

**UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL
CURSO DE DOUTORADO**

**CONTRIBUIÇÕES DA ANÁLISE GENÔMICA PARA O
CONTROLE DA TUBERCULOSE BOVINA**

Ana Beatriz Canevari Castelão

Campo Grande
Mato Grosso do Sul – Brasil
Março de 2016

**UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL
CURSO DE DOUTORADO**

**CONTRIBUIÇÕES DA ANÁLISE GENÔMICA PARA O
CONTROLE DA TUBERCULOSE BOVINA**

Contributions of genomic analysis for the control of bovine tuberculosis

Ana Beatriz Canevari Castelão

Orientador: Dr. Flávio Ribeiro de Araújo

Tese apresentada à
Universidade Federal de Mato
Grosso do Sul, como requisito
à obtenção do título de
Doutor em Ciência Animal.

Área de concentração: Saúde
Animal.

Campo Grande

Mato Grosso do Sul – Brasil

Março de 2016

Certificado de aprovação

ANA BEATRIZ CANEVARI CASTELÃO

Contribuições da análise genômica para o controle da tuberculose bovina

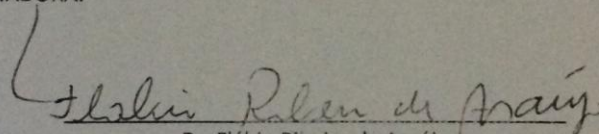
Contributions of genomic analysis for the control of bovine tuberculosis

Tese apresentada à Universidade Federal de Mato Grosso do Sul, como requisito à obtenção do título de doutora em Ciência Animal.

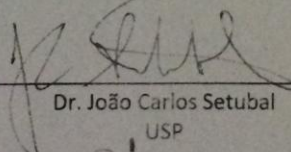
Área de concentração: Saúde Animal.

Aprovado(a) em: 01/03/2016

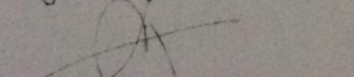
BANCA EXAMINADORA:



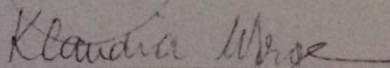
Dr. Flávio Ribeiro de Araújo
(EMBRAPA CNPQC) – (Orientador)



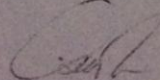
Dr. João Carlos Setubal
USP



Dr. Nalvo Franco de Almeida Junior
UFMS



Dra. Klaudia dos Santos Gonçalves Jorge
UFMS



Dr. Carlos Alberto do Nascimento Ramos
UFMS

A Felicidade exige Valentia

Augusto Cury (2003)

Posso ter defeitos, viver ansioso e ficar irritado algumas vezes, mas não esqueço de que minha vida é a maior empresa do mundo, e posso evitar que ela vá à falência.

Ser feliz é reconhecer que vale a pena viver apesar de todos os desafios, incompreensões e períodos de crise. Ser feliz é deixar de ser vítima dos problemas e se tornar um autor da própria história.

É atravessar desertos fora de si, mas ser capaz de encontrar um oásis no recôndito da sua alma.

É agradecer a Deus a cada manhã pelo milagre da vida.

Ser feliz é não ter medo dos próprios sentimentos.

É saber falar de si mesmo.

É ter coragem para ouvir um "não".

É ter segurança para receber uma crítica, mesmo que injusta.

Pedras no caminho?

Guardo todas, um dia vou construir um castelo...

Aos meus pais Clóvis e Marta por sempre estarem ao meu lado, pelo carinho, dedicação, paciência, incentivo, compreensão... por sempre investirem na minha formação e principalmente por me ensinarem o verdadeiro sentido da vida: o amor,

A vocês agradeço por tudo que sou!

Agradecimentos

Os meus sinceros agradecimentos ao Dr. Flávio Ribeiro de Araújo, meu orientador, pelos ensinamentos, auxílio, paciência, tolerância, exemplo profissional, convívio e amizade.

A toda equipe do Laboratório de Imunologia Aplicada da Embrapa Gado de Corte.

Aos Dr. Nalvo Almeida Júnior e Dr. João Carlos Setubal, equipe da Faculdade de Computação da Universidade Federal de Mato Grosso do Sul e Departamento de Bioquímica da Universidade de São Paulo por todo suporte científico na área de bioinformática e filogenia.

Ao pesquisador Dr. Martín José Zumárraga e equipe do Instituto de Biotecnologia, INTA Castelar, Argentina que participaram diretamente da realização deste trabalho.

Ao Dr. Tyler Thacker e equipe de Tuberculose Bovina do Departamento de Agricultura dos Estados Unidos (USDA, Ames) pelo carinho com que me receberam, conhecimento a mim compartilhado e experiência única de realização do doutorado sanduíche.

Ao Programa de Pós-graduação em Ciência Animal da Universidade Federal de Mato Grosso do Sul pelo apoio técnico, administrativo e pessoal.

A Embrapa Gado de Corte pelo agradável convívio e possibilidade de realização deste trabalho.

A Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul, pelo financiamento deste trabalho (TO0096/2012) e bolsa de doutorado (23/200.656/2013), bem como a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela bolsa de doutorado e bolsa de doutorado sanduíche, a Embrapa Gado de Corte (03.14.00.054.00.00, 02.13.16.002.00.00) e CNPq (443235/2014-7; 305857/2013-4).

Muito obrigada!

RESUMO

CASTELÃO, A. B. C. *Contribuições da análise genômica para o controle da tuberculose bovina*. 2016. 142f. Tese (Doutorado) – Faculdade de Medicina Veterinária e Zootecnia, Programa de Pós Graduação em Ciência Animal, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2016.

Mycobacterium bovis é o agente causador da tuberculose bovina, responsável por uma doença com perdas anuais para a agropecuária global de 3 bilhões de dólares além de sérias repercussões na saúde pública e saúde animal. O sequenciamento genômico e avanços nas técnicas de tipagem molecular têm contribuído para o estudo epidemiológico da tuberculose bovina, por meio da identificação e caracterização de isolados de *M. bovis*. O genoma de *M. bovis* é 99,95% idêntico em sequência de nucleotídeos ao genoma de *Mycobacterium tuberculosis*. Apesar da elevada relação genética global, essas espécies exibem uma grande variedade de fenótipos; a deleção da informação genética, juntamente com a presença de polimorfismos de um único nucleotídeo (SNPs) ao longo do cromossoma das espécies provavelmente explica a diferença na gama de hospedeiros, virulência e imunopatologia encontrada entre estas espécies. O objetivo deste trabalho foi sequenciar, montar, analisar e caracterizar geneticamente genomas micobacterianos, bem como analisá-los filogeneticamente, estabelecendo relações entre os mesmos e comparar um repertório de genes de virulência presentes nos genomas dos isolados *M. bovis* 04-303 e 534 que apresentam perfis patogênicos contrastantes. A sequência do genoma de AN5 foi obtida utilizando a tecnologia MiSeq e a montagem do genoma assistida pelo genoma referência: *M. bovis* AF2122/97. A montagem contém 70 contigs e N50 150.219. O genoma compreende 3.938 sequências codificadoras (CDs) e 48 genes de RNA (3 rRNAs e 45 tRNAs). Nenhuma deleção ou variação genética no isolado *M. bovis* AN5 utilizado para produção do PPDBR apoia a diferença proteômica encontrada quando comparado ao PPDUK. As sequências dos genomas de *M. bovis* 04-303 e 534 foram também obtidas utilizando a tecnologia MiSeq. Todos os genes de ambos genomas foram analisados, utilizando *M. bovis* AF2122/97 como referência. Mutações nucleotídicas foram encontradas em 31 genes de virulência no isolado 04-303: 21 mutações não sinônimas missense (substituição de aminoácidos), 9 mutações de deleção (supressão de aminoácidos) e uma mutação de inserção (inserção de aminoácido). Em relação ao isolado 534, mutações nucleotídicas foram encontradas em 47 genes de virulência: 24 mutações não sinônimas missense, 20 mutações de deleção e 3 mutações de inserção. Dezoito dessas mutações foram comuns em ambos os genomas. As mutações encontradas em genes de virulência estão em concordância com o fenótipo atenuado do isolado *M. bovis* 534. Estas mutações estão apenas parcialmente em consonância com o fenótipo virulento do isolado *M. bovis* 04-303. Análises filogenéticas e de processos evolutivos de 38 genomas sequenciados de *M. bovis* das Américas, Europa, Ásia e África apontaram que isolados norte-americanos estão evoluindo distintamente da maioria dos isolados sul-americanos, além de apresentarem menor pressão de seleção. Através de datação de divergência e da posição basal apresentada por isolados humanos nas árvores filogenéticas evidenciando a origem de *M. bovis* há aproximadamente 10.000 anos atrás, além de corroborar com a afirmativa de *Mycobacterium tuberculosis* como progenitor de *M. bovis*.

ABSTRACT

CASTELÃO, A. B. C. *Contributions of genomic analysis for the control of bovine tuberculosis control*. 2016. 142f. Tese (Doutorado) – Faculdade de Medicina Veterinária e Zootecnia, Programa de Pós Graduação em Ciência Animal, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2016.

Mycobacterium bovis is the causative agent of bovine tuberculosis, a disease responsible for annual losses to global agriculture of \$3 billion and with serious repercussions for public health and animal welfare. The genome sequencing and advances in molecular typing techniques have contributed to the epidemiological study of bovine tuberculosis, through the identification and characterization of *M. bovis* isolates. *M. bovis* genome is 99,95% identical in nucleotide sequence to *Mycobacterium tuberculosis*. Despite their high overall genetic relationship, this species exhibit a wide variety of phenotypes; the deletion of genetic information, together with the presence of a single nucleotide polymorphism (SNP) along chromosomes of the species, probably explain the different host ranges, virulence and immunopathology found among these species. The aim of this study was to sequence, assemble, analyze and characterize genetically mycobacterial genomes and analyze them phylogenetically, establishing relationships between them and compare a repertoire of virulence genes present in the genomes of *M. bovis* isolates 04-303 and 534 with contrasting pathogenic profiles. The genome sequence of AN5 was obtained using MiSeq technology and the assembly was performed with a reference-assisted genome: *M. bovis* AF2122/97 strain. The assembly contains 70 contigs and an N50 of 150,219. The *M. bovis* AN5 genome comprises 3,938 coding sequences (CDSs) and 48 RNA genes (3 rRNAs and 45 tRNAs). Neither gene deletion nor gene variation in *M. bovis* AN5 strains used for production of PPDBR supports the proteomic differences compared with PPDUK. The genome sequence of 04-303 and 534 were also obtained using MiSeq technology. All genes of both *M. bovis* genome were analyzed using AF2122/97 as a reference genome. Nucleotide mutations were found in 31 virulence genes of 04-303: 21 missense mutations (amino acid replacement), 9 deletion mutations (deletion of amino acid) and 1 insertion mutation (insertion of amino acid). Concerning the 534 strain, nucleotide mutations were found in 47 virulence genes: 24 missense mutations, 20 deletion mutations and 3 insertion mutations. Eighteen of these mutations were common in both genomes. Mutations found in virulence genes are in concordance with the attenuated phenotype of the *M. bovis* strain 534. These mutations are only partially in consonance with the hyper-virulent phenotype of the *M. bovis* 04-303 strain. Filogenetic analyses and evolutionary patterns in 38 sequenced genomes of *M. bovis* from the Americas, Europe, Asia, and Africa pointed that US strains are evolving distinctly from most South American lineages, besides presented lower selection pressure. Across divergence time and basal position of human isolates in filogenetic trees evidencing the *M. bovis* origin approximately of 10.000 years, and corroborate with the affirmative of *Mycobacterium tuberculosis* as the progenitor of *M. bovis*.

LISTA DE ABREVEATURAS

+G	Distribuição Gama
+I	Probabilidade de Sítios Invariáveis
AA	Aminoácidos
Ac4SGL	Sulfoglicolípídeos Tetracilados
BAAR-	Bacilos Álcool Ácido Resistentes
BCG	Bacilo Calmette-Guérin
cAMP	Adenosina Mono Fosfato cíclica
CDs	Sequências Codificadoras de Proteínas
CFP10	Proteína de Cultura Filtrada 10 kDa
CG	Citosina e Guanina
CMT	Complexo <i>Mycobacterium tuberculosis</i>
DEL	Deleção
dN	Taxa de Mutações Não Sinônimas
DNA	Ácido Desoxirribonucleico
dS	Taxa de Mutações Sinônimas
ECF	Função Extra Citoplasmática
ERK-NF-kB	Fatores Nucleares do Eixo kB
ESAT-6	Alvo Antigênico Secretor Precoce 6 kDa
ESSS	Tamanho Efetivo das Amostras
ETR	Repetições Exatas em Tandem
HGT	Transferência Horizontal de Genes
HPD	Maior Densidade Posterior

IL	Interleucina
INF- γ	Interferon gama
iNOS	Sintase de Óxido Nítrico
INS	Inserção
LCBs	Blocos Colineares Locais
MCE	Entrada em Células de Mamíferos
MIRU	Unidades Repetitivas de Interespaçadores Micobacterianos
ML	Máxima Verossimilhança
MP	Máxima Parcimônia
NADH	Nicotinamida-Adenina-Dinucleotídio
Nc	Número Efetivo de Códon
NGS	Sequenciamento de Nova Geração
NKT	Células Natural <i>Killer</i>
NSMS	Mutação <i>Non-Synonymous Missense</i>
Nu	Taxa de Novas Substituições devido à Recombinação
PB	Pares de Base
PCR	Reação em Cadeia da Polimerase
PDIM	Dimicocerosate Fitocerosol
PimA	Alfa Fosfatidilinositol manosiltransferase
PIMs	Sulfolipídeos e Fosfatilinositol Manosídeo
PKS	Poliketide Sintase
PNCEBT	Programa Nacional de Controle e Erradicação da Brucelose e Tuberculose
PPD	Derivado Protéico Purificado

PstS	Lipoproteína Transportadora de Fosfato Periplasmático
r/m	Razão entre a Quantidade estimada de Polimorfismos interpostos por um evento de Recombinação em relação a um evento de Mutação
RD	Regiões de Diferenciação
RDAf1	Complexo Clonal África 1
RDAf2	Complexo Clonal África 2
RDEu1	Complexo Clonal Europa 1
RDEu2	Complexo Clonal Europa 2
RecA	Recombinase A
RFLP	Polimorfismos dos Tamanhos de Fragmentos de Restrição
Rho/theta	Razão entre a Taxa de Recombinação e a Taxa de Mutação
RI	Índice de Retenção
RNA	Ácido Ribonucleico
RNAp	RNA polimerase
RNAr	RNA ribossômico
RNAt	RNA transportador
SL	Sulfolipídeos
SNP	Polimorfismo de Única Base
TA	Toxina-Antitoxina
TB	Tuberculose Bovina
TCC	Teste Cervical Comparativo
TCS	Teste Cervical Simples
TDM	Trealose 6,6'dimicolato
Th1	Linfócito T helper 1

TLR	Receptores <i>Toll Like</i>
TNF- α	Fator de Necrose Tumoral alfa
TPC	Teste da Prega Caudal
VNTR	Repetições Sucessivas de Número Variável
WGS	Sequenciamento do Genoma Total
WhiB3	Fator de Transcrição Acessório

SUMÁRIO

1 INTRODUÇÃO	13
1.1 AGENTE ETIOLÓGICO	15
1.2 GENOMA	17
1.3 EVOLUÇÃO E FILOGENIA	14
1.4 REGIÕES DE DIFERENÇA E COMPLEXOS CLONAIIS	20
1.5 VIRULÊNCIA.....	21
1.6 DIAGNÓSTICO.....	25
1.7 EPIDEMIOLOGIA MOLECULAR.....	30
1.7.1 RFLP-IS6110.....	31
1.7.2 Spoligotyping.....	28
1.7.3 MIRU-VNTR	33
1.8 SEQUENCIAMENTO GENÔMICO	34
1.9 OBJETIVOS.....	33
1.9.1 Objetivo geral	33
1.9.2 Objetivos específicos.....	38
REFERÊNCIAS	34
ARTIGO 1 - SEQUENCIAMENTO GENÔMICO DE <i>MYCOBACTERIUM BOVIS</i> CEPA AN5, UTILIZADA PARA A PRODUÇÃO DE DERIVADO PROTÉICO PURIFICADO	43
ARTIGO 2 - SEQUENCIAMENTO GENÔMICO DE <i>MYCOBACTERIUM</i> 04-303, UM ISOLADO ALTAMENTE VIRULENTO DA ARGENTINA.....	49
ARTIGO 3- COMPARAÇÃO DE GENES DE VIRULÊNCIA DE DUAS CEPAS DE <i>MYCOBACTERIUM BOVIS</i> COM PERFIS PATOGÊNICOS CONTRASTANTES	55
ARTIGO 4 - QUANDO SUBSTITUIÇÕES DE DNA SÃO POSITIVAMENTE ENGANOSAS: PAN-GENOMA E INDELS AO RESGATE PARA FILOGENÔMICA E DATAÇÃO DE DIVERGÊNCIAS EM <i>MYCOBACTERIUM BOVIS</i>.....	108

INTRODUÇÃO

A tuberculose, considerada um dos maiores problemas de saúde pública no mundo, é uma doença infecto-contagiosa causada por micobactérias. As micobactérias pertencem à Classe *Actinobacteria*, Ordem *Actinomycetales*, Família *Mycobacteriaceae*, Gênero *Mycobacterium* e fazem parte do complexo *Mycobacterium tuberculosis* (CMT), um grupo de espécies relacionadas constituído, atualmente, por nove espécies – *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium microti*, *Mycobacterium africanum*, *Mycobacterium caprae*, *Mycobacterium pinnipedii* – e, recentemente, três novas espécies foram incluídas ao complexo - *Mycobacterium mungi*, *Mycobacterium suricattae* e *Mycobacterium orygis* (FABRE et al., 2004; SMITH et al., 2006; ALEXANDER et al., 2010; Van INGEN et al., 2012).

A tuberculose bovina é uma zoonose, doença naturalmente transmissível entre os animais vertebrados e o homem, causada por *Mycobacterium bovis*, cujo hospedeiro primário é o bovino. Entretanto diversas espécies de mamíferos domésticos e silvestres, inclusive o homem, são também susceptíveis ao bacilo (BIET et al., 2005; CORNER, 2006). A tuberculose é uma doença de elevada importância econômica para a criação de bovinos uma vez que afeta diretamente a produtividade além de influenciar o comércio internacional dos animais e seus subprodutos (MICHEL et al., 2010). Estima-se que os prejuízos anuais com a tuberculose bovina são da ordem de três bilhões de dólares (GARNIER et al., 2003).

Em bovinos, *M. bovis* é transmitido principalmente por via respiratória, através da inalação de aerossóis contaminados pelo bacilo. Entretanto, além de aerossóis, o animal pode disseminar o agente por descarga nasal, leite, fezes, urina, secreções vaginais e uterinas, e pelo sêmen. O trato digestório também é porta de entrada de *M. bovis*, principalmente em bezerros alimentados com leite proveniente de vacas com mastite tuberculosa e em animais que ingerem água ou forragens contaminadas (POLLOCK et al., 2006). É uma das mais graves e antigas enfermidades infecto-

contagiosas conhecidas e caracteriza-se pela evolução geralmente crônica, inicialmente subclínica e desenvolvimento progressivo de lesões nodulares denominadas tubérculos (GUTIÉRREZ et al., 1998).

As lesões sugestivas de tuberculose bovina são frequentemente encontradas nos pulmões e em linfonodos adjacentes, e são visualizadas durante a necropsia ou durante a inspeção de carcaças em matadouros frigoríficos. Caracterizam-se por nódulos de diâmetro variável entre 1,0 a 3,0 cm ou mais, podendo ser confluyente, de aspecto purulento ou caseoso, de cor branca, cinza ou amarelado, de consistência firme e inodoro (BRASIL, 2006; OIE, 2009). A partir dessas lesões é realizado o isolamento micobacteriano em meio de cultura *Stonebrink*. O modelo característico de crescimento, a morfologia colonial em meios de cultura e a presença de bacilos álcool ácido resistente (BAAR) podem formular o diagnóstico preliminar da infecção por *M. bovis*. Tais colônias podem ser identificadas por meio de provas bioquímicas ou por técnicas de biologia molecular (OIE, 2009).

No Brasil, o Programa Nacional de Controle e Erradicação da Brucelose e Tuberculose Animal (PNCEBT) foi instituído em 2001 com o propósito de diminuir o impacto negativo dessas zoonoses na saúde humana e animal e aumentar a competitividade da pecuária nacional (BRASIL, 2001). O controle da enfermidade é baseado na identificação e eliminação dos animais infectados (teste-abate) (BRASIL, 2001), tomando como base testes alérgicos de tuberculinização intradérmica em bovinos e bubalinos utilizando-se tuberculinas PPD (Derivado Protéico Purificado) (MONAGHAN et al., 1994). O diagnóstico confirmatório é realizado após a morte do animal e requer o isolamento e a identificação do agente etiológico (CORNER, 1994).

As diferenças bioquímicas existentes entre as espécies de micobactérias podem ser utilizadas para confirmar o agente causal da doença. Essas provas bioquímicas apresentam alta especificidade, porém são trabalhosas, requerem culturas novas, puras e abundantes, condições só alcançadas após sucessivos repiques a partir do primo isolamento, que pode se estender por vários meses até a confirmação da espécie (MICHEL; MULLER; van HELDEN, 2010). As técnicas de tipagem molecular

ajudam a estratificar e refinar dados, fornecendo informações mais sensíveis e específicas, que facilitam as atividades epidemiológicas, incluindo a vigilância de doenças, investigações de surtos, identificando padrões de transmissão e fatores de risco entre casos aparentemente díspares, bem como pode revelar a presença de focos provocados por mais de uma cepa (FOXMAN; RILEY, 1998).

Com a globalização da economia, as barreiras à comercialização de produtos de origem animal, antes alfandegárias, são hoje de natureza sanitária. A cada dia, crescem as exigências dos mercados internacionais. No contexto da tuberculose bovina, diversos países já adotam a obrigatoriedade da identificação etiológica de lesões sugestivas da enfermidade detectadas em abatedouros (RADUNZ, 2006; KANEENE et al., 2006).

Nos últimos anos, os avanços da biologia molecular vêm contribuindo com êxito no diagnóstico de enfermidades animais. O sequenciamento de DNA e a identificação de marcadores moleculares têm sido o foco deste desenvolvimento, possibilitando a padronização de técnicas moleculares, que permitem distinguir rapidamente espécies e também cepas da mesma espécie.

1.1 AGENTE ETIOLÓGICO

A tuberculose em bovinos tem como agente etiológico o bacilo *Mycobacterium bovis*, pertencente ao CMT. Estes bacilos são pequenos (0,3 a 0,6 μm de largura por 1 a 4 μm de comprimento), intracelulares facultativos, microaerófilos, imóveis, não esporulados, não flagelados e se apresentam na forma de bastonetes reto ou encurvado. São BAAR, ou seja, uma vez corados pela fucsina a quente não se descoram na presença de uma mistura álcool-ácido em uma técnica denominada de Ziehl-Neelsen. Crescem em meios especiais a base de ovos e amido, enriquecidos com asparagina e contendo verde malaquita para inibir contaminantes. *Mycobacterium bovis* não é capaz de metabolizar glicerol em piruvato como fazem as outras espécies do CMT. A análise da sequência genômica dessa espécie confirmou que todos os

genes requeridos para a formação do piruvato não são funcionais. Desse modo, *M. bovis* cresce melhor na presença de piruvato como fonte de carbono e encontram condições ideais de crescimento à temperatura de 37°C e pH neutro (BARRERA, 2007).

Mycobacterium bovis é um patógeno que habita preferencialmente o meio intracelular de fagócitos mononucleares e, como as demais micobactérias do CMT, é hospedeiro-dependente e não pode multiplicar-se no meio ambiente. Porém, são capazes de se reproduzir *in vitro*. Em meios de cultura à base de ovo é típico o crescimento de colônias pequenas, arredondadas, de coloração amarelo-pálida, com borda irregular e superfície granular (BARRERA, 2007). Na baciloscopia, *M. bovis* aparece como cocobacilos de coloração vermelho intensa (CORNER, 1994).

O bacilo da tuberculose bovina apresenta um envelope celular altamente impermeável, contendo elevado conteúdo lipídico (50 a 60%) o que confere um caráter hidrofóbico e o torna refratário à atividade enzimática. A parede celular bem desenvolvida contém moléculas de peptidoglicano covalentemente ligados a polissacarídeos ramificados, arabinogalactano, cujas extremidades são esterificadas com ácidos graxos de alto peso molecular, os ácidos micólicos, que representam uma importante barreira lipídica. A composição e a quantidade dos componentes da parede celular afetam as taxas de crescimento e a virulência da bactéria (BARRERA, 2007).

Mycobacterium bovis é moderadamente resistente ao calor, permanecendo viável em pastos e esterco por até dois anos, em água por até um ano e até dez meses em produtos de origem animal contaminados. O bacilo também apresenta resistência a desinfetantes compostos por amônia quaternária e clorexidina. Entretanto, é sensível à exposição solar direta em ambiente seco, combatido pela pasteurização do leite e por produtos como hipoclorito de sódio, desinfetantes fenólicos, formólicos e álcool, observando a concentração do produto, tempo de exposição, temperatura e presença de matéria orgânica (ROXO, 1996).

1.2 GENOMA

O sequenciamento do genoma completo de *M. bovis* ocorreu em 2003 por Garnier et al.(2003) e abriu portas para a obtenção de informações genéticas antes não disponíveis. A cepa sequenciada foi *Mycobacterium bovis* AF2122/97, uma cepa virulenta que foi isolada na Grã-Bretanha em 1997 (GORDON et al., 1999). Este isolado de *M. bovis* possui 4.345.492 pb (pares de base) em seu genoma, organizado em um único cromossomo circular com conteúdo médio de guanina e citosina de 65%. Contém 3.952 genes, incluindo um profago e 42 elementos de inserção (*insertion sequence* -IS) (GARNIER et al., 2003).

O genoma de *M. bovis* é 99,95% geneticamente idêntico ao de *M. tuberculosis*, mostrando colinearidade e nenhuma evidência de grandes translocações, duplicações ou inversões. As diferenças são observadas na variação das redes de expressão gênica. O maior grau de variação entre as sequências dos bacilos bovino e humano é encontrado em genes que codificam proteínas da parede celular e proteínas secretoras (GARNIER et al., 2003). Apesar da similaridade genética com *M. tuberculosis*, o genoma de *M. bovis* apresenta-se reduzido devido a deleções, eventos considerados únicos e unidirecionais nessas bactérias. Essas deleções no DNA de *M. bovis* são as principais responsáveis pela diferenciação e afetam genes envolvidos no transporte, nas estruturas de superfície celular e no metabolismo intermediário; podendo remover genes que são dispensáveis para a adaptação em um determinado hospedeiro e conduzir para uma gama de hospedeiros diferentes. Mutações pontuais, também desempenham papel na definição do fenótipo, como é o caso da resistência a pirazinamida de *M. bovis* ou da mutação presente no gene da piruvato quinase que torna *M. bovis* incapaz de utilizar glicerol como fonte de energia (KEATING et al., 2005).

1.3 EVOLUÇÃO E FILOGENIA

Estudos genéticos recentes propõem uma via evolutiva para micobactérias do CMT, baseada em pequenas regiões do cromossomo e em sequências polimórficas de genes selecionados. Por muitos anos acreditava-se que *M. bovis* poderia ser o progenitor de *M. tuberculosis*. Porém análises de regiões do genoma micobacteriano denominadas regiões de diferenciação (RD) indicam que *M. bovis* não seja o progenitor de *M. tuberculosis*, uma vez que foi encontrada uma ampla linhagem de *M. tuberculosis* que antecedia as linhagens de *M. bovis* (GORDON et al. 2009). Os primeiros estudos filogenéticos das micobactérias foram realizados inicialmente com base em similaridades fenotípicas (BERGEY et al., 1923, citado por TORTOLI, 2011). O gene 16S RNA foi o primeiro, e durante muitos anos, o único alvo de sequenciamento genético em bactérias. Sua função essencial na síntese de proteínas e o fato de estar presente em todos os organismos desde o início da evolução contribuíram para que os estudos de sequenciamento se voltassem para este gene (PEIX et al., 2009).

Em 1963, Schuitemaker desenvolveu uma técnica de diferenciação de isolados do CMT pela exposição a uma seleção predefinida de vírus que podiam lisar bactérias (bacteriófagos). Apesar de apresentar baixo poder discriminatório, esta técnica revelou-se útil para diferenciar surtos específicos (SCHUITEMAKER, 1968 citado por SCHÜRCH; van SOOLINGEN, 2012). Outro método para diferenciação fenotípica de cepas por muito tempo utilizado foi o perfil de susceptibilidade a drogas anti-tuberculose que em combinação a características bioquímicas permite uma limitada subdivisão de isolados (COLLINS; YATES; GRANGE, 1982).

Um enorme salto qualitativo ocorreu a partir da década de 1960, com o progresso do conhecimento genético e mais recentemente com a criação de bancos de dados de domínio público que armazenam, em tempo real, sequências genéticas em toda parte do globo (TORTOLI, 2011). Nas décadas de 1980 e 1990, sob a influência dos avanços na biologia molecular, os métodos de tipagem molecular baseados no DNA foram desenvolvidos com melhor poder discriminatório em relação aos métodos

fenotípicos. No entanto, um elevado poder de discriminação não é a única característica de um sistema de tipagem ideal, características como: padronização, reprodutibilidade, tempo de execução e custo também devem ser consideradas (SCHÜRCH; van SOOLINGEN, 2012).

Após a conclusão do sequenciamento do genoma completo da cepa H37Rv de *M. tuberculosis* (COLE et al., 1998), foi possível realizar comparações *in silico* dos genomas, e a partir daí identificar diversos marcadores genômicos que pudessem contribuir com a análise evolutiva e com estudos epidemiológicos acerca dos membros do CMT. Estudos de genômica comparativa entre estirpes de *M. tuberculosis* e *M. bovis* identificaram mais de 140 genes cuja presença é facultativa e que estão relacionados com diferenças em nível de fenótipo, hospedeiro preferencial e virulência. Muitos desses genes ocorrem em RDs que não estão presentes em determinadas espécies (PYM et al., 2002).

Foi com base no estudo de deleções sucessivas de determinadas RD, que Brosch et al. (2002) traçaram um novo cenário evolutivo para o CMT. Segundo esses autores, o ancestral comum ao *M. bovis* e ao *M. caprae* teria perdido ao longo da evolução as regiões RD9, numa fase seguinte RD7, RD8 e RD10 e por fim RD12 e RD13 (figura 1). Ao contrário das cepas de amostras clínicas, as cepas vacinais de *M. bovis* bacilo Calmette-Guérin (BCG) apresentam a deleção da RD1, sugerindo um papel importante desta região na virulência das cepas patogênicas do CMT (PYM et al., 2002).

Baseado em 24 loci de unidades repetitivas de interespaçadores micobacterianos (MIRU), Wirth et al. (2008) propuseram que o CMT é composto de duas linhagens principais, sendo a linhagem 1 composta exclusivamente de isolados *M. tuberculosis stricto sensu* (todos de pacientes humanos) exceto o grupo do Leste Africano-Indiana (EAI) e a linhagem 2 composta principalmente de isolados animais (*M. microti*, *M. bovis*, *M. caprae* e *M. pinipedii*) e isolados do Leste Africano (*M. africanum* West African 1 e 2) (WIRTH et al., 2008). Essas duas linhagens são clones do progenitor *M. prototuberculosis* que emergiu aproximadamente há 40.000 anos. Este

progenitor acompanhou a migração dos humanos modernos para fora da África e sofreu alterações que levaram a sua diferenciação.

Acredita-se que as duas principais linhagens surgiram mais tarde cerca de 20.000 a 30.000 anos atrás. A linhagem 1 se espalhou exclusivamente entre os seres humanos, com ondas subsequentes de migração para Ásia, Europa e África continental. A linhagem 2 surgiu a partir da população *M. tuberculosis* e é a provável fonte de tuberculose animal (BROSCH et al., 2002; GARNIER et al., 2003). Assim, é provável que os seres humanos infectaram seus animais e não o contrário. A transição de humanos para os hospedeiros animais pode, assim, estar ligada à domesticação de animais, que ocorreu cerca de 13.000 anos atrás. (WIRTH; MEYER; ACHTMAN, 2005; WIRTH et al., 2008).

1.4 REGIÕES DE DIFERENCIAÇÃO E COMPLEXOS CLONAIIS

Em 2002, um trabalho com base em deleções sucessivas desenvolvido por Brosch et al. revelou 20 regiões de diferença no genoma das espécies do CMT. Catorze delas denominadas RD1-RD14 apresentaram uma variação de tamanho de 2 a 12,7 kb e se mostraram ausentes no BCG Pasteur quando comparadas com a cepa H37Rv de *M. tuberculosis*. Seis delas denominadas RvD1-RvD5 e TbD1 se mostraram ausentes na cepa H37Rv quando comparadas com outros membros do CMT. O ensaio foi realizado pela técnica da reação em cadeia da polimerase (PCR) utilizando 100 cepas e contou com amostras representativas de *M. africanum*, *M. microti*, *M. bovis*, *M. canneti*, o bacilo Calmette-guérin (BCG) e *M. tuberculosis*. Três tipos de deleções entre as 20 regiões reveladas foram determinadas e apresentaram sua importância como marcadores evolutivos (BROSCH et al., 2002). O primeiro marcador evolutivo, inclui elementos genéticos móveis como profagos phiRv1 (RD3) e phiRv2 (RD11) e sequências de inserção IS1532 (RD6) e IS6110 (RD5), cuja distribuição no bacilo da tuberculose foi altamente divergente.

O segundo tipo de deleção é mediado por recombinação homóloga entre elementos de inserção IS6110 adjacentes resultando na perda do segmento de DNA interveniente (RvD2, RvD3, RvD4 e RvD5), sendo variável de cepa para cepa. O terceiro e último tipo inclui exclusões cuja fronteira com as regiões genômicas normalmente não apresentam sequências repetitivas. Muitas vezes essas deleções ocorrem em regiões codificadoras, resultando no truncamento de genes que ainda estão intactos em outros membros do complexo. As regiões RD1, RD2, RD4, RD7, RD8, RD9, RD10, RD12, RD14 e TbD1 são representantes deste terceiro grupo (BROSCH et al., 2002).

Os resultados encontrados para *M. tuberculosis* mostraram que as RD citadas no terceiro grupo são altamente conservadas, e que essas RD podem ser utilizadas para diferenciar o bacilo humano independente da sua origem geográfica além de sugerir que essas regiões podem estar envolvidas na especificidade do seu hospedeiro. Os resultados apresentados para *M. bovis* indicaram uma variabilidade das regiões RD suprimidas em função dos diferentes hospedeiros que são susceptíveis ao bacilo bovino. Com exceção dos profagos, a cepa clássica de *M. bovis* ficou caracterizada pela deleção de oito RD (RD4, RD5, RD7, RD8, RD9, RD10, RD12 e RD13). Em contrapartida, as regiões RvDs se mostraram altamente conservadas e o TbD1 se mostrou presente em todos os isolados de *M. bovis* (BROSCH et al., 2002).

Mycobacterium bovis fica classificado como o último membro de uma linhagem separada representada por *M. africanum* (RD9) e *M. microti* (RD7, RD8, RD9 e RD10) e teria se ramificado do progenitor *M. tuberculosis*, representante exclusivo da RD9. As deleções observadas características de cada espécie podem ter contribuído para a expansão clonal e o aparecimento de patógenos melhores sucedidos em novos hospedeiros. Os resultados levaram a proposta de um novo cenário evolutivo para os membros do complexo e identificou *M. tuberculosis* como a espécie mais estreitamente relacionada com o ancestral comum do bacilo da tuberculose (BROSCH et al., 2002).

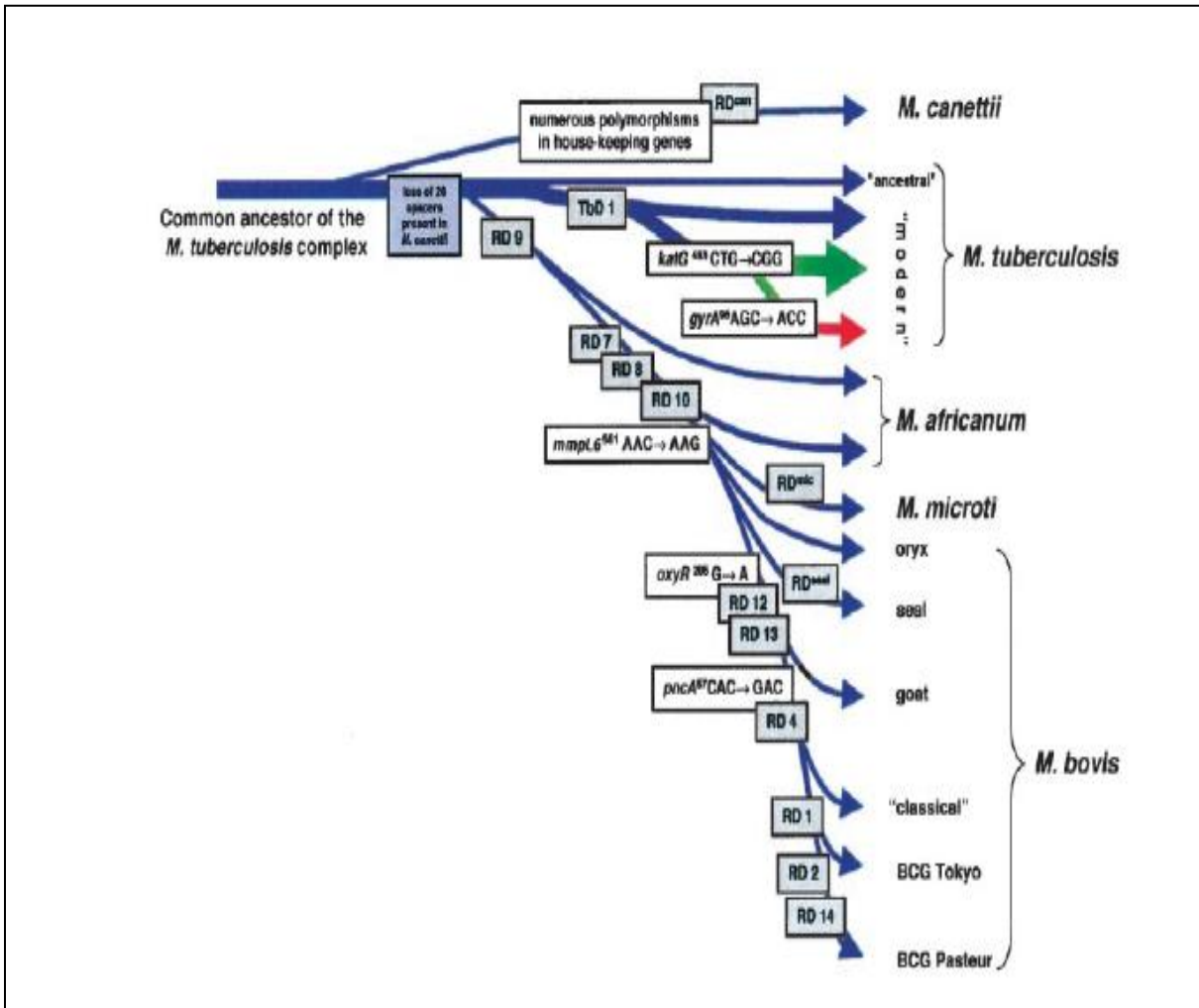


Figura 1. Filogenia e Complexos Clonais do Complexo *Mycobacterium tuberculosis*. Via evolutiva proposta para as micobactérias baseada em regiões de diferenciação. Fonte: Brosch et al. 2012.

Embora a tuberculose bovina tenha sido relatada em todos os continentes onde há bovinocultura, ficou assumido que este patógeno se originou em um único lugar, de uma só vez, e desde então tem sido distribuído em todo o mundo. Complexos clonais são definidos como grupos de cepas descendentes de uma única célula que foi o ancestral comum mais recente deste complexo clonal, e que apresenta todas as características derivadas deste ancestral comum (SMITH et al., 2011). Smith et al. (2003) descreveram a população de *M. bovis* como sendo uma população de estrutura

irregular e não uniforme, que consiste em uma série de grandes e pequenos complexos clonais.

Para bactérias que não apresentam a sua diversidade limitada pela reprodução sexuada, a diversidade genética pode ser decorrente de eventos como a recombinação genética, translocações, duplicações ou inversões no DNA. Entretanto, foi observado que nos membros do CMT não há troca de material genético entre linhagens (SMITH et al., 2011). As baixas taxas de homoplasia (quando duas linhagens desenvolvem independentemente a mesma característica) e a congruência entre os diferentes marcadores moleculares encontrados a partir de reconstruções filogenéticas dos membros do CMT foram observadas como uma consequência de clonalidade (COMAS; GAGNEUX, 2009).

Alguns complexos clonais de *M. bovis* foram definidos pela presença de deleções específicas, cada uma com o seu padrão específico de espoligotipo. Dois estão geograficamente localizados na África (RDAf1 e RDAf2), um terceiro apresenta distribuição global (RDEu1) e um quarto localizado principalmente na Península Ibérica (RDEu2). Em todos os estudos realizados, o *spoligotyping* foi realizado inicialmente como um *scanner* para identificar os membros de cada complexo clonal.

O complexo clonal Africa 1 (Af1) é dominante em Camarões, Nigéria, Mali e Chade e foram identificados pela deleção cromossômica específica de RDAf1 e pela ausência do espaçador 30 no *spoligotyping* (MULLER et al., 2009). Já o complexo clonal África 2 (Af2) foi definido pela deleção cromossômica RDAf2 e pela associação com a ausência dos espaçadores 3-7 no padrão de *spoligotyping*. O Af2 foi observado com alta frequência em isolados de *M. bovis* em Uganda, Burundi, Tanzânia e Etiópia. Um dado interessante observado foi que os membros do Af2 de *M. bovis* apresentaram em sua maioria quatro ou mais cópias da IS6110, em contraste com a maioria das outras cepas que possuem apenas uma ou poucas cópias. A deleção RDAf2 se mostrou intacta nas cepas do complexo clonal Af1, assim como a deleção de RDAf1 se mostrou intacta para as linhagens Af2, permitindo concluir que as cepas dos complexos clonais Af1 e Af2 são mutuamente exclusivas e não compartilham qualquer história

filogenética com o ancestral comum mais recente de cada complexo clonal, sugerindo que a mistura entre as populações de bovinos nesses países é pouco frequente (BERG et al., 2011 e SMITH et al., 2011).

O terceiro complexo clonal, identificado por Smith et al. (2011), apresenta distribuição global e é nomeado Europa 1 (Eu1). Este complexo é marcado pela exclusão da região cromossômica RDEu1. Cepas do complexo clonal Eu1 podem ser identificadas pela perda do espaçador 11 no padrão *spoligotyping*, embora esta característica não seja necessariamente específica para este complexo clonal. Dados mostram que cepas desse complexo são altamente observadas em isolados da República da Irlanda, Reino Unido, Estados Unidos, África do Sul, Nova Zelândia, Austrália, Canadá, Argentina, Chile, Equador, México, Coreia e do Cazaquistão e pouco observados no Brasil, França, Espanha, Portugal, Irã e Taiwan além de ser raro em outros países da comunidade europeia. A explicação mais simples para a distribuição desse complexo clonal é que ele foi disseminado por bovinos infectados do Reino Unido a bovinos de antigos parceiros comerciais (SMITH et al., 2011).

Um quarto complexo clonal Europa 2 (Eu2) também foi identificado. Ao contrário dos outros já mencionados, este complexo foi definido pela ausência do espaçador 21 associado a um SNP no gene *guaA*. Os isolados de *M. bovis* membros do Eu2 são provenientes da Península Ibérica onde são encontrados com alta frequência, enquanto na França, Itália e Ilhas Britânicas eles existem, mas em uma frequência muito baixa. A mutação no gene *guaA* não foi observada nas cepas de referência dos três complexos clonais descritos anteriormente (Af1, Af2 e Eu1), sugerindo que o Eu2 identifica um quarto complexo clonal de importância global (RODRIGUEZ-CAMPOS et al., 2001).

Em resumo, o CMT é um grupo altamente clonal de cepas (SMITH et al., 2006) e a diversidade genética entre os seus membros pode ser explicada em parte por exclusões decorrentes de eventos genéticos (GORDON et al., 1999). Essas exclusões denominadas polimorfismos de sequências longas têm sido exploradas com o objetivo de criar um novo cenário evolutivo para este complexo, baseado em regiões de diferenciação (BROSCH et al., 2002). Os SNPs também têm sido utilizados, porém

com uma aplicação mais direcionada para estudos de filogenia (GUTIERREZ et al., 2005 e SMITH et al., 2006). Análises acerca dos complexos clonais de *M. bovis* representam atualmente, uma nova ferramenta para a interpretação e exploração em massa de dados gerados constantemente na rotina de epidemiologistas moleculares, e ainda permite criar hipóteses para investigar a distribuição demográfica e a patogenicidade deste importante patógeno veterinário de importância global (SMITH et al., 2006).

1.5 VIRULÊNCIA

O que torna *M. bovis* virulento? Infelizmente, esta não é uma pergunta fácil de se responder. A definição de fatores de virulência, no geral, depende do tipo de abordagem utilizada. Do ponto de vista do patógeno a virulência pode ser determinada pelo número de organismos requeridos para estabelecer uma infecção, taxa de crescimento bacteriano, disseminação, persistência crônica e transmissibilidade. Do ponto de vista do hospedeiro a virulência pode ser verificada pela suscetibilidade a infecção, patologia tecidual, severidade dos sintomas da doença e frequência de mortalidade (COSMA et al., 2003). Em uma abordagem direta, Collins (2001) definiu para *M. bovis* fatores de virulência como propriedades especiais que permitem a micobactéria infectar, sobreviver, multiplicar e causar doença em um hospedeiro animal.

Entretanto, segundo Falkow (1998) a demonstração formal de que um gene específico é um gene de virulência exige o cumprimento dos Postulados Moleculares de Koch, que podem ser resumidos em três etapas. Primeiro, é necessário demonstrar que o referido gene codifica um fator de virulência ou controla a sua produção ou ainda promove a sua exportação ativa em cepas virulentas. Segundo, o organismo deve se tornar avirulento quando esse gene é interrompido. E terceiro, a complementação do gene inativado com um gene ativo deve restaurar a virulência.

Não é fácil satisfazer completamente todos estes requisitos, em alguns casos e, na prática, um padrão um pouco mais flexível de evidências é muitas vezes apresentada. E há ainda um outro fator a ser considerado: o período de tempo necessário para os estudos de virulência em *M. bovis*, uma vez que as micobactérias apresentam um tempo de crescimento muito maior do que o tempo requerido para bactérias patogênicas de crescimento rápido, como por exemplo, *Salmonella sp* (COLLINS, 2001).

Devido à elevada homologia entre os genomas de *M. bovis* e *M. tuberculosis* (COLE et al., 1998; GARNIER et al., 2003) esperava-se que os genes de virulência fossem em grande parte idênticos em ambas as espécies. Na prática, *M. bovis* causa doença em uma ampla gama de mamíferos, enquanto que a infecção por *M. tuberculosis* encontra-se quase que exclusivamente em seres humanos. Assim esta variação de especificidade do hospedeiro é contabilizada em grande parte por razões epidemiológicas. Entretanto há a hipótese de que parte da diferença genética entre estas duas espécies micobacterianas estaria envolvida nas propriedades de virulência de *M. bovis* e seria então considerada como genes de virulência específicos do hospedeiro (COLLINS & GICQUEL, 2000).

Até 1990, houve muitas sugestões de possíveis fatores de virulência para *M. bovis* e *M. tuberculosis*. Entretanto, as técnicas moleculares para investigar estas alegações não estavam disponíveis. Com o desenvolvimento há 25 anos de vetores que poderiam transportar genes entre *Escherichia coli* e micobactérias (JACOBS & BLOOM, 1994) tornando possível a busca por genes de virulência. Desde os primeiros protótipos e com o advento dos métodos biológicos moleculares tornou-se possível a construção dos primeiros mutantes de *M. bovis* o que facilitou os estudos de virulência (CALMETTE & GUERRIN, 1911, BERTHET et al., 1998; WILSON et al., 1996). A possibilidade de desativar genes específicos em *M. bovis* e *M. tuberculosis* utilizando micobacteriófagos recombinantes sensíveis à temperatura ou plasmídeos suicidas (PARISH & STOKER, 2000) também contribuiu no avanço do conhecimento de

virulência e assim uma gama completa de técnicas exigidas para experimentos de genes de virulência está disponível.

Os modelos utilizados para o estudo de virulência em *M. bovis* e *M. tuberculosis* são a cultura de tecidos, utilizando principalmente macrófagos, células dendríticas (BODNAR et al., 2001; HICKMAN et al., 2002) e pneumócitos (BERMUDEZ & GOODMAN, 1996) e modelos animais. Embora os modelos de cultura de tecidos sejam mais fáceis de trabalhar e deem resultados mais rápidos, eles são limitados para as fases iniciais de infecção. Assim, os modelos animais são melhores pois todas as fases da tuberculose podem ser estudadas. A escolha de modelos animais para estudos de virulência é importante, e os três principais modelos utilizados são camundongos, cobaias e coelhos (SMITH, 2003).

Estudos genômicos foram fundamentais para identificação e caracterização das regiões de diferenciação dos genomas micobacterianos. A RD1 é uma região do genoma micobacteriano responsável por codificar fatores intrínsecos à virulência. Análises comparativas do genoma de *M. tuberculosis* H37Rv, *M. tuberculosis* cepas H37Ra virulenta e avirulenta e *M. bovis* atenuada (BCG), possibilitaram constatar a importância desta região para a manutenção da virulência ao longo da evolução. Esta região é responsável por codificar as proteínas ESAT-6 e CFP10 (do inglês, 10 kDa Culture Filtrate Protein), consideradas de alta imunogenicidade (GANGULY et al. 2008).

Assim a combinação dos métodos elencados acima permitiram pesquisas que identificaram vários genes importantes em diversos aspectos de patogenicidade de *M. bovis* e conseqüentemente envolvidos na virulência desta micobactéria. A maioria desses genes de virulência codificam enzimas de diversas vias lipídicas, proteínas de superfície celular, reguladores e proteínas dos sistemas de sinais de transdução. Outro grupo de relevância é aquele envolvido na sobrevivência das micobactérias no interior do microambiente agressivo dos macrófagos do hospedeiro (FORRELLAD et al., 2013). Um outro ponto importante em relação a virulência de *M. bovis* é que *M.*

bovis, assim como *M. tuberculosis*, não apresenta clássicos fatores de virulência como aqueles que são as principais causas de doenças por outros agentes patogênicos bacterianos, como por exemplo, as toxinas produzidas por *Corynebacterium diphtheriae*, *Escherichia coli* O157: H7, *Shigella dysenteriae* and *Vibrio cholerae* (SMITH, 2003). Além disso, muitos dos genes de virulência das espécies do CMT são também conservados em micobactérias não-patogênicas. Estes resultados sugerem que espécies patogênicas adaptaram seus genomas de um estilo de vida livre para o ambiente intracelular com aquisição mínima de genes exclusivos de virulência (FORRELLAD et al., 2013).

Entender mecanismos de regulação da expressão gênica também é crucial para determinar a manutenção da virulência micobacteriana. A expressão gênica na bactéria é regulada primariamente ao nível da iniciação transcricional conferida por holoenzimas RNA polimerase (RNAP). Esta holoenzima é formada por cinco subunidades fixas e uma subunidade dissociável denominada subunidade sigma (σ). O fator σ é responsável por muitos dos determinantes da especificidade promotora da RNAP (SACHDEVA et al., 2009).

Para *M. tuberculosis* foi descrito que a expressão diferencial e temporal de reguladores controlados pela indução de um ou mais fatores σ permitem a este microrganismo manter-se por múltiplos estágios de interação com o hospedeiro, incluindo adesão, invasão, replicação intracelular e disseminação para outros órgãos (SMITH, 2003). *Mycobacterium tuberculosis* possui 13 fatores σ , sendo o fator sigma A (SigA ou σ^A) o mais importante, uma vez que este fator é indispensável ao crescimento bacteriano. Por meio da utilização de mutantes verificou-se que bacilos deficientes para o fator σ^A apresentavam atenuação da virulência, sendo o fator σ^A incapaz de interagir com o fator de transcrição acessório WhiB3. A inativação do fator WhiB3 leva à perda da virulência em modelos animais sem afetar o crescimento bacteriano. (WU et al., 2004).

Diante da existência de poucos trabalhos ou mesmo da não abrangência de todos os fatores de virulência e de resultados previamente obtidos, a compreensão destes fatores irá levar a avaliação do impacto da virulência e da patogenia da tuberculose bovina ao desenvolvimento de novas estratégias, bem como uma possível vacina efetiva para controle desta enfermidade.

1.6 DIAGNÓSTICO

O diagnóstico da tuberculose bovina pode ser realizado por meio dos métodos clínicos, alérgicos, bacteriológicos, sorológicos e/ou anatomopatológicos e moleculares.

O método diagnóstico *in vivo* preconizado para bovinos e prescrito para o comércio internacional é o teste alérgico de tuberculinização intradérmico (BRASIL, 2006). O teste consiste em uma prova cutânea indireta realizada pela inoculação intradérmica do PPD, um extrato de antígenos de *M. bovis*, com detecção do espessamento cutâneo local causado por uma reação de hipersensibilidade tardia após 72 horas da inoculação (TSCHOPP et al., 2011).

No Brasil os testes alérgicos de tuberculinização intradérmico utilizando tuberculina para o diagnóstico de tuberculose bovina podem ser de três tipos. O teste da prega caudal (TPC) utiliza tuberculina bovina e é utilizado como prova de triagem exclusivamente em gado de corte; o teste cervical simples (TCS) também utiliza tuberculina bovina e é a prova de rotina em gado de leite devido à sua boa sensibilidade; e o teste cervical comparativo (TCC) utiliza tuberculina bovina e aviária e é a prova confirmatória para animais reagentes ao TPC ou ao TCS. Todavia, o TCC também pode ser empregado como única prova diagnóstica em rebanhos com histórico de reações inespecíficas (BRASIL, 2006).

No exame *post mortem*, estudos bacteriológicos e histopatológicos das lesões macroscópicas encontradas em abatedouros-frigoríficos permitem controlar o diagnóstico realizado a campo e avaliar a eficácia das provas tuberculínicas (BRASIL, 2006). O exame histopatológico pode aumentar a confiança do diagnóstico, mas apenas o isolamento bacteriológico de *M. bovis* fornece um diagnóstico definitivo. A

bacteriologia clássica é considerada padrão-ouro no diagnóstico de *M. bovis*. Entretanto, as micobactérias patogênicas crescem lentamente (cerca de 90 dias, ou mais) sendo que 10% das amostras de *M. bovis* somente apresentam colônias visíveis após 60 dias de cultivo. A qualidade do meio de cultura, da descontaminação da amostra e das condições de incubação aliados ao fato de algumas amostras conterem apenas um pequeno número de bacilos vivos também podem comprometer o sucesso do isolamento. Além disso, a identificação da espécie por meio de provas bioquímicas é trabalhosa, requer cultura nova, pura e abundante, condição alcançada somente após sucessivos repiques a partir do primo isolamento (ZANINI et al., 2001; HADDAD et al., 2004).

O diagnóstico precoce é essencial para o controle da doença. Métodos de biologia molecular estão sendo desenvolvidos para detectar o agente diretamente em amostras clínicas e avaliar a variação genética dentro de espécies de micobactérias (BRASIL, 2006). A tuberculose bovina é um exemplo de doença infecciosa em que as técnicas de biologia molecular permitiram a obtenção de informações que seriam difíceis ou até mesmo impossíveis de serem obtidas através de métodos laboratoriais convencionais. A diferenciação das espécies do CMT depende de características fenotípicas, porém devido à elevada homologia do genoma, a identificação precisa de certas espécies se torna difícil e trabalhosa (SUPPLY et al., 2001).

1.7 EPIDEMIOLOGIA MOLECULAR

O aprimoramento da engenharia genética tem fornecido recursos baseados no DNA com grande potencial para elucidar os problemas relacionados à identificação das micobactérias. Os sistemas de tipagem molecular possuem a capacidade de distinguir entre isolados em decorrência da variação genética no DNA cromossômico de uma espécie bacteriana. Comumente, os métodos de genotipagem molecular recorrem à amplificação por PCR de regiões do genoma suficientemente estáveis que servem então como marcadores epidemiológicos, e que também apresentam potencial

polimórfico para diferenciação de cepas sem relação epidemiológica entre si (DUARTE et al., 2007).

A maioria das técnicas de genotipagem utilizadas para diferenciação de cepas do CMT são classificadas em duas categorias: técnicas baseadas no genoma completo e técnicas baseadas em sequências específicas do genoma (DURR; HEWINSON; CLIFTON-HADLEY, 2000). As técnicas que envolvem o genoma completo foram as primeiras técnicas descritas e possuem a vantagem da obtenção de todas as potenciais informações genéticas. No entanto, esses métodos (incluindo a análise de restrição de endonucleases e a eletroforese em campo pulsado) são tecnicamente exigentes e de difícil automação, portanto, não utilizadas frequentemente nas rotinas laboratoriais. As técnicas baseadas em sequências do genoma incluem técnicas de sequências repetidas específicas encontradas ao longo do genoma, que compreendem as sequências de inserção, as regiões de repetições diretas e os loci de repetições em sequência (RODRÍGUEZ-CAMPOS, 2012).

As técnicas mais utilizadas para tipagem molecular de membros do CMT são *Restriction Fragment Length Polymorphism* — RFLP-IS6110, *Spoligotyping* e *Mycobacterial Interspersed Repetitive Units* — MIRU-Variable Number Tandem Repeats — VNTR (HADDAD; MASSELOT; DURAND, 2004, DUARTE et al., 2007, COLLINS, 2011).

1.7.1 RFLP-IS6110

A análise de Polimorfismos dos Tamanhos de Fragmentos de Restrição (RFLP) é uma técnica que tem sido bastante utilizada para o CMT, sendo considerada padrão-ouro para diferenciação de cepas de *M. tuberculosis* (VAN SOOLINGEN et al., 1994). Este método utiliza endonucleases de restrição, enzimas que reconhecem sequências específicas de DNA, cortando-o em fragmentos de restrição específicos pela hibridação com sondas de DNA marcadas. Dependendo do grau de mobilidade e número de cópias destas sequências, são obtidos perfis de DNA (“DNA fingerprints”) que podem ser utilizadas para inferir relação entre cepas e traçar relação

epidemiológica (KAMERBEEK et al., 1997, COLLINS et al., 2011). A sonda tradicionalmente utilizada para diferenciar cepas de *M. tuberculosis* tem sido a IS6110, fragmento este que se apresenta com até 20 cópias no genoma desta micobactéria, entretanto o mesmo não ocorre para *M. bovis*, uma vez que este apresenta apenas uma ou duas cópias de IS6110, sendo o máximo de oito cópias encontradas em isolados desta espécie (NIEMANN et al., 1999, COLLINS et al., 2011).

1.7.2 *Spoligotyping*

A tipagem pelo *spoligotyping* baseia-se na amplificação por PCR do DNA presente no *locus* da região de repetição direta (DR), presente exclusivamente no genoma de micobactérias do CMT (KAMERBEEK et al., 1997). O *locus* DR, que caracteriza a técnica do *spoligotyping*, é composto de pequenas sequências repetidas de 36 pb que se intercalam com pequenas sequências espaçadoras variadas, que podem conter de 34 a 41 pb. A ordem dos espaçadores é aproximadamente a mesma em todos os isolados, a diferença está na ocorrência de inserções e deleções de espaçadores, de modo que o polimorfismo de vários isolados pode ser comprovado com base na presença ou ausência de um ou mais espaçadores (ARANAZ et al., 1996). O número de DRs varia entre as espécies do CMT. *Mycobacterium bovis* caracteriza-se pela ausência das sequências espaçadoras 3, 9, 16 e 39 a 43 (KAMERBEEK et al., 1997).

A técnica de *spoligotyping* é uma técnica de hibridação de linha reversa em *blot* (HADDAD et al., 2004) e envolve duas etapas. A primeira consiste na amplificação, por PCR, dos espaçadores presentes entre as regiões DR. Somente um par de iniciadores, *a* e *b*, específico às duas extremidades das sequências DR, é necessário para amplificar todos os espaçadores presentes. O iniciador *a* está ligado a uma enzima, a fim de permitir a detecção dos produtos amplificados. A segunda etapa tem como objetivo revelar quais espaçadores foram amplificados nos isolados que estão sendo analisados e é realizada através da hibridação dos espaçadores a uma membrana usando um *miniblotter*. Nesta membrana, oligonucleotídeos

correspondentes aos diferentes espaçadores estão covalentemente ligados. Os produtos amplificados são depositados e a presença de um espaçador é revelada por um ponto de inserção dos oligonucleotídeos às linhas de amplificação utilizando um método químico. (ARANAZ et al., 1996; HADDAD et al., 2004).

O *spoligotyping* fornece um sistema de tipificação que pode ser utilizado em um grande número de isolados e os resultados podem ser apresentados de uma forma binária padronizada, o que foi uma grande vantagem sobre as outras técnicas, uma vez que permitiu pela primeira vez a comparação de resultados entre laboratórios (SMITH, 2011). A criação de bancos de dados com perfis genotípicos dos membros do CMT denominados espoligotipos tem facilitado a comparação dos perfis genotípicos encontrados nas diferentes regiões do mundo, permitindo uma melhor compreensão da dinâmica da doença (FILLIOL et al., 2002).

1.7.3 MIRU-VNTR

Muitos loci possuem hipervariabilidade em seus números de repetições e são denominados de loci de repetições sucessivas de número variável (VNTR) (NAKAMURA et al., 1988). Os VNTR são encontrados nas regiões intergênicas e não intergênicas no genoma, interferindo nos processos de regulação da célula (transcrição e tradução) e nas trocas moleculares dos microrganismos (EPPLEN et al., 1997).

O primeiro locus VNTR identificado no genoma de *M. tuberculosis* consistiu de uma repetição em tandem de 75 pb localizada no gene *katG* (GOYAL et al., 1994). Outros loci foram encontrados dispersos no genoma de micobactérias, sendo que, Supply et al. (1997) denominaram estes loci de unidades micobacterianas repetitivas (MIRU), que variam em tamanho de 46-101 pb (RODRÍGUEZ-CAMPOS, 2012). Estudos posteriores de tipagem utilizando VNTR de isolados do CMT reconheceram diversas outras regiões, tais como: as repetições exatas em tandem (ETR) A a F, variando de tamanho de 53 e 79 pb (FROTHINGHAM; MEEKER-O'CONNELL, 1998), *SenX3-RegX3* (MAGDALENA et al., 1998), QUB11a, QUB11b, QUB18, QUB23, QUB26 (SKUCE et al., 2002). Segundo Price-Carter et al. (2011) o número de

marcadores utilizados pela técnica VNTR atualmente dispõe de 45 *loci VNTR*; sendo que o polimorfismo nos loci pode ocorrer tanto como mudança na sequência dos nucleotídeos, como na variação no número de unidades repetidas, conseqüentemente ambas criando uma variação alélica (van BELKUM et al., 1998).

Para revelar o número de repetições em sequência, a PCR é realizada com oligonucleotídeos iniciadores destinados a amplificar individualmente os loci de repetição em sequência selecionados (COWAN et al., 2002). Primeiramente são realizadas PCRs utilizando diferentes pares de iniciadores (um para cada MIRU-VNTR), para a determinação do número de repetições em cada locus. Após a amplificação de um dado locus, é determinado o tamanho (peso molecular em pares de bases) do amplificado por eletroforese em gel de agarose, em comparação com marcadores de tamanho molecular para deduzir o número de repetição em sequência presentes em cada locus. Para a interpretação dos resultados é utilizada uma tabela de alelos que relaciona o peso molecular obtido com o número de alelos repetidos em cada locus. O resultado final da tipagem MIRU-VNTR é um código numérico, que corresponde ao número de repetições do locus alvo e serve como uma impressão digital do respectivo isolado (RODRÍGUEZ-CAMPOS, 2012). Uma vez determinado o perfil de alelos para cada cepa, os dados são analisados empregando um aplicativo *web MIRU-VNTRplus* (<http://www.miru-vntrplus.org>).

1.8 SEQUENCIAMENTO GENÔMICO

Nos últimos anos, um progresso notável ocorreu na compreensão dos mecanismos de virulência e na patogênese de infecções micobacterianas particularmente pela aplicação do sequenciamento genômico e análise genômica comparativa. A ciência genômica tornou possível a elucidação do modelo genético completo de várias espécies de micobactérias de importância em medicina humana e veterinária, bem como espécies ambientais (OLSEN; BARLETTA; THOEN, 2010).

O sequenciamento genômico teve início em meados da década de 70, primeiramente com a elaboração da metodologia descrita por Frederick Sanger e

colaboradores em 1975, a qual ficou conhecida por método de Sanger, método do dideoxi ou terminação de cadeia (SANGER; COULSON, 1975). De forma independente, Maxam e Gilbert desenvolveram, em 1976, a técnica de sequenciamento químico, a qual foi preterida devido à maior dificuldade técnica para automatização (MAXAM; GILBERT, 1977). A metodologia de Sanger é fundamentada no princípio da terminação controlada da polimerização pela utilização de dideoxinucleotídeos marcados (modificados), que impedem a extensão de um fragmento de DNA em replicação pela DNA polimerase após sua incorporação, possibilitando assim determinar a ordem sequencial dos nucleotídeos que compõem fragmentos da molécula do DNA (SANGER; COULSON, 1975).

O sequenciamento do genoma humano pela técnica de Sanger demonstrou claramente a necessidade do desenvolvimento de novas metodologias capazes de gerar uma quantidade imensa de dados a baixo custo (VENTER et al., 2001). A partir de então a geração de dados genômicos tem crescido expressivamente devido ao desenvolvimento de novos processos de sequenciamento de DNA. O sequenciamento de nova geração – *Next Generation Sequencing*, ou simplesmente NGS – são processos de sequenciamento de DNA que utilizam metodologias diferentes da de Sanger, com o objetivo de acelerar e baixar o custo do processo de sequenciamento. Apesar de se diferenciarem consideravelmente entre si todos os sequenciadores de NGS se baseiam no processamento paralelo massivo de fragmentos de DNA.

Os sequenciadores de nova geração começaram a ser comercializados a partir de 2005, com o lançamento do equipamento 454 FLX pela *Roche*. A plataforma 454 realiza o sequenciamento baseado em síntese, o pirosequenciamento (RONAGHI et al., 1998). A leitura da sequência nesse sistema é realizada a partir de uma combinação de reações enzimáticas que se inicia com a liberação de um pirofosfato, oriundo da adição de um desoxinucleotídeo à cadeia. Em seguida, esse pirofosfato é convertido em ATP, pela ATP sulfúrilase, sendo este utilizado pela luciferase para oxidar a luciferina, produzindo um sinal de luz capturado por uma câmera CCD (*charge-coupled device*) acoplada ao sistema.

A partir de então, novas técnicas de sequenciamento evoluíram rapidamente. Em 2007, a empresa *Illumina* lançou a plataforma *Solexa GA*, seguido pela plataforma *SOLiD (Applied Biosystems)* e *Heliscope (Helicos)*. Essas tecnologias promovem o sequenciamento de milhares de pares de bases em uma única corrida (CARVALHO; SILVA, 2010). O sequenciamento realizado pela plataforma *Illumina* difere da estratégia utilizada no equipamento 454 e *SOLiD* especialmente por realizar a amplificação dos fragmentos *in situ*, ou seja, na própria superfície de clonagem (*flow cell*), também conhecida de PCR de fase sólida (ANDERSON; SCHRIJVER, 2010). Entretanto, assim como o sequenciamento de Sanger, o sequenciamento é realizado pela ação da enzima DNA polimerase capaz de incorporar nucleotídeos marcados por diferentes fluoróforos (ANSORGE, 2009).

Na *flow cell* adaptadores são fixados à superfície pela extremidade 5', deixando a extremidade 3' livre para servir na iniciação da reação de sequenciamento dos fragmentos imobilizados. Uma vez que os fragmentos de DNA da amostra também são ligados a adaptadores em ambas as extremidades durante o preparo da biblioteca genômica, é possível fixá-la a *flow cell* por hibridação a um dos adaptadores fixados.

No primeiro ciclo de amplificação, nucleotídeos não marcados são fornecidos para que ocorra a síntese da segunda fita do fragmento imobilizado no suporte. A alta densidade de adaptadores no suporte facilita a hibridação do adaptador livre dos fragmentos imobilizados à sua sequência complementar fixa perto do clone inicial durante o ciclo de anelamento. Após o ciclo de anelamento, o fragmento forma uma estrutura em forma de “ponte” na superfície de sequenciamento, e a extensão ocorre, formando a fita complementar também em “ponte”. No ciclo de desnaturação, as fitas são separadas e linearizadas. Esses ciclos são repetidos 35 vezes, e assim, obtém-se um *cluster* de sequenciamento com milhares de cópias geradas dos fragmentos (SHENDURE; JI, 2008).

Após os ciclos de desnaturação, nucleotídeos terminadores marcados são fornecidos para as reações de sequenciamento que ocorrem dentro de cada *cluster*. A alta densidade dos *clusters* de sequenciamento permite que o sinal de fluorescência, gerado com a incorporação de cada um dos nucleotídeos terminadores, tenha uma

intensidade suficiente para garantir sua detecção exata. Assim a leitura do sinal de fluorescência é realizada. A leitura das bases é feita pela análise sequencial das imagens capturadas em cada ciclo de sequenciamento (ANDERSON; SCHRIJVER, 2010).

No sistema *SOLiD* (do inglês, *Sequencing by Oligo Ligation and Detection*), lançado em 2007, diferentemente dos demais processos, a reação de sequenciamento é catalisada por uma DNA ligase, e não uma polimerase. Essa plataforma utiliza a incorporação de dinucleotídeos marcados por meio da DNA ligase, seguido da excitação do fluoróforo (sinal emitido é capturado por sensores) e a incorporação dos dinucleotídeos seguintes. Essa leitura gera um código de cores que é analisado por ferramenta de bioinformática e convertida a sequência de letras. Cada corrida na plataforma *SOLiD* leva aproximadamente cinco dias e produz de 3 a 4 Gb de sequências com o comprimento variando de 50 a 75 pb (MARDIS, 2008).

Em 2011, a *Applied Biosystems* lançou a plataforma *Ion Torrent*. Esta plataforma utiliza mecanismo semelhante ao 454 da *Roche*, no entanto, ao invés de usar o pirofosfato, utiliza o íon de hidrogênio que é liberado ao ser incorporado um novo nucleotídeo. Essa plataforma gera leituras de tamanho médio de até 200 pb, e apresenta ainda como principal limitação a significativa taxa de erro (1,5%) em homopolímeros (LOMAN et al., 2012).

Essas plataformas de nova geração vêm sendo amplamente empregadas no sequenciamento genômico de várias espécies bacterianas, incluindo *Mycobacterium tuberculosis* (COLE et al. 1998) e *Mycobacterium bovis* (GARNIER et al., 2003).

1.9 OBJETIVOS

1.9.1 Objetivo geral

Sequenciamento genômico de isolados de *Mycobacterium bovis* e realização de análises de virulência e filogenética.

1.9.2 Objetivos específicos

Sequenciar, montar e depositar genomas de isolados de *M. bovis*;

Caracterizar geneticamente por *Spoligotyping* e MIRU-VNTR isolados de *M. bovis* anteriormente sequenciados;

Realizar análises filogenéticas dos isolados de *M. bovis* sequenciados estabelecendo relações entre as mesmas;

Realizar análise de genes de virulência em isolados de *M. bovis* previamente selecionados;

REFERÊNCIAS

ALEXANDER, K. A. et al. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerging Infectious Diseases*, Atlanta, GA, v. 16, n. 8, p. 1.296-1.299, 2010.

ANDERSON, M. W.; SCHRIJVER, I. Next generation DNA sequencing and the future of genomic medicine. *Genes – Open Access Genetics & Genomics Journal*, Basel, v. 1, n. 1, p. 38-69, 2010.

ANSORGE, W. J. Next-generation DNA sequencing techniques. *New Biotechnology*, Amsterdam, v. 25, n. 4, p. 195-203, 2009.

ARANAZ, A. et al. Use of polymerase chain reaction in the diagnosis of tuberculosis in cats and dogs. *The Veterinary Record*, London, v. 138, n. 12, p. 276-280, 1996.

ARAÚJO, C. P. et al. Detection of *Mycobacterium bovis* in bovine and bubaline tissues using nested-PCR for TbD1. *PloS One*, San Francisco, CA, v. 11, n. 3, p. e91023, 2014.

BARRERA, L. The basis of clinical bacteriology. In: PALOMINO, J. C.; LEÃO, S. C.; RITACCO, V. (Ed.). *Tuberculosis 2007: From basic science to patient care*. p. 93-112.

BERG, S. et al. African 2, a clonal complex of *Mycobacterium bovis* epidemiologically important in East Africa. *Journal of Bacteriology*, Washington, DC, v. 193, n. 3, p. 670-678, 2011.

BERMUDEZ, L. E.; GOODMAN, J. *Mycobacterium tuberculosis* invades and replicates within type II alveolar cells. *Infection and Immunity*, v. 64, p. 1400-1406, 1996.

BERTHET, F. X. et al. Attenuation of virulence by disruption of the *Mycobacterium tuberculosis* *erp* gene. *Science*, v. 282, p. 759-762, 1998.

BIET, F. et al. Zoonotic aspects of *Mycobacterium bovis* and *Mycobacterium avium-intracellulare* complex (MAC). *Veterinary Research*, Paris, v. 36, n. 3, p. 411-436, 2005.

BODNAR, K. A., N. V. et al. Fate of *Mycobacterium tuberculosis* within murine dendritic cells. *Infection and Immunity*, v. 69, p. 800–809, 2001.

BRASIL. Ministério da Agricultura e do Abastecimento. Gabinete do Ministro. Instrução Normativa n. 2, de 10 de janeiro de 2001. Instituir o Programa Nacional de Controle e Erradicação da Brucelose e da Tuberculose Animal e atribuir ao Secretário de Defesa Agropecuária a incumbência de baixar o Regulamento Técnico do Programa e expedir as instruções necessárias à plena implementação das atividades de combate às supracitadas doenças no País. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 11 jan. 2001a.

BRASIL. Ministério da Agricultura e do Abastecimento. Secretário de Defesa Agropecuária. Instrução Normativa n. 2, de 10 de janeiro de 2001. Aprovar o Regulamento Técnico do Programa Nacional de Controle e Erradicação da Brucelose e Tuberculose Animal e delegar competência ao Diretor do Departamento de Defesa Animal para baixar instruções de serviços complementares a este Regulamento. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 16 jan. 2001. Republicada em 4 jun. 2001b.

BRASIL. Ministério de Agricultura, Pecuária e Abastecimento. *Programa Nacional de Controle e Erradicação da Brucelose e a Tuberculose Animal (PNCEBT)*. Brasília, DF: MAPA/SDA/DSA, 2006.

BROSCH, R. et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, DC, v. 99, n. 6, p. 3.684-3.689, 2002.

CALMETTE, A.; GUEHRIN, C. Recherches experimentales sur la defense de l'organisme contre l'infection tuberculose. *Annales Institute Pasteur*, v. 245, p. 625-647, 1911.

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. *Ciência Rural*, Santa Maria, v. 40, n. 3, p. 735-744, 2010.

- COLE, S. T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, London, v. 393, n. 6.685, p. 537-544, 1998.
- COLLINS, C. H.; YATES, M. D.; GRANGE, J. M. Subdivision of *Mycobacterium tuberculosis* into five variants for epidemiological purposes: Methods and nomenclature. *The Journal of Hygiene*, London, v. 89, n. 2, p. 235-242, 1982.
- COLLINS, D. M.; GICQUEL, B. Genetics of Mycobacterial Virulence. In: Hatfull G F, Jacobs Jr W R, eds. *Molecular Genetics of Mycobacteria*. Washington DC, *American Society for Microbiology Press*, p. 265-278, 2000.
- COLLINS, D. M. Virulence factors of *Mycobacterium bovis*. *Tuberculosis (Edinb)*, v. 81(1-2), p. 97-102, Review, 2001.
- COLLINS, D. Advances in molecular diagnostics for *Mycobacterium bovis*. *Veterinary Microbiology*, Amsterdam, v. 151, n. 1-2, p. 2-7, 2011.
- COMAS, I.; GAGNEUX, S. The past and future of tuberculosis research. *PLoS Pathogens*, San Francisco, CA, v. 5, n. 10, p. 1-7, 2009.
- CORNER, L. A. Post mortem diagnosis of *Mycobacterium bovis* infection in cattle. *Veterinary Microbiology*, Amsterdam, v. 40, n. 1-2, p. 53-63, 1994.
- CORNER, L. A. The role of wild animal populations in the epidemiology of tuberculosis in domestic animals: How to assess the risk. *Veterinary Microbiology*, Amsterdam, v. 112, n. 2-4, p. 303-312, 2006.
- COSMA, C.L. et al. The secret lives of the pathogenic mycobacteria. *Annual Review of Microbiology*, v. 57, p. 641-76, Review, 2003.
- COWAN, L. et al. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS6110 by using mycobacterial interspersed repetitive units. *Journal of Clinical Microbiology*, Washington, DC, v. 40, n. 5, p. 1.592-1.602, 2002.
- DUARTE, E. L. et al. Transmissão da tuberculose bovina entre espécies domésticas e silvestres em Portugal: primeiras evidências moleculares em isolados de *Mycobacterium bovis* de uma exploração no Alentejo. *Revista Portuguesa de Ciência Veterinárias*, Lisboa, v. 102, n. 563-564, p. 299-303, 2007.
- DURR, P.; HEWINSON, R. G.; CLIFTON-HADLEY, R. S. Molecular epidemiology of bovine tuberculosis – I. *Mycobacterium bovis* genotyping. *Revue Scientifique et Technique / Office International des Épizooties*, Paris, v. 19, n. 3, p. 675-688, 2000.

- EPPLEN, C. et al. On simple repetitive DNA sequences and complex diseases. *Electrophoresis*, Weinheim, v. 18, n. 9, p. 1.577-1.585, 1997.
- FABRE, M. et al. High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of *hsp65* gene polymorphism in a large collection of “*Mycobacterium canettii*” strains indicates that the *M. tuberculosis* complex is a recently emerged clone of “*M. canettii*”. *Journal of Clinical Microbiology*, Washington, DC, v. 42, n. 7, p. 3.248-3.255, 2004.
- FALKOW, S. Molecular Koch’s postulates applied to microbial pathogenicity. *Review of Infection Disease*, v. 10, p. S274-S276, 1998.
- FILLIOL, I. et al. Spoligotyping and IS6110-RFLP typing of *Mycobacterium tuberculosis* from French Guiana: A comparison of results with international databases underlines interregional transmission from neighboring countries. *Research in Microbiology*, v. 153, n. 2, p. 81-88, 2002.
- FORRELLAD, M. A. et al. Virulence factors of the *Mycobacterium* complex. *Virulence*, v. 4, n. 1, p. 3-66, 2013.
- FOXMAN, B.; RILEY, L. Molecular epidemiology: focus on infection. *American Journal of Epidemiology*, Baltimore, v. 153, n. 12, p. 1.135-1.141, 2001.
- FROTHINGHAM, R.; MEEKER-O’CONNELL, W. A. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology (Reading, England)*, Reading, v. 144, n. 5, p. 1189-1196, 1998.
- GANGULY, N., SIDDIQUI, I. & SHARMA, P. Role of *M. tuberculosis* RD-1 region encoded secretory proteins in protective response and virulence. *Tuberculosis (Edinb)*. v. 88, n. 6, p. 510-517, 2008.
- GARNIER, T. et al. The complete genome sequence of *Mycobacterium bovis*. *Proceedings of the National Academy of Sciences*, Washington, DC, v. 100, n. 13, p. 7877-7882, 2003.
- GORDON, S. V. et al. Pathogenicity in the tubercle bacillus: molecular and evolutionary determinants. *BioEssays: News and reviews in molecular, cellular and developmental biology*, Cambridge, UK, v. 31, n. 4, p. 378-388, 2009.
- GORDON, S. V.; BROSCHE, R.; BILLAULT, A. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Molecular Microbiology*, Boston, v. 32, n. 3, p. 643-655, 1999.
- GOYAL, M. et al. PCR amplification of variable sequence upstream of *katG* gene to subdivide strains of *Mycobacterium tuberculosis* complex. *Journal of Clinical Microbiology*, Washington, DC, v. 32, n. 12, p. 3.070-3.071, 1994.

GUTIERREZ, M. C. et al. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathogens*, San Francisco, CA, v. 1, n. 1, p. 55-61, 2005.

GUTIÉRREZ, M.; TELLECHEA, J.; GARCÍA MARÍN, J. F. Evaluation of cellular and serological diagnostic tests for the detection of *Mycobacterium bovis*-infected goats. *Veterinary Microbiology*, Amsterdam, v. 62, n. 4, p. 281-290, 1998.

HADDAD, N.; MASSELOT, M.; DURAND, B. Molecular differentiation of *Mycobacterium bovis* isolates: Review of main techniques and applications. *Research in Veterinary Science*, Oxford, UK, v. 76, n. 1, p. 1-18, 2004.

HICKMAN, S. P. et al. *Mycobacterium tuberculosis* induces differential cytokine production from dendritic cells and macrophages with divergent effects on naive T cell polarization. *Journal of Immunology*, v. 168, p. 4636-4642, 2002.

JACOBS, J. R. W. R.; BLOOM, B. R. Molecular genetic strategies for identifying virulence determinants of *Mycobacterium tuberculosis*. In: Bloom B R, ed. Tuberculosis: pathogenesis, protection and control. Washington DC, *American Society for Microbiology Press*, p. 253-268, 1994.

JORGE, K. S. G. et al. Tuberculose bovina: epidemiologia e controle. In: ALMEIDA, R. F. C.; SOARES, C. O.; ARAÚJO, F. R. (Ed.). *Brucelose e tuberculose bovina: epidemiologia, controle e diagnóstico*. Brasília, DF: Embrapa Informação Tecnológica, 2004. p. 45-59. KAMERBEEK, J. et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology*, Washington, DC, v. 35, n. 4, p. 907-914, 1997.

KANEENE, J. B.; MILLER, R.; MEYER, R. M. Abattoir surveillance: the U.S. experience. *Veterinary Microbiology*, Amsterdam, v. 112, n. 2-4, p. 273-282, 2006.

KEATING, L. A. et al. The pyruvate requirement of some members of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase: Implications for in vivo growth. *Molecular Microbiology*, Boston, v. 56, n. 1, p. 163-174, 2005.

LOMAN, N. J. et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, New York, v. 30, n. 5, p. 434-439, 2012.

MAGDALENA, J. et al. Identification of a new DNA region specific for members of *Mycobacterium tuberculosis* complex. *Journal of Clinical Microbiology*, Washington,

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, v. 24, n. 3, p. 133-141, 2008.

MAXAM, A. M.; GILBERT, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, DC, v. 74, n. 2, p. 560-564, 1977.

MICHEL, A.; MÜLLER, B.; VAN HELDEN, P. D. *Mycobacterium bovis* at the animal-human interface: A problem, or not? *Veterinary Microbiology*, Amsterdam, v. 140, n. 3-4, p. 371-381, 2010.

MOKADDAS, E.; AHMAD, S. Development and evaluation of a multiplex PCR for rapid detection and differentiation of *Mycobacterium tuberculosis* complex members from non-tuberculous mycobacteria. *Japanese Journal of Infectious Diseases*, Tokyo, v. 60, n. 2-3, p. 140-144, 2007.

MONAGHAN, M. L. et al. The tuberculin test. *Veterinary Microbiology*, Amsterdam, v. 40, n. 1-2, p. 111-124, 1994.

MÜLLER, B. et al. African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. *Journal of Bacteriology*, Washington, DC, v. 191, n. 6, p. 1.951-1.960, 2009.

NAKAMURA, Y. et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*, New York, v. 235, n. 4.796, p. 1.616-1.622, 1988.

NIEMANN, S.; RICHTER, E.; RÜSCH-GERDES, S. Stability of *Mycobacterium tuberculosis* IS6110 restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. *Journal of Clinical Microbiology*, Washington, DC, v. 37, n. 11, p. 409-412, 1999.

OLSEN, I.; BARLETTA, R. G.; THOEN, C. O. Mycobacterium. In: GYLES, C. L. et al. (Ed.). *Pathogenesis of bacterial infections in animals*. 4th. ed. Ames, Iowa: Blackwell Publishing, 2010. p. 113-132.

PARISH, T.; STOKER, N. G. Use of an exible cassette method to generate a double unmarked *Mycobacterium tuberculosis* *tlyA*. *Microbiology*, v. 146, p. 1969-1975, 2000.

PEIX, A.; RAMÍREZ-BAHENA, M.H., VELÁSQUEZ, E. Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID)*, Amsterdam, v. 9, n. 6, p. 1.132-1.147, 2009.

POLLOCK, J. M. et al. Pathogenesis of bovine tuberculosis: The role of experimental models of infection. *Veterinary Microbiology*, Amsterdam, v. 112, n. 2-4, p. 141-150, 2006.

PRICE-CARTER, M.; ROOKER, S.; COLLINS, D. M. Comparison of 45 variable number tandem repeat (VNTR) and two direct repeat (DR) assays to restriction

- endonuclease analysis for typing isolates of *Mycobacterium bovis*. *Veterinary Microbiology*, Amsterdam, v. 150, n. 1-2, p. 107-114, 2011.
- PYM, A. S. et al. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Molecular Microbiology*, Boston, v. 46, n. 3, p. 709-717, 2002.
- RADUNZ, B. Surveillance and risk management during the latter stages of eradication: Experiences from Australia. *Veterinary Microbiology*, Amsterdam, v. 112, n. 2-4, p. 283-290, 2006.
- RODRÍGUEZ, J. G. et al. Species-specific identification of *Mycobacterium bovis* by PCR. *Microbiology (Reading, England)*, Reading, v. 141, n. 9, p. 2.131-2.138, 1995.
- RODRÍGUEZ-CAMPOS, S. *Epidemiología molecular de Mycobacterium bovis y Mycobacterium caprae en España*. 2012. 356 f. Tesis (Doctoral Sanidad Animal) – Universidad Complutense de Madrid, Madrid, 2012.
- RODRIGUEZ-CAMPOS, S. et al. European 2 – A clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID)*, Amsterdam, v. 12, n. 4, p. 866-872, 2012.
- RONAGHI, M.; UHLÉN, M.; NYRÉN, P. A sequencing method based on real-time pyrophosphate. *Science*, New York, v. 281, n. 5.375, p. 363-365, 1998.
- ROXO, E. Tuberculose bovina: revisão. *Arquivos do Instituto Biológico*, São Paulo, v. 63, n. 2, p. 91-97, 1996.
- SACHDEVA, P. et al. The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS Journal*, v. 277, n. 3, p. 605-26, Review, 2009.
- SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, London, v. 94, n.3, p. 441-448, 1975.
- SCHÜRCH, A. C.; VAN SOOLINGEN, D. DNA fingerprinting of *Mycobacterium tuberculosis*: From phage typing to whole-genome sequencing. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID)*, Amsterdam, v. 12, n. 4, p. 602-609, 2012.
- SHENDURE, J.; JI, H. Next-generation DNA sequencing. *Nature Biotechnology*, New York, v. 26, n. 10, p. 1.135-1.145, 2008.

SKUCE, R. A. et al. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology (Reading, England)*, Reading, Microbiology-Sgm, v. 148, n. 2, p. 519-528, 2002.

SMITH I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical of Microbiology Review*, v. 16, n. 3, p. 463-96, Review, 2003.

SMITH, N. H. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID)*, Amsterdam, v. 12, n. 4, p. 857-865, 2012.

SMITH, N. H. et al. Bottlenecks and broomsticks: The molecular evolution of *Mycobacterium bovis*. *Nature Reviews Microbiology*, London, v. 4, n. 9, p. 670-681, 2006.

SMITH, N. H. et al. European 1: a globally important clonal complex of *Mycobacterium bovis*. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID)*, Amsterdam, v. 11, n. 6, p. 1.340-1.351, 2011.

SMITH, N. H. et al. The population structure of *Mycobacterium bovis* in Great Britain: Clonal expansion. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, DC, v. 100, n. 25, p. 15.271-15.275, 2003.

SUPPLY, P. et al. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *Journal of Clinical Microbiology*, Washington, DC, v. 39, n. 10, p. 3.563-3.571, 2001.

SUPPLY, P. et al. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Molecular Microbiology*, Boston, v. 26, n. 5, p. 991-1.003, 1997.

TORTOLI, E. Phylogeny of the genus *Mycobacterium*: many doubts, few certainties. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID)*, Amsterdam, v. 12, n. 4, p. 827-831, 2011.

TSCHOPP, R. et al. *Bovine tuberculosis* at a cattle-small ruminant-human interface in Meskan, Gurage region, Central Ethiopia. *BMC Infectious Diseases*, London, v. 11, n. 318, p. 1-14, 2011.

VAN BELKUM, A. et al. Short-sequence DNA repeats in prokaryotic genomes. *Microbiology and Molecular Biology Reviews: MMBR*, v. 62, n. 2, p. 275-293, 1998.

VAN INGEN, J. et al. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerging Infectious Diseases*, Atlanta, GA, v. 18, n. 4, p. 653-655, 2012.

VAN SOOLINGEN, D. et al. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. *Journal of Clinical Microbiology*, Washington, DC, v. 32, n. 10, p. 2.425-2.433, 1994.

VENTER, J. C. et al. The sequence of the human genome. *Science*, New York, v. 291, n. 5.507, p. 1.304-1.351, 2001.

WILSON, T. M.; COLLINS, D. M. *ahpC*, a gene involved in isoniazid resistance of the *Mycobacterium tuberculosis* complex. *Molecular Microbiology*, v. 19, p. 1025-1034, 1996.

WIRTH, T. et al. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathogens*, San Francisco, CA, v. 4, n. 9, p. 1-10, 2008.

WIRTH, T.; MEYER, A.; ACHTMAN, M. Deciphering host migrations and origins by means of their microbes. *Molecular Ecology*, Oxford, UK v. 14, n. 11, p. 3.289-3.306, 2005.

WORLD ORGANISATION FOR ANIMAL HEALTH. Bovine tuberculosis. In: _____. *Manual of diagnostic tests and vaccines for terrestrial animals*. 5. ed. Paris: OIE, 2009. v. 2. chap. 2.4.7. p. 1-16. Version adopted by the World Assembly of Delegates of the OIE in May 2009.

WU, S. et al. The principal sigma factor sigA mediates enhanced growth of *Mycobacterium tuberculosis* in vivo. *Molecular Microbiology*, v. 51, n. 6, p. 1551-1562, 2004.

ZANINI, M. S. et al. *Mycobacterium bovis*: Polymerase chain reaction identification in bovine lymphonode biopsies and genotyping in isolates from Southeast Brazil by spoligotyping and restriction fragment length polymorphism. *Memórias do Instituto Oswaldo Cruz*, Rio de Janeiro, v. 96, n. 6, p. 809-813, 2001.

ARTIGO 1

SEQUENCIAMENTO GENÔMICO DE *MYCOBACTERIUM BOVIS* CEPA AN5, UTILIZADA PARA A PRODUÇÃO DE DERIVADO PROTÉICO PURIFICADO

Publicado: Genome Announcements

Março/Abril 2014, vol 2, issue 2, e00277-14

Sequenciamento genômico de *Mycobacterium bovis* cepa AN5, utilizada para a produção de Derivado Protéico Purificado

Ana Beatriz Canevari Castelão^a, Christiane Nishibe^b, André Moura^c, Andrea Padilha de Alencar^c, Mariana de Azevedo Issa^c, Mikael Arrais Hodon^c, Pedro Moacyr Pinto Coelho Mota^c, Erica Bravo Sales^c, Antônio Augusto Fonseca Júnior^c, Nalvo Franco Almeida^b, Flávio Ribeiro de Araújo^d

Programa de Pós-Graduação em Ciência Animal, UFMS, Campo Grande, Mato Grosso do Sul, Brazil^a; Escola de Computação, UFMS, Campo Grande, Mato Grosso do Sul, Brazil^b; Laboratório Nacional Agropecuário–LANAGRO/MG, Ministério da Agricultura, Pecuária e Abastecimento, Pedro Leopoldo, Minas Gerais, Brazil^c; Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brasil^d

A cepa *Mycobacterium bovis* AN5 tem sido utilizada na produção de derivado proteico purificado (PPD) para o teste intradérmico da tuberculose bovina desde que foi introduzido em 1948. Este trabalho reporta o sequenciamento genômico de *M. bovis* cepa AN5, que é utilizado para a produção de PPD bovino no Brasil, bem como comparações com outras cepas de *M. bovis* e *M. tuberculosis*.

Mycobacterium bovis é o agente causador da tuberculose bovina, uma doença que ocasiona perdas anuais de U\$3 bilhões de dólares (1) e que representa uma ameaça para a saúde pública e bem estar animal (2,3). No Brasil e em muitos outros países, o programa de controle da tuberculose bovina envolve o teste intradérmico em bovinos com um antígeno preparado a partir da cepa *M. bovis* AN5, denominado de derivado protéico purificado (PPD), seguido do sacrifício dos animais reagentes. Essa cepa produz um elevado rendimento de massa celular em meio glicerinado, um fenótipo desejável que foi selecionado a partir de repetidas culturas do bacilo em meio laboratorial (4). O sequenciamento genômico da cepa *M. bovis* AN5 utilizada no Brasil na produção do PPD é relatada neste artigo. Comparações de genes que codificam proteínas que constituem o PPD em outras cepas de *M. bovis* (AF2122/97 e 04-303) e *Mycobacterium tuberculosis* H37Rv também são descritas.

A sequência do genoma foi obtida utilizando a tecnologia MiSeq (5), produzindo um total de 3.240.633 *paired-end reads*. Após filtragem dos dados, 2.269.762 *paired-end reads* e 372.399 *single-end reads* foram utilizados na montagem. Realizamos a montagem do genoma assistida por um genoma referência: *M. bovis* AF2122/97 (GenBank número de acesso NC_002945) utilizando os dados filtrados de *M. bovis* cepa AN5 utilizando o programa *Bowtie*. A montagem contém 70 *contigs* (de tamanhos maiores de 500 pb) e N50 150.219. A anotação do genoma da cepa *M. bovis* AN5 foi obtida através da utilização do NCBI Prokariotic Genome Annotation Pipeline (7) e compreende 3.938 sequências codificadoras de proteínas (CDs) e 48 genes codificadores de RNA (3 rRNAs e 45 tRNAs). Fora das 3.938 CDs, 2.830 (72%) possuem uma atribuição funcional.

Uma comparação dos genes que codificam proteínas encontradas no PPD bovino produzido no Reino Unido (PPDUK) e no PPD produzido no Brasil (PPDBR) por espectrometria de massa em cromatografia líquida revelou um total de 116 proteínas. Destas proteínas, 67 foram encontradas somente no PPDUK, 12 foram encontradas somente no PPDBR e 37 foram identificados em ambos PPDs (8). Este fato sugere que o isolado brasileiro *M. bovis* AN5 pode ter perdido alguns genes que codificavam proteínas do PPD ou que estes genes apresentam um elevado grau de variação. Os genes que codificam para todas as 116 proteínas encontrado em ambos os PPDs foram comparados com seus ortólogos em *M. bovis* AF2122/97 (1), *M. bovis* 04-303 (9), e *M. tuberculosis* H37Rv (10) e apresentaram um elevado grau de conservação, com identidades variando de 98 a 100%. Além disso não foram detectadas *frameshifts*. Os genes que codificam as proteínas imunodominantes MPB64, MPB70, MPB83, GroEL, GroES, HspX, CFP-10 e CFP-21 são 100% conservadas entre os ortólogos. O gene que codifica para ESAT-6 mostrou 100% de identidade com seus ortólogos em *M. tuberculosis* H37Rv e *M. bovis* AF2122/97 e 99% de identidade com o ortólogo em *M. bovis* 04-303. O gene que codifica para MPB53 mostrou 100% de identidade com ortólogos em *M. bovis* 04-303 e *M. bovis* AF2122/97 e 99% de identidade com *M. tuberculosis* H37Rv. Em conclusão, nenhuma deleção ou variação genética no isolado

M. bovis AN5 utilizada no PPDBR apoia a diferença proteômica encontrada quando comparado-o com o PPDUK.

Número de acesso da sequência nucleotídica: Este projeto de sequenciamento do genoma completo foi depositado no DDBJ/EMBL/GenBank acesso no. AWPL00000000. A versão descrita neste trabalho é a versão AWPL01000000.

AGRADECIMENTOS

Este trabalho foi financiado pelo CNPq (concessão 305503/2010-3 e 479394/ 2011-3), Fundect (concessão TO 0096/12 and 23/200.152/2009), e Embrapa (concessão 03.11.22.003.00.00).

REFERÊNCIAS

1. **Garnier T, Eiglmeier K, Camus JC, 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 PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG.** 2003. The complete genome sequence of *Mycobacterium bovis*. Proc. Natl. Acad. Sci. U. S. A. **100**: 7877–7882. <http://dx.doi.org/10.1073/pnas.1130426100>.
2. **Gutiérrez M, Tellechea J, García Marín JF.** 1998. Evaluation of cellular and serological diagnostic tests for the detection of *Mycobacterium bovis* infected goats. Vet. Microbiol. **62**:281–290. [http://dx.doi.org/10.1016/S0378-1135\(98\)00217-X](http://dx.doi.org/10.1016/S0378-1135(98)00217-X).
3. **Michel AL, Cooper D, Jooste J, de Klerk LM, Jolles A.** 2011. Approaches towards optimising the gamma interferon assay for diagnosing *Mycobacterium bovis* infection in African buffalo (*Syncerus caffer*). Prev. Med. Vet. **98**:142–151. <http://dx.doi.org/10.1016/j.prevetmed.2010.10.016>.

4. **Paterson AB.** 1948. The production of bovine tuberculoprotein. *J. Comp. Pathol. Ther.* **58**:302–313. [http://dx.doi.org/10.1016/S0368-1742\(48\)80027-5](http://dx.doi.org/10.1016/S0368-1742(48)80027-5).
5. **Bennett S.** 2004. Solexa Ltd. *Pharmacogenomics* **5**:433–438. <http://dx.doi.org/10.1517/14622416.5.4.433>.
6. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:357–359. <http://dx.doi.org/10.1038/nmeth.1923>.
7. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O.** 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *Omics* **12**: 137–141. <http://dx.doi.org/10.1089/omi.2008.0017>.
8. **Borsuk S, Newcombe J, Mendum TA, Dellagostin OA, McFadden J.** 2009. Identification of proteins from tuberculin purified protein derivative (PPD) by LC-MS/MS. *Tuberculosis (Edinb)* **89**:423–430. <http://dx.doi.org/10.1016/j.tube.2009.07.003>.
9. **Nishibe C, Canevari Castelão AB, Dalla Costa R, Pinto BJ, Varuzza L, Cataldi AA, Bernardelli A, Bigi F, Blanco FC, Zumárraga MJ, Almeida NF, Araújo FR.** 2013. Draft genome sequence of *Mycobacterium bovis* 04-303, a highly virulent strain from Argentina. *Genome Announc.* **1**(6): e00931-13. <http://dx.doi.org/10.1128/genomeA.00931-13>.
10. **Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J,**

Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544. <http://dx.doi.org/10.1038/31159>.

ARTIGO 2

SEQUENCIAMENTO GENÔMICO DE *MYCOBACTERIUM BOVIS* 04-303, UM ISOLADO ALTAMENTE VIRULENTO DA ARGENTINA

Publicado: Genome Announcements

Novembro/Dezembro 2013, vol 1, issue 6, e00931-13

Sequenciamento genômico de *Mycobacterium bovis* 04-303, um isolado altamente virulento da Argentina

Christiane Nishibe^a, Ana Beatriz Canevari Castelão^b, Ricardo Dalla Costa^c, Beatriz Jeronimo Pinto^c, Leonardo Varuzza^c, Angel Adrain Cataldi^d, Amelia Bernadelli^e, Fabiana Bigi^d, Federico Carlos Blanco^d, Martín José Zumárraga^d, Nalvo Franco Almeida^a, Flávio Ribeiro de Araújo^f

Escola de Computação, UFMS, Campo Grande, Mato Grosso do Sul, Brasil^a; Programa de Pós-Graduação em Ciência Animal, UFMS, Campo Grande, Mato Grosso do Sul, Brasil^b; Life Technologies do Brasil, São Paulo, São Paulo, Brasil^c; Instituto de Biotecnologia, CICyA, INTA Castelar, Buenos Aires, Argentina^d; Ceva Saúde Animal, Buenos Aires, Argentina^e, Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil^f

***Mycobacterium bovis* cepa 04-303 foi isolado de um javali selvagem vivendo em campo livre na Argentina. Este trabalho relata o sequenciamento genômico deste isolado altamente virulento e a comparação genômica de seus principais genes relacionados à virulência com aqueles de *M. bovis* cepa AF2122/97 e *Mycobacterium tuberculosis* cepa H37Rv.**

A tuberculose bovina (TB) é uma doença infecciosa crônica causada por *Mycobacterium bovis*, afetando bovinos (1), outras espécies de animais domésticos, animais selvagens e seres humanos, (2,3), ocasionando perdas agropecuárias mundias na ordem de 3 bilhões de dólares por ano (4). A TB também é uma zoonose, sendo uma problema de saúde pública, especialmente em países em desenvolvimento onde a interação homem-animal e a prevalência do HIV são elevadas (5). Os animais selvagens representam um sério entrave para o controle e a erradicação da TB e contribuem para a manutenção da infecção uma vez que atuam como hospedeiros reservatórios da doença (6).

Este trabalho apresenta o sequenciamento genômico da cepa *M. bovis* 04-303, isolado a partir de um javali em um campo livre em La Pampa, Argentina, em 2004. Em modelo murino este isolado induziu desenvolvimento de súbita pneumonia, áreas de

necrose massiva pulmonar, e morte após 6 semanas de infecção, com mais de duas unidades formadoras de colônias pulmonar do que camundongos infectados com o isolado controle *M. bovis* isolado AN5. Este isolado é caracterizado por apresentar o espoligotipo SB0140 e MIRU 232224263322 (7,8).

O sequenciamento genômico foi realizado utilizando as plataformas MiSeq (9) e Ion Torrent (10), produzindo 521.134 *paired-end reads* e 651.617 *reads* após a filtragem dos dados, totalizando 118 Mpb. O genoma foi montado tendo como referência o genoma *M. bovis* AF2122/97 (acesso. NC_002945) utilizando o programa *Bowtie* (11). Apresenta 169 *contigs* (o maior contendo 329.090 pb), com um N50 de 85.558 pb, conteúdo de CG de 63% e cobertura média de 27,3 x.

Anotação realizada com o NCBI POC (12) identificou um total de 3.988 sequências codificadoras de proteínas e 49 genes codificadores de RNA. Análises filogenética e comparativa do genoma de 04-303 a quatro outros genomas (*M. bovis* AF2122/97, *M. bovis* BCG, *Mycobacterium tuberculosis* H37Rv, e *Mycobacterium africanum* GM041182) apontou que *M. bovis* AF2122/97 é o isolado mais próximo de 04-303, compartilhando 99,4% dos seus genes codificadores de proteínas.

Comparações foram realizadas entre os isolados *M. bovis* 04-303, *M. bovis* AF2122/97, e *M. tuberculosis* H37Rv (13). Os genes ortólogos a *rv1759c* de H37Rv em 04-303 e AF2122/97 são pseudogenes. Este gene codifica uma proteína PE que apresenta sequências polimórficas rica em CG (PGRS) que se liga a fibronectina, e sabe-se que alterações nesta família de proteínas pode influenciar no hospedeiro preferencial ou no tropismo de tecidos. O gene *rpfA*, está associado ao retorno do período de dormência e da replicação do bacilo, apresentou uma deleção de 240 pb nos isolados 04-303 e AF2122/97, que conseqüentemente ocasiona à síntese de uma proteína menor quando comparada a proteína codificada pelo mesmo gene no isolado H37Rv. O gene *pks6*, envolvido com síntese de policetídeos, é interrompido em nos isolados 04-303 e AF2122/97. A inativação deste gene mostrou atenuar *M. tuberculosis* em modelo murino (14). O gene *mce3* está presente em H37Rv e ausente em 04-303 e AF2122/97. Curiosamente, isolados de *M. tuberculosis* que apresentam o

operon mce3 interrompido são atenuados em modelo murino (15). O gene *esxA* (ESAT6), que codifica um fator de virulência em *M. tuberculosis*, é conservado em AF2122/97, porém detectamos uma mutação única (A por G) em 04-303, resultando na substituição preditiva desfavorável de uma treonina por uma alanina (16).

Número de acesso da sequência nucleotídica: Este projeto de sequenciamento do genoma completo foi depositado na DDBJ/EMBL/GenBank acesso no. AVSW00000000. A versão descrita neste trabalho é a versão AVSW01000000.

AGRADECIMENTOS

Este trabalho foi financiado pelo CNPq (concessão 305503/2010-3 e 479394/ 2011-3), Fundect (concessão TO 0096/12 and 23/200.152/2009), e Embrapa (concessão 03.11.22.003.00.00).

REFERÊNCIAS

1. **Wobeser G.** 2009. Bovine tuberculosis in Canadian wildlife: an updated history. *Can. Vet. J.* **50**:1169–1176.
2. **Gutiérrez M, Tellechea J, García Marín JF.** 1998. Evaluation of cellular and serological diagnostic tests for the detection of *Mycobacterium bovis* infected goats. *Vet. Microbiol.* **62**:281–290.
3. **Michel AL, Cooper D, Jooste J, de Klerk LM, Jolles A.** 2011. Approaches towards optimising the gamma interferon assay for diagnosing *Mycobacterium bovis* infection in African buffalo (*Syncerus caffer*). *Prev. Vet. Med.* **98**:142–151. November/December 2013 Volume 1 Issue 6 e00931-13 Genome Announcements genomea.asm.org **1**
4. **Garnier T, Eiglmeier K, Camus JC, 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 PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG.** 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 7877–788.

5. **Joshi D, Harris NB, Waters R, Thacker T, Mathema B, Krieswirth B, Sreevatsan S.** 2012. Single nucleotide polymorphisms in the *Mycobacterium bovis* genome resolve phylogenetic relationships. *J. Clin. Microbiol.* **50**:3853–3861.
6. **Talbot EA, Williams DL, Frothingham R.** 1997. PCR identification of *Mycobacterium bovis* BCG. *J. Clin. Microbiol.* **35**:566–569.
7. **Zumárraga MJ, Martin C, Samper S, Alito A, Latini O, Bigi F, Roxo E, Cicuta ME, Errico F, Ramos MC, Cataldi A, van Soolingen D, Romano MI.** 1999. Usefulness of spoligotyping in molecular epidemiology of *Mycobacterium bovis*-related infections in South America. *J. Clin. Microbiol.* **37**:296–303.
8. **Aguilar León D, Zumárraga MJ, Jiménez Oropeza R, Gioffré AK, Bernardelli A, Orozco Estévez H, Cataldi AA, Hernández Pando R.** 2009. *Mycobacterium bovis* with different genotypes and from different hosts induce dissimilar immunopathological lesions in a mouse model of tuberculosis. *Clin. Exp. Immunol.* **157**:139–147.
9. **Bennett S.** 2004. Solexa Ltd. *Pharmacogenomics* **5**:433–438.
10. **Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J.** 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**:348–352.
11. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:357–359.
12. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, Kodira CD, Kyripides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O.** 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *Omics* **12**:137–141.
13. **Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG.** 1998.

Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.

14. **Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C.** 1999. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol. Microbiol.* **34**:257–267.

15. **Senaratne RH, Sidders B, Sequeira P, Saunders G, Dunphy K, Marjanovic O, Reader JR, Lima P, Chan S, Kendall S, McFadden J, Riley LW.** 2008. *Mycobacterium tuberculosis* strains disrupted in *mce3* and *mce4* operons are attenuated in mice. *J. Med. Microbiol.* **57**:164–170.

16. **Chen JM, Zhang M, Rybniker J, Boy-Röttger S, Dhar N, Pojer F, Cole ST.** 2013. *Mycobacterium tuberculosis* EspB binds phospholipids and mediates EsxA-independent virulence. *Mol. Microbiol.* **89**:1154–1166.

ARTIGO 3

COMPARAÇÃO DE GENES DE VIRULÊNCIA DE DUAS CEPAS DE *MYCOBACTERIUM BOVIS* COM PERFIS PATOGÊNICOS CONTRASTANTES

Artigo para submissão:

BMC Microbiology

**COMPARAÇÃO DE GENES DE VIRULÊNCIA DE DUAS CEPAS DE
MYCOBACTERIUM BOVIS COM PERFIS PATOGÊNICOS CONTRASTANTES**

AUTORES:

Ana Beatriz Canevari Castelão¹

Fabiana Bigi²

Martín José Zumárraga²

Angel Cataldi²

Carlos Alberto do Nascimento Ramos¹

Christiane Nishibe³

Augusto César de Aquino Ribas³

Nalvo Franco Almeida³

Flávio Ribeiro Araújo^{4*}

ENDEREÇO INSTITUCIONAL:

¹Faculdade de Medicina Veterinária e Zootecnia, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brasil.

²Instituto de Biotecnología, CICVyA/INTA, Los Reseros y las Cabañas, 1712 Castelar, Argentina

³Faculdade de Computação, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, CP 549, Brazil

⁴Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, 79106-550, Brazil

E-MAIL:

ABCC - anabia_85@yahoo.com.br

FB - bigi.fabiana@inta.gob.ar

MJZ - zumarraga.martin@inta.gob.ar

AC - cataldi.angeladrian@inta.gob.ar

CANR - carlosanramos@yahoo.com.br

CN - cnishibe@gmail.com

ACAR – ribas.aca@gmail.com

NFA - nalvo@facom.ufms.br

FRA - flabio.araujo@embrapa.br

Autor correspondente: FRA: flabio.araujo@embrapa.br

Abstract: *Mycobacterium bovis* é o agente causador da tuberculose bovina, uma doença infecciosa crônica que pode afetar gado, outras espécies domésticas, animais selvagens e seres humanos, com prejuízos econômicos expressivos em escala mundial. Duas cepas de *M. bovis* foram sequenciadas recentemente. Uma das cepas foi isolada a partir de um javali (04- 303) e a outra cepa foi isolada a partir de um bovino de leite (534), ambas na Argentina. Estudos anteriores demonstraram que o isolado 04-303 induziu 100% de mortalidade em camundongos 6 semanas após a infecção, ao passo que o isolado 534 não induziu alto percentual de mortalidade mesmo após 4 meses de infecção, com danos teciduais limitados. Objetivou-se neste trabalho comparar um repertório de genes de virulência presentes nesses genomas. Todos os genes de ambos genomas foram analisados, utilizando *M. bovis* AF2122/97 como genoma de referência. Mutações nucleotídicas foram encontradas em 30 genes de virulência no isolado 04-303: 21 mutações *non-synonymous missense* (troca de aminoácidos), 9 deleções (deleção de um ou mais nucleotídeos) e 1 inserção (inserção de um ou mais nucleotídeos). Em relação ao isolado 534, mutações nucleotídicas foram encontradas em 44 genes de virulência: 24 mutações *non-synonymous missense*, 20 deleções e 2 inserções. Dezoito destas mutações foram comuns em

ambos os genomas. As mutações encontradas nos genes de virulência estão em concordância com o fenótipo atenuado do isolado *M. bovis* 534, enquanto que as mutações encontradas nos genes de virulência estão apenas parcialmente em concordância com o fenótipo virulento do isolado 04-303. A comparação deste importante conjunto de genes relacionados à virulência nos levará a uma melhor compreensão dos aspectos de patogenicidade e virulência de *M. bovis*, essenciais para futuros estudos que visem o desenvolvimento de novas vacinas e técnicas de diagnóstico para tuberculose bovina.

Palavras chave: *Mycobacterium bovis*; genes de virulência; genoma.

Introdução

Mycobacterium bovis (*M. bovis*) é o agente causador da tuberculose bovina (TB), uma doença infecciosa crônica que está disseminada em todo o mundo e que infecta principalmente bovinos, embora possa afetar também outras espécies domésticas, animais selvagens e seres humanos (GUTIÉRREZ et al., 1998; WOBESER, 2009; MICHEL et al., 2011). *M. bovis* pertence ao Complexo *Mycobacterium tuberculosis* (CMT), tradicionalmente composto por *M. tuberculosis*, o principal agente da tuberculose humana, *Mycobacterium africanum*, *Mycobacterium microti*, *Mycobacterium canettii*, *M. bovis*, *Mycobacterium caprae*, *Mycobacterium pinnipedii*, e recentemente incorporado às espécies *Mycobacterium suricattae*, *Mycobacterium mungi* e *Mycobacterium orygis* (FABRE et al., 2004; SMITH et al., 2006; ALEXANDER et al., 2010, VANINGEN et al., 2012; PARSONS et al., 2013). Todas estas espécies são geneticamente muito semelhantes (FORRELLAD et al., 2013). Tem sido sugerido que os membros do CMT evoluíram a partir de um ancestral comum por meio de sucessivas deleções/inserções de DNA, resultando na presente especiação e em suas diferenças de patogenicidade (FABRE et al., 2004).

Apesar da elevada relação genética global, as espécies do CMT exibem uma grande variedade de fenótipos e gama de hospedeiros. A deleção da informação

genética, juntamente com a presença de polimorfismos de um único nucleotídeo (SNPs) ao longo do cromossoma das espécies do CMT, provavelmente explica a diferença na gama de hospedeiros, virulência e imunopatologia encontrados entre estas espécies.

A cepa *M. bovis* 04-303 foi isolada de um javali na província de La Pampa, Argentina em 2004. Esta cepa foi caracterizada com o espoligotipo SB0140 e padrão de MIRU 232224263322. A cepa *M. bovis* 534 foi isolada de um bovino de leite na província de Santa Fé, Argentina em 1997 e foi caracterizada com o espoligotipo SB0140, e padrão de MIRU 232224163322. Estas cepas foram utilizadas no modelo murino experimental BALB/c de tuberculose pulmonar progressiva, tal como descrito por Aguilar León et al. (2009). Neste estudo, os camundongos infectados com o isolado 04-303 apresentaram mortalidade elevada. Esses animais desenvolveram uma resposta inflamatória leve seguida de pneumonia súbita, com extensa necrose pulmonar. Camundongos infectados com o isolado 534 apresentaram taxa de sobrevivência elevada e dano tecidual limitado. Este isolado induziu expressão progressiva de INF- γ , bem como elevada expressão de IL-4, TNF- α e iNOS. Por outro lado, o isolado 04-303 induziu uma expressão significativamente mais baixa e estável destas citocinas. Com base nestes resultados, o isolado *M. bovis* 04-303 foi caracterizado como um isolado virulento enquanto o isolado 534 foi caracterizado como um isolado atenuado.

Blanco et al. (2009) definiram as diferenças de transcriptoma entre *M. bovis* 04-303 e *M. bovis* 534. Um total de 49 genes foram diferencialmente expressos entre ambas os isolados, 35 genes em níveis mais elevados em *M. bovis* 04-303, enquanto que 14 genes tiveram uma expressão mais elevada em *M. bovis* 534. Resultados de *microarray* e RT-qPCR mostraram a expressão acentuada dos genes *mce4D*, *Mb2607/Mb2608* no isolado virulento 04-303 enquanto *alkB*, *Mb3277c* e *Mb1077* foram expressos em níveis mais elevados no isolado atenuado 534. Em relação aos 35 genes diferencialmente expressos, a expressão desses mudou drasticamente entre ambos isolados durante o cultivo *in vitro*, exibindo um comportamento semelhante aos macrófagos em bovinos. A análise em nível de nucleotídeos revelou que estes genes

foram conservados em ambos isolados, sugerindo que as diferenças de expressão podem ser atribuídas a alterações tais como deleções ou a mutações em regiões de regulação da expressão gênica.

Um conhecimento mais aprofundado acerca dos fatores de virulência pode ser essencial para o desenvolvimento de novas vacinas e medicamentos para ajudar a controlar a doença e avançar para um mundo livre de tuberculose. Tendo em conta os fenótipos extremos patogênicos atribuídos aos isolados *M. bovis* 04-303 e *M. bovis* 534, lançamos a hipótese de que as diferenças de virulência entre esses isolados podem ser atribuídas a mutações nos genes de virulência. Com base nos diferentes fenótipos de expressão gênica e de virulência observados, a análise genômica comparativa foi realizada a fim de identificar um repertório de genes que podem desempenhar papel na virulência. A análise comparativa foi realizada utilizando o genoma *M. bovis* AF2122/97 (Garnier et al. 2003) como referência.

Resultados

Mutações nucleotídicas foram encontradas em 30 genes relacionados a virulência no genoma do isolado *M. bovis* 04-303 (Tabela 1), quando comparado ao genoma AF2122/97. Mutações *non-synonymous missense* foram encontradas em 21 genes: *dnaA* (*Mb0001*), *fgd2* (*Mb0137c*), *TetR* (*Mb0310*), *def* (*Mb0437c*), *MmcO* (*Mb0869c*), *mb1393* (*Mb1393*), *ansA* (*Mb1565c*) *pknE* (*Mb1772*), *eccD5* (*Mb1823*), *PPE32* (*Mb1837*), *PPE34* (*Mb1951c*), *ppsA* (*Mb2074c*), *uppP* (*Mb2160c*), *LppN* (*Mb2293*), *hrcA* (*Mb2395c*), *hpx* (*Mb3196c*), *mb3436c* (*Mb3436c*), *bphC* (*Mb3599c*), *mmpL8* (*Mb3853c*), *esat6* (*Mb3905*) e *sigM* (*Mb3941*). Além disso, nove genes exibiram deleções: *LuxR* (*Mb0201*), *mb0637* (*Mb0637*), *pstS1* (*Mb0959*), *mb1029c* (*Mb1029c*), *mb1393* (*Mb1393*), *pks10* (*Mb1688*), *ilvG* (*Mb1851*), *mb2608* (*Mb2608*), *mb3692c* (*Mb3692c*) e um gene exibiu uma inserção: *mb2056c* (*Mb2056c*).

O genoma do isolado *M. bovis* 534 exibiu mutações nucleotídicas em 44 genes relacionados à virulência (Tabela 1), quando comparado ao genoma AF2122/97. Mutações *non-synonymous missense* foram encontradas em 24 genes: *dnaA*, *fgd2*, *TetR*, *pknG* (*Mb0418c*), *def*, *vapC29* (*Mb0634*), *MmcO*, *glgP* (*Mb1363*), *irtA* (*Mb1383*), *mb1393*, *pks10* (*Mb1688*), *narX* (*Mb1765c*), *pknE*, *eccD5*, *mb1843c* (*Mb1843c*), *ppsA*, *uppP*, *LppN*, *clgR* (*Mb2765*), *PPE46* (*Mb3043c*), *tetR* (*Mb3277c*), *mb3436c*, *bphc* and *sigM*. Deleções em 22 genes: *mce1R* (*Mb0170c*), *ilvD* (*Mb0195c*), *rsIA* (*Mb0757*), *CpdA* (*Mb0828*), *vapC9* (*Mb0985*), *irtB* (*Mb1384*), *mb1393*, *pks10*, *mb1762c*, *ilvG*, *mb1966c* (*Mb1966c*), *pknJ* (*Mb2115*), *mbtE* (*Mb2401c*), *cysH* (*Mb2413*), *mb2608*, *pimA* (*Mb2642c*), *recA* (*Mb2756c*), *lpdA* (*Mb3331c*), *cut4* (*Mb3482*), *eccA1* (*Mb3898*) e inserções em dois genes: *mb1295* (*Mb1295*) and *mb2056c* (*Mb2056c*).

Dezoito mutações foram idênticas em ambos isolados, *M. bovis* 04-303 e 534, localizadas nos genes *dnaA*, *fgd2*, *TetR*, *def*, *MmcO*, *mb1393*, *pknE*, *eccD5*, *ilvG*, *mb1056c*, *ppsA*, *uppP*, *mb2608*, *mb3436c*, *bphc*, *sigM*, *LppN* e *pks10* (Figura 1).

*Todas as mutações (*synonymous*, *non-synonymous*, deleções e inserções) encontradas estão listadas no arquivo adicional.

Discussão

Recentemente, avanços consideráveis na compreensão de bases moleculares de patogenicidade, virulência e persistência de micobactérias foram alcançados. Uma contribuição significativa foi a identificação de genes essenciais relacionados à virulência micobacteriana. A maioria dos genes relacionados à virulência codificam enzimas de várias vias lipídicas, proteínas de superfície celular, reguladores e proteínas sinalizadoras da transdução. Estes resultados sugerem que as espécies patogênicas têm adaptado seus genomas de um estilo de vida livre para o meio intracelular com a aquisição mínima de genes exclusivos a virulência.

De acordo com Collins (2001) fatores de virulência de *M. bovis* são propriedades especiais que permitem a micobactéria infectar, sobreviver, multiplicar e

causar doença em um hospedeiro animal. No entanto, esta definição não é tão simples. A demonstração formal da propriedade de virulência de um gene requer o cumprimento dos postulados moleculares de Koch que podem ser resumidos nas três etapas seguintes. Primeiro, é necessário que o referido gene codifique o determinado fator de virulência ou controle a sua produção ou exportação em isolados virulentos. Em segundo, o organismo tem de se tornar avirulento quando esse gene é interrompido. E em terceiro, a complementação do gene inativo por um gene ativo deve restaurar a virulência do organismo (Falkow, 1998).

Como demonstrado anteriormente, 04-303 é uma cepa virulenta de *M. bovis*, enquanto 534 é uma cepa atenuada (Aguilar Leon et al., 2009, Blanco et al., 2009). Considerando estes fenótipos extremos, hipotetizamos que a diferença de virulência entre eles pode ser atribuída a mutações em genes relacionados à virulência. Neste contexto, a discussão é focada em genes que apresentaram mutações *non-synonymous*, deleções ou inserções nos genomas dos isolados 04-303 e 534, quando comparados ao genoma referência *M. bovis* (AF2122/97). No presente estudo, os determinantes de virulência foram agrupados de acordo com sua função, anotação da sequência de DNA, características moleculares e/ou localização celular em: (1) proteínas do envelope celular, (2) metabolismo de lipídeos e ácidos graxos, (3) proteína-quinases, (4) proteases, (5) proteínas transportadoras de metais, (6) reguladores da expressão gênica, (7) proteínas de função desconhecida e (8) outros fatores de virulência. Um resumo, de todos os genes relacionados à virulência aqui descritos, é apresentado na Tabela 1.

(1) Proteínas do envelope celular

A estrutura da parede celular das micobactérias é extremamente complexa, composta por uma estrutura de lipídeos e proteínas. O seu "núcleo" é constituída por peptidoglicanos ligados covalentemente a galactofuranos lineares, por sua vez ligados a vários cordões de arabinofurano altamente ramificados unidos a ácidos micólicos que são longas cadeias de α -alquil β -hidroxi ácidos graxos. Intercalados a este "núcleo" estão o dimicocerosate fitiocerol (PDIM), trealose 6,6'-dimicolato (TDM), sulfolipídeos e

fosfatilinositol manosídeo (PIMs). Fosfatilinositol manosídeo, lipomanana e lipoarabinomanana estão ancorados à membrana plasmática e se estendem para o exterior da parede celular. Estudos proteômicos da parede celular micobacteriana identificaram mais de quinhentas proteínas nesta estrutura celular e algumas delas são exclusivas a patogenicidade micobacteriana e, portanto, espera-se que sejam excelentes alvos para maiores investigações acerca da virulência de *M. bovis* (Corner, 1994; Smith, 2003; Forrellad et al., 2013).

(i) **Esat6:** ESAT-6 (do inglês, 6-kDa *early secretory antigenic target*) faz parte de uma família de proteínas denominada Esx que compreende 23 proteínas pequenas (cerca de 100 aminoácidos). Os primeiros membros desta família a serem descritos são codificados pelos genes adjacentes *esxA* e *esxB*, que estão localizados no locus *esx-1* (Sorensen et al., 1995). ESAT-6 de *M. bovis* tem sido descrita contígua a outras pequenas proteínas denominadas CFP-10 (do inglês, *culture filtrate antigen*) (Mb3904) como forte estimuladores da imunidade mediada por células (Jones et al., 2010) e, uma vez que não estão presentes em micobactérias não patogênicas, incluindo a cepa vacinal *M. bovis* BCG (Bacilo Calmette-Guerin) (Waters et al., 2004), elas são consideradas antígenos específicos para o diagnóstico da tuberculose bovina e proteínas relevantes associadas à virulência no CMT (Pollock et al, 1997; Vordermeier et al, 2001).

Recentemente, a regulação da resposta imune foi descrita em *M. tuberculosis* com a participação de ESAT-6, isolada ou no complexo com a proteína chaperona CFP-10 (ESAT-6:CFP-10). Verificou-se que ESAT-6 e ESAT-6:CFP-10 podem sequestrar a proteína do hospedeiro beta-2-microglobulina (b2M) inibindo a expressão na superfície celular do complexo MHC-I-b2M, prejudicando assim a resposta imune adaptativa do hospedeiro e conseqüentemente favorecendo o estabelecimento de uma infecção bem sucedida (SREEJIT et al., 2014). Foi também demonstrado que ESAT-6 provoca lise celular através da formação de poros na membrana celular (SMITH et al., 2008).

No presente estudo, uma única mutação (A para G) foi encontrada no gene *esat6* (*esxA*) no genoma do isolado *M. bovis* 04-303 quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3905*), resultando na substituição preditiva de uma treonina, um aminoácido polar com cadeias laterais neutras, por uma alanina, um aminoácido com cadeias laterais apolares. Uma vez que o papel de ESAT-6 na interação patógeno-hospedeiro durante a infecção por *M. tuberculosis* foi claramente estabelecida, é possível que as alterações em determinados domínios da proteína ESAT-6 possam aumentar sua propriedade de virulência. Na verdade, a principal causa da atenuação da cepa BCG é a ausência dos genes codificadores de CFP10 e ESAT6. Assim, dada a relevância de ESAT-6 na patogênese da infecção, uma mutação desfavorável como esta que foi descrita no genoma do isolado 04-303 poderia ter implicações para o fenótipo virulento.

(ii) **Mce1R:** As proteínas MCE (do inglês, *mammalian cell entry*) são um grande grupo de proteínas secretadas ou expostas à superfície da célula, organizadas em um grande *operon*, que recebem este nome por causa da relação demonstrada entre Mce1 e a capacidade das micobactérias de entrar em células de mamíferos e sobreviver no interior de macrófagos (Arruda et al., 1993). Shimono et al. (2003) demonstraram que cepas de *M. tuberculosis* com interrupção no *operon mce1* ocasionaram granulomas mal organizados nos pulmões de camundongos, e reduzida capacidade de estimulação de resposta imune do tipo Th1; além disso, os mutantes foram incapazes de estabelecer um estágio estabilizado de infecção pulmonar nos camundongos do referido estudo. Este *operon* é regulado negativamente pelo gene *mce1R* (*rv0165c* em *M. tuberculosis*) (Cole et al., 1998). Experimentos com *M. tuberculosis* H37Rv mutantes para o gene *mce1R* sugeriram que a expressão desregulada de produtos do *operon mce1* determinam resposta granulomatosa tecidual tardia (Casali et al., 2006). Assim o *operon mce1* pode homeostaticamente regular a arquitetura da parede celular *in vivo* e provocar resposta tecidual granulomatosa organizada permitindo a *M. tuberculosis* estabelecer infecção a longo prazo. Neste trabalho foi detectada uma única mutação no gene *mce1R* no genoma do isolado 534

(G para ., posição 194.983, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (mb0170c). A deleção de um nucleotídeo produz *frameshift*. Assim, uma possível falha na tradução desta proteína parece ser compatível ao fenótipo atenuado do isolado 534.

(iii) **PstS-1:** PstS-1 (do inglês, *periplasmic phosphate-binding lipoprotein-1*), pertence à família de transportadores de fosfato ABC, caracterizada como uma glicoproteína de adesão com diversas atividades biológicas (Andersen & Hansen, 1989; Sanchez et al., 2009), sendo uma delas a interação com receptores de manose de macrófagos e mediação de fagocitose (Esparza et al., 2015). A adesão a células hospedeiras é um ponto crítico e fator de virulência essencial de bactérias patogênicas, especialmente bactérias intracelulares (Klemm & Schembri, 2000). Além disso, a proteína PstS-1 está envolvida na importação de fosfato durante períodos de restrição de energia, garantindo a micobactéria permanência no interior do fagossoma (Rengarajan et al., 2005). Foi demonstrado ainda que esta proteína possui motivos estruturais capazes de interagir com *Toll-like receptors* (TLR) 2 e TLR 4, e assim, induzir a produção de citocinas pró-inflamatórias como o fator de necrose tumoral alfa (TNF- α) e interleucina (IL) 6 por monócitos humanos (Espitia et al., 1992). *Mycobacterium tuberculosis* nocauteado para os genes *pstS-1* e *pstS-2* apresentaram uma taxa de multiplicação significativamente reduzida no interior de macrófagos peritoneais de camundongos e apresentaram-se atenuados em modelo de infecção *in vivo* (Peirs et al., 2005). Neste estudo foi detectada uma única mutação no gene *pstS-1* no genoma do isolado 04-303 (G para ., posição 1.043.504, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0959*). A mutação descrita aqui para o gene *pstS-1* pode contribuir para a virulência do isolado 04-303 por ocasionar alterações na interação de *M. bovis* com o sistema imune do hospedeiro, em favor dos bacilos.

(iv) **Mb1029c:** Mb1029c (Rv1002c em *M. tuberculosis*) é uma dolichyl fosfatomanose manosiltransferase, enzima esta que atua no transporte de resíduos de manose de carreadores lipídicos a proteínas (Strahl-Bolsinger, Gentsch & Tanner,

1999). A expressão ectópica de Rv1002c em *M. smegmatis* demonstrou que esta proteína de membrana está acoplada ao sistema de exportação de proteínas Sec-dependente, interferindo no transporte apenas de proteínas extracitoplasmáticas (VanderVen et al., 2005). Rv1002c é caracterizado como um gene essencial para o ótimo crescimento de *M. tuberculosis in vitro*, e *M. tuberculosis* mutantes para o gene *rv1002c* apresentaram crescimento severamente impactado *in vitro* e *in cellulo* associado a atenuação de sua patogenicidade em camundongos imunocomprometidos (Liu et al., 2013). Neste trabalho foi detectada uma única mutação no gene *mb1029c* no genoma do isolado 04-303 (G para ., posição 1.120.062, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1029c*). Sabe-se que a ausência de atividade da enzima manosil transferase ou a alteração desta atividade pode modificar as propriedades imunogênicas de determinadas proteínas e, assim, alterar a interação de *M. bovis* com o sistema imune do hospedeiro. Assim, supomos que esta mutação pode ocasionar em, pelo menos, estratégia de evasão ao sistema imune do hospedeiro e desta forma contribuir para a virulência do isolado *M. bovis* 04-303.

(v) **Mb1762c:** Mb1762c (Rv1733c em *M. tuberculosis*) é uma proteína transmembrana caracterizada como antígeno altamente expresso por *M. tuberculosis* em estágio de latência e bem reconhecida pelas células T a partir de indivíduos infectados por *M. tuberculosis* em estágio de latência (Zhang et al., 2014). Neste estudo foi detectada uma única mutação no gene *mb1762c* no genoma do isolado 534 (G para ., posição 1.945.541, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1762c*). Células de camundongos imunizadas com peptídeos Rv1733c produziram IFN- γ /TNF α , IFN γ /CD4+ e anticorpos específicos a Rv1733c e apresentaram uma redução significativa da carga bacteriana nos pulmões de camundongos desafiados com *M. tuberculosis* (Coppola et al., 2015). Assim, os dados expostos confirmam a imunogenicidade desta proteína, a mutação aqui descrita pode ter contribuído para a atenuação do fenótipo 534, estudos adicionais são necessários para vincular estas informações.

(vi) **PimA:** PIMs são os fosfolipídeos mais proeminentes e distinguíveis de micobactérias. Em *M. bovis* BCG, fosfatidilinositol e PIMs representam mais de 56% de todos os fosfolipídeos da parede celular e 37% da membrana citoplasmática e são, assim, considerados importantes componentes estruturais que atuam como "cimento" do esqueleto da parede celular (Jackson et al., 2000). PimA (do inglês, alpha-(1-2)-phosphatidylinositol manosiltransferase) está localizada nas proximidades do gene responsável pela síntese de fosfatidilinositol e codifica uma manosil transferase envolvida nos primeiros passos da síntese de fosfatidilinositol manosil. Além disso PimA é uma importante proteína moduladora da resposta imune durante a infecção de tuberculose e lepra (Nigou et al., 2002), bem como chave de ligação nas interações entre *M. tuberculosis* e as células fagocíticas do hospedeiro (Gilléron et al., 2001).

Para determinar se PimA é uma enzima essencial às micobactérias, Kordulakova et al. (2002) construíram *M. smegmatis* mutantes para o gene *pimA*. A capacidade destes mutantes de síntese da enzima manosil-transferase foi dependente a presença de uma cópia funcional do gene *pimA* transportado em um plasmídeo de resgate sensível à temperatura. Eles demonstraram que os mutantes para o gene *pimA* são incapazes de crescer a uma temperatura mais elevada do que a temperatura limite do plasmídeo. Assim, a síntese de fosfatidilinositol manosil e de seus derivados em *M. smegmatis* parece ser dependente da presença do gene *pimA* e essencial para o crescimento. Pirson et al. (2015) isolaram moléculas puras de PIM (do inglês, phosphatidylinositol dimannoside [PIM2], acylphosphatidylinositol dimannoside [AcPIM2], diacyl-phosphatidylinositol dimannoside [Ac2PIM2], acylphosphatidylinositol hexamannoside [AcPIM6], e diacylphosphatidylinositol hexamannoside [Ac2PIM6]) a partir de isolados virulentos de *M. tuberculosis* a fim de avaliar o potencial destas moléculas para estimular resposta de células mononucleares de sangue periférico (PBMC) de bovinos infectados por *M. bovis*. Destas moléculas, AcPIM6 induziu níveis significativos de IFN- γ em PBMC, enquanto, AcPIM6, Ac2PIM2 e Ac2PIM6 conduziram proliferação significativa em PBMC. AcPIM6 foi subsequentemente utilizado em ensaio de proliferação celular por citometria de fluxo. Esta análise demonstrou que AcPIM6 foi predominantemente reconhecidos por células CD3+, CD335+ e células NKT.

Neste estudo, uma única mutação no gene *pimA* (G para ., posição 2.905.671, deleção) foi detectada no genoma do isolado 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2642c*). Devido à elevada homologia entre os genomas de *M. tuberculosis* e *M. bovis* especula-se que este gene desempenhe a mesma função em ambas espécies. Considerando a capacidade destas moléculas de lipídeos em interagir com populações de linfócitos de bovinos, e ativar especificamente células NKT em bovinos e provável função desta proteína, é possível especular que a mutação aqui descrita para o gene *pimA* pode ter contribuído para a atenuação do isolado *M. bovis* 534.

(vii) **Mb0637:** Mb0637 (Rv0621 em *M. tuberculosis*) é uma proteína de membrana. Neste estudo foi detectada uma única mutação do gene *mb0637* no genoma do isolado 04-303 (G para ., posição 716.130, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0637*). Dan et al. (2009) sugeriram que Rv0621 pode estar envolvido na regulação da síntese de ácidos graxos e na resistência de *M. tuberculosis* a defesa ao surfactante do hospedeiro. A relevância desta mutação no fenótipo do isolado 04-303 ainda não está clara e estudos adicionais são necessários.

(2) Metabolismo de lipídeos e ácidos graxos

Micobactérias são únicas entre os patógenos bacterianos que exibem uma grande variedade de lipídeos complexos e lipoglicanos em sua superfície celular. Estes lipídeos exclusivos de parede celular são conhecidos por desempenhar um papel importante na patogênese, por conseguinte, os genes responsáveis pela sua biossíntese, transporte e degradação são potenciais fatores de virulência (Forrellad et al., 2013, Székely & Cole, 2016).

(i) **Pks10:** O genoma de *M. bovis* revelou uma notável variedade de poliketide sintases (PKS). A maioria dos genes codificantes de PKS estão implicados na biossíntese de lipídeos complexos, especialmente do ácido micocerosídeo, encontrado no envelope de *M. bovis* (Rousseau et al., 2003). No presente estudo, o

gene *pks10* exibiu uma dupla mutação (A para G, posição 1.860.533, *non-synonymous missense* e G para ., posição 1.859.643, deleção) no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1688*). A primeira mutação descrita resultou na substituição preditiva desfavorável de uma serina (aminoácido polar com cadeias laterais neutras) por uma glicina (aminoácido apolar, alifático). *Mycobacterium tuberculosis* mutantes para o gene *pks10* foram descritos apresentando síntese defeituosa de PDIM, um importante lipídeo de superfície de micobactérias patogênicas. Estes mutantes *pks10* também apresentaram-se significativamente atenuados para o crescimento em modelo experimental murino (Sirakova et al., 2003). Assim, as enzimas envolvidas na biossíntese de PDIM representam alvos potenciais que podem afetar a patogênese de micobactérias e o fenótipo de virulência, parecendo ser compatível com o fenótipo atenuado do isolado 534.

(ii) **MmpL8:** *Mmpl8*, membro de uma família de proteínas de membrana potencialmente envolvidas com transporte de lipídeos e requerida para a síntese de sulfolipídeos (SL-1) (Domenech, Reed & Barry, 2005). Além disso, esta proteína é fundamental para a montagem completa de sulfoglicolipídeos tetracilados (Ac4SGL) (Layre 2011), que compoem grande parte dos lipídeos de parede celular de micobactérias patogênicas. Brodin et al. (2010) sugeriram que Ac4SGL aceleram a maturação do fagossoma, favorecendo o tráfego de fagossomas contendo *M. tuberculosis* para compartimentos mais acidificados. Foi reportado que a formação de fagolisossomas ácidos favorece a apresentação de antígenos de *M. tuberculosis* e conseqüente eliminação da bactéria. Assim, especulamos que a mutação aqui descrita no gene *mmpL8* afeta negativamente a produção de Ac4SGL no isolado *M. bovis* 04-303 e conseqüentemente aumenta a persistência do isolado no interior de macrófagos e evita a exposição do antígeno ao sistema imune do hospedeiro. Por outro lado, Domenech et al. (2004) demonstraram que mutantes para o gene *MmpL8* foram capazes de estabelecer infecção normalmente porém apresentaram-se significamente

atenuados. Assim, estudos complementares são necessários para relacionar esta mutação com a virulência de *M. bovis*.

(iii) **EccA1:** EccA1 é uma ATPase que pertence ao sistema de exportação ESX-1 (Abdallah et al., 2007) requerido para a secreção de proteínas cruciais a virulência e patogenia de micobactérias, incluindo *M. tuberculosis* (Rv3868) e *M. bovis* (Bitter et al., 2009). Este sistema inclui três ATPases: EccCa1/EccCb1, cuja função está relacionada ao transporte de proteínas através da membrana citoplasmática (Abdallah et al., 2007) e EccA1 uma proteína essencial ao sistema ESX-1 (Gao et al., 2004) além de fator de virulência, uma vez que participa da secreção e regulação da síntese de lipídeos e ácidos micólicos (Bitter et al., 2009). Joshi et al. (2012) relataram síntese defeituosa de ácido micólico em mutantes para o gene codificador da enzima EccA1-ATPase, correlacionado a diminuição da virulência e do crescimento intracelular *in vivo*. Neste estudo, o genoma do isolado *M. bovis* 534 exibiu uma deleção de um nucleotídeo no gene *eccA1* (C para ., posição 4.279.990), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3898*). Considerando que a maioria dos lipídeos de superfície, assim como TDM, PDIM e glicolipídeos fenólicos, estão implicados na virulência micobacteriana (Chopra and Gokhale, 2009; Minnikin et al., 2002), nosso resultado é compatível com o perfil atenuado do isolado *M. bovis* 534.

(iv) **Hpx:** O gene *hpx* gene é uma possível não heme haloperoxidase requerida ao crescimento da micobactéria no colesterol (Griffin et al., 2011). No gene *hpx*, uma única mutação (A para G, posição 3.497.977, *non-synonymous missense*) foi detectada no genoma do isolado *M. bovis* 04-303, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3196c*). Esta mutação resulta na substituição preditiva desfavorável de uma aspartato (aminoácido polar com cadeias laterais carregadas negativamente) por uma glicina. Estudos são necessários para entender a relevância desta mutação, uma vez que não há estudos na literatura descrevendo o papel do gene *hpx* na virulência micobacteriana.

(3) Proteínas quinases

A fosforilação de proteínas é o principal mecanismo pelo qual os sinais extracelulares são traduzidos em respostas celulares. Estas reações são efetuadas por proteínas quinases específicas e fazem par a reações de desfosforilação realizadas por fosfatases (PPases).

(i) **PknG:** PknG é uma quinase funcional capaz de fosforilar o seu domínio quinase, bem como o seu próprio domínio tetratricopeptídeo. Este gene está localizado em um *operon* que contém *glnH*, um gene que codifica uma proteína potencialmente envolvida na captação de glutamina. No entanto, a participação de PknG na regulação do metabolismo do glutamato é controversa (Forrellad et al., 2013). Enquanto, *M. tuberculosis* mutantes para o gene *pknG* apresentaram redução de sobrevivência *in vitro* e *in vivo*, acúmulo de glutamato e glutamina e redução de 50% na síntese de glutamina apoiando a hipótese de que PknG medeia a transferência de sinais de detecção de defasagem nutricional em *M. tuberculosis* e as traduzem em adaptação metabólica (Cowley et al., 2004); Nguyen et al. (2005) relataram que *M. bovis* BCG mutantes para o gene *pknG* não apresentaram diferenças no nível intracelular de glutamina quanto na captação deste aminoácido. Estudos de infecção de macrófagos por *M. tuberculosis* revelam que os primeiros 73 resíduos de aminoácidos desta proteína, o motivo Trx, e o resíduo de treonina 63 são independentemente essenciais para modulação de PknG e mediação da sobrevivência das micobactérias no seu hospedeiro (Tiwari et al. 2009). Foi relatado que PknG é a única enzima que de fato sofre autofosforilação sobre resíduos de triptofano situados na porção N terminal, embora esta função não pareça estar relacionada com a regulação de sua atividade quinase, é essencial para impedir o tráfego de *M. bovis* BCG para o interior de lisossomas e para a sobrevivência das micobactérias em macrófagos derivados de medula óssea em modelo murino (Scherr et al., 2009). Neste trabalho, foi detectada uma única mutação no gene *pknG* (C para T, posição 497.411, *non-synonymous missense*) no genoma do isolado *M. bovis* 04-303, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0418c*). Esta mutação resulta na substituição preditiva desfavorável de uma glicina por um glutamato (aminoácido polar com cadeias laterais carregadas negativamente). Análises confirmativas devem ser

realizadas a fim de compreender a verdadeira função deste gene na virulência de micobactérias bem como correlacionar a mutação aqui descrita ao perfil virulento do isolado *M. bovis* 04-303.

(ii) PknJ: PknJ é uma quinase funcional que não possui ortólogo detectável em micobactérias que não pertençam ao CMT, sugerindo possível adaptação das espécies do CMT à vida intracelular (Narayan et al., 2007). Sobre sua função, estudos recentes sugerem que a enzima glicolítica piruvato quinase A seja substrato potencial de pknJ (Cohen-Gonsaud et al., 2009). Singh et al. (2014) estabeleceram papel importante de pknJ no crescimento de micobactérias, descreveram aumento da sobrevivência de *M. tuberculosis* em linhagem celular THP-1 e definiram que esta resposta é regulada pela fosforilação de piruvato quinases. Em nosso estudo, uma única mutação no gene *pknJ* (C para ., posição 2.327.145, deleção) foi detectado no genoma do isolado 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2115*), o que sugere que esta mutação pode ter contribuído com o fenótipo atenuado 534.

(4) Proteases

Proteases desempenham papel crucial na homeostase celular, controlando proteínas envolvidas na transcrição, regulação, metabolismo e virulência. Os agentes patogênicos microbianos frequentemente utilizam proteases extracelulares como fatores de virulência que podem desempenhar diferentes funções na destruição de tecidos ou na modulação da resposta imune através da inativação de moléculas de defesa do hospedeiro, tais como imunoglobulinas e componentes do complemento. Elas também podem ativar proteínas ou peptídeos reguladores chave e adquirir nutrientes que regulam a expressão gênica (Zhao & Xie, 2011).

(i) **CpdA:** o gene *mb0828* em *M. bovis* (*Rv0805* em *M. tuberculosis*) codifica uma metalofosfoesterase que apresenta atividade hidrolítica à cAMP. Adenosina monofosfato cíclico (cAMP) é um modulador chave da fisiologia bacteriana regulando

uma variedade de processos que vão desde o metabolismo do carbono até mecanismos de virulência (Gazdik et al., 2009). Neste trabalho foi detectada uma única mutação no gene *CpdA* no genoma do isolado 534 (G para ., posição 900.385, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0828*). A super expressão de Rv0805 tem sido usada como ferramenta de redução de níveis intracelulares de cAMP para assim elucidar as funções de cAMP em micobactérias. Níveis de cAMP em *M. tuberculosis* foram reduzidos em aproximadamente 30% após a super expressão de Rv0805, e níveis de transcrição de determinado número de genes, incluindo genes relacionados a virulência, e o ciclo de metil citrato foram alterados. Mutantes para o gene *rv0805* demonstraram que a assinatura transcricional da super expressão de Rv0805 é uma combinação de efeitos dependentes e independentes de catalisação, e que a estrutura flexível da porção C-terminal de Rv0805 é crucial para o efeito independente de catalisação da proteína (Mantage et al., 2013). Assim, a mutação aqui descrita no gene *CpdA* pode ter contribuído para o fenótipo atenuado do isolado *M. bovis* 534.

(ii) **ClgR:** O gene *clgR* codifica uma proteína que desempenha importante papel na resposta ao *stress redox*, e é um dos aproximadamente 100 genes que participam da resposta de reaeração, mecanismos pelos quais as micobactérias retornam a um estado de replicação após re-exposição a condições favoráveis (Sherrid et al., de 2010, McGillivray et al., 2014). Uma única mutação no gene *clgR* foi descrita no genoma do isolado 534 (C para A, posição 3.014.181, *nonsynonymous missense*) quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2765*), conduzindo à substituição preditiva desfavorável de uma asparagina (aminoácido polar com cadeias laterais não carregadas) por aspartato. Sherrid et al. (2010) induziram um modelo experimental *in vitro* para avaliar mecanismos envolvidos com a resposta de reativação de *M. tuberculosis*. O estudo demonstrou que a protease ClgR é regulada positivamente durante a transição de *M. tuberculosis* da bacteriostase para o crescimento. Além disso, como demonstrado por McGillivray et al. (2014), mutantes isogênicos de *M. tuberculosis* para o gene *rv2745c* são mais sensíveis a alterações do

stress redox in vitro gerados por diamida. Assim, a mutação detectada no gene *clgR* também pode ter contribuído para atenuação do isolado 534.

(iii) **Mb3692c:** Mb3992c (Rv3668c em *M. tuberculosis*) é uma serina protease que poderia interromper a produção de citocinas dos macrófagos e implicar na ação de fármacos e dormência (Cerdeira-Maira & Darwin 2009). A localização na parede celular da micobactéria desta protease suporta seu papel na interação com o hospedeiro (Armstrong 2006). *Mycobacterium smegmatis* recombinantes-Rv3668c induziram positivamente a secreção de citocinas pró-inflamatórias TNF- α , IL-1 e IL-6 e negativamente a secreção de citocinas anti-inflamatórias e IL-10 por células U937. Peptídeos Rv3668c ocasionaram a sobrevivência prolongada de células U937 e aceleraram a morte de células do hospedeiro. Experimentos inibitórios mostraram que quinases reguladas por sinais extracelulares de fatores nucleares do eixo kB (ERK-NF-kB) resultaram no desencadeamento de alterações em Rv3668c, TNF- α e IL-1. Assim, Zhao et al. (2014) forneceram evidências que comprovam o envolvimento de Rv3668c na interação entre *Mycobacterium* e hospedeiro. Neste estudo foi detectada uma única mutação no gene *mb3992c* no genoma do isolado 04-303 (G para ., posição 4.047.766, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3692c*). Portanto, a mutação no gene *mb3992c* pode ter interferido na modulação da resposta imune do hospedeiro a favor do patógeno.

(5) Proteínas transportadoras de metais

Os metais fazem parte de grupos protéticos ou são co-factores de determinadas enzimas e são essenciais para a sobrevivência de procariotos e eucariotos, tanto no ambiente como no interior da célula. Entretanto estas proteínas não são fatores de virulência reais que proporcionam às bactérias vantagens significativas para adaptação e sobrevivência em células hospedeiras.

(i) **IrtA/IrtB:** Um dos principais mecanismos através dos quais as bactérias obtêm nutrientes essenciais a partir do hospedeiro é a aquisição de ferro. O ferro pode

ser adquirido pelo contato direto entre as bactérias e as moléculas carreadoras deste metal do hospedeiro, seguida por sua remoção por redução e subsequente absorção. Alternativamente, as micobactérias utilizam moléculas, de baixo peso molecular, sequestrantes de ferro chamadas sideróforos. Neste método alternativo, três proteínas estão envolvidas na importação e exportação de ferro através da membrana e consequente absorção. IrtA é um sideróforo exportador e ambas as proteínas do sistema IrtB são importantes sideróforos importadores de ferro (Farhana et al., 2008).

Uma única mutação no gene *irtA* (G para T, posição 1.512.766, *non synonymous missense*) foi detectada no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1383*), conduzindo à substituição preditiva desfavorável de uma arginina (aminoácido polar com cadeias laterais carregadas positivamente) por uma leucina (aminoácido não polar, alifático), enquanto uma deleção de um nucleotídeo no gene *irtB* (C para ., posição 1.515.055) foi detectada no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1384*). Farhana et al. (2008) demonstraram que *M. smegmatis* nocauteados para os genes *irtA* e *irtB* exibiram crescimento extremamente reduzido devido à absorção incompleta de ferro. Sabe-se que para patógenos intracelulares a assimilação de ferro é essencial para contornar a escassez de produção além de determinante de virulência. Assim, é possível especular que as mutações aqui encontradas nos genes *irtA/irtB* são parcialmente responsáveis pelo fenótipo atenuada do isolado *M. bovis* 534.

(ii) MbtE: MbtE (Rv2380c em *M. tuberculosis*) é codificada pelo *cluster mbt* predicto como um peptídeo não-ribossomal sintetase. Sabe-se que mutação em qualquer gene *mbt* interrompe a síntese de algum determinado sideróforos, os quais, por sua vez, capacitam a bactéria em adquirir determinado metal a partir do meio. Consequentemente, as bactérias não conseguem sobreviver na célula hospedeira (Voss et al., 2000). Reddy et al. (2003) relataram que a interrupção da biossíntese de micobactina ocasiona a atenuação de *M. tuberculosis* para crescimento e virulência. Este grupo demonstrou também que *M. tuberculosis* mutantes para o gene *mbtE* foram

incapazes de sintetizar micobactinas, exibiram colônias com morfologia alterada e apresentaram-se atenuados para crescimento em meio de cultura e em macrófagos. Microscopia eletrônica de transmissão revelou que *M. tuberculosis* mutantes para o gene *mbtE* exibiram alterações na permeabilidade da parede celular. Além disso, cobaias infectados com *M. tuberculosis* mutantes para o gene *mbtE* exibiram redução significativa na carga bacilar e em danos histopatológicos nos órgãos, quando comparados com animais infectados com *M. tuberculosis*. Neste estudo, identificamos dupla mutação no gene *mbtE* no genoma do isolado 534 (G para ., C para ., posições 2.630.194, 2.634.471, deleção) quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2401c*). Assim, frente a importância deste gene na virulência das micobactérias e fenótipo atenuado de mutantes para este gene descritos na literatura, as mutações aqui descritas parecem ser compatíveis com o fenótipo atenuado do isolado *M. bovis* 534.

(6) Reguladores da expressão gênica

(i) **HrcA:** Um importante conjunto de proteínas que tem sido classificado entre os membros do CMT é a família de proteínas de choque térmico (do inglês, heat shock). Estes fatores altamente conservados são induzidos tanto no patógeno como no hospedeiro durante o processo de infecção, auxiliando na manutenção da integridade celular, contribuindo com a sinalização imune (Asea et al., 2000) e reconhecimento do patógeno (Castellino et al., 2000). HrcA (Rv2374c em *M. tuberculosis*) é uma proteína de choque térmico que atua como repressor transcricional na resposta Hsp60 (GroE) em *M. tuberculosis* (Stewart et al., 2002).

Stewart et al. (2001) demonstraram que a ruptura parcial da regulação da resposta de choque térmico em *M. tuberculosis* tem impacto importante sobre a virulência micobacteriana, prejudicando a capacidade da bactéria em persistir durante a fase crônica da infecção. Além disso, as proteínas de choque térmico desempenham papel importante na apoptose (Stuart et al., 1998; Lang et al., 2000; Wang et al., 2002) e no mecanismo implicado na patogênese de granulomas (Rutherford et al., 2001;

Stridh et al., 2002; Xaus et al., 2003). Postulou-se que o equilíbrio entre fatores pró-apoptóticos e anti-apoptóticos influencia a resolução do processo inflamatório ou o desenvolvimento da enfermidade (Dubaniewicz et al., 2006). No presente estudo, o isolado virulento *M. bovis* 04-303 exibiu uma mutação *non-synonymous missense* no gene *hrcA* (G para A, posição 2.622.933) quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2395c*), ocasionando a substituição preditiva desfavorável de um aspartato para uma asparagina (aminoácido polar de cadeias laterais não carregadas). A mutação aqui descrita no gene *hrcA* no genoma do isolado *M. bovis* 04-303, um repressor da transcrição de proteínas de choque térmico, pode ocasionar uma super expressão destas proteínas associadas a virulência.

(ii) **TetR:** TetR é membro de uma família de reguladores da transcrição que atuam frequentemente como repressores e são amplamente distribuídos entre as bactérias (Yang et al., 2012). Em *M. tuberculosis*, EthR (Rv3855) regula a expressão de um gene mono-oxigenase que catalisa a ativação de etionamida, um antibiótico utilizado no tratamento da tuberculose e repressor da família TetR/CamR (Dover et al., 2004). Em *M. bovis*, a proteína TetR é um regulador transcricional codificada pelo *operon* Mb3277c/Mb3280c (Blanco et al., 2009; Liu et al., 2014). No genoma do isolado *M. bovis* 534 foi detectada uma mutação no gene *TetR* (T para C, posição 3.584.552, *non-synonymous missense*), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3277c*), resultando na substituição preditiva desfavorável de uma serina, por uma prolina (aminoácido apolar, alifático). Blanco et al. (2009) demonstraram que os genes *mb1077c* e *mb3277c*, codificadores de fatores regulatórios, foram regulados negativamente no isolado virulento 04-303. Fato este que sugere que estes reguladores podem controlar negativamente a expressão de relevantes proteínas durante a infecção de *M. bovis* em modelo murino. Portanto, é possível especular que a mutação aqui descrita no gene *TetR* pode alterar a expressão deste gene, contribuindo assim para a atenuação do isolado *M. bovis* 534.

(iii) **LuxR:** LuxR pertence a uma família de fatores de transcrição. Neste estudo, o genoma do isolado virulento *M. bovis* 04-303 exibiu uma deleção de um

nucleotídeo no gene *luxR* (G para ., posição 231.486), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0201*). Fang et al. (2013) demonstraram que a proteína Rv0195 desempenha papel fundamental na dormência de *M. tuberculosis* além de ser importante para a virulência micobacteriana em células THP-1 em modelo murino. A exclusão do gene *rv0195* em *M. tuberculosis* não afetou o crescimento bacteriano e a sobrevivência a longo prazo em condições de aerobiose porém, diminuiu a sobrevivência das células, a capacidade de recuperação do estágio de dormência em modelo *in vitro* de anaerobiose e a capacidade de sobrevivência intracelular em condições de hipóxia e estresse. Além disso, a exclusão do gene *rv0195* diminuiu a virulência micobacteriana em macrófagos humanos e resultou em redução da sobrevivência bacteriana e patogenicidade em modelo murino (C57BL/6) de infecção (Canino et al., 2013). Portanto, podemos inferir que a mutação aqui descrita no gene *luxR* interfere na tradução desta proteína (*frameshift*) contribuindo com a virulência do isolado *M. bovis* 04-303.

(iv) **RslA:** O genoma de *M. bovis* codifica alguns fatores sigma dos quais a grande maioria se enquadra na subfamília de função extracitoplasmática (ECF). Um destes fatores sigma é o fator SigI que está codificado adjacente a um gene codificador de uma proteína de membrana (*mb0757* em *M. bovis* e *rv0736* em *M. tuberculosis*) e apresenta a sequência conservada HXXXCXXC. Este motivo é encontrado em fatores anti-sigma que regulam vários fatores sigma ECF, incluindo aqueles que controlam as respostas ao estresse oxidativo (Song et al., 2003). SigI e Rv0736 foram descritos cotranscritos, e o domínio intracelular de Rv0736 foi capaz de interagir especificamente com SigL sugerindo assim que *rv0736* codifique um fator anti-sigma SigI (*rslA*). Este fator sigma ECF (*rslA*) parece regular a síntese de lipídeos do envelope celular e controlar a secreção de proteínas modificadas. Hahn et al. (2005) demonstraram também que *M. tuberculosis* mutantes para o gene *rsIA* são severamente atenuados em modelo murino de infecção. SigI e RslA interagem e RslA parece funcionar como regulador negativo de SigI. Em conjunto, estes resultados sugerem um modelo no qual RslA pode detetar sinais extracelulares deste fator e assim diminuir sua ligação a SigI.

O aumento da atividade dos genes regulados por SigI alteram as características da superfície de *Mycobacterium*. Neste estudo, o genoma do isolado *M. bovis* 534 exibiu deleção de um nucleotídeo no gene *rsIA* (C para ., posição 830.473), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0757*). Frente ao exposto, é possível especular que a deleção de um nucleotídeo no gene *rsIA* descrita aqui pode ter interferido com a regulação e expressão de SigI, contribuindo para a atenuação do isolado *M. bovis* 534.

(v) **Mb1966c:** A proteína Mb1966c (Rv1931c em *M. tuberculosis*) é anotada como provável regulador de transcrição que pertence à família AraC, estima-se que sua função seja regular genes de funções diversas, variando desde o metabolismo do carbono até virulência micobacteriana (Martin & Rosner 2001). Frota et al. (2004) demonstraram que *M. tuberculosis* mutante para o gene *rv1931c* apresentaram diminuição da sobrevivência tanto em macrófagos como em modelo murino, com estabelecimento da sobrevivência com a complementação do gene *rv1931c* ativo. Estes resultados sugerem que Rv1931c regula genes importantes relacionados a virulência de *M. tuberculosis*. Neste estudo, uma única mutação no gene *mb1966c* (C para ., posição 2.174.242, deleção) foi detectada no genoma do isolado 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1966c*). Assim esta mutação neste proteína reguladora pode alterar o perfil de seus reguladores de expressão, resultando em atenuação do isolado 534.

(7) Proteínas de função desconhecida

O advento da genômica permitiu sequenciar o genoma de muitas espécies de *Mycobacterium*. Fato este, somado a estudos de proteômica, transcriptômica, metabólicos e análises de bioinformática permitiram definir ou inferir a função biológica de muitas ORFs. Apesar disso, a função fisiológica de muitas proteínas relacionadas a virulência de *Mycobacterium* permanece desconhecida.

Família PE/PPE: As famílias PE/PPE foram descritas apenas em micobactérias e consistem em duas grandes famílias distintas de proteínas com domínios ácidos e ricas em glicina cujos genes estão agrupados (Cole et al., 1998). Muitos destes genes estão localizados ao redor ou fazem parte do *operon ESX*. O nome PE deriva do motivo Pro-Glu, nos resíduos 8 e 9 na extremidade N-terminal, ao passo que PPE representa o motivo Pro-Pro-Glu. A maioria destas proteínas estão localizadas na superfície da célula e/ou são secretadas, e induzem forte resposta imune no hospedeiro (Forrellad et al., 2013). A função da maioria das proteínas PE/PPE é desconhecida, porém as propriedades de algumas delas incluem associações com o envelope celular (Bani et al., 2002), exposição na superfície da célula micobacteriana, efeitos sobre a agregação bacteriana e interações com células do hospedeiro e/ou sistema imune (Brennan et al., 2001). O elevado grau de polimorfismo entre os membros destas famílias de proteínas sugere papel relacionado ao mecanismo de evasão da resposta imune do hospedeiro, embora uma ampla variedade de funções, também foram atribuídas a elas (Mukhopadhyay & Balaji, 2011). No entanto, ainda é incerto se as proteínas PE/PPE modulam o sistema imune do hospedeiro em favor dos bacilos, uma vez que somente algumas proteínas PE/PPE têm mostrado estar, certamente, relacionada a virulência do CMT (Forrellad et al., 2013).

Neste estudo uma única mutação no gene *ppe32* (G para A, posição 2.041.061, *non-synonymous missense*, conduzindo à substituição de um triptofano a um *stop codon*) e uma única mutação no gene *ppe34* (C para T, posição 2154292, *non-synonymous missense*, conduzindo à substituição preditivo desfavorável de uma glicina por uma arginina) foram detectados no genoma do isolado *M. bovis* 04-303, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1837*, *mb1951c*, respectivamente). No genoma do isolado *M. bovis* 534 uma única mutação foi detectada no gene *ppe46* (C para G, posição 3.333.886, *non-synonymous missense* conduzindo à substituição preditivo de uma valina [aminoácido apolar, alifático] por uma leucina), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3043c*). Estes genes estão localizados em uma região com elevado grau de

variação por isso não podemos fazer nenhuma inferência sobre essas mutações e suas consequências na virulência destes isolados.

(i) **NarX**: NarX (Mb1765c; Rv1736c em *M. tuberculosis*) é uma proteína de função desconhecida embora tenha sido anotada como "redutase de nitrato" devido à sua homologia com genes de redução de nitrato em outros organismos. Entretanto, um estudo demonstrou que a capacidade de redução de nitrato em aerobiose e anaerobiose em *M. tuberculosis* não é afetada pela ausência de *NarX* (Cole et al., 1998). Por outro lado, outro estudo com *M. bovis* BCG demonstrou relação entre o gene *NarX* e a capacidade de redução de nitrato e sobrevivência durante período de dormência em condições de anaerobiose (Honaker et al., 2008). No presente estudo, o gene *NarX* apresentou uma única mutação (C para T, posição 1.949.403, *non-synonymous missense*) no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1765c*). A mutação descrita resultou na substituição preditiva desfavorável de uma glicina por uma arginina. Assim, mesmo com a função obscura de *NarX* hipotetizamos que a mutação aqui descrita parece ser compatível com o fenótipo atenuado do isolado *M. bovis* 534.

(ii) **Mb1843c**: Mb1843c (Rv1813c em *M. tuberculosis*) codifica uma proteína hipotética conservada com função desconhecida. Bretl et al. (2012) realizaram uma caracterização inicial de Rv1813c em *M. tuberculosis*, incluindo sua regulação por MprAB e DosRST. Os resultados destes estudos indicaram que Rv1813c e Rv1812c são cotranscritos e ambos regulados por MprAB e DosRST. *Mycobacterium tuberculosis* mutantes para o gene *rv1813c* apresentaram-se atenuados em modelo murino de aerossol de baixa dose de tuberculose, carga bacteriana reduzida, redução na sobrevivência e diminuição na capacidade de estimular citocinas pró-inflamatórias IL-1 e IL-12. No presente estudo, uma única mutação (C para T) foi descrita no gene *mb1843c* no genoma do isolado *M. bovis* 534, resultando na substituição preditiva desfavorável de uma alanina por uma treonina. Portanto, uma mutação no gene *mb1843c* poderia ter contribuído para o fenótipo atenuado do isolado *M. bovis* 534.

Outras proteínas relacionadas a virulência

(i) **VapC29 and VapC9:** VapC é uma proteína toxina-antitoxina (TA) que está disposta em um *operon* bicistrônico (Anantharaman e Aravind, 2003) e codifica uma toxina estável e uma antitoxina lábil: a toxina atua em diversos processos celulares enquanto a antitoxina restringe a ação da toxina (Gerdes, 2000). A toxina VapC atua como ribonuclease inibindo o processo de tradução (Zhang et al., 2003). Em tempos de estresse e privação de nutrientes, proteases celulares degradam as antitoxinas e, assim, permitem que as toxinas livres, retardem o crescimento da célula ou mesmo matem a célula (Hayes, 2003).

Mb0634 ou vapC29 (Rv0617 em *M. tuberculosis*) apresenta um domínio PIM ribonuclease (Ahidjo et al., 2011). Em *vapC29*, uma única mutação (T para C, posição 712.319, *non-synonymous missense*) foi detectada no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0634*). Esta mutação resulta na substituição preditiva desfavorável de uma valina por uma metionina (aminoácido apolar e alifático).

Mb0985 ou vapC9 (Rv0960 em *M. tuberculosis*) também é uma ribonuclease. Em *vapC9*, uma única mutação (C para ., posição 1.074.361, deleção) foi detectada no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0985*). Atráves de análises de perfil de expressão em microarranjos, Jain et al. (2006) encontraram 33 genes em *M. tuberculosis* regulados positivamente durante as fases iniciais da invasão da barreira hematoencefálica por *M. tuberculosis*, sendo que 18 deles pertencem a uma ilha genômica descrita anteriormente (Rv0960-Rv1001). Além disso, foi demonstrado que a supressão das ribonucleases MazF, membro do complexo TA, reduz a virulência (Tiwari et al., 2015). O papel de *vapC29* e *vapC9* no complexo *M. tuberculosis* não é bem definido e o impacto das mutação aqui descritas nos genes *vapC29* e *vapC9* no genoma do isolado *M. bovis* 534 merece investigação adicional.

(ii) **IlvD:** Dihidroxiácido desidratase, enzima chave envolvida na biossíntese de aminoácidos de cadeia ramificada (BCAA), catalisa a síntese de cetoácidos a partir

de dihidroxiácidos. Esta enzima é codificado em *M. tuberculosis* pelo gene *rv0189c*, enquanto que em *M. bovis* pelo gene *mb0195c*. Estudos apontaram *Rv0189c* como gene essencial para o crescimento de *M. tuberculosis in vitro*. *Rv0189c*, provavelmente, seja necessária para o crescimento intracelular, como é evidente pela sua regulação positiva *ex vivo* e *in vivo*, e pela sua capacidade de lidar com as condições de estresse (Singh et al., 2011). Neste estudo, foi detectada uma única mutação do gene *ilvD* no genoma do isolado 534 (C para ., posição 221.361, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb0195c*). Esta deleção produz uma *frameshift*, e conseqüentemente ocasiona falha na tradução desta proteína, sendo compatível com o fenótipo atenuado do isolado 534.

(iii) **Mb1295:** Mb1295 (*Rv1264* em *M. tuberculosis*) é uma adenil ciclase envolvida no metabolismo energético. Esta enzima atua em pH ácido (6,0), ou seja, pH o qual *M. tuberculosis* encontra no interior do fagossoma. Conseqüentemente *Rv1264* tem sido apontado como detector do meio fagossomal resultando em adaptação de *M. tuberculosis* no ambiente intracelular. *M. tuberculosis* mutantes para o gene *rv1264*, no entanto, apresentaram fenótipo selvagem de virulência (Dittrich et al., 2006). Neste estudo, uma única mutação no gene *mb1295* no genoma do isolado 534 (. para AG, posição 1.411.239, inserção) foi detectado, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1295*). Deste modo, esta inserção deve alterar a estrutura da proteína expressa (*frameshift*) e contribuir para o fenótipo atenuado apresentado pelo isolado 534.

(iv) **AnsA:** AnsA (*Rv1538c* em *M. tuberculosis*) é uma asparaginase que atua em conjunto com os transportadores AnsP1/AnsP2 permitindo a absorção de asparagina e a desaminação. Gouzy et al. (2014) forneceram evidências de que AnsA é liberada para o meio de cultura celular de *M. tuberculosis in vitro* e para o interior do fagossoma das micobactérias. Assim AnsA é importante para a migração da bactéria para o interior de fagossomos ácidos e sobrevivência em macrófagos, atuando então como fator de virulência. No presente estudo, uma única mutação (T para C) foi encontrada no gene *ansA* no genoma do isolado *M. bovis* 04-303, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb1565c*), resultando na

substituição preditiva de uma metionina por uma valina. Além de AnsA ser crucial na migração da micobactéria para fagossomos ácidos e replicação intracelular, *M. tuberculosis* com supressão de asparaginase apresentou fenótipo atenuado em macrófagos em modelo murino. Estudos adicionais são necessários para compreender a relevância desta mutação e relacioná-la com a virulência micobacteriana.

(v) **CysH:** 5 ϵ -adenosinofosfosulfato redutase (CysH), enzima essencial para a produção de metabólitos contendo enxofre reduzido e crítica a sobrevivência de *M. tuberculosis* em fase de infecção crônica em camundongos (Loiseau et al., 2003). Disrupção do gene *cysH* gerou *M. tuberculosis* auxotróficos para cisteína e metionina, e virulência atenuada em camundongos BALB/c e C57BL/6 imunocompetentes. *Mycobacterium tuberculosis* mutantes e selvagens replicaram de forma semelhante durante a fase aguda de infecção, porém os mutantes apresentaram diminuição da viabilidade durante a fase de infecção persistente. Os mutantes também apresentaram-se susceptíveis a peroxinitrito e peróxido de hidrogênio *in vitro*. Estes resultados demonstraram que CysH é importante para a proteção de *M. tuberculosis* durante a fase crônica de infecção, a resistência à nitrosativos e ao estresse oxidativo podendo ser um mecanismo de proteção micobacteriano (Senaratne et al., 2006). Neste trabalho foi detectada uma única mutação do gene *mb2413* no genoma do isolado 534 (G para ., posição 2.654.933, deleção), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2413*). A literatura citada suporta a associação desta mutação com o fenótipo atenuado do isolado 534, entretanto novos estudos são necessários para comprovar esta relação.

(vi) **RecA:** RecA (recombinase A) é uma enzima chave para o reinício da replicação (HR, do inglês, restart replication), que promove a troca de cadeias entre as moléculas de DNA. Durante o RH, um complexo de DNAs (DNA de cadeia simples) revestido por RecA reconhece homologia em DNAs (DNA de cadeia dupla), as “invadem”, e, como consequência, catalisam a troca de cadeias (Kowalczykowski, 1994; Sasseti & Rubin, 2003). Szulc-Kielbik et al. (2015) relataram que em *M. tuberculosis* RecA é indispensável a inibição da ativação do mitógeno de proteína quinase com atividade bactericida dependente de macrófagos THP-1 *in vitro* e

sugeriram que RecA contribua para a virulência de *M. tuberculosis*. Observaram também uma diminuição na ativação de ERK1/2 em macrófagos infectados por *M. tuberculosis* enfatizando que *recA* represente um importante mecanismo para que as cepas selvagens escapem da morte intracelular. Neste trabalho, foi descrita a deleção de dois nucleotídeos no gene *recA* no genoma do isolado 534 (GG para .., posição 3.007.225), quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb2765c*). Pouco se sabe sobre a função deste gene em *M. bovis* e seu impacto na virulência, e assim novos estudos são necessários para determinar esta relação. No entanto, a mutação descrita aqui parece ser compatível com o fenótipo atenuado do isolado *M. bovis* 534.

(vii) **LpdA:** LpdA é anotada como uma provável lipoamido desidrogenase e recentemente foi relatado que os genes *mb3331c/rv3303c* codificam um nova variedade de flavoproteína dissulfureto redutase que catalisa a redução de fosfato NADH ou NAD(P)H (Argyrou et al., 2004). A capacidade de micobactérias em sobreviver sobre condições de estresse oxidativo *in vivo* é um aspecto importante da patogênese. Tem sido demonstrado que NAD(P)H redutase administra o estresse oxidativo em outras bactérias patogênicas. Akhtar et al. (2006) relataram que peptídeos Rv3303c de *M. tuberculosis* possuem a capacidade de proteger contra o estresse oxidativo *in vivo* e contribuir para a virulência em modelo murino. Neste estudo, uma única mutação no gene *mb3331c* (C para ., posição 3.645.254, deleção) foi detectado no genoma do isolado 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3331c*). Dada a provável importância deste gene na virulência de *M. bovis* e a deleção de nucleotídeos aqui descrita, estudos adicionais devem ser desenvolvidos, a fim de confirmar a associação desta mutação e o perfil atenuado do isolado *M. bovis* 534.

(viii) **Cut4:** As enzimas que pertencem à família de cutinase são serinas ativas e apresentam um amplo painel de substratos tais como a cutina, os triacilgliceróis e os fosfolipídeos (Shue et al., 2010). As cutinases são enzimas extracelulares ECT envolvidas na virulência de bactérias e fungos fitopatogênicos uma vez que hidrolisam cutina, um poliéster de plantas responsável pela proteção, limitando a evaporação e

protegendo-a de várias agressões (Kolattukudy, 2001). No genoma de *M. tuberculosis* H37Rv sete genes que codificam para cutinases semelhantes foram identificados com propriedades imunogênicas, sugerindo um potencial papel como candidatas vacinais. Foi demonstrado que estas enzimas apresentam várias funções fisiológicas, desde a formação da parede celular até efeitos citotóxicos sobre as células hospedeiras (Cole et al., 1998). Uma delas, a cutinase 4 é uma fosfolipase A2, que apresenta efeitos citotóxicos sobre macrófagos. No gene *cut4*, uma única mutação (G para ., posição 3.821.778, deleção) foi detectada no genoma do isolado *M. bovis* 534, quando comparado ao gene ortólogo no genoma referência AF2122/97 (*mb3482*). Muitas enzimas, tais como fosfolipase C e fosfolipase A2, estão envolvidos na virulência em muitas espécies bacterianas (Titball 1998; Istivan & Coloe 2006). O papel de Cut4 no CMT não é claro e o impacto dessa deleção no isolado atenuado *M. bovis* 534 merece mais investigações.

(ix) **GlgP:** Mb1363 (Rv1328 em *M. tuberculosis*) codifica a enzima GlgP, uma fosforilase catabólica de glicogênio (Schneider et al., 2000). Esta enzima pertence a um grupo estruturalmente relacionado e onipresente de proteínas que são geralmente consideradas como enzimas degradadoras de glucano que catalisam a produção de glucose-1-fosfato pela clivagem reversível nas extremidades não poliglucanos, tais como maltodextrinas, amido e glicogênio (Buchbinder et al., 2001; Chia et al., 2004). Para compreender a função biológica bacteriana de GlgP Casajús Alonso et al. (2006) produziram e caracterizaram células de *E. coli* com expressão nula de *glgP* ou expressão alterada. GlgP mutantes não apresentaram atividade de fosforilase de glicogênio, indicando que toda a atividade enzimática é dependente do produto *glgP*. Os resultados globais assim demonstrados comprovam a participação de GlgP na catalisação da quebra de glicogênio e consequente alteração na estrutura do glicogênio através da remoção de unidades de glucose das cadeias de polissacarídeos externos em *E. coli*. No genoma do isolado *M. bovis* 534, uma mutação no gen *glgP* (A para C, posição 1.495.014, *non-synonymous missense*) foi detectada, em comparação ao gene ortólogo no genoma referência AF2122/97, resultando na substituição preditiva de uma metionina por uma leucina. Outros estudos são necessários para compreender a

relevância desta mutação, bem como a função desta enzima no metabolismo de *M. bovis*.

Diferentes tipos de mutações foram identificados nos genomas dos isolados de perfis patogênicos contrastantes. Todas as mutações aqui descritas codificam proteínas relacionadas à patogênese de *M. tuberculosis* ou *M. bovis*. Entretanto, o impacto desses polimorfismos na virulência dos fenótipos micobacterianos ainda necessita de uma análise tridimensional *in silico* destas proteínas e de demonstrações experimentais.

Materiais e Métodos

Isolados bacterianos

M. bovis 04-303 e *M. bovis* 534 foram isolados de um javali em um campo aberto e de um bovino leiteiro na região úmida dos Pampas, respectivamente, ambos na Argentina. O DNA dos isolados foi extraído e purificado como descrito por Van Embden et al., (1992).

Sequenciamento genômico e anotação

Ambos genomas, *M. bovis* 04-303 (Nishibe et al. 2013) e 534, foram depositadas no DDBJ/EMBL/GenBank, sob os números de acesso GenBank: AVSW01000000 e GenBank: JQEM00000000, respectivamente. A montagem do genoma foi realizada com um total de 169 *contigs* (cobertura média de 27,3 x) para o isolado *M. bovis* 04-303 e com 72 *contigs* (cobertura média de 50,5 x) para o isolado *M. bovis* 534.

A fim de avaliar a qualidade da montagem dos genomas, realizamos uma série de análises com o programa CheckM (Parks et al. 2014), um método automatizado que utiliza um conjunto mais amplo de marcadores de genes específicos para a posição de um genoma dentro de uma árvore de genomas referência e também informações sobre a colocação destes genes. De acordo com limites padrão CheckM, ambas montagens

apresentam boa qualidade. A densidades de codificação de ambos os genomas estão dentro do intervalo esperado em genomas bacterianos. A Tabela 2 apresenta os principais resultados obtidos através do programa CheckM para ambas montagem.

Análise de genes relacionados à virulência

Todos os genes de ambos os genomas foram analisados, tendo como base a revisão de Forrellad et al., (2013) e genes listados na base de dados do servidor TubercuList (<http://genolist.pasteur.fr/TubercuList>).

Polimorfismos de único nucleotídeo

Todos os SNPs encontrados nos genomas dos isolados 04-303 e 534, quando comparados ao genoma AF2122/97 (acesso NC_002945), e posicionados em qualquer gene relacionado a virulência foram analisados. Os SNPs foram encontrados utilizando a ferramenta SAMtools Versão 1.2 (Li et al. 2009) e depois convertidos para um formato vcf através de BCFtools Versão 1.2 (Danecek et al., 2011), estas ferramentas são conhecidas e largamente empregadas para identificar e manipular “chamada de variantes” (*variant calls*) em genomas. Os SNPs foram classificados mutações *synonymous* (sinônimas ou silenciosas), *non-synonymous* (não sinônimas)(*missense* ou *nonsense*), deleção ou inserção. Concentramos nossa atenção em mutações *non-synonymous*, deleções e inserções, uma vez que as mesmas alteram a sequência primária de aminoácidos de proteínas “mutantes”. No entanto, as mutações *synonymous* podem gerar códons raros e assim conduzir a uma tradução ocasional de genes mutantes (arquivo adicional 1). Além disso, os *codons* raros podem ser convertidos em *codons* mais frequentes, favorecendo assim a tradução.

Em ordem de determinar um *threshold* (limite) para a qualidade dos SNPs, calculamos, para cada isolado, a taxa de transições por transversões de todos os escores de qualidade gerados na chamada de variantes. O valor esperado para esta taxa é de entre 2,0 a 2,1 para as análises de genoma (DePristo et al.2011). Além disso,

Liu e colaboradores (Liu et al., 2006) relataram uma razão igual a 2,06 para *M. tuberculosis*. Assim determinamos o valor desta taxa como 2,0, resultando no *threshold* de escore de qualidade de 98 para 04-303 e 43 para 534, como indicado na Figura 2. Assim, todos os SNPs cujo índice de qualidade foram inferiores aos limites correspondentes foram descartados.

Arquivo adicional:

Arquivo adicional 1: Lista de todos os genes relacionados a virulência de *Mycobacterium bovis*, assim como revisado por Forrellad et al., (2013) e/ou listados na base de dados do servidor *TubercuList*, analisados para os genomas dos isolados *M. bovis* 04-303 e 534, utilizando AF2122/97 como genoma referência.

Financiamento

Este projeto foi financiado pela Embrapa (03.14.00.054.00.00, 02.13.16.002.00.00), Fundect (TO0096/2012; 23/200.656/2013) e CNPq (443235/2014-7; 305857/2013-4).

Referências:

Abdallah, A. M., Gey van Pittius, N. C., Champion, P. A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., and Bitter, W. (2007). Type VII secretion—mycobacteria show the way. *Nat. Rev. Microbiol.* **5**, 883–891.

Aguilar León, D., Zumárraga, M. J., Jiménez Oropeza, R., Gioffré, A. K., Bernardelli, A., Orozco Estévez, H., Cataldi, A. A., Hernández Pando, R. (2009). *Mycobacterium bovis* with different genotypes and from different hosts induce dissimilar immunopathological lesions in a mouse model of tuberculosis. *Clin. Exp. Immunol.* **157**, 139-147.

Ahidjo, B.A., Kuhnert, D., McKenzie, J. L., Machowski, E. E., Gordhan, B. G., Arcus, V., Abrahams, G. L., Mizrahi, V. (2011). VapC toxins from *Mycobacterium tuberculosis* are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins. *PLoS One* **6**(6):e21738.

Akhtar, P., Srivastava, S., Srivastava, A., Srivastava, M., Srivastava, B. S., Srivastava, R. (2006). Rv3303c of *Mycobacterium tuberculosis* protects tubercle bacilli against oxidative stress in vivo and contributes to virulence in mice. *Microbes Infect.* **8**(14-15), 2855-2862.

Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R. M. (2010). Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg. Infect. Dis.* **16**, 1296–1299.

Alonso-Casajús, N., Dauvillée, D., Viale, A. M., Muñoz, F. J., Baroja-Fernández, E., Morán-Zorzano, M. T., Eydallin, G., Ball, S., Pozueta-Romero, J. (2006). Glycogen phosphorylase, the product of the *glgP* Gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*. *J. Bacteriol.* **188**(14), 5266-5272.

Anantharaman, V. and Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.* **4**(12):R81.

Andersen, A. B., Hansen, E.B. (1989). Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecular-weight protein of *Mycobacterium tuberculosis*. *Infect. Immun.* **57**(8), 2481-2488.

Argyrou, A., Vetting, M. W., Blanchard, J. S. (2004). Characterization of a new member of the flavoprotein disulfide reductase family of enzymes from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **279**, 52694e52702.

Armstrong, P. B. (2006). Proteases and protease inhibitors: a balance of activities in host-pathogen interaction. *Immunobio.* **211**(4), 263-281. Review.

Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T., Riley, L.W. (1993) Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* **261**: 1454–1457.

Asea, A., Kraeft, S. K., Kurt-Jones, E. A., Stevenson, M. A., Chen, L. B., Finberg, R. W., Koo, G. C., Calderwood, S. K. (2000). HSP70 stimulates cytokine production through a CD14-dependent pathway, demonstrating its dual role as a chaperone and cytokine. *Nat. Med.* **6**, 435-442.

Banu, S., Honoré, N., Saint-Joanis, B., Philpott, D., Prévost, M. C., Cole, S. T. (2002). Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol. Microbiol.* **44**(1), 9-19.

Bitter, W., Houben, E.N., Bottai, D., Brodin, P., Brown, E.J., Cox, J.S., Derbyshire, K., Fortune, S.M., Gao, L.Y., Liu, J., et al. (2009). Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog.* **5**, e1000507.

Blanco, F. C., Nunez-Garcia, J., Garcia-Pelayo, C., Soria, M., Bianco, M. V., Zumárraga, M., Golby, P., Cataldi, A.A., Gordon, S.V., Bigi, F. (2009). Differential

transcriptome profiles of attenuated and hypervirulent strains of *Mycobacterium bovis*. *Microbes Infect.* **11**, 956-963.

Brennan, M. J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M., Jacobs, W. R. Jr. (2001). Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect. Immun.* **69**(12), 7326-7333.

Bretl, D. J., He, H., Demetriadou, C., White, M. J., Penoske, R. M., Salzman, N. H., Zahrt, T. C. (2012). MprA and DosR coregulate a *Mycobacterium tuberculosis* virulence operon encoding Rv1813c and Rv1812c. *Infect. Immun.* **80**(9), 3018-3033.

Brodin, P., Poquet, Y., Levillain, F., Peguillet, I., Larrouy-Maumus, G., Gilleron, M., Ewann, F., Christophe, T., Fenistein, D., Jang, J., Jang, M. S., Park, S. J., Rauzier, J., Carralot, J. P., Shrimpton, R., Genovesio, A., Gonzalo-Asensio, J. A., Puzo, G., Martin, C., Brosch, R., Stewart, G. R., Gicquel, B., Neyrolles, O. (2010). High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. *PLoS Pathog.* **9**;6(9):e1001100, doi: 10.1371/journal.ppat.1001100.

Buchbinder, J. L., V. L. Rath, and R. J. Fletterick. (2001). Structural relationships among regulated and unregulated phosphorylases. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 191–209.

Casali, N., White, A.M., Riley, L.W. (2006) Regulation of the *Mycobacterium tuberculosis mce1* operon. *J. Bacteriol.* **188**: 441–449.

Castellino, F., Boucher, P. E., Eichelberg, K., Mayhew, M., Rothman, J. E., Houghton, A. N., Germain, R. N. (2000). Receptor-mediated uptake of antigen heat shock protein complexes results in major histocompatibility complex class I antigen presentation via two distinct processing pathways. *J. Exp. Med.* **191**, 1957-1964.

Cerda-Maira, F. and Darwin, K. H. (2009). The *Mycobacterium tuberculosis* proteasome: more than just a barrel-shaped protease. *Microbes Infect.* **11**(14-15), 1150-1155. Review.

Chia, T., D. Thorneycroft, A. Chapple, G. Messerli, J. Chen, S. C. Zeeman, S. M. Smith, and A. M. Smith. (2004). A cytosolic glucosyltransferase is required for conversion of starch to sucrose in Arabidopsis leaves at night. *Plant J.* **37**, 853–863.

Chopra, T., and Gokhale, R.S. (2009). Polyketide versatility in the biosynthesis of complex mycobacterial cell wall lipids. *Methods Enzymol.* **459**, 259–294.

Cohen-Gonsaud, M., Barthe, P., Canova, M. J., Stagier-Simon, C., Kremer, L., Roumestand, C., et al. (2009). The *Mycobacterium tuberculosis* Ser/Thr kinase

substrate Rv2175c is a DNA-binding protein regulated by phosphorylation. *J. Biol. Chem.* **284**, 19290-19300.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544.

Collins, D. M. (2001) Virulence factors of *Mycobacterium bovis*. *Tubercul.* 81(1-2), 97-102. Review.

Corner, L. A. (1994) Post mortem diagnosis of *Mycobacterium bovis* infection in cattle. *Vet. Microbiol.* **40**(1-2), 53-63. Review.

Coppola, M., van den Eeden, S. J., Wilson, L., Franken, K. L., Ottenhoff, T. H., Geluk, A. (2015). Synthetic Long Peptide Derived from *Mycobacterium tuberculosis* Latency Antigen Rv1733c Protects against Tuberculosis. *Clin. Vaccine Immunol.* **22**(9), 1060-1069.

Cowley, S., Ko, M., Pick, N., Chow, R., Downing, K. J., Gordhan, B. G., Betts, J. C., Mizrahi, V., Smith, D. A., Stokes, R. W., Av-Gay, Y. (2004). The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Mol. Microbiol.* **52**(6), 1691-702.

Dan, L., Jianping, X., Ruzhen, G., Honghai, W. (2009) Cloning and characterization of Rv0621 gene related to surfactant stress tolerance in *Mycobacterium tuberculosis*. *Mol. Biol. Rep.* **36**(7), 1811-1817.

Dan, L., Jianping, X., Ruzhen, G., Honghai, W. (2009) Cloning and characterization of Rv0621 gene related to surfactant stress tolerance in *Mycobacterium tuberculosis*. *Mol. Biol. Rep.* **36**(7), 1811-1817.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. (2011). 1000 Genomes Project Analysis Group, Collaborators. The variant call format and VCFtools. *Bioinformatics* **27**(15), 2156-2158.

dePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M. et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* **43**(5):491–498.

De Voss, J. J., Rutter, K., Schroeder, B. G., Su, H., Zhu, Y., Barry, C. E. (2000). The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc. Natl. Acad. Sci. USA* **97**, 1252–1257.

Dittrich, D., Keller, C., Ehlers, S., Schultz, J. E., Sander, P. (2006) Characterization of a *Mycobacterium tuberculosis* mutant deficient in pH-sensing adenylate cyclase Rv1264. *Int. J. Med. Microbiol.* **296**(8), 563-566.

Domenech, P., Reed, M. B., Dowd, C. S., Manca, C., Kaplan, G., Barry, C. E. (2004). The role of MmpL8 in sulfatide biogenesis and virulence of *Mycobacterium tuberculosis*. *J. Biol. Chem.* **279**(20): 21257-21265.

Domenech, P., Reed, M.B., and Barry, C.E. (2005) Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. *Infect. Immun.* **73**: 3492–3501.

Dover, L. G., Corsino, P. E., Daniels, I. R., Cocklin, S. L., Tatituri, V., Besra, G. S., Futterer, K. (2004). Crystal structure of the TetR/CamR family repressor *Mycobacterium tuberculosis* EthR implicated in ethionamide resistance. *J. Mol. Biol.* **340**, 1095–1105.

Dubaniewicz, A., Trzonkowski, P., Dubaniewicz-Wybieralska, M., Dubaniewicz, A., Singh, M., Myśliwski, A. (2006). Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J. Clin. Immunol.* **26**(3), 243-250.

Esparza, M., Palomares, B., García, T., Espinosa, P., Zenteno, E., Mancilla, R. (2015) PstS-1, the 38-kDa *Mycobacterium tuberculosis* glycoprotein, is an adhesin, which binds the macrophage mannose receptor and promotes phagocytosis. *Scand. J. Immunol.* **81**(1), 46-55.

Espitia, C., Elinos, M., Hernandez-Pando, R., Mancilla, R. (1992). Phosphate starvation enhances expression of the immunodominant 38-kilodalton protein antigen of *Mycobacterium tuberculosis*: demonstration by immunogold electron microscopy. *Infect. Immun.* **60**, 2998–3001.

Fabre, M., Koeck, J. L., Le Flèche, P., Simon, F., Hervé, V., Vergnaud, G., Pourcel, C. (2004). High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of "*Mycobacterium canettii*" strains indicates that the *M. tuberculosis* complex is a recently emerged clone of "*M. canettii*". *J. Clin. Microbiol.* **42**(7), 3248-3255.

Falkow, S. (1998) Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* **10**, S274-S276.

Fang, H., Yu, D., Hong, Y., Zhou, X., Li, C., Sun, B. (2013). The LuxR family regulator Rv0195

modulates *Mycobacterium tuberculosis* dormancy and virulence. *Tuberculosis* (Edinb). **93**(4), 425-431.

Farhana, A., Kumar, S., Rathore, S. S., Ghosh, P. C., Ehtesham, N. Z., Tyagi, A. K., Hasnain, S. E. (2008). Mechanistic insights into a novel exporter-importer system of *Mycobacterium tuberculosis* unravel its role in trafficking of iron. *PLoS One* **3**(5):e2087.

Forrellad, M. A., Klepp, L. I., Gioffré, A., García, J. S., Morbidoni, H. R., Santangelo, M. P., Cataldi, A. A., Bigi, F. (2013). Virulence factors of the *Mycobacterium* complex. *Virulence* **4**(1), 3-66.

Frota, C. C., Papavinasasundaram, K. G., Davis, E. O., Colston, M. J. (2004). The AraC family transcriptional regulator Rv1931c plays a role in the virulence of *Mycobacterium tuberculosis*. *Infect. Immun.* **72**(9), 5483-5486.

Gao, L.Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J.N., and Brown, E.J. (2004). A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol. Microbiol.* **53**, 1677–1693.

Gazdik, M. A., Bai, G., Wu, Y., McDonough, K. A. (2009). Rv1675c (cmr) regulates intramacrophage and cyclic AMP-induced gene expression in *Mycobacterium tuberculosis*-complex mycobacteria. *Mol. Microbiol.* **71**(2), 434-448.

Gerdes, K. (2000). Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J. Bacteriol.* **182**(3), 561-572. Review.

Gilleron, M., Ronet, C., Mempel, M., Monsarrat, B., Gachelin, G., Puzo, G. (2001) Acylation state of the phosphatidylinositol mannosides from *Mycobacterium bovis* bacillus Calmette Guérin and ability to induce granuloma and recruit natural killer T cells. *J. Biol. Chem.* **276**(37), 34896-34904.

Gouzy, A., Larrouy-Maumus, G., Bottai, D., Levillain, F., Dumas, A., Wallach, J. B., Caire-Brandli, I., de Chastellier, C., Wu, T. D., Poincloux, R., Brosch, R., Guerquin-Kern, J. L., Schnappinger, D., Sório de Carvalho, L. P., Poquet, Y., Neyrolles, O. (2014). *Mycobacterium tuberculosis* exploits asparagine to assimilate nitrogen and resist acid stress during infection. *PLoS Pathog.* **10**(2):e1003928. doi: 10.1371/journal.ppat.1003928.

Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J., Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **7**(9):e1002251.

Gutiérrez, M., Tellechea, J., García Marín, J. F. (1998). Evaluation of cellular and serological diagnostic tests for the detection of *Mycobacterium bovis* infected goats. *Vet. Microbiol.* **62**, 281–290.

Hahn, M. Y., Raman, S., Anaya, M., Husson, R. N. (2005). The *Mycobacterium tuberculosis* extracytoplasmic-function sigma factor SigL regulates polyketide synthases and secreted or membrane proteins and is required for virulence. *J. Bacteriol.* **187**(20), 7062-7071.

Hayes, C. S. and Sauer, R. T. (2003). Toxin-antitoxin pairs in bacteria: killers or stress regulators? *Cell* **112**(1), 2-4.

Honaker, R. W., Stewart, A., Schittone, S., Izzo, A., Klein, M. R., Voskuil, M. I. (2008). *Mycobacterium bovis* BCG vaccine strains lack narK2 and narX induction and exhibit altered phenotypes during dormancy. *Infect. Immun.* **76**(6), 2587-2593.

Istivan, T. S., Coloe, P. J. (2006). Phospholipase A in Gram-negative bacteria and its role in pathogenesis. *Microbiol.* **152**, 1263–1274.

Jackson, M., Crick, D. C. and Brennan, P. J. (2000). Phosphatidylinositol is an essential phospholipid of mycobacteria. *J. Biol. Chem.* **275**, 30092–30099.

Jain, S. K., Paul-Satyaseela, M., Lamichhane, G., Kim, K. S., Bishai, W. R. (2006). *Mycobacterium tuberculosis* invasion and traversal across an in vitro human blood-brain barrier as a pathogenic mechanism for central nervous system tuberculosis. *J. Infect. Dis.* **193**(9), 1287-1295.

Jones, G. J., Gordon, S. V., Hewinson, R. G., Vordermeier, H. M. (2010). Screening of Predicted Secreted Antigens from *Mycobacterium bovis* Reveals the Immunodominance of the ESAT-6 Protein Family. *Infect. Immun.* **78**, 1326-1332.

Joshi, S. A., Ball, D. A., Sun, M. G., Carlsson, F., Watkins, B. Y., Aggarwal, N., McCracken, J. M., Huynh, K. K., Brown, E. J. (2012). EccA1, a component of the *Mycobacterium marinum* ESX-1 protein virulence factor secretion pathway, regulates mycolic acid lipid synthesis. *Chem. Biol.* **19**(3), 372-380.

Klemm, P., Schembri, M. A. (2000). Bacterial adhesins: function and structure. *Int. J. Med. Microbiol.* **290**, 27–35.

Kolattukudy, P. E. (2001). Polyesters in higher plants. *Adv. Biochem. Eng. Biotechnol.* **71**, 1-49.

Korduláková, J., Gilleron, M., Mikusova, K., Puzo, G., Brennan, P. J., Gicquel, B., Jackson, M. (2002). Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis. PimA is essential for growth of mycobacteria. *J. Biol. Chem.* **277**(35), 31335-31344.

Kowalczykowski, S. C. (1994). In vitro reconstitution of homologous recombination reactions. *Experientia* **50**, 204–215.

- Lang, D., Hubrich, A., Dohle, F., Terstesse, M., Saleh, H., Schmidt, M., Pauels, H. G., Heidenreich, S.** (2000). Differential expression of heat shock protein 70 (hsp70) in human monocytes rendered apoptotic by IL-4 or serum deprivation. *J. Leukoc. Biol.* **68**(5), 729-736.
- Layre, E., Paepe, D. C., Larrouy-Maumus, G., Vaubourgeix, J., Mundayoor, S., Lindner, B., Puzo, G., Gilleron, M.** (2011). Deciphering sulfoglycolipids of *Mycobacterium tuberculosis*. *J. Lipid Res.* **52**(6), 1098-1110.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.** (2009). The sequence alignment/map format and SAM tools. *Bioinformatics* **25**, 2078–2079.
- Li, H.** (2011). Improving SNP discovery by base alignment quality. *Bioinformatics* **27**(8), 1157–1158.
- Liu, X., Gutacker, M. M., Musser, J. M., Fu, Y.** (2006) Evidence for Recombination in *Mycobacterium tuberculosis*. *J. Bacteriol.* **188**(23).
- Liu, C. F., Tonini, L., Malaga, W., Beau, M., Stella, A., Bouyssié, D., Jackson, M. C., Nigou, J., Puzo, G., Guilhot, C., Burlet-Schiltz, O., Rivière, M.** (2013). Bacterial protein-O-mannosylating enzyme is crucial for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **110**(16), 6560-6565.
- Liu, L., Huang, C., He, Z. G.** (2014). A TetR family transcriptional factor directly regulates the expression of a 3-methyladenine DNA glycosylase and physically interacts with the enzyme to stimulate its base excision activity in *Mycobacterium bovis* BCG. *J. Biol. Chem.* **289**, 9065-9075.
- Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M., and Barras, F.** (2003). Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J. Biol. Chem.* **278**: 38352–38359.
- Martin, R. G., and J. L. Rosner.** (2001). The AraC transcriptional activators. *Curr. Opin. Microbiol.* **4**, 132–137.
- Matange, N., Hunt, D. M., Buxton, R. S., Visweswariah, S. S.** (2013). Overexpression of the Rv0805 phosphodiesterase elicits a cAMP-independent transcriptional response. *Tuberculosis (Edinb).* **93**(5), 492-500.
- McGillivray, A., Golden, N. A., Gautam, U. S., Mehra, S., Kaushal, D.** (2014). The *Mycobacterium tuberculosis* Rv2745c plays an important role in responding to redox stress. *Plos One* **9**(4):e93604.

- Michel, A. L., Cooper, D., Jooste, J., de Klerk, L. M., Jolles, A.** (2011). Approaches towards optimising the gamma interferon assay for diagnosing *Mycobacterium bovis* infection in African buffalo (*Syncerus caffer*). *Prev. Vet. Med.* **98**, 142–151.
- Minnikin, D.E., Kremer, L., Dover, L.G., and Besra, G.S.** (2002). The methylbranched fortifications of *Mycobacterium tuberculosis*. *Chem. Biol.* **9**, 545–553.
- Mukhopadhyay, S. and Balaji, K. N.** (2011). The PE and PPE proteins of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **91**(5), 441-447.
- Narayan, A., Sachdeva, P., Sharma, K., Saini, A. K., Tyagi, A. K., Singh, Y.** (2007). Serine threonine protein kinases of mycobacterial genus : phylogeny to function. *Physiol. Genomics.* **29**, 66–75.
- Nguyen, L., Walburger, A., Houben, E., Koul, A., Muller, S., Morbitzer, M., Klebl, B., Ferrari, G., Pieters, J.** (2005). Role of protein kinase G in growth and glutamine metabolism of *Mycobacterium bovis* BCG. *J. Bacteriol.* **187**(16), 5852-5856.
- Nigou, J., Gilleron, M., Rojas, M., Garcia, L. F., Thurnher, M., and Puzo, G.** (2002). Mycobacterial lipoarabinomannans: modulators of dendritic cell function and the apoptotic response. *Microbes Infect.* **4**, 945–953.
- Nishibe, C., Canevari Castelão, A. B., Dalla Costa, R., Pinto, B. J., Varuzza, L., Cataldi, A. A., Bernardelli, A., Bigi, F., Blanco, F. C., Zumárraga, M. J., Almeida, N. F., Araújo, F. R.** (2013). Draft Genome Sequence of *Mycobacterium bovis* 04-303, a highly virulent strain from Argentina. *Genome Announc.* **27**:1(6).
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., Tyson, G. W.** (2014). Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, **25**: 1043-1055.
- Parsons, S. D., Drewe, J. A., Gey van Pittius, N. C., Warren, R. M., van Helden, P. D.** (2013). Novel cause of tuberculosis in meerkats, South Africa. *Emerg. Infect. Dis.* **19**, 2004-2007.
- Peirs, P., Lefevre, P., Boarbi, S., Wang, X. M., Denis, O., et al.** (2005). *Mycobacterium tuberculosis* with disruption in genes encoding the phosphate binding proteins PstS1 and PstS2 is deficient in phosphate uptake and demonstrates reduced in vivo virulence. *Infect. Immun.* **73**, 1898–1902.
- Pirson, C., Engel, R., Jones, G. J., Holder, T., Holst, O., Vordermeier, H. M.** (2015) Highly purified mycobacterial phosphatidylinositol mannosides drive cell-mediated responses and activate NKT cells in cattle. *Clin. Vaccine Immunol.* **22**, 178-184.

- Pollock, J. M. and Andersen, P.** (1997). Predominant recognition of the ESAT-6 protein in the first phase of interferon with *Mycobacterium bovis* in cattle. *Infect. Immun.* **65**, 2587-2592.
- Reddy, P. V., Puri, R. V., Chauhan, P., Kar, R., Rohilla, A., Khera, A., Tyagi, A. K.** (2013). Disruption of mycobactin biosynthesis leads to attenuation of *Mycobacterium tuberculosis* for growth and virulence. *J. Infect. Dis.* **208**(8), 1255-1265.
- Rengarajan, J., Bloom, B. R., Rubin, E. J.** (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. USA.* **102**(23), 8327-8332.
- Rousseau, C., Sirakova, T. D., Dubey, V. S., Bordat, Y., Kolattukudy, P. E., Gicquel, B., Jackson, M.** (2003). Virulence attenuation of two Mas-like polyketide synthase mutants of *Mycobacterium tuberculosis*. *Microbiol.* **149**, 1837-1847.
- Rutherford, R. M., Kehren, J., Staedtler, F., Chibout, S. D., Egan, J. J., Tamm, M., Gilmartin, J. J., Brutsche, M. H.** (2001). Functional genomics in sarcoidosis--reduced or increased apoptosis? *Swiss Med. Wkly.* **131**(31-32), 459-470.
- Sanchez, A., Espinosa, P., Esparza, M. A., Colon, M., Bernal, G., Mancilla, R.** (2009). *Mycobacterium tuberculosis* 38-kDa lipoprotein is apoptogenic for human monocyte-derived macrophages. *Scand. J. Immunol.* **69**, 20–28.
- Sasseti, C. M. & Rubin, E. J.** (2003). Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. USA* **100**, 12989–12994.
- Scherr, N., Müller, P., Perisa, D., Combaluzier, B., Jenö, P., Pieters, J.** (2009). Survival of pathogenic mycobacteria in macrophages is mediated through autophosphorylation of protein kinase G. *J. Bacteriol.* **191**(14), 4546-4554.
- Schneider, D., Bruton, C.J., and Chater, K.F.** (2000). Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor*. *Mol. Gen. Genet.* **263**, 543–553.
- Schue, M., Maurin, D., Dhouib, R., Bakala, N', Goma, J. C., Delorme, V., et al.** (2010). Two cutinase-like proteins secreted by *Mycobacterium tuberculosis* show very different lipolytic activities reflecting their physiological function. *FASEB J.* **24**, 1893–1903.
- Senaratne, R. H., De Silva, A. D., Williams, S. J., Mougous, J. D., Reader, J. R., Zhang, T., Chan, S., Sidders, B., Lee, D. H., Chan, J., Bertozzi, C. R., Riley, L.W.** (2006). 5'-Adenosinephosphosulphate reductase (CysH) protects *Mycobacterium*

tuberculosis against free radicals during chronic infection phase in mice. *Mol. Microbiol.* **59**(6), 1744-1753.

Sherrid, A. M., Rustad, T. R., Cangelosi, G. A., Sherman, D. R. (2010). Characterization of a Clp protease gene regulator and the reoxygenation response in *Mycobacterium tuberculosis*. *PLoS One* **5**(7):e11622.

Shimono, N., Morici, L., Casali, N., Cantrell, S., Sidders, B., Ehrt, S., Riley, L.W. (2003). Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the *mce1* operon. *Proc Natl Acad Sci USA* **100**: 15918– 15923.

Singh, V., Chandra, D., Srivastava, B. S., Srivastava, R. (2011). Downregulation of Rv0189c, encoding a dihydroxyacid dehydratase, affects growth of *Mycobacterium tuberculosis* in vitro and in mice. *Microbiol.* **157**(Pt 1), 38-46.

Singh, D. K., Singh, P. K., Tiwari, S., Singh, S. K., Kumari, R., Tripathi, D. K., Srivastava, K. K. (2014). Phosphorylation of pyruvate kinase A by protein kinase J leads to the altered growth and differential rate of intracellular survival of mycobacteria. *Appl. Microbiol. Biotechnol.* **98**(24), 10065-10076.

Sirakova, T. D., Dubey, V. S., Cynamon, M. H., Kolattukudy, P. E. (2003). Attenuation of *Mycobacterium tuberculosis* by disruption of a mas-like gene or a chalcone synthase-like gene, which causes deficiency in dimycocerosyl phthiocerol synthesis. *J. Bacteriol.* **185**, 2999–3008.

Smith, N. H., Gordon, S. V., de La Rúa-Domenech, R., Clifton-Hadley, R. S., Hewinson, G. (2006). Bottleneck and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nature* **4**, 670-681.

Smith, I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin. Microbiol. Rev.* **16**, 463-96.

Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, J., Liu, J., McDonald, K. L., Szyk, A., LaRonde-LeBlanc, N., Gao, L. Y. (2008). Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. *Infect. Immun.* **76**, 5478-5487.

Song, T., S. L. Dove, K. H. Lee, and R. N. Husson. (2003). RshA, an anti-sigma factor that regulates the activity of the mycobacterial stress response sigma factor SigH. *Mol. Microbiol.* **50**, 949–959.

Sorensen, A. L., Nagai, S., Houen, G., Andersen, P., Andersen, A. B. (1995). Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect. Immun.* **63**, 1710-1717.

Sreejit, G., Ahmed, A., Parveen, N., Jha, V., Valluri, V. L., Ghosh, S., Mukhopadhyay, S. (2014). The ESAT-6 protein of *Mycobacterium tuberculosis* interacts with beta-2-microglobulin (β 2M) affecting antigen presentation function of macrophage. PLoS Pathog. **10**(10):e1004446.

Stewart, G. R., Snewin, V. A., Walzl, G. (2001). Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. Nat. Med. **7**, 732-737.

Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Butcher, P. D., Young, D. B. (2002). The heat shock response of *Mycobacterium tuberculosis*: linking gene expression, immunology and pathogenesis. Comp. Funct. Genomics **3**, 348-351.

Strahl-Bolsinger, S., Gentsch, M., Tanner, W. (1999). Protein O-mannosylation. Biochim. Biophys. Acta. **1426**(2), 297–307.

Stridh, H., Planck, A., Gigliotti, D., Eklund, A., Grunewald, J. (2002). Apoptosis resistant bronchoalveolar lavage (BAL) fluid lymphocytes in sarcoidosis. Thorax. **57**(10), 897-901.

Stuart, J. K., Myszka, D. G., Joss, L., Mitchell, R. S., McDonald, S. M., Xie, Z., Takayama, S., Reed, J. C., Ely, K. R. (1998). Characterization of interactions between the anti-apoptotic protein BAG-1 and Hsc70 molecular chaperones. J. Biol. Chem. **273**(35), 22506-22514.

Székely, R. and Cole, S.T. (2016). Mechanistic insight into mycobacterial MmpL protein function. Mol. Microbiol. doi: 10.1111/mmi.13306.

Szulc-Kielbik, I., Brzezinska, M., Kielbik, M., Brzostek, A., Dziadek, J., Kania, K., Sulowska, Z., Krupa, A., Klink, M. (2015). *Mycobacterium tuberculosis* RecA is indispensable for inhibition of the mitogen-activated protein kinase-dependent bactericidal activity of THP-1-derived macrophages in vitro. FEBS J. **282**(7), 1289-1306.

Titball, R. W. (1998). Bacterial phospholipases. Symp. Ser. Soc. Appl. Microbiol. **27**, 127S-137S.

Tiwari, D., Singh, R. K., Goswami, K., Verma, S. K., Prakash, B., Nandicoori, V. K. (2009) Key residues in *Mycobacterium tuberculosis* protein kinase G play a role in regulating kinase activity and survival in the host. J. Biol. Chem. **284**(40): 27467-27479.

Tiwari, P., Arora, G., Singh, M., Kidwai, S., Narayan, O. P., Singh, R. (2015). MazF ribonucleases promote *Mycobacterium tuberculosis* drug tolerance and virulence in guinea pigs. Nat. Commun. **22**;6:6059. doi: 10.1038/ncomms7059.

- Van Embden, J. D., van Soolingen, D., Small, P. M., Hermans, P. W.** (1992). Genetic markers for the epidemiology of tuberculosis. *Res. Microbiol.* **143**(4), 385-391.
- VanderVen, B. C., Harder, J. D., Crick, D. C., Belisle, J. T.** (2005). Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science* **309**(5736), 941–943.
- Van Ingen, J., Rahim, Z., Mulder, A., Boeree, M. J., Simeoni, R., Brosch, R., van Soolingen, D.** (2012). Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerg. Inf. Dis.* **18**(4), 653-655.
- Vordermeier, H. M., Whelan, A., Cockle, P. J., Farrant, L., Palmer, N., Hewinson, R. G.** (2001). Use of synthetic peptides derived from the antigens ESAT-6 and CFP-10 for differential diagnosis of bovine tuberculosis in cattle. *Clin. Diagn. Lab. Immunol.* **8**, 571-578.
- Wang, Y., Kelly, C. G., Singh, M., McGowan, E. G., Carrara, A. S., Bergmeier, L. A., Lehner, T.** (2002). Stimulation of Th1-polarizing cytokines, C-C chemokines, maturation of dendritic cells, and adjuvant function by the peptide binding fragment of heat shock protein 70. *J. Immunol.* **169**(5), 2422-2429.
- Waters, W. R., Nonnecke, B. J., Palmer, M. V., Robbe-Austermann, S., Bannantine, J. P., Stabel, J. R., Whipple, D. L., Payeur, J. B., Estes, D. M., Pitzer, J. E., Minion, F. C.** (2004). Use of Recombinant ESAT-6:CFP-10 Fusion Protein for Differentiation of Infections of Cattle by *Mycobacterium bovis* and *M. avium* subsp. *paratuberculosis*. *Clin. Diag. Lab. Immun.* **11**, 729-735.
- Wobeser, G.** (2009). Bovine tuberculosis in Canadian wildlife: an updated history. *Can. Vet. J.* **50**, 1169–1176.
- Xaus, J., Besalduch, N., Comalada, M., Marcoval, J., Pujol, R., Mañá, J., Celada, A.** (2003). High expression of p21 Waf1 in sarcoid granulomas: a putative role for long-lasting inflammation. *J. Leukoc. Biol.* **74**(2), 295-301.
- Yang, M., Gao, C., Cui, T., An, J., He, Z. G.** (2012). A TetR-like regulator broadly affects the expressions of diverse genes in *Mycobacterium smegmatis*. *Nucleic Acids Res.* **40**(3), 1009-1020.
- Zhang, W., Jiang, H., Bai, Y. L., Kang, J., Xu, Z. K., Wang, L. M.** (2014). Construction and immunogenicity of the DNA vaccine of *Mycobacterium Tuberculosis* dormancy antigen rv1733c. *Scand. J. Immunol.* **79**(5), 292-298.
- Zhao, Q. J., Xie, J. P.** (2011). *Mycobacterium tuberculosis* proteases and implications for new antibiotics against tuberculosis. *Crit. Rev. Eukaryot. Gene Expr.* **21**, 347-61.

Zhao, Q., Li, W., Chen, T., He, Y., Deng, W., Luo, H., Xie, J. (2014). *Mycobacterium tuberculosis* serine protease Rv3668c can manipulate the host-pathogen interaction via Erk-NF- κ B axis-mediated cytokine differential expression. *J. Interferon Cytokine Res.* **34**(9), 686-698.

Genoma	Bases Ambíguas	Contigs	Genes Preditos	Scaffolds	Coding Density	CG	Tamanho do Genoma	Contig mais longo	Média do tamanho dos Contigs	N50
<i>M.bovis</i> ARG_04-303	0	169	4,158	169	0.905	0.655	4,313,075	329,090	25,521.15	85,558
<i>M.bovis</i> ARG_534	0	72	4,083	72	0.905	0.655	4,325,986	333,163	60,083.13	178,004

Tabela 2. Dados do sequenciamento genômico dos isolados *M. bovis* 04-303 e *M. bovis* 534 utilizando o programa CheckM.

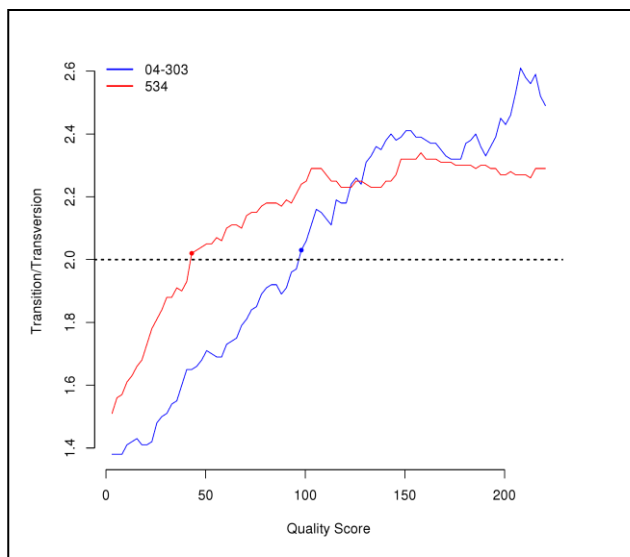


Figura 2. Relação de transições por transversões para todas as qualidades de escores gerados na variante chamando para ambos genomas 04-303 e 534. Para cada conjunto, o ponto corresponde ao valor da relação mais próximo de 2,0 (eixo Y), que determina o limite da pontuação de qualidade (eixo X).

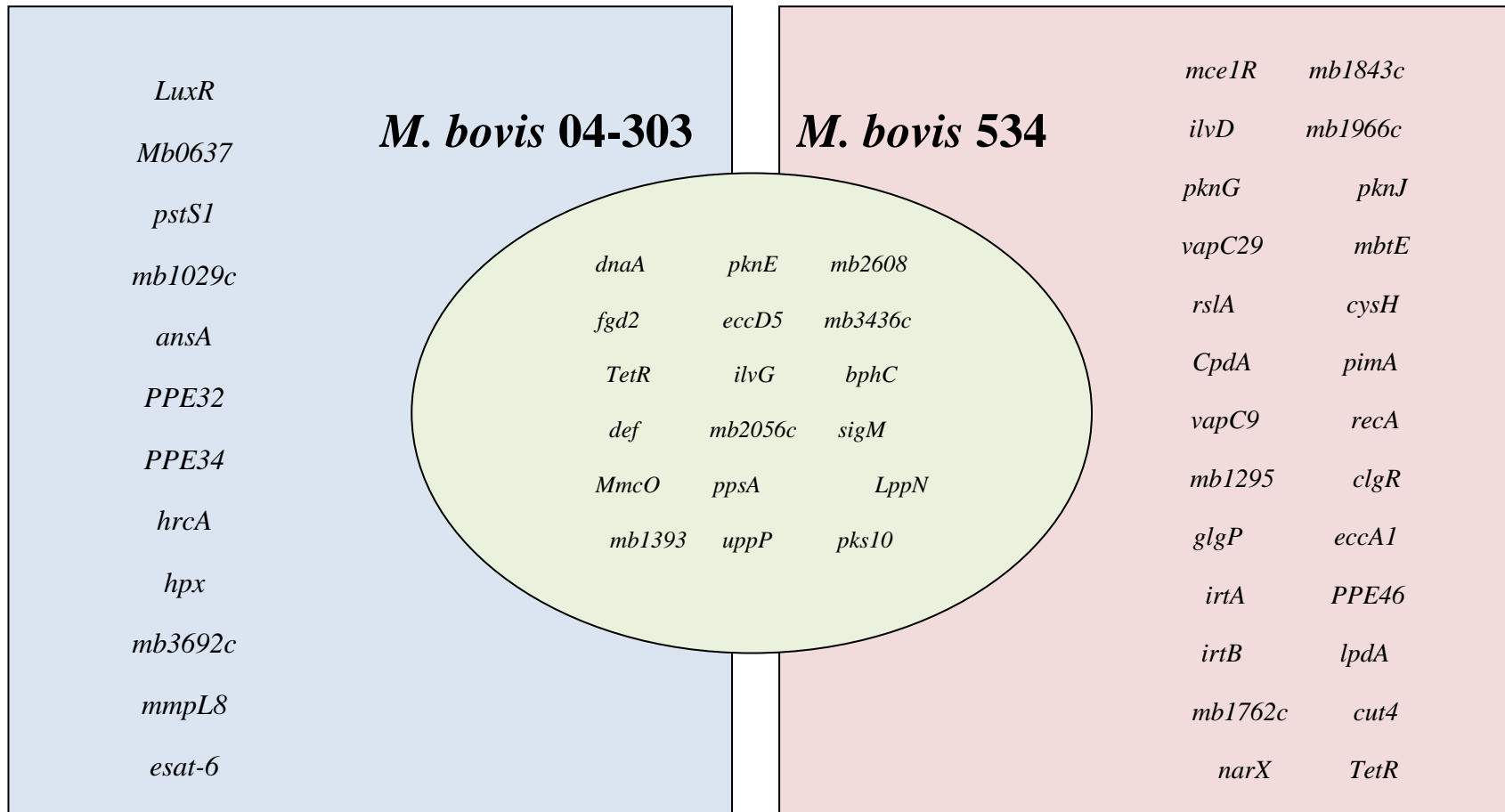


Figura 1. Diagrama de Venn ilustrando genes relacionados à virulência que apresentaram mutações (idênticas ou específicas) nos genomas dos isolados *M. bovis* 04-303 e 534 utilizando AF2122/97 como genoma referência.

Gene	AF2122/97	Proteína	Função	<i>M. bovis</i> 04-303				<i>M. bovis</i> 534				Referência
				Posição	Mutação	AA	Tipo	Posição	Mutação	AA	Tipo	
Rv0001	Mb0001	dnaA	chromosome replication initiator DnaA	1057	A-G	Ile-Val	NSMS	1057	A-G	Ile-Val	NSMS	Masiewicz et al. 2014
Rv0132c	Mb0137c	fgd2	F420-dependent glucose-6-phosphate dehydrogenase	160535	T-C	Lys-Glu	NSMS	160535	T-C	Lys-Glu	NSMS	Purwantini et al. 2013
Rv0165c	Mb0170c	mce1R	transcriptional regulator Mce1R	-	-	-	-	194983	G-	NE	DEL	Uchida et al. 2007
Rv0189c	Mb0195c	ilvD	dihydroxy-acid dehydratase	-	-	-	-	221361	C-	NE	DEL	Singh et al. 2011
Rv0195	Mb0201	LuxR	LuxR family transcriptional regulator	231486	G-	NE	DEL	-	-	-	-	Fang et al. 2013
Rv0302	Mb0310	TetR	TetR/ACRR family transcriptional regulator	366022	T-C	Val-Ala	NSMS	366022	T-C	Val-Ala	NSMS	Chou et al. 2015
Rv0410c	Mb0418c	pknG	serine/threonine-protein kinase G	-	-	-	-	497411	C-T	Gly-Glu	NSMS	Kirksey et al. 2011
Rv0429c	Mb0437c	def	peptide deformylase	519412	T-C	Thr-Ala	NSMS	519412	T-C	Thr-Ala	NSMS	Narayanan et al. 2011
Rv0617	Mb0634	vapC29	ribonuclease VapC29	-	-	-	-	712319	T-C	Val-Met	NSMS	Ahidjo et al. 2011
Rv0621	Mb0637	mb0637	membrane protein	716130	G-	NE	DEL	-	-	-	-	Dan et al. 2008
Rv0736	Mb0757	rslA	anti-sigma-L factor RslA	-	-	-	-	830473	C-	NE	DEL	Hahn et al. 2005
Rv0805	Mb0828	CpdA	3',5'-cyclic adenosine monophosphate phosphodiester	-	-	-	-	900385	G-	NE	DEL	Matange et al. 2014
Rv0846c	Mb0869c	MmcO	multicopper oxidase	943719	G-A	His-Tyr	NSMS	943719	G-A	His-Tyr	NSMS	Rowland et al. 2013
Rv0934	Mb0959	pstS1	periplasmic phosphate-binding lipoprotein PSTS1	1043504	G-	NE	DEL	-	-	-	-	Esparza et al. 2015
Rv0960	Mb0985	vapC9	ribonuclease VapC9	-	-	-	-	1074361	C-	NE	DEL	Jain et al. 2006
Rv1002c	Mb1029c	mb1029c	dolichyl-phosphate-mannose protein mannosyltransfe	1120062	G-	NE	DEL	-	-	-	-	Liu et al. 2003
Rv1264	Mb1295	mb1295	adenyl cyclase	-	-	-	-	1411239	.-AG	NE	INS	Dittrich et al. 2006
Rv1328	Mb1363	glgP	glycogen phosphorylase GlgP	-	-	-	-	1495014	A-C	Met-Leu	NSMS	Ryoo et al. 2007
Rv1348	Mb1383	irtA	Drugs-transport transmembrane ATP-binding protein A	-	-	-	-	1512766	G-T	Arg-Leu	NSMS	Rodriguez et al. 2006
Rv1349	Mb1384	irtB	drugs-transport transmembrane ATP-binding protein A	-	-	-	-	1515055	C-	NE	DEL	Farhana et al. 2008
Rv1358	Mb1393	mb1393	transcriptional regulator	1524738	T-G	Val-Gly	NSMS	1524738	T-G	Val-Gly	NSMS	Santos et al. 2015
Rv1358	Mb1393	mb1393	transcriptional regulator	1526476	G-	NE	DEL	1526476	G-	NE	DEL	Santos et al. 2015
Rv1538c	Mb1565c	ansA	L-aporaginase ansA	1725323	T-C	Met-Val	NSMS	-	-	-	-	Cai et al. 2012
Rv1660	Mb1688	pkS10	Chalcone SMthase	-	-	-	-	1860533	A-G	Ser-Gly	NSMS	Sirakova et al.2003
Rv1660	Mb1688	pkS10	Chalcone SMthase	-	-	-	-	1859643	G-	NE	DEL	Sirakova et al.2003
Rv1660	Mb1688	pkS10	Chalcone SMthase	1859795	C-	NE	DEL	1859795	C-	NE	DEL	Sirakova et al.2003
Rv1733c	Mb1762c	mb1762c	transmembrane protein	-	-	-	-	1945541	G-	NE	DEL	Coppola et al. 2015
Rv1736c	Mb1765c	narX	nitrate reductase-like protein NarX	-	-	-	-	1949403	C-T	Gly-Arg	NSMS	Honaker et al. 2008

Tabela 1. (parte 1). Genes relacionados à virulência de *M. bovis* que apresentaram mutações nos genomas dos isolados *M. bovis* 04-303 e 534 utilizando AF2122/97 como genoma referência (NSMS – *non-synonymous missense*, DEL – deleção, INS – inserção, NE – não se aplica).

Gene	AF2122/97	Proteína	Função	<i>M. bovis</i> 04-303				<i>M. bovis</i> 534				Referência
				Posição	Mutação	AA	Tipo	Posição	Mutação	AA	Tipo	
Rv1743	Mb1772	pknE	Ser/Thr protein kinase	1954568	A-T	Asp-Val	NSMS	1954568	A-T	Asp-Val	NSMS	Jayakumar et al. 2008
Rv1795	Mb1823	eccD5	ESX conserved component EccD5. ESX-5 type VII secret	2023460	G-A	Ala-Thr	NSMS	2023460	G-A	Ala-Thr	NSMS	Bottai et al. 2012
Rv1808	Mb1837	PPE32	PPE family protein	2041061	G-A	Trp-	NSMS	-	-	-	-	Deng et al. 2014
Rv1813c	Mb1843c	mb1843c	hypothetical protein	-	-	-	-	2046255	C-T	Ala-Thr	NSMS	Bretl et al. 2012
Rv1820	Mb1851	ilvG	acetolactate synthase large subunit IlvG	2056269	G-	NE	DEL	2056269	G-	NE	DEL	Singh et al. 2011
Rv1917c	Mb1951c	PPE34	PPE family protein	2154292	C-T	Gly-Arg	NSMS	-	-	-	-	Bansal et al. 2010
Rv1931c	Mb1966c	mb1966c	transcriptional regulator	-	-	-	-	2174242	C-	NE	DEL	Frota et al. 2004
Rv2030c	Mb2056c	mb2056c	hypothetical protein	2260638	-C	NE	INS	2260638	-C	NE	INS	Wieczorek et al. 2014
Rv2931	Mb2074c	ppsA	phthiocerol synthesis polyketide synthase type I	8825, 2278	G-C, G-T,	p-Glu, His	NSMS	8825, 2278	G-C, G-T,	p-Glu, His	NSMS	Bisson et al. 2012
Rv2088	Mb2115	pknJ	transmembrane serine/threonine-protein kinase PknJ	-	-	-	-	2327145	C-	NE	DEL	Singh et al. 2014
Rv2136c	Mb2160c	uppP	Undecaprenyl pyrophosphate phosphatase	2377026	G-T	Val-Leu	NSMS	2377026	G-T	Val-Leu	NSMS	Vandal et al. 2009
Rv2270	Mb2293	LppN	lipoprotein LppN	2524106	T-C	Leu-Pro	NSMS	2524106	T-C	Leu-Pro	NSMS	Ocampo et al. 2014
Rv2374c	Mb2395c	hrcA	Probable heat shock protein transcriptional repressor H	2622933	G-A	Asp-Asn	NSMS	-	-	-	-	Stewart et al. 2002
Rv2380c	Mb2401c	mbtE	peptide synthetase	-	-	-	-	2630194	G-	NE	DEL	Reddy et al. 2013
Rv2380c	Mb2401c	mbtE	peptide synthetase	-	-	-	-	2634471	C-	NE	DEL	Reddy et al. 2013
Rv2392	Mb2413	cysH	phosphoadenosine phosphosulfate reductase	-	-	-	-	2654933	G-	NE	DEL	Senaratne et al. 2006
Rv2577	Mb2608	mb2608	hypothetical protein	2869753	C-	NE	DEL	2869753	C-	NE	DEL	Blanco et al. 2009
Rv2610c	Mb2642c	pimA	alpha-(1-2)-phosphatidylinositol mannosyltransferase	-	-	-	-	2905671	G-	NE	DEL	Kordula'kova' et al. 2002
Rv2737c	Mb2756c	recA	recombinase A	-	-	-	-	3007225	GG-..	NE	DEL	Szulc-Kielbik et al. 2015
Rv2745c	Mb2765	clgR	Transcriptional regulatory protein ClgR	-	-	-	-	3014181	C-A	Gln-Lys	NSMS	Cox et al. 1999
Rv3018c	Mb3043c	PPE46	PPE family protein	-	-	-	-	3333886	C-G	Val-Leu	NSMS	Chaitra et al. 2008
Rv3171c	Mb3196c	hpx	Possible non-heme haloperoxidase Hpx	3497977	A-G	Asp-Gly	NSMS	-	-	-	-	Murphy et al. 2005
Rv3249c	Mb3277c	TetR	Possible transcriptional regulatory protein (probably T	-	-	-	-	3584552	T-C	Ser-Pro	NSMS	Griffin et al. 2011
Rv3303c	Mb3331c	lpdA	NAD(P)H quinone reductase LpdA	-	-	-	-	3645254	C-	NE	DEL	Akhtar et al. 2006
Rv3402c	Mb3436c	mb3436c	hypothetical protein	3774060	A_G	Cys-Arg	NSMS	3774060	A_G	Cys-Arg	NSMS	Li et al. 2014
Rv3452	Mb3482	cut4	cutinase	-	-	-	-	3821778	G-	NE	DEL	Dedieu et al. 2013
Rv3568c	Mb3599c	bphC	Biphenyl-2,3-diol 1,2-dioxygenase BPHC	3952885	G-A	Arg-Gln	NSMM	3952885	G-A	Arg-Gln	NSMS	Flores-Valdez et al. 2009
Rv3668c	Mb3692c	mb3692c	protease	4047766	G-	NE	DEL	-	-	-	-	Quanju Zhao et al. 2014
Rv3823c	Mb3853c	mmpL8	Conserved integral membrane transport protein MmpL	4226452	C-G	Leu-Val	NSMS	-	-	-	-	Wards et al 2000
Rv3868	Mb3898	eccA1	ESX-1 secretion system protein EccA1	-	-	-	-	4279990	C-	NE	DEL	Joshi et al. 2012
Rv3875	Mb3905	esat-6	6 kDa early secretory antigenic target EsxA (ESAT-6)	4289115	A-G	Thr-Ala	NSMS	-	-	-	-	de Jonge et al. 2007
Rv3911	Mb3941	sigM	RNA polymerase sigma factor SigM	4315, 4334	A-G, T-G	Ala, Ser-	NSMS	4334624	T-G	Ser-Ala	NSMS	Sahadevan Raman et al. 2006

Tabela 1. (parte 2). Genes relacionados à virulência que apresentaram mutações no genoma dos isolados *M. bovis* 04-303 e 534 utilizando AF2122/97 como genoma referência (NSMS – *non-synonymous missense*, DEL – deleção, INS – inserção, NE – não se aplica).

ARTIGO 4

QUANDO SUBSTITUIÇÕES DE DNA SÃO POSITIVAMENTE ENGANOSAS: PAN-GENOMA E *INDELS* AO RESGATE PARA FILOGENÔMICA E DATAÇÃO DE DIVERGÊNCIAS EM *MYCOBACTERIUM BOVIS*

Artigo para submissão

BMC Genomics

**QUANDO SUBSTITUIÇÕES DE DNA SÃO POSITIVAMENTE ENGANOSAS: PAN-
GENOMA E *INDELS* AO RESGATE PARA FILOGENÔMICA E DATAÇÃO DE
DIVERGÊNCIAS EM *MYCOBACTERIUM BOVIS***

José Patané¹

Ana Beatriz Canevari Castelão²

Joaquim Martins Junior¹

Martín José Zumárraga³

Fabiana Bigi³

Angel Cataldi³

Tyler Thacker⁴

Luciana Montera⁵

Christiane Nishibe⁵

Nalvo Franco Almeida⁵

Flávio Ribeiro de Araújo⁶

João Carlos Setubal¹

Endereço institucional:

¹Instituto de Bioquímica, Universidade de São Paulo, São Paulo, São Paulo, Brasil.

²Faculdade de Medicina Veterinária e Zootecnia, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brasil.

³Instituto de Biotecnología, CICVyA/INTA, Los Reseros y las Cabañas, 1712 Castelar, Argentina.

⁴ Departamento de Agricultura dos Estados Unidos, NADC USDA, Ames, Iowa, Estados Unidos.

⁵ Faculdade de Computação, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brasil.

⁶ Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brasil.

Autores correspondentes: joão.c.setubal@gmail.com e flabio.araujo@embrapa.br

Resumo: *Mycobacterium bovis* é um importante agente patogênico animal em escala global, infectando animais vertebrados domésticos e selvagens, bem como os seres humanos. Neste trabalho investigamos os padrões e processos evolutivos de 38 genomas sequenciados de *M. bovis* das Américas, Europa, Ásia e África. Observamos que os isolados americanos estão evoluindo distintamente (embora com baixa magnitude) da maioria das linhagens da América do Sul em relação ao tamanho do *core coding*, conteúdo CG, e *codon usage bias*, e empregamos uma nova definição de acordo com esse padrão, como proxy para confiabilidade filogenética. Três tipos de dados foram utilizados para inferência desta árvore: 1. cópia única de genes do *core coding*; 2. blocos *INDEL* codificados como caracteres numéricos (*Indels*); 3. presença/ausência de genes codificados semelhantemente (pan-genoma). O desempenho do *core coding* foi abaixo do esperado considerando que usamos mais de 2.000 genes, porém *indels* e pan-genoma estimaram individualmente árvores razoáveis, com taxas de transferência horizontal de genes elevadas, como 200x a taxa de substituição de DNA. Ao questionarmos o porque do fraco desempenho do *core coding*, excluímos vieses de metodologia filogenética, recombinação homóloga e seleção positiva e purificadora como possíveis causas, enquanto consideramos retenção transitória de mutações deletérias levemente aleatórias devido à diversificação recente, e também pseudogeneização em curso, ambas podendo ser plausíveis. Embora o índice de retenção tenha sido tão bom quanto os dados de presença/ausência individualmente identificados nas amostras dos EUA, adicionalmente identificamos um

clado sul-americano com a maioria das linhagens excluindo apenas duas amostras brasileiras. Inferimos datação de divergência utilizando *indels* associado com pan-genoma devido à falta de informações do *core coding*, recuperando estimativas de dados empregados em estudos anteriores sobre a origem de *M. bovis* (aproximadamente 10.000 anos), independentemente de outros dados (tempo estrito vs. tempo não correlacionado, uniforme vs. exponencial, diferença de idade/taxa padrão de desvio), evidenciando a adequação e robustez insuspeita de utilizar estritamente dados genômicos/ausência de marcação de tempo, enquanto ao mesmo tempo, decisivamente rejeitar um rígido marco para nosso conjunto de dados. Todas as árvores suportam uma posição relativamente basal de linhagens africanas, isoladas de *Homo sapiens*, sugerindo que esta foi uma importante região para a diversificação e confirmando que fomos um dos primeiros anfitriões desta espécie. A evolução de isolados dos EUA foi afetada pela aquisição lateral de 57 genes exclusivos, incluindo genes importantes para a virulência. Um clado de três localidades dos EUA, geograficamente distantes, está sob influência de taxas ainda mais rápidas evolutivas suportados pelo *core coding*, *indels* e pan-genoma.

Introdução

A tuberculose bovina é uma doença significativa mundial de bovinos. *Mycobacterium bovis*, o agente causador, é também um organismo patogênico para seres humanos e vários outros animais domésticos e selvagens (Thoen et al., 2006). *M. bovis* pertence ao complexo *Mycobacterium tuberculosis* (CMT), que consiste, tradicionalmente, de *M. tuberculosis*, o principal agente da tuberculose humana, *M. africanum*, *M. canettii*, *M. microti*, *M. bovis*, *M. caprae*, *M. pinnipedii*, *M. mungii*, *M. suricatte* e *M. orygis* (Fabre et al., 2004; Smith et al., 2006; Alexander et al., 2010; Van Ingen et al., 2012). Todas estas espécies são geneticamente muito semelhantes (Forrellad et al., 2013). Tem sido sugerido que os membros do CMT evoluíram a partir de um ancestral comum através de sucessivas deleções/inserções resultantes na

presente especiação de *Mycobacterium* e suas diferenças de patogenicidade (Fabre et al., 2004).

A fim de analisar a evolução da infecção por *M. bovis* no espaço e no tempo, é essencial estudar a diferenciação genética de isolados e suas relações filogenéticas, em conjunto com o movimento do rebanho e ecologia dos animais selvagens. Uma compreensão mais clara da epidemiologia de *M. bovis* pode contribuir para um controle mais efetivo ou para a erradicação da doença. Dados epidemiológicos de alta qualidade aumentam a eficiência dos programas de controle da tuberculose bovina. Com novas ferramentas disponíveis de epidemiologia molecular, há a crescente possibilidade de se encontrar respostas para questões importantes como a transmissão desta enfermidades entre animais e fatores de risco associados, bem como o papel de animais silvestres como reservatórios desta doença (Hilty et al., 2005).

O uso de tecnologias de sequenciamento de genoma total (do inglês, *whole genome sequencing* - WGS) para estudar doenças infecciosas bacterianas resultou em avanços sem precedentes na capacidade de resolução de dados epidemiológicos em escala global (Harris et al. 2010), além de proporcionar novas perspectivas sobre os processos de replicação da micobactéria no interior do hospedeiro (Ford et al., 2011), e corroborar com importante identificação das cadeias de transmissão (Gardy et al., 2011). O WGS de muitos isolados da mesma espécie pode fornecer a potente estatística necessária para definir associações entre genótipos e fenótipos, virulência, patogênese, gama de hospedeiros e variabilidade antigênica.

A fim de correlacionar as bases genéticas e ambientais da tuberculose bovina, realizamos uma análise de genomas de 38 isolados de *M. bovis*, de diferentes hospedeiros (bovinos, javalis, alces e veados) de diferentes lugares do mundo, combinado com metadados clínicos (epidemiologia, *spoligotyping* e MIRU [unidades repetitivas de interespaçadores micobacterianos]) de alta qualidade. Assim, realizamos o primeiro estudo abordando aspectos da evolução do genoma de *M. bovis* em quatro continentes (África, Américas, Ásia, Europa). Objetivamos examinar a adequação de diferentes tipos de dados na reconstrução da filogenia de *M. bovis*, definir o tipo de

recombinação (homóloga ou HGT [transferência horizontal de genes]) presente e seus impactos, inferir a origem de *M. bovis* e discutir acerca de processos associados com a evolução de *M. bovis*.

Materiais e Métodos

Isolados bacterianos

Vinte e oito isolados de *M. bovis* foram sequenciados e incluídos neste estudo: 07 da Argentina, 10 do Brasil e 11 dos Estados Unidos (Tabela 1), isolados de bovinos e animais silvestres de diferentes estados destes três países, tentando ser representativos de diferentes localizações geográficas. As amostras foram descontaminadas e cultivadas em meios de cultura *Stonebrink* e incubadas a 37°C. A partir das colônias, 2 µg de DNA de cada isolado foram suspensos em 150 µl de água destilada e fervidos (100°C) durante 15 min.

Outros dez isolados sequenciados anteriormente e encontrados depositados no GenBank foram utilizados neste trabalho para efeito de comparações filogenéticas. Nove isolados de *M. bovis* dos seguintes países: Coréia do Norte (1), Coréia do Sul (1), Reino Unido (3), Uruguai (1), Uganda (2) e Mali (1) e um isolado de *Mycobacterium tuberculosis* (1) (Tabela 2).

Sequenciamento genômico

Todos os isolados de *M. bovis* neste estudo foram sequenciados utilizando plataforma Illumina (MiSeq), de acordo com orientações do fabricante. A qualidade dos *reads* foi avaliada utilizando o programa FastQC (versão 0.10.1) {disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>}, e a triagem efetuada utilizando o programa prinseq-lite (versão 0.20.3) {disponível em: <http://prinseq.sourceforge.net/>}. Os genomas foram montados tendo como referência o genoma *M. bovis* AF2122/97 (acesso. NC_002945) utilizando o programa *Bowtie* (disponível em: <http://bowtie-bio.sourceforge.net/index.shtml>) e anotados utilizando o programa NCBI POC (Boldt, 2003). Os genomas brasileiros e argentinos já encontram-

se depositados no GenBank. A genotipagem dos isolados foi realizada por *Spoligotyping* segundo Kamerbeek et al. (1997) e MIRU segundo Supply et al. (2006).

Inferência de ortólogos

A inferência das famílias dos homólogos foi realizada por meio do programa Get_Homologues v2.0 (Contreras-Moreira & Vinuesa 2013), utilizando o *script* get_homologues.pl ('-m') ativando o agrupamento OrthoMCL (Li et al., 2003) com valores padrão. Subsequentemente, genes estritamente ortólogos, cópias únicas dos genes do core-genoma (*core-coding*) foram estimadas pelo mesmo programa com o *script* compare_clusters.pl.

Alinhamento múltiplo

Alinhamentos individuais de genes do *core coding* foram realizados utilizando Guidance v2.0 (Sela et al. 2015) ancorado por códons usando MAFFT v7.0 (Kato & Standley 2013) com guia de 100 árvores pseudoreplicativas. A concatenação do alinhamento múltiplo foi realizada por meio do programa FasConCat (Kück & Meusemann 2010). A correção manual foi então realizada em Aliview (Larsson 2014), e sempre que regiões locais foram desalinhadas usamos a opção "realinhar bloco selecionado" e atualizávamos as informações sobre início e fim de cada gene. Informações sobre a quantidade de variáveis locais e parcimônias foram obtidas em Mega (Tamura et al., 2013). A fração de substituições não sinônimas relativa ao número total de substituições foi calculada como a razão entre o número de sítios de aminoácidos (AA) variáveis pelo número de sítios variáveis no DNA.

Padrões genômicos

Verificamos ao longo de todo genoma entre os diferentes indivíduos que alguns padrões poderiam servir como *proxy* para a confiabilidade filogenética. O primeiro parâmetro que consideramos foi o tamanho do genoma, uma vez que alguns genomas não estão 100% fechados, o que pode ocasionar artefatos. Para superar isso, utilizamos o alinhamento concatenado do *core coding*, assumindo que este é menos

sujeito a artefatos pela variação do número de *contigs*. Estimamos também dois outros parâmetros relacionados ao tamanho do *core coding*: conteúdo de CG e número efetivo de códons (Nc), utilizando o programa Dambe v5.0 (Xia 2013).

Indels e Pan-genoma

Baseando-se na importância de dados de presença/ausência de *M. bovis*, utilizamos-os como um conjunto de dados complementares aos polimorfismos de DNA. A matriz de *indels* foi obtida a partir de genes do *core coding* pela aplicação de códigos complexos disponíveis em SeqState (Muller 2005), que gera como resultado uma matriz de multi-estado de presença/ausência de blocos de *indels* contíguos. A matriz de presença/ausência do conjunto de genes do pan-genoma foi obtida executando novamente o *script* compare_clusters.pl adicionando a opção '-t 0', que inclui todas as famílias de genes; se um gene está presente em um dado genoma, mais de uma cópia pode estar presente (duplicações), e nestes casos estados > 1 foram investigados.

Inferências filogenômicas

Foram realizadas análises filogenéticas por máxima verossimilhança (ML), IQTree v1.3.12 (Nguyen et al. 2015), fixando *M. tuberculosis* H37Rv como *outgroup*. Cada conjunto de dados foi utilizado separadamente bem como em conjunto (*core coding*, *indels*, pan-genoma). Para testar o melhor esquema de particionamento para o *core-coding*, utilizamos subconjuntos das três posições dos códons, e em seguida estimamos o melhor esquema de particionamento entre eles e o respectivo modelo para cada partição em IQTree. Para testar a variabilidade de taxas ao longo do genoma, além de usar a distribuição típica gama (+G) e a probabilidade de sítios invariáveis (+I), também utilizamos modelos GTR com distribuições mistas para cada análise, de duas a cinco variáveis por modelo (+MM). *Indels* e pan-genoma também foram testadas para o melhor modelo multi-estado, incluindo as possibilidades de +G, +I, e +MM. Para testar os vieses do método filogenético, também inferimos a topologia do *core coding* sobre máxima parcimônia (MP) em Mega (Kumar et al., 2004).

A filogenia empregando os três conjuntos de dados em conjunto foi reconstruída, uma vez que supostamente tal análise poderia ser mais esclarecedora e robusta. A análise foi composta por cinco partidas (três por códon, *indels* e pan-genoma), a partir do melhor modelo anteriormente sugerido para cada uma. IQTree também definiu as taxas relativas entre as partições. Suporte para cada análise filogenética foi calculada a partir de 1000 *UFBoot* pseudo-replicativas (Minh et al., 2013). A fim de testarmos a concordância entre os dados filogenéticos e geográficos, primeiramente definimos o caráter "localização geográfica" atribuindo uma entre cinco opções para cada genoma analisado (América do Sul, América do Norte, Europa, Ásia e África). Posteriormente, calculamos o índice de retenção (RI)(PAUP v4b10 [Swofford 2013]) para este caráter em cada topologia estimada, com valores mais elevados indicando melhor correlação desta árvore com a localização geográfica.

Recombinação homóloga

Os dez maiores blocos colineares locais (LCBs) entre genomas foram obtidos pelo algoritmo Progressive Mauve (Darling et al., 2010). Foram utilizados os três testes disponíveis no pacote PHIPHACK (PHI, NSS, e MaxChi) para determinar a recombinação em cada um destes blocos (Bruen & Bruen 2005). Para sermos conservadores, assumiu-se um LCB livre de recombinação se nenhum dos três testes fossem significativos ($p=0,1$). Blocos em que recombinação não foi detectado foram concatenados e ML estimada em IQTree como descrito acima. Assumimos esta prerrogativa a fim de avaliar se a recombinação estava ou não influenciando as estimativas de ML do *core coding*, e para verificar se havia algum bloco com sinais de recombinação; embora LCBs não correspondam aos genes isolados do *core coding*. Como outra forma de avaliar o efeito da recombinação, executamos ClonalFrame ML (Didelot & Wilson 2015) com os LCBs fixando a combinação da topologia de evidência. Estimamos assim contribuições relativas entre recombinação homóloga e mutações pontuais entre os LCBs em todos os ramos da árvore.

Seleção positiva e Seleção purificadora

Utilizamos genes do *core coding* para testar seleção positiva. Pretendíamos utilizar FUBAR (Murrell et al., 2013) pacote Hyphy (Pond & Muse 2005) para estimar taxas de mutações não sinônimas por sinônimas (dN/dS), mas devido à pequena variabilidade existente entre os genomas de *M. bovis*, decidimos realizar esta análise a partir dos genes do *core coding* como um todo, para darmos maior sustentação ao teste; no entanto, mesmo assim, este software interrompia as análises. Assim, utilizamos o método de FEL (Pond & Geada 2005) em Hyphy, assumindo três classes de taxas diferentes para dN e outras três para dS, a fim de identificar e remover ainda mais genes selecionados positivamente que poderiam causar vieses filogenéticos. Também calculamos a topologia do *core coding* assumindo um modelo de códon no IQTree, avaliando o melhor modelo antes de realizar a estimativa ML e o suporte das ramificações, sendo uma segunda estratégia para superar os efeitos de seleção sem ter que remover genes selecionados de forma positiva.

Transferência horizontal de genes (HGT)

A fim de testar a HGT, avaliamos como taxas mais rápidas/mais lentas de *indels* e pan-genoma estão em *M. bovis* em comparação a substituições nucleotídicas. Além disso, avaliamos presença/ausência de sinapomórficos em clados relevantes e observamos se estes estariam agrupados ou espalhados ao longo dos genomas.

Datação de divergência

Tentamos estimar datação de divergência dos genes do *core coding*, porém três ensaios distintos não convergiram mesmo depois da geração de 1×10^9 MCMC em BEAST v1.8.2 (Drummond et al., 2012), independentemente dos dados ou taxas utilizados. No entanto, a datação de divergência baseada em *indels* + pan-genoma apresentou boa convergência em um ensaio piloto com taxa de tempo fixo em 1,0 (parâmetro strict.clock). As duas partidas apresentam cada uma (proporcionalmente) taxa relativa ponderada pelo número de colunas em cada conjunto de dados, como é

comumente efetuado para os dados de DNA regulares. Primeiramente conservadamente determinamos que a prévia deveria atingir taxas das mais diferentes magnitudes. Decidindo assim por taxas que abrangem 0,001-100 *state changes/site/branch/My* (sc/s/b/My) em seis ordens de magnitude, desde taxas de nucleotídeos mais lentos, até 100x as maiores taxas de nucleotídeos. Para uma idade máxima prévia sobre a origem de *M. bovis* utilizamos 10.000 anos (0,01 Mya), de acordo com Wirth et al. (2008), e um mínimo de 0,0 y.

Depois de decidir sobre estes intervalos de tempo e taxas, uma distribuição prévia específicas foi assumida. Iniciamos com uma corrida de tempo fixo (dois MCMC executados para cada prévia) assumindo para as taxas de tempo fixo=Uniform[0.001, 100.0]sc/s/b/My e Lognormal (0,00404; 0,7) e para hora de origem (95% máximo flexível = 10000 y) que é um lognormal enviesada, com média no espaço real. Também testamos prévia taxas exponenciais dentro deste mesmo intervalo, uma vez que as prévias uniformes em geral apresentam média perto de metade da largura, o que, portanto, pode inclinar para a direita e não provar adequadamente baixas taxas. Então também implementamos uma lognormal não correlacionada prévia da taxa de distribuição para avaliar a variabilidade da frequência da taxa de variabilidade de ramificação através da própria ramificação (distribuição ucl), com média ucl uniforme [0.001, 100.0] ou exponencial (espelhamento da análise de tempo fixo) e ucl.stdev seguindo BEASTv1.8.2 padrão anterior. Testamos ainda uma média ucl uniforme com Lognormal de (0.006387; 0,3) para o tempo, com o mesmo limite superior, porém com uma distribuição lognormal menos assimétrica. Finalmente avaliamos uma média ucl uniforme com exponencial (10,0) para ucl.stdev ao invés do padrão BEAST antes aplicado. Corridas com diferentes antecedentes foram comparados pelo critério AICM e verificando também se houve ou não sobreposição entre os diferentes parâmetros (distribuições não-sobrepostas podem ser interpretada como sendo significativamente diferentes em probabilidade). O cronograma final foi obtido pelo log de combinação de corridas com menor AICM em LogCombiner com remoção de *burner*, e a árvore final calculada pelo TreeAnnotator. A convergência das corridas, tamanhos efetivo das

amostras (ESSS), e 95% HPD (maior densidade posterior) dos parâmetros foram avaliados em v1.6 Tracer.

Resultados e Discussão

Core-genoma e Pan-genoma

Um total de 6.189 famílias de homólogos (pan-genoma) foi obtido por Get_Homologues. Entre esses, 2.550 foram inferidas como genes de única cópia do *core-coding* e, posteriormente, alinhados multiplamente (sequências de nucleotídeos) por Guidance (usando Mafft), que descartou 67 delas por problemas de alinhamento e/ou por conter sequências que não eram múltiplas de três. E assim, um total de 2.483 grupos de *core coding* foram validados para análises subsequentes.

Alinhamento múltiplo

O alinhamento múltiplo final dos genes do *core coding* apresentou 2.491.902 pb, com 6.707 sítios variáveis (0,2%) e 1.761 sítios informativos parcimônias (0,07%). Quando o alinhamento foi traduzido para aminoácidos, 3.847 sítios variáveis (0,46%) e 1.134 sítios informativos parcimônias (0,14%) foram detectados. A proporção de sítios variáveis de aminoácidos por sítios variáveis de DNA foi de 0,57 (ou seja, 57% dos sítios variáveis de DNA incluem substituições não sinônimas, uma taxa relativamente elevada de acumulação não sinônima).

Tendências genômicas

Foi observada em baixa magnitude, entretanto ainda significativa, diferenças no tamanho do core, conteúdo de CG e Nc (inversamente proporcional ao *codon usage bias*) entre as amostras dos EUA quando comparadas a amostras de outras regiões do mundo (Figura 1). A correlação positiva entre o conteúdo CG e a polarização do *codon usage bias* (portanto correlação inversa ao Nc; Figura 1) é comumente observada em muitas bactérias (Chen et al., 2004; Li et al., 2015.). Contrariamente, a relação inversa entre o tamanho do *core coding* e o conteúdo CG seguiu o padrão oposto típico de

Actinobactérias (filó ao qual pertence as micobactérias, e conseqüentemente *M. bovis*), que geralmente apresenta correlação positiva entre estes dois parâmetros (Li & Du 2014).

Uma possível explicação, embora com pouca evidência em bactérias (Venton 2013), seria um provável diferencial na velocidade de seleção de replicação: por algum motivo, as amostras não-EUA podem estar em um processo de seleção mais intenso, o que pode resultar em menores traduções, e assim, em genomas replicantes mais rápidos e melhores adaptados, os quais devem beneficiar o acoplamento rigoroso entre o códon mais comum para cada um dos aminoácidos e o seu respectivo RNAt para taxas mais rápidas (como descrito por Sharp et al., 2010). No outro extremo, os isolados dos EUA podem apresentar um processo de seleção menos intenso, ou seja menor pressão para ambas viéses dos códons (ou seja, utilizando, em média, mais códons por aminoácido do que o tolerado), e menor necessidade de ir contra a decadência de CG devido à impulsão bioquímica da pressão de mutação AT (Hildebrand et al., 2010) (novamente, não sem exceções).

Assim estes padrões indicam que as características genômicas estão correlacionados com a amostragem geográfica, tornando-se razoável assumir que os “padrões” são resultados de uma história evolutiva comum (como em: Mostowy et al., 2005). Por isso, foi utilizado o local de amostragem como critério de confiabilidade na inferência da árvore de *M. bovis*. Para quantificar a concordância foi utilizado o índice de retenção, que avalia o quanto informativo é um padrão específico (continente de origem, no nosso caso) no suporte aos clados inferidos filogeneticamente. Valores próximos a 1,0 indicaram uma boa concordância filogenética, enquanto valores próximos de zero podem indicar algum tipo de viés na análise. Os cálculos foram realizados com PAUP v4b10.

Filogenômica

Teste para o melhor esquema de partição foi efetuado para os cinco subconjuntos assumido a priori (três posições de códons, *indels*, pan-genoma). O

particionamento escolhido foi uma das cinco partições diferentes, com o melhor modelo para cada calculado concomitantemente por IQTree: K3Pu + I + G4 (1º códon), TVM + I + G4 (2º códon), TVM + I (3º códon), ordenados + FQ + ASC + R4 (*indels*), GTR2 + FO + R2 (pan-genoma), e as taxas relativas entre as cinco partições foram, respectivamente, 0,439, 0,471, 0,585, 122,196, 160,847. Portanto, a taxa média de dados de presença/ausência da taxa evolutiva de nucleotídeos foi de 189,33, quase que duas ordens de magnitude maior do que a taxa de substituições de DNA. Estes dados suportam dados de estudos prévios que sugerem a perda de muitas regiões genômicas típicas de *M. bovis* em relação a outros isolados do CMT (Mostowy et al., 2005).

Entre os parâmetros *core coding*, *indels*, e pan-genoma, o primeiro apresentou pior desempenho filogenético (Tabela 2; Figura 2), ainda que reuniu amostras da Coréia e do Reino Unido, este padrão não foi recuperado pelo restante das análises filogenéticas. A árvore filogenética baseada na evidência total (*core-coding* + *indels* + pan-genoma) obteve resultados semelhantes às árvores apresentadas pelos dados de *indels* e pan-genoma separadamente, de acordo com RI. No entanto, percebe-se que a árvore de evidência total resgata tanto um clado dos Estados Unidos como um clado mais inclusivo da América do Sul que não está presente nas árvores remanescentes (com a exceção de dois isolados brasileiros e um isolado uruguaio). E assim, mais uma vez destacamos a importância da utilização de diferentes fontes de dados para inferências de relações filogenéticas. O pior desempenho do *core coding* é um fenômeno bastante interessante, uma vez que analisamos genomas completos (o que supostamente, permitiria que mesmo baixas taxas de mutações poderiam, em princípio, ser suficientes para estimar árvores de boa qualidade), e o porque de análises filogenéticas anteriores utilizando conjuntos restritos de loci identificaram diferentes grupos de isolados divergentes dentro de uma escala excelente com um bom suporte de dados (por exemplo, spoligotyping, MIRU) (Allix et al., 2006).

E então, será este desempenho evidência de algum tipo de viés filogenético, ou ele está associado a uma verdadeira falta de informatividade dos dados de DNA?

Responder a esta pergunta é importante não só para lidar com a evolução de *M. bovis*, mas também para suportar elevadas ramificações, que geralmente são consideradas boa evidência para a existência de ramos reais e são bastante comuns em análises de genomas completos e ainda enganosa, em muitos casos, como a inferência de *core coding* (Figura 2). Inconsistência filogenética foi relatada em outros lugares e em muitos estudos, e pode ser grosseiramente subdividida em duas categorias principais: 1) viés de metodologia, tais como zonas de inconsistência filogenética associadas a escolha do método e 2) tendência inerente às peculiaridades evolutivas do conjunto de dados, incluindo efeitos de seleção positiva, seleção negativa, recombinação e outras forças.

Desempenho do DNA

Comparamos a árvore filogenética do *core coding* com a árvore filogenética de MP utilizando o mesmo conjunto de dados para testar vieses metodológicas. Optamos pelo modelo de ML após estudos dirrigidos e o particionamento dos dados, porém sempre há a possibilidade de que o modelo escolhido (a melhor opção) seja contudo, ainda longe de ser representativo do processo evolutivo real. O modelo MP geral necessita de caracteres de ponderação e taxas completas (assim de modo implícito assumi-se que ambas são iguais em todo alinhamento, o que normalmente é pouco provável em cenários evolutivos moleculares realistas) e implica em uma comparação mais adequada do que uma árvore Bayesiana, mesmo esta sendo correlacionada com ML uma vez que também é uma metodologia paramétrica. A árvore MP do *core coding* apresentou 7.108 passos, e precisão semelhante a ML utilizando o mesmo conjunto de dados (mesmo RI), sugerindo assim que o viés identificado não está relacionado a metodologia empregada.

A recombinação homóloga é conhecida por ser um fenômeno de forças invasivas em bactérias. Para testar se este fenômeno está causando qualquer viés filogenético, utilizamos o conjunto de dados LCBs para estimar a recombinação homóloga por quatro métodos diferentes. Os três primeiros métodos (PHI, NSS, MaxChi) avaliaram se cada bloco (individualmente) apresentava sinais de recombinação e, em seguida,

adotando uma abordagem conservadora, excluímos os que apresentaram sinais leves de recombinação. Três dos LCBs (blocos 1, 5 e 10) não apresentaram sinais de recombinação, por isso, concatenamo-os e executamos a estimativa ML, para um total de 174.900 pb, com apenas 350 sítios polimórficos (0,2% sítios variáveis, 0,07% sítios informativos parcimôniaais, espelhando o alinhamento do *core coding*). A árvore (Figura 2) apresentou RI ainda menor do que a árvore original do *core coding* (Tabela 3).

O quarto método de detecção de recombinação (ClonalFrameML) avaliou o percentual de recombinação em relação às mutações na topologia de evidência total utilizando os 10 LCBs concatenados (550.634 pb). Rho/theta (razão entre a taxa de recombinação e a taxa de mutação) e r/m (razão entre a quantidade estimada de polimorfismos interpostos por um evento de recombinação em relação a um evento de mutação) apresentam valores médios através de ramos semelhantes em magnitude, com $Rho/theta=0,92$ e $r/m=0,82$, evidenciando a taxa relativamente baixa de recombinação, e, além disso, os parâmetros delta (comprimento do trato média de um evento de recombinação) e ν (taxa de novas substituições devido à recombinação) estão abaixo de 1,0, sugerindo baixo impacto da recombinação em valores absolutos. No geral, as análises que enfocam a influência da recombinação na evolução de *M. bovis* indicam baixa magnitude; no entanto, não esperamos que a remoção das regiões supostamente recombinadas agravaria esta topologia, porém pode ser uma questão de falta de poder da análise devido ao número relativamente pequeno de sítios concatenados (apenas 123 sítios informativos parcimôniaais em um total de 174.900 pb). Portanto, não podemos atribuir o baixo RI do *core coding* em relação aos dados de presença/ausência aos efeitos de recombinação.

A avaliação de seleção positiva indicou que 114 dos 2483 (4%) genes estão sob seleção positiva. Para testar se é um fator afeta a filogenia do *core coding*, removemos em primeiro lugar estes 114 genes do alinhamento e reestimamos a topologia. Essa manobra não mostrou nenhum efeito, a topologia apresentou um novo RI similar ao anterior. O segundo teste manteve os 114 genes sob seleção positiva, porém estimou uma otimização na filogenia na relação dN/dS (ômega) em códons ao longo dos braços

da árvore, portanto, foi capaz (pelo menos em teoria) de superar preceitos filogenéticos uma vez que corrigiu a pressão de seleção durante a análise. No entanto, a topologia dos códons continuou a apresentar o mesmo IR baixo para o *core coding*, e assim a seleção positiva também não interferiu com a reconstrução filogenética.

A seleção purificadora poderia explicar a baixa variabilidade global de dados de DNA, e neste caso uma filogenia “ruim” pode ser reflexo da falta de poder de dados com poucos sítios informativos (taxas de recombinação, como já foi discutido acima). No entanto, isso tenderia a manter a relação dN/dS global relativamente perto de 0,0 uma vez que substituições não sinônimas iriam acumular muito menos do que as substituições sinônimas. No entanto, a relação omega obtida pela análise anterior (aproximadamente 0,6) é demasiadamente elevada para ser explicada pela seleção purificadora em bactérias (Ran et al., 2014.). Além disso, estes dados tenderiam a manter a ordem típica das taxas de substituições relativas entre as três partições de códons, $3 > 1 > 2$, mas ao invés disso, nossos dados revelaram $3 > 2 > 1$, de modo que esta possibilidade também pode ser descartada.

Aparentemente, após descartarmos os vieses metodológicos, a recombinação homóloga, a seleção positiva e a seleção purificadora como fatores que influenciam o comportamento filogenética do *core coding*, algumas outras opções permanecem, como a pseudogeneização, a fim de contribuir com contínuas novas taxas de substituições, aparentemente aleatórias. Outra possibilidade, mais parcimoniosa, é que a seleção purificadora ainda não teve tempo suficiente para eliminar as substituições moderadamente deletérias (Rocha et al., 2006) que ainda ocorrem ao longo dos genomas (em taxas de mutação baixas, dada a escassez de sítios informativos), então elas estão se acumulando lentamente em um padrão aparentemente aleatória desde o surgimento de *M. bovis* até os dias atuais, e assim não contribuindo filogeneticamente. A HGT, por outro lado (aqui observada pelas análises de *indels* e pan-genoma) apresentou-se mais informativa em relação as localizações geográficas.

Estimativas de datação de divergência utilizando *indels* e pan-genoma

Avaliações de divergência são comumente realizadas utilizando dados de DNA. No entanto, neste estudo, não poderíamos confiar em dados de DNA genômico isolados devido ao desempenho filogenético tendencioso, então recorremos a diferentes testes modelos de datação com dados de presença/ausência em BEAST (*indels* + pan-genoma). Para nossa surpresa, os resultados de todos os testes realizados convergiram embora tenhamos usado uma taxa bem conservadora abrangendo seis ordens de magnitude. Este fato pode ser possivelmente explicado pela informatividade na presença/ausência como evidenciado pela sua melhor eficácia filogenética, associado aos menores conjuntos de dados quando comparado com ao tamanho do DNA genômico nos alinhamentos, o que permite economia de tempo computacional. Estes resultados ainda informaram a hipótese de rejeição de dados específicos (Tabela 4), tal como menor probabilidade de modelos de tempo fixo. Além disso, a origem de *M. bovis* aqui estimada baseada em um estudo anterior, localiza-se a sua origem, no máximo, há 10.000 anos (Wirth et al., 2008). Embora nenhuma dessas observações forneçam garantias em relação ao fato dos dados de presença/ausência calcularem com precisão eventos antigos de *M. bovis*, este tipo de análise foi útil para a exploração dos dados, tanto filogeneticamente (RI) como também pela estimativa robusta de divergência, que pode representar uma força especial nos casos em que a datação da divergência é relativamente recente para informar, como parece ser o caso aqui descrito.

Evolução de *M. bovis* e o impacto da HGT

Todas as árvores da figura 2 concordam no posicionamento mais estreito das amostras africanas nas ramificações basais. Essas são amostras isoladas de seres humanos, sugerindo um importante início de diversificação e ao mesmo tempo, corroborando evidências de que *M. bovis* tenha evoluído a partir de *M. tuberculosis* ao invés do contrário (Wirth et al., 2008). Aspectos interessantes foram observados a partir de análises do clado dos EUA, uma vez que HGT afetou fortemente a evolução deste

grupo, levando a um grande número de ramificações destes isolados (Figura 2). Para melhor inferirmos sobre este aparente impacto da HGT no clado norte-americano, buscamos por ganhos ou perdas sinapomórficas neste grupo. Observamos que os ganhos e perdas deste grupo de genes estão espalhados por todo o genoma, sem um padrão claro. Através da realização de busca de sequências por BLASTn nos 57 ganhos sinapomórficos em linhagens dos EUA, descobrimos que esses encontros eram de isolados da Rússia, Cazaquistão, China, Coréia do Sul, derivados de BCG (vacina de tuberculose baseados em *M. bovis*) ou de cepas multirresistentes, evidenciando locais da Ásia, como origem dos genes específicos dos isolados dos EUA e levantando alguma preocupação sobre a possibilidade de novos surtos da doença no território americano. Além disso, muitos destes genes pertencem a família PE/PPE, os quais são fatores genéticos conhecidos associados a virulência da espécie. Tais aquisições podem ter favorecido as micobactérias nas relações com seu hospedeiro.

Há também um sub-clado dentro dos isolados norte-americanos com taxas evolutivas aparentemente mais elevadas, de acordo com os comprimentos dos ramos de todos os três tipos de dados (DNA, *indels* e pan-genoma)(Figura 2). Uma vez que estes isolados estão localizados em estados distantes (NM, PA e MN), este fato sugere que há risco de surtos epidêmicos nos EUA, caso as medidas de controle não sejam continuamente mantidas. Um fator que pode ser associado com tais potenciais de disseminação aparentes nos EUA é a ampla presença de reservatórios sinantrópicos, como pode ser visto através de nossas amostras, bovinos, veados e alces em convívio por exemplo (Figuras 2 e 3). Evidências semelhantes de contaminação cruzada pode ser observado em outras partes da árvore filogenética de evidência total (Figuras 2 e 3), por exemplo, clados bovino na América do Sul e ovelhas do Brasil evolutivamente perto de clados bovinos do Reino Unido e isolado de javali da Coréia.

Financiamento

Este projeto foi financiado pela Embrapa (03.14.00.054.00.00, 02.13.16.002.00.00), Fundect (TO 096/2012; 085/2015; 23/200.656/2013), CNPq (443235/2014-7; 305857/2013-4) e Fundação Bill & Melinda Gates (edital Marketplace ID 1611).

Referências

1. **Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R.M.** (2010). Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg. Infect. Dis.* **16**, 1296–1299.
2. **Allix, C., Walravens, K., Saegerman, C., Godfroid, J., Supply, P., & Fauville-Dufaux, M.** (2006). 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. *J. Clin. Microbiol.* **44**(6), 1951-1962.
3. **Boldt J.** (2003). [Point of care (POC) monitoring in anesthesia and intensive care--an overview of available POC systems]. *Anesthesiol Intensivmed Notfallmed Schmerzther* **38**(3):158-64. Review.
4. **Bruen, T. & Bruen, T.** (2005). PhiPack: PHI test and other tests of recombination. McGill University, Montréal, Québec, Canada.
5. **Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., McAdams, H. H.** (2004). Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Nat. Acad. Sci. USA.* **101**(10), 3480-3485.
6. **Contreras-Moreira, B. & Vinuesa, P.** (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pan-genoma analysis. *Appl. Environm. Microbiol.* **79**(24), 7696-7701.
7. **Darling, A. E., Mau, B., Perna, N. T.** (2010). ProgressiveMAUVE: multiple genome alignment with gene gain, loss and rearrangement. *PLoS one*, **5**(6), e11147.
8. **Didelot, X. & Wilson, D. J.** (2015). ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS Comput. Biol.* **11**(2), e1004041.
9. **Drummond, A. J., Suchard, M. A., Xie, D., Rambaut, A.** (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molec. Biol. Evol.* **29**(8),1969-1973.
10. **Fabre, M., Koeck, J. L., Le Flèche, P., Simon, F., Hervé, V., Vergnaud, G., Pourcel, C.** (2004). High genetic diversity revealed by variable-number tandem repeat

genotyping and analysis of hsp65 gene polymorphism in a large collection of "*Mycobacterium canettii*" strains indicates that the *M. tuberculosis* complex is a recently emerged clone of "*M. canettii*". J. Clin. Microbiol. **42**(7):3248-3255.

11. Ford, C. B., Lin, P. L., Chase, M. R., Shah, R. R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T. R., Sacchettini, J. C., Lipsitch, M., Flynn, J. L., Fortune, S. M. (2011). Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat. Genet. **43**: 482–486.

12. Forrellad, M. A., Klepp, L. I., Gioffré, A., García, J. S., Morbidoni, H. R., Santangelo, M. P., Cataldi, A. A., Bigi, F. (2013). Virulence factors of the *Mycobacterium* complex. Virulence **4**(1):3-66.

13. Gardy, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L. N., Brodtkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M. K., Elwood, K., Jones, S. J., Brinkman, F. S., Brunham, R. C., Tang, P. (2011). Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. N. Engl. J. Med. **364**: 730–739.

14. Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J. A., Edgeworth, J. D., de Lencastre, H., Parkhill, J., Peacock, S. J., Bentley, S. D. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. Science **327**:469–474.

15. Hildebrand, F., Meyer, A., Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. PLoS Genet, **6**(9), e1001107.

16. Hilty, M., Diguimbaye, C., Schelling, E., Baggi, F., Tanner, M., Zinsstag, J. (2005). Evaluation of the discriminatory power of variable number tandem repeat (VNTR) typing of *Mycobacterium bovis* strains. Vet. Microbiol. **109**:217–222.

17. Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J. (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35**, 907-914.

18. Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molec. Biol. Evol. **30**(4), 772-780.

19. Kück, P. & Meusemann, K. (2010). FASconCAT: convenient handling of data matrices. Molec. Phylogen. Evol. **56**(3), 1115-1118.

20. Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinf. **30**(22), 3276-3278.

21. **Li, J., Zhou, J., Wu, Y., Yang, S., Tian, D.** (2015). GC-content of synonymous codons profoundly influences amino acid usage. *G3: Genes Genomes Genetics* **5**(10), 2027-2036.
22. **Li, L., Stoeckert, C. J., Roos, D. S.** (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**(9), 2178-2189.
23. **Li, X.-Q., Du, D.** (2014). Variation, Evolution, and Correlation Analysis of C+G Content and Genome or Chromosome Size in Different Kingdoms and Phyla. *PLoS one* **9**(2), e88339.
24. **Minh, B. Q., Nguyen, M. A. T., von Haeseler, A.** (2013). Ultrafast approximation for phylogenetic bootstrap. *Molec. Biol. Evol.* mst024.
25. **Mostowy, S., Inwald, J., Gordon, S., Martin, C., Warren, R., Kremer, K., Behr, M. A.** (2005). Revisiting the evolution of *Mycobacterium bovis*. *J. Bacteriol.* **187**(18), 6386-6395.
26. **Müller, K.** (2005). SeqState. *Appl. Bioinf.* **4**(1), 65-69.
27. **Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S. L. K., & Scheffler, K.** (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molec. Biol. Evol.* mst030.
28. **Nguyen, L. T., Schmidt, H. A., von Haeseler, A., Minh, B. Q.** (2015). IQ-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molec. Biol. Evol.* **32**(1), 268-274.
29. **Pond, S. L. K. & Frost, S. D.** (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molec. Biol. Evol.* **22**(5), 1208-1222.
30. **Pond, S. L. K. & Muse, S. V.** (2005). HyPhy: hypothesis testing using phylogenies. In *Statistical methods in molecular evolution* (pp. 125-181). Springer New York.
31. **Ran, W., Kristensen, D. M., Koonin, E. V.** (2014). Coupling between protein level selection and codon usage optimization in the evolution of bacteria and archaea. *Mol. Bio.* **5**(2), e00956-14.
32. **Rocha, E. P., Smith, J. M., Hurst, L. D., Holden, M. T., Cooper, J. E., Smith, N. H., Feil, E. J.** (2006). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theoret. Biol.* **239**(2), 226-235.

33. **Sela, I., Ashkenazy, H., Katoh, K., Pupko, T.** (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* gkv318.
34. **Sharp, P. M., Emery, L. R., Zeng, K.** (2010). Forces that influence the evolution of codon bias. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**(1544), 1203-1212.
35. **Smith., N. H., Gordon, S. V., de La Rua-Domenech, R., Clifton-Hadley, R. S., Hewinson, G.** (2006) Bottleneck and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nature* **4**:670-681.
36. **Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsç-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.** (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **44**, 4498-4510.
37. **Swofford, D. L.** (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
38. **Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S.** (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molec. Biol. Evol.* **30**(12), 2725-2729.
39. **Thoen, C., Lobue, P., de Kantor, I.** (2006). The importance of *Mycobacterium bovis* as a zoonosis. *Vet. Microbiol.* **112**(2-4):339-345. Review.
40. **Van Ingen, J., Rahim, Z., Mulder, A., Boeree, M. J., Simeoni, R., Brosch, R., van Soolingen, D.** (2012). Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerg. Inf. Dis.* **18**(4): 653-655.
41. **Venton, D.** (2013). Highlight: Bacteria Shed DNA as They Adapt to Hot Temperatures. *Gen. Biol. Evol.* **5**(5), 987-988.
42. **Wirth, T., Hildebrand, F., Allix-Béguet, C., Wölbelling, F., Kubica, T., Kremer, K., van Soolingen, D., Rüsç-Gerdes, S., Locht, C., Brisse, S., Meyer, A., Supply, P., Niemann, S.** (2008). Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog.* **19**;4(9): e1000160.
43. **Xia, X.** (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molec. Biol. Evol.* **30**(7), 1720-1728.

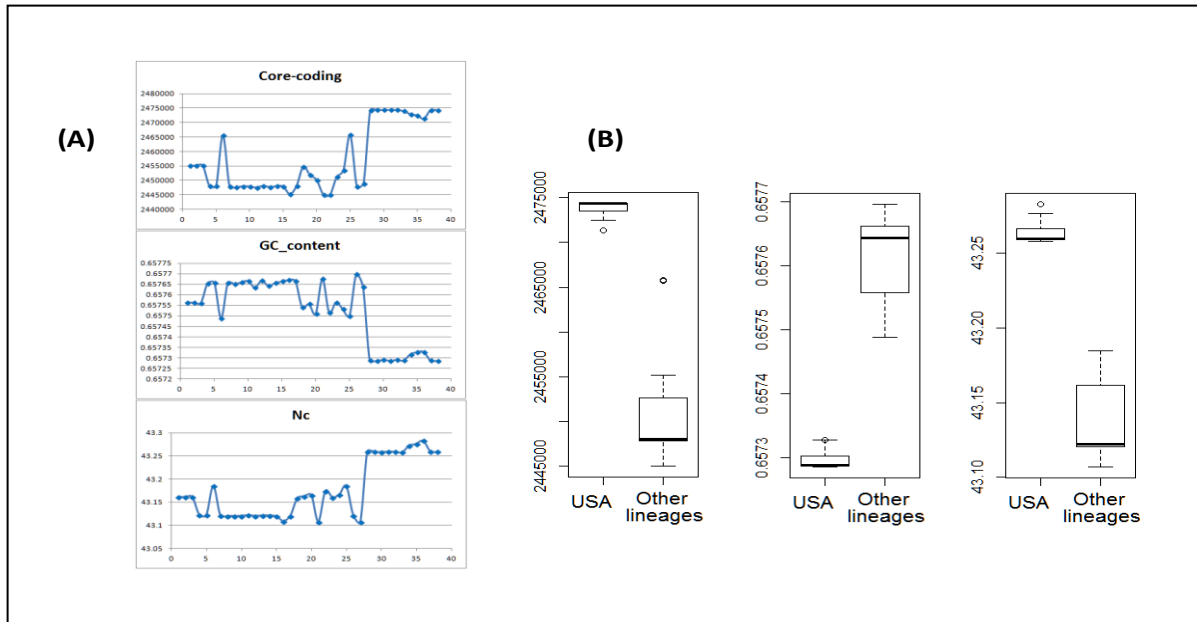


Figura 1. *Box-plots* de quantidades estimadas a partir dos 28 genomas analisados. (A) Padrão geral do tamanho do *core coding*, conteúdo CG e do número efetivo de códon em todos os 28 genomas (os últimos 11 da direita são amostras dos EUA). (B) *Box-plots* ilustrando a distribuição de valores dos tamanhos do *core coding*, conteúdo CG e número efetivo de códon respectivamente.

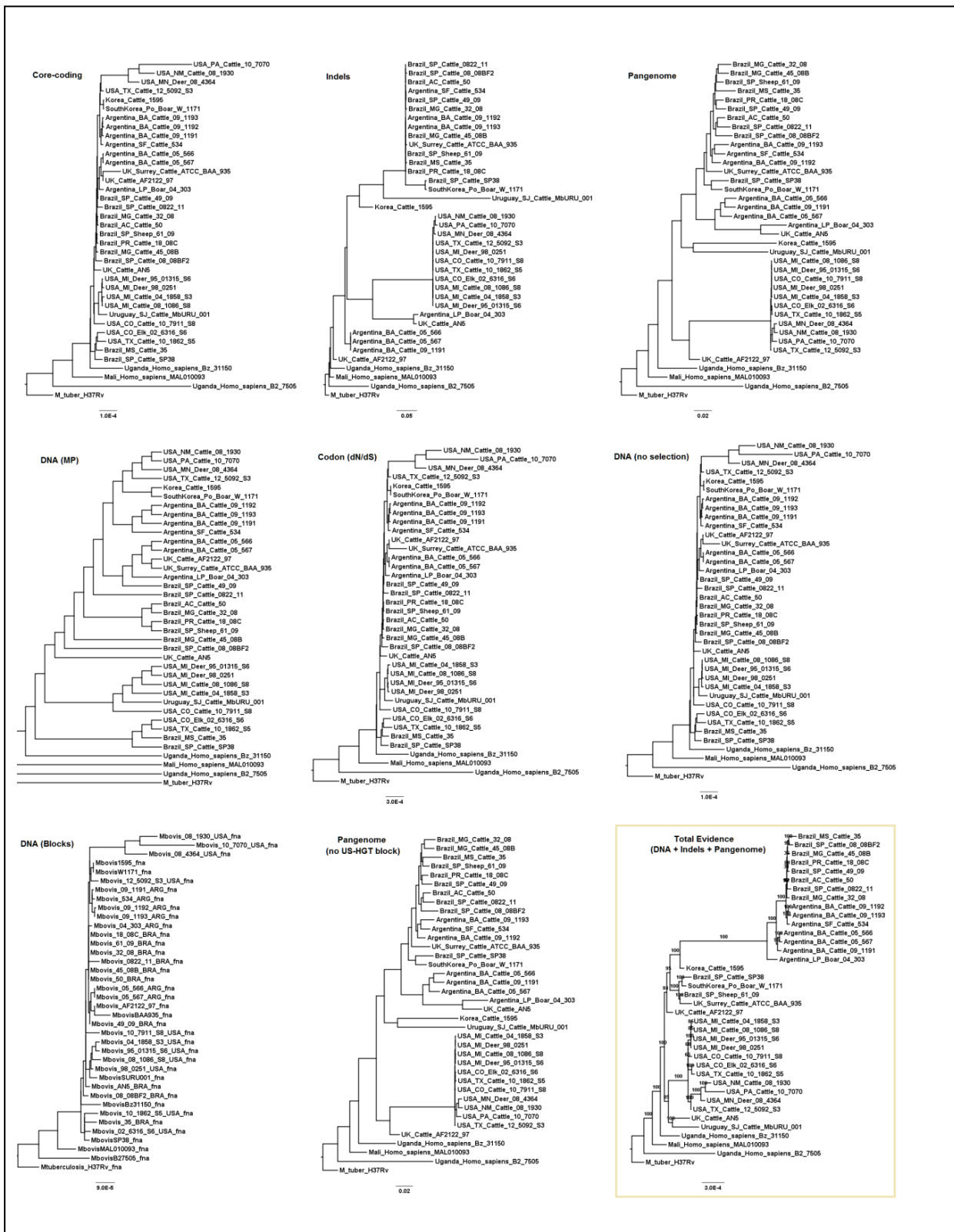


Figura 2. Árvores filogenéticas construídas com metodologia de máxima parcimônia (MP) para diferentes parâmetros com diferentes conjuntos de dados.

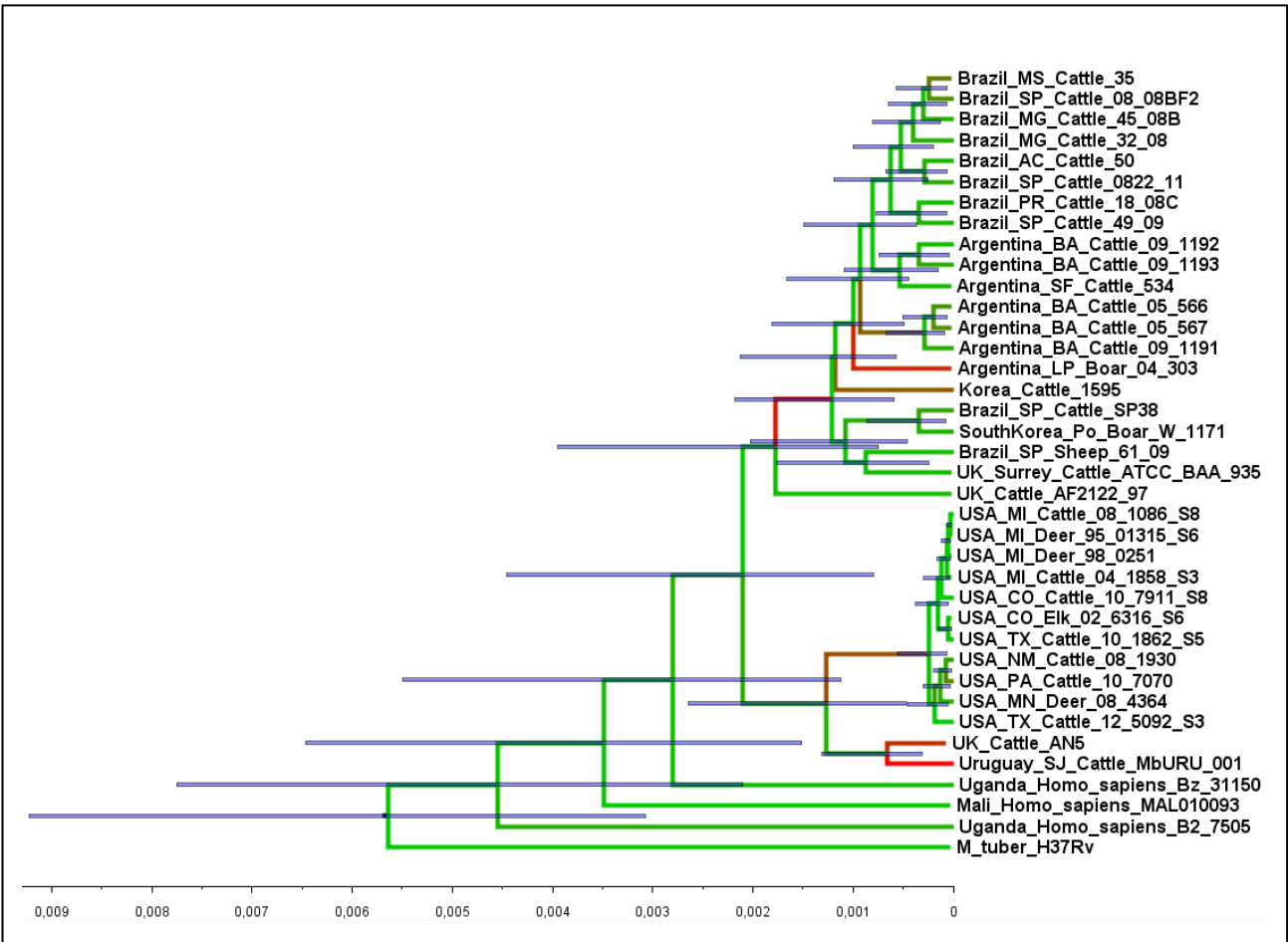


Figura 3. Cronograma Bayesiano de acordo com análise MCMC assumindo o modelo de divergência de dados (não correlação da distribuição lognormal das taxas através dos ramos), com base em *indels* + pan-genoma e topologia de evidência total ML fixa. Ramos coloridos refletem taxa evolutiva média dos ramos (verde=inferior, vermelho=superior), com um aumento significativo na ramificação referente ao clado EUA e também em alguns outros ramos. 95% HPD, barras mostradas em azul, indicando que *M. bovis* pode ter se originado no passado há aproximadamente 10.000 anos.

Sample	Country	State	Isolate	Species	Year of Isolation	Spoligotype	VNTR
ARG_91191	Argentina	Buenos Aires	91191	Cattle	2009	SB0140	232224253322
ARG_V2D	Argentina	Buenos Aires	V2D	Cattle	2000	SB0140	232224253322
ARG_V2A	Argentina	Buenos Aires	V2A	Cattle	2000	SB0140	232224253322
ARG_4303	Argentina	La Pampa	04-303	Boar	2004	SB0140	232224263322
ARG_534	Argentina	Santa Fé	534	Cattle	1994	SB0140	232224163322
ARG_91193	Argentina	Buenos Aires	91193	Cattle	2009	SB0140	232324253322
ARG_91192	Argentina	Buenos Aires	91192	Cattle	2009	SB0140	232324253322
BRA_45-08B	Brazil	Minas Gerais	45-08B	Cattle	2008	SB1050	unavailable
BRA_18-08C	Brazil	Paraná	18-08C	Cattle	2008	SB0121	unavailable
BRA_61-09	Brazil	São Paulo	61-09	Cattle	2009	SB1869	unavailable
BRA_32-08	Brazil	Minas Gerais	32-08	Cattle	2008	SB0140	unavailable
BRA_49-09	Brazil	São Paulo	49-09	Cattle	2009	SB0120	unavailable
BRA_08-08BF2	Brazil	São Paulo	08-08BF2	Cattle	2008	SB1328	unavailable
BRA_50	Brazil	Acre	50	Cattle	2009	SB1137	unavailable
BRA_35	Brazil	Mato Grosso do Sul	35	Cattle	2003	SB0881	unavailable
BRA_AN5	Brazil	PPD lab strain	AN5	Cattle	unavailable	SB0268	unavailable
BRA_0822-11	Brazil	São Paulo	0822-11	Cattle	2012	unavailable	unavailable
USA_02-6316	United States	Colorado	02-6316_S6	Elk	2002	SB0265	23335432534
USA_04-1858	United States	Michigan	04-1858_S3	Cattle	2003	SB0292	23237552533
USA_08-1086	United States	Michigan	08-1086_S8	Cattle	2007	SB0145	3237552533
USA_08-1930	United States	New Mexico	08 1930	Cattle	2007	SB0673	25237452534
USA_08-4364	United States	Minnisota	08 4364	Deer	2008	SB0271	25237452534
USA_10-1862	United States	Texas	10-1862_S5	Cattle	2009	SB0121	23336442535
USA_10-7911	United States	Colorado	10-7911_S8	Cattle	2010	SB0211	25147452323
USA_10-7070	United States	Pennsylvania	10 7070	Cattle	2010	SB0265	23335432534
USA_12-5092	United States	Texas	12-5092_S3	Cattle	2012	SB0271	253473453214
USA_95-01315	United States	Michigan	95-01315_S6	Deer	1994	SB0145	23237552533
USA_08-0251	United States	Michigan	98-0251	Deer	1997	SB0145	2334755113

Tabela 1. Amostras utilizadas neste estudo, informações de país de origem, isolamento (local e ano) e dados de *Spoligotyping* e MIRU disponíveis.

SeqName	Core-coding	Nc	GC content	Total bases sequenced	Contigs
Argentina_BA_Cattle_05_566	2455143	43.1614	0.657563735	4308143	89
Argentina_BA_Cattle_05_567	2455143	43.1615	0.65756292	4328889	69
Argentina_BA_Cattle_09_1191	2455158	43.1614	0.657560532	4327459	86
Argentina_BA_Cattle_09_1192	2448051	43.1222	0.657655008	4334537	56
Argentina_BA_Cattle_09_1193	2448051	43.1219	0.657657867	4308937	89
Argentina_LP_Boar_04_303	2465685	43.1849	0.657489095	4313075	169
Argentina_SF_Cattle_534	2447877	43.1214	0.657657227	4325986	72
Brazil_AC_Cattle_50	2447661	43.1202	0.657651938	4320450	84
Brazil_MG_Cattle_32_08	2447979	43.1202	0.657660053	4314192	83
Brazil_MG_Cattle_45_08B	2447916	43.1206	0.657666358	4320055	72
Brazil_MS_Cattle_35	2447568	43.1219	0.657634844	4275484	94
Brazil_PR_Cattle_18_08C	2448102	43.1219	0.657667042	4329318	63
Brazil_SP_Cattle_08_08BF2	2447727	43.1219	0.65764401	4311106	108
Brazil_SP_Cattle_0822_11	2448000	43.1219	0.657655637	4301774	85
Brazil_SP_Cattle_49_09	2447925	43.1219	0.657664348	4333756	61
Brazil_SP_Cattle_SP38	2445258	43.1219	0.657670479	4370360	36
Brazil_SP_Sheep_61_09	2448099	43.1219	0.657666214	4332493	68
Korea_Cattle_1595	2454768	43.1219	0.657540346	4351712	1
M_tuber_H37Rv	2451837	43.1219	0.657556763	4411532	1
Mali_Homo_sapiens_MAL010093	2450121	43.1219	0.657509976	4381900	4
SouthKorea_Po_Boar_W_1171	2445051	43.1219	0.657676261	4304865	50
Uganda_Homo_sapiens_B2_7505	2445030	43.1219	0.657515041	4166298	243

Uganda_Homo_sapiens_Bz_31150	2451225	43.1219	0.657563463	4271367	136
UK_Cattle_AF2122_97	2453499	43.1219	0.65753318	4345492	1
UK_Cattle_AN5	2465802	43.1219	0.657499669	4325667	70
UK_Surrey_Cattle_ATCC_BAA_935	2447790	43.1219	0.657696943	4358088	1
Uruguay_SJ_Cattle_MbURU_001	2448981	43.1219	0.657637605	4295606	100
USA_CO_Cattle_10_7911_S8	2474358	43.1219	0.657288072	4345492	1
USA_CO_Elk_02_6316_S6	2474379	43.1219	0.657285727	4345492	1
USA_MI_Cattle_04_1858_S3	2474379	43.1219	0.657290173	4345492	1
USA_MI_Cattle_08_1086_S8	2474379	43.1219	0.657286131	4345492	1
USA_MI_Deer_95_01315_S6	2474379	43.1219	0.657290173	4345492	1
USA_MI_Deer_98_0251	2473992	43.1219	0.657287089	4345492	1
USA_MN_Deer_08_4364	2472948	43.1219	0.657315884	4345492	1
USA_NM_Cattle_08_1930	2472507	43.1219	0.657327563	4345492	1
USA_PA_Cattle_10_7070	2471358	43.1219	0.657328076	4345492	1
USA_TX_Cattle_10_1862_S5	2474376	43.1219	0.657286928	4345492	1
USA_TX_Cattle_12_5092_S3	2474343	43.1219	0.657286399	4345492	1

Tabela 2. Número de genes do *core coding*, número efetivo de códons (NC), proporção do conteúdo CG, número total de bases sequenciadas e número de *contigs* obtidos (quantidades estimadas para cada um dos genomas).

Tree	RI
DNA	0.6875
Indels	0.75
Pan-genoma	0.75
DNA (maximum parsimony)	0.6875
DNA Omega (dN/dS)	0.6875
DNA (no rec/sel)	0.6875
DNA Blocks	0.5625
DNA + Indels + Pan-genoma	0.75
Pan-genoma (no US-HGT block)	0.75

Tabela 3. Índice de retenção (RI) das reconstruções filogenéticas com base em diferentes conjuntos de dados. (DNA: genes do *core coding*; Omega: estimativa média de ramo global das mutações não sinônimas por mutações sinônimas [DN/DS]; Blocos: blocos localmente colineares; Rec: recombinação homóloga; Sel: genes selecionados de forma positiva; bloco US-HGT: região genômica única contendo apenas genes dos isolados dos EUA).

Test	AICm
Test 1: clock type and different clock priors	
uclid.mean - Uniform [0, 100] (b)	83004.79
uclid.mean - Exponential [0, 100] (b)	83004.87
uclid.mean - Uniform [0, 100] (a)	83005.35
uclid.mean - Exponential [0, 100] (a)	83007.93
clock.rate - Uniform [0, 100] (a)	85661.31
clock.rate - Exponential [0, 100] (a)	85661.57
clock.rate - Uniform [0, 100] (b)	85663.07
clock.rate - Exponential [0, 100] (b)	85669.61
Test 2: Age prior (larger vs. smaller log stdv)	
uclid.mean - Uniform [0, 100] age_prior_std=0.3 (b)	83004.56 *
uclid.mean - Uniform [0, 100] (b)	83004.79
uclid.mean - Uniform [0, 100] (a)	83005.35
uclid.mean - Uniform [0, 100] age_prior_std=0.3 (a)	83009.74
Test 3: Rate stdv prior (larger vs. smaller uclid.stdv)	
uclid.mean - Uniform [0, 100] (b)	83004.79
uclid.mean - Uniform [0, 100] (a)	83005.35
uclid.mean - Uniform [0, 100] uclid.stdev=Exp(10) (a)	83028.64
uclid.mean - Uniform [0, 100] uclid.stdev=Exp(10) (b)	83021.24

Tabela 4: Resultados de testes utilizando dados de presença/ausência (*indels* + pan-genoma) para estimar divergências em BEAST. Primeiro avaliamos diferentes modelos

de tempo (estrito vs. não correlacionados) em conjunto com distribuições anteriores alternativas para a taxa de tempo média (uniforme vs. exponencial). Em seguida, comparamos o melhor modelo de tempo/média da taxa (modelo não correlacionado de taxas), com diferentes antecedentes de idade (maior vs. menor), e, finalmente, em comparação com diferentes desvios de taxas entre todas as sucursais (mais forte contra mais suave) . Melhor modelo geral está marcado com "*". As comparações foram realizadas em v1.6 Tracer utilizando o critério AICM com 100 *bootstraps*.