃO DO DNA

Para o sequenciamento do genoma, uma estirpe de *M. bovis* (SP38) foi escolhida aleatoriamente do banco de amostras do Laboratório de Zoonoses Bacterianas do Departamento de Medicina Veterinária Preventiva e Saúde Animal (VPS), Universidade de São Paulo (USP), local onde também foi realizado o primo-isolamento. Essa estirpe foi isolada de uma lesão granulomatosa adquirida de um bovino em abatedouro no Estado de São Paulo em 2010 (MORATO et al., 2016). No estudo de Ikuta (2011), a amostra foi acondicionada em um homogeneizador de amostras, em solução salina 0,85% por 1 minuto. Para descontaminação da amostra, ou seja, inativação de outras bactérias, adicionou-se cloreto de hexadecilpiridínio (HPC) 1,5%, incubou-se a temperatura ambiente por 30 minutos e então centrifugou-se a 3.000 xg por 15 minutos. Após descarte do sobrenadante, o pellet formado foi suspenso em solução de salina estéril 0,85% (AMBROSIO et al., 2008), semeado em meios Stonebrink e Lowenstein-Jensen e incubado a 37 °C por 60 dias (Centro Panamericano de Zoonosis, 1985). Colônias resultantes do crescimento foram identificadas como sendo de *M. bovis* por métodos moleculares (Ho et al. 2004; Warren et al. 2006). Esse primo-isolamento foi armazenado a -20 °C em meio 7H9 até ser utilizado nesse estudo.

O primo-isolado foi descongelado e semeado em meio Stonebrink. Uma colônia do primo-isolado foi então selecionada e novamente propagada em meio Stonebrink para posterior extração do DNA. Após crescimento em incubação, o DNA da estirpe foi extraído de acordo com um protocolo adaptado de Van Soolingen et al., (1994) e Bemer-Melchior; Drugeon, (1999). Primeiramente, para inativação, retirou-se uma alçada de massa bacteriana do meio de cultura que foi transferido para um microtubo de 1,5 mL com 400 µL de tampão TE (10mM Tris-HCl e 1mM ácido etilenodiaminotetracético - EDTA, pH 8,0) e agitado e incubado a 100 °C por 5 minutos. Em seguida, adicionou-se 50 µL de lisozima (10mg/mL) e incubou-se a 37 °C por aproximadamente 18 horas. Após esse período, foi adicionado 5 µL de proteinase K (10mg/mL) e 70 µL de SDS (dodecil sulfato de sódio) a 10%, para agitar e incubar a 65 °C por 10 minutos. Então, acrescentou-se 100 µL de NaCl 5M (cloreto de sódio), pré-aquecido a 65 °C, e 100 µL de solução CTAB (brometo de cetiltrimetilamônio), seguindo de agitação e incubação a 65 °C por 10 minutos. Posteriormente, acrescentou-se 750 µL de clorofórmio:álcool isoamílico (24:1) e o microtubo agitado e centrifugado a 12.000 xg por 7 minutos. Do sobrenadante, 500 µL foi transferido para outro microtubo e acrescentado 500 µL de isopropanol. A amostra ficou mantida a -20 °C por 30 minutos para então ser novamente

centrifugada, 12.000 xg por 15 minutos, e o sobrenadante desprezado para acrescentar 1.000 µL de etanol 70% gelado, seguido de centrifugação a 12.000 xg por 5 minutos. O sobrenadante foi desprezado e, após a secagem do microtubo (56 °C por 10 minutos), o DNA foi ressuspensionado com 50 µL de tampão TE e armazenado em freezer -20 °C.

4.2 BIBLIOTECAS GENÔMICAS E SEQUENCIAMENTO

O DNA extraído foi enviado ao Centro de Genômica da Universidade de Purdue, West Lafayette, Indiana, EUA, para sequenciamento. Para isso, a qualidade e a concentração do DNA foi mensurada por Nanodrop 2000c (Thermo Fischer Scientific, Massachusetts, EUA), resultando em razões das absorvâncias de 260/280 de 2,10 nm e 260/230 de 0,67 nm e concentração de 286,86 ng/µl. Em seguida o DNA foi quantificado em gel de agarose 0,8% com marcador de massa molecular, resultando em uma concentração de aproximadamente 120 ng/µL. Uma análise final foi realizada com DNA High Sensitivity Chip de Agilent 2100 Bioanalyzer (Agilent Technologies, Califórnia, EUA) para verificação da concentração e fragmentação da amostra. Em seguida, a biblioteca genômica foi construída no modelo *paired-end* com o kit TruSeq DNA PCR-free sample preparation kit (Illumina, Califórnia, EUA) de acordo com as instruções do fabricante. Um total de 20% de uma *lane* (Illumina v3 chemistry) do HiSeq2500 (Illumina) foi utilizado para sequenciar a biblioteca genômica.

4.3 MONTAGEM E FECHAMENTO DO GENOMA

Na tentativa de complementar resultados para facilitar o fechamento do genoma, as leituras (*reads*) geradas pelo sequenciador HiSeq2500 foram montadas (*assembled*) *de novo* utilizando software ABySS (SIMPSON et al., 2009) com diferentes *k-mers* (50, 60, 70, 80 e 90) e CLC Genomics Workbench 8 (QIAGEN, Venlo, Holanda) (*De Novo Assembly Paired Data*) de acordo com as instruções do fabricante (CLC bio, 2015). Adicionalmente, essas leituras foram mapeadas contra o genoma de referência *M. bovis* AF2122/97 utilizando a função *Map Reads to Reference* do CLC Genomics Workbench 8 (QIAGEN). Contigs gerados pelo software ABySS foram mapeados e ordenados contra o mesmo genoma de referência utilizando o algoritmo de montagem online Projector2 (VAN HIJUM et al., 2005) (valores padrões com ferramenta BLAT). Os finais 3' e 5' destes *contigs* foram avaliados quanto a possíveis sobreposições utilizando alinhamentos locais de duas sequências Smith-Waterman. Quando as diferentes sequências apresentavam sobreposições claras e sem elementos repetitivos, seguindo de acordo com o genoma de referência, os *contigs* eram manualmente conectados. Para manualmente preencher *gaps* internos contendo múltiplos N's e checar as junções dos *contigs*,

resultados obtidos com a montagem *de novo* pelo CLC Genomics Workbench 8 (QIAGEN) e com o mapeamento das leituras contra o genoma de referência foram utilizados.

4.4 PREDIÇÃO DE GENES E ANOTAÇÃO

Para a identificação de genes e anotação, a sequência do genoma foi submetida ao NCBI *Prokaryotic Genome Annotation Pipeline* (PGAP, http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) para anotação automática. Para visualização do genoma completo de *M. bovis* SP38 na forma de um mapa circular foi utilizado o programa ClicO FS (CHEONG et al., 2015) disponível *online* (<http://clifofs.codoncloud.com>). Esse programa permite a inserção de genoma no formato “gbk”, e em cada círculo formado evidencia os genes nas fitas senso e anti-senso, genes RNAs, distribuição de GC ao longo do genoma e conteúdo de GC (GC-skew) [calculado por $(G-C)/(G+C)$].

4.5 GENÔMICA COMPARATIVA

4.5.1 Escolha dos genomas

Primeiramente, todos os 31 genomas de *M. bovis* virulentos depositados no GenBank foram pesquisados para confirmação das espécies. A técnica descrita por Warren et al., (2006), que se baseia na identificação da presença ou ausência das RDs (1, 4, 9, 12, 1mic e 2seal) para determinação de espécies do MTBC foi utilizada *in silico*. Após essa análise, 28 genomas de *M. bovis* virulentos confirmados foram selecionados para este estudo, incluindo *M. bovis* SP38, em um total de 4 genomas completos e 24 *drafts*. Desses genomas, 10 são provenientes do Brasil, sete da Argentina, três da Espanha, três da Coreia do Sul, e um do Reino Unido, China, Uruguai, Uganda e França e envolvem como hospedeiros 21 bovinos, cinco porcos selvagens, um ovino e um chimpanzé (Apêndice A).

Para os genomas de *M. bovis* BCGs, dos 37 genomas depositados no GenBank, apenas genomas completos e com melhor montagem foram selecionados, evitando genomas com grande número de *contigs*. Assim, cinco genomas foram eliminados por apresentarem grande quantidade de *contigs* ($n > 321$), restando 32 genomas que foram incluídos, dos quais 11 são completos e 21 na forma de *drafts* (Apêndice B).

Em relação aos genomas de *M. tuberculosis*, são mais de cinco mil genomas depositados, sendo 61 completos. Para abranger estirpes geograficamente representativas, foram selecionados 23 genomas completos, correspondendo às linhagens 1 (Filipinas e Oceano Índico; antiga), 2 (Leste Asiático; moderna), 3 (Leste Africano; moderna) e 4 (Europa,

Américas e África; moderna), além de estirpes sensíveis e resistentes à antibióticos (Apêndice C).

Além desses, e apenas para análise filogenômica, foram escolhidos genomas dos demais membros do MTBC, sendo esses: *M. africanum* GM041182, *M. africanum* MAL010070, *M. microti* 12, *M. caprae* MB2, *M. orygis* 112400015, *M. mungi* BM22813 e *M. suricattae* (Apêndice D). Apenas *M. suricattae* não está depositado no GenBank; suas *reads* estão disponíveis no banco de dados europeu ENA (*European Nucleotide Archive*) e a montagem *de novo* foi realizada no CLC Genomics Workbench 8, como descrito na montagem do genoma de *M. bovis* SP38. Finalmente, um total de 91 genomas pertencentes ao MTBC foram incluídos nas análises, que constam descritos nos apêndices A a D.

4.5.2 Marcadores epidemiológicos

Os programas SpolPred (COLL et al., 2012), que utiliza *reads* provenientes do sequenciamento, e SpoTyping (XIA; TEO; ONG, 2016), que permite diferentes formatos de arquivos de inserção como genoma completo, *assembled contigs* e *reads*, foram utilizados para determinação dos espoligotipos, seguindo recomendações descritas. Para *M. bovis* SP38 utilizou-se SpolPred, já para os demais genomas utilizou-se SpoTyping. Os resultados dos padrões de espaçadores foram então inseridos no banco de dados *Mycobacterium bovis* Spoligotype Database (www.mbovis.org) que gerou o padrão do espoligotipo e o número SB.

A verificação de MIRU-VNTR (*Mycobacterial Interspersed Repetitive Unit - Variable Number Tandem Repeats*), foi realizada segundo descrito previamente (SUPPLY et al., 2006). Para isso, os grupos de *primers* referentes aos 24 *loci* foram pesquisados *in silico* no genoma de *M. bovis* SP38 e os tamanhos referentes à cada região foram conferidos com o protocolo. Os resultados foram inseridos juntamente com o padrão do espoligotipo no servidor MIRU-VNTRplus (www.miru-vntrplus.org) (ALLIX-BÉGUEC et al., 2008; WENIGER et al., 2010) que permitiu a análise e comparação dessa estirpe em relação às outras depositadas, visualizando através de dendograma.

4.5.3 Complexo Clonal

A análise dos quatro Complexos Clonais descritos (MÜLLER et al., 2009; BERG et al., 2011; SMITH et al., 2011; RODRIGUEZ-CAMPOS et al., 2012) foi realizada nos 28 genomas de *M. bovis*. Para isso, foram avaliados *in silico*: um par de *primer* para pesquisa de presença ou ausência da RDEu1 e deleção do espaçador 11 no espoligotipo para Complexo Clonal Europeu 1; identificação de SNP no gene *guaA* na posição 3.765.573 do genoma de referência

M. bovis AF2122/97 e deleção do espaçador 21 para Europeu 2; três *primers* para verificação de RDAf1 intacto ou deletado além de deleção de espaçador 30 para Africano 1; e três *primers* para verificação de RDAf2 intacto ou deletado e ausência de espaçadores três a sete para Africano 2 (Tabela 2).

Tabela 2 - Regiões de diferença (RDs) avaliadas para determinação dos Complexos clonais Europeu 1, Africano 1 e Africano 2 de *Mycobacterium bovis*.

Marcador	Perfil	Complexo clonal
RDEu1	Intacto - 1.207 pb; Deletado - 400 pb	Europeu 1 - deleção RDEu1
RDAf1	Intacto - 350 pb; Deletado - 531 pb	Africano 1 - deleção RDAf1
RDAf2	Intacto - 451 pb; Deletado - 711 pb	Africano 2 - deleção RDAf2

Fonte: Elaborado pela autora.

Legenda: RDs: regiões de diferença; pb: pares de base.

Para determiná-los, utilizou-se os padrões de espaçadores juntamente com RDs e mutações, para separação dos genomas de *M. bovis* conforme seus complexos clonais. Para garantir que não ocorresse erros, todos os parâmetros (RDs e espaçadores) foram pesquisados em todos os genomas, evitando sobreposições de resultados que prejudicassem a determinação do complexo.

4.5.4 Parálogos, ortólogos e sintenia gênica

Para identificação de parálogos nos genomas de interesse, a na ferramenta BLASTClust, disponível na plataforma *online* MPI bioinformatics Toolkit (ALVA et al., 2016), foi utilizada selecionando valores de 70% de cobertura e 30% de identidade das proteínas. A ferramenta para comparação de genomas, SYBIL, disponibilizada pelo *Institute for Genome Sciences* (IGS), Universidade de Maryland (EUA), foi utilizada para verificação de genes ortólogos e mapas de sintenia gênica entre as espécies de *M. bovis*, *M. bovis* BCG e *M. tuberculosis*.

Sybil, juntamente com programas associados, identifica grupos de proteínas utilizando uma forma modificada de melhor BLAST recíproco (ALTSCHUL et al., 1990) corrigindo o agrupamento para parálogos (CRABTREE et al., 2007). As comparações incluem a procura de proteínas no modo todos *versus* todos, de maneira *pairwise* (parâmetros de e-value de 1e-05, identidade e cobertura de pelo menos 80% e 70%, respectivamente), no BLASTp, que permite gerar pares de genes com melhores *hits* em cada genoma estudado. Assim, os genes parálogos são agrupados baseando-se no coeficiente de similaridade Jaccard, aplicando um *cutoff* de 0,6 para cada par de proteína. O resultado é o agrupamento de genes parálogos e *singletons* (únicos), e esses são então utilizados para identificar os grupos de genes ortólogos ao longo dos genomas (RILEY et al., 2012). Após a identificação e análise dos ortólogos no presente estudo,

a confirmação de *decay* genômico e sequências de DNA codificante (CDSs) únicos foi realizada utilizando a plataforma *online* BLASTp e tBLASTn, disponível pelo NCBI.

Por sua vez, a construção do mapa de sintenia gênica no SYBIL fundamenta-se no desenho do genoma de maneira linear, dispondo as CDSs em barras verticais. Um genoma de referência é utilizado, e as CDSs são representadas em cores que variam de amarelo à azul, da esquerda para a direita, indicando conservação gênica na comparação com os demais genomas agrupados e também regiões com deleções e rearranjos (RILEY et al., 2012). Contudo, esse programa não consegue lidar com muitos genomas simultaneamente e também não consegue montar um gradiente de sintenia correto utilizando *drafts*. Assim, utilizou-se apenas genomas completos de *M. bovis*, um de *M. bovis* BCG (Pasteur) e um de *M. tuberculosis* (H37Rv), alternando as referências entre *M. tuberculosis* H37Rv e *M. bovis* AF2122/97.

4.5.5 Identificação e análise de sítios polimórficos

A detecção de sítios polimórficos nas comparações foi realizada pelo programa kSNP3, seguindo recomendações descritas (GARDNER; SLEZAK; HALL, 2015). Os mesmos genomas citados anteriormente foram divididos em três grupos para a análise: grupo de *M. bovis* virulentos, grupo de *M. bovis* BCG e grupo de *M. tuberculosis*. Brevemente, o programa *Kchooser*, disponível no pacote do kSNP3, permitiu a detecção do melhor valor de *kmer* para cada grupo, que então foi utilizado para as comparações. Além disso, o kSNP3 possibilitou a classificação dessas mutações, identificando-as como sinônimas ou não sinônimas, e a anotação do gene envolvido. Para as anotações utilizou-se os genomas completos de cada grupo (quatro genomas completos de *M. bovis*, 11 de *M. bovis* BCG e 23 de *M. tuberculosis*) como referências. Após a detecção e anotação dos genes com sítios polimórficos, esses foram categorizados conforme os COGs (*Cluster of Orthologous Groups*) (<https://www.ncbi.nlm.nih.gov/COG>) (TATUSOV et al., 2000).

Com o objetivo de avaliar qual espécie apresentaria maior número de SNPs, dois genomas por vez (*pairwise*) foram inseridos no kSNP3 de forma a comparar todos *M. bovis* BCG entre si (totalizando 496 análises), todos *M. bovis* (377 análises) e todos *M. tuberculosis* (253 análises). Os resultados foram compilados no programa estatístico GraphPad Prism 6 (GraphPad Software Inc, La Jolla, California, EUA) e visualizados em gráfico *box plot*. A análise estatística para comparação desses três grupos foi executada pelo mesmo programa, através do teste não-paramétrico de Kruskal-Wallis com pós teste de Dunn para detecção das diferenças de dois grupos. Valores de $p \leq 0,05$ foram considerados estatisticamente significativos.

Para identificar se os sítios polimórficos ocorrem em genes comuns (homólogos) aos diferentes grupos bacterianos (*M. bovis*, *M. tuberculosis* e *M. bovis* BCG), todas as sequências das proteínas com sítios polimórficos foram selecionadas. Replicatas (i.e. proteínas que apareciam duas ou mais vezes por possuírem mais de um sítio polimórfico) foram removidas para evitar redundância. Essas proteínas foram então inseridas, juntamente, na ferramenta BLASTClust, obedecendo os mesmos parâmetros descritos acima, para identificação de grupos de ortólogos entre as diferentes espécies e proteínas espécie-específicas. A mesma análise foi realizada com proteínas contendo sítios polimórficos não sinônimos. Esses grupos de genes ortólogos e proteínas espécie-específicas foram funcionalmente categorizados de acordo com COGs.

4.5.6 Filogenômica

Para construção da árvore filogenética baseada em SNPs, os genomas de *M. bovis* juntamente com um membro de cada espécie do MTBC foram submetidos ao kSNP3 como descrito acima. O documento disponibilizado após a análise denominado “SNPs_all_matrix.fasta”, que contém as sequências de SNPs, foi utilizado para construir um alinhamento múltiplo no programa MEGA 7 (KUMAR; STECHER; TAMURA, 2016) utilizando algoritmo MUSCLE (EDGAR, 2004). O alinhamento resultante foi utilizado para construir uma árvore filogenética empregando métodos *Neighbor-Joining* (NJ) e máxima verossimilhança (*Maximum Likelihood* - ML), ambos com *bootstrap* de 1.000 e modelo de substituição Jukes-Cantor.

5 RESULTADOS E DISCUSSÃO

5.1 O GENOMA DE *M. BOVIS* SP38, UMA TRADICIONAL MICOBACTÉRIA TUBERCULOSA

Um total de 29.651.856 *reads* contendo 2.858.321.177 bases foram geradas pelo Illumina HiSeq2500. Após filtragem, que permitiu reduzir *reads* repetidas, resultando em 5.930.382 *reads* (571.709.745 bases) que foram utilizadas pelos bioinformatas da Universidade de Purdue para fazer a montagem do genoma em software ABySS. Montagens com diferentes tamanhos de *kmer* (50, 60, 70, 80, 90) foram disponibilizadas (Apêndice E). Dentre eles, o *kmer* de tamanho 90 foi escolhido para as análises, por apresentar melhores valores, como maiores *contigs* e valor N50 alto, em comparação aos demais, assim como recomendado (MÄKINEN; SALMELA; YLINEN, 2012; CHIKHI; MEDVEDEV, 2014). Nesse caso, a montagem resultou em 51 *contigs*, totalizando 4.373.651 bases. Um N50 de 208.012 bases foi obtido, o que significa que 50% das bases do genoma todo estão contidas em *contigs* iguais ou maiores a esse valor. Um n:N50 de 8 também foi obtido, ou seja, 8 sequências apresentaram tamanho de pelo menos 208.012 bases (Apêndice F). Além disso, a montagem apresentou 68 *gaps* visualizados com a presença de "Ns" ao longo dos 51 *contigs* disponibilizados, resultando em 119 *gaps*.

Essa montagem foi então submetida ao Projector2 (VAN HIJUM et al., 2005), alinhando-a contra o genoma de referência *M. bovis* AF2122/97, gerando um mapa de orientação da ordem dos *contigs* que revelou possíveis sobreposições. A utilização de alinhamentos Smith-Waterman e análises em nBLAST contra o genoma de referência, aliado ao alinhamento manual entre os próprios *contigs*, resultaram no fechamento da maioria dos *gaps*, restando 5 *gaps*. Esses últimos foram fechados com a utilização de outras duas montagens no software CLC Genomics Workbench, *de novo* e alinhamento das *reads* contra o genoma de referência *M. bovis* AF2122/97. Genomas de micobactérias são difíceis de serem fechados devido às áreas repetitivas e ao alto conteúdo de GC (> 60%) (KUMAR; KAUR, 2014). Neste contexto, a utilização de múltiplos algoritmos de montagem facilitou o fechamento do presente genoma, uma vez que não existe uma montagem dita como ótima (WENCES et al., 2015).

Dessa maneira, um genoma único, circular e completo de 4.347.646 pb foi finalmente obtido para *M. bovis* SP38. O genoma completo dessa estirpe foi submetido ao GenBank para atualizar o genoma incompleto (GUIMARAES et al., 2015) e disponibilizar para acesso público sob código de acesso NZ_CP015773.1. A anotação fornecida pela PGAP/NCBI revelou um genoma típico de micobactérias tuberculosas, com alto conteúdo de GC (65,6%) e 4.216 genes, incluindo 154 pseudogenes, 3 genes de rRNA (RNA ribossomal), 45 de tRNA (RNA transportador), 2 de ncRNA (RNA não codificante), 1 tmRNA (RNA transferência-mensageiro)

e 4.011 CDSs. Dentre as CDSs, a maioria (2.805/4.011; 69,93%) possui função definida e 1.206 (30,07%) são classificados como hipotéticos. A distribuição das CDSs, genes tRNA e genes rRNAs nas fitas senso e anti senso do genoma de *M. bovis* SP38 pode ser visualizada na Figura 2. A distribuição negativa e positiva do GC-skew confirma a possível origem de replicação (oriC) próximo a base 1+ deste genoma, que corresponde a primeira base do gene da *dnaA*, e ao sítio de terminação da replicação (TerC) no lado oposto. Uma breve comparação (Tabela 3) das características gerais deste genoma com os genomas de referência de *M. tuberculosis* (estirpe H37Rv) e *M. bovis* (estirpe AF2122/97) confirma a grande semelhança da estirpe SP38 com genomas de micobactérias tuberculosas.

Figura 2 - Mapa circular do genoma completo de *Mycobacterium bovis* estirpe SP38

Fonte: Acervo pessoal.

Legenda: Representação gráfica do genoma completo de *M. bovis* SP38. Base 1+ do gene *dnaA* está na posição 1. De fora para dentro: círculo 1 (marrom): numeração das bases; círculo 2 (azul): CDSs (sequência de DNA codificante) em fita senso; círculo 3 (roxo): CDSs em fita anti-senso; círculo 4: tRNAs (RNA transportadores) (barras laranja sobre o círculo amarelo); círculo 5: rRNA (RNA ribossomais) (barras vermelhas sobre o círculo laranja); círculo 6: conteúdo de GC (verde); círculo interno: distribuição de GC (GC-skew) (vermelho/preto).

Tabela 3 - Comparação das características gerais dos genomas completos de *Mycobacterium bovis* SP38, *M. bovis* AF2122/97 e *M. tuberculosis* H37Rv

Características	<i>M. bovis</i> SP38	<i>M. bovis</i> AF2122/97	<i>M. tuberculosis</i> H37Rv
Tamanho do genoma (pb)	4.347.646	4.345.492	4.411.532
Conteúdo GC (%)	65,6	65,6	65,6
Nº de genes	4.216	4.001	4.008
Nº de tRNAs	45	45	45
Nº de genes 5S rRNA	1	1	1
Nº de genes 16S rRNA	1	1	1
Nº de genes 23S rRNA	1	1	1
CDSs	4.011	3.918	3.906
CDSs com função	2.805 (69,93%)	2.321 (59,23%)	2.851 (72,99%)
CDSs hipotéticas	1.206 (30,07%)	1.597 (40,76%)	1.055 (27,00%)
CDSs em PGFs	1.049	1.062	1.119
Nº of PGFs (membros)	307 (2-100)	304 (2-113)	324 (2-110)
PGFs (CDSs hipotéticas)	135 (12,86%)	299 (28,15%)	143 (12,78%)
PGFs (CDSs com função)	914 (87,13%)	763 (71,84%)	976 (87,22%)

Fonte: Elaborada pela autora.

Legenda: pb: pares de base, tRNA: RNA transportador; rRNA: RNA ribossomal; CDSs; sequência de DNA codificante, PGF: família de genes parálogos.

O espótipo de *M. bovis* SP38 demonstrou ser o SB0121, descrito como o mais prevalente do Brasil (ROCHA, 2013; CARVALHO et al., 2016) e o padrão MIRU-VNTR confirmou essa bactéria como sendo *M. bovis*, como descrito no Apêndice G.

5.2 GENÔMICA COMPARATIVA

5.2.1 Nem todos os genomas depositados no GenBank são *M. bovis* virulentos

A análise da presença ou ausência de RDs em genomas de *M. bovis* depositados no GenBank mostrou que o genoma *M. bovis* ATCC BAA-935, descrito como proveniente da estirpe de *M. bovis* AF2122/97, apresentou padrões de RD1, RD4 e RD9 ausentes, características que classificam como *M. bovis* BCG. Já *M. bovis* B2 7505, um isolado de humano (WANZALA et al., 2015), possui padrões RD1, RD4, RD9 e RD12 presentes, equivalendo à um genoma de *M. tuberculosis*. Por fim, *M. bovis* MAL010093, também isolado de humano, exibiu RD1, RD4, RD12 e RD1mic presentes, RD9 e RD2seal ausentes, sendo classificado como *M. africanum* (Apêndice H). Dessa forma, dos 31 genomas primeiramente selecionados, apenas 28 permaneceram nas análises. Esses dados evidenciam que a classificação das micobactérias deve ser criteriosa e padronizada nos diversos laboratórios. Para evitar erros de depósito em bancos de dados, a confirmação dos genomas é necessária, sugerindo programas como RD-analyzer criado por Faksri et al. (2016), que utiliza as *reads* para identificar os diversos padrões de RDs. A especiação possivelmente equivocada desses isolados levou Patané et al. (2017) a sugerir que os primeiros isolados de *M. bovis* poderiam ter surgido em humanos na África e que *M. bovis* ATCC BAA-935, uma BCG, possui áreas de

recombinação genética e polimorfismos de sequência grande quando comparado a *M. bovis* virulentos. Esses resultados devem ser avaliados com cuidado, pois a presença de extensas regiões de recombinação é frequentemente considerada ausente em membros do MTBC por diversos autores (GUTIERREZ et al., 2005; SMITH et al., 2006a, 2006b; PEPPERELL et al., 2013; GALAGAN, 2014).

5.2.2 O espoligotipo mais comum dos genomas de *M. bovis* virulentos depositados no GenBank é o SB0140

Dezenove (67,86%) dos 28 genomas de *M. bovis* depositados no GenBank possuem o espoligotipo SB0140. Os espoligotipos SB0339 (2/28, 7,14%), SB0120 (1/28, 3,57%), SB0145 (1/28, 3,57%), SB0134 (1/28, 3,57%) e SB0121 (1/28, 3,57%; estirpe SP38) também foram encontrados (Tabela 4). A classificação realizada nesse estudo não identificou os padrões de cinco isolados de *M. bovis*: três genomas que resultaram em padrões ainda desconhecidos (*M. bovis* B_322, *M. bovis* Bz 31150 e *M. bovis* 30) e outros dois (*M. bovis* MB1 e *M. bovis* MB3), que puderam ser então identificados por meio de publicação prévia (DE LA FUENTE et al., 2015). Duas possibilidades podem explicar o porquê esses espoligotipos não puderam ser identificados. Primeiro, o programa SpoTyping, que se baseia em concatenar a sequência dos espaçadores ao longo dos *contigs* (XIA; TEO; ONG, 2016), pode não ter encontrado essas regiões na análise dos *drafts* dos genomas. Segundo, é a existência de padrões ainda não descritos.

Tabela 4 - Padrões binários, OCTcode e SB referentes aos espoligotipos dos genomas de *Mycobacterium bovis* depositados no GenBank

(continua)			
Genoma	Padrão binário	Padrão OCTcode	Padrão SB
<i>M. bovis</i> AF2122/97	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> SP38	1101111101111101111011111111111111110000	67677367777600	SB0121
<i>M. bovis</i> 1595	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 30	1101011101011101101111111101111111110000	65657337737600	?
<i>M. bovis</i> Bz 31150	00000000000000000000111111000000000010000	000000176000200	?
<i>M. bovis</i> 04-303	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 09-1191	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 05-566	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 05-567	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 49-09	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 32-08	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 18-08C	1101101000001110111111111111111111110000	66407377777600	SB0140
<i>M. bovis</i> 35	1101101000001110111111111000011111110000	664073776077600	SB0140
<i>M. bovis</i> 08-08BF2	1101101000001110111111111111111111110000	66407377777600	SB0140

5.2.3 Complexo clonal Europeu 2 é encontrado no Brasil

Os Complexos Clonais dos genomas de *M. bovis* depositados no GenBank encontram-se listados na tabela 5. Não foi possível classificar apenas dois genomas, *M. bovis* D_10_02315 e *M. bovis* MB3, ambos isolados de porcos selvagens na França e Espanha, respectivamente. Além disso, entre os genomas, não se detectou a presença do complexo clonal Africano 1, possivelmente devido a baixa representatividade geográfica da amostra.

Tabela 5 - Resultados dos perfis que determinaram os Complexos Clonais dos genomas de *Mycobacterium bovis* depositados no Genbank

Genoma	Perfil	Complexo clonal
<i>M. bovis</i> AF2122/97	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> SP38	Mutação gene guaA e ausência espaçador 11, 21 e 30	Europeu 2
<i>M. bovis</i> 1595	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 30	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> Bz 31150	RDAf1 ausente (711 pb) e ausência espaçador 3 a 7, 11 e 30	Africano 2
<i>M. bovis</i> 04-303	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 09-1191	RDEu2 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 05-566	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 05-567	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 49-09	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 32-08	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 18-08C	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 35	Mutação gene guaA e ausência espaçador 11	Europeu 2
<i>M. bovis</i> 08-08BF2	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 09-1193	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 534	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 0822-11	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 61-09	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 45-08b	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 09-1192	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> 50	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> W-1171	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> MbURU-001	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> MB4	Mutação gene guaA e ausência espaçador 21	Europeu 2
<i>M. bovis</i> B-3222	RDEu1 ausente (401 pb) e ausência espaçador 11	Europeu 1
<i>M. bovis</i> D_10_02315	RDEu1 intacto (1.207 pb), RDAf1 intacto (350 pb), RDAf2 intacto (451 pb), sem mutação no gene guaA, ausência espaçadores 3 a 7, 11, 30 e presença do 21	?
<i>M. bovis</i> MB1	Mutação gene guaA e ausência espaçador 11, 21 e 30	Europeu 2
<i>M. bovis</i> MB3	RDEu1 intacto (1.207 pb), RDAf1 intacto (350 pb), RDAf2 intacto (451 pb), sem mutação no gene guaA	?

Fonte: Elaborada pela autora.

Legenda: pb: pares de base; ?: complexo não determinado.

A estirpe SP38 de *M. bovis*, sequenciada no presente estudo, e a estirpe 35 de *M. bovis*

constituem o primeiro relato do Complexo Clonal Europeu 2 no Brasil. Contrariamente ao descrito por Smith et al., (2011), em que apenas 6 (16,66%) dos 36 isolados de *M. bovis* do Brasil foram identificados como pertencentes ao Complexo Clonal Europeu 1, nesse estudo esse complexo foi o mais encontrado em isolados de *M. bovis* brasileiros depositados no GenBank (8 de 10 isolados). Porém, convém ressaltar que esses dados devem ser interpretados com muita cautela, uma vez que essa é uma amostra viciada e pouco representativa. Mais estudos devem ser realizados para confirmar a distribuição dos Complexos Clonais no território nacional. Até o momento, pode-se concluir que os Complexos Clonais Europeus 1 e 2 são encontrados no país.

De maneira geral, os Complexos Clonais apresentam certa correlação com os espoligotipos. Apesar do SB0140 ser relacionado com o Europeu 1 (SMITH et al., 2011), variações nos padrões de espoligotipos têm sido descritas dentro de cada complexo clonal (MÜLLER et al., 2009; BERG et al., 2011; SMITH et al., 2011; RODRIGUEZ-CAMPOS et al., 2012), sendo que apenas os espoligotipos não podem ser usados como características únicas para essa categorização. Ressaltando isso, uma estirpe com esse mesmo número de SB (*M. bovis* 35) que foi enquadrada nas características do Europeu 2 no presente estudo. Além disso, a não determinação dos complexos em dois isolados nos permite o questionamento se há outras linhagens que ainda não foram descritas ou se o padrão de investigação dos complexos não representa todas as estirpes e deveriam ser revistos.

5.2.4 Regiões sem sintenia gênica são observadas entre os genomas de *M. bovis* virulentos

Grandes regiões de recombinação e/ou transversão são consideradas ausentes em membros do MTBC. Além do mais, esses micro-organismos possuem grande identidade genética, onde regiões alinháveis dos genomas são > 99,95% idênticas (FROTHINGHAM; HILLS; WILSON, 1994; BROSCHE et al., 2002; GARNIER et al., 2003; SMITH et al., 2006a, 2006b). Assim, foi possível confirmar uma acentuada conservação da sintenia gênica entre os genomas completos de *M. bovis* virulentos e genomas de referência de *M. tuberculosis* e *M. bovis* BCG (Figura 3A). Porém, observam-se blocos sem sintenia gênica, sugestivo de regiões de polimorfismos de grandes sequências (LSP, *large sequence polymorphisms*; deleções) ou ausência de genes ortólogos (e.g. por truncamento prematuro de genes) entre espécies de *M. bovis* em relação a *M. tuberculosis* H37Rv (Figura 3A) e entre as espécies de *M. bovis* virulentos (Figura 3B). Dessas, *M. bovis* 30 obedece a um padrão ligeiramente diferente dos outros *M. bovis* virulentos. Dentre deleções já conhecidas, destacam-se aquelas previamente descritas

como RD1, caracteristicamente ausente em *M. bovis* BCG, e RD4, ausente em micobactérias animais (Figura 3A). Essas regiões sem sintenia gênica evidenciam o processo evolutivo de perda de genes sofrido por *M. bovis* em relação a *M. tuberculosis*, concentrando-se na região 2,2 Mb a 4,3 Mb (Figura 3A). Mais estudos devem ser feitos para analisar cada uma dessas regiões individualmente e descrever as CDSs presentes nas mesmas, de maneira a inferir distintos perfis de linhagens de *M. bovis*, como já descrito para *M. tuberculosis*, em que as diferentes linhagens apresentam distintos perfis de RDs assim como mutações em determinados genes (RODRIGUEZ-CAMPOS et al., 2014).

Figura 3 - Gradiente de sintenia com genomas de *Mycobacterium bovis*, *M. bovis* BCG e *M. tuberculosis*

Fonte: Acervo pessoal.

Legenda: O gradiente baseia-se em cores, sendo amarelo (início) ao azul (final), cor branca significa região ausência de gene ortólogo no genoma de referência. (A) Utilizando na comparação *M. tuberculosis* H37Rv (Mtb H37Rv) como referência. Mb: estirpes de *Mycobacterium bovis*; BCG Pasteur: *Mycobacterium bovis* BCG Pasteur 1173P2; Flecha preta significa RD4, ausente em todos *M. bovis* e *M. bovis* BCG e flecha vermelha RD1 ausente no genoma de *M. bovis* BCG. (B) Comparação entre os quatro genomas completos de *M. bovis*, utilizando *M. bovis* AF2122/97 como referência.

5.2.5 Grupos de genes ortólogos

5.2.5.1 Genoma *core* de *M. bovis* apresenta um grande número de CDSs de função desconhecida

O genoma *core* de *M. bovis*, isto é, CDSs presentes em todos 28 genomas de *M. bovis* analisados, é constituído de 1.491 grupos de genes ortólogos. Esses grupos foram classificados de acordo com COG, listados na tabela 6.

Tabela 6 - Genes ortólogos presentes nos 28 genomas *Mycobacterium bovis* categorizados pelos COGs (*Cluster of Orthologous*)

(continua)	
COGs	<i>M. bovis</i>
Controle do ciclo celular, divisão celular, partição cromossômica	10 (0,67%)
Biogênese de parede celular / membrana / envelope	38 (2,54%)
Motilidade celular	1 (0,06%)

(conclusão)	
COGs	<i>M. bovis</i>
Modificação pós-translacional, turnover proteico e chaperonas	28 (1,87%)
Mecanismos de transdução de sinal	17 (1,14%)
Tráfego intracelular, secreção e transporte vesicular	6 (0,40%)
Mecanismos de defesa	24 (1,60%)
Estruturas extracelulares	1 (0,06%)
<i>Mobilome</i> : transposons e profagos	14 (0,93%)
Processamento e modificação do RNA	1 (0,06%)
Tradução, estrutura ribossomal e biogênese	31 (2,07%)
Transcrição	69 (4,62%)
Replicação, recombinação e reparação	31 (2,07%)
Produção e conversão de energia	78 (5,23%)
Transporte e metabolismo de aminoácidos	50 (3,35%)
Transporte e metabolismo de nucleotídeos	21 (1,40%)
Transporte e metabolismo de carboidratos	34 (2,28%)
Transporte e metabolismo de coenzimas	41 (2,74%)
Transporte e metabolismo de lipídios	69 (4,62%)
Transporte e metabolismo de íons inorgânicos	41 (2,74%)
Biossíntese de metabólitos secundários, transporte e catabolismo	40 (2,68%)
Somente predição de função geral	89 (5,96%)
Função desconhecida	799 (53,58%)
Total de grupos de genes ortólogos	1.491

Fonte: Elaborada pela autora.

Nota: O total de COGs é ligeiramente maior que o número de grupos porque uma proteína pode estar envolvida em mais de um COG.

A maioria dos grupos de ortólogos do genoma *core* de *M. bovis* está categorizado como função desconhecida (53,58%), sendo representados por proteínas hipotéticas e das famílias PE e PPE (prolina-glutamato e prolina-prolina-glutamato, respectivamente). Semelhante a *M. tuberculosis* (VOSKUIL et al., 2004), 3,54% das CDSs (142/4.011) são anotados como pertencentes às famílias PE e PPE. Acredita-se que essas famílias de parálogos sejam importantes fatores de virulência, atuando na evasão do sistema imune por meio da variabilidade genética (FISHBEIN et al., 2015; PHELAN et al., 2016). Em seguida destacam-se os genes envolvidos em funções de *housekeeping*, como produção e conservação de energia (5,23%), citados como fundamentais para o momento da infecção no hospedeiro em análises de transcriptoma (SHARMA et al., 2017), transporte e metabolismo de lipídeos (4,62%), que contém genes essenciais para a sobrevivência nos macrófagos (FISHER; PLIKAYTIS; SHINNICK, 2002; RUSSELL et al., 2010), e transcrição (4,62%), que são primordiais para manutenção bacteriana no hospedeiro (RAMAN et al., 2004).

5.2.6 Pesquisa de *singletons* de *M. bovis* demonstra *decay* genômico

A plataforma SYBIL identificou 1.718 proteínas como sendo provenientes de genes únicos, isto é, genes presentes em apenas 1 dos 28 genomas de *M. bovis* (i.e. sem ortólogos). Destes, 1.416 (82%) haviam sido anotados como pseudogenes pelo GenBank, e assim sua sequência proteica não é disponibilizada para análise. Desta forma, pseudogenes não entram no agrupamento de ortólogos. Das 302 proteínas restantes, análises de BLASTp contra membros do MTBC (incluindo *M. bovis*) indicam que, apesar da alta identidade da sequência, 249 (82,45%) apresentam um tamanho menor que outras proteínas de membros do complexo, impedindo seu grupamento como ortólogas. Por fim, as 53 proteínas restantes não apresentaram *hits* significativos no BLASTp contra outros membros do MTBC (Tabela 7). No entanto, no tBLASTn (proteína *versus* DNA traduzido), todas as suas respectivas sequências gênicas foram encontradas no genoma de membros do MTBC. Pode-se concluir que essas proteínas são então produtos oriundos de mutações *frameshift* ou classificadas como pseudogenes nos outros genomas do complexo. Tais características são sugestivas de um processo de *decay* genômico (perda de genes) e/ou pseudogenização intra-espécie de *M. bovis* e em comparação a outros membros do MTBC. Esses processos têm sido descritos em *M. tuberculosis* e associados ao comportamento especialista deste patógeno em relação ao hospedeiro (BÄUMLER; FANG, 2013; BOLOTIN; HERSHBERG, 2015). O impacto desta remodelagem genômica na virulência e adaptabilidade hospedeira do MTBC é desconhecido. Convém ressaltar que a anotação de pseudogenes pode ser maior em genomas na forma de *drafts*, pois as CDSs localizadas nas extremidades dos *contigs* e sem *stop codons* acabam sendo anotados como pseudogenes. Uma análise mais aprofundada do genoma acessório de *M. bovis* permitirá identificar o impacto do processo de *decay* genômico nesta espécie bacteriana.

Tabela 7 - Proteínas que não apresentam homólogos e são encontradas apenas em um dos 28 genomas de *Mycobacterium bovis*

(continua)				
COGs	Genoma encontrado	Proteínas	Número de acesso	Tamanho (aa)
Biogênese de parede celular / membrana / envelope	<i>M. bovis</i> 30	ABC de glutamina de ligação de ATP	AMC53313.1	76
Mecanismos de transdução de sinal	<i>M. bovis</i> Bz 31150	Proteína de resposta à hipóxia	KAN85348.1	221
Produção e conversão de energia	<i>M. bovis</i> AF2122/97	Proteína de transporte de manganês MntH	NP_854605.1	684
	<i>M. bovis</i> 30	Uroporfirina-III C-metiltransferase	AMC53783.1	923
	<i>M. bovis</i> 30	Açúcar quinase	AMC53937.1	526

COGs	Genoma encontrado	Proteínas	Número de acesso	(conclusão)
				Tamanho (aa)
Biossíntese de metabólitos secundários, transporte e catabolismo	<i>M. bovis</i> 30	Policetídeos sintase	AMC54895.1	85
	<i>M. bovis</i> 30	Policetídeos sintase	AMC54894.1	50
Somente predição de função geral	<i>M. bovis</i> 30	Proteína rica em alanine/arginina	AMC56345.1	71
	<i>M. bovis</i> 30	Translocase subunidade SecF	AMC56017.1	837
	<i>M. bovis</i> 30	Lipoproteína LpqB	AMC56699.1	200
	<i>M. bovis</i> 30	Proteína Mce4F	AMC56970.1	78
	<i>M. bovis</i> 30	Translocase	AMC56019.1	51
Função desconhecida	Diversos	Hipotéticas (n=41)	Diversos	Diversos

Fonte: Elaborada pela autora.

Legenda: aa: aminoácidos.

5.2.6.1 Grupos de genes ortólogos entre os genomas de *M. bovis*, *M. bovis* BCG e *M. tuberculosis*: não existem CDSs exclusivamente presente em todos os genomas de *M. bovis*

Os diagramas de venn (Figura 4) ilustram os grupos de genes ortólogos presentes em todos os genomas de *M. bovis*, *M. bovis* BCG e *M. tuberculosis* analisados. Grupos de genes ortólogos presentes exclusivamente em todos os genomas de *M. bovis* não foram detectados (Figura 4A,B). É interessante também observar que há mais grupos de genes ortólogos em comum a todos genomas de *M. bovis* e de *M. tuberculosis* quando genomas de *M. bovis* BCG são excluídos (Figura 4A,B). A mesma análise utilizando apenas genomas completos dos três grupos bacterianos também demonstrou um aumento no número de grupos de ortólogos ao considerar apenas *M. bovis* virulentos e *M. tuberculosis* (dados não mostrados). Assim, é possível sugerir que o processo de atenuação da *M. bovis* BCG levou a perda de CDSs em comparação aos *M. bovis* virulentos e *M. tuberculosis*.

O único grupo de genes ortólogos encontrado exclusivamente entre *M. bovis* e *M. tuberculosis* (Figura 4A) e ausente em *M. bovis* BCG consiste na proteína ESAT-6. *Mycobacterium bovis* BCG é uma forma atenuada de *M. bovis* devido a deleção RD1. Essa região possui aproximadamente 9 kb e codifica nove CDSs que compreendem proteínas da família PE/PPE e sistema de secreção do importante fator de virulência ESAT-6 (MAHAIRAS et al., 1996; BROSCHE et al., 2002). Além disso, *M. tuberculosis* apresentou 36 grupos de genes ortólogos que não estavam presentes em *M. bovis* e *M. bovis* BCG (Figura 4B, Tabela 8). Catorze (38,89%) dos 36 grupos de genes ortólogos incluem CDSs da região RD4, que possui tamanho aproximado de 12 kb e contém proteínas hipotéticas, de membrana e EpiA e GmdA (BROSCHE et al., 2002, 2007) (Tabela 8). Em decorrência da deleção de RD4 estar entre as últimas descritas evolutivamente, essa região é encontrada nos demais membros do complexo, exceto em *M. bovis* virulentos e *M. bovis* BCG (BROSCHE et al., 2002; SMITH et al., 2006a,

2006b). Os demais grupos de genes ortólogos foram então analisados por BLASTp contra *M. bovis*, evidenciando que três grupos, fator anti-anti-sigma RsfB, ESAT-6 e um citocromo P450, eram encontrados em *M. bovis*, mas com menor tamanho. Os demais grupos de ortólogos não apresentaram *hits* significativos no BLASTp. Tais observações demonstram mais uma vez possível processo de *decay* genômico e/ou pseudogenização de *M. bovis* em relação ao *M. tuberculosis*. Dentre esses grupamentos ortólogos, a deleção das CDSs do operon 3 da família Mce (*mammalian cell entry*) nos genomas de *M. bovis* chama atenção pois este é composto por invasinas/adesinas envolvidas na invasão e sobrevivência de *M. tuberculosis* nos macrófagos humanos (HAILE et al., 2002; AHMAD et al., 2004).

Figura 4 - Diagrama de Venn demonstrando compartilhamento de genes ortólogos entre *Mycobacterium bovis*, *M. bovis* BCG e *M. tuberculosis*

Fonte: Acervo pessoal.

Legenda: *Mycobacterium bovis* BCG (traçado), *M. bovis* (cinza claro) e *M. tuberculosis* (cinza escuro).

A: Compartilhamento de genes ortólogos analisando *M. bovis*, *M. bovis* BCG e *M. tuberculosis*

B: Compartilhamento de genes ortólogos analisando *M. bovis* e *M. tuberculosis*.

Tabela 8 - Número de grupos de proteínas homólogas encontrados unicamente nos genomas de *Mycobacterium tuberculosis*

Nome (número do gene)*	Número de grupos de ortólogos	Presente na RD4
Proteínas hipotéticas ou PE/PPE	15	Sim (n=12)
GDP-D-manose-desidratase GmdA (Rv1511)	1	Sim
Epimerase nucleotídeo-açúcar EpiA (Rv1512)	1	Sim
Fator anti-anti-sigma RsfB (Rv3687c)	1	-
6-fosfogluconato desidrogenase Gnd1 (Rv1844c)	1	-
Citocromo P450 (Rv1256c; Rv3518c; Rv3121)	3	-
Proteína de regulação transcricional (Rv1255c)	1	-
ESAT-6 EsxF (Rv3905c)	1	-
Profago (Rv2655c; Rv2658c; Rv2659c)	3	-
Família Mce (Rv1971; Rv1966; Rv1969; Rv1970; Rv1967; Rv0176; Rv1973)	7	-
Oxidoredutase (Rv2073c)	1	-
Piridoxamina 5-fosfato oxidase (Rv1155_F420)	1	-
Total de grupos de genes ortólogos	36	-

Fonte: Elaborada pela autora.

Legenda: * Número de acesso representativo da estirpe de *M. tuberculosis* H37Rv; PE/PPE: prolina-glutamato e prolina-prolina-glutamato.

5.2.7 Sítios Polimórficos

5.2.7.1 Estirpes de *M. tuberculosis* possuem um maior número de sítios polimórficos quando comparado a estirpes de *M. bovis* virulentos e *M. bovis* BCGs

Um total de 3.448, 1.088 e 8.335 sítios polimórficos foram detectados quando comparando todos os genomas de *M. bovis* virulentos, *M. bovis* BCG e *M. tuberculosis*, respectivamente. Análises *pairwise* de cada grupo bacteriano indicam que o número de sítios polimórficos é maior em *M. tuberculosis* seguido de *M. bovis* e *M. bovis* BCG ($p \leq 0,05$) (Figura 5). Mesmo com seu comportamento generalista, genomas de *M. bovis* apresentaram uma menor variabilidade genética quando comparado aos genomas de *M. tuberculosis* neste estudo. Uma explicação provável para esses valores é decorrente da amostragem, uma vez que a representatividade de genomas de *M. tuberculosis* escolhidos foi maior (distribuídas em diversos países), compreendendo as linhagens 1 a 4 (antigas e modernas) e fenótipos de susceptibilidade e resistência aos antimicrobianos. Porém, é também possível que *M. tuberculosis* apresente uma maior variabilidade genética por estar exposto a populações humanas de diferentes fundos (*background*) genéticos, que permitiu a co-evolução dessa micobactéria com esse hospedeiro (GAGNEUX, 2012), aumentando a pressão seletiva e consequente taxa de mutações. O sequenciamento de genomas de *M. bovis* isolados de diferentes reservatórios animais, que não somente bovinos, poderá auxiliar no esclarecimento desta diferença. Por fim, o menor número de sítios polimórficos em *M. bovis* BCG evidencia que essa estirpe vacinal está sujeita a uma menor pressão de seleção que *M. bovis* virulentos e *M. tuberculosis*, como esperado. Mesmo assim, a presença de 1.088 sítios polimórficos demonstra a presença de uma variabilidade genética que pode refletir no sucesso vacinal das diferentes estirpes (GARCIA PELAYO et al., 2009).

Figura 5 - Padrão de distribuição de sítios polimórficos com avaliação *pairwise* entre genomas de *Mycobacterium bovis* BCG, *M. bovis* e *M. tuberculosis*

Fonte: Acervo pessoal.

Legenda: Na análise de sítios polimórficos nos 32 genomas de *M. bovis* BCG foi obtido uma média de 76,83 sítios e uma mediana de 69. Nos 28 genomas de *M. bovis* foi obtido uma média de 381,8 e uma mediana de 378,5. Nos 23 genomas de *M. tuberculosis* foi obtido uma média de 1.192 e uma mediana de 1.134,2. Comparação das diferenças de SNPs entre os três grupos de espécies (***) p-valor: 0,0001). Análise estatística realizada por Kruskal-Wallis, assumindo p-valor $\leq 0,05$.

5.2.7.2 Anotação funcional dos sítios polimórficos: *M. bovis* BCG está sujeito a pressões evolutivas distintas daquelas de micobactérias virulentas.

Dos 3.448 sítios polimórficos de *M. bovis*, 1.658 (48,08%) sítios apresentavam mutações sinônimas e 1.790 (51,92%) apresentavam mutações não sinônimas (Tabela 9). A anotação permitiu também visualizar que desses sítios polimórficos, 2.804 (81,32%) estavam contidos em proteínas, 342 (9,92%) eram presentes em genomas não incluídos como referências, 24 (0,70%) em regiões não codificadoras de proteínas e 278 (8,06%) em regiões não anotadas (genes rRNA e tRNA).

Tabela 9 - Categorização e as determinações de mutações sinônimas e não sinônimas dos sítios polimórficos encontrados em *Mycobacterium bovis* de acordo com as funções dos COGs (*Cluster of Orthologous*)

(continua)

COGs	<i>M. bovis</i>	Sinônimo	Não sinônimo
Controle do ciclo celular, divisão celular, partição cromossômica	21 (0,60%)	10 (0,60%)	11 (0,61%)
Biogênese de parede celular / membrana / envelope	81 (2,34%)	34 (2,05%)	47 (2,62%)
Motilidade celular	12 (0,34%)	6 (0,36%)	6 (0,33%)

(conclusão)

COGs	<i>M. bovis</i>	Sinônimo	Não sinônimo
Modificação pós-translacional, turnover proteico e chaperonas	73 (2,11%)	30 (1,81%)	43 (2,40%)
Mecanismos de transdução de sinal	102 (2,95%)	42 (2,53%)	60 (2,35%)
Tráfego intracelular, secreção e transporte vesicular	8 (0,23%)	6 (0,36%)	2 (0,11%)
Mecanismos de defesa	52 (1,50%)	17 (1,02%)	35 (1,95%)
Estruturas extracelulares	1 (0,03%)	1 (0,06%)	0 (0,00%)
<i>Mobilome</i> : transposons e profagos	37 (1,07%)	14 (0,84%)	23 (1,28%)
Processamento e modificação do RNA	3 (0,08%)	2 (0,12%)	1 (0,05%)
Tradução, estrutura ribossomal e biogênese	84 (2,43%)	34 (2,05%)	50 (2,79%)
Transcrição	119 (3,45%)	58 (3,50%)	61 (3,40%)
Replicação, recombinação e reparação	113 (3,27%)	39 (2,35%)	74 (4,13%)
Produção e conversão de energia	157 (4,55%)	57 (3,43%)	100 (5,58%)
Transporte e metabolismo de aminoácidos	137 (3,97%)	48 (2,89%)	83 (4,63%)
Transporte e metabolismo de nucleotídeos	46 (1,33%)	20 (1,20%)	23 (1,28%)
Transporte e metabolismo de carboidratos	97 (2,81%)	34 (2,05%)	63 (3,52%)
Transporte e metabolismo de coenzimas	106 (3,07%)	42 (2,53%)	64 (3,57%)
Transporte e metabolismo de lipídios	173 (5,01%)	65 (3,92%)	108 (6,03%)
Transporte e metabolismo de íons inorgânicos	94 (2,72%)	35 (2,11%)	59 (3,29%)
Biossíntese de metabólitos secundários, transporte e catabolismo	218 (6,32%)	77 (4,64%)	141 (7,87%)
Somente predição de função geral	280 (8,12%)	94 (5,67%)	186 (10,39%)
Função desconhecida	1.093 (31,70%)	364 (21,95%)	729 (40,72%)
Total de sítios polimórficos em proteínas	2.804 (81,32%)	1.014 (61,15%)	1.790 (100,00%)
Total de sítios polimórficos	3.448	1.658 (48,08%)	1.790 (51,92%)

Fonte: Elaborada pela autora.

Nota: O total de COGs é ligeiramente maior que o número de proteínas porque uma proteína pode estar envolvida em mais de um COG.

Dos 8.335 sítios polimórficos encontrados em *M. tuberculosis*, 3.374 (40,48%) possuem mutações sinônimas e 4.961 (59,52%) possuem mutações não sinônimas (Tabela 10). Assim como em *M. bovis* virulentos, a maioria desses sítios estão localizados em regiões codificadoras de proteína (8.293, 99,50%). Apenas dois sítios (0,02%) em genoma não anotado, 10 sítios (0,12%) em região não codificadora de proteínas e 30 (0,36%) em região não anotada foram encontrados.

Tabela 10 - Categorização e as determinações de mutações sinônimas e não sinônimas dos sítios polimórficos encontrados em *Mycobacterium tuberculosis* de acordo com as funções dos COGs (*Cluster of Orthologous*)

(continua)

COGs	<i>M. tuberculosis</i>	Sinônimo	Não sinônimo
Controle do ciclo celular, divisão celular, partição cromossômica	37 (0,44%)	13 (0,38%)	24 (0,48%)
Biogênese de parede celular / membrana / envelope	232 (2,78%)	85 (2,52%)	147 (2,96%)
Motilidade celular	37 (0,44%)	17 (0,50%)	20 (0,40%)

(conclusão)			
COGs	<i>M. tuberculosis</i>	Sinônimo	Não sinônimo
Modificação pós-translacional, turnover proteico e chaperonas	157 (1,88%)	68 (2,01%)	89 (1,80%)
Mecanismos de transdução de sinal	213 (2,55%)	82 (2,43%)	131 (2,64%)
Tráfego intracelular, secreção e transporte vesicular	21 (0,25%)	12 (0,35%)	9 (0,18%)
Mecanismos de defesa	156 (1,87%)	67 (1,98%)	89 (1,80%)
Estruturas extracelulares	3 (0,03%)	1 (0,03%)	2 (0,04%)
<i>Mobilome</i> : transposons e profagos	92 (1,10%)	37 (1,09%)	55 (1,10%)
Tradução, estrutura ribossomal e biogênese	230 (2,76%)	83 (2,46%)	120 (2,41%)
Transcrição	255 (3,06%)	99 (2,93%)	156 (3,14%)
Replicação, recombinação e reparação	203 (2,43%)	83 (2,46%)	120 (2,41%)
Produção e conversão de energia	367 (4,40%)	156 (4,62%)	205 (4,13%)
Transporte e metabolismo de aminoácidos	281 (3,37%)	110 (3,26%)	171 (3,44%)
Transporte e metabolismo de nucleotídeos	74 (0,88%)	36 (1,06%)	38 (0,76%)
Transporte e metabolismo de carboidratos	217 (2,60%)	82 (2,43%)	135 (2,72%)
Transporte e metabolismo de coenzimas	227 (2,72%)	90 (2,66%)	137 (2,76%)
Transporte e metabolismo de lipídios	415 (4,98%)	185 (5,48%)	240 (4,83%)
Transporte de íons inorgânicos e metabolismo	252 (3,02%)	87 (2,57%)	165 (3,32%)
Biossíntese de metabólitos secundários, transporte e catabolismo	450 (5,40%)	197 (5,83%)	253 (5,10%)
Somente predição de função geral	486 (5,83%)	207 (6,13%)	279 (5,62%)
Função desconhecida	4.276 (51,30%)	1.657 (49,11%)	2.619 (52,80%)
Total de sítios polimórficos em proteínas	8.293 (99,50%)	3.332 (98,75%)	4.961 (100,00%)
Total de sítios polimórficos	8.335	3.374 (40,48%)	4.961 (59,52%)

Fonte: Elaborada pela autora.

Nota: O total de COGs é ligeiramente maior que o número de proteínas porque uma proteína pode estar envolvida em mais de um COG.

Nesse contexto, a categorização das funções (de acordo com os COGs) dos sítios polimórficos evidencia que *M. bovis* e *M. tuberculosis* contêm padrões de mutações similares. Além das proteínas identificadas sem função ou função geral, proteínas envolvidas em “biossíntese de metabólitos secundários, transporte e catabolismo”, “transporte e metabolismo de lipídios”, “produção e conversão de energia” e “transporte e metabolismo de aminoácidos” destacam-se. Esse padrão compartilhado pode ser sugerido devido a pressão de seleção natural intra-hospedeira, uma vez que são espécies virulentas isoladas de diferentes hospedeiros. Mesmo assim, não é possível desconsiderar que as diferenças nos fenótipos dessas bactérias sejam decorrentes da presença variável de SNPs em cada gene individualmente (FILLIOL et al., 2006; BIGI et al., 2016, 2017).

Dos 1.088 sítios polimórficos descritos em BCG, 556 (51,10%) constituem mutações sinônimas e 532 (48,90%) mutações não sinônimas (Tabela 11). Ainda, 757 (69,58%) foram

identificados em proteínas, 178 (16,36%) em genoma não anotado, 52 (4,78%) em região não codificadora de proteínas e 101 (9,28%) em região não anotada. Observa-se uma menor proporção de mutações não sinônimas e uma proporção maior de sítios polimórficos em regiões intergênicas (não codificadoras) quando comparado às micobactérias virulentas, esse último podendo refletir em alterações de expressão gênica na espécie atenuada. Adicionalmente, diferentemente das espécies de micobactérias virulentas, o padrão funcional de sítios polimórficos encontrados segue com proteínas envolvidas em “produção e conversão de energia”, seguido de “controle do ciclo celular, divisão celular, partição cromossômica”, “transcrição” e “mecanismos de transdução de sinal”. Essas características refletem a pressão seletiva sofrida *in vitro* em *M. bovis* BCG, contrastando com a realidade intra-hospedeira das micobactérias virulentas.

Tabela 11 - Categorização e as determinações de mutações sinônimas e não sinônimas dos sítios polimórficos encontrados em *Mycobacterium bovis* BCG de acordo com as funções dos COGs (*Cluster of Orthologous*)

COGs	<i>M. bovis</i> BCG	Sinônimo	Não sinônimo
Controle do ciclo celular, divisão celular, partição cromossômica	37 (3,4%)	13 (2,33%)	24 (4,51%)
Modificação pós-translacional, turnover proteico e chaperonas	12 (1,10%)	5 (0,90%)	7 (1,31%)
Mecanismos de transdução de sinal	29 (2,66%)	2 (0,36%)	25 (4,70%)
Tráfego intracelular, secreção e transporte vesicular	3 (0,29%)	2 (0,36%)	1 (0,18%)
Mecanismos de defesa	1 (0,09%)	0 (0,00%)	1 (0,18%)
<i>Mobilome</i> : transposons e profagos	11 (1,01%)	4 (0,72%)	7 (1,31%)
Tradução, estrutura ribossomal e biogênese	19 (1,74%)	5 (0,90%)	14 (2,63%)
Transcrição	30 (2,75%)	6 (1,08%)	24 (4,51%)
Replicação, recombinação e reparação	10 (0,92%)	1 (0,18%)	9 (1,70%)
Produção e conversão de energia	38 (3,49%)	10 (1,80%)	28 (5,26%)
Transporte e metabolismo de aminoácidos	19 (1,74%)	5 (0,90%)	14 (2,63%)
Transporte e metabolismo de nucleotídeos	10 (0,92%)	2 (0,36%)	8 (1,50%)
Transporte e metabolismo de carboidratos	13 (1,19%)	2 (0,36%)	11 (2,06%)
Transporte e metabolismo de coenzimas	16 (1,47%)	2 (0,36%)	14 (2,63%)
Transporte e metabolismo de lipídios	27 (2,48%)	6 (1,08%)	21 (3,94%)
Transporte de íons inorgânicos e metabolismo	16 (1,47%)	3 (0,54%)	13 (2,44%)
Biossíntese de metabólitos secundários, transporte e catabolismo	22 (2,02%)	6 (1,08%)	16 (3,00%)
Somente predição de função geral	27 (2,48%)	12 (2,15%)	15 (2,82%)
Função desconhecida	471 (43,30%)	155 (27,87%)	316 (59,40%)
Total de SNPs em proteínas	757 (69,57%)	225 (40,46%)	532 (100,00%)
Total de SNPS	1.088	556 (51,10%)	532 (48,90%)

Fonte: Elaborada pela autora.

Nota: O total de COGs é ligeiramente maior que o número de proteínas porque uma proteína pode estar envolvida em mais de um COG.

5.2.7.3 Genes ortólogos envolvidos em sítios polimórficos: *M. bovis* e *M. tuberculosis* compartilham um maior número de genes sob pressão evolutiva entre si do que com *M. bovis* BCG

Um total de 2.804 proteínas de *M. bovis*, 757 de *M. bovis* BCG e 8.293 de *M. tuberculosis* foram identificadas contendo sítios polimórficos (total: 11.854 proteínas). Após remoção de proteínas repetidas (i.e. proteínas que apareciam mais de uma vez por possuírem mais de um sítio polimórfico), um total de 6.023 sequências de proteínas foram analisadas na plataforma BLASTClust para identificação de homologia. As proteínas, com homólogos ou não, foram classificadas de acordo com sua presença nas diferentes espécies bacterianas (Figura 6). Foram identificados 1.461 grupos de proteínas homólogas, com 2 a 483 proteínas cada, totalizando 4.226 proteínas. As 1.797 proteínas remanescentes não foram agrupadas em grupos de homólogos, sendo 323 proteínas encontradas em *M. bovis*, 1.436 em *M. tuberculosis* e 38 em *M. bovis* BCG (Figura 6). Apenas 24,88% (1.499/6.023) das proteínas com sítios polimórficos analisadas são comuns entre *M. bovis*, *M. tuberculosis* e *M. bovis* BCG. Esse número sobe para 64,86% (3.907/6.023) quando se considera apenas *M. bovis* virulentos e *M. tuberculosis*, indicando que esses micro-organismos compartilham um número maior de proteínas sujeitas a pressão evolutiva por mutações. Mesmo assim, existem ainda números consideráveis de proteínas (323 de *M. bovis* e 1.436 de *M. tuberculosis*) sob pressão evolutiva espécie-específicas.

Figura 6 - Diagrama de Venn demonstrando proteínas únicas e compartilhamento de proteínas ortólogas envolvidas em sítios polimórficos, entre genomas de *Mycobacterium bovis* BCG, *M. bovis* e *M. tuberculosis*

Fonte: Acervo pessoal.

Legenda: *M. bovis* BCG identificado pelo traçado, *M. bovis* cinza escuro e *M. tuberculosis* cinza claro.

Dessas 6.023 proteínas contendo sítios polimórficos, a maioria (4.514/6.023, 74,94%) possui pelo menos um sítio polimórfico de mutação não sinônima. Essas proteínas mostram-se

categorizadas entre as espécies bacterianas e também por COGs na figura 7. É possível observar que a maioria dessas proteínas está categorizada como função desconhecida ou predição de função geral (2.557/4.514, 56,64%). Entre as proteínas com função categorizada (Figura 7, em negrito na tabela), “transporte e metabolismo de coenzimas”, “transporte e metabolismo de carboidratos”, “transporte e metabolismo de lipídeos e “biossíntese de metabólitos secundários, “transporte e catabolismo”, destacam-se como sendo as categorias com o maior número de proteínas sob pressão seletiva em *M. bovis* e *M. tuberculosis*, sendo essas as mesmas categorias com importante relevância descritas em outros estudos (DAS et al., 2009; YIMER et al., 2016; JIA et al., 2017). Quando analisamos proteínas exclusivamente sob pressão evolutiva em *M. bovis* (n=345) ou *M. tuberculosis* (n=1.324), é possível observar proteínas nas mesmas categorias listadas acima, mas que por não serem compartilhadas entre essas espécies sugerem padrões evolutivos espécie-específicos dentro de cada categoria funcional (Figura 7, em negrito na tabela). Destaca-se o surgimento da categoria de “produção e conservação de energia” (*M. bovis*) e “transcrição” (*M. tuberculosis*) como categorias de importante número (Figura 7, em negrito na tabela), sugerindo que funções relacionadas a expressão gênica são importantes para evolução espécie-específica desses micro-organismos, como já descrito anteriormente (GOLBY et al., 2007; REHREN et al., 2007; PEPPERELL et al., 2013).

Figura 7 - Diagrama de Venn demonstrando proteínas únicas e compartilhamento de proteínas ortólogas com mutações não sinônimas, entre *M. bovis* BCG, *M. bovis* e *M. tuberculosis*, e sua categorização em COGs (*Cluster of Orthologous Groups*)

COG	A	C	E	B	D	F	G	D+G
Controle do ciclo celular, divisão celular, partição cromossômica	0	0	8	0	17	0	0	17
Biogênese de parede celular / membrana / envelope	4	7	50	2	51	11	10	61
Motilidade celular	0	0	5	0	5	0	5	10
Modificação pós-translacional, turnover proteico e chaperonas	1	11	36	0	24	11	8	32
Mecanismos de transdução de sinal	3	6	35	0	21	10	43	64
Tráfego intracelular, secreção e transporte vesicular	1	2	5	0	1	0	0	1
Mecanismos de defesa	0	7	15	0	51	0	3	54
Estruturas extracelulares	0	0	2	0	0	0	0	0
<i>Mobilome</i> : transposons e profagos	0	8	15	0	19	1	10	29
Processamento e modificação do RNA	0	0	0	0	1	0	0	1
Tradução, estrutura ribossomal e biogênese	5	13	40	2	0	8	4	4
Transcrição	5	11	57	0	54	13	23	77
Replicação, recombinação e reparação	0	11	30	8	58	5	8	66
Produção e conversão de energia	3	17	47	1	72	16	59	131
Transporte e metabolismo de aminoácidos	2	13	55	1	79	8	18	97
Transporte e metabolismo de nucleotídeos	1	8	14	3	12	0	5	17
Transporte e metabolismo de carboidratos	1	9	36	1	60	7	13	173
Transporte e metabolismo de coenzimas	0	12	45	0	61	19	13	174
Transporte e metabolismo de lipídios	2	15	36	3	100	8	70	170
Transporte de íons inorgânicos e metabolismo	1	8	32	0	81	8	19	100
Biossíntese de metabólitos secundários, transporte e catabolismo	2	11	30	0	82	2	71	153
Somente predição de função geral	4	33	72	0	141	13	45	186
Função desconhecida	22	157	740	13	676	96	545	1.221
Total*	57	345	1.324	36	1.603	227	922	2.525

Legenda: *M. bovis* BCG (traçado - A), *M. bovis* e *M. bovis* BCG (B), *M. bovis* (cinza claro - C), *M. bovis* e *M. tuberculosis* (D), *M. tuberculosis* (cinza escuro - E), *M. bovis* BCG e *M. tuberculosis* (F), *M. bovis*, *M. bovis* BCG e *M. tuberculosis* (G). * Total de COGs >4.514 porque uma proteína pode estar envolvida em mais de um COG.

5.2.8 Filogenia inferida por SNPs agrupa linhagens de *M. bovis* conforme os Complexos Clonais

A árvore filogenética baseada em alelos de SNPs suporta a classificação em clados de *M. bovis* por possíveis linhagens condizentes com os complexos clonais e não com hospedeiros de origem dos isolados de *M. bovis* (Figura 8). A presença de SNPs nos genomas é uma importante ferramenta para a construção de árvores filogenéticas e é sugerido como nova forma de categorização de linhagens para o MTBC. Essa metodologia já vem sendo utilizada para diferenciar as linhagens de *M. tuberculosis* (GUTACKER et al., 2002; NAMOUCHI et al., 2012; COSCOLLA; GAGNEUX, 2014; PHELAN et al., 2016) e também como uma forma de marcador filogenético capaz de distinguir as espécies dentro desse complexo (COLL et al., 2014). Contudo, são poucos estudos descritos envolvendo genômica comparativa apenas entre *M. bovis* (GARCIA PELAYO et al., 2009; JOSHI et al., 2012, DE LA FUENTE et al., 2015; PATANÉ et al., 2017), e a separação por diferentes complexos clonais, Europeu 1 e 2 e Africano 1 e 2, é recente (MÜLLER et al., 2009; BERG et al., 2011; SMITH et al., 2011; RODRIGUEZ-CAMPOS et al., 2012). Este é o primeiro estudo envolvendo identificação de SNPs, analisando genomas completos e *drafts* de *M. bovis*, depositados no GenBank, que permitiu separação conforme os diferentes Complexos Clonais através da utilização dos SNPs identificados. Assim pode-se sugerir a utilização da pesquisa de SNPs como um modelo de marcador genotípico de *M. bovis*, permitindo a separação de estirpes conforme diferentes linhagens suportadas pelos Complexos Clonais.

Figura 8 - Árvore filogenética obtida pelo método *Neighbor-joining*, utilizando SNPs (polimorfismo de nucleotídeo único) de membros de Complexo *Mycobacterium tuberculosis* (MTBC). Os genomas de *Mycobacterium bovis* são acompanhados da identificação espécie hospedeira afetada, local de isolamento, número SB do *spoligotyping* e agrupados conforme os Complexos Clonais

Fonte: Acervo pessoal.

Legenda: Relação filogenética de membros do MTBC baseada em SNPs. Ramos que concordam com os alinhamentos pelos métodos de *maximum likelihood* e *neighbor-joining* estão marcados com asterisco. *Mycobacterium tuberculosis* H37Rv foi selecionado como *outgroup*.

** Números SB do *spoligotyping* não categorizados.

† Complexos clonais não determinados.

7 CONSIDERAÇÕES FINAIS

Este é o primeiro estudo a sequenciar por completo o genoma de uma estirpe nacional de *M. bovis*, assim como o primeiro a realizar uma análise de genômica comparativa abrangente dos genomas desta espécie, depositados em um banco de dados público. O genoma de *M. bovis* SP38 mostrou características tradicionais de um genoma de micobactéria tuberculosa, e foi categorizado como Complexo Clonal Europeu 2, sendo este o primeiro relato deste complexo no Brasil.

Dentre os genomas de *M. bovis* depositados no GenBank, a investigação das RDs identificou a ocorrência de genomas do MTBC depositados equivocadamente como essa espécie. Ainda, dentre esses genomas, o espótipo mais presente foi o SB0141, inclusive dos isolados brasileiros, e apenas a estirpe *M. bovis* SP38 demonstrou o padrão mais prevalente do Brasil, o SB0121.

A análise comparativa das CDSs dos genomas de *M. bovis* indica que essa espécie está sob um processo de *decay* genômico intra-específico e também em comparação ao *M. tuberculosis*. Também não existem CDSs exclusivamente presentes em todos os genomas de *M. bovis*, suportando o aspecto clonal do complexo bacteriano. Sugere-se então que a taxa de pseudogenização e o processo de *decay* genômico seja explorado em estudos futuros de forma a prover dados que possam auxiliar na compreensão da adaptação de membros do MTBC ao hospedeiro, uma vez que o processo de pseudogenização tem sido associado à especialização hospedeira de bactérias.

A ocorrência de uma maior variabilidade genética de *M. tuberculosis* em comparação ao *M. bovis* observada neste estudo contrasta com a capacidade deste último em infectar múltiplos hospedeiros. É imprescindível que mais estudos sejam realizados com um maior número de genomas de *M. bovis*, isolados de diferentes hospedeiros, para excluir o viés geográfico e de linhagens da amostra utilizada neste estudo. Também é possível concluir que o *M. bovis* BCG sofre uma pressão evolutiva diferente dessas micobactérias virulentas, e a presença de sítios polimórficos, mesmo que em menor grau que *M. bovis* virulentos e *M. tuberculosis*, sugere que essas mutações possam afetar o sucesso vacinal das diferentes estirpes.

Apesar dos Complexos Clonais descreverem linhagens de *M. bovis*, nem todos os *M. bovis* podem ser classificados dentre as opções existentes, evidenciando lacunas do conhecimento a serem preenchidas. A utilização da filogenia baseada em SNPs poderá auxiliar no delineamento das diferentes linhagens de *M. bovis*, podendo ser utilizada como um marcador possivelmente geográfico e não de hospedeiro de origem dos isolados.

REFERÊNCIAS

ABOUBAKER OSMAN, D.; BOUZID, F.; CANAAN, S.; DRANCOURT, M. Smooth Tubercle Bacilli: Neglected Opportunistic Tropical Pathogens. **Frontiers in public health**, v. 3, p. 283, 2015.

AHMAD, S.; EL-SHAZLY, S.; MUSTAFA, A. S.; AL-ATTIYAH, R. Mammalian Cell-Entry Proteins Encoded by the mce3 Operon of *Mycobacterium tuberculosis* are Expressed During Natural Infection in Humans. **Scandinavian Journal of Immunology**, v. 60, n. 4, p. 382–391, out. 2004.

ALEXANDER, K. A.; LAVER, P. N.; MICHEL, A. L.; WILLIAMS, M.; VAN HELDEN, P. D.; WARREN, R. M.; GEY VAN PITTIUS, N. C. Novel *Mycobacterium tuberculosis* Complex Pathogen, *M. mungi*. **Emerging Infectious Diseases**, v. 16, n. 8, p. 1296–1299, ago. 2010.

ALLIX-BÉGUEC, C.; HARMSSEN, D.; WENIGER, T.; SUPPLY, P.; NIEMANN, S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. **Journal of clinical microbiology**, v. 46, n. 8, p. 2692–9, ago. 2008.

ALLIX, C.; WALRAVENS, K.; SAEGERMAN, C.; GODFROID, J.; SUPPLY, P.; FAUVILLE-DUFAUX, M. Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of *Mycobacterium bovis* and comparison of the method with IS6110 restriction fragment length polymorphism analysis and spoligotyping. **Journal of clinical microbiology**, v. 44, n. 6, p. 1951–62, 1 jun. 2006.

ALVA, V.; NAM, S.-Z.; SÖDING, J.; LUPAS, A. N. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. **Nucleic Acids Research**, v. 44, n. W1, p. W410–W415, 8 jul. 2016.

AMBROSIO, S. R.; OLIVEIRA, E. M. D. D.; RODRIGUEZ, C. A. R.; NETO, J. S. F.; AMAKU, M. Comparison of three decontamination methods for *Mycobacterium bovis* isolation. **Brazilian Journal of Microbiology**, v. 39, p. 241–244, 2008.

ARANAZ, A.; JUAN, L. De; MONTERO, N.; SÁNCHEZ, C.; GALKA, M.; DELSO, C.; ÁLVAREZ, J.; ROMERO, B.; BEZOS, J.; VELA, A. I.; BRIONES, V.; MATEOS, A.; DOMÍNGUEZ, L. Bovine tuberculosis (*Mycobacterium bovis*) in wildlife in Spain. **Journal of Clinical Microbiology**, v. 42, n. 6, p. 2602–2608, 2004.

BÄUMLER, A.; FANG, F. C. Host specificity of bacterial pathogens. **Cold Spring Harbor perspectives in medicine**, v. 3, n. 12, p. a010041, 1 dez. 2013.

BEHR, M. A.; GORDON, S. V. Why doesn't *Mycobacterium tuberculosis* spread in animals? **Trends in Microbiology**, v. 23, n. 1, p. 1–2, 2015.

BEMER-MELCHIOR, P.; DRUGEON, H. B. Inactivation of *Mycobacterium tuberculosis* for DNA typing analysis. **Journal of clinical microbiology**, v. 37, n. 7, p. 2350–1, jul. 1999.

BERG, S.; GARCIA-PELAYO, M. C.; MÜLLER, B.; HAILU, E.; ASIIMWE, B.; KREMER, K.; DALE, J.; BONIOTTI, M. B.; RODRIGUEZ, S.; HILTY, M.; RIGOUTS, L.; FIRDESSA, R.; MACHADO, A.; MUCAVELE, C.; NGANDOLO, B. N. R.; BRUCHFELD, J.; BOSCHIROLI, L.; MÜLLER, A.; SAHRAOUI, N.; PACCIARINI, M.; CADMUS, S.; JOLOBA, M.; VAN SOOLINGEN, D.; MICHEL, A. L.; DJØNNE, B.; ARANAZ, A.; ZINSSTAG, J.; VAN HELDEN, P.; PORTAELS, F.; KAZWALA, R.; KÄLLENIUS, G.; HEWINSON, R. G.; ASEFFA, A.; GORDON, S. V.; SMITH, N. H. African 2, a clonal complex of *Mycobacterium bovis* epidemiologically important in East Africa. **Journal of bacteriology**, v. 193, n. 3, p. 670–8, fev. 2011.

BIEK, R.; O'HARE, A.; WRIGHT, D.; MALLON, T.; MCCORMICK, C.; ORTON, R. J.; MCDOWELL, S.; TREWBY, H.; SKUCE, R. A.; KAO, R. R. Whole Genome Sequencing Reveals Local Transmission Patterns of *Mycobacterium bovis* in Sympatric Cattle and Badger Populations. **PLoS Pathogens**, v. 8, n. 11, p. e1003008, 29 nov. 2012.

BIGI, M. M.; BLANCO, F. C.; ARAÚJO, F. R.; THACKER, T. C.; ZUMÁRRAGA, M. J.; CATALDI, A. A.; SORIA, M. A.; BIGI, F. Polymorphisms of 20 regulatory proteins between *Mycobacterium tuberculosis* and *Mycobacterium bovis*. **Microbiology and Immunology**, v. 60, n. 8, p. 552–560, ago. 2016.

BIGI, M. M.; LOPEZ, B.; BLANCO, F. C.; SASIAIN, M. D. C.; DE LA BARRERA, S.; MARTI, M. A.; SOSA, E. J.; FERNÁNDEZ DO PORTO, D. A.; RITACCO, V.; BIGI, F.; SORIA, M. A. Single nucleotide polymorphisms may explain the contrasting phenotypes of two variants of a multidrug-resistant *Mycobacterium tuberculosis* strain. **Tuberculosis (Edinburgh, Scotland)**, v. 103, p. 28–36, mar. 2017.

BOLOTIN, E.; HERSHBERG, R. Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. **Genome Biology and Evolution**, v. 7, n. 8, p. 2173–2187, 10 ago. 2015.

BOS, K. I.; HARKINS, K. M.; HERBIG, A.; COSCOLLA, M.; WEBER, N.; COMAS, I.; FORREST, S. A.; BRYANT, J. M.; HARRIS, S. R.; SCHUENEMANN, V. J.; CAMPBELL, T. J.; MAJANDER, K.; WILBUR, A. K.; GUICHON, R. A.; WOLFE STEADMAN, D. L.; COOK, D. C.; NIEMANN, S.; BEHR, M. A.; ZUMARRAGA, M.; BASTIDA, R.; HUSON, D.; NIESELT, K.; YOUNG, D.; PARKHILL, J.; BUIKSTRA, J. E.; GAGNEUX, S.; STONE, A. C.; KRAUSE, J. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. **Nature**, v. 514, n. 7523, p. 494–7, 23 out. 2014.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Programa Nacional de Controle e Erradicação da Brucelose e da Tuberculose Animal (PNCEBT). Brasília, 2006. 184 p.

BROSCH, R.; GORDON, S. V.; MARMIESSE, M.; BRODIN, P.; BUCHRIESER, C.; EIGLMEIER, K.; GARNIER, T.; GUTIERREZ, C.; HEWINSON, G.; KREMER, K.; PARSONS, L. M.; PYM, A. S.; SAMPER, S.; VAN SOOLINGEN, D.; COLE, S. T. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. **Proceedings of the National Academy of Sciences**, v. 99, n. 6, p. 3684–3689, 19 mar. 2002.

BROSCH, R.; GORDON, S. V.; GARNIER, T.; EIGLMEIER, K.; FRIGUI, W.; VALENTI, P.; DOS SANTOS, S.; DUTHOY, S.; LACROIX, C.; GARCIA-PELAYO, C.; INWALD, J. K.; GOLBY, P.; GARCIA, J. N.; HEWINSON, R. G.; BEHR, M. A.; QUAIL, M. A.; CHURCHER, C.; BARRELL, B. G.; PARKHILL, J.; COLE, S. T. Genome plasticity of BCG and impact on vaccine efficacy. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 13, p. 5596–601, 27 mar. 2007.

CARVALHO, R. C. T.; VASCONCELLOS, S. E. G.; ISSA, M. de A.; SOARES FILHO, P. M.; MOTA, P. M. P. C.; ARAÚJO, F. R. de; CARVALHO, A. C. da S.; GOMES, H. M.; SUFFYS, P. N.; FIGUEIREDO, E. E. de S.; PASCHOALIN, V. M. F. Molecular Typing of *Mycobacterium bovis* from Cattle Reared in Midwest Brazil. **PLOS ONE**, v. 11, n. 9, p. e0162459, 15 set. 2016.

Centro Panamericano de Zoonosis. Manual de normas y procedimientos técnicos para la bacteriología de la tuberculosis. Parte II. El cultivo del *Mycobacterium tuberculosis*. Nota técnica 27. Martínez: Organización Panamericana de la Salud; 1985.

CHEONG, W.-H.; TAN, Y.-C.; YAP, S.-J.; NG, K.-P. ClicO FS: an interactive web-based service of Circos. **Bioinformatics**, v. 31, n. 22, p. 3685–3687, 15 nov. 2015.

CHIKHI, R.; MEDVEDEV, P. Informed and Automated k -Mer Size Selection for Genome Assembly. **Bioinformatics**, v. 30, n. 1, p. 31–37, 2014.

CLC bio. Manual for Genomics Benchwork 8, 2015. Disponível online www.clcbio.com. Acessado em 07 de junho de 2016.

COLE, S. T. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex a. **Microbiology**, v. 148, n. 10, p. 2919–2928, 1 out. 2002.

COLEMAN, J. D.; COOKE, M. M. *Mycobacterium bovis* infection in wildlife in New Zealand. **Tuberculosis**, v. 81, n. 3, p. 191–202, jun. 2001.

COLL, F.; MALLARD, K.; PRESTON, M. D.; BENTLEY, S.; PARKHILL, J.; MCNERNEY, R.; MARTIN, N.; CLARK, T. G. SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. **Bioinformatics**, v. 28, n. 22, p. 2991–2993, 15 nov. 2012.

COLL, F.; MCNERNEY, R.; GUERRA-ASSUNÇÃO, J. A.; GLYNN, J. R.; PERDIGÃO, J.; VIVEIROS, M.; PORTUGAL, I.; PAIN, A.; MARTIN, N.; CLARK, T. G. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. **Nature Communications**, v. 5, p. 4812, 1 set. 2014.

COSCOLLA, M.; GAGNEUX, S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. **Seminars in Immunology**, v. 26, n. 6, p. 431–444, dez. 2014.

COSCOLLA, M.; LEWIN, A.; METZGER, S.; MAETZ-RENNING, K.; CALVIGNAC-SPENCER, S.; NITSCHKE, A.; DABROWSKI, P. W.; RADONIC, A.; NIEMANN, S.; PARKHILL, J.; COUACY-HYMAN, E.; FELDMAN, J.; COMAS, I.; BOESCH, C.; GAGNEUX, S.; LEENDERTZ, F. H. Novel *Mycobacterium tuberculosis* Complex Isolate from a Wild Chimpanzee. **Emerging Infectious Diseases**, v. 19, n. 6, p. 969–976, jun. 2013.

CRABTREE, J.; ANGIUOLI, S. V.; WORTMAN, J. R.; WHITE, O. R. Sybil: Methods and Software for Multiple Genome Comparison and Visualization. **Methods in molecular biology (Clifton, N.J.)**, v. 408, p. 93–108, 2007.

CRISPELL, J.; ZADOKS, R. N.; HARRIS, S. R.; PATERSON, B.; COLLINS, D. M.; DELISLE, G. W.; LIVINGSTONE, P.; NEILL, M. A.; BIEK, R.; LYCETT, S. J.; KAO, R. R.; PRICE-CARTER, M. Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. **BMC genomics**, v. 18, n. 1, p. 180, 16 fev. 2017.

DAS, S.; YENNAMALLI, R. M.; VISHNOI, A.; GUPTA, P.; BHATTACHARYA, A. Single-nucleotide variations associated with *Mycobacterium tuberculosis* KwaZulu-Natal strains. **Journal of biosciences**, v. 34, n. 3, p. 397–404, set. 2009.

DE GARINE-WICHATITSKY, M.; CARON, A.; KOCK, R.; TSCHOPP, R.; MUNYEME, M.; HOFMEYER, M.; MICHEL, A. A review of bovine tuberculosis at the wildlife–livestock–human interface in sub-Saharan Africa. **Epidemiology and Infection**, v. 141, n. 7, p. 1342–1356, 15 jul. 2013.

DE JONG, B. C.; ANTONIO, M.; GAGNEUX, S. *Mycobacterium africanum* - Review of an Important Cause of Human Tuberculosis in West Africa. **PLoS Neglected Tropical Diseases**, v. 4, n. 9, p. e744, 28 set. 2010.

DE KANTOR, I. N.; AMBROGGI, M.; POGGI, S.; MORCILLO, N.; DA SILVA TELLES, M. A.; OSÓRIO RIBEIRO, M.; GARZÓN TORRES, M. C.; LLERENA POLO, C.; RIBÓN, W.; GARCÍA, V.; KUFFO, D.; ASENCIOS, L.; VÁSQUEZ CAMPOS, L. M.; RIVAS, C.; DE WAARD, J. H. Human *Mycobacterium bovis* infection in ten Latin American countries. **Tuberculosis**, v. 88, n. 4, p. 358–365, jul. 2008.

DE LA FUENTE, J.; DÍEZ-DELGADO, I.; CONTRERAS, M.; VICENTE, J.; CABEZAS-CRUZ, A.; TOBES, R.; MANRIQUE, M.; LÓPEZ, V.; ROMERO, B.; BEZOS, J.; DOMINGUEZ, L.; SEVILLA, I. A.; GARRIDO, J. M.; JUSTE, R.; MADICO, G.; JONES-LÓPEZ, E.; GORTAZAR, C. Comparative Genomics of Field Isolates of *Mycobacterium bovis* and *M. caprae* Provides Evidence for Possible Correlates with Bacterial Viability and Virulence. **PLoS Neglected Tropical Diseases**, v. 9, n. 11, p. 1–22, 2015.

DE LA RUA-DOMENECH, R.; GOODCHILD, A. T.; VORDERMEIER, H. M.; HEWINSON, R. G.; CHRISTIANSEN, K. H.; CLIFTON-HADLEY, R. S. Ante mortem diagnosis of tuberculosis in cattle: A review of the tuberculin tests, γ -interferon assay and other ancillary diagnostic techniques. **Research in Veterinary Science**, v. 81, n. 2, p. 190–210, 2006.

DE LISLE, G. W.; BENGIS, R. G.; SCHMITT, S. M.; O'BRIEN, D. J. Tuberculosis in free-ranging wildlife: detection, diagnosis and management. **Revue scientifique et technique (International Office of Epizootics)**, v. 21, n. 2, p. 317–34, ago. 2002.

DIPPENAAR, A.; PARSONS, S. D. C.; SAMPSON, S. L.; VAN DER MERWE, R. G.; DREWE, J. A.; ABDALLAH, A. M.; SIAME, K. K.; GEY VAN PITTIUS, N. C.; VAN HELDEN, P. D.; PAIN, A.; WARREN, R. M. Whole genome sequence analysis of *Mycobacterium suricattae*. **Tuberculosis (Edinburgh, Scotland)**, v. 95, n. 6, p. 682–8, dez. 2015.

DONNELLY, C. A.; WOODROFFE, R.; COX, D. R.; BOURNE, J.; GETTINBY, G.; LE FEVRE, A. M.; MCINERNEY, J. P.; MORRISON, W. I. Impact of localized badger culling on tuberculosis incidence in British cattle. **Nature**, v. 426, n. 6968, p. 834–837, 18 dez. 2003.

DOS VULTOS, T.; MESTRE, O.; RAUZIER, J.; GOLEC, M.; RASTOGI, N.; RASOLOFO, V.; TONJUM, T.; SOLA, C.; MATIC, I.; GICQUEL, B. Evolution and Diversity of Clonal Bacteria: The Paradigm of *Mycobacterium tuberculosis*. **PLoS ONE**, v. 3, n. 2, p. e1538, 6 fev. 2008.

DUNN, P. L.; NORTH, R. J. Virulence ranking of some *Mycobacterium tuberculosis* and *Mycobacterium bovis* strains according to their ability to multiply in the lungs, induce lung pathology, and cause mortality in mice. **Infection and immunity**, v. 63, n. 9, p. 3428–37, set. 1995.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792–7, 2004.

EL-SAYED, A.; EL-SHANNAT, S.; KAMEL, M.; CASTAÑEDA-VAZQUEZ, M. A.; CASTAÑEDA-VAZQUEZ, H. Molecular Epidemiology of *Mycobacterium bovis* in Humans and Cattle. **Zoonoses and Public Health**, v. 63, n. 4, p. 251–264, jun. 2016.

FAKSRI, K.; XIA, E.; TAN, J. H.; TEO, Y.-Y.; ONG, R. T.-H. In silico region of difference (RD) analysis of *Mycobacterium tuberculosis* complex from sequence reads using RD-Analyzer. **BMC genomics**, v. 17, n. 1, p. 847, 2 nov. 2016.

FILLIOL, I.; MOTIWALA, A. S.; CAVATORE, M.; QI, W.; HAZBÓN, M. H.; BOBADILLA DEL VALLE, M.; FYFE, J.; GARCÍA-GARCÍA, L.; RASTOGI, N.; SOLA, C.; ZOZIO, T.; GUERRERO, M. I.; LEÓN, C. I.; CRABTREE, J.; ANGIUOLI, S.; EISENACH, K. D.; DURMAZ, R.; JOLOBA, M. L.; RENDÓN, A.; SIFUENTES-OSORNIO, J.; PONCE DE LEÓN, A.; CAVE, M. D.; FLEISCHMANN, R.; WHITTAM, T. S.; ALLAND, D. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP s. **Journal of bacteriology**, v. 188, n. 2, p. 759–72, 15 jan. 2006.

FISHBEIN, S.; VAN WYK, N.; WARREN, R. M.; SAMPSON, S. L. Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity. **Molecular Microbiology**, v. 96, n. 5, p. 901–916, 2015.

FISHER, M. A.; PLIKAYTIS, B. B.; SHINNICK, T. M. Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. **Journal of bacteriology**, v. 184, n. 14, p. 4025–32, jul. 2002.

FORRELLAD, M. A.; KLEPP, L. I.; GIOFFRÉ, A.; SABIO Y GARCÍA, J.; MORBIDONI, H. R.; DE LA PAZ SANTANGELO, M.; CATALDI, A. A.; BIGI, F. Virulence factors of the *Mycobacterium tuberculosis* complex. **Virulence**, v. 4, n. 1, p. 3–66, 1 jan. 2013.

FROTHINGHAM, R.; HILLS, H. G.; WILSON, K. H. Extensive DNA sequence conservation throughout the *Mycobacterium tuberculosis* complex. **Journal of clinical microbiology**, v. 32, n. 7, p. 1639–43, jul. 1994.

GAGNEUX, S. Host-pathogen coevolution in human tuberculosis. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 367, n. 1590, p. 850–9, 19 mar. 2012.

GALAGAN, J. E. Genomic insights into tuberculosis. **Nature Reviews Genetics**, v. 15, n. 5, p. 307–320, 25 mar. 2014.

GARBACCIO, S.; MACIAS, A.; SHIMIZU, E.; PAOLICCHI, F.; PEZZONE, N.; MAGNANO, G.; ZAPATA, L.; ABDALA, A.; TARABLA, H.; PEYRU, M.; CAIMI, K.; ZUMÁRRAGA, M.; CANAL, A.; CATALDI, A. Association between spoligotype-VNTR types and virulence of *Mycobacterium bovis* in cattle. **Virulence**, v. 5, n. 2, p. 297–302, 15 fev. 2014.

GARCIA PELAYO, M. C.; UPLEKAR, S.; KENIRY, A.; MENDOZA LOPEZ, P.; GARNIER, T.; NUNEZ GARCIA, J.; BOSCHIROLI, L.; ZHOU, X.; PARKHILL, J.; SMITH, N.; HEWINSON, R. G.; COLE, S. T.; GORDON, S. V. A comprehensive survey of single nucleotide polymorphisms (SNPs) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent *M. bovis* strains and *M. bovis* BCG strains. **Infection and immunity**, v. 77, n. 5, p. 2230–8, maio 2009.

GARDNER, S. N.; SLEZAK, T.; HALL, B. G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome: Table 1. **Bioinformatics**, v. 31, n. 17, p. 2877–2878, 1 set. 2015.

GARNIER, T.; EIGLMEIER, K.; CAMUS, J.-C.; MEDINA, N.; MANSOOR, H.; PRYOR, M.; DUTHOY, S.; GRONDIN, S.; LACROIX, C.; MONSEMPE, C.; SIMON, S.; HARRIS, B.; ATKIN, R.; DOGGETT, J.; MAYES, R.; KEATING, L.; WHEELER, P. R.; PARKHILL, J.; BARRELL, B. G.; COLE, S. T.; GORDON, S. V.; HEWINSON, R. G. The complete genome sequence of *Mycobacterium bovis*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 13, p. 7877–82, 24 jun. 2003.

GLASER, L.; CARSTENSEN, M.; SHAW, S.; ROBBE-AUSTERMAN, S.; WUNSCHMANN, A.; GREAR, D.; STUBER, T.; THOMSEN, B. Descriptive Epidemiology and Whole Genome Sequencing Analysis for an Outbreak of Bovine Tuberculosis in Beef Cattle and White-Tailed Deer in Northwestern Minnesota. **PLOS ONE**, v. 11, n. 1, p. e0145735, 19 jan. 2016.

GOLBY, P.; HATCH, K. A.; BACON, J.; COONEY, R.; RILEY, P.; ALLNUTT, J.; HINDS, J.; NUNEZ, J.; MARSH, P. D.; HEWINSON, R. G.; GORDON, S. V. Comparative transcriptomics reveals key gene expression differences between the human and bovine pathogens of the *Mycobacterium tuberculosis* complex. **Microbiology**, v. 153, n. 10, p. 3323–3336, 1 out. 2007.

GOLBY, P.; NUNEZ, J.; WITNEY, A.; HINDS, J.; QUAIL, M. a; BENTLEY, S.; HARRIS, S.; SMITH, N.; HEWINSON, R. G.; GORDON, S. V. Genome-level analyses of *Mycobacterium bovis* lineages reveal the role of SNPs and antisense transcription in differential

gene expression. **BMC genomics**, v. 14, p. 710, jan. 2013.

GONZALO-ASENSIO, J.; MALAGA, W.; PAWLIK, A.; ASTARIE-DEQUEKER, C.; PASSEMAR, C.; MOREAU, F.; LAVAL, F.; DAFFÉ, M.; MARTIN, C.; BROSC, R.; GUILHOT, C. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 31, p. 11491–6, 5 ago. 2014.

GUIMARAES, A. M.; ZIMPEL, C. K.; IKUTA, C. Y.; DO NASCIMENTO, N. C.; DOS SANTOS, A. P.; MESSICK, J. B.; HEINEMANN, M. B.; FERREIRA NETO, J. S.; BRANDAO, P. E. Draft Genome Sequence of *Mycobacterium bovis* Strain SP38, a Pathogenic Bacterium Isolated from a Bovine in Brazil. **Genome Announc.**, v. 3, n. 3, 2015.

GUTACKER, M. M.; SMOOT, J. C.; MIGLIACCIO, C. A. L.; RICKLEFS, S. M.; HUA, S.; COUSINS, D. V.; GRAVISS, E. A.; SHASHKINA, E.; KREISWIRTH, B. N.; MUSSER, J. M. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. **Genetics**, v. 162, n. 4, p. 1533–43, dez. 2002.

GUTIERREZ, M. C.; BRISSE, S.; BROSC, R.; FABRE, M.; OMAÏS, B.; MARMIESSE, M.; SUPPLY, P.; VINCENT, V. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. **PLoS pathogens**, v. 1, n. 1, p. e5, set. 2005.

HAILE, Y.; CAUGANT, D. A.; BJUNE, G.; WIKER, H. G. *Mycobacterium tuberculosis* mammalian cell entry operon (*mce*) homologs in *Mycobacterium* other than tuberculosis (MOTT). **FEMS Immunology & Medical Microbiology**, v. 33, n. 2, p. 125–132, jun. 2002.

HAUER, A.; DE CRUZ, K.; COCHARD, T.; GODREUIL, S.; KAROUI, C.; HENAULT, S.; BULACH, T.; BAÑULS, A.-L.; BIET, F.; BOSCHIROLI, M. L. Genetic Evolution of *Mycobacterium bovis* Causing Tuberculosis in Livestock and Wildlife in France since 1978. **PLOS ONE**, v. 10, n. 2, p. e0117103, 6 fev. 2015.

HO, H.-T.; CHANG, P.-L.; HUNG, C.-C.; CHANG, H.-T. Capillary Electrophoretic Restriction Fragment Length Polymorphism Patterns for the *Mycobacterium* hsp65 Gene. **JOURNAL OF CLINICAL MICROBIOLOGY**, v. 42, n. 8, p. 3525–3531, 2004.

IKUTA, C. Y. **Comparação entre meios de cultura e condições de incubação para o primo isolamento de *Mycobacterium bovis* de bovinos brasileiros**. 2011. Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, São Paulo, 2011.

ILINA, E. N.; SHITIKOV, E. A.; IKRYANNIKOVA, L. N.; ALEKSEEV, D. G.;

KAMASHEV, D. E.; MALAKHOVA, M. V.; PARFENOVA, T. V.; AFANAS'EV, M. V.; ISCHENKO, D. S.; BAZALEEV, N. A.; SMIRNOVA, T. G.; LARIONOVA, E. E.; CHERNOUSOVA, L. N.; BELETSKY, A. V.; MARDANOV, A. V.; RAVIN, N. V.; SKRYABIN, K. G.; GOVORUN, V. M. Comparative Genomic Analysis of *Mycobacterium tuberculosis* Drug Resistant Strains from Russia. **PLoS ONE**, v. 8, n. 2, p. e56577, 20 fev. 2013.

JIA, X.; YANG, L.; DONG, M.; CHEN, S.; LV, L.; CAO, D.; FU, J.; YANG, T.; ZHANG, J.; ZHANG, X.; SHANG, Y.; WANG, G.; SHENG, Y.; HUANG, H.; CHEN, F. The Bioinformatics Analysis of Comparative Genomics of *Mycobacterium tuberculosis* Complex (MTBC) Provides Insight into Dissimilarities between Intraspecific Groups Differing in Host Association, Virulence, and Epitope Diversity. **Frontiers in cellular and infection microbiology**, v. 7, p. 88, 2017.

JOSHI, D.; HARRIS, N. B.; WATERS, R.; THACKER, T.; MATHEMA, B.; KRIESWIRTH, B.; SREEVATSAN, S. Single nucleotide polymorphisms in the *Mycobacterium bovis* genome resolve phylogenetic relationships. **Journal of clinical microbiology**, v. 50, n. 12, p. 3853–61, dez. 2012.

KANEENE, J. B.; MILLER, R.; KANTOR, I. N. De; THOEN, C. O. Tuberculosis in wild animals. **The International Journal of Tuberculosis and Lung Disease**, v. 14, n. 12, p. 1508–1512, 2010.

KANTOR, I. N. De; LOBUE, P. A.; THOEN, C. O. Tuberculosis: a re-emerging disease in animals and humans Human tuberculosis caused by *Mycobacterium bovis* in the United States, Latin America and the Caribbean. **International Journal of Tuberculosis and Lung Disease**, v. 14, n. 11, p. 1369–1373, 2010.

KATO-MAEDA, M.; RHEE, J. T.; GINGERAS, T. R.; SALAMON, H.; DRENKOW, J.; SMITTIPAT, N.; SMALL, P. M. Comparing genomes within the species *Mycobacterium tuberculosis*. **Genome research**, v. 11, n. 4, p. 547–54, abr. 2001.

KIESER, K. J.; RUBIN, E. J. How sisters grow apart: mycobacterial growth and division. **Nature reviews. Microbiology**, v. 12, n. 8, p. 550–62, 7 ago. 2014.

KIM, N.; JANG, Y.; KIM, J. K.; RYOO, S.; KWON, K. H.; KANG, S. S.; BYEON, H. S.; LEE, H. S.; LIM, Y.-H.; KIM, J.-M. Complete Genome Sequence of *Mycobacterium bovis* Clinical Strain 1595, Isolated from the Laryngopharyngeal Lymph Node of South Korean Cattle. **Genome announcements**, v. 3, n. 5, 1 out. 2015.

KIM, N.; JANG, Y.; KIM, J. K.; RYOO, S.; KWON, K. H.; KIM, M.; KANG, S. S.; BYEON, H. S.; LEE, H. S.; LIM, Y.-H.; KIM, J.-M. Molecular and genomic features of *Mycobacterium*

bovis strain 1595 isolated from Korean cattle. **Journal of veterinary science**, 6 abr. 2017.

KUMAR, A.; KAUR, J. Primer Based Approach for PCR Amplification of High GC Content Gene: *Mycobacterium* Gene as a Model. **Molecular biology international**, v. 2014, p. 937308, 2014.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870–1874, jul. 2016.

LIU, F.; HU, Y.; WANG, Q.; LI, H.; GAO, G. F.; LIU, C.; ZHU, B. Comparative genomic analysis of *Mycobacterium tuberculosis* clinical isolates. **BMC Genomics**, v. 15, n. 1, p. 469, 13 jun. 2014.

MAHAIRAS, G. G.; SABO, P. J.; HICKEY, M. J.; SINGH, D. C.; STOVER, C. K. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. **Journal of bacteriology**, v. 178, n. 5, p. 1274–82, mar. 1996.

MÄKINEN, V.; SALMELA, L.; YLINEN, J. Normalized N50 assembly metric using gap-restricted co-linear chaining. **BMC Bioinformatics**, v. 13, n. 1, p. 255, 2012.

MALAMA, S.; MUMA, J. B.; GODFROID, J. A review of tuberculosis at the wildlife-livestock-human interface in Zambia. **Infectious diseases of poverty**, v. 2, n. 1, p. 13, jan. 2013.

MCGUIRE, A.; WEINER, B.; PARK, S.; WAPINSKI, I.; RAMAN, S.; DOLGANOV, G.; PETERSON, M.; RILEY, R.; ZUCKER, J.; ABEEL, T.; WHITE, J.; SISK, P.; STOLTE, C.; KOEHRSEN, M.; YAMAMOTO, R. T.; IACOBELLI-MARTINEZ, M.; KIDD, M. J.; MAER, A. M.; SCHOOLNIK, G. K.; REGEV, A.; GALAGAN, J. Comparative analysis of *Mycobacterium* and related actinomycetes yields insight into the evolution of *Mycobacterium tuberculosis* pathogenesis. **BMC Genomics**, v. 13, n. 1, p. 120, 2012.

MEDINA, E.; RYAN, L.; LACOURSE, R.; NORTH, R. J. Superior virulence of *Mycobacterium bovis* over *Mycobacterium tuberculosis* (Mtb) for Mtb-resistant and Mtb-susceptible mice is manifest as an ability to cause extrapulmonary disease. **Tuberculosis**, v. 86, n. 1, p. 20–27, jan. 2006.

MICHEL, A. L.; BENGIS, R. G.; KEET, D. F.; HOFMEYER, M.; KLERK, L. M. de; CROSS, P. C.; JOLLES, A. E.; COOPER, D.; WHYTE, I. J.; BUSS, P.; GODFROID, J. Wildlife tuberculosis in South African conservation areas: Implications and challenges. **Veterinary Microbiology**, v. 112, n. 2–4, p. 91–100, 25 fev. 2006.

MIGNARD, S.; PICHAT, C.; CARRET, G. *Mycobacterium bovis* Infection, Lyon, France. **Emerging Infectious Diseases**, v. 12, n. 9, p. 1431–1433, 2006.

MILLER, R. S.; SWEENEY, S. J. *Mycobacterium bovis* (bovine tuberculosis) infection in North American wildlife: current status and opportunities for mitigation of risks of further infection in wildlife populations. **Epidemiology and infection**, v. 141, n. 7, p. 1357–70, jul. 2013.

MONTALI, R. J.; MIKOTA, S. K.; CHENG, L. I. *Mycobacterium tuberculosis* in zoo and wildlife species. **Revue scientifique et technique (International Office of Epizootics)**, v. 20, n. 1, p. 291–303, abr. 2001.

MORATO, F.; IKUTA, C. Y.; GONÇALES, A. P.; SOUZA, G. O. de; AMAKU, M.; CORTEZ, A.; HEINEMANN, M. B.; FERREIRA NETO, J. S.; NETO, J. S. F. Evaluation of optimal storage time of tuberculous lesions stored in sodium borate. **Semina: Ciências Agrárias**, v. 37, n. 5Supl2, p. 3685, 9 nov. 2016.

MOSTOWY, S.; INWALD, J.; GORDON, S.; MARTIN, C.; WARREN, R.; KREMER, K.; COUSINS, D.; BEHR, M. A. Revisiting the evolution of *Mycobacterium bovis*. **Journal of bacteriology**, v. 187, n. 18, p. 6386–95, set. 2005.

MÜLLER, B.; HILTY, M.; BERG, S.; GARCIA-PELAYO, M. C.; DALE, J.; BOSCHIROLI, M. L.; CADMUS, S.; NGANDOLO, B. N. R.; GODREUIL, S.; DIGUIMBAYE-DJAIBÉ, C.; KAZWALA, R.; BONFOH, B.; NJANPOP-LAFOURCADE, B. M.; SAHRAOUI, N.; GUETARNI, D.; ASEFFA, A.; MEKONNEN, M. H.; RAZANAMPARANY, V. R.; RAMAROKOTO, H.; DJØNNE, B.; OLOYA, J.; MACHADO, A.; MUCAVELE, C.; SKJERVE, E.; PORTAELS, F.; RIGOUTS, L.; MICHEL, A.; MÜLLER, A.; KÄLLENIS, G.; VAN HELDEN, P. D.; HEWINSON, R. G.; ZINSSTAG, J.; GORDON, S. V.; SMITH, N. H. African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. **Journal of bacteriology**, v. 191, n. 6, p. 1951–60, mar. 2009.

NAMOUCHE, A.; DIDELOT, X.; SCHÖCK, U.; GICQUEL, B.; ROCHA, E. P. C. After the bottleneck: Genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. **Genome research**, v. 22, n. 4, p. 721–34, 1 abr. 2012.

NEDELTCHEV, G. G.; RAGHUNAND, T. R.; JASSAL, M. S.; LUN, S.; CHENG, Q.-J.; BISHAI, W. R. Extrapulmonary Dissemination of *Mycobacterium bovis* but Not *Mycobacterium tuberculosis* in a Bronchoscopic Rabbit Model of Cavitory Tuberculosis. **Infection and Immunity**, v. 77, n. 2, p. 598–603, 1 fev. 2009.

OLEA-POPELKA, F.; MUWONGE, A.; PERERA, A.; DEAN, A. S.; MUMFORD, E.; ERLACHER-VINDEL, E.; FORCELLA, S.; SILK, B. J.; DITIU, L.; EL IDRISSE, A.; RAVIGLIONE, M.; COSIVI, O.; LOBUE, P.; FUJIWARA, P. I. Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis* - a call for action. **The Lancet Infectious Diseases**, v. 17, n. 1, p. e21–e25, jan. 2017.

PAI, M.; BEHR, M. A.; DOWDY, D.; DHEDA, K.; DIVANGAHL, M.; BOEHME, C. C.; GINSBERG, A.; SWAMINATHAN, S.; SPIGELMAN, M.; GETAHUN, H.; MENZIES, D.; RAVIGLIONE, M. Tuberculosis. **Nature Reviews Disease Primers**, v. 2, p. 16076, 27 out. 2016.

PALMER, M. V. *Mycobacterium bovis*: Characteristics of Wildlife Reservoir Hosts. **Transboundary and Emerging Diseases**, v. 60, p. 1–13, nov. 2013.

PATANÉ, J. S. L.; MARTINS, J.; CASTELÃO, A. B.; NISHIBE, C.; MONTERA, L.; BIGI, F.; ZUMÁRRAGA, M. J.; CATALDI, A. A.; JUNIOR, A. F.; ROXO, E.; OSÓRIO, A. L. A. R.; JORGE, K. S.; THACKER, T. C.; ALMEIDA, N. F.; ARAÚJO, F. R.; SETUBAL, J. C.; SETUBAL, J. C. Patterns and Processes of *Mycobacterium bovis* Evolution Revealed by Phylogenomic Analyses. **Genome Biology and Evolution**, v. 9, n. 3, p. 521–535, 13 mar. 2017.

PEPPERELL, C. S.; CASTO, A. M.; KITCHEN, A.; GRANKA, J. M.; CORNEJO, O. E.; HOLMES, E. C.; BIRREN, B.; GALAGAN, J.; FELDMAN, M. W. The Role of Selection in Shaping Diversity of Natural *M. tuberculosis* Populations. **PLoS Pathogens**, v. 9, n. 8, p. e1003543, 15 ago. 2013.

PERIWAL, V.; PATOWARY, A.; VELLARIKKAL, S. K.; GUPTA, A.; SINGH, M.; MITTAL, A.; JEYAPPAUL, S.; CHAUHAN, R. K.; SINGH, A. V.; SINGH, P. K.; GARG, P.; KATOCH, V. M.; KATOCH, K.; CHAUHAN, D. S.; SIVASUBBU, S.; SCARIA, V. Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of *Mycobacterium tuberculosis* Pangenome. **PLOS ONE**, v. 10, n. 4, p. e0122979, 8 abr. 2015.

PHELAN, J. E.; COLL, F.; BERGVAL, I.; ANTHONY, R. M.; WARREN, R.; SAMPSON, S. L.; GEY VAN PITTIUS, N. C.; GLYNN, J. R.; CRAMPIN, A. C.; ALVES, A.; BESSA, T. B.; CAMPINO, S.; DHEDA, K.; GRANDJEAN, L.; HASAN, R.; HASAN, Z.; MIRANDA, A.; MOORE, D.; PANAIOTOV, S.; PERDIGAO, J.; PORTUGAL, I.; SHEEN, P.; DE OLIVEIRA SOUSA, E.; STREICHER, E. M.; VAN HELDEN, P. D.; VIVEIROS, M.; HIBBERD, M. L.; PAIN, A.; MCNERNEY, R.; CLARK, T. G. Recombination in *pe/ppe* genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. **BMC genomics**, v. 17, n. 1, p. 151, 29 fev. 2016.

RAMAN, S.; HAZRA, R.; DASCHER, C. C.; HUSSON, R. N. Transcription regulation by the *Mycobacterium tuberculosis* alternative sigma factor SigD and its role in virulence. **Journal of bacteriology**, v. 186, n. 19, p. 6605–16, 1 out. 2004.

RAMOS, D. F.; TAVARES, L.; DA SILVA, P. E. A.; DELLAGOSTIN, O. A. Molecular typing of *Mycobacterium bovis* isolates: a review. **Brazilian Journal of Microbiology: [publication of the Brazilian Society for Microbiology]**, v. 45, n. 2, p. 365–72, 2014.

REHREN, G.; WALTERS, S.; FONTAN, P.; SMITH, I.; ZÁRRAGA, A. M. Differential gene expression between *Mycobacterium bovis* and *Mycobacterium tuberculosis*. **Tuberculosis (Edinburgh, Scotland)**, v. 87, n. 4, p. 347–59, jul. 2007.

RENEWICK, A. R.; WHITE, P. C. L.; BENGIS, R. G. Bovine tuberculosis in southern African wildlife: a multi-species host-pathogen system. **Epidemiology and infection**, v. 135, n. 4, p. 529–40, maio 2007.

RILEY, D. R.; ANGIUOLI, S. V.; CRABTREE, J.; DUNNING HOTOPP, J. C.; TETTELIN, H. Using Sybil for interactive comparative genomics of microbes on the web. **Bioinformatics (Oxford, England)**, v. 28, n. 2, p. 160–6, 15 jan. 2012.

ROCHA, V. C. F. **Avaliação do spoligotyping, miru-vntr e multispacer sequence typing na discriminação de isolados autóctones de *mycobacterium bovis***. 2013. Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, São Paulo, 2013.

RODRÍGUEZ, S.; ROMERO, B.; BEZOS, J.; DE JUAN, L.; ÁLVAREZ, J.; CASTELLANOS, E.; MOYA, N.; LOZANO, F.; GONZÁLEZ, S.; SÁEZ-LLORENTE, J. L.; MATEOS, A.; DOMÍNGUEZ, L.; ARANAZ, A.; SPANISH NETWORK ON SURVEILLANCE AND MONITORING OF ANIMAL TUBERCULOSIS. High spoligotype diversity within a *Mycobacterium bovis* population: Clues to understanding the demography of the pathogen in Europe. **Veterinary Microbiology**, v. 141, n. 1–2, p. 89–95, 24 fev. 2010.

RODRIGUEZ-CAMPOS, S.; SCHÜRCH, A. C.; DALE, J.; LOHAN, A. J.; CUNHA, M. V.; BOTELHO, A.; CRUZ, K. De; BOSCHIROLI, M. L.; BONIOTTI, M. B.; PACCIARINI, M.; GARCIA-PELAYO, M. C.; ROMERO, B.; DE JUAN, L.; DOMÍNGUEZ, L.; GORDON, S. V.; VAN SOOLINGEN, D.; LOFTUS, B.; BERG, S.; HEWINSON, R. G.; ARANAZ, A.; SMITH, N. H. European 2 – A clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. **Infection, Genetics and Evolution**, v. 12, n. 4, p. 866–872, jun. 2012.

RODRIGUEZ-CAMPOS, S.; NAVARRO, Y.; ROMERO, B.; DE JUAN, L.; BEZOS, J.; MATEOS, A.; GOLBY, P.; SMITH, N. H.; HEWINSON, G. R.; DOMÍNGUEZ, L.; GARCÍA-DE-VIDEIRA, D.; ARANAZ, A. Splitting of a Prevalent *Mycobacterium bovis* Spoligotype by Variable-Number Tandem-Repeat Typing Reveals High Heterogeneity in an Evolving Clonal Group. **Journal of Clinical Microbiology**, v. 51, n. 11, p. 3658–3665, nov. 2013.

RODRIGUEZ-CAMPOS, S.; SMITH, N. H.; BONIOTTI, M. B.; ARANAZ, A. Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics

and legislation of bovine tuberculosis. **Research in veterinary science**, v. 97 Suppl, p. S5–S19, out. 2014.

RUE-ALBRECHT, K.; MAGEE, D. A.; KILLICK, K. E.; NALPAS, N. C.; GORDON, S. V.; MACHUGH, D. E. Comparative functional genomics and the bovine macrophage response to strains of the mycobacterium genus. **Frontiers in immunology**, v. 5, p. 536, 2014.

RUSSELL, D. G.; VANDERVEN, B. C.; LEE, W.; ABRAMOVITCH, R. B.; KIM, M.; HOMOLKA, S.; NIEMANN, S.; ROHDE, K. H. *Mycobacterium tuberculosis* wears what it eats. **Cell host & microbe**, v. 8, n. 1, p. 68–76, 22 jul. 2010.

SHARMA, S.; RYNDAK, M. B.; AGGARWAL, A. N.; YADAV, R.; SETHI, S.; MASI, S.; LAAL, S.; VERMA, I. Transcriptome analysis of mycobacteria in sputum samples of pulmonary tuberculosis patients. **PLOS ONE**, v. 12, n. 3, p. e0173508, 10 mar. 2017.

SHIMIZU, E.; MACÍAS, A.; PAOLICCHI, F.; MAGNANO, G.; ZAPATA, L.; FERNÁNDEZ, A.; CANAL, A.; GARBACCIO, S.; CATALDI, A.; CAIMI, K.; ZUMÁRRAGA, M.; SHIMIZU, E.; MACÍAS, A.; PAOLICCHI, F.; MAGNANO, G.; ZAPATA, L.; FERNÁNDEZ, A.; CANAL, A.; GARBACCIO, S.; CATALDI, A.; CAIMI, K.; ZUMÁRRAGA, M. Genotyping *Mycobacterium bovis* from cattle in the Central Pampas of Argentina: temporal and regional trends. **Memórias do Instituto Oswaldo Cruz**, v. 109, n. 2, p. 236–245, abr. 2014.

SILVA, M. R.; ROCHA, A. da S.; DA COSTA, R. R.; DE ALENCAR, A. P.; DE OLIVEIRA, V. M.; FONSECA JÚNIOR, A. A.; SALES, M. L.; ISSA, M. de A.; FILHO, P. M. S.; PEREIRA, O. T. V.; DOS SANTOS, E. C.; MENDES, R. S.; FERREIRA, A. M. de J.; MOTA, P. M. P. C.; SUFFYS, P. N.; GUIMARÃES, M. D. C.; SUFFYS, P. N.; GUIMARÃES, M. D. C. Tuberculosis patients co-infected with *Mycobacterium bovis* and *Mycobacterium tuberculosis* in an urban area of Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 108, n. 3, p. 321, maio 2013.

SIMPSON, J. T.; WONG, K.; JACKMAN, S. D.; SCHEIN, J. E.; JONES, S. J. M.; BIROL, I. ABySS: A parallel assembler for short read sequence data. **Genome Research**, v. 19, n. 6, p. 1117–1123, 2009.

SMITH, N. H.; KREMER, K.; INWALD, J.; DALE, J.; DRISCOLL, J. R.; GORDON, S. V.; VAN SOOLINGEN, D.; HEWINSON, R. G.; SMITH, J. M. Ecotypes of the *Mycobacterium tuberculosis* complex. **Journal of theoretical biology**, v. 239, n. 2, p. 220–5, 21 mar. 2006a.

SMITH, N. H.; GORDON, S. V.; DE LA RUA-DOMENECH, R.; CLIFTON-HADLEY, R. S.; HEWINSON, R. G. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. **Nature reviews. Microbiology**, v. 4, n. 9, p. 670–81, set. 2006b.

SMITH, N. H.; HEWINSON, R. G.; KREMER, K.; BROSCHE, R.; GORDON, S. V. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. **Nature Reviews Microbiology**, v. 7, n. 7, p. 537–544, 1 jul. 2009.

SMITH, N. H.; BERG, S.; DALE, J.; ALLEN, A.; RODRIGUEZ, S.; ROMERO, B.; MATOS, F.; GHEBREMICHAEL, S.; KAROUI, C.; DONATI, C.; MACHADO, A. da C.; MUCAVELE, C.; KAZWALA, R. R.; HILTY, M.; CADMUS, S.; NGANDOLO, B. N. R.; HABTAMU, M.; OLOYA, J.; MULLER, A.; MILIAN-SUAZO, F.; ANDRIEVSKAIA, O.; PROJAHN, M.; BARANDIARÁN, S.; MACÍAS, A.; MÜLLER, B.; ZANINI, M. S.; IKUTA, C. Y.; RODRIGUEZ, C. A. R.; PINHEIRO, S. R.; FIGUEROA, A.; CHO, S.-N.; MOSAVARI, N.; CHUANG, P.-C.; JOU, R.; ZINSSTAG, J.; VAN SOOLINGEN, D.; COSTELLO, E.; ASEFFA, A.; PROAÑO-PEREZ, F.; PORTAELS, F.; RIGOUTS, L.; CATALDI, A. A.; COLLINS, D. M.; BOSCHIROLI, M. L.; HEWINSON, R. G.; NETO, J. S. F.; SURUJBALLI, O.; TADYON, K.; BOTELHO, A.; ZÁRRAGA, A. M.; BULLER, N.; SKUCE, R.; MICHEL, A.; ARANAZ, A.; GORDON, S. V.; JEON, B.-Y.; KÄLLENIUS, G.; NIEMANN, S.; BONIOTTI, M. B.; VAN HELDEN, P. D.; HARRIS, B.; ZUMÁRRAGA, M. J.; KREMER, K. European 1: A globally important clonal complex of *Mycobacterium bovis*. **Infection, Genetics and Evolution**, v. 11, n. 6, p. 1340–1351, ago. 2011.

SMITH, N. H. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. **Infection, Genetics and Evolution**, v. 12, n. 4, p. 857–865, jun. 2012.

SOBRAL, L. F.; DUARTE, R. S.; VIEIRA, G. B. de O.; SILVA, M. G. da; BOECHAT, N.; FONSECA, L. de S. Identificação de *Mycobacterium bovis* em cepas micobacterianas isoladas espécimes clínicos humanos em um complexo hospitalar na cidade do Rio de Janeiro. **Jornal Brasileiro de Pneumologia**, v. 37, n. 5, p. 664–668, out. 2011.

SUPPLY, P.; ALLIX, C.; LESJEAN, S.; CARDOSO-OELEMANN, M.; RÜSCH-GERDES, S.; WILLERY, E.; SAVINE, E.; DE HAAS, P.; VAN DEUTEKOM, H.; RORING, S.; BIFANI, P.; KUREPINA, N.; KREISWIRTH, B.; SOLA, C.; RASTOGI, N.; VATIN, V.; GUTIERREZ, M. C.; FAUVILLE, M.; NIEMANN, S.; SKUCE, R.; KREMER, K.; LOCHT, C.; VAN SOOLINGEN, D. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. **Journal of clinical microbiology**, v. 44, n. 12, p. 4498–510, dez. 2006.

SUPPLY, P.; MARCEAU, M.; MANGENOT, S.; ROCHE, D.; ROUANET, C.; KHANNA, V.; MAJLESSI, L.; CRISCUOLO, A.; TAP, J.; PAWLIK, A.; FIETTE, L.; ORGEUR, M.; FABRE, M.; PARMENTIER, C.; FRIGUI, W.; SIMEONE, R.; BORITSCH, E. C.; DEBRIE, A.-S.; WILLERY, E.; WALKER, D.; QUAIL, M. A.; MA, L.; BOUCHIER, C.; SALVIGNOL, G.; SAYES, F.; CASCIOFERRO, A.; SEEMANN, T.; BARBE, V.; LOCHT, C.; GUTIERREZ, M.-C.; LECLERC, C.; BENTLEY, S. D.; STINEAR, T. P.; BRISSE, S.; MÉDIGUE, C.; PARKHILL, J.; CRUVEILLER, S.; BROSCHE, R. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. **Nature genetics**, v. 45, n. 2, p. 172–9, 6 fev. 2013.

TATUSOV, R. L.; GALPERIN, M. Y.; NATALE, D. A.; KOONIN, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. **Nucleic acids research**, v. 28, n. 1, p. 33–6, 1 jan. 2000.

UNE, Y.; MORI, T. Tuberculosis as a zoonosis from a veterinary perspective. **Comparative Immunology, Microbiology and Infectious Diseases**, v. 30, n. 5–6, p. 415–425, 2007.

VAN HIJUM, S. A. F. T.; ZOMER, A. L.; KUIPERS, O. P.; KOK, J. Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. **Nucleic Acids Research**, v. 33, n. Web Server, p. W560–W566, 1 jul. 2005.

VAN SOOLINGEN, D.; DE HAAS, P. E.; HERMANS, P. W.; VAN EMBDEN, J. D. DNA fingerprinting of *Mycobacterium tuberculosis*. **Methods in enzymology**, v. 235, p. 196–205, 1994.

VOSKUIL, M. I.; SCHNAPPINGER, D.; RUTHERFORD, R.; LIU, Y.; SCHOOLNIK, G. K. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. **Tuberculosis**, v. 84, n. 3, p. 256–262, 2004.

WANZALA, S. I.; NAKAVUMA, J.; TRAVIS, D. A.; KIA, P.; OGWANG, S.; SREEVATSAN, S. Draft Genome Sequences of *Mycobacterium bovis* BZ 31150 and *Mycobacterium bovis* B2 7505, Pathogenic Bacteria Isolated from Archived Captive Animal Bronchial Washes and Human Sputum Samples in Uganda. **Genome Announcements**, v. 3, n. 5, p. e01102-15, 29 out. 2015.

WARREN, R. M.; GEY VAN PITTIUS, N. C.; BARNARD, M.; HESSELING, A.; ENGELKE, E.; DE KOCK, M.; GUTIERREZ, M. C.; CHEGE, G. K.; VICTOR, T. C.; HOAL, E. G.; VAN HELDEN, P. D. Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference. **International Journal of Tuberculosis and Lung Disease**, v. 10, n. 7, p. 818–822, 2006.

WENCES, A. H.; SCHATZ, M. C.; HERNANDEZ WENCES, A.; SCHATZ, M. C. Metassembler: Merging and optimizing de novo genome assemblies. **Genome Biology**, v. 16, n. 1, p. 207, 2015.

WENIGER, T.; KRAWCZYK, J.; SUPPLY, P.; NIEMANN, S.; HARMSSEN, D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. **Nucleic Acids Research**, v. 38, n. Web Server, p. W326–W331, 1 jul. 2010.

WHO. **Global Tuberculosis Report 2016**. Genebra, Suíça, 214p, 2016.

XIA, E.; TEO, Y.-Y.; ONG, R. T.-H. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. **Genome medicine**, v. 8, n. 1, p. 19, 17 fev. 2016.

YIMER, S. A.; NAMOUCHI, A.; ZEGEYE, E. D.; HOLM-HANSEN, C.; NORHEIM, G.; ABEBE, M.; ASEFFA, A.; TØNJUM, T. Deciphering the recent phylogenetic expansion of the originally deeply rooted *Mycobacterium tuberculosis* lineage 7. **BMC Evolutionary Biology**, v. 16, n. 1, p. 146, 30 dez. 2016.

ZHANG, J.; ABADIA, E.; REFREGIER, G.; TAJAJ, S.; BOSCHIROLI, M. L.; GUILLARD, B.; ANDREMONT, A.; RUIMY, R.; SOLA, C. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of “spoligotyping” with new spacers and a microbead-based hybridization assay. **Journal of Medical Microbiology**, v. 59, n. 3, p. 285–294, 1 mar. 2010.

ZHU, L.; ZHONG, J.; JIA, X.; LIU, G.; KANG, Y.; DONG, M.; ZHANG, X.; LI, Q.; YUE, L.; LI, C.; FU, J.; XIAO, J.; YAN, J.; ZHANG, B.; LEI, M.; CHEN, S.; LV, L.; ZHU, B.; HUANG, H.; CHEN, F. Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. **Nucleic Acids Research**, v. 44, n. 2, p. 730–743, 29 jan. 2016.

ZIMPEL, C. K.; BRUM, J. S.; DE SOUZA FILHO, A. F.; BIONDO, A. W.; PEROTTA, J. H.; DIB, C. C.; BONAT, M.; NETO, J. S. F.; BRANDÃO, P. E.; HEINEMANN, M. B.; GUIMARAES, A. M. S. *Mycobacterium bovis* in a European bison (*Bison bonasus*) raises concerns about tuberculosis in Brazilian captive wildlife populations: a case report. **BMC Research Notes**, v. 10, n. 1, p. 91, 10 dez. 2017.

ZUMÁRRAGA, M. J.; ARRIAGA, C.; BARANDIARAN, S.; COBOS-MARÍN, L.; DE WAARD, J.; ESTRADA-GARCIA, I.; FIGUEIREDO, T.; FIGUEROA, A.; GIMÉNEZ, F.; GOMES, H. M.; GONZALEZ-Y-MERCHAND, J. A.; MACÍAS, A.; MILIÁN-SUAZO, F.; RODRÍGUEZ, C. A. R.; SANTILLÁN, M. A.; SUFFYS, P. N.; TRANGONI, M. D.; ZÁRRAGA, A. M.; CATALDI, A. Understanding the relationship between *Mycobacterium bovis* spoligotypes from cattle in Latin American Countries. **Research in Veterinary Science**, v. 94, n. 1, p. 9–21, 2013.

APÊNDICES

APÊNDICE A - Lista dos genomas de *Mycobacterium bovis* utilizados no estudo, número de acesso do GenBank, identificação do hospedeiro de cada isolado, local isolado e forma disponível no banco de dados

Genomas de <i>M. bovis</i>	Número de acesso	Hospedeiro	Local	Forma disponível	Número de contigs
AF2122/97	NC_002945.3	Bovino	Reino Unido	Completo	1
SP38	NZ_CP015773.1	Bovino	Brasil	Completo	1
1595	NZ_CP012095.1	Bovino	Coréia do Sul	Completo	1
30	CP010332.1	Bovino	China	Completo	1
Bz 31150	NZ_JKAM00000000.1	Chimpanzé	Uganda	Draft	136
04-303	NZ_AVSW00000000.1	Porco selvagem	Argentina	Draft	169
09-1191	NZ_JPPF00000000.1	Bovino	Argentina	Draft	86
05-567	NZ_JPFQ00000000.1	Bovino	Argentina	Draft	69
05-566	NZ_JPFR00000000.1	Bovino	Argentina	Draft	89
49-09	NZ_JQES00000000.1	Bovino	Brasil	Draft	61
32-08	NZ_JQER00000000.1	Bovino	Brasil	Draft	83
18-08C	NZ_JQEQ00000000.1	Bovino	Brasil	Draft	63
35	NZ_JQEV00000000.1	Bovino	Brasil	Draft	94
08-08BF2	NZ_JQET00000000.1	Bovino	Brasil	Draft	108
09-1193	NZ_JQEN00000000.1	Bovino	Argentina	Draft	89
534	NZ_JQEM00000000.1	Bovino	Brasil	Draft	72
0822-11	NZ_JQEW00000000.1	Ovino	Argentina	Draft	85
61-09	NZ_JQEX00000000.1	Bovino	Brasil	Draft	68
45-08B	NZ_JQEP00000000.1	Bovino	Brasil	Draft	72
09-1192	NZ_JQEO00000000.1	Bovino	Argentina	Draft	56
50	NZ_JQEU00000000.1	Bovino	Brasil	Draft	84
W-1171	NZ_JXTK00000000.1	Porco selvagem	Coréia do Sul	Draft	50
MbURU-001	NZ_LFGY00000000.1	Bovino	Uruguai	Draft	100
MB4	NZ_CDHE00000000.1	Porco selvagem	Espanha	Draft	118
B-3222	NZ_LNOF00000000.1	Bovino	Coréia do Sul	Draft	56
D-10-02315	NZ_MINA00000000.1	Porco selvagem	França	Draft	159
MB1	NZ_CDHf00000000.1	Bovino	Espanha	Draft	49
MB3	NZ_CDHH00000000.1	Porco selvagem	Espanha	Draft	94

APÊNDICE B - Lista dos genomas de *Mycobacterium bovis* BCG utilizados no estudo, número de acesso do GenBank e forma disponível no banco de dados

(continua)

Genoma	Número de acesso	Forma disponível
<i>M. bovis</i> BCG Pasteur 1173P2	NC_008769.1/AM408590.1	Completo
<i>M. bovis</i> BCG Tokyo 172	NC_012207.1/AP010918.1	Completo
<i>M. bovis</i> BCG ATCC 35743	NZ_CP003494.1/CP003494.1	Completo
<i>M. bovis</i> BCG Mexico	NC_016804.1/CP002095.1	Completo
<i>M. bovis</i> BCG Korea 1168P	NC_020245.2/CP003900.2	Completo
<i>M. bovis</i> BCG Moreau RDJ	NZ_AM412059.1/AM412059.2	Completo

(conclusão)		
Genoma	Número de acesso	Forma disponível
<i>M. bovis</i> BCG_3281	NZ_CP008744.1/CP008744.1	Completo
<i>M. bovis</i> BCG Russia 368	NZ_CP009243.1/CP009243.1	Completo
<i>M. bovis</i> BCG-1 (Russia)	NZ_CP013741.1/CP013741.1	Completo
<i>M. bovis</i> BCG 26	CP010331.1	Completo
<i>M. bovis</i> BCG Tokyo 172 TRCS	NZ_CP014566.1/CP014566.1	Completo
<i>M. bovis</i> BCG-1	NZ_CP011455.1/CP011455.1	Draft
<i>M. bovis</i> BCG China	NZ_AEZE00000000.1	Draft
<i>M. bovis</i> BCG ATCC 35733	NZ_AEZF00000000.1	Draft
<i>M. bovis</i> BCG ATCC 35740	NZ_AEZG00000000.1	Draft
<i>M. bovis</i> BCG Frappier	NZ_AKYQ00000000.1	Draft
<i>M. bovis</i> BCG_jnaf	NZ_JNAF00000000.1	Draft
<i>M. bovis</i> BCG_cys	NZ_CYST00000000.1	Draft
<i>M. bovis</i> BCG Phipps	NZ_CUWN00000000.1	Draft
<i>M. bovis</i> BCG Moreau	NZ_CUWK00000000.1	Draft
<i>M. bovis</i> BCG Birkhaug	NZ_CUWE00000000.1	Draft
<i>M. bovis</i> BCG Sweden	NZ_CUWP00000000.1	Draft
<i>M. bovis</i> BCG Danish	NZ_CUWH00000000.1	Draft
<i>M. bovis</i> BCG Pasteur	NZ_CUWL00000000.1	Draft
<i>M. bovis</i> BCG Glaxo	NZ_CUWJ00000000.1	Draft
<i>M. bovis</i> BCG Russia	NZ_CUWO00000000.1	Draft
<i>M. bovis</i> BCG Prague	NZ_CUWM00000000.1	Draft
<i>M. bovis</i> BCG China	NZ_CUWG00000000.1	Draft
<i>M. bovis</i> BCG Connaught	NZ_CUWF00000000.1	Draft
<i>M. bovis</i> BCG Tice	NZ_CUWQ00000000.1	Draft
<i>M. bovis</i> BCG Japan	NZ_CUWR00000000.1	Draft
<i>M. bovis</i> BCG Frappier	NZ_CUWI00000000.1	Draft

APÊNDICE C - Lista dos genomas de *Mycobacterium tuberculosis* utilizados no estudo, número de acesso do GenBank, local e características dos isolados

(continua)			
Genoma	Número de acesso	Local	Característica
<i>M. tuberculosis</i> H37Rv	NC_000962.3	Reino Unido	Linhagem 4
<i>M. tuberculosis</i> CDC1551	NC_002755.2/AE000516.2	EUA	Linhagem 4
<i>M. tuberculosis</i> F11	NC_009565.1/CP000717.1	África do Sul	Suscetível a drogas
<i>M. tuberculosis</i> KZN 1435	NC_012943.1/CP001658.1	África do Sul	Multi-droga Resistente
<i>M. tuberculosis</i> Haarlem	NC_022350.1/CP001664.1	-	Multi-droga Resistente
<i>M. tuberculosis</i> CCDC5180	NC_017522.1/CP001642.1	China	Multi-droga Resistente
<i>M. tuberculosis</i> 7199-99	NC_020089.1/HE663067.1	Alemanha	-
<i>M. tuberculosis</i> Beijing/NITR203	NC_021054.1/CP005082.1	Índia	-
<i>M. tuberculosis</i> CCDC5079	NC_021251.1/CP002884.1	China	Sensível a drogas

(conclusão)			
Genoma	Número de acesso	Local	Característica
<i>M. tuberculosis</i> EAI5	NC_021740.1/CP006578.1	Índia	-
<i>M. tuberculosis</i> KIT87190	NZ_CP007809.1/CP007809.1	Coréia do Sul	-
<i>M. tuberculosis</i> 96075	NZ_CP009426.1/CP009426.1	China	Linhagem 2
<i>M. tuberculosis</i> 96121	NZ_CP009427.1/CP009427.1	Filipinas	Linhagem 1
<i>M. tuberculosis</i> Kurono	NZ_AP014573.1/AP014573.1	Japão	Linhagem 4
<i>M. tuberculosis</i> SCAID 187.0	NZ_CP012506.2/CP012506.2	Cazaquistão	Multi-droga Resistente
<i>M. tuberculosis</i> 2242	CP010335.1	China	Linhagem 2
<i>M. tuberculosis</i> 2279	CP010336.1	China	Linhagem 2
<i>M. tuberculosis</i> 22115	NZ_CP010337.1/CP010337.1	China	Linhagem 4
	NZ_CP010338.1/CP010338.1	China	Linhagem 4
<i>M. tuberculosis</i> 37004			
<i>M. tuberculosis</i> 26105	NZ_CP010340.1/CP010340.1	China	Linhagem 3
<i>M. tuberculosis</i> Beijing-like	CP010873.1	Colômbia	-
<i>M. tuberculosis</i> ZMC13-264	NZ_CP009100.1/CP009100.1	China	Multi-droga Resistente
<i>M. tuberculosis</i> ZMC13-88	NZ_CP009101.1/CP009101.1	China	Multi-droga Resistente

APÊNDICE D - Lista dos outros genomas do MTBC e *Mycobacterium canetti* utilizados no estudo, número de acesso em bancos de dados públicos, forma disponível no banco de dados e características

Genoma	Número de acesso	Forma disponível	Característica
<i>M. africanum</i> GM041182	NC_015758.1/FR878060.1	Completo	Linhagem 6
<i>M. africanum</i> MAL010070	NZ_JLAZ000000000.1	Draft	Linhagem 5
<i>M. canettii</i> CIPT 140010059	NC_015848.1/HE572590.1	Completo	-
<i>M. microti</i> 12	CP010333.1	Draft	-
<i>M. caprae</i> MB2	NZ_CDHG000000000.1	Draft	-
<i>M. orygis</i> 112400015	NZ_APKD000000000.1	Draft	-
<i>M. mungi</i> BM22813	NZ_LXTB000000000.1	Draft	-
<i>M. suricattae</i>	ERX1047373*	Reads	-

Nota: Números de acesso provenientes do banco de dados GenBank do NCBI, exceto o marcado com asterisco *M. suricattae* proveniente do ENA.

APÊNDICE E - Resultado dos diferentes *k-mers* obtidos na montagem pelo software ABySS

<i>k-mer</i>	Qualidade	n:N50	N50	Tamanho mínimo dos <i>contigs</i>	Tamanho máximo dos <i>contigs</i>
50	Boa	9	196.003	500	392.411
60	Boa	8	214.934	500	392.981
70	Melhor	8	214.974	500	395.792
80	Melhor	7	215.014	500	500.869
90	Melhor	8	208.012	500	500.955

APÊNDICE F - Resultado da montagem do genoma *Mycobacterium bovis* SP38 pelo software ABySS

Características de montagem	Genoma <i>M. bovis</i> SP38
Nº de <i>contigs</i>	51
Total de bases	4.373.351
Menor sequência (bases)	500
Maior sequência (bases)	500.955
N50	208.012
N:N50	8
<i>k-mer</i>	90

APÊNDICE G - Dendograma de *Mycobacterium bovis* SP38 realizado com padrão de espoligotipo e MIRU-VNTR com outros genomas de *M. bovis* disponíveis no banco de dados (<http://www.miru-vntrplus.org>)

Apêncide H - Genomas de disponíveis como *Mycobacterium bovis* no GenBank que apresentaram padrões de RDs (Regiões de diferença) identificando-os como outras espécies

Genomas de <i>M. bovis</i>	Número de acesso	Perfil de RDs	Genoma corrigido
ATCC BAA-935	NZ_CP009449.1/ CP009449.1	RD1 ausente (192 pb) RD4 ausente (267 pb) RD9 ausente (107 pb)	<i>M. bovis</i> BCG
B2 7505	NZ_JKAL00000000.1	RD1 presente (146 pb) RD4 presente (172 pb) RD9 presente (235 pb) RD12 presente (369 pb)	<i>M. tuberculosis</i>
MAL_010093	NZ_JLAP00000000.1	RD1 presente (146 pb) RD4 presente (172 pb) RD9 ausente (108 pb) RD12 presente (369 pb) 1mic presente (195 pb) 2seal ausente (sem sequência)	<i>M. africanum</i>

APÊNDICE I - Dendograma demonstrando árvore filogenética analisando os padrões de *spoligotyping* dos genomas de *M. bovis*

NJ-Tree, MINU-VNTN (24): Categorical (1), Spoligo: Categorical (1), RD: Categorical (1)