TAIANA TAINÁ SILVA PEREIRA
Sequenciamento, anotação e análise do genoma de <i>Mycobacterium pinnipedii</i> e genômica comparativa de espécies do Complexo <i>Mycobacterium tuberculosis</i>
Sequenciamento, anotação e análise do genoma de <i>Mycobacterium pinnipedii</i> e genômica comparativa de espécies do Complexo <i>Mycobacterium tuberculosis</i>
Sequenciamento, anotação e análise do genoma de Mycobacterium pinnipedii e genômica comparativa de espécies do Complexo Mycobacterium tuberculosis
Sequenciamento, anotação e análise do genoma de <i>Mycobacterium pinnipedii</i> e genômica comparativa de espécies do Complexo <i>Mycobacterium tuberculosis</i>
Sequenciamento, anotação e análise do genoma de <i>Mycobacterium pinnipedii</i> e genômica comparativa de espécies do Complexo <i>Mycobacterium tuberculosis</i>
Sequenciamento, anotação e análise do genoma de Mycobacterium pinnipedii e genômica comparativa de espécies do Complexo Mycobacterium tuberculosis
Sequenciamento, anotação e análise do genoma de Mycobacterium pinnipedii e genômica comparativa de espécies do Complexo Mycobacterium tuberculosis

TAIANA TAINÁ SILVA PEREIRA

Sequenciamento, anotação e análise do genoma de *Mycobacterium pinnipedii* e genômica comparativa de espécies do Complexo *Mycobacterium tuberculosis*

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

Departamento:

Medicina Veterinária Preventiva e Saúde Animal

Área de concentração:

Epidemiologia Experimental Aplicada à Zoonoses

Orientador:

Prof. Dr. Ana Marcia de Sá Guimarães

De acordo: AnaMarciadeSaG

Orientador

São Paulo 2019

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

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

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO

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

T. 3841 FMVZ Pereira, Taiana Tainá Silva

Sequenciamento, anotação e análise do genoma de Mycobacterium pinnipedii e genômica comparativa de espécies do Complexo Mycobacterium tuberculosis / Taiana Tainá Silva Pereira. – 2019.

155 f. : il.

Dissertação (Mestrado) – Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Medicina Veterinária Preventiva e Saúde Animal, São Paulo, 2019.

 ${\bf Programa\ de\ P\'os\hbox{-}Gradua\'ç\~ao}.\ {\bf Epidemiologia\ Experimental\ Aplicada\ \`as\ Zoonoses}.$

Área de concentração: Epidemiologia Experimental Aplicada às Zoonoses.

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

1. Complexo Mycobacterium tuberculosis. 2. Genoma. 3. Mycobacterium pinnipedii. 4. Fatores de virulência. 5. Epítopos. I. Título.

Ficha catalográfica elaborada pela bibliotecária Denise Yamashita, CRB-8/8931, da FMVZ/USP.





Comissão de Ética no Uso de Animais

São Paulo, 27th May 2019

CERTIFIED

We certify that the Research "Sequencing, annotation and analysis of the Mycobacterium genome Pinnipedii and comparative genomic species of the Mycobacterium tuberculosis complex", protocol number CEUAx 3653040717 (ID 000795), under the responsibility Ana Marcia De Sá Guimarães, agree with Ethical Principles in Animal Research adopted by Ethic Committee in the Use of Animals of School of Veterinary Medicine and Animal Science (University of São Paulo), and was approved in the meeting of day August 03, 2017.

Certificamos que o protocolo do Projeto de Pesquisa intitulado "Sequenciamento, anotação e análise do genoma de Mycobacterium pinnipedii e genômica comparativa de espécies do complexo Mycobacterium tuberculosis", protocolado sob o CEUAx nº 3653040717, sob a responsabilidade de Ana Marcia De Sá Guimarães, está de acordo com os princípios éticos de experimentação animal da Comissão de Ética no Uso de Animais da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, e foi aprovado na reunião de 03 de agosto de 2017.

Profa. Dra. Anneliese de Souza Traldi Presidente da Comissão de Ética no Uso de Animais

Annelien Tealor

Faculdade de Medicina Veterinária e Zootecnia da Universidade Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo

Roseli da Costa Gomes Secretária

de São Paulo

FOLHA DE AVALIAÇÃO

Autor: SILVA-PEREIRA, Taina Tainá

Título: Sequenciamento, anotação e análise do genoma de *Mycobacterium pinnipedii* e genômica comparativa de espécies do Complexo *Mycobacterium tuberculosis*

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

Data:/		
	Banca Examinadora	
Prof. Dr		
	Julgamento:	
Prof. Dr		
	Julgamento:	
Prof. Dr.		
Instituição:	Julgamento:	

DEDICATÓRIA

Aos meus pais, Adalci e Benedito. Às minhas irmãs, Allana e Jéssica. À Dona Nice, Seu Adalberto e Adenir. Aos ancestrais que derramaram sangue e abriram caminhos.

AGRADECIMENTO ESPECIAL

Ao conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) número de processo 134266/2017-0, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) de número 001, 1721455, à Morris Animal Foundation de número D17ZO-307 e Fundação de Amparo à Pesquisa do Estado do Estado de São Paulo (FAPESP) de número 2016/26108-0 pelo apoio financeiro que permitiu a realização desse estudo.

AGRADECIMENTOS

À Universidade de São Paulo, funcionários do ICB II e do VPS-FMVZ pelo suporte estrutural para o desenvolvimento da pesquisa.

Ao CNPq pelo apoio financeiro via bolsa de mestrado.

À professora Ana Marcia pela confiança, orientação, ensinamentos, motivação, paciência e dedicação. Por ser um exemplo profissional que contribuiu de forma significativa para muito além da minha formação acadêmica. Se você não fosse exatamente quem é, o meu caminho da pós-graduação teria sido mais tortuoso do que ele já é por si só.

Ao professor Lucas, por acreditar e ver qualquer potencial em mim.

À todos os professores que contribuíram para a minha formação.

À Lorena Jardim por contribuir para que eu tenha sempre a melhor vesão de mim. Agradeço por permanecer. Por ser suporte, por ser família.

Aos amigos que permaneceram apesar da distância geográfica. À Maíra e Beatriz por estarem aqui pra tudo, todo dia.

À Aléxia, Samuel, Lennon, Bianca, Dalva e Hélic. Vocês fizeram com que São Paulo tivesse cheiro de casa. Foram imprescindíveis para a minha adaptação tão longe de casa.

À Naila Soler, pelo carinho, pelo colo e cuidado independente de qualquer coisa. Por dizer exatamente o que eu preciso ouvir. Por fincar meus pés no chão.

À Cristina, Maysa e Mara, pelo norte, pelo amparo, pelo riso, por fazerem com que as coisas fossem mais leves.

Ao Felipe e à Kerstin por mostrar que existe amor em SP.

Ao LaPAM, Alexandre, Giovanni, Camilla e Luan.

Por último, e mais importante, agradeço àqueles que dão todo peso ao real significado da palavra família. Aos meus pais, Adalci e Benedito, por fazerem o impossível em diferentes situações para me dar oportunidades que lhes foram negadas. São guerreiros a me ensinar. Meus amores, minha bênção.

Às minhas irmãs, Allana e Jéssica, por serem espelhos, referências. Vocês são a vida minha da minha vida. São pedaços de mim passeando pelo mundo.

À Dona Nice, Seu Adalberto e Adenir. À todxs aqueles que vieram antes de mim e abriram caminhos, tornando possível que alguém como eu ocupasse um lugar como este. Chegar aqui é desafiar a lei a gravidade.

"A ciência e a vida cotidiana não podem e não devem ser separadas."

Rosalind Franklin

RESUMO

SILVA-PEREIRA, T. T. Sequenciamento, anotação e análise do genoma de *Mycobacterium pinnipedii* e genômica comparativa de espécies do Complexo *Mycobacterium tuberculosis*. [Sequencing, annotation and analysis of the *Mycobacterium pinnipedii* genome and comparative genomics of *Mycobacterium tuberculosis* complex species]. 2019. 155 f. Dissertação (Mestrado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2019.

O Complexo Mycobacterium tuberculosis (MTBC) causa tuberculose em humanos e animais e é composto de 12 espécies bacterianas com diferentes tropismos por hospedeiros e perfis de virulência. O MTBC se originou na África, e evoluiu clonalmente por meio de mutações pontuais e deleções genéticas de até 12Kb. Apesar das variações fenotípicas, sequências homólogas de genomas do MTBC são 99,95% idênticas e transferências horizontais de genes são consideradas ausentes. Estudos compreensivos de genômica comparativa incluindo todas as espécies de MTBC a fim de identificar os determinantes genéticos dessas diferenças fenotípicas não estão disponíveis. Paralelamente, a recente descoberta de uma estirpe relacionada a Mycobacterium pinnipedii infectando múmias peruanas pré-colombianas, isto é, antes da introdução europeia da tuberculose nas Américas, levantou dúvidas quanto ao papel do M. pinnipedii na evolução do MTBC e sobre a co-evolução desses patógenos com populações humanas. Desta forma, esta dissertação tem como objetivos: (i) sequenciar e anotar os genomas completos de estirpes de M. pinnipedii isoladas de um leão marinho no Brasil e compará-los com outras estirpes da mesma espécie depositadas em bancos de dados públicos; (ii) produzir uma avaliação de genômica comparativa de todos as espécies descritas do MTBC a fim de identificar genes relacionados à virulência e à adaptabilidade ao hospedeiro. No primeiro estudo, foi realizado o sequenciamento dos genomas completos de dois isolados de M. pinnipedii obtidos da carcaça de um leão marinho (Otaria flavescens) encontrada no Rio Grande do Sul. Análises comparativas das mutações encontradas nesses genomas indica que esse animal estava infectado por duas estirpes diferentes desta bactéria. Este achado constitui o primeiro relato de infecção mista por M. pinipedii nesta espécie hospedeira e sugere que o patógeno é altamente endêmico na população de origem deste animal. Em contraste com estirpes de M. tuberculosis, a genômica comparativa entre estirpes modernas e antigas de M. pinnipedii depositadas em bancos de dados públicos indica proteomas altamente conservados e a ocorrência de um decaimento gênico através de deleções e pseudogenização, em um processo ativo de redução do genoma que vem ocorrendo há, pelo menos, 1.000 anos. Por fim, duas linhagens filogenéticas, modernas de M. pinnipedii foram detectadas globalmente, circunscritas por geografia e, consequentemente, por espécies hospedeiras. No segundo estudo, o objetivo foi investigar a presença e conservação de epítopos de células T e fatores de virulência no genoma core, acessório e espécie específico de 205 estirpes do MTBC. A estrutura do pan-genoma do MTBC mostra que a maioria das proteínas (832.234/866.916, 96,00%) está no genoma core, que é composto por proteínas importantes para a sobrevivência bacteriana. A análise da arquitetura do genoma acessório de MTBC com a distribuição dos epítopos mostrou, pela primeira vez, um mecanismo de variação antigênica de epítopos de células T baseado na perda de genes. Adicionalmente, foi possível observar variações significativas na quantidade de genes relacionados aos fatores de virulência, principalmente em estirpes de M. pinnipedii e M. africanum, sugerindo que a perda gênica relacionada à virulência possui importante papel nos fenótipos bacterianos. Em conclusão, os resultados apresentados contribuem para o entendimento da interação patógeno-hospedeiro de M. pinnipedii e outros membros do MTBC, sugerindo marcadores genéticos importantes dessa interação, e elucidando efeitos de pressões seletivas na determinação dos fenótipos de adaptabilidade hospedeira e virulência do grupo clonal MTBC.

Palavras-chave: Complexo *Mycobacterium tuberculosis*, genoma, *Mycobacterium pinnipedii*, fatores de virulência, epítopos.

ABSTRACT

SILVA-PEREIRA, T. T. Sequencing, annotation and analysis of the *Mycobacterium* pinnipedii genome and comparative genomics of *Mycobacterium tuberculosis* complex species. 2019. 155 f. Dissertação (Mestrado em Ciências) — Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2019.

The Mycobacterium tuberculosis Complex (MTBC) causes tuberculosis in humans and animals and is composed of 12 bacterial species with different host tropisms and virulence profiles. MTBC originated in Africa and evolved clonally through point mutations and genetic deletions of up to 12Kb. Despite phenotypic variations, homologous sequences of MTBC genomes are 99.95% identical and horizontal gene transfers are considered absent. Comprehensive comparative genomics studies including all MTBC species to identify the genetic determinants of these phenotypic differences are not available. At the same time, the recent discovery of a strain related to Mycobacterium pinnipedii infecting pre-Columbian Peruvian mummies, i.e. prior to the European introduction of tuberculosis in the Americas, raised doubts about the role of M. pinnipedii in the evolution of MTBC and the co-evolution of these pathogens with human populations. Thus, the aims of this study were to: (i) sequence and annotate the complete genomes of M. pinnipedii strains isolated from a sea lion in Brazil and compare them with other strains of the same species deposited in public databases; (ii) produce a comparative genomic evaluation of all described MTBC species to identify genes related to virulence and host adaptability. In the first study we sequenced the complete genomes of two M. pinnipedii isolates obtained from the carcass of a sea lion (Otaria flavescens) found in Rio Grande do Sul. Comparative analysis of mutations found in these genomes indicates that this animal was infected with two different strains of this bacterium. This finding is the first report of mixed M. pinipedii infection in this host species and suggests that the pathogen is highly endemic in the source population of this animal. In contrast to M. tuberculosis strains, the comparative genomics of ancient and modern M. pinnipedii strains deposited in public databases indicate highly conserved proteomes and the occurrence of gene decay through deletions and pseudogenization in an active process of genome reduction occurring for at least 1,000 years. Finally, two modern phylogenetic strains of M. pinnipedii were detected globally, circumscribed by geography and, consequently, by host species. In the second study, the aim was to investigate the presence and conservation of T cell epitopes and virulence factors in the core, accessory and strain-specific genomes of 205 MTBC strains. The MTBC pan-genome structure shows that most proteins (832,234/866,916, 96.00%) are in the core genome, which is composed of proteins important for bacterial survival. Analysis of the MTBC accessory genome architecture with epitope distribution showed, for the first time, a mechanism of antigenic T cell epitope variation based on gene loss. Additionally, it was possible to observe significant variations in the number of genes related to virulence factors, mainly in M. pinnipedii and Mycobacterium africanum strains, suggesting that virulencerelated gene loss plays an important role in determining bacterial phenotypes. In conclusion, the results presented contribute to the understanding of the pathogen-host interaction of M. pinnipedii and other MTBC members, suggesting important genetic markers of this interaction, and elucidating the effects of selective pressures in the determination of host adaptability and virulence phenotypes of the clonal group MTBC.

Keywords: *Mycobacterium tuberculosis* complex, genome, *Mycobacterium pinnipedii*, virulence factors, epitope

LISTA DE FIGURAS

CAPÍTULO 1

- Figura 1 Mapa circular do proteoma dos genomas do *Mycobacterium pinnipedii* comparado ao genoma de referência do *Mycobacterium tuberculosis* H37Rv. Do anel externo ao anel interno: posição do genoma, *M. tuberculosis* H37Rv, *M. pinnipedii* MP1, *M. pinnipedii* MP2, *M. pinnipedii* G01222, *M. pinnipedii* G01491, *M. pinnipedii* G01492, *M. pinnipedii* G01498, *M. pinnipedii* ATCC BAA-688. Escala de cores representa identidade de proteína. Contigs foram ordenados usando MAUVE (RISSMAN et al., 2009) (Rissman et al., 2009) e o mapa gerado em recursos da web Patric 3.5.18 (WATTAM et al., 2017).
- Figura 2 Rede de interação de proteínas em regiões de CDS (Sequências de DNA codificante) com função conhecida contendo SNPs (polimorfismo de nucleotídeo único) não-sinônimos. CDSs com mutações foram identificadas entre as linhagens MP1 e MP2 de *Mycobacterium pinnipedii*. A rede foi gerada usando o banco de dados STRING, com parâmetros padrão.
- Figura 3 Agrupamentos de proteínas ortólogas de estirpes de *Mycobacterium pinnipedii*.

 (A) Diagrama de Venn de proteínas ortólogas. (B) *Cluster of Orthologous Groups* (COG) com base na função dos clusters de proteínas do genoma *core* das estirpes de *M. pinnipedii*. (C) classificação COG sem classe de função desconhecida (S). O agrupamento de proteínas foi realizado usando OrthoMCL como disponível na plataforma Kbase (ARKIN et al., 2018). * O número de proteínas gênicas ortólogas compartilhadas entre esses grupos está disponível no Apêndice C. O tamanho do cluster variou de duas a quarenta e três proteínas.
- Figura 4 Polimorfismos de sequência grande (PSGs) de estirpes de *Mycobacterium pinnipedii*. Os valores indicam o número de nucleotídeos que abrangem cada região excluída (zero significa que a região não é excluída na linhagem bacteriana correspondente). Início e fim: posições de nucleotídeos de acordo com o genoma de referência de *M. tuberculosis* H37Rv. Os genes são anotados de acordo com o genoma de referência. * Antigas estirpes sul-americanas de *M. pinnipedii*. Deleções de cinza são aquelas que já foram descritas em estirpes de *M. pinnipedii*.
- Figura 5 Árvore filogenética baseada em SNPs (polimorfismo de nucleotídeo único) do Complexo *Mycobacterium tuberculosis* (MTBC) usando o modelo de máxima verossimilhança (ML). Caixa verde: cluster moderno 1; caixa laranja: cluster moderno 2; caixa rosa: cluster antigo; círculo preto: deleções MiD3 e MiD4; triângulo vermelho: deleção RD2seal. Mycobacterium tuberculosis H37Rv foi usado como grupo externo. Um script customizado em Python foi usado para construir uma matriz de SNPs, que foi usada para inferir uma árvore ML usando o RAxML com o modelo GTRCAT e o autoMRE para a árvore ML de melhor pontuação e um máximo de 1.000 inferências de bootstrap usando o software RaxML. (STAMATAKIS, 2014).

CAPÍTULO 2

- Figura 1 Distribuição gráfica do pan-genoma de 205 estirpes do Complexo *Mycobacterium tuberculosis*. A: Percentual de grupos de proteínas de genes ortólogos. B: Percentual de número de proteínas. O pan-genoma pode ser dividido em genoma *core* (proteínas comuns a 95% das estirpes), genoma acessório (proteínas presentes em 2 a 204 estirpes) e genoma espécie-específico (proteínas presentes apenas em uma estirpe).
 - Figura 2 Cluster of Orthologous Groups (COG) baseada na função de proteínas agrupadas. As categorias de COG são: D: Controle do ciclo celular, divisão celular, partição de cromossomos; M: Biogênese da parede celular / membrana / envelope; N: motilidade celular; O: modificação pós-translacional, turnover de proteína e acompanhantes; T: Mecanismos de transdução de sinal; U: Tráfico, secreção e transporte vesicular intracelular; V: Mecanismos de defesa; Z: citoesqueleto; A: Processamento e modificação de RNA; J: Tradução, estrutura ribossômica e biogênese; K: Transcrição; L: Replicação, recombinação e reparo; C: Produção e conversão de energia; E: Transporte e metabolismo de aminoácidos; F: Transporte e metabolismo de nucleotídeos; G: Transporte e metabolismo de carboidratos ; H: transporte e metabolismo de co-enzima; I: transporte lipídico e metabolismo; P: Transporte e metabolismo de íons inorgânicos; Q: Biossíntese, transporte e catabolismo de metabólitos secundários; S: Função desconhecida e X: componentes do mobiloma.
 - Figura 3 Rede de interações proteína-proteína com representantes da classe funcional de Transporte e Metabolismo de Carboidratos do genoma *core* do Complexo *Mycobacterium tuberculosis*. A rede foi gerada usando o banco de dados STRING, com parâmetros padrão.
 - Figura 4 Rede de interações proteína-proteína com representantes da classe funcional de Biossíntese, transporte e catabolismo de metabólitos secundários (Q) do genoma acessório do Complexo *Mycobacterium tuberculosis*. A rede foi gerada usando o banco de dados STRING, com parâmetros padrão.
 - Figura 5 Distribuição do número de epítopos em proteínas do genoma *core* (A), genoma acessório (B) e Genoma Espécie-Específico (C) de espécies do Complexo *Mycobacterium tuberculosis*. O eixo x refere-se ao número de epítopos encontrados e o eixo y ao número de proteínas.
 - Figura 6 Taxa de variação de epítopos preditos de proteínas do genoma *core* e genoma acessório de espécies do MTBC. Em azul estão os índices de epítopos conservados e em vermelho os de epítopos variantes.
 - Figura 7 Alinhamento de epítopos preditos de proteínas do genoma *core* de espécies do MTBC. A figura foi criada usando o WebLogo (CROOKS et al., 2004).

Figura 8 - *Heatmap* apresentando fatores de virulência de espécies do Complexo *Mycobacterium tuberculosis*. A escala gráfica de cor varia do vermelho ao verde, nos quais quanto mais próximo do verde maior a quantidade de genes relacionados àquele fator de virulência nos determinados genomas. As barras coloridas abaixo indicam as linhagens de cada espécie bacteriana. Azul claro: Linhagem 1 (L1) *M. tuberculosis*; azul: linhagem 2 (L2) *M.tuberculosis*; azul escuro: linhagem 4 (L4) *M. tuberculosis*; cinza: linhagem 5 (L5) *Mycobacterium africanum*; preto: linhagem 6 (L6) *M. africanum*; amarelo: linhagem 1 (L1) *M. bovis*; laranja: linhagem 2 (Lb2) *M. bovis*; vermelho: linhagem 3 (Lb3) *M. bovis*.

* Bactérias isoladas em animais selvagens.

LISTA DE ABREVIATURAS E SIGLAS

ACGs - acr-coregulados

ANN - "rede neural artificial"

AT - Toxina-antitoxina

ATP - Adenosina trifosfato

BWA - Burrows-Wheeler

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior,

CDS - Sequências de DNA codificante

CEFAP - Centro de Facilidades de Apoio à Pesquisa

CNPq - Conselho Nacional de Pesquisa Científica

CoA - Coenzima A

COG - Cluster of Orthologous Groups

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

FVs - Fatores de virulência

GDP - Guanosina difosfato

GPLs - Glicopeptidolipídios

HLA - Human leukocyte antigen

Indels - Inserções e deleções

LaCTAD - Laboratório Central de Tecnologias de Alto Desempenho

MCE - Mammalian cell entry

MCL - Markov Cluster

MHC - Complexo principal de histocompatibilidade

MIRUs - Unidades micobacterianas repetitivas intercaladas

ML - Máxima verossimilhança (Maximum Likelihood)

MmpL - Mycobacterial membrane protein Large

mRNA - Ácido ribonucleico mensageiro

MTBC - Complexo Mycobacterium tuberculosis

NCBI - National Center for Biotechnology Information

NO - Óxido nítrico

OMS - Organização Mundial de Saúde

ONU - Organização das Nações Unidas

pb - Pares de bases

PE - Prolina-glutamato

PGAP - Prokaryotic Genome Annotation Pipeline

PKS - Policetídeos sintase

PPE - Prolina-prolina-glutamato

PSGs - Polimorfismos de sequência grande

PSGs - Polimorfismos de sequência grande

RDs - "regions of difference"

RNA - ácido ribonucleico

SL-1 - Sulfolipídio-1

SNPs - Polimorfismo de nucleotídeo único

SRA - Sequence Read Archive

TB - Tuberculose

tRNA - Ácido ribonucleico transportador

UNICAMP - Universidade Estadual de Campinas

VFDB - Banco de Dados de Fatores de Virulência

VNTRs - Número variável de repetições em tandem

SUMÁRIO

1.	INTRODUÇÃO	18
2.	REVISÃO BIBLIOGRÁFICA	21
2.1.	TUBERCULOSE E SEUS HOSPEDEIROS	21
2.2.	EVOLUÇÃO DO MTBC	23
2.3. HO	ASSINATURA GENÉTICAS DE ADAPTABILIDADE BACTERIANA AO SPEDEIRO	25
2.4.	GENÔMICA COMPARATIVA DO MTBC	26
3.	OBJETIVOS	29
3.1.	OBJETIVOS GERAIS	29
3.2.	OBJETIVOS ESPECÍFICOS	29
	CAPÍTULO 1: Sequenciamento do genoma de estirpes de <i>Mycobacterium pinniped</i> acterização genética completa e evidência de superinfecção em um leão-marinho sul-cricano (<i>Otaria flavescens</i>)	
RES	SUMO	30
4.1.	INTRODUÇÃO	32
4.2.	MATERIAL E MÉTODOS	34
4.2.	1. Seleção de genomas	34
4.2.	2. Isolamento e extração de DNA bacteriano	34
4.2.	3. Sequenciamento de isolados de M. pinnipedii	35
4.2.	4. Montagem e anotação do genoma	35
4.2.	5. Mapa circular dos genomas de M. pinnipedii	35
4.2.	6. Mapeamento e identificação de variantes	36
4.2.	7. Redes de interação de proteínas	36
4.2.	8. Spoligotyping in silico e Número Variável de Repetição em Tandem (VNTR	36. (1
4.2.	9. Agrupamentos de proteínas de genes ortólogos e parálogos de M. pinnipedii	37
4.2.	10. Polimorfismos de sequência grandes	37
4.2.	11. Análise filogenética	38
4.3.	RESULTADOS E DISCUSSÃO	39
4.3.	1. Primeira descrição de infecção mista por M. pinnipedii em um pinípede	39
4.3.	2. Estirpes de <i>M. pinnipedii</i> têm um proteoma altamente conservado	42
4.3.	3. As estirpes modernas de <i>M. pinnipedii</i> têm marcadores únicos de deleção	45
4.3.	4. Dois agrupamentos distintos de estirpes modernas de <i>M. pinnipedii</i>	48
4.4.	CONSIDERAÇÕES FINAIS	49

5. CAPÍTULO 2: Influência da arquitetura gênica na distribuição e conservação de	
epítopos lineares de células T e fatores de virulência do Complexo Mycobacterium	
tuberculosis	
RESUMO	
5.1. INTRODUÇÃO	
5.2. MATERIAL E MÉTODOS	
5.2.1. Seleção dos genomas de MTBC	
5.2.2. Construção dos genomas core, acessório e estirpe-específico	
5.2.3. Predição de epítopos lineares	64
5.2.3.1. Distribuição e conservação de epítopos lineares	65
5.2.4. Fatores de virulência	65
5.3. RESULTADOS E DISCUSSÃO	66
5.4. CONSIDERAÇÕES FINAIS	82
Referências	84
6. CONSIDERAÇÕES FINAIS	97
REFERÊNCIAS1	00
APÊNDICE A – SNPs que diferenciam as estirpes MP1 e MP2 de M. pinnipedii1	24
APÊNDICE B – Padrão de MIRU-VNTR e <i>Spoligotyping</i> das estirpes MP1 e MP2 de <i>Mycobacterium pinnipedii</i> 1	32
APÊNDICE C – Agrupamentos de proteínas de genes ortólogos entre estirpes de <i>Mycobacterium pinnipedii</i> 1	32
APÊNDICE D – Lista do conjunto de genomas do Complexo <i>Mycobacterium tuberculosis</i> (MTBC) utilizados neste estudo, nome da estirpe, número de acesso do RefSeq do NCBI e linhagem bacteriana	35
APÊNDICE E – Epítopos preditos das proteínas do genoma acessório do Complexo Mycobacterium tuberculosis encontrados no IEDB	40
APÊNDICE F - Lista dos grupos de fatores de virulência e genes relacionados seguindo a ordem no heatmap	47

1. INTRODUÇÃO

A tuberculose (TB) é uma doença infectocontagiosa primariamente pulmonar, podendo atingir outros órgãos, considerada um grave problema de saúde pública. Segundo o último relatório global da TB, a doença causou 1,6 milhões de mortes e foi reportada em 10 milhões de pessoas (entre 9 a 11,1 milhões) somente em 2017 (WHO 2018). A TB tem como agentes etiológicos um grupo de bactérias álcool-ácido resistentes conhecido como Complexo Mycobacterium tuberculosis (MTBC), cujo principal microrganismo, o bacilo Mycobacterium tuberculosis, foi identificado como causador da TB por Robert Koch em 1882. O MTBC são microrganismos intracelulares facultativos, aeróbicos ou microaerófilos, de crescimento lento e com alta similaridade genética (SMITH et al., 2006; DOS VULTOS et al., 2008; GALAGAN, 2014; KIESER; RUBIN, 2014). Membros deste Complexo possuem como principal célula alvo os macrófagos alveolares, mas podem ser fagocitados também por neutrófilos e células dendríticas quando entram em contato com os alvéolos pulmonares (WORBOYS, 1990; DORHOI; KAUFMANN, 2015; LERNER; BOREL; GUTIERREZ, 2015). O Complexo compreende os principais agentes da TB humana, Mycobacterium tuberculosis e Mycobacterium africanum, além de membros que tem como hospedeiros várias espécies de animais domésticos e selvagens, sendo eles Mycobacterium bovis, Mycobacterium caprae, Mycobacterium pinnipedii, Mycobacterium microti, Mycobacterium orygis, Mycobacterium suricattae, Mycobacterium mungi, "dassie bacillus" e "chimpanzee bacillus" (COSCOLLA; GAGNEUX, 2014).

As espécies e linhagens bacterianas do MTBC evoluem clonalmente e compartilham 99,9% de identidade nas sequências de DNA em regiões alinháveis (BRITES; GAGNEUX, 2015). Essa evolução clonal, a partir de um ancestral comum com *Mycobacterium canettii*, ocorre apenas por mutações nucleotídicas pontuais, deleções de regiões de 2 a 12.7 Kb denominadas de *regions of difference* (RDs), elementos de inserção (IS – *insertion sequences*) e duplicação de um número limitado de famílias de genes parálogos. Transferências horizontais de genes e grandes recombinações são consideradas ausentes. Atualmente, existem duas teorias quanto à origem evolutiva do MTBC, mas que possuem consenso em afirmar que o complexo surgiu no Norte e Leste da África. A primeira estimativa data a origem do MTBC em 70 mil anos atrás, com a disseminação do patógeno pelo mundo por meio da migração de hominídeos para fora do continente Africano (COMAS et al., 2013). A segunda estimativa, mais aceita (MENARDO et al., 2019; SABIN et al., 2019), indica que o MTBC surgiu entre 4.000 e 5.000 anos atrás (BOS et al., 2014; O'NEILL et al., 2019; SABIN et al., 2019).

No Brasil, *M. tuberculosis* e *M. bovis* são as espécies do MTBC mais comumente encontradas. Estirpes de *M. tuberculosis* circulantes atualmente nas Américas são filogeneticamente próximas às estirpes circulantes na Europa. Por este motivo, acredita-se que a introdução desse patógeno no continente tenha ocorrido após contato com os Europeus. Porém, a existência de lesões ósseas compatíveis com TB em múmias peruanas précolombianas colocava em xeque essa afirmação. Utilizando técnicas especiais para recuperação de DNA antigo, o genoma da bactéria causando tais lesões foi recentemente sequenciado (BOS et al., 2014). Surpreendentemente, o mesmo mostrou-se filogeneticamente próximo ao patógeno de focas e leões marinhos, *M. pinnipedii*, incitando dúvidas quanto ao papel deste patógeno na evolução e dispersão do MTBC.

Estirpes de *M. pinnipedii* têm sido isoladas de espécies de pinípedes restritas ao hemisfério sul, sendo primeiramente detectadas em leões marinhos australianos (*Neophoca cinerea*) e da Nova Zelândia (*Arctocephalus forsteri*), tanto em animais de cativeiro quanto de vida livre (COUSINS, 1987; BASTIDA et al., 1999). Além de pinípedes, a TB causada por *M. pinnipedii* já foi diagnosticada em outras oito espécies de animais em zoológicos, bovinos e humanos (COUSINS, 1987; BASTIDA et al., 1999; KRIZ et al., 2011; BOARDMAN et al., 2014; LOEFFLER et al., 2014). A capacidade de *M. pinnipedii* em causar doença em outras espécies hospedeiras ressalta a importância desse micro-organismo como um patógeno multi-hospedeiro (i.e. possivelmente generalista), mas de impacto desconhecido em espécies não-pinípedes. No Brasil, a carcaça de um leão marinho foi recentemente encontrada na praia de Capão da Canoa, no Rio Grande do Sul. Como o Brasil não possui colônias reprodutivas desses animais, espécimes encontradas no litoral são provavelmente oriundas do Uruguai. Esse animal apresentava lesões compatíveis com TB de onde o *M. pinnipedii* foi isolado, constituindo o primeiro relato no litoral brasileiro (DE AMORIM et al., 2014). Naquele momento, os genomas dos isolados não foram sequenciados, impedindo sua caracterização genética completa.

A alta similaridade genética dos membros do MTBC e a ausência de transferência horizontal de genes contrasta com os fenótipos variados de tropismo por hospedeiros e virulência. Entretanto, foi evidenciado que SNPs (polimorfismos de nucleotídeos únicos) e indels (inserções e deleções) podem gerar tanta variabilidade gênica (por meio do truncamento e deleção de genes, i.e. perda gênica) em patógenos clonais quanto àquela observada em patógenos sujeitos a transferência horizontal de genes (BOLOTIN; HERSHBERG, 2015). Para explorar essas variações, deve-se analisar o conteúdo gênico através da construção do pangenoma, genoma *core* e genoma acessório. Porém, a grande maioria dos estudos em MTBC é baseada somente na identificação e avaliação de SNPs e indels, sem considerar seus efeitos no

conjunto gênico (SALAMON et al., 2000; KATO-MAEDA et al., 2001; GAO et al., 2005; ISAZA et al., 2012; DEJESUS; SACCHETTINI; IOERGER, 2013; SUPPLY et al., 2013; LIU et al., 2014; DIPPENAAR et al., 2015; PERIWAL et al., 2015; JIA et al., 2017; XIONG et al., 2017; YANG et al., 2018). Além do mais, não existem estudos sobre o conteúdo gênico que incluam todos as espécies e linhagens conhecidas do MTBC. As consequências da perda gênica no repertório global de fatores de virulência e na expressão de antígenos contendo epítopos de células T em MTBC, por exemplo, são desconhecidas. Assim, acreditamos que estudos sobre a arquitetura gênica do complexo poderão fornecer importantes informações a respeito da adaptabilidade hospedeira e virulência dos patógenos do MTBC.

Nesta dissertação, dois estudos serão apresentados. O primeiro envolve o sequenciamento e análise dos genomas completos de dois isolados de *M. pinnipedii* obtidos da carcaça do leão-marinho (*Otaria flavescens*) encontrado no Rio Grande do Sul e sua comparação genômica com outras estirpes da mesma espécie depositadas em bancos de dados públicos, constituindo o primeiro estudo compreensivo da caracterização genética deste patógeno. O segundo estudo envolve uma análise do conteúdo gênico (pan-genoma, genoma *core* e genoma acessório) de 205 genomas do MTBC, incluindo o maior número de espécies já sequenciadas do complexo. Tal conteúdo gênico foi avaliado, de maneira comparativa, quanto às suas funções, conservação de epítopos de células T e repertório de fatores de virulência, de maneira a gerar informações a respeito das diferenças gênicas subjacentes aos distintos fenótipos do MTBC.

2. REVISÃO BIBLIOGRÁFICA

2.1.TUBERCULOSE E SEUS HOSPEDEIROS

A TB tem feito vítimas durante grande parte da história da humanidade, com registros arqueológicos do Egito antigo datados entre 3.500 e 2.650 a.C (ZINK et al., 2001, 2003a, 2003b, 2007). Atualmente, a doença é responsável pelo maior índice de mortalidade humana causada por um único agente infeccioso no mundo. A TB tem incidência anual de 10 milhões de casos e 1,6 milhões de mortes globalmente, sendo o Brasil o 17º país colocado no ranking da doença (WHO, 2018). Em 1993, a enfermidade foi declarada pela Organização Mundial de Saúde (OMS) como questão de urgência à saúde pública global, por representar 26% das mortes previsíveis e 7% de todas as mortes no mundo, sendo que acabar com a epidemia mundial até 2030 faz parte dos objetivos de desenvolvimento sustentável da Organização das Nações Unidas (ONU). A TB pode também afetar os animais, sendo que os bovinos e certas espécies de animais selvagens de cativeiro e vida livre são os que mais sofrem com a doença. Rebanhos bovinos infectados geram sérios prejuízos econômicos e sociais aos pequenos e grandes produtores, principalmente em países em desenvolvimento onde os programas de controle da enfermidade são pouco eficientes. A TB em animais selvagens pode também ser devastadora, constituindo uma das enfermidades de mais difícil controle para medicina da conservação.

Apesar dos índices relacionados à TB terem diminuído em todo mundo, as estratégias utilizadas não parecem suficientes para alcançar as primeiras metas (em 2020) da campanha global da OMS "End TB Strategy". Os objetivos são que em 2020 a taxa de incidência da doença (casos/100.000 habitantes) caia 4-5% por ano e a taxa de mortalidade seja de 10% na população doente. Em 2017, a proporção de mortes por TB foi de 16% e a média global de declínio na taxa de incidência foi de 2%, não sendo suficiente para atingir a meta de redução. Além disso, casos de TB resistentes aos medicamentos tem chamado a atenção, com 558.000 novos casos de TB multirresistente ou TB resistente a rifampicina (MDR/RR-TB) estimados em 2017 e uma taxa relativamente baixa de sucesso do tratamento, com 55% em todo o mundo. Para diminuir o ônus da doença e fazer com que a estratégia para o fim da TB como epidemia e um grande problema de saúde pública torne-se viável, é necessário fechar lacunas no diagnóstico e tratamento da doença e também fortalecer estratégias para o gerenciamento de casos de TB resistente aos antimicrobianos (WORLD HEALTH ORGANIZATION, 2018).

A TB humana é causada pelo *Mycobacterium tuberculosis*, uma bactéria altamente adaptada a humanos, mas que já foi identificada em uma grande variedade de espécies hospedeiras, incluindo primatas não-humanos, elefantes, cães, gatos, gado, papagaio, antas,

gambás, carneiros, porcos, cabras, camundongos, texugos, roedores, focas, rinocerontes dentre outros animais selvagens e domésticos (CORPER, 1947; COUSINS et al., 1994; PARRA et al., 2003; ALEXANDER et al., 2010; COSCOLLA et al., 2013; DE GARINE-WICHATITSKY et al., 2013; BOARDMAN et al., 2014). Entretanto, a doença em animais ocorreu na maior parte dentro de ambiente de cativeiro ou doméstico, com os animais servindo de hospedeiros acidentais. A persistência do patógeno em populações de mamíferos de vida livre não ocorre, sendo assim, não existem reservatórios animais estabelecidos de *M. tuberculosis*. Por outro lado, a TB bovina, causada por *Mycobacterium bovis*, é amplamente disseminada por diversas espécies hospedeiras, sendo as espécies pertencentes a família Bovidae os hospedeiros mais afetados (ASHFORD et al., 2001; MONTALI et al., 2001). Humanos podem também se infectar com *M. bovis* (TB zoonótica), sendo essa uma forma negligenciada da TB nesta população, com estimados 145 mil casos e 12.500 mortes anualmente (WORLD HEALTH ORGANIZATION, 2018). Animais selvagens infectados com *M. bovis* podem servir como reservatórios e fonte de infecção de TB aos humanos, bovinos e outros animais selvagens, tornando-se um problema tanto na saúde pública quanto na saúde animal e conservação.

As espécies de micobactérias tuberculosas fazem parte do complexo Mycobacterium tuberculosis (MTBC). Esse complexo, além de conter os causadores da TB humana e animal, M. tuberculosis e M. bovis, respectivamente, possuem mais outras 10 espécies geneticamente relacionadas que causam, em sua maioria, TB em diversas espécies de animais selvagens e podem infectar humanos. Os membros MTBC têm diferentes fenótipos de adaptação hospedeira e patogenicidade, sendo influenciadas pela espécie animal infectada. Dentre esses membros, M. tuberculosis e Mycobacterium africanum são adaptados aos seres humanos, enquanto que M. bovis, Mycobacterium caprae, Mycobacterium microti, Mycobacterium pinnipedii, Mycobacterium orygis, Mycobacterium mungi, Mycobacterium suricattae, "dassie bacillus", e "chimpanzee bacillus" são adaptados aos animais (COUSINS et al., 1994, 2003; COSCOLLA; GAGNEUX, 2014). As diferentes espécies de micobactérias "animais" ganham seu nome de acordo com o hospedeiro que foram primeiramente identificadas ou que ocorrem com maior frequência. Com exceção de M. bovis, M. caprae, M. pinnipedii e M. orygis, as outras micobactérias animais infectam animais selvagens restritos ao continente africano, como suricato (M. suricatte), hyrax ("dassie bacillus"), mangusto (M. mungi) e chimpanzés ("chimpanzee bacillus"). Definições sobre tropismo por espécies hospedeiras e virulência são limitadas devido a escassez de estudos com essas espécies bacterianas.

2.2. EVOLUÇÃO DO MTBC

Atualmente, é sabido que as micobactérias tuberculosas se disseminaram pelo mundo por meio das migrações humanas e evoluíram para infectar outros animais (GALAGAN, 2014). É de amplo conhecimento que o MTBC possui uma evolução clonal, com nenhuma evidência de transferência horizontal de genes ou grandes recombinações (HIRSH et al., 2004). O complexo compreende espécies e linhagens bacterianas compartilhando 99,95% de identidade de sequência de DNA em regiões alinháveis (BRITES; GAGNEUX, 2015). Essa evolução clonal, a partir de um ancestral comum com Mycobacterium canettii, ocorreu então apenas por mutações nucleotídicas pontuais, deleções de regiões de 2 a 12.7 Kb denominadas de regions of difference (RDs), elementos de inserção e duplicação de um número limitado de famílias de genes parálogos. Porém, mesmo com essa grande similaridade genética, as estirpes de MTBC apresentam grandes variações no fenótipo, incluindo a interação com o hospedeiro, virulência e imunorreatividade, que levam a distinção das mesmas em diferentes espécies bacterianas pelos taxonomistas. Uma ampla gama de mutações são descritas em genomas do MTBC, incluindo SNPs (polimorfismos de nucleotídeo único), deleções, duplicações e elementos móveis, todos os quais podem ter impacto em fenótipos bacterianos (COSCOLLA; GAGNEUX, 2014). Essas variações podem ser o resultado da co-evolução a longo prazo com diversas populações de seres humanos e animais, relacionando genótipos restritos a diferentes regiões geográficas com os diferentes fenótipos encontrados, levando assim a tropismos distintos por hospedeiros e variações em persistência populacional (DUNN; NORTH, 1995; RENWICK; WHITE; BENGIS, 2007; JIA et al., 2017).

A co-evolução entre patógeno e hospedeiro é caracterizada por mudanças adaptativas recíprocas nas espécies que interagem (WOOLHOUSE et al., 2002). Acredita-se que as espécies que compõem o MTBC têm sido associadas aos seus hospedeiros há muito tempo, e a ocorrência de co-evolução entre *M. tuberculosis* e humanos tem sido sugerida (GAGNEUX, 2012; BRITES; GAGNEUX, 2015). Assim, mudanças na demografia humana poderiam ter desempenhado um importante papel na evolução das linhagens de MTBC, levando à identificação de ao menos 7 linhagens distintas de *M. tuberculosis* e *M. africanum* (*M. tuberculosis* L1 a L4 e L7, entre estirpes "modernas" e "antigas", e *M. africanum* L5 e L6). Essa linhagens são frequentemente encontradas associadas aos diferentes perfis genéticos das populações humanas, sendo estes os únicos hospedeiros nos quais as linhagens são capazes de manter o ciclo de infecção, incluindo desenvolvimento da doença e transmissão (ERNST, 2012). A distribuição dessas linhagens difere-se fortemente, algumas apresentando restrições geográficas e outras com distribuição global (GAGNEUX et al., 2006). A linhagem 1 (L1)

ocorre em toda região do Oceano índico; a linhagem 2 (L2) ocorre no Leste Asiático; a linhagem 3 (L3) apresenta predominância nas regiões Central e Sul da Ásia e Leste da África; a linhagem 4 (L4), também conhecida como linhagem euro-americana, tem ampla distribuição global. As linhagens 5 e 6 (*M. africanum*) são restritas ao Leste da África (DE JONG; ANTONIO; GAGNEUX, 2010) e a linhagem 7 ocorre apenas na Etiópia (FIRDESSA et al., 2013). Acreditase que a maior expansão geográfica das linhagens "modernas" (L2, L3 e L4) pode ser explicada por ganhos evolutivos das mesmas em relação às linhagens "antigas" (L1, L5, L6, L7).

Hipóteses sobre a origem e história evolutiva do MTBC foram testadas usando filogenias calibradas pelo tempo (COMAS et al., 2013; PEPPERELL et al., 2013; BOS et al., 2014); contudo, os resultados obtidos nesses estudos se diferem quanto a possível datação da diversificação desses micro-organismos. Um estudo conduzido por Comas et al. (2013) sugeriu co-divergência entre populações humanas e M. tuberculosis e M. africanum, comparando diversidades genômicas dessas espécies e de haplogrupos mitocondriais de populações humanas através de análises filogenéticas. Associado ao sequenciamento completo da primeira estirpe de M. bovis (GARNIER et al., 2003), esse estudo indica que, ao contrário da ideia inicial, M. tuberculosis não tem origem animal, e também mostra que a doença tem afetado os humanos antes mesmo da Transição Demográfica Neolítica, propondo tempo coalescente para linhagens de MTBC de 70.000 (entre 50.000-96.000 anos) como o mais provável (COMAS et al., 2013). Porém, outro estudo baseado no cromossomo Y humano (PEPPERELL et al., 2013) não obteve resultados para uma associação equivalente, sugerindo o surgimento do MTBC por volta de 2.190 anos atrás (entre 1.331–3.142 anos). Uma análise recente, que utilizou uma calibração independente baseada em genomas de M. pinnipedii extraídos de amostras de múmias peruanas do período pré-colombiano (BOS et al., 2014), também mostrou uma estimativa diferente da encontrada por Comas et al, 2013 (COMAS et al., 2013). As estimativas de Bos e colaboradores sugerem que o último ancestral de todas as linhagens MTBC atuais ocorreu há 4.000 à 5.000 anos atrás e que o ancestral comum de todas as linhagens modernas ocorreu entre 3.500 e 2.200 anos atrás (BOS et al., 2014). Devido à proximidade filogenética entre o DNA isolado das múmias e M. pinnipedii, os dados suportam a hipótese de uma introdução do MTBC no continente americano através dos pinípedes; porém os resultados não são suficientes para comprovar a disseminação deste patógeno pelo continente no período pré-colombiano.

2.3.ASSINATURA GENÉTICAS DE ADAPTABILIDADE BACTERIANA AO HOSPEDEIRO

Alguns patógenos bacterianos mais virulentos podem apresentar baixa diversidade genética. Não obstante, seus genomas mostram alguns sinais de degradação e remodelagem que estão frequentemente relacionados ao processo de especialização hospedeira — transição da bactéria do ambiente para o hospedeiro. Nessa busca pela adaptação, as bactérias patogênicas podem ser classificadas como especialistas (aquelas que infectam uma ou poucas espécies hospedeiras; e.g. *M. tuberculosis*) ou generalistas (aquelas que infectam várias espécies hospedeiras; e.g. *M. bovis*). Assinaturas genômicas associadas com a evolução de um generalista a um especialista incluem: rearranjos genômicos, transferências horizontais de genes e *decay* (decaimento) genômico (perda de genes e formação de pseudogenes) (BAUMLER; FANG, 2013). Como os dois primeiros estão ausentes no grupo clonal MTBC, o *decay* (decaimento) genômico, polimorfismos de nucleotídeo único (SNPs), inserções e deleções (Indels) são as forças que conduzem a diversidade genética nestas espécies (BOLOTIN; HERSHBERG, 2015).

As RDs, que são deleções de aproximadamente 2 a 12Kb do MTBC, fazem parte deste decay (decaimento) genômico, mas não explicam facilmente as variações em hospedeiro e virulência das espécies e linhagens do MTBC. Mais recentemente, SNPs em PhoPR, um regulador de fatores de virulência, foi descrito como responsável pela atenuação de M. bovis e M. africanum em macrófagos humanos e em um modelo de experimental de camundongo quando comparados ao M. tuberculosis (GONZALO-ASENSIO et al., 2014). Embora estas mutações tentam explicar a atenuação de M. africanum e porquê M. bovis não é transmitido entre os seres humanos, a diminuição da virulência não é sinônimo de baixa transmissibilidade (BEHR; GORDON, 2015), e múltiplos outros SNPs e modulações do conteúdo gênico poderiam desempenhar papéis distintos em fenótipos do MTBC. Neste sentido, um estudo recente mostrou que estirpes de M. tuberculosis, uma espécie especialista, possuem um pangenoma com uma grande proporção de pseudogenes no genoma acessório (BOLOTIN; HERSHBERG, 2015). A perda de genes impulsionada por pseudogenização ou truncamento gênico e seu papel nos fenótipos do MTBC, entretanto, não foram avaliados.

Mecanismos moleculares de especificidade bacteriana ao hospedeiro incluem aderência microbiana às células hospedeiras, replicação intracelular e fatores de virulência (FVs) (BAUMLER; FANG, 2013). Apesar de vários FVs terem sido descritos em *M. tuberculosis* (FORRELLAD et al., 2013; GOVENDER; RAMSUGIT; PILLAY, 2014), sua conservação nunca foi analisada dentro do grupo MTBC. Muitas das análises comparativas, até então

publicadas, foram focadas na comparação entre estirpes de *M. tuberculosis* virulentas e não virulentas; assim, ainda faltam estudos genômicos comparativos que abordem as variações da associação com diferentes hospedeiros, virulência e epítopos (JIA et al., 2017).

2.4. GENÔMICA COMPARATIVA DO MTBC

Os primeiros genomas de MTBC a serem sequenciados, *M. tuberculosis* H37Rv e *M. bovis* AF2122/97 (COLE et al., 1998; GARNIER et al., 2003), serviram como base para a definição de RDs que determinaram as espécies do MTBC. Ao longo dos anos, os genomas de outras espécies do complexo *M. tuberculosis* tornaram-se disponíveis para análises mais aprofundadas (BENTLEY et al., 2012; COSCOLLA et al., 2013; DE LA FUENTE et al., 2015; DIPPENAAR et al., 2015; ALEXANDER et al., 2016; WINGLEE et al., 2016; ZHU et al., 2016). Infelizmente, os genomas do *M. suricattae* e "chimpanzee bacillus" não estão montados e anotados. Além disso, os genomas de *M. pinnipedii* e "dassie bacillus" nunca foram analisados por genômica comparativa.

Análises de genômica comparativa podem fornecer informações quanto às especificidades de cada espécie, adaptação e evolução da bactéria e sua relação com os hospedeiros. Apesar de estudos de genômica comparativa entre as espécies que compõem o MTBC terem sido realizados, descrevendo as relações filogenéticas entre as diferentes espécies e linhagens, identificando mutações pontuais, inserções e deleções genômicas (KATO-MAEDA et al., 2001; BENTLEY et al., 2012; NAMOUCHI et al., 2012; COSCOLLA et al., 2013; ILINA et al., 2013; COSCOLLA; GAGNEUX, 2014; LIU et al., 2014; RAHMAN et al., 2014; DE LA FUENTE et al., 2015; DIPPENAAR et al., 2015), a maior parte dos estudos realizados até o momento não foram conduzidos de forma abrangente para todos os membros do MTBC, e poucos analisaram o conteúdo gênico dessas bactérias ou diferenças metabólicas inferidas por anotação (SALAMON et al., 2000; KATO-MAEDA et al., 2001; GAO et al., 2005; ISAZA et al., 2012; DEJESUS; SACCHETTINI; IOERGER, 2013; SUPPLY et al., 2013; LIU et al., 2014; DIPPENAAR et al., 2015; PERIWAL et al., 2015; JIA et al., 2017; XIONG et al., 2017; YANG et al., 2018).

A maioria das abordagens utilizando SNPs e indels compara sequências de *M. tuberculosis* para avaliar diferenças entre linhagens quanto a estirpes multirresistentes a antimicrobianos comparados com a estirpe H37Rv, além de variações genéticas específicas associadas a virulência e resistência a antimicrobianos e comparações genômicas quanto a estrutura filogenética (KATO-MAEDA et al., 2001; DAUM et al., 2012; ALI et al., 2015;

BLACK et al., 2015; REGMI et al., 2015; PHELAN et al., 2016; YAR et al., 2018). Por exemplo, inúmeros estudos avaliaram isolados de *M. tuberculosis* para identificar regiões de deleções genômicas e mutações pontuais específicas com potencial relevância para a patogênese da doença (BROSCH et al., 1999; FLEISCHMANN et al., 2002; UPLEKAR et al., 2011; LIU et al., 2014; LUO et al., 2014; BROWN et al., 2015; RODRÍGUEZ-CASTILLO et al., 2017; FAKSRI et al., 2018). A comparação genômica entre eles também permitiu estruturar filogenias robustas de linhagens distribuídas globalmente, sugerindo uma relação patógenohospedeiro estável ao longo do tempo evolutivo, considerando as associações geográficas das estirpes (BROSCH et al., 2002). Por outro lado, comparações envolvendo outras espécies do MTBC são normalmente resultantes de análises filogenéticas para investigar genes candidatos a promoverem a diferença entre as espécies, descrevendo relações evolutivas e a proximidade genética de estirpes adaptadas aos humanos e representantes de estirpes adaptadas aos animais (BENTLEY et al., 2012; DE LA FUENTE et al., 2015; BRITES et al., 2018).

Após as primeiras avaliações envolvendo a identificação dos SNPs e indels, alguns estudos se concentraram no conteúdo gênico entre os membros do MTBC. Devido à alta taxa de similaridade entre as espécies, foram encontradas pequenas, porém notáveis diferenças no repertório de genes entre as estirpes, algumas vezes relacionados a patogenêse bacteriana. Também há um foco voltado para a espécie *M. tuberculosis*, descrevendo categorias de genes presentes em todas as estirpes de analisadas e genes parcialmente compartilhados entre eles (GAO et al., 2005; YANG et al., 2018). Estudos que incluíram nas análises outras espécies do MTBC foram realizados com o objetivo de compreender genes de resistência a antimicrobianos relacionada às diferentes espécies (PERIWAL et al., 2015), e descrever a presença de genes não essenciais para o crescimento de estirpes do complexo *M. tuberculosis*, mas que a sua expressão pode influenciar no grau de virulência bacteriana (SALAMON et al., 2000; KATO-MAEDA et al., 2001).

Por fim, embora resultados importantes tenham sido obtidos, a maioria dos estudos realizados não utilizaram plataformas de anotação gênica que incluísse a detecção de pseudogenes nas sequências de interesse (ISAZA et al., 2012; DEJESUS; SACCHETTINI; IOERGER, 2013; SUPPLY et al., 2013; LIU et al., 2014; DIPPENAAR et al., 2015; JIA et al., 2017; XIONG et al., 2017; YANG et al., 2018) o que pode ter impacto significativo nos resultados encontrados. Pseudogenes e regiões truncadas localizados no final de contigs do genomas incompletos (i.e. artefatos de montagem) podem ser reportados como genes verdadeiros, fazendo com que o número de proteínas ortólogas encontradas no genoma acessório e genoma estirpe-específico seja superestimado. Por tratar-se de um grupo bacteriano

com alta taxa de similaridade entre as espécies, utilizar ferramentas robustas é imprescindível para encontrar diferenças reais entre as estirpes.

3. OBJETIVOS

3.1.OBJETIVOS GERAIS

- (i) Sequenciar, anotar e analisar os genomas completos de dois isolados de *Mycobacterium pinnipedii* obtidos de uma carcaça de leões marinhos da América do Sul e compará-los com outras estirpes de *M. pinnipedii* depositadas em bancos de dados públicos;
- (ii) Comparar genomas das espécies do complexo *Mycobacterium tuberculosis* a fim de identificar variações genômicas relacionadas ao metabolismo, epítopos de células T e virulência.

3.2.OBJETIVOS ESPECÍFICOS

- (i) Sequenciar, montar e anotar os genomas completos de dois isolados de *M. pinnipedii* (MP1 e MP2) previamente obtidos do pulmão e do linfonodo mesentérico do leão-marinho sul-americano (*Otaria flavescens*);
- (ii) Avaliar a distância genética entre *M. pinnipedii* MP1 e MP2 para determinar a ocorrência de microevolução ou superinfecção (infecção mista) no leão marinho;
- (iii) Realizar uma análise genômica comparativa abrangente de *M. pinnipedii* MP1 e MP2 e outras cepas de *M. pinnipedii* depositadas em bancos de dados públicos;
- (iv) Determinar e caracterizar funcionalmente os genomas *core*, acessório e estirpeespecífico de espécies sequenciadas do MTBC.
- (v) Investigar a presença de fatores de virulência nos genomas *core*, acessório e estirpe-específico de espécies do MTBC;
- (vi) Comparar o número, distribuição e conservação de epítopos de células T nos genomas *core*, acessório e estirpe-específico de espécies do MTBC.

4. CAPÍTULO 1: Sequenciamento do genoma de estirpes de *Mycobacterium pinnipedii*: caracterização genética completa e evidência de superinfecção em um leão-marinho sulamericano (*Otaria flavescens*)

RESUMO

Mycobacterium pinnipedii, membro do Complexo Mycobacterium tuberculosis (MTBC), é capaz de infectar várias espécies hospedeiras, incluindo humanos. Recentemente, o DNA antigo desse microrganismo foi recuperado de múmias pré-colombianas do Peru, revelando uma discussão sobre a origem e frequência da tuberculose nas Américas antes da colonização europeia. Neste trabalho, apresentamos o primeiro estudo de genômica comparativa dessa espécie bacteriana a partir do sequenciamento completo do genoma de dois isolados de M. pinnipedii (MP1 e MP2) obtidos de diferentes órgãos de um leão marinho sul-americano (Otaria flavescens). Nossos resultados indicam que MP1 e MP2 diferem em 113 SNPs (polimorfismos de nucleotídeo único) e 46 deleções curtas, constituindo o primeiro relato de uma infecção mista por M. pinnipedii em um leão marinho. A anotação dos SNPs e análise de redes de interação de proteínas indicam que os genes da família VapBC, um sistema toxina-antitoxina, estão sob pressão evolutiva para a alteração da sequência proteica nestas estirpes. Análises com OrthoMCL dos genomas desses isolados com os genomas de mais cinco isolados modernos de M. pinnipedii depositados em bancos de dados públicos mostram que estas estirpes possuem proteomas altamente conservados. Variações genéticas foram associadas apenas a proteínas hipotéticas e proteínas da família gênica prolina-glutamato (PE)/prolina-prolina-glutamato (PPE). Também foram detectados 28 polimorfismos de sequência grande (deleção ≥ 500 pb) em estirpes de M. pinnipedii antigas e modernas, incluindo 21 nunca descritos e 15 ocorrendo apenas em estirpes modernas, indicando um processo de redução do genoma ocorrido nos últimos 1.000 anos. Nossas análises filogenéticas sugerem a existência de dois grupos modernos de M. pinnipedii associados à localização geográfica e, possivelmente, às espécies hospedeiras, e um nó basal associado às estirpes antigas de M. pinnipedii. As deleções MiD3 e MiD4 descritas anteriormente podem ter ocorrido de forma independente, duas vezes, ao longo do curso evolutivo do MTBC. Em conclusão, a presença de superinfecção (i.e. infecção por estirpe mista) neste leão marinho sugere que M. pinnipedii é altamente endêmico nesta população. Em contraste com as estirpes de M. tuberculosis adaptadas ao homem, os proteomas de M. pinnipedii dos isolados estudados apresentaram alto grau de conservação, apesar de estarem sob redução genômica quando comparados ao M. tuberculosis. Este achado indica que genomas adicionais precisam ser sequenciados e analisados para aumentar as chances de encontrar genes variavalmente presentes entre as linhagens ou que o remodelamento do genoma de *M. pinnipedii* ocorreu antes da especiação bacteriana e se mantém com poucas alterações.

Palavras-chave: *Mycobacterium pinnipedii*, superinfecção, genômica comparativa, Complexo *Mycobacterium tuberculosis*, genoma.

4.1.INTRODUÇÃO

A tuberculose é uma doença infecto-contagiosa, e frequentemente grave, causada por membros do Complexo Mycobacterium tuberculosis (MTBC). O MTBC é um grupo bacteriano clonal composto por 11 espécies que podem ser divididas em dois grandes grupos: os altamente adaptados a humanos, denominados Mycobacterium tuberculosis e Mycobacterium africanum linhagem 5 (L5) e linhagem 6 (L6), e aqueles adaptados a hospedeiros animais, denominados Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti, Mycobacterium pinnipedii, Mycobacterium orygis, Mycobacterium mungi, Mycobacterium suricattae, "dassie bacillus" e "chimpanzee bacillus" (COUSINS et al., 1994, 2003; COSCOLLA; GAGNEUX, 2014). Estas espécies compartilham grande semelhança genômica, com 99,95% de identidade de nucleotídeos em regiões alinhadas, e nenhuma evidência de transferência horizontal de genes ou maiores eventos de recombinação (BRITES; GAGNEUX, 2015). Polimorfismos de nucleotídeo único (SNPs) e deleções de regiões genômicas variando de 2 a 12,7 Kb, denominados regions of difference (RDs), permitem a diferenciação de espécies. Apesar dessa alta similaridade genética, membros de MTBC variam em seu tropismo hospedeiro e capacidade de causar doenças, que são prováveis consequências dessas diferenças genéticas pontuais, além da presença de duplicações e elementos genéticos móveis (COSCOLLA; GAGNEUX, 2014).

Recentemente, especial atenção tem sido dada à espécie do MTBC conhecida por infectar pinípedes, o M. pinnipedii, devido ao fato de seu DNA ter sido detectado em três esqueletos humanos pré-colombianos de 1.000 anos de idade da América do Sul (Peru). Este estudo foi a primeira evidência genética de micobactérias tuberculosas infectando seres humanos antes do primeiro contato dos europeus com o Novo Mundo (BOS et al., 2014). Além disso, foi também o primeiro relato do sequenciamento do genoma de estirpes modernas de M. pinnipedii, que ainda não foram totalmente comparadas e analisadas. Nos dias de hoje, a tuberculose em pinípedes foi relatada pela primeira vez em 1913 (BLAIR WR, 1913) com M. pinnipedii sendo descrito como "bacilo de focas" no início dos anos 90 (COUSINS et al., 1990, 1993) e proposta como um novo membro do MTBC em 2003 (COUSINS et al., 2003). Estirpes de M. pinnipedii foram isoladas de pinípedes em todo o mundo, especialmente em cativeiro, mas também de animais de vida livre do hemisfério sul (COUSINS, 1987; BASTIDA et al., 1999). Surpreendentemente, desde sua primeira descrição, M. pinnipedii foi detectado em uma variedade de espécies hospedeiras, incluindo seis espécies diferentes de pinípedes (FORSHAW; PHELPS, 1991; BERNARDELLI et al., 1996; KIERS et al., 2008), camelos-bactrianos, leopardos-das-neves, leopardos-de-amur, gado, lhama, gorilas das planícies, antas-malaia, golfinho-de-hector e possivelmente humanos, o que sugere um comportamento generalista para o tropismo hospedeiro (COUSINS, 1987;

THOREL et al., 1998; BASTIDA et al., 1999; KRIZ et al., 2011; BOARDMAN et al., 2014; LOEFFLER et al., 2014). Pinípedes infectados com *M. pinnipedii* em jardins zoológicos e parques marinhos são a principal fonte de infecção para outros animais e seres humanos. Relatos de bovinos infectados foram associados a um canal de água contaminada conectando-se diretamente às áreas de pastagens oceânicas ou de praia onde os pinípedes foram encontrados (LOEFFLER et al., 2014). No entanto, o impacto real da tuberculose em espécies de pinípedes permanece em grande parte desconhecido, principalmente porque sua prevalência em focas e leões marinhos de vida livre é completamente inexplorada.

A carcaça de um leão-marinho sul-americano (Otaria flavescens) foi recentemente recuperada do litoral sul do estado do Rio Grande do Sul, Brasil. Como colônias reprodutivas desses animais estão ausentes no Brasil, os espécimes encontrados no litoral são provavelmente do Uruguai. Na necropsia do animal foram encontradas lesões compatíveis com tuberculose em diferentes órgãos, das quais foram obtidos isolados de M. pinnipedii, constituindo o primeiro relato desse patógeno na costa brasileira (DE AMORIM et al., 2014). Infelizmente, o sequenciamento do genoma dos isolados não foi realizado na época, impedindo a oportunidade de explorar a composição genética dessas bactérias, incluindo a possibilidade de variantes clonais bacterianas intra-hospedeiras (i.e., microevolução) ou infecção por estirpe mista (i.e. superinfecção). Até agora, as sequências do genoma de M. pinnipedii nunca foram totalmente caracterizadas pela genômica comparativa. Dados genômicos de estirpes modernas de M. pinnipedii descritas por Bos e colaboradores foram usados apenas para comparar com os antigos genomas de M. pinnipedii (BOS et al., 2014). Durante o desenvolvimento deste estudo, outra estirpe de M. pinnipedii foi sequenciada (RIOJAS et al., 2017), mas foi aplicada apenas para avaliar a posição filogenômica desta espécie dentro do MTBC. Portanto, os objetivos do presente estudo foram sequenciar e comparar dois isolados de *M. pinnipedii* obtidos de diferentes órgãos desta carcaça de leões marinhos (DE AMORIM et al., 2014), e realizar uma análise genômica comparativa com outras estirpes de M. pinnipedii disponíveis em banco de dados públicos.

4.2.MATERIAL E MÉTODOS

4.2.1. Seleção de genomas

Conjuntos de leituras (reads) de seis estirpes modernas de M. pinnipedii [G01222 da Argentina; G01491, G01492, G01498 da Austrália; e 7739 e 7011 de um zoológico da Alemanha (SRR1239336, SRR1239337, SRR1239338, SRR1239339, SRR1239341 e SRR1239340)] e três estirpes antigas de M. pinnipedii de múmias peruanas de mil anos foram selecionadas para este estudo [58, 54 e 64 (SRR1238557, SRR1238558 e SRR1238559)] (BOS et al., 2014). Todos os conjuntos de leituras foram recuperados do Sequence Read Archive (SRA) do National Center for Biotechnology Information (NCBI). Contigs montados da estirpe ATCC BAA-688 (NZ_MWXB01000000) de M. pinnipedii da Austrália depositados no RefSeq também foram selecionados (RIOJAS et al., 2017). Conjuntos de leituras desta estirpe ATCC (BAA-688) não estão depositados em bancos de dados públicos e não foram incluídos em nenhuma análise que requer este tipo de arquivo de entrada. Além disso, as estirpes 7739 e 7011 de M. pinnipedii tinham apenas conjuntos de leituras não pareadas (single reads) publicamente disponíveis. Estas não eram apropriadas para a montagem do genoma e não foram incluídas nas análises que necessitavam de montagens. O genoma completo de M. tuberculosis H37Rv (NC_000962.3) foi usado como referência. Para a análise filogenômica, leituras de representantes do MTBC ["chimpanzee bacillus" (ERR150046), "dassie bacillus" (SRR3745458), M. africanum GM041182 (ERR234255), M. africanum MAL010070 (SRR998578), M. bovis SP38 (SRR6705904), M. caprae EPDC02 (DRR120409), M. microti 94-2272 (ERR027298), M. mungi BM22813 (SRR3500411), M. orygis IDR1100020842 (SRR5642712), M. suricattae ERS798580 (ERR970409)] foram também incluídos.

4.2.2. Isolamento e extração de DNA bacteriano

Dois isolados de *M. pinnipedii* (MP1 e MP2) obtidos do pulmão e linfonodos mesentéricos de um leão marinho sul-americano (*Otaria flavescens*) encontrado morto em Capão da Canoa, Rio Grande do Sul, Brasil (DE AMORIM et al., 2014) foram mantidos a -80 °C em meio 7H9 com 20% de glicerol. Para este estudo, ambos os isolados foram reativados em meio Stonebrink e uma única colônia de cada isolado foi sub-cultivada para extração de DNA como descrito anteriormente (VAN SOOLINGEN et al., 1994; BEMER-MELCHIOR; DRUGEON, 1999). A qualidade e quantidade do DNA foram medidas por Nanodrop 2000c (Thermo Scientific, MA, EUA) seguindo as razões de absorbância de 260/280 e 260/230 e em gel de agarose a 0,8% com marcador de massa molecular. Uma análise final foi realizada com o Agilent 2100 High Sensitivity Chip DNA Bioanalyzer (Agilent Technologies, CA, EUA) para

concentração e fragmentação da amostra. Todos os procedimentos envolvendo material infeccioso foram realizados no Laboratório de Biossegurança 3+ (BSL-3 + Prof. Dr. Klaus Eberhard Stewien) do Departamento de Microbiologia do Instituto de Ciências Biomédicas - Universidade de São Paulo, Brasil.

4.2.3. Sequenciamento de isolados de M. pinnipedii

Bibliotecas genômicas foram construídas utilizando o kit de preparação de amostras de PCR TruSeq DNA (Illumina, CA, EUA) de acordo com as instruções do fabricante. Utilizou-se HiSeq2500 Illumina v3 para sequenciar as bibliotecas genômicas com leituras pareadas de 100 pares de bases (pb). Esses procedimentos foram realizados no Laboratório Central de Tecnologias de Alto Desempenho (LaCTAD) da Universidade Estadual de Campinas (UNICAMP) - Brasil. As leituras de sequenciamento da Illumina foram depositadas no SRA, NCBI sob os números de acesso: SRR7693584 e SRR7693090.

4.2.4. Montagem e anotação do genoma

As leituras pareadas de 100 pb de cada biblioteca foram primeiro filtradas por qualidade e presença de adaptadores usando a versão 0.36 do software Trimmomatic (BOLGER; LOHSE; USADEL, 2014). Os genomas MP1 e MP2 de *M. pinnipedii* foram montados *de novo* utilizando o software SPAdes (Bankevich et al., 2012). A identificação e anotação gênica foram realizadas automaticamente pelo *Prokaryotic Genome Annotation Pipeline* (PGAP) e os genomas montados foram depositados como GCA_003027795.2 (MP1) e GCA_003027895.2 (MP2).

4.2.5. Mapa circular dos genomas de M. pinnipedii

Para visualizar a similaridade entre as proteínas preditas de diferentes linhagens de *M. pinnipedii*, os genomas das estirpes G01222, G01491, G01492 e G01498 foram montados utilizando-se o software CLC Genomics WorkBench (Qiagen, Holanda) ou SPAdes (BANKEVICH et al., 2012) e anotado com RAST (AZIZ et al., 2008). Contigs resultantes das estirpes G01222, G01491, G01492 e G01498 e contigs das estirpes MP1, MP2 e ATCC BAA-688 de *M. pinnipedii* depositadas no GenBank foram reordenadas com MAUVE (RISSMAN et al., 2009) usando *M. tuberculosis* H37Rv (GenBank NC000962. 3) como o genoma de referência. Posteriormente, um mapa circular foi construído usando a ferramenta de "Comparação de Proteomas" implementado nos recursos da plataforma Patric 3.5.18 (WATTAM et al., 2017).

4.2.6. Mapeamento e identificação de variantes

Utilizou-se o programa Burrows-Wheeler (BWA) (LI; DURBIN, 2009) para mapear leituras de *M. pinnipedii* MP1 e MP2 contra o genoma de referência *M. tuberculosis* H37Rv. As saídas do BWA foram analisadas com SAMtools (LI et al., 2009) para classificar de acordo com filtros de qualidade de mapeamento da escala de Phred 20 e remoção de leituras duplicadas. Detecção de SNPs e inserções e deleções foi realizada com FreeBayes (GARRISON; MARTH, 2012) com os seguintes parâmetros: qualidade mínima de mapeamento de 10, qualidade de base mínima em uma posição de 10, profundidade mínima de leitura em uma posição de 5. Os SNPs identificados foram então anotados usando SNPEff (CINGOLANI et al., 2012). Variantes anotadas em regiões relacionadas a elementos transponíveis, família 13E12, fagos ou família de proteínas PE/PPE (prolina-glutamato e prolina-prolina-glutamato, respectivamente) foram removidos para evitar resultados falso-positivos.

Para identificar SNPs únicos de *M. pinnipedii*, leituras genômicas representantes do MTBC ("chimpanzee bacillus", "dassie bacillus", *M. africanum* GM041182, *M. africanum* MAL010070, *M. bovis*, *M. caprae*, *M. microti*, *M. mungi*, *M. orygis*, *M. suricattae*) foram incluídos na análise e mapeados contra *M. tuberculosis* H37Rv. As variantes presentes apenas nas 11 estirpes de *M. pinnipedii* foram consideradas únicas para essas espécies bacterianas.

4.2.7. Redes de interação de proteínas

Sequências de DNA codificante (CDSs) identificadas com mutações não-sinônimas quando comparadas as estirpes MP1 e MP2 foram analisadas usando o banco de dados STRING com configurações padrão para a previsão de associações de rede entre proteínas (JENSEN et al., 2008). Os resultados foram utilizados para identificar vias metabólicas específicas ou interações de proteínas sob pressão evolutiva. Quando necessário, sequências proteicas foram pesquisadas contra o Banco de Dados de Fatores de Virulência (VFDB) (CHEN et al., 2004) para identificar fatores de virulência.

4.2.8. Spoligotyping in silico e Número Variável de Repetição em Tandem (VNTR)

Para a identificação dos espoligotipos, as leituras dos isolados de *M. pinnipedii* foram analisadas com SpoTyping (XIA; TEO; ONG, 2016). Os padrões identificados foram submetidos ao banco de dados Spoligotyping de *M. bovis* (www.mbovis.org) para identificação do número do padrão. Para identificar o número variável de repetições em tandem (VNTRs) de elementos genéticos denominados unidades micobacterianas repetitivas intercaladas (MIRUs), os 24 loci MIRU-VNTR foram identificados individualmente *in silico* utilizando primers previamente

descritos (SUPPLY et al., 2006) e o resultado dos padrões foram analisados no banco de dados MIRU-VNTRplus (ALLIX-BÉGUEC et al., 2008; WENIGER et al., 2010).

4.2.9. Agrupamentos de proteínas de genes ortólogos e parálogos de M. pinnipedii

Para identificar grupos de proteínas de genes ortólogos, sete estirpes de M. pinnipedii foram utilizadas (G01222, G01491, G01492, G01498, MP1, MP2 e ATCC BAA-688). Os contigs montados foram anotados ou re-anotados com o programa RAST (AZIZ et al., 2008) e os CDSs previstos com ≤ 150 pb foram excluídos. As sequências de proteínas foram agrupadas usando OrthoMCL (LI; STOECKERT; ROOS, 2003) disponível na plataforma KBase (ARKIN et al., 2018). Resumidamente, pares homólogos de sequências foram encontrados usando o algoritmo BLASTp todos-contra-todos com um e-value <1e-5. O OrthoMCL converteu os resultados do BLASTp em uma matriz de similaridade normalizada que foi analisada pelo algoritmo Markov Cluster (MCL) para agrupamento de sequências ortólogas. O índice de inflação de 1,5 foi usado para regular o cluster. Como o algoritmo RAST reporta pseudogenes e genes tuncados localizados no final de contigs como genes verdadeiros, pode haver uma superestimação do número de CDSs estirpe-específicos. Essa superestimação foi verificada manualmente usando BLASTp e/ou tBLASTn (ALTSCHUL et al., 1990) comparando os resultados obtidos com a anotação das estirpes MP1, MP2 e ATCC BAA-688 de M. pinnipedii, que estão depositadas no Genbank e anotadas com PGAP do NCBI. Grupos de proteínas ortólogas foram classificadas com eggNOG (HUERTA-CEPAS et al., 2016) usando um algoritmo a metodologia COG (Cluster of Orthologous Groups) (TATUSOV et al., 2000).

4.2.10. Polimorfismos de sequência grandes

Polimorfismos de sequência grandes (PSGs) foram detectados com CLC Genomics Workbench 11 (QIAGEN, Venlo, Holanda) por mapeamento de leituras de estirpes antigas (54, 58, 64) e modernas (G01222, G01491, G01492, G01498, 7739, 7011, MP1, MP2) de *M. pinnipedii* contra *M. tuberculosis* H37Rv. Ao mapear as amostras antigas contra o genoma de referência, foi dado tratamento especial às leituras antes e durante o mapeamento. Primeiro, as leituras foram analisadas com parâmetros de qualidade de probabilidade de erro de chamada de base de 0,05 e presença de dois ou mais nucleotídeos ambíguos nas extremidades usando o algoritmo modificado de corte Mott. Segundo, o mapeamento de leituras em mais de uma região do genoma foi excluído. Regiões mapeadas com menos de 10 leituras alinhadas e/ou contendo regiões de baixa qualidade foram consideradas ausentes no genoma de interesse. Os PSGs foram considerados quando essa região ultrapassava 500 pb.

4.2.11. Análise filogenética

Os membros do MTBC e onze estirpes de *M. pinnipedii* descritas acima foram utilizados para construir uma árvore filogenética. Resumidamente, todos os conjuntos de leituras foram mapeados contra o genoma de referência *M. tuberculosis* H37Rv e as variantes verdadeiras foram denominadas como descrito acima, com exclusão de SNPs localizados em regiões repetitivas. Um *script* personalizado em linguagem python foi então usado para construir uma matriz posicional de SNPs identificados em genomas de todas as estirpes/espécies. A matriz foi utilizada como entrada para o software RAxML (STAMATAKIS, 2014) para a construção da árvore filogenética usando o algoritmo de Máxima Verossimilhança (ML) com modelo GTRCAT e autoMRE para ML de melhor pontuação e um máximo de 1.000 inferências de *bootstrap*.

4.3.RESULTADOS E DISCUSSÃO

4.3.1. Primeira descrição de infecção mista por M. pinnipedii em um pinípede

O sequenciamento completo do genoma e a montagem dos dois isolados de *M. pinnipedii* (denominados MP1 e MP2) resultaram em 9.766.154 (cobertura de 228x) e 9.433.048 leituras (cobertura de 220x), montadas em 102 e 106 contigs (~ 4,28 Mb/genoma), respectivamente. A anotação do genoma com PGAP indica um conteúdo GC de 65,4% e 45 tRNA para ambas as linhagens, e 4.269 e 4.276 genes, 3.993 e 3.996 CDSs, e 152 e 158 pseudogenes para MP1 e MP2, respectivamente. Genomas completos de ambas os isolados mostraram características similares a outros genomas da mesma espécie disponíveis no banco de dados (Figura 1).

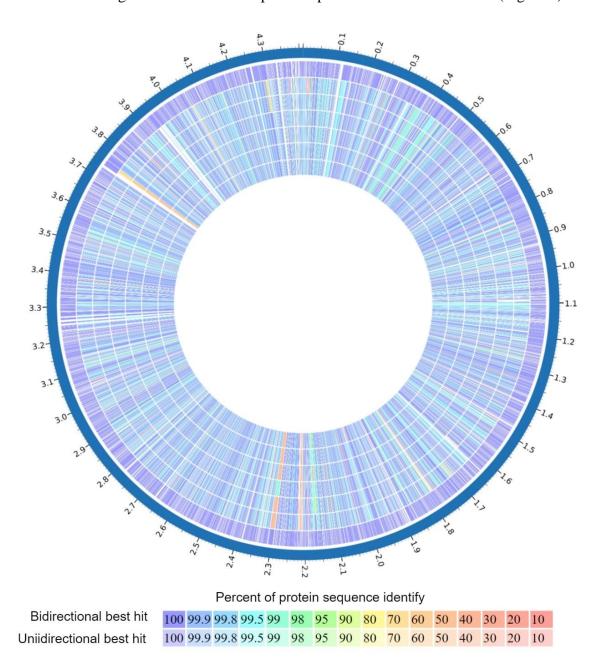


Figura 1. Mapa circular do proteoma dos genomas do *Mycobacterium pinnipedii* comparado ao genoma de referência do *Mycobacterium tuberculosis* H37Rv. Do anel externo ao anel interno: posição do genoma, *M. tuberculosis*, H37Rv, *M. pinnipedii* MP1, *M. pinnipedii* MP2, *M. pinnipedii* G01222, *M. pinnipedii* G01491, *M. pinnipedii* G01492, *M. pinnipedii* G01498, *M. pinnipedii* ATCC BAA-688. Escala de cores representa identidade de proteína conforme apresentado na legenda. Contigs foram ordenados usando MAUVE (RISSMAN et al., 2009) e o mapa foi gerado com recursos da plataforma Patric 3.5.18 (WATTAM et al., 2017).

O mapeamento das leituras de M. pinnipedii MP1 e MP2 com o genoma de referência M. tuberculosis H37Rv resultou em uma cobertura de 99,08% e 99,03% e 2,312 e 2,350 variantes, respectivamente. Após a exclusão de potenciais áreas falso-positivas, os isolados compartilharam 1.912 SNPs (M. pinnipedii MP1) e 2.206 SNPs (M. pinnipedii MP2) com M. tuberculosis H37Rv. Além disso, foram detectados 181 (de 2 a 36 pb) (M. pinnipedii MP1) e 208 indels curtos (M. pinnipedii MP2). Considerando as diferenças de nucleotídeos, a distância entre MP1 e MP2 de M. pinnipedii foi então de 113 SNPs e 46 deleções curtas. Um total de 122 (76,73%) destes 159 SNPs e deleções curtas foram localizados em CDSs com função conhecida, com base na anotação funcional COG (APÊNDICE A) relacionadas à produção e conversão de energia (11,95%, 19/159), transporte e metabolismo lipídico (9,43%, 15/159), e biogênese da parede celular/membrana/envelope (7,55%, 12/159), que são conhecidos por desempenhar papéis cruciais na patogenicidade de espécies MTBC. Curiosamente, a maioria dos SNPs detectados são não-sinônimos (86,72%, 98/113), o que significa que os CDSs relacionados estão sob pressão seletiva para a mudança da sequência de proteínas, tendo o potencial de promover a diversidade funcional. Ao analisar SNPs não-sinônimos dos CDSs de função conhecida, identificamos fortes interações de proteínas entre membros específicos da família de proteínas VapBC (Figura 2). O fato de que múltiplas proteínas interagindo umas com as outras apresentarem SNPs nãosinônimos indica que essa interação, em particular, está sob pressão evolutiva por mudança. Em M. tuberculosis, a família VapBC é responsável por mais da metade da proporção de sistemas toxina-antitoxina (AT) (47/88 sistemas AT putativos) (RAMAGE; CONNOLLY; COX, 2009) e regiões genômicas contendo VapBC estão fortemente associadas à virulência e fatores de patogenicidade (ARCUS; RAINEY; TURNER, 2005). Os CDSs identificados fazem parte de um sistema AT tipo II relacionado à regulação da tradução do mRNA. Mais especificamente, VapC25 e VapC29 são ribonucleases que demonstraram inibir o crescimento de células de Mycobacterium smegmatis quando induzidas (isto é, têm efeitos bacteriostáticos) (WINTHER et al., 2016). Em conjunto, estes resultados sugerem que MP1 e MP2 podem ter diferentes capacidades de sobrevivência.

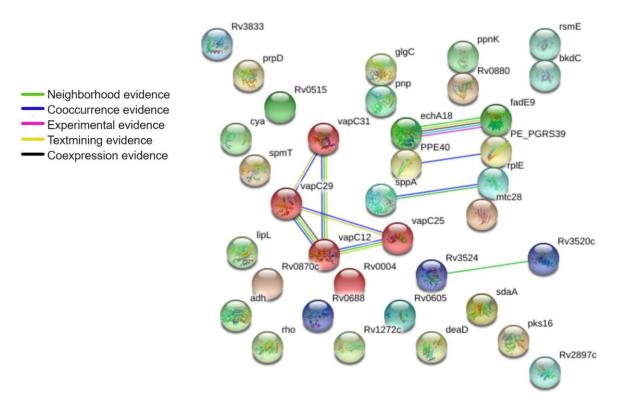


Figura 2. Rede de interação de proteínas em regiões de CDS (sequências de DNA codificante) com função conhecida contendo SNPs (polimorfismo de nucleotídeo único) não-sinônimos. CDSs com mutações foram identificadas entre as linhagens MP1 e MP2 de *Mycobacterium pinnipedii*. A rede foi gerada usando o banco de dados STRING, com parâmetros padrão.

Dois espoligotipos, SB0155 e SB2455, e dois perfis MIRU-VNTR distintos, VNTR580 e VNTR960, foram identificados em *M. pinnipedii* MP1 e MP2 (APÊNDICE B), respectivamente. Em contraste com o espoligotipo SB0155 (COUSINS et al., 2003), o perfil SB2455 nunca foi descrito em estirpes de *M. pinnipedii*. A presença de múltiplas bandas de locus MIRU-VNTR indica heterogeneidade na multiplicidade de números de cópias de alelos. A variação em vários loci normalmente indica infecções por estirpes mistas (COHEN et al., 2011). Portanto, com base nos parâmetros atuais de distinção entre microevolução e superinfecção (infecção mista) em relação a SNPs/indels (WALKER et al., 2013) e MIRU-VNTR (COHEN et al., 2012), o leão marinho aqui analisado encontrava-se infectado com duas estirpes diferentes de *M. pinnipedii*.

Este é o primeiro relato de superinfecção (infecção mista) por *M. pinnipedii* em um leão marinho. Como esses animais normalmente vivem em aglomerados populacionais densos, eles são altamente vulneráveis à epizootias de doenças infecciosas que podem ser transmitidas por contato direto de animal-animal. A superinfecção causada por *M. tuberculosis* em humanos é

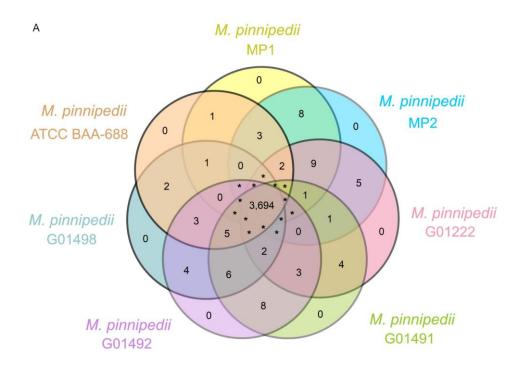
normalmente relatada em países altamente endêmicos, nos quais as pessoas são expostas a múltiplas estirpes de *M. tuberculosis* ao longo de suas vidas e o HIV desempenha um papel importante na formação da incidência da doença (COHEN et al., 2012). Assim, ao traçar um paralelo, é possível que o *M. pinnipedii* seja altamente endêmico na população de onde veio este leão marinho, representado por diferentes estirpes circulantes, o que pode ter efeitos sem precedentes na conservação dessa espécie. Condições ambientais adversas, nutrição insuficiente e estresse crônico devido à perturbação ou competição podem agir para suprimir o sistema imunológico e, portanto, os pinípedes podem estar predispostos a doenças causadas por vários patógenos. No Brasil, o encalhe de leões marinhos ocorre normalmente devido a um comprometimento da saúde animal quando viajam longas distâncias durante o período migratório (RUOPPOLO, 2003). O estado enfraquecido com que esses animais chegam ao litoral, muitas vezes infectados por grandes cargas bacterianas, é o fator que mais leva esses animais à morte (HIGGINS, 2000; GONZALES-VIERA et al., 2013). A contribuição da infecção por *M. pinnipedii* neste contexto ainda precisa ser elucidada.

Embora o *M. pinnipedii* tenha sido descrito em animais de cativeiro e de vida livre, ainda faltam estudos sistemáticos de vigilância populacional. Em 2011, descobriu-se que um leão marinho em cativeiro da América do Sul, alojado em um zoológico da República Tcheca, importado de um zoológico alemão (não especificado), estava infectado por *M. pinnipedii*. Os pais do animal também morreram de tuberculose na Alemanha e foram capturados como juvenis de águas costeiras do Uruguai em 1992 (KRIZ et al., 2011), sugerindo que a infecção veio da vida selvagem. Surtos da doença em leões marinhos sul-americanos selvagens e/ou em cativeiro alojados no zoológico de Heidelberg, na Alemanha, e no zoológico de Le Pal, na França, também foram descritos (JURCZYNSKI et al., 2011). Adicionado a este relato de superinfecção, fica claro que a *M. pinnipedii* é endêmico em populações selvagens de leões marinhos da América do Sul e está sendo introduzida em jardins zoológicos, o que justifica uma necessidade urgente de avaliar a extensão da tuberculose em pinípedes de vida livre do Hemisfério sul.

4.3.2. Estirpes de M. pinnipedii têm um proteoma altamente conservado

A Figura 3 ilustra grupos de proteínas ortólogas presentes em sete estirpes modernas de M. pinnipedii que foram depositadas em NCBI (M. pinnipedii MP1, MP2, ATCC BAA-688, G01222, G01491, G01492, G01498). Um total de 3.986 agrupamentos de proteínas (ou seja, grupos de \geq 2 proteínas) foram identificados entre as estirpes analisadas, incluindo 3.694 (92,67%) grupos de proteínas do genoma core presentes em todos os sete genomas (Figura 3A).

De acordo com a classificação COG, um terço (1.147/3.694; 31,05%) dos clusters de proteínas do genoma *core* têm função desconhecida, enquanto a maioria deles possui anotação funcional (2.547/3.694; 68,95%). As cinco principais classes funcionais foram: biogênese da parede celular/membrana/envelope (M; 426 / 3.694, 11,5%), transporte e metabolismo lipídico (I; 284/3.694, 7; 7%), produção e conversão de energia (C; 280 / 3,694, 7,6%), transcrição (K; 205 / 3,694, 5,5%) e transporte e metabolismo de aminoácidos (E, 181 / 3,694, 4,9%) (Figura 3B-C).



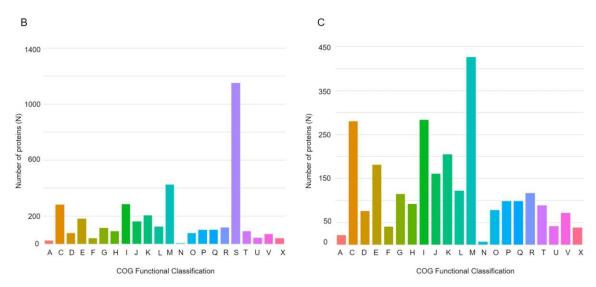


Figura 3. Agrupamentos de proteínas ortólogas de estirpes de *Mycobacterium pinnipedii*. (A) Diagrama de Venn de proteínas ortólogas. (B) *Cluster of Orthologous Groups* (COG) com base na função dos clusters de proteínas do genoma *core* das estirpes de *M. pinnipedii*. (C) classificação COG sem classe de função desconhecida (S). O agrupamento de proteínas foi realizado usando OrthoMCL como disponível na plataforma Kbase (ARKIN et al., 2018). * O

número de proteínas gênicas ortólogas compartilhadas entre esses grupos está disponível no Apêndice C. O tamanho do cluster variou de duas a quarenta e três proteínas.

Não existem agrupamentos de proteínas (isto é, grupos de ≥ 2 proteínas) únicos para cada estirpe de M. pinnipedii (Figura 3A). No entanto, existiram proteínas não categorizadas em grupos ortólogos e que são então consideradas únicas de cada estirpe de M. pinnipedii (MP1 = 32, MP2 = 37, ATCC BAA-688 = 58, G01498 = 103, G01492 = 68, G01491 = 52, G01222 = 80). Como a ferramenta de anotação usada aqui (RAST) não diferencia entre pseudogenes/genes truncados e genes verdadeiros, usamos BLASTp e/ou tBLASTn para pesquisar a anotação PGAP do NCBI das estirpes MP1, MP2 e ATCC BAA-688 usando suas "proteínas únicas" como entrada. Daquelas com função anotada (i.e. não hipotéticos ou genes PE/PPE), nossa busca BLAST revelou que todos os CDSs funcionais são considerados pseudogenes, inexistentes ou estão presentes em contigs menores que 700 pb (com qualidade duvidosa de sequenciamento/montagem) na anotação PGAP. Como o RAST reporta CDSs quebrados no final de contigs como genes verdadeiros, estes fragmentos de CDS (isto é, artefatos de montagens de genomas) não conseguiram agrupar-se em grupos ortólogos. No entanto, ainda existem vários genes hipotéticos e de PE/PPE que não foram capazes de distinguir entre pseudogenes e genes verdadeiros usando as ferramentas propostas, indicando a possibilidade de fenótipos distintos dentro das estirpes de M. pinnipedii. Essas variações podem ser consequência de truncamento ou pseudogenização de genes, o que pode levar à neo-funcionalização ou perda gênica respectivamente, bem como eventos de duplicação, fenômenos comumente observados no MTBC (BOLOTIN; HERSHBERG, 2015).

Ao considerar os grupos de proteínas do genoma acessório (n = 292) compartilhados entre duas e até seis estirpes de *M. pinnipedii*, resultados semelhantes foram obtidos com todos os CDSs funcionais identificados como pseudogenes ou inexistentes com base na anotação do PGAP, e um número desconhecido de possíveis genes hipotéticos e PE/PPE que podem estar variavelmente presentes entre as estirpes de *M. pinnipedii*. Em geral, nossos resultados indicam que as estirpes de *M. pinnipedii* analisadas apresentam um alto nível de conservação de proteomas, o que contrasta com uma análise recente de pan-genoma de estirpes de *M. tuberculosis* que detectaram pelo menos 1.122 CDSs no genoma acessório e 964 CDSs no genoma estirpe-específico em 36 genomas (YANG et al., 2018). No entanto, o pan-genoma *M. tuberculosis* nesse estudo também foi analisado usando RAST, e os pseudogenes e/ou artefatos de montagem não foram levados em consideração.

4.3.3. As estirpes modernas de M. pinnipedii têm marcadores únicos de deleção

Na análise de PSGs, 28 regiões deletadas, incluindo sete descritas anteriormente (RD2seal, MiD3, MiD4, RD7, RD8, RD9, RD10), foram encontradas em estirpes de *M. pinnipedii*, variando de 548 pb a 14.630 pb (Figura 4). Destes, 19 (67,9%) estão associados a pelo menos um gene da família PE/PPE. Embora a cobertura do genoma antigo de *M. pinnipedii* em relação ao *M. tuberculosis* H37Rv variou de 31,94% a 47,83% (após o filtro de qualidade), foi possível observar 15 regiões nessas estirpes de *M. pinnipedii* que não são deletadas em comparação às estirpes modernas de *M. pinnipedii*, e uma região deletada que está ausente apenas em algumas estirpes modernas de *M. pinnipedii*. A maioria dessas regiões (15/16), no entanto, está associada aos genes da família PE/PPE e deve ser interpretada com cautela, dada a qualidade do sequenciamento de DNA antigo e viés de duplicação gênica. A região não-PE/PPE compreende uma proteína hipotética e uma oxireductase (Rv3530c-Rv3531c). Este achado contribui para o entendimento da evolução do *M. pinnipedii* ao longo do tempo, indicando um processo ativo de redução do genoma que vem ocorrendo há pelo menos 1.000 anos.

Start	End	Sample 58°	Sample 54*	Sample 64*	Strain G0122	2 Strain G01491	Strain G0149	Strain G01498	Strain 7011	1 Strain 7739	Strain MP1	Strain MP2	Affected Gene	
149912	150861	0	0	0	949	949	949	949	949	949	949	949	PE_PGRS2	
264755	266656	1901	1901	1901	1901	1901	1901	1901	1901	1901	1901	1901	Rv0221-echA1-Rv0223c	RD10
361791	363013	0	0	0	1222	1222	1222	1222	1222	1222	1222	1222	PE_PGRS5	
836162	837957	0	0	0	1795	1795	1795	1795	1795	1795	1795	1795	PE_PGRS9	
838880	840587	0	0	0	1707	1707	1707	1707	1707	1707	1707	1707	PE_PGRS10	
889030	891403	2373	2373	2373	2373	2373	2373	2373	2373	2373	2373	2373	Rv0797 (transposase)	
925379	927486	0	0	0	2107	2107	2107	2107	2107	2107	2107	2107	PE_PGRS13	
1212087	1213736	0	0	0	1649	1649	1649	1649	1649	1649	1649	1649	PE_PGRS21	
1216975	1218635	0	0	0	1660	1660	1660	1660	1660	1660	1660	1660	PE_PGRS22	
1488407	1489522	0	0	0	1115	1115	1115	1115	1115	1115	1115	1115	PE_PGRS24	
1572228	1573484	0	0	0	1256	1256	1256	1256	1256	1256	1256	1256	PE_PGRS25	
1618335	1619139	0	0	0	0	0	0	0	804	804	804	804	PE_PGRS26	
1633475	1634610	1135	1135	1135	1135	1135	1135	1135	1135	1135	1135	1135	PE_PGRS27	
1637197	1637963	766	766	766	766	766	766	766	766	766	766	766	PE_PGRS28	
1779279	1789536	10257	10257	10257	10257	10257	10257	10257	10257	10257	10257	10257	Rv1573-Rv1588c (phage)	
1996102	1998197	2095	2095	2095	2095	2095	2095	2095	2095	2095	2095	2095	Rv1765c (hypothetical protein)	
2061403	2062190	0	0	0	787	787	787	787	787	787	787	787	PE_PGRS33	
2208006	2220724	12718	12718	12718	12718	12718	12718	12718	12718	12718	12718	12718	yrbE3A/B-mce3A/B/C/D-lprM-mceF-Rv1972-Rv1977	RD7
2220948	2222975	2027	2027	2027	2027	2027	2027	2027	2027	2027	2027	2027	Rv1978-Rv1979c (permease)	RD2seal
2330074	2332101	2027	2027	2027	2027	2027	2027	2027	2027	2027	2027	2027	cobL-Rv2073c-Rv2075c	RD9
2423531	2424245	0	0	0	714	714	714	714	714	714	714	714	PE_PGRS38	
3377696	3380674	0	0	0	2978	2978	2978	2978	2978	2978	2978	2978	PPE46-esxR-esxS-PPE47-48	MiD4
3730466	3735795	5329	5329	5329	5329	5329	5329	5329	5329	5329	5329	5329	PPE54	
3741146	3755776	0	0	0	14630	14630	14630	14630	14630	14630	14630	14630	PE_PGRS50-PPE55-Rv3348-Rv3349c	MiD3
3381373	3382712	1339	1339	1339	1339	1339	1339	1339	1339	1339	1339	1339	Rv3023c (transposase)	
3967681	3968967	0	0	0	1286	1286	1286	1286	1286	1286	1286	1286	Rv3530c-Rv3531c (oxiredutase)	
4036915	4037463	0	0	0	548	548	548	548	548	548	548	548	PE_PGRS59	
4056841	4062730	5889	5889	5889	5889	5889	5889	5889	5889	5889	5889	5889	ephA-Rv3618-esxV-esxW-PPE65-PE32-lpqG	RD8
								Size(bp)						
0	500 15	00 2000	2500 3000	3500 400	0 4500 50	000 5500 60	00 6500 70	000 7500 800	00 8500	9000 9500	10000 10	0500 11000	11500 12000 12500 13000 13500 14000 14	500 15000

Figura 4. Polimorfismos de sequência grande (PSGs) de estirpes de *Mycobacterium pinnipedii*. Os valores indicam o número de nucleotídeos que abrangem cada região deletadas (zero significa que a região não é deletada na linhagem bacteriana correspondente). Início e fim: posições de nucleotídeos de acordo com o genoma de referência de *M. tuberculosis* H37Rv. Os genes são anotados de acordo com o genoma de referência. * Estirpes sul-americanas antigas de *M. pinnipedii*. Deleções em cinza são aquelas que já foram descritas em estirpes de *M. pinnipedii*.

As famílias dos genes PE/PPE correspondem a aproximadamente 10% da capacidade de codificação do genoma de *M. tuberculosis* (COLE et al., 1998) e mostram um alto grau de variação entre os membros do MTBC e entre as linhagens da mesma espécie. As proteínas das famílias PPEs e PEs frequentemente interagem entre si e são encontradas intimamente associadas com os domínios ESX do sistema de secreção tipo VII nos genomas micobacterianos, atuando como importantes fatores de virulência (GEY VAN PITTIUS et al., 2006). Assim como na família ESAT-6, a perda de genes que codificam as proteínas PE/PPE pode desempenhar um papel importante no processo de atenuação de *M. pinnipedii* para alguns hospedeiros (BRENNAN, 2017).

Regiões deletadas específicas de *M. microti* denominadas MiD3 e MiD4 foram encontradas apenas nas linhagens modernas de *M. pinnipedii* (Figura 5). Esta observação sugere que as deleções de MiD3 e MiD4 podem ter ocorrido independentemente, duas vezes, ao longo do curso evolutivo. *Mycobacterium microti* foi descrito pela primeira vez em ratos (WELLS, 1937), mas desde então tem sido isolado de porcos, lhamas, gatos, javalis e humanos imunossuprimidos, ficando restrito à Eurásia. Isso levanta questões sobre o verdadeiro alcance de hospedeiros desses patógenos intimamente relacionados, que pode ou não ter mudado menos de 1.000 anos atrás.

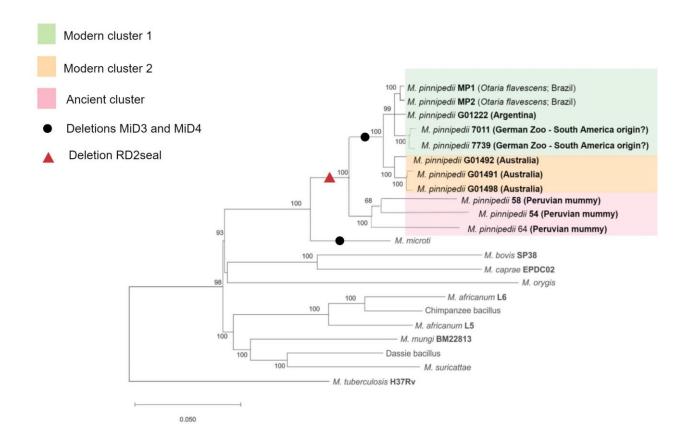


Figura 5. Árvore filogenética baseada em SNPs (polimorfismo de nucleotídeo único) do Complexo *Mycobacterium tuberculosis* (MTBC) usando o modelo de máxima verossimilhança (ML). Caixa verde: cluster moderno 1; caixa laranja: cluster moderno 2; caixa rosa: cluster antigo; círculo preto: deleções MiD3 e MiD4; triângulo vermelho: deleção RD2seal. *Mycobacterium tuberculosis* H37Rv foi usado como grupo externo. Um script customizado em Python foi usado para construir uma matriz de SNPs, que foi usada para inferir uma árvore ML usando o RAxML com o modelo GTRCAT e o autoMRE para a árvore ML de melhor pontuação e um máximo de 1.000 inferências de *bootstrap* usando o software RaxML. (STAMATAKIS, 2014).

4.3.4. Dois agrupamentos distintos de estirpes modernas de M. pinnipedii

A Figura 5 ilustra a árvore filogenética baseada em SNPs do MTBC, incluindo onze estirpes modernas e três antigas de *M. pinnipedii*. Verificou-se que um total de 1.698 posições polimórficas eram únicas das estirpes de *M. pinnipedii*, isto é, não estavam presentes em nenhuma outra espécie de MTBC. Além disso, estirpes de *M. pinnipedii* apareceram distribuídas em três grupos principais, com as estirpes antigas constituindo o nó mais basal (Figura 5). Curiosamente, as estirpes modernas de *M. pinnipedii* apareceram divididas em dois grupos de acordo com a localização geográfica e, possivelmente, com as espécies hospedeiras: o cluster moderno 1, composto por isolados de origem sul-americana, e o cluster moderno 2, composto por isolados da Austrália. Estes resultados estão de acordo com os achados das linhagens *M. tuberculosis* e *M. bovis* que também estão associados a localizações geográficas distintas (FILLIOL et al., 2006; GAGNEUX, 2012; ZIMPEL et al., 2017). Não se sabe se estes estão ou não associados com diferentes fenótipos de virulência.

4.4.CONSIDERAÇÕES FINAIS

Este é o primeiro relato de infecção mista por estirpe de *M. pinnipedii* em pinípedes. Este achado, em conjunto com relatos anteriores na literatura, sugere que a infecção por *M. pinnipedii* é endêmica em populações de vida livre de leões marinhos da América do Sul. As diferenças genéticas entre estas estirpes foram associadas a fatores de virulência e enzimas necessárias para manutenção intracelular e modelagem de micobactérias. Os efeitos reais desses fenômenos para o desfecho da doença e conservação das espécies animais e potenciais implicações para população são desconhecidos. No entanto, é urgente avaliar a extensão da doença nesses animais e considerar a tuberculose em pinípedes como uma das principais preocupações de saúde para essas espécies. Investigações da infecção e da doença clínica devem ser conduzidas ao introduzir animais em instalações de zoológicos e estudos envolvendo populações livres devem ser encorajados.

Tal como com outras espécies de MTBC, os genomas de *M. pinnipedii* estão em decaimento evolutivo através da perda de regiões genômicas específicas e da perda de genes devido a pseudogenização. Estes resultados foram ainda apoiados pelo achado de PSGs que não estão presentes nos genomas antigos de *M. pinnipedii* quando comparados com as estirpes modernas de *M. pinnipedii*. O remodelamento do genoma de *M. pinnipedii* afeta os conhecidos fatores de virulência da MTBC, com potencial impacto na adaptabilidade do hospedeiro e no desfecho da doença, como a deleção de MiD3 e MiD4 e regiões contendo genes de PE/PPE. No entanto, o proteoma das estirpes estudadas apresentou alto grau de conservação, indicando que genomas adicionais precisam ser sequenciados e analisados para aumentar a chance de encontrar diferenças significativas ou que o remodelamento do genoma de *M. pinnipedii* ocorreu antes da especiação bacteriana. Além disso, deve-se dar atenção especial a possíveis pseudogenes e CDSs truncados, a fim de padronizar as anotações do genoma e orientar futuras análises baseadas em genes. E finalmente, estirpes de *M. pinnipedii*, como em outras espécies de MTBC, provavelmente se agruparam com base na ocorrência geográfica, que coincidirá com espécies de pinípedes, já que essas populações de animais já estão segregadas por geografia.

Referências

ALLIX-BÉGUEC, C.; HARMSEN, D.; WENIGER, T.; SUPPLY, P.; NIEMANN, S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. **Journal of clinical microbiology**, v. 46, n. 8, p. 2692–9, 1 ago. 2008.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.

ARCUS, V. L.; RAINEY, P. B.; TURNER, S. J. The PIN-domain toxin–antitoxin array in mycobacteria. Trends in Microbiology, v. 13, n. 8, p. 360–365, 1 ago. 2005. ARKIN, A. P.; COTTINGHAM, R. W.; HENRY, C. S.; HARRIS, N. L.; STEVENS, R. L.; MASLOV, S.; DEHAL, P.; WARE, D.; PEREZ, F.; CANON, S.; SNEDDON, M. W.; HENDERSON, M. L.; RIEHL, W. J.; MURPHY-OLSON, D.; CHAN, S. Y.; KAMIMURA, R. T.; KUMARI, S.; DRAKE, M. M.; BRETTIN, T. S.; GLASS, E. M.; CHIVIAN, D.; GUNTER, D.; WESTON, D. J.; ALLEN, B. H.; BAUMOHL, J.; BEST, A. A.; BOWEN, B.; BRENNER, S. E.; BUN, C. C.; CHANDONIA, J.-M.; CHIA, J.-M.; COLASANTI, R.; CONRAD, N.; DAVIS, J. J.; DAVISON, B. H.; DEJONGH, M.; DEVOID, S.; DIETRICH, E.; DUBCHAK, I.; EDIRISINGHE, J. N.; FANG, G.; FARIA, J. P.; FRYBARGER, P. M.; GERLACH, W.; GERSTEIN, M.; GREINER, A.; GURTOWSKI, J.; HAUN, H. L.; HE, F.; JAIN, R.; JOACHIMIAK, M. P.; KEEGAN, K. P.; KONDO, S.; KUMAR, V.; LAND, M. L.; MEYER, F.; MILLS, M.; NOVICHKOV, P. S.; OH, T.; OLSEN, G. J.; OLSON, R.; PARRELLO, B.; PASTERNAK, S.; PEARSON, E.; POON, S. S.; PRICE, G. A.; RAMAKRISHNAN, S.; RANJAN, P.; RONALD, P. C.; SCHATZ, M. C.; SEAVER, S. M. D.; SHUKLA, M.; SUTORMIN, R. A.; SYED, M. H.; THOMASON, J.; TINTLE, N. L.; WANG, D.; XIA, F.; YOO, H.; YOO, S.; YU, D. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnology, v. 36, n. 7, p. 566–569, 2018.

AZIZ, R. K.; BARTELS, D.; BEST, A. A.; DEJONGH, M.; DISZ, T.; EDWARDS, R. A.; FORMSMA, K.; GERDES, S.; GLASS, E. M.; KUBAL, M.; MEYER, F.; OLSEN, G. J.; OLSON, R.; OSTERMAN, A. L.; OVERBEEK, R. A.; MCNEIL, L. K.; PAARMANN, D.; PACZIAN, T.; PARRELLO, B.; PUSCH, G. D.; REICH, C.; STEVENS, R.; VASSIEVA, O.; VONSTEIN, V.; WILKE, A.; ZAGNITKO, O. The RAST Server: Rapid Annotations using Subsystems Technology. **BMC Genomics**, v. 9, n. 1, p. 75, 2008.

BANKEVICH, A.; NURK, S.; ANTIPOV, D.; GUREVICH, A. A.; DVORKIN, M.; KULIKOV, A. S.; LESIN, V. M.; NIKOLENKO, S. I.; PHAM, S.; PRJIBELSKI, A. D.; PYSHKIN, A. V; SIROTKIN, A. V; VYAHHI, N.; TESLER, G.; ALEKSEYEV, M. A.; PEVZNER, P. A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology**: a journal of computational molecular cell biology, v. 19, n. 5, p. 455–77, 2012.

BASTIDA, R.; LOUREIRO, J.; QUSE, V.; BERNARDELLI, A.; RODRÍGUEZ, D.; COSTA, E. Tuberculosis in a Wild Subantarctic Fur Seal from Argentina. **Journal of Wildlife Diseases**, v. 35, p. 796–798, 1999.

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

BERNARDELLI, A.; BASTIDA, R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; BERNARDELLI BASTIDA, A. R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; MUNDO MARINO, F.; CLEMENTE DEL TUYÚ, S.; BERNARDELLI, A.; BASTIDA, R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; BERNARDELLI BASTIDA, A. R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; MUNDO MARINO, F.; CLEMENTE DEL TUYÚ, S.; BERNARDELLI, A.; BASTIDA, R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E. Tuberculosis in sea lions and fur seals from the south-western Atlantic coast. **Revue scientifique et technique (International Office of Epizootics)**, v. 15, n. 3, p. 985–1005, 1996.

BLAIR WR. Report of the veterinarian on the mammals. **17th Annual Report of the New York Zoological Society**, v. 74, p. 73–77, 1913.

BOARDMAN, W. S. J.; SHEPHARD, L.; BASTIAN, I.; GLOBAN, M.; FYFE, J. A. M.; COUSINS, D. V.; MACHADO, A.; WOOLFORD, L. Mycobacterium pinnipedii Tuberculosis in a Free-Ranging Australian Fur Seal (Arctocephalus pusillus doriferus) In South Australia. **Journal of Zoo and Wildlife Medicine**, v. 45, n. 4, p. 970–972, 2014.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics (Oxford, England)**, v. 30, n. 15, p. 2114–20, 2014.

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

BOS, K. I.; HARKINS, K. M.; HERBIG, A.; COSCOLLA, M.; WEBER, N.; COMAS, I. I. I.; FORREST, S. A.; BRYANT, J. M.; HARRIS, S. R.; SCHUENEMANN, V. J.; CAMPBELL, T. J.; MAJANDER, K.; WILBUR, A. K.; GUICHON, R. A.; WOLFE STEADMAN, D. L.; COOK, D. C.; NIEMANN, S.; BEHR, M. A.; ZUMARRAGA, M.; BASTIDA, R.; HUSON, D.; NIESELT, K.; YOUNG, D.; PARKHILL, J.; BUIKSTRA, J. E.; GAGNEUX, S.; STONE, A. C.; KRAUSE, J. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. **Nature**, v. 514, n. 7523, p. 494–497, 2014.

BRENNAN, M. J. The Enigmatic PE/PPE Multigene Family of Mycobacteria and Tuberculosis Vaccination. **Infection and Immunity**, v. 85, n. 6, p. e00969-16, 2017.

BRITES, D.; GAGNEUX, S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. **Immunological Reviews**, v. 264, n. 1, p. 6–24, 2015.

CHEN, L.; YANG, J.; YU, J.; YAO, Z.; SUN, L.; SHEN, Y.; JIN, Q. VFDB: a reference database for bacterial virulence factors. **Nucleic Acids Research**, v. 33, n. Database issue, p. D325–D328, 2004.

CINGOLANI, P.; PLATTS, A.; WANG, L. L.; COON, M.; NGUYEN, T.; WANG, L.; LAND, S. J.; LU, X.; RUDEN, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. **Fly**, v. 6, n. 2, p. 80–92, 2012.

COHEN, T.; VAN HELDEN, P. D.; WILSON, D.; COLIJN, C.; MCLAUGHLIN, M. M.; ABUBAKAR, I.; WARREN, R. M. Mixed-Strain Mycobacterium tuberculosis Infections and the Implications for Tuberculosis Treatment and Control. **Clinical microbiology reviews**, v. 25, n. 4, 2012.

COHEN, T.; WILSON, D.; WALLENGREN, K.; SAMUEL, E. Y.; MURRAY, M. Mixedstrain Mycobacterium tuberculosis infections among patients dying in a hospital in KwaZulu-Natal, South Africa. **Journal of clinical microbiology**, v. 49, n. 1, p. 385–8, 2011.

COLE, S. T.; BROSCH, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON, S. V.; EIGLMEIER, K.; GAS, S.; BARRY, C. E.; 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, M. A.; RAJANDREAM, M.-A.; ROGERS, J.; RUTTER, S.; SEEGER, K.; SKELTON, J.; SQUARES, R.; SQUARES, S.; SULSTON, J. E.; TAYLOR, K.; WHITEHEAD, S.; BARRELL, B. G. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. **Nature**, v. 393, n. 6685, p. 537–544, 1998.

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

COUSINS, D. V.; BASTIDA, R.; CATALDI, A.; QUSE, V.; REDROBE, S.; DOW, S.; DUIGNAN, P.; MURRAY, A.; DUPONT, C.; AHMED, N.; COLLINS, D. M.; BUTLER, W. R.; DAWSON, D.; RODRÍGUEZ, D.; LOUREIRO, J.; ROMANO, M. I.; ALITO, A.; ZUMARRAGA, M.; BERNARDELLI, A. Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 53, n. 5, p. 1305–1314, 2003.

COUSINS, D. V.; WILLIAMS, S. N.; REUTER, R.; FORSHAW, D.; CHADWICK, B.; COUGHRAN, D.; COLLINS, P.; GALES, N. Tuberculosis in wild seals and characterisation of the seal bacillus. **Australian veterinary journal**, v. 70, n. 3, p. 92–97, 1993.

COUSINS, D. V. ELISA for detection of tuberculosis in seals. **The Veterinary record**, v. 121, n. 13, p. 305, 1987.

COUSINS, D. V; FRANCIS, B. R.; GOW, B. L.; COLLINS, D. M.; MCGLASHAN, C. H.; GREGORY, A.; MACKENZIE, R. M. Tuberculosis in captive seals: bacteriological studies on an isolate belonging to the Mycobacterium tuberculosis complex. **Research in veterinary science**, v. 48, n. 2, p. 196–200, 1990.

COUSINS, D. V. V; PEET, R. L. L.; GAYNOR, W. T. T.; WILLIAMS, S. N. N.; GOW, B. L. L. Tuberculosis in imported hyrax (Procavia capensis) caused by an unusual variant belonging to the Mycobacterium tuberculosis complex. **Veterinary microbiology**, v. 42, n. 2–3, p. 135–45, 1994.

DE AMORIM, D. B.; CASAGRANDE, R. A.; ALIEVI, M. M.; WOUTERS, F.; DE OLIVEIRA, L. G. S.; DRIEMEIER, D.; TAVARES, M.; IKUTA, C. Y.; TELLES, E. O.; FERREIRA-NETO, J. S. Mycobacterium pinnipedii in a stranded South American sea lion (Otaria byronia) in Brazil. **Journal of Wildlife Diseases**, v. 50, n. 2, p. 419-422, 2014.

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

FORSHAW, D.; PHELPS, G. R. TUBERCULOSIS IN A CAPTIVE COLONY OF PINNIPEDS. **Journal of Wildlife Diseases**, v. 27, n. 2, p. 288–295, 24 abr. 1991.

GAGNEUX, S. Host—pathogen coevolution in human tuberculosis. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 367, n. 1590, p. 850-859, 2012.

GARRISON, E.; MARTH, G. Haplotype-based variant detection from short-read sequencing. **arXiv preprint arXiv:1207.3907**, 17 jul. 2012. Disponível em: http://arxiv.org/abs/1207.3907>.

GEY VAN PITTIUS, N. C.; SAMPSON, S. L.; LEE, H.; KIM, Y.; VAN HELDEN, P. D.; WARREN, R. M. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. **BMC Evolutionary Biology**, v. 6, n. 1, p. 95, 2006.

GONZALES-VIERA, O.; MARIGO, J.; RUOPPOLO, V.; ROSAS, F. C. W.; KANAMURA, C. T.; TAKAKURA, C.; FERNÁNDEZ, A.; CATÃO-DIAS, J. L. Toxoplasmosis in a Guiana dolphin (Sotalia guianensis) from Paraná, Brazil. **Veterinary Parasitology**, v. 191, n. 3–4, p. 358–362, 2013.

HIGGINS, R. Bacteria and fungi of marine mammals: a review. **The Canadian veterinary journal = La revue veterinaire canadienne**, v. 41, n. 2, p. 105–16, 2000.

HUERTA-CEPAS, J.; SZKLARCZYK, D.; FORSLUND, K.; COOK, H.; HELLER, D.; WALTER, M. C.; RATTEI, T.; MENDE, D. R.; SUNAGAWA, S.; KUHN, M.; JENSEN, L. J.; VON MERING, C.; BORK, P. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic acids research**, v. 44, n. D1, p. D286-93, 2016.

JENSEN, L. J.; SIMONOVIC, M.; BORK, P.; VON MERING, C.; MULLER, J.; STARK, M.; DOERKS, T.; ROTH, A.; JULIEN, P.; KUHN, M.; CHAFFRON, S.; CREEVEY, C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. **Nucleic Acids Research**, v. 37, n. Database, p. D412–D416, 2008.

JURCZYNSKI, K.; LYASHCHENKO, K. P.; GOMIS, D.; MOSER, I.; GREENWALD, R.; MOISSON, P. Pinniped Tuberculosis in Malayan Tapirs (*Tapirus indicus*) and its Transmission to Other Terrestrial Mammals. **Journal of Zoo and Wildlife Medicine**, v. 42, n. 2, p. 222–227, 2011.

KIERS, A.; KLARENBEEK, A.; MENDELTS, B.; VAN SOOLINGEN, D.; KOËTER, G. Transmission of Mycobacterium pinnipedii to humans in a zoo with marine mammals. **The International Journal of Tuberculosis and Lung Disease**, v. 12, p. 1469–1473, 2008.

KRIZ, P.; KRALIK, P.; SLANY, M.; SLANA, I.; SVOBODOVA, J.; PARMOVA, I.; BARNET, V.; JUREK, V.; PAVLIK, I. Case Report Mycobacterium pinnipedii in a captive Southern sea lion (Otaria flavescens): a case report Private veterinary practitioner, Czech Republic. **Veterinarni Medicina**, v. 56, n. 6, p. 307–313, 2011.

- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754–1760, 2009.
- LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R.; 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.
- LI, L.; STOECKERT, C. J.; ROOS, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. **Genome Research**, v. 13, n. 9, p. 2178–2189, 2003.
- LOEFFLER, S. H.; DE LISLE, G. W.; NEILL, M. A.; COLLINS, D. M.; PRICE-CARTER, M.; PATERSON, B.; CREWS, K. B. The seal tuberculosis agent, Mycobacterium pinnipedii, infects domestic cattle in New Zealand: epidemiologic factors and DNA strain typing.

 Journal of Wildlife Diseases, v. 50, n. 2, p. 180–187, 2014.
- RAMAGE, H. R.; CONNOLLY, L. E.; COX, J. S. Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. **PLoS Genetics**, v. 5, n. 12, p. e1000767, 2009.
- RIOJAS, M. A.; MCGOUGH, K. J.; RIDER-RIOJAS, C. J.; RASTOGI, N.; HAZBÓN, M. H. Phylogenomic analysis of the species of the Mycobacterium tuberculosis complex demonstrates that Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of Mycob. **International Journal of Systematic and Evolutionary Microbiology**, v. 68, n. 1, p. 324–332, 2017.
- RISSMAN, A. I.; MAU, B.; BIEHL, B. S.; DARLING, A. E.; GLASNER, J. D.; PERNA, N. T. Reordering contigs of draft genomes using the Mauve Aligner. **Bioinformatics**, v. 25, n. 16, p. 2071–2073, 2009.
- RUOPPOLO, V. **Patologia comparada de cetáceos e pinípedes.** [Comparative pathology of cetaceans and pinnipeds]. 2003. 131 f. Dissertação (Mestrado em Patologia Experimental e Comparada) Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2003.
- STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, v. 30, n. 9, p. 1312–1313, 2014.
- SUPPLY, P.; ALLIX, C.; LESJEAN, S.; CARDOSO-OELEMANN, M.; RUSCH-GERDES, S.; WILLERY, E.; SAVINE, E.; DE HAAS, P.; VAN DEUTEKOM, H.; RORING, S.; BIFANI, P.; KUREPINA, N.; KREISWIRTH, B.; SOLA, C.; RASTOGI, N.; VATIN, V.;

GUTIERREZ, M. C.; FAUVILLE, M.; NIEMANN, S.; SKUCE, R.; KREMER, K.; LOCHT, C.; VAN SOOLINGEN, D. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. **Journal of Clinical Microbiology**, v. 44, n. 12, p. 4498–4510, 2006.

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

THOREL, M. F.; KAROUI, C.; VARNEROT, A.; FLEURY, C.; VINCENT, V. Isolation of Mycobacterium bovis from baboons, leopards and a sea-lion. **Veterinary research**, v. 29, n. 2, p. 207–12, 1998.

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

WALKER, T. M.; IP, C. L.; HARRELL, R. H.; EVANS, J. T.; KAPATAI, G.; DEDICOAT, M. J.; EYRE, D. W.; WILSON, D. J.; HAWKEY, P. M.; CROOK, D. W.; PARKHILL, J.; HARRIS, D.; WALKER, A. S.; BOWDEN, R.; MONK, P.; SMITH, E. G.; PETO, T. E. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. **The Lancet Infectious Diseases**, v. 13, n. 2, p. 137–146, 2013.

WATTAM, A. R.; DAVIS, J. J.; ASSAF, R.; BOISVERT, S.; BRETTIN, T.; BUN, C.; CONRAD, N.; DIETRICH, E. M.; DISZ, T.; GABBARD, J. L.; GERDES, S.; HENRY, C. S.; KENYON, R. W.; MACHI, D.; MAO, C.; NORDBERG, E. K.; OLSEN, G. J.; MURPHY-OLSON, D. E.; OLSON, R.; OVERBEEK, R.; PARRELLO, B.; PUSCH, G. D.; SHUKLA, M.; VONSTEIN, V.; WARREN, A.; XIA, F.; YOO, H.; STEVENS, R. L. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. **Nucleic Acids Research**, v. 45, n. D1, p. D535–D542, 2017.

WELLS, A. Q. Tuberculosis In Wild Voles. **The Lancet**, v. 229, n. 5934, p. 1221, 1937.

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

WINTHER, K.; TREE, J. J.; TOLLERVEY, D.; GERDES, K. VapCs of Mycobacterium tuberculosis cleave RNAs essential for translation. **Nucleic Acids Research**, v. 44, n. 20, p. 9860–9871, 2016.

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

YANG, T.; SHENG, Y.; YUE, L.; DING, N.; WANG, G.; ZHONG, J.; XIAO, J.; YU, X.; LI, C.; CHEN, F.; ZHANG, J.; SUN, Y.; LIANG, Q.; JIA, X.; HUANG, H.; MA, G. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. **Frontiers in Microbiology**, v. 9, p. 1–12, 2018.

ZIMPEL, C. K.; BRANDÃO, P. E.; SOUZA FILHO, A. F. de; DE SOUZA, R. F.; YKUTA, C. Y.; SOARES FERREIRA NETO, J.; SOLER CAMARGO, N. C.; BRYAN HEINEMANN, M.; GUIMARAES, A. M. S. Complete genome sequencing of Mycobacterium bovis SP38 and comparative genomics of Mycobacterium bovis and M. tuberculosis strains. **Frontiers in Microbiology**, v. 8, p. 2389, 2017.

5. CAPÍTULO 2: Influência da arquitetura gênica na distribuição e conservação de epítopos lineares de células T e fatores de virulência do Complexo *Mycobacterium tuberculosis*

RESUMO

O Complexo Mycobacterium tuberculosis (MTBC) trata-se de um grupo de bactérias conhecido por causar tuberculose (TB) em humanos e animais. Comparações abrangentes da influência do conteúdo gênico de todos os membros do MTBC no metabolismo, tropismo por hospedeiros, conservação e distribuição de epítopos de células T e fatores de virulência ainda não estão disponíveis. Por isso, o objetivo desse trabalho foi definir o pan-genoma, genoma core, genoma acessório e genoma estirpe-específico de 205 genomas de MTBC, e investigar a influência dessa arquitetura genômica na distribuição de epítopos lineares de células T preditos in silico e de fatores de virulência nessas bactérias. As 866.916 proteínas dos 205 genomas de estirpes representantes do MTBC analisadas estruturaram um pan-genoma com 4.528 grupos de proteínas ortólogas, sendo que 3.944 (87,10%; total de proteínas: 832.234/96,00%) grupos de proteínas ortólogas compõem o genoma *core*, 575 (12,70%; total de proteínas: 30.286/3,49%) grupos de proteínas ortólogas compõem o genoma acessório, e 9 grupos ortólogos (com um total de 160 proteínas) e 4.236 proteínas únicas (0,20%, total de proteínas: 4.396/0,51%) compõem o genoma estirpe-específico. Apesar de algumas vias metabólicas relacionadas ao metabolismo de lipídeos e metabólitos secundários é conservado no genoma core, outras são variavelmente presente entre as estirpes, sugerindo a importância da modulação lipídica do envelope celular na variabilidade fenotípica de estirpes do MTBC. A análise de predição de epítopos mostrou que 12.970 epítopos foram encontrados no genoma core, 961 epítopos no genoma acessório e 320 no genoma estirpe-específico. Nossos resultados evidenciam, pela primeira vez, um mecanismo de variação antigênica de epítopos de célula T do MTBC baseado no silenciamento/perda de genes. A presença de 961 epítopos no genoma acessório sugere que os diferentes isolados de MTBC são capazes de evadir o sistema imune por meio de variação antigênica de fase. A evasão do sistema imune por meio da variação de epítopos de células T não é tão insignificante como se pensava, e a identificação de epítopos variáveis deverá ser avaliada no desenvolvimento de estratégias vacinais para o controle da TB. A análise comparativa entre diferentes bactérias quanto a presença dos fatores de virulência (FVs) em espécies do MTBC mostrou que, apesar da grande similaridade compartilhada entre o grupo, há variações significativas na quantidade de genes relacionados aos fatores de virulência principalmente em espécies de Mycobacterium pinnipedii e algumas estirpes de Mycobacterium

59

africanum. Destaca-se que algumas estirpes de MTBC apresentam variações no número de

genes associados aos quatro operons Mce (1-4), em menor ou maior escala dependendo do

operon. Importantes variações no número de FVs observadas intra- e entre-espécies do MTBC

estão diretamente relacionadas ao sistema de secreção do tipo VII (ESX), resistência às

perturbações ambientais e modulação do envoltório bacteriano via metabolismo de lipídeos,

sugerindo que tais genes desempenham papel importante no estabelecimento dos diferentes

fenótipos do MTBC. Resultados obtidos nesse trabalho mostram que M. pinnipedii, a única

espécie do MTBC infectando animais marinhos, possui variações no número de genes

relacionados ao regulon DevR/DosR, associado ao mecanismo de dormência, do que as outras

espécies do MTBC. Este e outros achados descritos neste trabalho pode estar relacionado a uma

co-evolução hospedeira, permitindo que o M. pinnipedii sobreviva às condições fisiológicas

dos pinípedes.

Palavras-chave: Complexo Mycobacterium tuberculosis, epítopos, fatores de virulência.

5.1. INTRODUÇÃO

O Complexo Mycobacterium tuberculosis (MTBC) trata-se de um grupo de bactérias conhecido por causar tuberculose (TB) em humanos e animais, e por apresentar altos índices de mortalidade, com 1,6 milhões de mortes e 10 milhões de novos casos de TB em humanos somente em 2017 (WORLD HEALTH ORGANIZATION, 2018). Apesar das espécies e linhagens do MTBC possuírem grande similaridade entre suas sequências genômicas alinháveis (>99,95%) (BROSCH et al., 2000), essas bactérias exibem diferentes fenótipos que promovem diversidade nos mecanismos de associação ao hospedeiros (COMAS et al., 2013; GALAGAN, 2014), além de diferentes graus de virulência (PIROFSKI; CASADEVALL, 2012). Por possuir uma evolução clonal, as diferenças genéticas do MTBC se dão apenas por polimorfismos de nucleotídeo único (SNPs; transições e transversões), indels (inserções e deleções curtas), deleções longas de até ~12Kb, elementos de inserção (IS - insertion sequences) e duplicação gênica de poucas famílias de genes parálogos. Diferente da maioria das bactérias, os membros do MTBC não realizam transferência horizontal de genes (BORITSCH, 2016), por causa disso, regiões genômicas perdidas devido a deleções não podem ser readquiridas (GAGNEUX et al., 2006). Porém, análises comparativas de conteúdo gênico sugerem que a perda e ganho de genes é uma importante fonte de variação genética em espécies clonais, trazendo tanta diversidade no conteúdo gênico quanto a transferência horizontal de genes. Tal fenômeno ocorre devido a processos de decay (decaimento) genômico (por pseudogenização/truncamento de genes ou deleções) e duplicações gênicas, ambos já relatados em MTBC (BAUMLER; FANG, 2013; BOLOTIN; HERSHBERG, 2015; ZIMPEL et al., 2017). A pseudogenização e o truncamento de genes (com possível neofuncionalização), por sua vez, se dão pela ocorrência de SNPs e indels, e acredita-se que possam estar associados à adaptação do patógeno ao hospedeiro e padrões de virulência. Essas variações genéticas, quando acompanhadas de efeitos positivos, podem fornecer vantagem seletiva e serem fixadas na população (FARHAT et al., 2013; ELDHOLM et al., 2015).

A perda e ganho de genes só pode ser estimada por meio do estudo do conteúdo gênico de cada estirpe em um contexto populacional. Assim, a análise do pan-genoma constitui uma ferramenta importante para estimar a diversidade do repertório gênico, explorando características gerais e específicas de diferentes grupos bacterianos (VERNIKOS et al., 2015). A grande maioria dos dados relacionados ao MTBC se concentram em análises para detecção de indels e SNPs para compreender particularidades do *M. tuberculosis* (BROSCH et al., 1999; KATO-MAEDA et al., 2001; FLEISCHMANN et al., 2002; UPLEKAR et al., 2011; DAUM et al., 2012; LIU et al., 2014; LUO et al., 2014; ALI et al., 2015; BLACK et al., 2015; REGMI

et al., 2015; BROWN et al., 2015; PHELAN et al., 2016; RODRÍGUEZ-CASTILLO et al., 2017; YAR et al., 2018; FAKSRI et al., 2018). Até o momento, estudos que avaliaram o conteúdo gênico do MTBC foram focados apenas em estirpes de M. tuberculosis, M. africanum, M. canettii, M. microti e/ou M. bovis (BENTLEY et al., 2012; NAMOUCHI et al., 2012; COSCOLLA et al., 2013; RAHMAN et al., 2014; DE LA FUENTE et al., 2015; DIPPENAAR et al., 2015; PERIWAL et al., 2015; ZIMPEL et al., 2017; YANG et al., 2018). Além do mais, as espécies M. pinnipedii e "dassie bacillus" nunca foram analisadas por genômica comparativa de SNPs/indels e muito menos por repertório gênico. A maioria dos estudos que envolveram o conteúdo gênico do MTBC (ISAZA et al., 2012; DEJESUS; SACCHETTINI; IOERGER, 2013; SUPPLY et al., 2013; LIU et al., 2014; DIPPENAAR et al., 2015; JIA et al., 2017; XIONG et al., 2017; YANG et al., 2018) ainda utilizaram plataformas de anotação gênica que não possuem detecção de pseudogenes e que consideram genes truncados ao final de contigs (i.e. artefatos de montagem) como verdadeiros, o que acaba superestimando o número de genes estirpe-específicos e do genoma acessório. Assim, comparações abrangentes da influência do conteúdo gênico de todos os membros do MTBC no metabolismo, tropismo por hospedeiros, conservação e distribuição de epítopos de células T e fatores de virulência ainda não estão disponíveis.

Análises comparativas de estirpes geneticamente e geograficamente representativas de M. tuberculosis indicam que os epítopos lineares de células T conhecidos e preditos são altamente conservados nestes genomas, isto é, não existem variações em suas sequências de amino ácidos (COMAS et al., 2010; COSCOLLA et al., 2015). No entanto, o efeito da perda gênica na expressão desses epítopos não foi avaliado. Não se sabe se esses epítopos estariam presentes em genes com tendência a truncamento ou pseudogenização, o que levaria à um possível silenciamento destes epítopos e modulação da resposta imune frente ao patógeno. No processo de invasão do organismo hospedeiro, a interação patógeno-hospedeiro promove o desenvolvimento de respostas imune adaptativas mediadas por células T, consideradas primordial para o combate a infecção por MTBC (ERNST, 2012). Esse reconhecimento do patógeno pelas células especializadas do sistema imune acontece de forma específica e restrita, considerando a afinidade da ligação epítopo-alelo MHC. Variações em sequências ou de fase desses epítopos poderiam impedir o reconhecimento bacteriano, prejudicando o desenvolvimento da resposta imune adaptativa mediada pelas células T e, consequentemente, o desfecho da infecção (ERNST, 2012, 2017). Portanto, o objetivo desse trabalho foi definir o pan-genoma, genoma core, genoma acessório e genoma estirpe-específico de 205 genomas de MTBC, compreendendo a maioria das espécies do Complexo, e investigar a influência dessa arquitetura genômica na distribuição de epítopos lineares de células T preditos *in silico* e de fatores de virulência nessas bactérias.

5.2.MATERIAL E MÉTODOS

5.2.1. Seleção dos genomas de MTBC

Para garantir uma melhor qualidade e padronização de predição gênica e anotação genômica dos genomas selecionados, todos os genomas completos e incompletos (drafts) foram recuperados do banco de dados Reference Sequence (RefSeq) disponível na plataforma do National Center for Biotechnology Information (NCBI). Os genomas incompletos utilizados e os números mínimo e máximo de *contigs* por espécie estão sumarizados na Tabela 1. O conjunto de dados públicos fornecido pelo RefSeq (NCBI) é construído a partir das sequências disponíveis no Genbank que são curados pela equipe do NCBI, nas quais mutações desconhecidas, erros de sequenciamento, artefatos de clonagem e anotações errôneas são evitadas. O Refseq fornece registros de CDSs precisos e completos, analisando as sequências com base nos seguintes parâmetros de qualidade: o CDS deve ter mais de 100 aminoácidos; não pode apresentar códon de início inválido ou códon de parada interno; avalia se há baixa complexidade da sequência; avalia se as sequências tem ou não semelhança com outros registros de proteínas; e avalia se o transcrito tem inserções, deleções ou transições/transversões em relação à sequência do genoma de referência. Esses parâmetros são levados em consideração em passos posteriores para analisar se o erro relatado deve ser ignorado ou se houve um erro de montagem na determinada região, definindo também quais devem ser considerados pseudogenes e não serem reportados (PRUITT; TATUSOVA; MAGLOTT, 2007)

Assim, todos os genomas de MTBC disponíveis no RefSeq em outubro de 2018 que foram anotados com a pipeline do *Prokaryotic Genome Annotation Pipeline* (PGAP) do NCBI e correspondiam à espécie indicada no depósito (as espécies foram confirmadas por meio da identificação *in silico* das RDs correspondentes) foram incluídos neste estudo. Assim, 113 genomas de *Mycobacterium tuberculosis* (todos completos), 51 genomas de *Mycobacterium dividos* (dois completos e 49 incompletos), 28 genomas de *Mycobacterium africanum* (um completo e 27 incompletos), quatro genomas de "dassie bacillus" (todos incompletos), três genomas de *Mycobacterium pinnipedii* (todos incompletos), três genomas de *Mycobacterium orygis*, e um genoma incompleto de *Mycobacterium orygis*, e um genoma incompleto de *Mycobacterium microti* foram incluídos na análise (APÊNDICE D), totalizando 205 genomas. Infelizmente, genomas de *M. suricattae* e "chimpanzee bacillus" estão disponíveis apenas como *reads*, não tendo sido montados e anotados com PGAP para serem disponibilizados no RefSeq.

Tabela 1. Número	mínimo e máxim	no de contigs e	encontrados nos	genomas incompletos
por espécie				

	Número de contigs				
Espécie	Mínimo	Máximo			
Mycobacterium pinnipedii	116	226			
Mycobacterium bovis	28	191			
Mycobacterium africanum	5	97			
Dassie bacillus	193	260			
Mycobacterium caprae	2	141			
Mycobacterium mungi*	110	-			
Mycobacterium orygis*	108	-			

^{*} As espécies apresentam apenas um genoma disponível no banco de dados

5.2.2. Construção dos genomas core, acessório e estirpe-específico

O proteoma de cada genoma foi recuperado do RefSeq e as relações de homologia entre as sequências de proteínas foram determinadas pelo programa OrthoFinder (EMMS; KELLY, 2015). As análises permitiram a construção do pan-genoma (todas as proteínas presentes nas estirpes, i.e. grupos de ortólogos e proteínas estirpe-específicas), genoma *core* (proteínas comuns a 95% das estirpes ou mais, isto é, presentes em 195 até 205 estirpes), genoma acessório (proteínas presentes em duas ou até 195 estirpes) e genoma espécie-específico (proteínas presentes em apenas uma estirpe). Devido a presença de genomas na forma de *drafts* (incompletos), optou-se por computar o genoma *soft-core* (95% das estirpes) ao invés do *hard-core* (99% das estirpes). A análise da classificação funcional das proteínas que compõem o pangenoma foi realizado com o programa eggNOG (HUERTA-CEPAS et al., 2016) que classifica cada proteína seguindo a metodologia de *Cluster of Orthologous Groups* (COG) (TATUSOV et al., 2000).

5.2.3. Predição de epítopos lineares

O método de predição padrão de epítopos lineares do NetMHCpan-4.0 (JURTZ et al., 2017) foi utilizado para a predição nas sequências proteicas de interesse. O documento de *input* foi o conjunto de proteínas dos genomas *core*, acessório e estirpe específico, separadamente. O programa utiliza o método "rede neural artificial" (ANN) para realizar análises quanto à afinidade de possíveis epítopos lineares com o alelo supertipo representativo da molécula de *Human Leukocyte Antigen* (HLA) (HLA-A02:01), selecionando os epítopos com 9

aminoácidos. Foram selecionados apenas os epítopos que apresentaram potencial de forte ligação à molécula HLA, definidos a partir de um valor de score (% Rank < 0,5).

5.2.3.1.Distribuição e conservação de epítopos lineares

A análise da distribuição do número de epítopos lineares preditos por proteína foi realizada utilizando um *script* personalizado escrito em *python* para manipular e contabilizar os dados obtidos. O pacote do R Balcony 0.2.10 (PŁUCIENNIK et al., 2018) foi utilizado para a construção de um *script* desenvolvido para a análise de variabilidade de epítopos em sequências de proteínas dos genomas *core* e acessório do MTBC. Os epítopos preditos da mesma proteína foram analisados quanto a presença de variações nas sequências das diferentes estirpes. As avaliações gráficas foram realizadas usando o pacote ggplot2 (GINESTET, 2011) compatível com o programa R (R CORE TEAM, 2013). As variações de aminoácidos em epítopos preditos selecionados foram visualizadas com o WebLogo 3 (CROOKS et al., 2004).

5.2.4. Fatores de virulência

O programa VFanalyzer (LIU et al., 2019) foi utilizado para identificar fatores de virulência nos 205 genomas analisados. Foram realizadas buscas por similaridade nas sequências de interesse usadas como *input* (arquivo *gbff* de cada estirpe) dentro do banco de dados de fatores de virulência (VFDB, *virulence fator database*) (CHEN et al., 2004). Os resultados obtidos a partir do conjunto de dados do VFDB foram utilizados para estruturar uma tabela numérica baseada na quantidade de genes relacionados aos fatores de virulência presentes em todas as estirpes selecionadas no estudo. O *heatmap* foi gerado usando o pacote gplots R (WARNES et al., 2016).

5.3.RESULTADOS E DISCUSSÃO

As 866.916 proteínas dos 205 genomas de estirpes representantes do MTBC analisadas estruturaram um pan-genoma com 4.528 grupos de proteínas ortólogas, contendo 862.680 proteínas, e 4.236 proteínas estirpe-específicas. Destes, 3.944 (87,10%; total de proteínas: 832.234/96,00%) grupos de proteínas ortólogas compõem o genoma core, 575 (12,70%; total de proteínas: 30.286/3,49%) grupos de proteínas ortólogas compõem o genoma acessório, e 9 grupos ortólogos (com um total de 160 proteínas) e 4.236 proteínas únicas (0,20%, total de proteínas: 4.396/0,51%) compõem o genoma estirpe-específico (Figura 1). Resultados encontrados numa análise estrutural do pan-genoma de M. tuberculosis corroboram com nossos resultados, concordando com a hipótese de que trata-se de uma espécie clonal, no qual o genoma *core* representou 80% da média em um conjunto de dados com 1.595 estirpes de M. tuberculosis (KAVVAS et al., 2018). Em nosso estudo, mesmo após incluir outras espécies do MTBC, a proporção de proteínas no genoma core continuou alta. Por outro lado, Periwal e colaboradores analisaram 96 genomas de MTBC e Mycobacterium canettii e reportaram um genoma core razoavelmente menor do que o genoma acessório, com 3.374 e 4.725 grupos de ortólogos, respectivamente (PERIWAL et al., 2015). Provavelmente este resultado está relacionado ao fato de que genomas do M. canettii foram incluídos nas análises. Apesar da proximidade entre essas espécies, M. canettii apresenta mecanismos diferentes de recombinação e transferência horizontal de genes, sugerindo apenas uma relação de ancestralidade com o MTBC (SUPPLY et al., 2013). Outro fator que pode ter influenciado nas análises de Periwal e colaboradores é o fato de a anotação utilizada considerar pseudogenes como genes verdadeiros. Pseudogenes (uma média aproximada de 120 por genoma) não são reportados nos proteomas de genomas depositados no NCBI, que incorpora em sua pipeline PGAP mecanismos especiais de detecção de pseudogenes. Neste estudo, a comparação de genomas anotados em uma única plataforma e oriundos do RefSeq garantem que os dados gerados são provenientes da melhor anotação possível com os dados disponíveis. Assim, conclui-se que os números apresentados aqui retratam de maneira mais adequada o conjunto de genes do MTBC e as consequências de sua evolução clonal na variabilidade do conteúdo gênico.

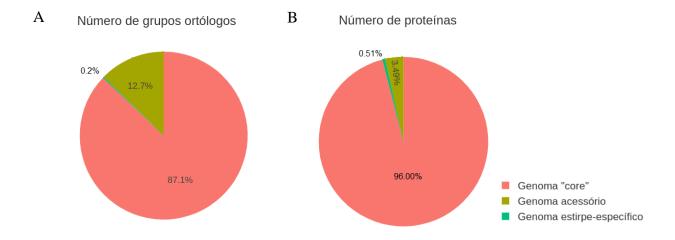
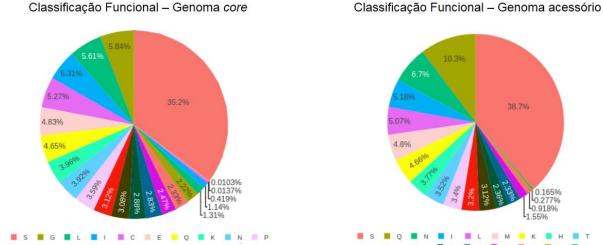
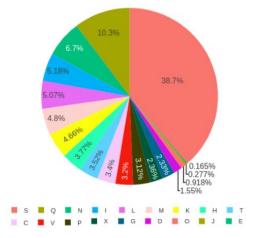


Figura 1. Distribuição gráfica do pan-genoma de 205 estirpes do Complexo *Mycobacterium tuberculosis*. A: Percentual de grupos de proteínas de genes ortólogos. B: Percentual de número de proteínas. O pan-genoma pode ser dividido em genoma *core* (proteínas comuns a 95% das estirpes ou mais), genoma acessório (proteínas presentes em 2 a 195 estirpes) e genoma espécie-específico (proteínas presentes apenas em uma estirpe).

A Figura 2 mostra a classificação funcional dos grupos de proteínas ortólogas do genoma *core* (Figura 2A), genoma acessório (Figura 2B) e do genoma estirpe-específico (Figura 2C). Em todos os grupos destacaram-se as proteínas anotadas com função desconhecida (S). No genoma *core*, a segunda classe funcional mais representativa é que envolve o transporte e metabolismo de carboidratos [G (1702/29.133, 5,84%)], com proteínas como fosfomanomutase, glicogênio fosforilase e gliceraldeído desidrogenase-3-fosfato, que participam ativamente de vias metabólicas envolvendo carboidratos (Figura 3). A alta taxa de conectividade da rede de interação de proteínas envolvidas no metabolismo de carboidratos (Figura 3) evidencia a conservação e importância dessas vias metabólicas para as micobactérias tuberculosas, constituindo um potencial alvo terapêutico.





Classificação Funcional - Genoma estirpe-específico 2 31% 2.21% 1.81% 1.71% 1.61% 65 4% 1% 0.803% 0.803% 0.703% 0.602% 0.502% 0.502 0.19 G

Figura 2. Cluster of Orthologous Groups (COG) baseados na função de proteínas agrupadas de 205 genomas do Complexo Mycobacterium tuberculosis. As categorias de COG são: D: Controle do ciclo celular, divisão celular, partição de cromossomos; M: Biogênese da parede celular / membrana / envelope; N: motilidade celular; O: modificação pós-translacional, turnover de proteína e acompanhantes; T: Mecanismos de transdução de sinal; U: Tráfico, secreção e transporte vesicular intracelular; V: Mecanismos de defesa; Z: citoesqueleto; A: Processamento e modificação de RNA; J: Tradução, estrutura ribossômica e biogênese; K: Transcrição; L: Replicação, recombinação e reparo; C: Produção e conversão de energia; E: Transporte e metabolismo de aminoácidos; F: Transporte e metabolismo de nucleotídeos; G: Transporte e metabolismo de carboidratos ; H: transporte e metabolismo de co-enzima; I: transporte lipídico e metabolismo; P: Transporte e metabolismo de íons inorgânicos; Q: Biossíntese, transporte e catabolismo de metabólitos secundários; S: Função desconhecida e X: componentes do mobiloma.

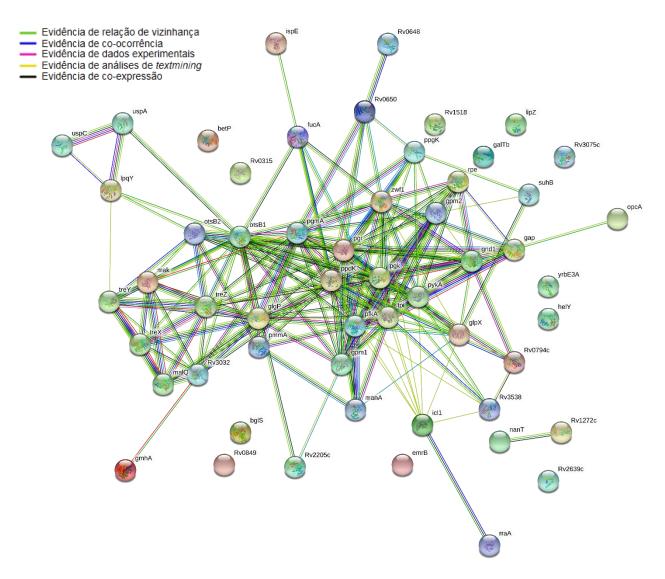


Figura 3. Rede de interações proteína-proteína com representantes da classe funcional de Transporte e Metabolismo de Carboidratos (G) do genoma *core* do Complexo *Mycobacterium tuberculosis*. A rede foi gerada usando o banco de dados STRING, com parâmetros padrão.

Por outro lado, a segunda classe funcional mais representativa do genoma acessório do MTBC é representada por 10,30% (1.559/15.136) dos grupos ortólogos, sendo essas envolvidas no processo de biossíntese, transporte e catabolismo de metabólitos secundários (classe Q). Grande parte dessa classe é composta pelas policetídeos sintases (PKS) (n=469 das 1.559 proteínas dessa classe; 30,08%), que são enzimas importantes do metabolismo de conjugados lipídicos e glicolipídios da parede celular do MTBC, atuando na síntese de ácidos graxos (HERTWECK, 2009). A rede de interação das proteínas do genoma acessório relacionados a esta classe mostra ampla conectividade funcional separada em dois grupos (Figura 4). O primeiro grupo inclui yrbEA e yrbEB, permeases dos transportadores *mammalian cell entry* (*mce*) 1 a 4, e outras proteínas relacionadas aos transportadores Mce2 e Mce4. Foi descrito que

mce2 está ausente em algumas estirpes de M. bovis classificadas como Complexo Clonal Africano 2 (BERG et al., 2011), e encontra-se na forma de pseudogene nas estirpes CDC1551, Haarlem, F11 e KZN1435 de M. tuberculosis (ZHANG; XIE, 2011). Genes do transportador Mce3, por sua vez, estão deletados em todas as micobactérias "animais". Os transportadores Mce1, Mce2 e Mce4 são transportadores relacionados à importação de moléculas hidrofóbicas, possivelmente, ácidos micólicos, sulfolipídeos e colesterol, respectivamente (PANDEY; SASSETTI, 2008; MARJANOVIC; IAVARONE; RILEY, 2011). Não existem informações quanto ao substrato do transportador Mce3. No segundo grupo estão incluídas proteínas policetídeo sintases (PKS, PPS, Mbt), Salicil-AMP ligase (MbtA), citocromo P450 (Cyp) e a p-ácido hidrobenzoico-AMP ligase (FadD22). A maioria desses genes tem sido associados a vias metabólicas para produção de glicolipídios fenol-phthiocerol e phthiocerol, e mycobactina, sendo este último um sideróforo para sequestro de ferro (CONSTANT et al., 2002; FERRERAS et al., 2008; QUADRI, 2014). O cluster gênico de 24 kilobases de M. tuberculosis correspondendo aos genes *mbtA-J*, encontrados nesta rede de interação (Figura 4), contém os componentes necessários para biogênese da mycobactina (QUADRI, 2014). Além disso, estão presentes tioesterase, dienelactona hidrolase e carboximetilenobutenolidase, que catalisam a hidrólise de ácidos graxos modulando o metabolismo lipídico (SWARBRICK et al., 2017). O destaque desta classe no genoma acessório indica a importância da modulação lipídica na variabilidade fenotípica de estirpes do MTBC.

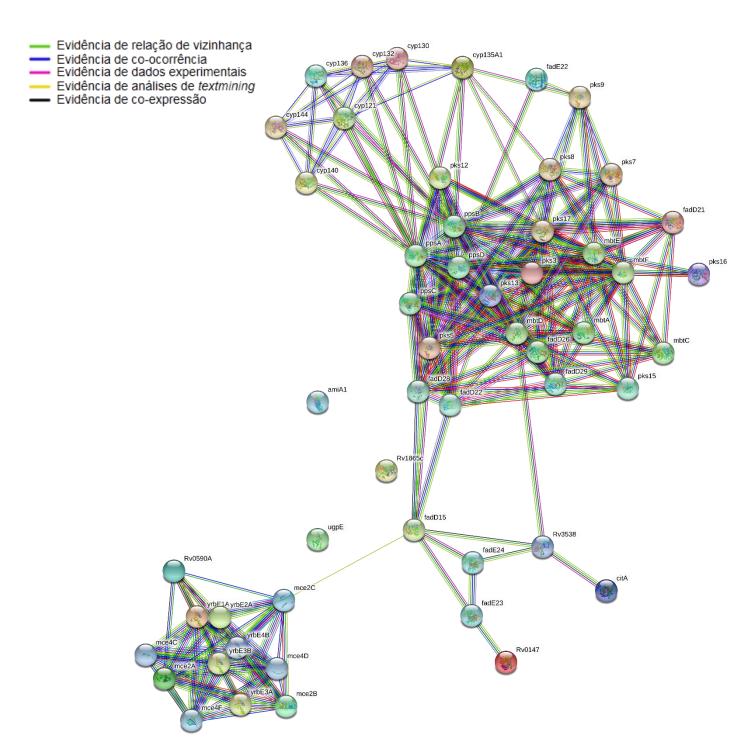


Figura 4. Rede de interações proteína-proteína com representantes da classe funcional de Biossíntese, transporte e catabolismo de metabólitos secundários (Q) do genoma acessório do Complexo *Mycobacterium tuberculosis*. A rede foi gerada usando o banco de dados STRING, com parâmetros padrão.

Enquanto que parte dos genes relacionados ao metabolismo de lipídeos e metabólitos secundários é conservado no genoma *core* (classes Q e I), proteínas de determinadas vias metabólicas dessas classes são variavelmente presente entre as estirpes (i.e. são parte do

genoma acessório). Proteínas envolvidas no transporte e metabolismo de lipídios (Classe I) estão entre as 5 classes funcionais mais representativas no genoma acessório. A parede celular dos representantes do MTBC é constituída por peptidoglicano, ácido micólico, arabinogalactana, ácidos graxos e lipídios livres intercalados por manosídio fosfatidilinositol (PIM), lipomanana (LM) e lipoarabinomanana (LAM). Os lipídios representam 60% da massa seca do envelope celular, impactando no processo de apresentação de antígenos e na resposta de citocinas (ASTARIE-DEQUEKER et al., 2010). A estrutura e composição da parece celular tem influência direta na viabilidade bacteriana (HETT; RUBIN, 2008) e na interação com a célula hospedeira (BRENNAN, 2003). O componente manosilados LAMs (manLAMs) apresenta variações em estirpes da linhagem 2 de M. tuberculosis, e foi demonstrado que estas cresceram mais lentamente quando comparados com as estirpes H37Rv e Erdman da linhagem 4 (TORRELLES et al., 2008), o que também pode apresentar diferenças na patogênese (TORRELLES; SCHLESINGER, 2010). Estudos apontam que os ácidos micólicos atuam na resistência bacteriana durante o estresse oxidativo, além de modular a resposta imune do hospedeiro durante a infecção (VANDER BEKEN et al., 2011). Portevin e colaboradores descreveram uma variabilidade na composição de ácidos micólicos na parede celular entre diferentes linhagens de MTBC (PORTEVIN et al., 2014), sugerindo que tais variações podem provocar diferentes níveis de adaptabilidade a momentos de estresse oxidativo, além de induzir padrões inflamatórios variáveis em diferentes estirpes (PORTEVIN et al., 2011; VANDER BEKEN et al., 2011). Estudos futuros envolvendo a comparação detalhada dessas vias metabólicas conservadas e variáveis entre as espécies e linhagens poderá auxiliar no entendimento dos diferentes fenótipos do MTBC como resultado da modulação lipídica.

No genoma estirpe-específico destacaram-se as proteínas com funções desconhecidas (3.256/4.975, 65,44%). Outra função de destaque foi a relacionada à mecanismos de transdução de sinal [T (374/4.975, 7,51%)], que inclui genes da família de fosfatases (123/374, 32,89%). Neste grupo incluem proteínas tirosina fosfatase PtpA, descritas como importante na patogenicidade de estirpes de *M. tuberculosos in vivo*. PtpA atua na desfosforilação de proteínas do citoplasma de células hospedeiras, provocando a supressão da imunidade inata durante o processo de infecção (COWLEY; BABAKAIFF; AV-GAY, 2002). Proteínas PtpA podem provocar a desfosforilação da proteína hospedeira VPS33B responsável pela fusão fagossomolisossomo, e impedir a acidificação dos fagossomos. Além disso, essas proteínas também podem suprimir a ativação do NF-kB (WANG et al., 2015), responsável por regular a resposta imune inata a partir do controle da expressão de moléculas imunorreguladoras (DEV et al.,

2011). O número limitado de proteínas e grupos ortólogos estirpe-específicos é um reflexo da ausência da transferência horizontal de genes e do comportamento de um grupo clonal.

Um total de 12.970 epítopos foram preditos para sequências de proteínas do genoma *core*, 961 epítopos no genoma acessório e 320 foram encontrados no genoma estirpe-específico, totalizando 14.251 epítopos. Os resultados da distribuição da quantidade de epítopos preditos por proteína mostrou que tanto no genoma *core* (Figura 5A) quanto no genoma acessório (Figura 5B) há uma prevalência de um a dois epítopos na maioria das proteínas. Nas sequências do genoma estirpe-específico a maioria das proteínas apresentam apenas um epítopo (Figura 5C).

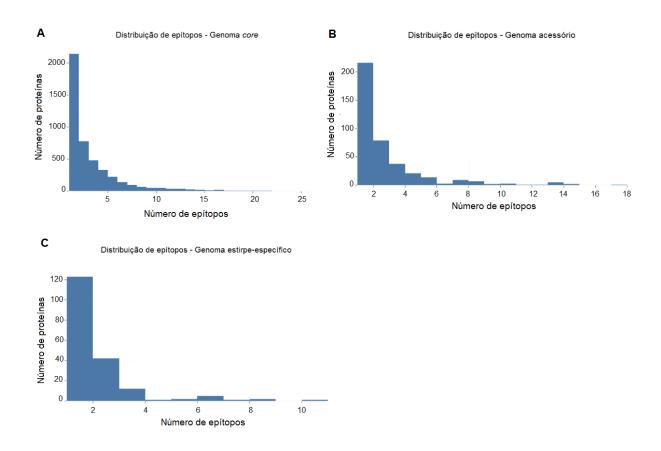


Figura 5. Distribuição do número de epítopos em proteínas do Genoma *core* (A), Genoma acessório (B) e Genoma Espécie-Específico (C) de espécies do Complexo *Mycobacterium tuberculosis* (MTBC). O eixo x refere-se ao número de epítopos encontrados e o eixo y ao número de proteínas.

Análises quanto a variabilidade dos epítopos preditos mostram que, embora haja um nível alto de conservação, há variações antigênicas dessas regiões nas estirpes analisadas do MTBC (Figura 6). No genoma *core*, um total de 93,46% (12.121/12.970) dos epítopos

mostraram-se conservados em todas as sequências analisadas, isto é, com 100% de identidade em todas as proteínas dos 205 genomas. Por outro lado, 12,28% (118/961) dos epítopos localizados em proteínas do genoma acessório apresentaram uma variabilidade de sequência, com mudanças de um a três sítios de aminoácidos, como demonstrado nos exemplos na Figura 7. A variação de sequência é uma das estratégias adotadas por patógenos para evadir a resposta imune do hospedeiro, e é caracterizada por mudanças nas sequência de amino ácidos que podem alterar o reconhecimento de epítopos pelos receptores MHC (complexo principal de histocompatibilidade) (ERNST, 2017).

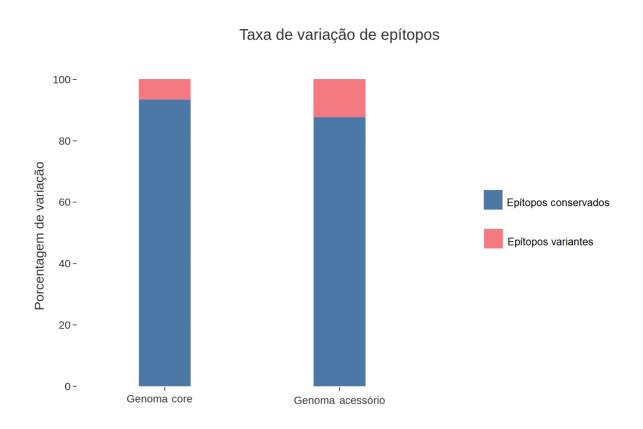


Figura 6. Taxa de variação de sequência de epítopos de células T preditos de proteínas do Genoma *core* e Genoma acessório de espécies do Complexo *Mycobacterium tuberculosis*. Em azul estão os índices de epítopos conservados e em vermelho os de epítopos variantes (i.e., com pelo menos um sítio com variantes de aminoácidos).

Nossos resultados evidenciam, pela primeira vez, um mecanismo de variação antigênica de epítopos de célula T do MTBC baseado no silenciamento/perda de genes. A presença de 961 epítopos no genoma acessório, que constituem 2,24% do número total de epítopos preditos, sugere que os diferentes isolados de MTBC são capazes de evadir o sistema imune por meio de

variação antigênica de fase (*phase antigenic variation*). Em outras palavras, a expressão desses 961 epítopos é variável entre as estirpes do MTBC, uma vez que seus genes foram perdidos por deleções ou truncamentos. Além do mais, foi possível identificar que 256 dos 961 (26,64%) epítopos identificados no genoma acessório já foram validados experimentalmente e são reportados no banco de dados de epítopos (IEDB) (VITA et al., 2019), suportando a veracidade deste mecanismo de evasão do sistema imune. A lista desses epítopos do genoma acessório encontrados no IEDB e em quais genes eles foram encontrados estão disponíveis no Apêndice E.

Como a grande maioria dos epítopos investigados apresentaram-se hiperconservados em sequência, nossos resultados corroboram o encontrado por Comas et al., 2010, que avaliaram 21 estirpes representativas de M. tuberculosis e M. africanum (L5 e L6) para presença de 491 epítopos de células T humanas experimentalmente validados. Naquele estudo, os pesquisadores mostraram que os 491 epítopos de células T são evolutivamente hiperconservados, sendo estas regiões mais conservadas em sequência do que os genes "essenciais" dos genomas de MTBC (COMAS et al., 2010). Na mesma linha, Coscolla et al. (2015) posteriormente encontrou apenas raros epítopos hipervariáveis (n=7) ao avaliar 216 estirpes de M. tuberculosis. Em ambos os estudos, porém, os pesquisadores não avaliaram o silenciamento de epítopos por meio da perda de genes, deixando de associar a arquitetura do genoma acessório com a distribuição de epítopos. Desta forma, por mais que a grande maioria dos epítopos ainda esteja localizada no genoma core e esses epítopos sejam hiperconservados em sequência, negligenciou-se a possibilidade de variação antigênica de fase. Assim concluise que a evasão do sistema imune por meio da variação de epítopos de células T não é tão insignificante como se pensava, e a identificação de epítopos variáveis deverá ser avaliada no desenvolvimento de estratégias vacinais para o controle da TB. É possível que a inclusão desses epítopos variáveis em vacinas aumentem sua eficácia de proteção. Adicionalmente, a identificação de epítopos variáveis, como os encontrados neste estudo, podem fornecer melhorias no entendimento da resposta imune por células T, contribuindo para o entendimento da doença e sua transmissão, principalmente em regiões endêmicas e com altos índices de coinfecção com HIV em humanos. Estudos futuros deverão ser conduzidos para se avaliar em quais proteínas e em quais espécies do MTBC esses epítopos são encontrados.

Epítopos preditos de proteínas presentes no genoma acessório do MTBC mostraram-se relacionados com regiões de genes PE/PPE em 21,96% (211/961) dos resultados encontrados neste estudo, sendo estes, 9,48% (20/211) já verificados experimentalmente com base em dados recuperados do IEDB. Regiões genômicas pertencentes à PE/PPE são repetitivas, e foram

excluídas de estudos realizados para análise de epítopos de células T em representantes do MTBC (COMAS et al., 2010; COSCOLLA et al., 2015) devido às limitações técnicas para o mapeamento de leituras (*reads*) duplicadas em genomas de referência. A família gênica PE/PPE é caracterizada pela presença dos resíduos prolina-glutamato (PE) e prolina-prolina-glutamato (PPE) na região N-terminal dos seus transcritos (BRENNAN; DELOGU, 2002). Apesar da sua função pouco conhecida, a presença exclusiva desses genes em micobactérias patogênicas tem chamado a atenção (GEY VAN PITTIUS et al., 2006). Estudos apontam que proteínas PE/PPE estão associadas com o processo de apoptose e necrose de células hospedeiras a partir da indução de citocinas com a interação das proteínas com o receptor *toll like receptor-2* (TLR2) (BASU et al., 2007), e com a indução de respostas imunes perfil Th1 e Th17 com a ativação de células dendríticas via TLR2 (SU et al., 2018), apresentando um papel importante na sobrevivência de *M. tuberculosis* (RAMAKRISHNAN; FEDERSPIEL; FALKOW, 2000; SASSETTI; RUBIN, 2003). Além disso, PE/PPE estão envolvidos com a atividade do sistema de secreção tipo VII, responsável pela secreção de proteínas como ESAT-6, potente antígeno de células T (GEY VAN PITTIUS et al., 2006; ATES; HOUBEN; BITTER, 2016).

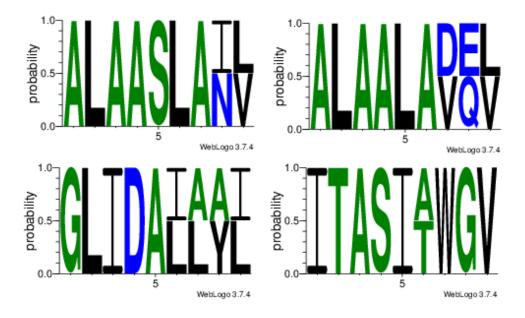


Figura 7. Exemplos de epítopos variáveis preditos neste estudo. Alinhamento de epítopos preditos de proteínas do genoma *core* de espécies do Complexo *Mycobacterium tuberculosis*. A figura foi criada usando o WebLogo (CROOKS et al., 2004).

Realizamos um estudo comparativo entre diferentes bactérias para analisar a presença dos FVs em espécies do MTBC. Apesar da grande similaridade compartilhada entre o grupo, foi possível observar variações na quantidade de genes relacionados aos fatores de virulência (Figura 8), principalmente em espécies de *M. pinnipedii* e algumas estirpes de *M. africanum* (MAL010131, MAL010137, MAL020107). Um total de 211 genes relacionados a FVs foram detectados nas estirpes de MTBC, e estes foram divididos em 88 grupos que estão separados em 20 classes de acordo com as funções exercidas, sendo elas: catabolismo de colesterol, metabolismo de aminoácidos, adaptação ao estresse, componentes da superfície celular, metabolismo de lipídios, sistema de secreção, proteínas secretadas, regulação gênica, operons Mce, respiração anaeróbica, absorção de cobre, fatores anti-apoptose, proteases, absorção de magnésio, metabolismo do fagossomo, enzimas, aderência, virulência nutricional e resistência ao estresse ácido. Vinte e oito dos 88 grupos de FVs são caracterizados como sendo do mesmo *operon* ou grupo gênico, e assim possuem um número variável de genes.

Com a exceção das estirpes de M. pinnipedii e as estirpes MAL010131, MAL010137, MAL020107 de *M. africanum* é possível observar três clusters de FVs altamente conservados entre os genomas do MTBC (clusters A, B e C na figura 8). Esses clusters são compostos pelos mais variados FVs, identificados em ordem de aparecimento no heatmap da figura 7, do número 63 ao 22, no apêndice F. É possível notar a deleção do operon Mce3, característica de MTBC animais e M. africanum L6, chamada de RD17 (FV 42). Destaca-se que algumas estirpes de M. tuberculosis (CDC1551, F11, KZN 1435, Haarlem, KZN 4207) também foram encontradas contendo variações no número de genes associados a este operon. De maneira geral, os quatro operons Mce (1-4) apresentam certo grau de variabilidade, onde foi possível observar: (i) 10 FVs relacionados ao operon Mce1 na estirpe Beijing-like/50148 de M. tuberculosis, enquanto as outras estirpes da mesma espécie apresentaram seis FVs relacionados a este mesmo operon; (ii) três FVs relacionados ao operon Mce2 nas espécies M. caprae, M. orygis e M. pinnipedii, sendo que para as outras espécies do MTBC foram encontrados seis FVs relacionado ao mesmo operon; e (iii) ausência de FVs relacionados ao Mce4 em 17 estirpes da linhagem 4 do M. tuberculosis (HN-024, Beijing-like/1104, BT2, KZN 4207, CSV4519, CSV4644, CSV10399, CSV11678, LE486, LE492, LN180, ME1473, TBDM425, BV5000, TBV5362, TBV5365, SLM036).

Em seguida, é possível notar um cluster hipervariável (cluster D, figura 8) (marcados em ordem, do 53 ao 21 no apêndice F) composto por 6 FVs e mais dois outros FVs também hipervariáveis (FVs 30 e 35). Esses são: proteína RegX3 de sistema regulatório de doiscomponentes, mycobactina, transportador ABC reciclador de trehalose, ESX-1 (T7SS), ESX-2 (T7SS), operon MymA, regulador dependente de ferro IdeR e fosfolipase C. Esses FVs são relacionados às respostas contra perturbações ambientais [variações na disponibilidade de fosfato (GLOVER et al., 2007), acificação (MymA; SINGH et al., 2003) e disponibilidade de

ferro (mycobactina, IdeR)], citotoxicidade (ESX-1, ESX-2 e fosfolipase C) e remodelagem do envelope celular (transportador ABC reciclador de trehalose e MymA; KALSCHEUER et al., 2010).

Por último, observa-se um grande cluster de variabilidade intermediária (cluster E, figura 8) composto por 26 conjuntos de FVs (marcados em ordem, do 79 ao 34 no apêndice F), sendo que destes, os quatro grupos mais representativos foram seis conjuntos de FVs (23,08%, 6/26) relacionados a componentes da superfície celular, cinco (19,23%, 5/26) ao sistema de secreção do tipo VII, três (11,54%, 3/26) ao metabolismo de aminoácidos e três ao catabolismo de colesterol. Assim, conclui-se que importantes variações nos FVs observadas intra- e entreespécies do MTBC estão diretamente relacionadas ao sistema de secreção do tipo VII (ESX), resistência às perturbações ambientais e modulação do envoltório bacteriano via metabolismo de lipídeos. Essas categorias sugerem que essas funções são de extrema importância para o estabelecimento dos diferentes fenótipos de tropismo por hospedeiros e virulência do MTBC. Resultados derivados de um estudo que analisou a distribuição e variabilidade de um conjunto de fatores de virulência no pan-genoma estruturado a partir de 1.595 estirpes de M. tuberculosis mostrou que 93,00% (414/445) dos fatores de virulência analiados foram encontrados no genoma core e 7,00% (31/445) no genoma acessório. Também foram encontrados FVs relacionados com o sistema de secreção tipo VII (ESX), esxG e esxH, que estavam presentes apenas em algumas das estipes analisadas (KAVVAS et al., 2018).

As estirpes de *M. pinnipedii* e as estirpes MAL010131, MAL010137, MAL020107 de *M. africanum* apresentaram diferenças significativas em todos os *clusters* de FVs quando comparadas aos outros membros do MTBC, mesmo nos clusters mais conservados entre os membros. Isso se dá pela variabilidade no número de genes de cada *operon* classificado como FV, possivelmente devido ao truncamento de genes. De maneira geral, é possível notar que estirpes de *M. pinnipedii* possui um número maior de genes em cada conjunto de FV, com exceção de 18 conjuntos de FVs (marcados em ordem, do 37 ao 45 na figura 7 apêndice F) que correspondem, principalmente, ao sistema de secreção do tipo VII (ESX) (44,44%, 4/9).

Embora *M. pinnipedii* seja considerado um patógeno generalista (i.e. infecta várias espécies hospedeiras), ele está mais relacionado com infecção em mamíferos marinhos, como leões-marinhos e focas. Estes animais apresentam características fisiológicas e metabólicas que permitem uma adaptação a um ambiente que promove situações de colapso alveolar, fluxo respiratório extremo, hiperóxia transitória, hipóxia, hipo e hipertensão intravasculares (OLSEN; HALE; ELSNER, 1969; KEREM; KYLSTRA; SALTZMAN, 1975; FAHLMAN; MADIGAN, 2016). Sugere-se que a variabilidade no número de genes relacionados a fatores

de virulência de estresse ambiental em *M. pinnipedii* ocorra como resultado do processo de adaptação desses animais ao ambiente.

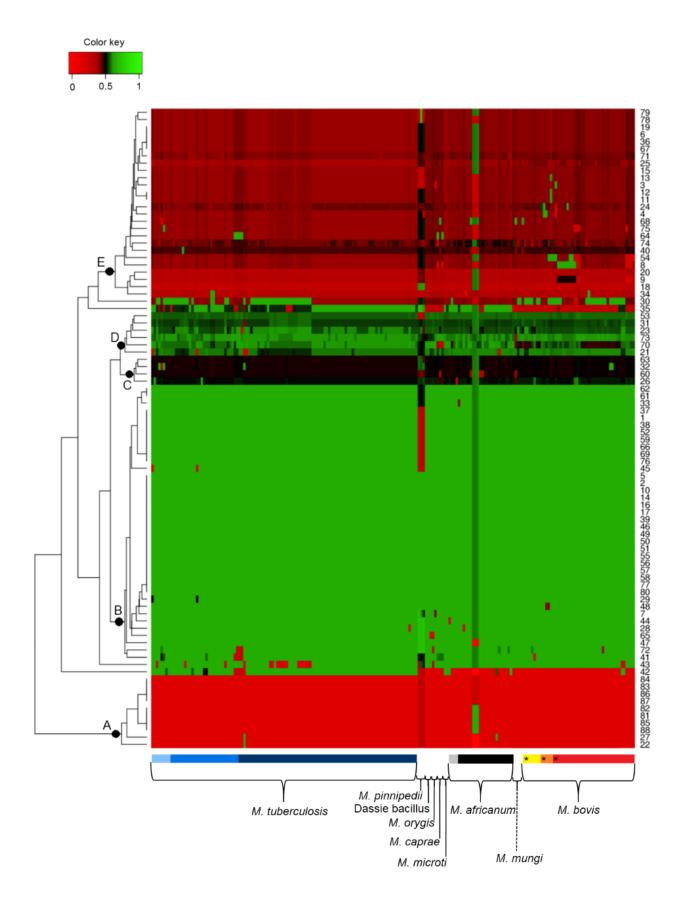


Figura 8. Heatmap apresentando fatores de virulência de espécies do Complexo Mycobacterium tuberculosis. A escala gráfica de cor varia do vermelho ao verde, nos quais quanto mais próximo do verde maior a quantidade de genes relacionados àquele fator de virulência nos determinados genomas. As barras coloridas abaixo indicam as linhagens de cada espécie bacteriana. Azul claro: Linhagem 1 (L1) M. tuberculosis; azul: linhagem 2 (L2) M.tuberculosis; azul escuro: linhagem 4 (L4) M. tuberculosis; cinza: linhagem 5 (L5) Mycobacterium africanum; preto: linhagem 6 (L6) M. africanum; amarelo: linhagem 1 (L1) M. bovis; laranja: linhagem 2 (Lb2) M. bovis; vermelho: linhagem 3 (Lb3) M. bovis. * Bactérias isoladas em animais selvagens. A lista de fatores de virulência representado por cada número está disponível no Apêndice F.

Resultados obtidos nesse trabalho mostram que as estirpes de *M. pinnipedii* possuem variações no número de genes relacionados ao *regulon* DevR/DosR em relação a outras espécies do MTBC (FV 63 no apêndice F). DevR/DosR é um *regulon* composto por uma variedade complexa de genes que são capazes de responder a diferentes sinais de estresses ambientais. A regulação positiva do *regulon* ocorre em resposta à hipóxia e baixos níveis de óxido nítrico (NO) (SHERMAN et al., 2001; SCHNAPPINGER et al., 2003; VOSKUIL et al., 2003). Um subconjunto do *regulon* DevR/DosR inclui genes *acr*-coregulados (ACGs). Todos os membros da família ACGs ativam e coordenam positivamente os macrófagos, em condições de crescimento com baixos níveis de oxigênio (PARK et al., 2003; KENDALL et al., 2004). Apesar de mais estudos serem necessários para esclarecer as funções biológicas exercidas pelo DevR/DosR, sabe-se que os genes que compõem essas vias metabólicas são regulados em resposta a diferentes condições externas associadas à infecção ao hospedeiro (KUMAR et al., 2008; SOUSA; GONZALEZ; GILLES-GONZALEZ, 2017) e participam do metabolismo de dormência de *M. tuberculosis*.

Mycobacterium pinnipedii e 4 estirpes da linhagem 4 (L4) de *M. tuberculosis* (CDC1551, F11, KZN1435, Haarlem) apresentaram mais genes relacionados a síntese e hidrólise (p)ppGpp dos que as outras estirpes do MTBC (FV 64 no apêndice F). A síntese dessa molécula é mediada pelos substratos guanosina difosfato (GDP) e adenosina trifosfato (ATP), regulada pelo gene *rel* (DAVID AVARBOCK; ANDREW AVARBOCK; RUBIN, 2000). O gene *rel* demonstrou ser de suma importância na sobrevivência bacteriana em situações de baixos recursos energéticos e hipóxia (WAYNE; HAYES, 1996). Estudos realizados em *M. smegmatis*, uma micobactéria modelo não-tuberculosa, mostrou que a superexpressão de *rel* afeta a fase de dormência na bactéria, sugerindo envolvimento na persistência da bactéria dentro dos granulomas do hospedeiro (OJHA; MUKHERJEE; CHATTERJI, 2000). A versatilidade quanto a adaptações que garantem a persistência do patógeno em estado não-replicativo em

condições de microaerofilia contribui para explicar a latência da TB a longo prazo (NYKA, 1974).

Os resultados obtidos nesse estudo chamaram a atenção para três estirpes (MAL010131, MAL010137, MAL020107) de *M. africanum* por apresentarem resultados que se diferem de todas as outras estirpes, de forma geral. Entre 2006 e 2010 foi realizado um estudo descritivo com dados de ocorrência de TB pulmonar em centros de referências de Mali, localizado no oeste Africano (TRAORE et al., 2012). Desses isolados, 92 estavam viáveis e foram submetidos ao sequenciamento e análise, construindo a "Coleção de Mali" de isolados de *M. africanum* (WINGLEE et al., 2016). O sequenciamento resultou em genomas com montagens e anotações gênicas com boa qualidade para todos os isolados, considerando a curadoria manual realizada pelo RefSeq (NCBI). Apesar dos resultados encontradas que estão ilustradas no *heatmap*, a princípio, não há características nesses genomas que justifiquem tais diferenças.

5.4.CONSIDERAÇÕES FINAIS

Este estudo traz evidências, pela primeira vez, de um mecanismo de evasão do sistema imune exercido pelo MTBC por meio da variação de fase da expressão de epítopos lineares de células T. Estudos futuros indicarão se essa perda de epítopos é associada a diferentes espécies do MTBC e consequentemente a mudança de tropismos por hospedeiro. Foi possível também observar variações importantes no repertório de FVs do MTBC relacionados à resistência aos estressores ambientais, sistema de secreção do tipo VII e modulação do envoltório lipídico, que podem estar diretamente associados aos diferentes fenótipos de tropismo por hospedeiro e virulência observados nas espécies e linhagens do MTBC. Essas variações confirmam que o processo de evolução clonal do MTBC baseado em SNPs, indels, deleções e duplicações é capaz de modular o conteúdo gênico das micobactérias tuberculosas, com possíveis efeitos fenotípicos que poderão ser futuramente explorados.

A análise do pan-genoma utilizando um *pipeline* robusto de anotação genômica mostra que o proteoma do MTBC é altamente conservado, mas que pelo menos 4% das proteínas são variavelmente presentes entre as estirpes e espécies. É importante ressaltar que neste estudo, os pseudogenes identificados pelo PGAP não foram incluídos, e esses constituem uma fonte adicional de variabilidade genética do MTBC (CAMARGO, 2018). Mesmo assim, os genomas acessório e estirpe-específico indicam que certas vias do metabolismo de lipídeos e metabólitos secundários são primordias na modulação dos fenótipos do MTBC, mostrando sua influência na constituição do envoltório bacteriano. Estudos mais aprofundados, possivelmente envolvendo aprendizado de máquina, poderão identificar quais vias metabólicas estão sob modulação nas diferentes estirpes e linhagens, buscando por padrões conservados e diferenças.

Curiosamente, estirpes de *M. pinnipedii* e algumas estirpes da linhagem 6 de *M. africanum* mostraram um padrão distinto de FVs de outras bactérias do MTBC. O *M. pinnipedii* é a única micobactéria tuberculosa frequentemente encontrada em animais marinhos, o que poderia explicar a necessidade de diferentes perfis gênicos por meio da co-evolução patógenohospedeiro. Os FVs variavelmente presentes nesta espécie indicam que estudos sobre os mecanismos de dormência, respiração celular e persistência intracelular serão de grande valia para compreensão da biologia do *M. pinnipedii*. Por outro lado, a linhagem 6 de *M. africanum* é geralmente negligenciada e muito pouco estudada. Dada sua proximidade filogenética com as micobactérias "animais", reservatórios animais ou ambientais para essa *M. africanum* L6 já foram sugeridos na literatura a partir de isolamentos dessa espécie em animais (THOREL,

1980a, 1980b; CADMUS et al., 2010). No desconhecimento da fonte de infecção que deu origem a estes isolados, não é possível concluir sobre sua epidemiologia.

Referências

ALI, A.; HASAN, Z.; MCNERNEY, R.; MALLARD, K.; HILL-CAWTHORNE, G.; COLL, F.; NAIR, M.; PAIN, A.; CLARK, T. G.; HASAN, R. Whole genome sequencing based characterization of extensively drug-resistant mycobacterium tuberculosis isolates from pakistan. **PLoS ONE**, v. 10, n. 2, p. e0117771, 2015.

ASTARIE-DEQUEKER, C.; NIGOU, J.; PASSEMAR, C.; GUILHOT, C. The role of mycobacterial lipids in host pathogenesis, **Drug Discovery Today: Disease Mechanisms**, n. 7, n. 1, p. e33-e41. 2010.

ATES, L. S.; HOUBEN, E. N. G.; BITTER, W. Type VII Secretion: A Highly Versatile Secretion System. **Microbiology spectrum**, v. 4, n. 1, p. 357–384, 2016.

BASU, S.; PATHAK, S. K.; BANERJEE, A.; PATHAK, S.; BHATTACHARYYA, A.; YANG, Z.; TALARICO, S.; KUNDU, M.; BASU, J. Execution of macrophage apoptosis by PE_PGRS33 of Mycobacterium tuberculosis is mediated by toll-like receptor 2-dependent release of tumor necrosis factor-α. **Journal of Biological Chemistry**, v. 282, n. 2, p. 1039–1050, 2007.

BAUMLER, A.; FANG, F. C. Host Specificity of Bacterial Pathogens. **Cold Spring Harbor Perspectives in Medicine**, v. 3, n. 12, p. a010041–a010041, 2013.

BENTLEY, S. D.; COMAS, I.; BRYANT, J. M.; WALKER, D.; SMITH, N. H.; HARRIS, S. R.; THURSTON, S.; GAGNEUX, S.; WOOD, J.; ANTONIO, M.; QUAIL, M. A.; GEHRE, F.; ADEGBOLA, R. A.; PARKHILL, J.; DE JONG, B. C. The Genome of Mycobacterium Africanum West African 2 Reveals a Lineage-Specific Locus and Genome Erosion Common to the M. tuberculosis Complex. **PLoS Neglected Tropical Diseases**, v. 6, n. 2, p. e1552, 2012.

BERG, S.; GARCIA-PELAYO, M. C.; MÜLLER, B.; HAILU, E.; ASIIMWE, B.; KREMER, K.; DALE, J.; BONIOTTI, M. B.; RODRIGUEZ, S.; HILTY, M.; RIGOUTS, L.; FIRDESSA, R.; MACHADO, A.; MUCAVELE, C.; NGANDOLO, B. N. R.; BRUCHFELD, J.; BOSCHIROLI, L.; MÜLLER, A.; SAHRAOUI, N.; PACCIARINI, M.; CADMUS, S.; JOLOBA, M.; VAN SOOLINGEN, D.; MICHEL, A. L.; DJØNNE, B.; ARANAZ, A.; ZINSSTAG, J.; VAN HELDEN, P.; PORTAELS, F.; KAZWALA, R.; KÄLLENIUS, G.; HEWINSON, R. G.; ASEFFA, A.; GORDON, S. V.; SMITH, N. H. African 2, a clonal complex of Mycobacterium bovis epidemiologically important in East Africa. **Journal of Bacteriology**, v. 193, n. 3, p. 670–678, 2011.

BLACK, P. A.; DE VOS, M.; LOUW, G. E.; VAN DER MERWE, R. G.; DIPPENAAR, A.; STREICHER, E. M.; ABDALLAH, A. M.; SAMPSON, S. L.; VICTOR, T. C.; DOLBY, T.; SIMPSON, J. A.; VAN HELDEN, P. D.; WARREN, R. M.; PAIN, A. Whole genome

sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates. **BMC Genomics**, v. 16, n. 1, p. 857, 2015.

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

BRENNAN, M. J.; DELOGU, G. The PE multigene family: A "molecular mantra" for mycobacteria. **Trends in Microbiology**, v. 10 n. 5, p. 246-249, 2002.

BRENNAN, P. J. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. **Tuberculosis**, v. 83, n. 1–3, p. 91–97, 2003.

BROSCH, R.; GORDON, S. V.; PYM, A.; EIGLMEIER, K.; GARNIER, T.; COLE, S. T. Comparative genomics of the mycobacteria. **International Journal of Medical Microbiology**, v. 290, n. 2, p. 143–152, 2000.

BROSCH, R.; PHILIPP, W. J.; STAVROPOULOS, E.; COLSTON, M. J.; COLE, S. T.; GORDON, S. V. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. **Infection and immunity**, v. 67, n. 11, p. 5768–74, 1999.

BROWN, A. C.; BRYANT, J. M.; EINER-JENSEN, K.; HOLDSTOCK, J.; HOUNIET, D. T.; CHAN, J. Z. M.; DEPLEDGE, D. P.; NIKOLAYEVSKYY, V.; BRODA, A.; STONE, M. J.; CHRISTIANSEN, M. T.; WILLIAMS, R.; MCANDREW, M. B.; TUTILL, H.; BROWN, J.; MELZER, M.; ROSMARIN, C.; MCHUGH, T. D.; SHORTEN, R. J.; DROBNIEWSKI, F.; SPEIGHT, G.; BREUER, J. Rapid whole-genome sequencing of mycobacterium tuberculosis isolates directly from clinical samples. **Journal of Clinical Microbiology**, v. 53, n. 7, p. 2230–2237, 2015.

CAMARGO, N. C. S. **Investigação da taxa de pseudogenização em genomas do complexo Mycobacterium tuberculosis.** 2018. 39 f. Trabalho de Conclusão de Curso (Graduação em Bacharelado em Biotecnologia) - Universidade Federal de São Paulo.

CHEN, L.; YANG, J.; YU, J.; YAO, Z.; SUN, L.; SHEN, Y.; JIN, Q. VFDB: a reference database for bacterial virulence factors. **Nucleic Acids Research**, v. 33, n. Database issue, p. D325–D328, 2004.

COMAS, I.; CHAKRAVARTTI, J.; SMALL, P. M.; GALAGAN, J.; NIEMANN, S.; KREMER, K.; ERNST, J. D.; GAGNEUX, S. Human T cell epitopes of Mycobacterium

tuberculosis are evolutionarily hyperconserved. **Nature Genetics**, v. 42, n. 6, p. 498–503, 2010.

COMAS, I.; COSCOLLA, M.; LUO, T.; BORRELL, S.; HOLT, K. E.; KATO-MAEDA, M.; PARKHILL, J.; MALLA, B.; BERG, S.; THWAITES, G.; YEBOAH-MANU, D.; BOTHAMLEY, G.; MEI, J.; WEI, L.; BENTLEY, S.; HARRIS, S. R.; NIEMANN, S.; DIEL, R.; ASEFFA, A.; GAO, Q.; YOUNG, D.; GAGNEUX, S. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. **Nature genetics**, v. 45, n. 10, p. 1176–82, 2013.

CONSTANT, P.; PEREZ, E.; MALAGA, W.; LANÉELLE, M.-A.; SAUREL, O.; DAFFÉ, M.; GUILHOT, C. Role of the pks15/1 Gene in the Biosynthesis of Phenolglycolipids in the Mycobacterium tuberculosis Complex. **Journal of Biological Chemistry**, v. 277, n. 41, p. 38148–38158, 2002.

COSCOLLA, M.; COPIN, R.; SUTHERLAND, J.; GEHRE, F.; DE JONG, B.; OWOLABI, O.; MBAYO, G.; GIARDINA, F.; ERNST, J. D.; GAGNEUX, S. M. tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens. **Cell host & microbe**, v. 18, n. 5, p. 538–48, 2015.

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

COWLEY, S. C.; BABAKAIFF, R.; AV-GAY, Y. Expression and localization of the Mycobacterium tuberculosis protein tyrosine phosphatase PtpA. **Research in Microbiology**, v. 153, n. 4, p. 233–241, 2002.

CROOKS, G. E.; HON, G.; CHANDONIA, J. M.; BRENNER, S. E. WebLogo: A sequence logo generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, 2004.

DAUM, L. T.; RODRIGUEZ, J. D.; WORTHY, S. A.; ISMAIL, N. A.; OMAR, S. V.; DREYER, A. W.; FOURIE, P. B.; HOOSEN, A. A.; CHAMBERS, J. P.; FISCHER, G. W. Next-generation ion torrent sequencing of drug resistance mutations in Mycobacterium tuberculosis strains. **Journal of Clinical Microbiology**, v. 50, n. 12, p. 3831–3837, 2012.

DAVID AVARBOCK; ANDREW AVARBOCK, A.; RUBIN, H. Differential Regulation of Opposing RelMtb Activities by the Aminoacylation State of a tRNA·Ribosome·mRNA·RelMtb Complex. **Biochemistry**, v. 39, n. 38, p. 11640–11648, 2000.

DE LA FUENTE, J.; DIEZ-DELGADO, I.; CONTRERAS, M.; VICENTE, J.; CABEZAS-CRUZ, A.; TOBES, R.; MANRIQUE, M.; LOPEZ, V.; ROMERO, B.; BEZOS, J.; DOMINGUEZ, L.; SEVILLA, I. A.; GARRIDO, J. M.; JUSTE, R.; MADICO, G.; JONES-LOPEZ, E.; GORTAZAR, C. Comparative Genomics of Field Isolates of Mycobacterium bovis and M. caprae Provides Evidence for Possible Correlates with Bacterial Viability and Virulence. **PLOS Neglected Tropical Diseases**, v. 9, n. 11, p. e0004232, 2015.

DEJESUS, M. A.; SACCHETTINI, J. C.; IOERGER, T. R. Reannotation of translational start sites in the genome of Mycobacterium tuberculosis. **Tuberculosis**, v. 93, n. 1, p. 18–25, 2013.

DEV, A.; IYER, S.; RAZANI, B.; CHENG, G. NF-κB and innate immunity. **Current Topics** in **Microbiology and Immunology**, v. 349, p. 115–143, 2011.

DIPPENAAR, A.; PARSONS, S. D. C.; SAMPSON, S. L.; VAN DER MERWE, R. G.; DREWE, J. A.; ABDALLAH, A. M.; SIAME, K. K.; GEY VAN PITTIUS, N. C.; VAN HELDEN, P. D.; PAIN, A.; WARREN, R. M. Whole genome sequence analysis of Mycobacterium suricattae. v. 95, n. 6, 2015.

ELDHOLM, V.; MONTESERIN, J.; RIEUX, A.; LOPEZ, B.; SOBKOWIAK, B.; RITACCO, V.; BALLOUX, F. Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. **Nature Communications**, v. 6, n. 1, p. 7119, 2015.

EMMS, D. M.; KELLY, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. **Genome Biology**, v. 16, n. 1, p. 157, 2015.

ERNST, J. D. The immunological life cycle of tuberculosis. **Nature Reviews Immunology**, v. 12, n. 8, p. 581, 2012.

ERNST, J. D. Antigenic variation and immune escape in the MTBC. In: **Advances in Experimental Medicine and Biology**. [s.l.] NIH Public Access, 2017. 1019p. 171–190. FAHLMAN, A.; MADIGAN, J. Respiratory function in voluntary participating patagonia sea lions (Otaria flavescens) in sternal recumbency. **Frontiers in Physiology**, v. 7, n. NOV, p. 528, 2016.

FAKSRI, K.; XIA, E.; ONG, R. T. H.; TAN, J. H.; NONGHANPHITHAK, D.; MAKHAO, N.; THAMNONGDEE, N.; THANORMCHAT, A.; PHURATTANAKORNKUL, A.; RATTANARANGSEE, S.; RATANAJARAYA, C.; SURIYAPHOL, P.; PRAMMANANAN, T.; TEO, Y. Y.; CHAIPRASERT, A. Comparative whole-genome sequence analysis of

Mycobacterium tuberculosis isolated from tuberculous meningitis and pulmonary tuberculosis patients. **Scientific Reports**, v. 8, n. 1, p. 4910, 2018.

FARHAT, M. R.; SHAPIRO, B. J.; KIESER, K. J.; SULTANA, R.; JACOBSON, K. R.; VICTOR, T. C.; WARREN, R. M.; STREICHER, E. M.; CALVER, A.; SLOUTSKY, A.; KAUR, D.; POSEY, J. E.; PLIKAYTIS, B.; OGGIONI, M. R.; GARDY, J. L.; JOHNSTON, J. C.; RODRIGUES, M.; TANG, P. K. C.; KATO-MAEDA, M.; BOROWSKY, M. L.; MUDDUKRISHNA, B.; KREISWIRTH, B. N.; KUREPINA, N.; GALAGAN, J.; GAGNEUX, S.; BIRREN, B.; RUBIN, E. J.; LANDER, E. S.; SABETI, P. C.; MURRAY, M. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. **Nature Genetics**, v. 45, n. 10, p. 1183–1189, 2013.

FERRERAS, J. A.; STIRRETT, K. L.; LU, X.; RYU, J. S.; SOLL, C. E.; TAN, D. S.; QUADRI, L. E. E. N. Mycobacterial Phenolic Glycolipid Virulence Factor Biosynthesis: Mechanism and Small-Molecule Inhibition of Polyketide Chain Initiation. **Chemistry and Biology**, v. 15, n. 1, p. 51–61, 2008.

FLEISCHMANN, R. D.; ALLAND, D.; EISEN, J. A.; CARPENTER, L.; WHITE, O.; PETERSON, J.; DEBOY, R.; DODSON, R.; GWINN, M.; HAFT, D.; HICKEY, E.; KOLONAY, J. F.; NELSON, W. C.; UMAYAM, L. A.; ERMOLAEVA, M.; SALZBERG, S. L.; DELCHER, A.; UTTERBACK, T.; WEIDMAN, J.; KHOURI, H.; GILL, J.; MIKULA, A.; BISHAI, W.; JACOBS, W. R.; VENTER, J. C.; FRASER, C. M. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. **Journal of bacteriology**, v. 184, n. 19, p. 5479–90, 2002.

GAGNEUX, S.; DERIEMER, K.; VAN, T.; KATO-MAEDA, M.; DE JONG, B. C.; NARAYANAN, S.; NICOL, M.; NIEMANN, S.; KREMER, K.; GUTIERREZ, M. C.; HILTY, M.; HOPEWELL, P. C.; SMALL, P. M. Variable host—pathogen compatibility in Mycobacterium tuberculosis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 8, p. 2869-2873, 2006.

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

GEY VAN PITTIUS, N. C.; SAMPSON, S. L.; LEE, H.; KIM, Y.; VAN HELDEN, P. D.; WARREN, R. M. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx.) gene cluster regions. **BMC Evolutionary Biology**, v. 6, n. 1, p. 95, 2006.

GINESTET, C. ggplot2: Elegant Graphics for Data Analysis. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, v. 174, n. 1, p. 245–246, 2011.

GLOVER, R. T.; KRIAKOV, J.; GARFORTH, S. J.; BAUGHN, A. D.; JACOBS, W. R. The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in Mycobacterium smegmatis. **Journal of Bacteriology**, v. 189, n. 15, p. 5495–5503, 2007.

HERTWECK, C. The biosynthetic logic of polyketide diversity. **Angewandte Chemie International Edition**, v. 48, n. 26, p. 4688-4716, 2009.

HETT, E. C.; RUBIN, E. J. Bacterial growth and cell division: a mycobacterial perspective. **Microbiology and molecular biology reviews**, v. 72, n. 1, p. 126–56, 2008.

HUERTA-CEPAS, J.; SZKLARCZYK, D.; FORSLUND, K.; COOK, H.; HELLER, D.; WALTER, M. C.; RATTEI, T.; MENDE, D. R.; SUNAGAWA, S.; KUHN, M.; JENSEN, L. J.; VON MERING, C.; BORK, P. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic acids research**, v. 44, n. D1, p. D286-93, 2016.

ISAZA, J. P.; DUQUE, C.; GOMEZ, V.; ROBLEDO, J.; BARRERA, L. F.; ALZATE, J. F. Whole genome shotgun sequencing of one Colombian clinical isolate of Mycobacterium tuberculosis reveals DosR regulon gene deletions. **FEMS Microbiology Letters**, v. 330, n. 2, p. 113–120, 2012.

JIA, X.; YANG, L.; DONG, M.; CHEN, S.; LV, L.; CAO, D.; FU, J.; YANG, T.; ZHANG, J.; ZHANG, X.; SHANG, Y.; WANG, G.; SHENG, Y.; HUANG, H.; CHEN, F. The bioinformatics analysis of comparative genomics of Mycobacterium tuberculosis complex (MTBC) provides insight into dissimilarities between intraspecific groups differing in host association, virulence, and epitope diversity. **Frontiers in Cellular and Infection Microbiology**, v. 7, p. 88, 2017.

JURTZ, V.; PAUL, S.; ANDREATTA, M.; MARCATILI, P.; PETERS, B.; NIELSEN, M. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. **The Journal of Immunology**, v. 199, n. 9, p. 3360–3368, 2017.

KALSCHEUER, R.; WEINRICK, B.; VEERARAGHAVAN, U.; BESRA, G. S.; JACOBS JR, W. R. Trehalose-Recycling ABC Transporter LpqY-SugA-SugB-SugC Is Essential for Virulence of Mycobacterium Tuberculosis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 50, p. 21761–21766, 2010.

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

KAVVAS, E. S.; CATOIU, E.; MIH, N.; YURKOVICH, J. T.; SEIF, Y.; DILLON, N.; HECKMANN, D.; ANAND, A.; YANG, L.; NIZET, V.; MONK, J. M.; PALSSON, B. O. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. **Nature Communications**, v. 9, n. 1, p. 4306, 2018.

KENDALL, S. L.; MOVAHEDZADEH, F.; RISON, S. C. G.; WERNISCH, L.; PARISH, T.; DUNCAN, K.; BETTS, J. C.; STOKER, N. G. The Mycobacterium tuberculosis dosRS two-component system is induced by multiple stresses. **Tuberculosis**, v. 84, n. 3–4, p. 247–255, 2004.

KEREM, D. H.; KYLSTRA, J. A.; SALTZMAN, H. A. Respiratory flow rates in the sea lion. **Undersea Biomedical Research**, v. 2, n. 1, p. 20–27, 1975.

KUMAR, A.; DESHANE, J. S.; CROSSMAN, D. K.; BOLISETTY, S.; YAN, B.-S.; KRAMNIK, I.; AGARWAL, A.; STEYN, A. J. C. Heme Oxygenase-1-derived Carbon Monoxide Induces the Mycobacterium tuberculosis Dormancy Regulon. **The Journal of biological Chemistry**, v. 283, p. 18032–18039, 2008.

LIU, B.; ZHENG, D.; JIN, Q.; CHEN, L.; YANG, J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. **Nucleic Acids Research**, v. 47, n. D1, p. D687–D692, 2019.

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

LUO, T.; YANG, C.; PENG, Y.; LU, L.; SUN, G.; WU, J.; JIN, X.; HONG, J.; LI, F.; MEI, J.; DERIEMER, K.; GAO, Q. Whole-genome sequencing to detect recent transmission of Mycobacterium tuberculosis in settings with a high burden of tuberculosis. **Tuberculosis**, v. 94, n. 4, p. 434–440, 2014.

MARJANOVIC, O.; IAVARONE, A. T.; RILEY, L. W. Sulfolipid accumulation in Mycobacterium tuberculosis disrupted in the mce2 operon. **Journal of Microbiology**, v. 49, n. 3, p. 441–447, 2011.

NAMOUCHI, A.; DIDELOT, X.; SCHOCK, U.; GICQUEL, B.; ROCHA, E. P. C. After the bottleneck: Genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. **Genome Research**, v. 22, n. 4, p. 721–734, 2012.

NYKA, W. Studies on the effect of starvation on mycobacteria. **Infection and Immunity**, v. 9, n. 5, p. 843–850, 1974.

OJHA, A. K.; MUKHERJEE, T. K.; CHATTERJI, D. High intracellular level of guanosine tetraphosphate in Mycobacterium smegmatis changes the morphology of the bacterium. **Infection and Immunity**, v. 68, n. 7, p. 4084–4091, 2000.

OLSEN, C. R.; HALE, F. C.; ELSNER, R. Mechanics of ventilation in the pilot whale. **Respiration Physiology**, v. 7, n. 2, p. 137–149, 1969.

PANDEY, A. K.; SASSETTI, C. M. Mycobacterial persistence requires the utilization of host cholesterol. **Proceedings of the National Academy of Sciences**, v. 105, n. 11, p. 4376–4380, 2008.

PARK, H.-D.; GUINN, K. M.; HARRELL, M. I.; LIAO, R.; VOSKUIL, M. I.; TOMPA, M.; SCHOOLNIK, G. K.; SHERMAN, D. R. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. **Molecular Microbiology**, v. 48, n. 3, p. 833–843, 2003.

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

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

PIROFSKI, L. anne; CASADEVALL, A. Q and A What is a pathogen? A question that begs the point. **BMC Biology**, v. 10, n. 1, p. 6, 2012.

PŁUCIENNIK, A.; STOLARCZYK, M.; BZÓWKA, M.; RACZYŃSKA, A.; MAGDZIARZ, T.; GÓRA, A. BALCONY: an R package for MSA and functional compartments of protein variability analysis. **BMC Bioinformatics**, v. 19, n. 1, p. 300, 2018.

PORTEVIN, D.; GAGNEUX, S.; COMAS, I.; YOUNG, D. Human Macrophage Responses to Clinical Isolates from the Mycobacterium tuberculosis Complex Discriminate between Ancient and Modern Lineages. **PLoS Pathogens**, v. 7, n. 3, p. e1001307, 2011.

PORTEVIN, D.; SUKUMAR, S.; COSCOLLA, M.; SHUI, G.; LI, B.; GUAN, X. L.; BENDT, A. K.; YOUNG, D.; GAGNEUX, S.; WENK, M. R. Lipidomics and genomics of Mycobacterium tuberculosis reveal lineage-specific trends in mycolic acid biosynthesis. **MicrobiologyOpen**, v. 3, n. 6, p. 823–35, 2014.

PRUITT, K. D.; TATUSOVA, T.; MAGLOTT, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. **Nucleic acids research**, v. 35, n. Database issue, p. D61-5, 2007.

QUADRI, L. E. N. Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. Critical Reviews in Biochemistry and Molecular Biology, Biosynthesis of mycobacterial lipids by polyketide synthases and beyond, v. 49, n. 3, p. 179-211, 2014.

R CORE TEAM. R: A language and environment for statistical computing. 2013.

RAHMAN, S. A.; SINGH, Y.; KOHLI, S.; AHMAD, J.; EHTESHAM, N. Z.; TYAGI, A. K.; HASNAIN, S. E. Comparative analyses of nonpathogenic, opportunistic, and totally pathogenic mycobacteria reveal genomic and biochemical variabilities and highlight the survival attributes of Mycobacterium tuberculosis. **mBio**, v. 5, n. 6, p. e02020-14, 2014.

RAMAKRISHNAN, L.; FEDERSPIEL, N. A.; FALKOW, S. Granuloma-Specific Expression of Mycobacterium Virulence Proteins from the Glycine-Rich PE-PGRS Family. **Science**, v. 288, n. 5470, p. 1436–1439, 2000.

REGMI, S. M.; CHAIPRASERT, A.; KULAWONGANUNCHAI, S.; TONGSIMA, S.; COKER, O. O.; PRAMMANANAN, T.; VIRATYOSIN, W.; THAIPISUTTIKUL, I. Whole genome sequence analysis of multidrug-resistant Mycobacterium tuberculosis Beijing isolates from an outbreak in Thailand. **Molecular Genetics and Genomics**, v. 290, n. 5, p. 1933–1941, 2015.

RODRÍGUEZ-CASTILLO, J. G.; PINO, C.; NIÑO, L. F.; ROZO, J. C.; LLERENA-POLO, C.; PARRA-LÓPEZ, C. A.; TAUCH, A.; MURCIA-ARANGUREN, M. I. Comparative genomic analysis of Mycobacterium tuberculosis Beijing-like strains revealed specific genetic variations associated with virulence and drug resistance. **Infection, Genetics and Evolution**, v. 54, p. 314–323, 2017.

SASSETTI, C. M.; RUBIN, E. J. Genetic requirements for mycobacterial survival during infection. **Proceedings of the National Academy of Sciences**, v. 100, n. 22, p. 12989–12994, 2003.

SCHNAPPINGER, D.; EHRT, S.; VOSKUIL, M. I.; LIU, Y.; MANGAN, J. A.; MONAHAN, I. M.; DOLGANOV, G.; EFRON, B.; BUTCHER, P. D.; NATHAN, C.; SCHOOLNIK, G. K. Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: Insights into the phagosomal environment. **Journal of Experimental Medicine**, v. 198, n. 5, p. 693–704, 2003.

SHERMAN, D. R.; VOSKUIL, M.; SCHNAPPINGER, D.; LIAO, R.; HARRELL, M. I.; SCHOOLNIK, G. K. Regulation of the mycobacterium tuberculosis hypoxic response gene encoding α-crystallin. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 13, p. 7534–7539, 2001.

SINGH, A.; JAIN, S.; GUPTA, S.; DAS, T.; TYAGI, A. K. mymA operon of Mycobacterium tuberculosis: its regulation and importance in the cell envelope. **FEMS Microbiology Letters**, v. 227, n. 1, p. 53–63, 2003.

SOUSA, E. H. S.; GONZALEZ, G.; GILLES-GONZALEZ, M.-A. Target DNA stabilizes Mycobacterium tuberculosis DevR/DosR phosphorylation by the full-length oxygen sensors DevS/DosS and DosT. **The FEBS Journal**, v. 284, n. 22, p. 3954–3967, 2017.

SU, H.; ZHANG, Z.; LIU, Z.; PENG, B.; KONG, C.; WANG, H.; ZHANG, Z.; XU, Y. Mycobacterium tuberculosis PPE60 antigen drives Th1/Th17 responses via Toll-like receptor 2-dependent maturation of dendritic cells. **The Journal of biological chemistry**, v. 293, n. 26, p. 10287–10302, 2018.

SUPPLY, P.; MARCEAU, M.; MANGENOT, S.; ROCHE, D.; ROUANET, C.; KHANNA, V.; MAJLESSI, L.; CRISCUOLO, A.; TAP, J.; PAWLIK, A.; FIETTE, L.; ORGEUR, M.; FABRE, M.; PARMENTIER, C.; FRIGUI, W.; SIMEONE, R.; BORITSCH, E. C.; DEBRIE, A.-S.; WILLERY, E.; WALKER, D.; QUAIL, M. A.; MA, L.; BOUCHIER, C.; SALVIGNOL, G.; SAYES, F.; CASCIOFERRO, A.; SEEMANN, T.; BARBE, V.; LOCHT, C.; GUTIERREZ, M.-C.; LECLERC, C.; BENTLEY, S. D.; STINEAR, T. P.; BRISSE, S.; MÉDIGUE, C.; PARKHILL, J.; CRUVEILLER, S.; BROSCH, R. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. **Nature Genetics**, v. 45, n. 2, p. 172–179, 2013.

SWARBRICK, C. M. D.; BYTHROW, G. V.; ARAGAO, D.; GERMAIN, G. A.; QUADRI, L. E. N.; FORWOOD, J. K. Mycobacteria Encode Active and Inactive Classes of TesB Fatty-Acyl CoA Thioesterases Revealed through Structural and Functional Analysis. **Biochemistry**, v. 56, n. 10, p. 1460–1472, 2017.

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

TORRELLES, J. B.; KNAUP, R.; KOLARETH, A.; SLEPUSHKINA, T.; KAUFMAN, T. M.; KANG, P.; HILL, P. J.; BRENNAN, P. J.; CHATTERJEE, D.; BELISLE, J. T.; MUSSER, J. M.; SCHLESINGER, L. S. Identification of Mycobacterium tuberculosis clinical isolates with altered phagocytosis by human macrophages due to a truncated lipoarabinomannan. **Journal of Biological Chemistry**, v. 283, n. 46, p. 31417–31428, 2008.

TORRELLES, J. B.; SCHLESINGER, L. S. Diversity in Mycobacterium tuberculosis mannosylated cell wall determinants impacts adaptation to the host. **Tuberculosis**, v. 90, n. 2, p. 84-93, 2010.

TRAORE, B.; DIARRA, B.; DEMBELE, B. P. P.; SOMBORO, A. M.; HAMMOND, A. S.; SIDDIQUI, S.; MAIGA, M.; KONE, B.; SARRO, Y. S.; WASHINGTON, J.; PARTA, M.; COULIBALY, N.; M'BAYE, O.; DIALLO, S.; KOITA, O.; TOUNKARA, A.; POLIS, M. A. Molecular strain typing of Mycobacterium tuberculosis complex in Bamako, Mali. **International Journal of Tuberculosis and Lung Disease**, v. 16, n. 7, p. 911–916, 2012.

UPLEKAR, S.; HEYM, B.; FRIOCOURT, V.; ROUGEMONT, J.; COLE, S. T. Comparative Genomics of esx Genes from Clinical Isolates of Mycobacterium tuberculosis Provides Evidence for Gene Conversion and Epitope Variation †. **Infection and immunity**, v. 79, n. 10, p. 4042–4049, 2011.

VANDER BEKEN, S.; AL DULAYYMI, J. R.; NAESSENS, T.; KOZA, G.; MAZA-IGLESIAS, M.; ROWLES, R.; THEUNISSEN, C.; DE MEDTS, J.; LANCKACKER, E.; BAIRD, M. S.; GROOTEN, J. Molecular structure of the Mycobacterium tuberculosis virulence factor, mycolic acid, determines the elicited inflammatory pattern. **European Journal of Immunology**, v. 41, n. 2, p. 450–460, 2011.

VERNIKOS, G.; MEDINI, D.; RILEY, D. R.; TETTELIN, H. Ten years of pan-genome analyses. **Current Opinion in Microbiology**, v. 23, p. 148–154, 2015.

VITA, R.; MAHAJAN, S.; OVERTON, J. A.; DHANDA, S. K.; MARTINI, S.; CANTRELL, J. R.; WHEELER, D. K.; SETTE, A.; PETERS, B. The Immune Epitope Database (IEDB): 2018 update. **Nucleic Acids Research**, v. 47, n. D1, p. D339–D343, 2019.

VOSKUIL, M. I.; SCHNAPPINGER, D.; VISCONTI, K. C.; HARRELL, M. I.; DOLGANOV, G. M.; SHERMAN, D. R.; SCHOOLNIK, G. K. Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program. **The Journal of**

experimental medicine, v. 198, n. 5, p. 705–13, 2003.

WANG, J.; LI, B. X.; GE, P. P.; LI, J.; WANG, Q.; GAO, G. F.; QIU, X. B.; LIU, C. H. Mycobacterium tuberculosis suppresses innate immunity by coopting the host ubiquitin system. **Nature Immunology**, v. 16, n. 3, p. 237–245, 2015.

WARNES, G. R.; BOLKER, B.; BONEBAKKER, L.; GENTLEMAN, R.; LIAW, W. H. A.; LUMLEY, T.; MAECHLER, M.; MAGNUSSON, A.; MOELLER, S.; SCHWARTZ, M.; VENABLES, B. gplots: Various R Programming Tools for Plotting Data. R package version 3.0-1. **The Comprehensive R Archive Network**, 2016.

WAYNE, L. G.; HAYES, L. G. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. **Infection and immunity**, v. 64, n. 6, p. 2062–9, 1996.

WINGLEE, K.; MANSON MCGUIRE, A.; MAIGA, M.; ABEEL, T.; SHEA, T.; DESJARDINS, C. A.; DIARRA, B.; BAYA, B.; SANOGO, M.; DIALLO, S.; EARL, A. M.; BISHAI, W. R. Whole Genome Sequencing of Mycobacterium africanum Strains from Mali Provides Insights into the Mechanisms of Geographic Restriction. **PLOS Neglected Tropical Diseases**, v. 10, n. 1, p. e0004332, 2016.

WORLD HEALTH ORGANIZATION. **WHO Tuberculosis Report 2018**. Geneva: World Health Organization: World Health Organization, 2018.

XIONG, X.; WANG, R.; DENG, D.; CHEN, Y.; LIU, H.; WANG, T.; WANG, J.; ZHU, X.; ZHU, X.; ZHU, Y.; LU, X.; CHEN, H.; ZHENG, H.; GUO, A. Comparative Genomics of a Bovine Mycobacterium tuberculosis Isolate and Other Strains Reveals Its Potential Mechanism of Bovine Adaptation. **Frontiers in Microbiology**, v. 8, p. 2500, 2017.

YANG, T.; SHENG, Y.; YUE, L.; DING, N.; WANG, G.; ZHONG, J.; XIAO, J.; YU, X.; LI, C.; CHEN, F.; ZHANG, J.; SUN, Y.; LIANG, Q.; JIA, X.; HUANG, H.; MA, G. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. **Frontiers in Microbiology**, v. 9, n. August, p. 1–12, 2018.

YAR, A. M.; ZAMAN, G.; HUSSAIN, A.; CHANGHUI, Y.; RASUL, A.; HUSSAIN, A.; BO, Z.; BOKHARI, H.; IBRAHIM, M. Comparative Genome Analysis of 2 Mycobacterium Tuberculosis Strains from Pakistan: Insights Globally Into Drug Resistance, Virulence, and Niche Adaptation. **Evolutionary Bioinformatics**, v. 14, p. 1176934318790252, 2018.

ZHANG, F.; XIE, J.-P. Mammalian cell entry gene family of Mycobacterium tuberculosis. **Molecular and Cellular Biochemistry**, v. 352, n. 1–2, p. 1–10, 2011.

ZIMPEL, C. K.; BRANDÃO, P. E.; DE SOUZA FILHO, A. F.; DE SOUZA, R. F.; IKUTA, C. Y.; NETO, J. S. F.; SOLER CAMARGO, N. C.; HEINEMANN, M. B.; GUIMARÃES, A. M. S. Complete genome sequencing of Mycobacterium bovis SP38 and comparative genomics of Mycobacterium bovis and M. tuberculosis strains.

Frontiers in Microbiology, v. 8, n. DEC, p. 2389, 2017.

6. CONSIDERAÇÕES FINAIS

Este estudo trata-se do primeiro relato de superinfecção por M. pinnipedii em um leãomarinho, assim como o primeiro estudo a realizar uma análise da presença e variabilidade de epítopos e repertório de fatores de virulência em todas as espécies do MTBC disponíveis no RefSeq (NCBI). Foi possível observar que as estirpes de M. pinnipedii MP1 e MP2 isoladas do mesmo hospedeiro apresentaram características comuns a outras micobactérias, e que sua comparação com outros genomas da mesma espécie disponíveis em bancos de dados públicos indica que as estirpes apresentam alto nível de conservação do proteoma. Análises de diferenças nucleotídicas das estirpes MP1 e MP2 mostraram que se tratava de dois isolados infectando o mesmo animal (i.e. superinfecção ou infecção mista). Dados das anotações funcionais apontaram que a maioria dos SNPs detectados eram não-sinônimos, o que siginifica que há uma potencial promoção de diversidade funcional devido a pressão seletiva para mudança da sequência de proteínas nos CDSs relacionados. Foram identificadas fortes interações relacionadas à proteínas da família VapBC, indicando que essas proteínas específicas estão sob pressão evolutiva por mudança. Surpreendentemente, no estudo comparativo realizado entre diferentes espécies do MTBC para analisar a presença de fatores devirulência, as estirpes de M. pinnipedii apresentaram diferenças significativas em todos os clusters de FVs quando comparados às outras estirpes do complexo, inclusive nos clusters que se apresentaram mais conservados. Este achado sugere uma possível co-evoluação entre bactéria e hospedeiro, uma vez que o M. pinnipedii é a única espécie do MTBC que se mantêm em mamíferos não terrestres. Estudos futuros envolvendo sequenciamentos de mais isolados de M. pinnipedii são necessários para compor o conjunto de dados e permitir análises que possam contribuir para encontrar diferenças entre os proteomas de isolados da mesma espécies. Além disso, estudos voltados para análise do conteúdo gênico e avaliação de infecções in vitro com essas estirpes podem contribuir no entendimento do efeito da pressão seletiva da família de proteínas VapBC e desse repertório diferenciado de FVs.

Nossos estudos seugerem, pela primeira vez, um mecanismo de variação antigênica de epítopos de células T do MTBC baseado no silenciamento/perda de genes, devido a presença de epítopos preditos em proteínas presentes apenas em algumas estirpes analisadas. A análise de variabilidade de sequência dos epítopos corrobora estudos passados quanto a hiperconservação, mas indica também que há pequenas variações antigênicas nas proteínas analisadas, principalmente naquelas que compõem o genoma acessório. A predição de epítopos em proteínas do genoma acessório mostra a presença considerável de epítopos em proteínas da

família PE/PPE, que geralmente são regiões excluídas de análises em representantes do MTBC devido a limitações técnicas da utilização de *reads*/leituras. Um próximo passo consistirá em analisar se estes epítopos estão em proteínas variáveis entre as espécies do MTBC, podendo assim contribuir para variabilidade do tropismo por hospedeiro, uma vez que esses são epítopos preditos como sendo humanos. Além do mais, é imprescindível que mais estudos sejam realizados sobre as proteínas da família PE/PPE, com todas as espécies do MTBC, que apesar das dificuldades técnicas, podem promover novas discussões no que tange a interação patógeno-hospedeiro na TB.

A estrutura do pan-genoma e repertório de fatores de virulência do complexo *M. tuberculosis* mostraram-se variáveis quanto a componentes do metabolismo de lipídios e metabólitos secundários, indicando que essas categorias desempenham funções importantes nos diferentes fenótipos e tropismo por hospedeiros e virulência do MTBC. Estudos futuros que realizem uma análise mais profunda nessas e em outras variantes específicas para elucidar determinantes metabólicos de virulência podem oferecer mais informações sobre a patogenicidade dessas bactérias e sobre os diferentes fenótipos do MTBC. O envoltório bacteriano parece ser crucial nessa interação patógeno-hospedeiro, como determinante do tropismo diferencial e também da virulência. A presença variável de genes no genoma acessório e genes únicos do genoma estirpe-específico pode representar uma alternativa para o diagnóstico bacterianos baseado em marcadores específicos ou para estratégias terapêuticas baseadas em inibidores de virulência específicos.

Há evidências de que lacunas do conhecimento precisam ser preenchidas para melhor compreensão do processo infeccioso de representantes do MTBC. Sabe-se que o combate a TB é uma preocupação global, sendo um dos focos da OMS com o *End TB Strategy* que objetiva acabar com epidemias de TB. Uma futura comparação das características metabólicas compartilhadas entre estirpes que causam TB podem fornecer infomações adicionais, oferecendo novas maneiras de inibir a disseminação epidemiológica dessas estirpes e permitindo a identificação de mecanismos determinantes da patogênese relacionados à fatores de virulência, mecanismo de evasão da resposta imune, entre outros. Novas tecnologias de sequenciamento de DNA de alto rendimento e técnicas de análises computacionais continuarão a melhorar nossa capacidade de caracterizar patógenos. Estudos em bioinformática com foco em determinantes metabólicos de virulência podem fornecer ferramentas poderosas para interpretar informações obtidas a partir de dados genômicos para compreender os diferentes fenótipos dessas bactérias, com potencial contribuição na descoberta de novos alvos farmacológicos, e fenótipos de virulência e adaptabilidade hospedeiras que podem servir como

base no desenvolvimento de vacinas, uma vez que a atual formulação utilizada no controle da TB em humanos é produzida a partir do *M. bovis* (BCG, *Bacille Calmette-Guérin*).

REFERÊNCIAS

ALEXANDER, K. A.; LARSEN, M. H.; ROBBE-AUSTERMAN, S.; STUBER, T. P.; CAMP, P. M. Draft genome sequence of the Mycobacterium tuberculosis complex Pathogen M. mungi, identified in a banded Mongoose (Mungos mungo) in Northern Botswana. **Genome Announcements**, v. 4, n. 4, p. e00471-16, 2016.

ALEXANDER, K. A.; LAVER, P. N.; MICHEL, A. L.; WILLIAMS, M.; VAN HELDEN, P. D.; WARREN, R. M.; GEY VAN PITTIUS, N. C. Novel Mycobacterium tuberculosis complex pathogen, M. mungi. **Emerging infectious diseases**, v. 16, n. 8, p. 1296–9, 2010.

ALI, A.; HASAN, Z.; MCNERNEY, R.; MALLARD, K.; HILL-CAWTHORNE, G.; COLL, F.; NAIR, M.; PAIN, A.; CLARK, T. G.; HASAN, R. Whole genome sequencing based characterization of extensively drug-resistant mycobacterium tuberculosis isolates from pakistan. **PLoS ONE**, v. 10, n. 2, p. e0117771, 2015.

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

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 1990.

ARCUS, V. L.; RAINEY, P. B.; TURNER, S. J. The PIN-domain toxin–antitoxin array in mycobacteria. **Trends in Microbiology**, v. 13, n. 8, p. 360–365, 2005.

ARKIN, A. P.; COTTINGHAM, R. W.; HENRY, C. S.; HARRIS, N. L.; STEVENS, R. L.; MASLOV, S.; DEHAL, P.; WARE, D.; PEREZ, F.; CANON, S.; SNEDDON, M. W.; HENDERSON, M. L.; RIEHL, W. J.; MURPHY-OLSON, D.; CHAN, S. Y.; KAMIMURA, R. T.; KUMARI, S.; DRAKE, M. M.; BRETTIN, T. S.; GLASS, E. M.; CHIVIAN, D.; GUNTER, D.; WESTON, D. J.; ALLEN, B. H.; BAUMOHL, J.; BEST, A. A.; BOWEN, B.; BRENNER, S. E.; BUN, C. C.; CHANDONIA, J.-M.; CHIA, J.-M.; COLASANTI, R.; CONRAD, N.; DAVIS, J. J.; DAVISON, B. H.; DEJONGH, M.; DEVOID, S.; DIETRICH, E.; DUBCHAK, I.; EDIRISINGHE, J. N.; FANG, G.; FARIA, J. P.; FRYBARGER, P. M.; GERLACH, W.; GERSTEIN, M.; GREINER, A.; GURTOWSKI, J.; HAUN, H. L.; HE, F.; JAIN, R.; JOACHIMIAK, M. P.; KEEGAN, K. P.; KONDO, S.; KUMAR, V.; LAND, M. L.; MEYER, F.; MILLS, M.; NOVICHKOV, P. S.; OH, T.; OLSEN, G. J.; OLSON, R.; PARRELLO, B.; PASTERNAK, S.; PEARSON, E.; POON, S. S.; PRICE, G. A.; RAMAKRISHNAN, S.; RANJAN, P.; RONALD, P. C.; SCHATZ, M. C.; SEAVER, S. M. D.; SHUKLA, M.; SUTORMIN, R. A.; SYED, M. H.; THOMASON, J.; TINTLE, N. L.; WANG, D.; XIA, F.; YOO, H.; YOO, S.; YU, D. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnology, v. 36, n. 7, p. 566–569,

2018.

ASHFORD, D. A.; WHITNEY, E.; RAGHUNATHAN, P.; COSIVI, O. Epidemiology of selected mycobacteria that infect humans and other animals. **Revue scientifique et technique** (**International Office of Epizootics**), v. 20, n. 1, p. 325–37, 2001.

ASTARIE-DEQUEKER, C.; NIGOU, J.; PASSEMAR, C.; GUILHOT, C. The role of mycobacterial lipids in host pathogenesis. **Drug Discovery Today: Disease Mechanisms**, v. 7, n. 1, p. e33-e41, 2010.

ATES, L. S.; HOUBEN, E. N. G.; BITTER, W. Type VII Secretion: A Highly Versatile Secretion System. **Microbiology spectrum**, v. 4, n. 1, p. 357–384, 2016.

AZIZ, R. K.; BARTELS, D.; BEST, A. A.; DEJONGH, M.; DISZ, T.; EDWARDS, R. A.; FORMSMA, K.; GERDES, S.; GLASS, E. M.; KUBAL, M.; MEYER, F.; OLSEN, G. J.; OLSON, R.; OSTERMAN, A. L.; OVERBEEK, R. A.; MCNEIL, L. K.; PAARMANN, D.; PACZIAN, T.; PARRELLO, B.; PUSCH, G. D.; REICH, C.; STEVENS, R.; VASSIEVA, O.; VONSTEIN, V.; WILKE, A.; ZAGNITKO, O. The RAST Server: Rapid Annotations using Subsystems Technology. **BMC Genomics**, v. 9, n. 1, p. 75, 2008.

BANKEVICH, A.; NURK, S.; ANTIPOV, D.; GUREVICH, A. A.; DVORKIN, M.; KULIKOV, A. S.; LESIN, V. M.; NIKOLENKO, S. I.; PHAM, S.; PRJIBELSKI, A. D.; PYSHKIN, A. V; SIROTKIN, A. V; VYAHHI, N.; TESLER, G.; ALEKSEYEV, M. A.; PEVZNER, P. A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology**, v. 19, n. 5, p. 455–77, 2012.

BASTIDA, R.; LOUREIRO, J.; QUSE, V.; BERNARDELLI, A.; RODRÍGUEZ, D.; COSTA, E. Tuberculosis in a Wild Subantarctic Fur Seal from Argentina. **Journal of Wildlife Diseases**, v. 35, p. 796–798, 1999.

BASU, S.; PATHAK, S. K.; BANERJEE, A.; PATHAK, S.; BHATTACHARYYA, A.; YANG, Z.; TALARICO, S.; KUNDU, M.; BASU, J. Execution of macrophage apoptosis by PE_PGRS33 of Mycobacterium tuberculosis is mediated by toll-like receptor 2-dependent release of tumor necrosis factor-α. **Journal of Biological Chemistry**, v. 282, n. 2, p. 1039–1050, 2007.

BAUMLER, A.; FANG, F. C. Host Specificity of Bacterial Pathogens. **Cold Spring Harbor Perspectives in Medicine**, v. 3, n. 12, p. a010041–a010041, 2013.

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

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

BENTLEY, S. D.; COMAS, I.; BRYANT, J. M.; WALKER, D.; SMITH, N. H.; HARRIS, S. R.; THURSTON, S.; GAGNEUX, S.; WOOD, J.; ANTONIO, M.; QUAIL, M. A.; GEHRE, F.; ADEGBOLA, R. A.; PARKHILL, J.; DE JONG, B. C. The Genome of Mycobacterium Africanum West African 2 Reveals a Lineage-Specific Locus and Genome Erosion Common to the M. tuberculosis Complex. **PLoS Neglected Tropical Diseases**, v. 6, n. 2, p. e1552, 2012.

BERG, S.; GARCIA-PELAYO, M. C.; MÜLLER, B.; HAILU, E.; ASIIMWE, B.; KREMER, K.; DALE, J.; BONIOTTI, M. B.; RODRIGUEZ, S.; HILTY, M.; RIGOUTS, L.; FIRDESSA, R.; MACHADO, A.; MUCAVELE, C.; NGANDOLO, B. N. R.; BRUCHFELD, J.; BOSCHIROLI, L.; MÜLLER, A.; SAHRAOUI, N.; PACCIARINI, M.; CADMUS, S.; JOLOBA, M.; VAN SOOLINGEN, D.; MICHEL, A. L.; DJØNNE, B.; ARANAZ, A.; ZINSSTAG, J.; VAN HELDEN, P.; PORTAELS, F.; KAZWALA, R.; KÄLLENIUS, G.; HEWINSON, R. G.; ASEFFA, A.; GORDON, S. V.; SMITH, N. H. African 2, a clonal complex of Mycobacterium bovis epidemiologically important in East Africa. **Journal of Bacteriology**, v. 193, n. 3, p. 670–678, 2011.

BERNARDELLI, A.; BASTIDA, R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; BERNARDELLI BASTIDA, A. R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; MUNDO MARINO, F.; CLEMENTE DEL TUYÚ, S.; BERNARDELLI, A.; BASTIDA, R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; BERNARDELLI BASTIDA, A. R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E.; MUNDO MARINO, F.; CLEMENTE DEL TUYÚ, S.; BERNARDELLI, A.; BASTIDA, R.; LOUREIRO, J.; MICHELIS, H.; ROMANO, M. I.; CATALDI, A.; COSTA, E. Tuberculosis in sea lions and fur seals from the south-western Atlantic coast. **Revue scientifique et technique (International Office of Epizootics)**, v. 15, n. 3, p. 985–1005, 1996.

BLACK, P. A.; DE VOS, M.; LOUW, G. E.; VAN DER MERWE, R. G.; DIPPENAAR, A.; STREICHER, E. M.; ABDALLAH, A. M.; SAMPSON, S. L.; VICTOR, T. C.; DOLBY, T.; SIMPSON, J. A.; VAN HELDEN, P. D.; WARREN, R. M.; PAIN, A. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates. **BMC Genomics**, v. 16, n. 1, p. 857, 2015.

BLAIR WR. Report of the veterinarian on the mammals. **17th Annual Report of the New York Zoological Society**, v. 74, p. 73–77, 1913.

BOARDMAN, W. S. J.; SHEPHARD, L.; BASTIAN, I.; GLOBAN, M.; FYFE, J. A. M.; COUSINS, D. V.; MACHADO, A.; WOOLFORD, L. Mycobacterium pinnipedii Tuberculosis in a Free-Ranging Australian Fur Seal (Arctocephalus pusillus doriferus) In South Australia. **Journal of Zoo and Wildlife Medicine**, v. 45, n. 4, p. 970–972, 2014.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics** (**Oxford, England**), v. 30, n. 15, p. 2114–20, 1 ago. 2014. BOLOTIN, E.; HERSHBERG, R. Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. **Genome Biology and Evolution**, v. 7, n. 8, p. 2173–2187, 2015.

BOS, K. I.; HARKINS, K. M.; HERBIG, A.; COSCOLLA, M.; WEBER, N.; COMAS, I. I. I.; FORREST, S. A.; BRYANT, J. M.; HARRIS, S. R.; SCHUENEMANN, V. J.; CAMPBELL, T. J.; MAJANDER, K.; WILBUR, A. K.; GUICHON, R. A.; WOLFE STEADMAN, D. L.; COOK, D. C.; NIEMANN, S.; BEHR, M. A.; ZUMARRAGA, M.; BASTIDA, R.; HUSON, D.; NIESELT, K.; YOUNG, D.; PARKHILL, J.; BUIKSTRA, J. E.; GAGNEUX, S.; STONE, A. C.; KRAUSE, J. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. **Nature**, v. 514, n. 7523, p. 494–497, 2014.

BRENNAN, M. J. The Enigmatic PE/PPE Multigene Family of Mycobacteria and Tuberculosis Vaccination. **Infection and Immunity**, v. 85, n. 6, p. e00969-16, 2017.

BRENNAN, M. J.; DELOGU, G. The PE multigene family: A "molecular mantra" for mycobacteria. **Trends in Microbiology**, v. 10, n. 5, p. 246-249, 2002.

BRENNAN, P. J. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. **Tuberculosis**, v. 83, n. 1–3, p. 91–97, 2003.

BRITES, D.; GAGNEUX, S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. **Immunological Reviews**, v. 264, n. 1, p. 6–24, 2015.

BRITES, D.; LOISEAU, C.; MENARDO, F.; BORRELL, S.; BONIOTTI, M. B.; WARREN, R.; DIPPENAAR, A.; PARSONS, S. D. C.; BEISEL, C.; BEHR, M. A.; FYFE, J. A.; COSCOLLA, M.; GAGNEUX, S. A New Phylogenetic Framework for the Animal-Adapted Mycobacterium tuberculosis Complex. **Frontiers in Microbiology**, v. 9, p. 2820, 2018.

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

BROSCH, R.; GORDON, S. V.; PYM, A.; EIGLMEIER, K.; GARNIER, T.; COLE, S. T. Comparative genomics of the mycobacteria. **International Journal of Medical Microbiology**, v. 290, n. 2, p. 143–152, 2000.

- BROSCH, R.; PHILIPP, W. J.; STAVROPOULOS, E.; COLSTON, M. J.; COLE, S. T.; GORDON, S. V. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. **Infection and immunity**, v. 67, n. 11, p. 5768–74, 1999.
- BROWN, A. C.; BRYANT, J. M.; EINER-JENSEN, K.; HOLDSTOCK, J.; HOUNIET, D. T.; CHAN, J. Z. M.; DEPLEDGE, D. P.; NIKOLAYEVSKYY, V.; BRODA, A.; STONE, M. J.; CHRISTIANSEN, M. T.; WILLIAMS, R.; MCANDREW, M. B.; TUTILL, H.; BROWN, J.; MELZER, M.; ROSMARIN, C.; MCHUGH, T. D.; SHORTEN, R. J.; DROBNIEWSKI, F.; SPEIGHT, G.; BREUER, J. Rapid whole-genome sequencing of mycobacterium tuberculosis isolates directly from clinical samples. **Journal of Clinical Microbiology**, v. 53, n. 7, p. 2230–2237, 2015.
- CADMUS, S. I. B.; YAKUBU, M. K.; MAGAJI, A. A.; JENKINS, A. O.; VAN SOOLINGEN, D. Mycobacterium bovis, but also M. africanum present in raw milk of pastoral cattle in north-central Nigeria. **Tropical Animal Health and Production**, v. 42, n. 6, p. 1047–1048, 2010.
- CAMARGO, N. C. S. **Investigação da taxa de pseudogenização em genomas do complexo Mycobacterium tuberculosis.** 2018. 39 f. Trabalho de Conclusão de Curso (Graduação em Bacharelado em Biotecnologia) Universidade Federal de São Paulo.
- CHEN, L.; YANG, J.; YU, J.; YAO, Z.; SUN, L.; SHEN, Y.; JIN, Q. VFDB: a reference database for bacterial virulence factors. **Nucleic Acids Research**, v. 33, n. Database issue, p. D325–D328, 2004.
- CINGOLANI, P.; PLATTS, A.; WANG, L. L.; COON, M.; NGUYEN, T.; WANG, L.; LAND, S. J.; LU, X.; RUDEN, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. **Fly**, v. 6, n. 2, p. 80–92, 2012.
- COHEN, T.; VAN HELDEN, P. D.; WILSON, D.; COLIJN, C.; MCLAUGHLIN, M. M.; ABUBAKAR, I.; WARREN, R. M. Mixed-Strain Mycobacterium tuberculosis infections and the implications for tuberculosis treatment and control. **Clinical microbiology reviews**, v. 25, n. 4, p. 708-719, 2012.
- COHEN, T.; WILSON, D.; WALLENGREN, K.; SAMUEL, E. Y.; MURRAY, M. Mixed-strain Mycobacterium tuberculosis infections among patients dying in a hospital in KwaZulu-Natal, South Africa. **Journal of clinical microbiology**, v. 49, n. 1, p. 385–8, 2011.
- COLE, S. T.; BROSCH, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON, S. V.; EIGLMEIER, K.; GAS, S.; BARRY, C. E.; 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, M. A.; RAJANDREAM, M.-A.; ROGERS, J.; RUTTER, S.; SEEGER, K.; SKELTON, J.; SQUARES, R.; SQUARES, S.; SULSTON, J. E.; TAYLOR, K.; WHITEHEAD, S.; BARRELL, B. G. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. **Nature**, v. 393, n. 6685, p. 537–544, 1998.

COMAS, I.; CHAKRAVARTTI, J.; SMALL, P. M.; GALAGAN, J.; NIEMANN, S.; KREMER, K.; ERNST, J. D.; GAGNEUX, S. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. **Nature Genetics**, v. 42, n. 6, p. 498–503, 2010.

COMAS, I.; COSCOLLA, M.; LUO, T.; BORRELL, S.; HOLT, K. E.; KATO-MAEDA, M.; PARKHILL, J.; MALLA, B.; BERG, S.; THWAITES, G.; YEBOAH-MANU, D.; BOTHAMLEY, G.; MEI, J.; WEI, L.; BENTLEY, S.; HARRIS, S. R.; NIEMANN, S.; DIEL, R.; ASEFFA, A.; GAO, Q.; YOUNG, D.; GAGNEUX, S. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. **Nature genetics**, v. 45, n. 10, p. 1176–82, 2013.

CONSTANT, P.; PEREZ, E.; MALAGA, W.; LANÉELLE, M.-A.; SAUREL, O.; DAFFÉ, M.; GUILHOT, C. Role of the pks15/1 Gene in the Biosynthesis of Phenolglycolipids in the Mycobacterium tuberculosis Complex. **Journal of Biological Chemistry**, v. 277, n. 41, p. 38148–38158, 2002.

CORPER, H. J. The Murine Type of Tubercle Bacillus (The Vole Acid-Fast Bacillus). **American Journal of Clinical Pathology**, v. 17, n. 11, p. 886.1-886, 1947.

COSCOLLA, M.; COPIN, R.; SUTHERLAND, J.; GEHRE, F.; DE JONG, B.; OWOLABI, O.; MBAYO, G.; GIARDINA, F.; ERNST, J. D.; GAGNEUX, S. M. tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens. **Cell host & microbe**, v. 18, n. 5, p. 538–48, 2015.

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

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

COUSINS, D. V.; BASTIDA, R.; CATALDI, A.; QUSE, V.; REDROBE, S.; DOW, S.;

DUIGNAN, P.; MURRAY, A.; DUPONT, C.; AHMED, N.; COLLINS, D. M.; BUTLER, W. R.; DAWSON, D.; RODRÍGUEZ, D.; LOUREIRO, J.; ROMANO, M. I.; ALITO, A.; ZUMARRAGA, M.; BERNARDELLI, A. Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 53, n. 5, p. 1305–1314, 2003.

COUSINS, D. V.; WILLIAMS, S. N.; REUTER, R.; FORSHAW, D.; CHADWICK, B.; COUGHRAN, D.; COLLINS, P.; GALES, N. Tuberculosis in wild seals and characterisation of the seal bacillus. **Australian veterinary journal**, v. 70, n. 3, p. 92–97, 1993.

COUSINS, D. V. ELISA for detection of tuberculosis in seals. **The Veterinary record**, v. 121, n. 13, p. 305, 1987.

COUSINS, D. V; FRANCIS, B. R.; GOW, B. L.; COLLINS, D. M.; MCGLASHAN, C. H.; GREGORY, A.; MACKENZIE, R. M. Tuberculosis in captive seals: bacteriological studies on an isolate belonging to the Mycobacterium tuberculosis complex. **Research in veterinary science**, v. 48, n. 2, p. 196–200, 1990.

COUSINS, D. V. V; PEET, R. L. L.; GAYNOR, W. T. T.; WILLIAMS, S. N. N.; GOW, B. L. L. Tuberculosis in imported hyrax (Procavia capensis) caused by an unusual variant belonging to the Mycobacterium tuberculosis complex. **Veterinary microbiology**, v. 42, n. 2–3, p. 135–45, 1994.

COWLEY, S. C.; BABAKAIFF, R.; AV-GAY, Y. Expression and localization of the Mycobacterium tuberculosis protein tyrosine phosphatase PtpA. **Research in Microbiology**, v. 153, n. 4, p. 233–241, 2002.

CROOKS, G. E.; HON, G.; CHANDONIA, J. M.; BRENNER, S. E. WebLogo: A sequence logo generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, 2004.

DAUM, L. T.; RODRIGUEZ, J. D.; WORTHY, S. A.; ISMAIL, N. A.; OMAR, S. V.; DREYER, A. W.; FOURIE, P. B.; HOOSEN, A. A.; CHAMBERS, J. P.; FISCHER, G. W. Next-generation ion torrent sequencing of drug resistance mutations in Mycobacterium tuberculosis strains. **Journal of Clinical Microbiology**, v. 50, n. 12, p. 3831–3837, 2012.

DAVID AVARBOCK; ANDREW AVARBOCK, A.; RUBIN, H. Differential Regulation of Opposing RelMtb Activities by the Aminoacylation State of a tRNA·Ribosome·mRNA·RelMtb Complex. **Biochemistry**, v. 38, p. 11640–11648, 2000.

DE AMORIM, D. B.; CASAGRANDE, R. A.; ALIEVI, M. M.; WOUTERS, F.; DE OLIVEIRA, L. G. S.; DRIEMEIER, D.; TAVARES, M.; IKUTA, C. Y.; TELLES, E. O.;

FERREIRA-NETO, J. S. Mycobacterium pinnipedii in a stranded South American sea lion (Otaria byronia) in BrazilJournal of Wildlife Diseases. **Journal of wildlife diseases**, v. 50, n. 2, p. 409-422, 2014.

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

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

DE LA FUENTE, J.; DIEZ-DELGADO, I.; CONTRERAS, M.; VICENTE, J.; CABEZAS-CRUZ, A.; TOBES, R.; MANRIQUE, M.; LOPEZ, V.; ROMERO, B.; BEZOS, J.; DOMINGUEZ, L.; SEVILLA, I. A.; GARRIDO, J. M.; JUSTE, R.; MADICO, G.; JONES-LOPEZ, E.; GORTAZAR, C. Comparative Genomics of Field Isolates of Mycobacterium bovis and M. caprae Provides Evidence for Possible Correlates with Bacterial Viability and Virulence. **PLOS Neglected Tropical Diseases**, v. 9, n. 11, p. e0004232, 2015.

DEJESUS, M. A.; SACCHETTINI, J. C.; IOERGER, T. R. Reannotation of translational start sites in the genome of Mycobacterium tuberculosis. **Tuberculosis**, v. 93, n. 1, p. 18–25, 2013. DEV, A.; IYER, S.; RAZANI, B.; CHENG, G. NF-κB and innate immunity. **Current Topics in Microbiology and Immunology**, v. 349, p. 115–143, 2011.

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

DORHOI, A.; KAUFMANN, S. H. E. Versatile myeloid cell subsets contribute to tuberculosis-associated inflammation. **European Journal of Immunology**, v. 45, n. 8, p. 2191–2202, 2015.

DOS VULTOS, T.; MESTRE, O.; RAUZIER, J.; GOLEC, M.; RASTOGI, N.; RASOLOFO, V.; TONJUM, T.; SOLA, C.; MATIC, I.; GICQUEL, B. Evolution and diversity of clonal bacteria: The paradigm of Mycobacterium tuberculosis. **PLoS ONE**, v. 3, n. 2, p. 1538, 2008.

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

- ELDHOLM, V.; MONTESERIN, J.; RIEUX, A.; LOPEZ, B.; SOBKOWIAK, B.; RITACCO, V.; BALLOUX, F. Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. **Nature Communications**, v. 6, n. 1, p. 7119, 2015.
- EMMS, D. M.; KELLY, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. **Genome Biology**, v. 16, n. 1, p. 157, 2015.
- ERNST, J. D. The immunological life cycle of tuberculosis. **Nature Reviews Immunology**, v. 12, n. 8, p. 581, 2012.
- ERNST, J. D. Antigenic variation and immune escape in the MTBC. In: **Advances in Experimental Medicine and Biology**, Springer, p. 171–190, 2017.
- FAHLMAN, A.; MADIGAN, J. Respiratory function in voluntary participating patagonia sea lions (Otaria flavescens) in sternal recumbency. **Frontiers in Physiology**, v. 7, p. 528, 2016.
- FAKSRI, K.; XIA, E.; ONG, R. T. H.; TAN, J. H.; NONGHANPHITHAK, D.; MAKHAO, N.; THAMNONGDEE, N.; THANORMCHAT, A.; PHURATTANAKORNKUL, A.; RATTANARANGSEE, S.; RATANAJARAYA, C.; SURIYAPHOL, P.; PRAMMANANAN, T.; TEO, Y. Y.; CHAIPRASERT, A. Comparative whole-genome sequence analysis of Mycobacterium tuberculosis isolated from tuberculous meningitis and pulmonary tuberculosis patients. **Scientific Reports**, v. 8, n. 1, p. 4910, 2018.
- FARHAT, M. R.; SHAPIRO, B. J.; KIESER, K. J.; SULTANA, R.; JACOBSON, K. R.; VICTOR, T. C.; WARREN, R. M.; STREICHER, E. M.; CALVER, A.; SLOUTSKY, A.; KAUR, D.; POSEY, J. E.; PLIKAYTIS, B.; OGGIONI, M. R.; GARDY, J. L.; JOHNSTON, J. C.; RODRIGUES, M.; TANG, P. K. C.; KATO-MAEDA, M.; BOROWSKY, M. L.; MUDDUKRISHNA, B.; KREISWIRTH, B. N.; KUREPINA, N.; GALAGAN, J.; GAGNEUX, S.; BIRREN, B.; RUBIN, E. J.; LANDER, E. S.; SABETI, P. C.; MURRAY, M. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. **Nature Genetics**, v. 45, n. 10, p. 1183–1189, 2013.
- FERRERAS, J. A.; STIRRETT, K. L.; LU, X.; RYU, J. S.; SOLL, C. E.; TAN, D. S.; QUADRI, L. E. E. N. Mycobacterial Phenolic Glycolipid Virulence Factor Biosynthesis: Mechanism and Small-Molecule Inhibition of Polyketide Chain Initiation. **Chemistry and Biology**, v. 15, n. 1, p. 51–61, 2008.
- FILLIOL, I.; MOTIWALA, A. S.; CAVATORE, M.; QI, W.; HAZBÓN, M. H.; BOBADILLA DEL VALLE, M.; FYFE, J.; GARCÍA-GARCÍA, L.; RASTOGI, N.; SOLA, C.; ZOZIO, T.; GUERRERO, M. I.; LEÓN, C. I.; CRABTREE, J.; ANGIUOLI, S.;

EISENACH, K. D.; DURMAZ, R.; JOLOBA, M. L.; RENDÓN, A.; SIFUENTES-OSORNIO, J.; PONCE DE LEÓN, A.; CAVE, M. D.; FLEISCHMANN, R.; WHITTAM, T. S.; ALLAND, D.; HAZBON, M. H.; BOBADILLA DEL VALLE, M.; FYFE, J.; GARCIA-GARCIA, L.; RASTOGI, N.; SOLA, C.; ZOZIO, T.; GUERRERO, M. I.; LEON, C. I.; CRABTREE, J.; ANGIUOLI, S.; EISENACH, K. D.; DURMAZ, R.; JOLOBA, M. L.; RENDON, A.; SIFUENTES-OSORNIO, J.; PONCE DE LEON, A.; CAVE, M. D.; FLEISCHMANN, R.; WHITTAM, T. S.; ALLAND, D. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. **Journal of Bacteriology**, v. 188, n. 2, 2006.

FIRDESSA, R.; BERG, S.; HAILU, E.; SCHELLING, E.; GUMI, B.; ERENSO, G.; GADISA, E.; KIROS, T.; HABTAMU, M.; HUSSEIN, J.; ZINSSTAG, J.; ROBERTSON, B. D.; AMENI, G.; LOHAN, A. J.; LOFTUS, B.; COMAS, I.; GAGNEUX, S.; TSCHOPP, R.; YAMUAH, L.; HEWINSON, G.; GORDON, S. V; YOUNG, D. B.; ASEFFA, A. Mycobacterial lineages causing pulmonary and extrapulmonary Tuberculosis, Ethiopia. **Emerging Infectious Diseases**, v. 19, n. 3, p. 460–463, 2013.

FLEISCHMANN, R. D.; ALLAND, D.; EISEN, J. A.; CARPENTER, L.; WHITE, O.; PETERSON, J.; DEBOY, R.; DODSON, R.; GWINN, M.; HAFT, D.; HICKEY, E.; KOLONAY, J. F.; NELSON, W. C.; UMAYAM, L. A.; ERMOLAEVA, M.; SALZBERG, S. L.; DELCHER, A.; UTTERBACK, T.; WEIDMAN, J.; KHOURI, H.; GILL, J.; MIKULA, A.; BISHAI, W.; JACOBS, W. R.; VENTER, J. C.; FRASER, C. M. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. **Journal of bacteriology**, v. 184, n. 19, p. 5479–90, 2002.

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

FORSHAW, D.; PHELPS, G. R. TUBERCULOSIS IN A CAPTIVE COLONY OF PINNIPEDS. **Journal of Wildlife Diseases**, v. 27, n. 2, p. 288–295, 24 abr. 1991.

GAGNEUX, S. Host—pathogen coevolution in human tuberculosis. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 367, n. 1590, p. 850-859, 2012.

GAGNEUX, S.; DERIEMER, K.; VAN, T.; KATO-MAEDA, M.; DE JONG, B. C.; NARAYANAN, S.; NICOL, M.; NIEMANN, S.; KREMER, K.; GUTIERREZ, M. C.; HILTY, M.; HOPEWELL, P. C.; SMALL, P. M. Variable host—pathogen compatibility in Mycobacterium tuberculosis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 8, p. 2869 LP-2873, 2006.

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

GAO, Q.; KRIPKE, K. E.; SALDANHA, A. J.; YAN, W.; HOLMES, S.; SMALL, P. M. Gene expression diversity among Mycobacterium tuberculosis clinical isolates. **Microbiology**, v. 151, n. 1, p. 5–14, 2005.

GARNIER, T.; EIGLMEIER, K.; CAMUS, J.-C.; MEDINA, N.; MANSOOR, H.; PRYOR, M.; DUTHOY, S.; GRONDIN, S.; LACROIX, C.; MONSEMPE, C.; SIMON, S.; HARRIS, B.; ATKIN, R.; DOGGETT, J.; MAYES, R.; KEATING, L.; WHEELER, P. R.; PARKHILL, J.; BARRELL, B. G.; COLE, S. T.; GORDON, S. V.; HEWINSON, R. G. The complete genome sequence of Mycobacterium bovis. **Proceedings of the National Academy of Sciences**, v. 100, n. 13, p. 7877–7882, 2003.

GARRISON, E.; MARTH, G. Haplotype-based variant detection from short-read sequencing. **arXiv preprint arXiv:1207.3907**, 2012.

GEY VAN PITTIUS, N. C.; SAMPSON, S. L.; LEE, H.; KIM, Y.; VAN HELDEN, P. D.; WARREN, R. M. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. **BMC Evolutionary Biology**, v. 6, n. 1, p. 95, 2006.

GINESTET, C. ggplot2: Elegant Graphics for Data Analysis. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, v. 174, n. 1, p. 245–246, 2011.

GLOVER, R. T.; KRIAKOV, J.; GARFORTH, S. J.; BAUGHN, A. D.; JACOBS, W. R. The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in Mycobacterium smegmatis. **Journal of Bacteriology**, v. 189, n. 15, p. 5495–5503, 2007.

GONZALES-VIERA, O.; MARIGO, J.; RUOPPOLO, V.; ROSAS, F. C. W.; KANAMURA, C. T.; TAKAKURA, C.; FERNÁNDEZ, A.; CATÃO-DIAS, J. L. Toxoplasmosis in a Guiana dolphin (Sotalia guianensis) from Paraná, Brazil. **Veterinary Parasitology**, v. 191, n. 3–4, p. 358–362, 2013.

GONZALO-ASENSIO, J.; MALAGA, W.; PAWLIK, A.; ASTARIE-DEQUEKER, C.; PASSEMAR, C.; MOREAU, F.; LAVAL, F.; DAFFE, M.; MARTIN, C.; BROSCH, R.; GUILHOT, C. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. **Proceedings of the National Academy of Sciences**, v. 111, n. 31, p. 11491–11496, 2014.

GOVENDER, V. S.; RAMSUGIT, S.; PILLAY, M. Mycobacterium tuberculosis adhesins: potential biomarkers as anti-tuberculosis therapeutic and diagnostic targets. **Microbiology**, v. 160, n. Pt_9, p. 1821–1831, 2014.

HERTWECK, C. The biosynthetic logic of polyketide diversity. **Angewandte Chemie International Edition**, v. 48, n. 26, 2009.

HETT, E. C.; RUBIN, E. J. Bacterial growth and cell division: a mycobacterial perspective. **Microbiology and molecular biology reviews : MMBR**, v. 72, n. 1, p. 126–56, table of contents, 2008.

HIGGINS, R. Bacteria and fungi of marine mammals: a review. **The Canadian veterinary journal = La revue veterinaire canadienne**, v. 41, n. 2, p. 105–16, 2000.

HIRSH, A. E.; TSOLAKI, A. G.; DERIEMER, K.; FELDMAN, M. W.; SMALL, P. M. Stable association between strains of Mycobacterium tuberculosis and their human host populations. **Proc. Natl. Acad. Sci.**, v. 101, n. 14, 2004.

HUERTA-CEPAS, J.; SZKLARCZYK, D.; FORSLUND, K.; COOK, H.; HELLER, D.; WALTER, M. C.; RATTEI, T.; MENDE, D. R.; SUNAGAWA, S.; KUHN, M.; JENSEN, L. J.; VON MERING, C.; BORK, P. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic acids research**, v. 44, n. D1, p. D286-93, 2016.

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

ISAZA, J. P.; DUQUE, C.; GOMEZ, V.; ROBLEDO, J.; BARRERA, L. F.; ALZATE, J. F. Whole genome shotgun sequencing of one Colombian clinical isolate of Mycobacterium tuberculosis reveals DosR regulon gene deletions. **FEMS Microbiology Letters**, v. 330, n. 2, p. 113–120, 2012.

JENSEN, L. J.; SIMONOVIC, M.; BORK, P.; VON MERING, C.; MULLER, J.; STARK, M.; DOERKS, T.; ROTH, A.; JULIEN, P.; KUHN, M.; CHAFFRON, S.; CREEVEY, C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. **Nucleic Acids Research**, v. 37, n. Database, p. D412–D416, 2008.

JIA, X.; YANG, L.; DONG, M.; CHEN, S.; LV, L.; CAO, D.; FU, J.; YANG, T.; ZHANG, J.; ZHANG, X.; SHANG, Y.; WANG, G.; SHENG, Y.; HUANG, H.; CHEN, F. The bioinformatics analysis of comparative genomics of Mycobacterium tuberculosis complex (MTBC) provides insight into dissimilarities between intraspecific groups differing in host association, virulence, and epitope diversity. **Frontiers in Cellular and Infection Microbiology**, v. 7, p. 88, 2017.

JURCZYNSKI, K.; LYASHCHENKO, K. P.; GOMIS, D.; MOSER, I.; GREENWALD, R.; MOISSON, P. Pinniped Tuberculosis in Malayan Tapirs (*Tapirus indicus*) and its Transmission to Other Terrestrial Mammals. **Journal of Zoo and Wildlife Medicine**, v. 42, n. 2, p. 222–227, 2011.

JURTZ, V.; PAUL, S.; ANDREATTA, M.; MARCATILI, P.; PETERS, B.; NIELSEN, M. NetMHCpan-4.0: Improved Peptide—MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. **The Journal of Immunology**, v. 199, n. 9, p. 3360–3368, 2017.

KALSCHEUER, R.; WEINRICK, B.; VEERARAGHAVAN, U.; BESRA, G. S.; JACOBS JR, W. R. Trehalose-Recycling ABC Transporter LpqY-SugA-SugB-SugC Is Essential for Virulence of Mycobacterium Tuberculosis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 50, p. 21761–21766, 2010.

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

KAVVAS, E. S.; CATOIU, E.; MIH, N.; YURKOVICH, J. T.; SEIF, Y.; DILLON, N.; HECKMANN, D.; ANAND, A.; YANG, L.; NIZET, V.; MONK, J. M.; PALSSON, B. O. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. **Nature Communications**, v. 9, n. 1, p. 4306, 2018.

KENDALL, S. L.; MOVAHEDZADEH, F.; RISON, S. C. G.; WERNISCH, L.; PARISH, T.; DUNCAN, K.; BETTS, J. C.; STOKER, N. G. The Mycobacterium tuberculosis dosRS two-component system is induced by multiple stresses. **Tuberculosis**, v. 84, n. 3–4, p. 247–255, 2004.

KEREM, D. H.; KYLSTRA, J. A.; SALTZMAN, H. A. Respiratory flow rates in the sea lion. **Undersea Biomedical Research**, v. 2, n. 1, p. 20–27, 1975.

KIERS, A.; KLARENBEEK, A.; MENDELTS, B.; VAN SOOLINGEN, D.; KOËTER, G. Transmission of Mycobacterium pinnipedii to humans in a zoo with marine mammals. **The**

International Journal of Tuberculosis and Lung Disease, v. 12, p. 1469–1473, 2008.

KIESER, K. J.; RUBIN, E. J. How sisters grow apart: Mycobacterial growth and division. **Nature Reviews Microbiology**, v. 12, n. 8, 2014.

KRIZ, P.; KRALIK, P.; SLANY, M.; SLANA, I.; SVOBODOVA, J.; PARMOVA, I.; BARNET, V.; JUREK, V.; PAVLIK, I. Case Report Mycobacterium pinnipedii in a captive Southern sea lion (Otaria flavescens): a case report Private veterinary practitioner, Czech Republic. **Veterinarni Medicina**, v. 56, n. 6, p. 307–313, 2011.

KUMAR, A.; DESHANE, J. S.; CROSSMAN, D. K.; BOLISETTY, S.; YAN, B.-S.; KRAMNIK, I.; AGARWAL, A.; STEYN, A. J. C. Heme Oxygenase-1-derived Carbon Monoxide Induces the Mycobacterium tuberculosis Dormancy Regulon. **The Journal of biological Chemistry**, v. 283, p. 18032–18039, 2008.

LERNER, T. R.; BOREL, S.; GUTIERREZ, M. G. The innate immune response in human tuberculosis. **Cellular Microbiology**, v. 17, n. 9, p. 1277–1285, 2015.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754–1760, 2009.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R.; 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.

LI, L.; STOECKERT, C. J.; ROOS, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. **Genome Research**, v. 13, n. 9, p. 2178–2189, 2003.

LIU, B.; ZHENG, D.; JIN, Q.; CHEN, L.; YANG, J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. **Nucleic Acids Research**, v. 47, n. D1, p. D687–D692, 2019.

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

LOEFFLER, S. H.; DE LISLE, G. W.; NEILL, M. A.; COLLINS, D. M.; PRICE-CARTER, M.; PATERSON, B.; CREWS, K. B. The seal tuberculosis agent, Mycobacterium pinnipedii, infects domestic cattle in New Zealand: epidemiologic factors and DNA strain typing.

Journal of Wildlife Diseases, v. 50, n. 2, p. 180–187, 2014.

LUO, T.; YANG, C.; PENG, Y.; LU, L.; SUN, G.; WU, J.; JIN, X.; HONG, J.; LI, F.; MEI, J.; DERIEMER, K.; GAO, Q. Whole-genome sequencing to detect recent transmission of Mycobacterium tuberculosis in settings with a high burden of tuberculosis. **Tuberculosis**, v. 94, n. 4, p. 434–440, 2014.

MARJANOVIC, O.; IAVARONE, A. T.; RILEY, L. W. Sulfolipid accumulation in Mycobacterium tuberculosis disrupted in the mce2 operon. **Journal of Microbiology**, v. 49, n. 3, p. 441–447, 2011.

MENARDO, F.; DUCHENE, S.; BRITES, D.; GAGNEUX, S. The molecular clock of Mycobacterium tuberculosis. **bioRxiv**, p. 532390, 2019.

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

NAMOUCHI, A.; DIDELOT, X.; SCHOCK, U.; GICQUEL, B.; ROCHA, E. P. C. After the bottleneck: Genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. **Genome Research**, v. 22, n. 4, p. 721–734, 2012.

NYKA, W. Studies on the effect of starvation on mycobacteria. **Infection and Immunity**, v. 9, n. 5, p. 843–850, 1974.

O'NEILL, M. B.; SHOCKEY, A.; ZARLEY, A.; AYLWARD, W.; ELDHOLM, V.; KITCHEN, A.; PEPPERELL, C. S. Lineage specific histories of Mycobacterium tuberculosis dispersal in Africa and Eurasia. **Molecular Ecology**, n. February, p. 1–16, 2019.

OJHA, A. K.; MUKHERJEE, T. K.; CHATTERJI, D. High intracellular level of guanosine tetraphosphate in Mycobacterium smegmatis changes the morphology of the bacterium. **Infection and Immunity**, v. 68, n. 7, p. 4084–4091, 2000.

OLSEN, C. R.; HALE, F. C.; ELSNER, R. Mechanics of ventilation in the pilot whale. **Respiration Physiology**, v. 7, n. 2, p. 137–149, 1969.

PANDEY, A. K.; SASSETTI, C. M. Mycobacterial persistence requires the utilization of host cholesterol. **Proceedings of the National Academy of Sciences**, v. 105, n. 11, p. 4376–4380,

2008.

PARK, H.-D.; GUINN, K. M.; HARRELL, M. I.; LIAO, R.; VOSKUIL, M. I.; TOMPA, M.; SCHOOLNIK, G. K.; SHERMAN, D. R. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. **Molecular Microbiology**, v. 48, n. 3, p. 833–843, 2003.

PARRA, A.; FERNÁNDEZ-LLARIO, P.; TATO, A.; LARRASA, J.; GARCÍA, A.; ALONSO, J. .; HERMOSO DE MENDOZA, M.; HERMOSO DE MENDOZA, J. Epidemiology of Mycobacterium bovis infections of pigs and wild boars using a molecular approach. **Veterinary Microbiology**, v. 97, n. 1–2, p. 123–133, 2003.

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

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

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

PIROFSKI, L. A.; CASADEVALL, A. Q and A What is a pathogen? A question that begs the point. **BMC Biology**, v. 10, p. 6, 2012.

PŁUCIENNIK, A.; STOLARCZYK, M.; BZÓWKA, M.; RACZYŃSKA, A.; MAGDZIARZ, T.; GÓRA, A. BALCONY: an R package for MSA and functional compartments of protein variability analysis. **BMC Bioinformatics**, v. 19, n. 1, p. 300, 2018.

PORTEVIN, D.; GAGNEUX, S.; COMAS, I.; YOUNG, D. Human Macrophage Responses to Clinical Isolates from the Mycobacterium tuberculosis Complex Discriminate between

Ancient and Modern Lineages. **PLoS Pathogens**, v. 7, n. 3, p. e1001307, 2011.

PORTEVIN, D.; SUKUMAR, S.; COSCOLLA, M.; SHUI, G.; LI, B.; GUAN, X. L.; BENDT, A. K.; YOUNG, D.; GAGNEUX, S.; WENK, M. R. Lipidomics and genomics of Mycobacterium tuberculosis reveal lineage-specific trends in mycolic acid biosynthesis. **MicrobiologyOpen**, v. 3, n. 6, p. 823–35, 2014.

PRUITT, K. D.; TATUSOVA, T.; MAGLOTT, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. **Nucleic acids research**, v. 35, n. Database issue, p. D61-5, 2007.

QUADRI, L. E. N. Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. **Critical Reviews in Biochemistry and Molecular Biology,** v. 49, n. 3, p. 179-211, 2014.

R CORE TEAM. R: A language and environment for statistical computing. 2013.

RAHMAN, S. A.; SINGH, Y.; KOHLI, S.; AHMAD, J.; EHTESHAM, N. Z.; TYAGI, A. K.; HASNAIN, S. E. Comparative analyses of nonpathogenic, opportunistic, and totally pathogenic mycobacteria reveal genomic and biochemical variabilities and highlight the survival attributes of Mycobacterium tuberculosis. **mBio**, v. 5, n. 6, p. e02020-14, 2014.

RAMAGE, H. R.; CONNOLLY, L. E.; COX, J. S. Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. **PLoS Genetics**, v. 5, n. 12, p. e1000767, 2009.

RAMAKRISHNAN, L.; FEDERSPIEL, N. A.; FALKOW, S. Granuloma-Specific Expression of Mycobacterium Virulence Proteins from the Glycine-Rich PE-PGRS Family. **Science**, v. 288, n. 5470, p. 1436–1439, 2000.

REGMI, S. M.; CHAIPRASERT, A.; KULAWONGANUNCHAI, S.; TONGSIMA, S.; COKER, O. O.; PRAMMANANAN, T.; VIRATYOSIN, W.; THAIPISUTTIKUL, I. Whole genome sequence analysis of multidrug-resistant Mycobacterium tuberculosis Beijing isolates from an outbreak in Thailand. **Molecular Genetics and Genomics**, v. 290, n. 5, p. 1933–1941, 2015.

RENWICK, A. R.; WHITE, P. C. L.; BENGIS, R. G. Bovine tuberculosis in southern African wildlife: a multi-species host–pathogen system. **Epidemiology and Infection**, v. 135, n. 04, p. 529, 2007.

RIOJAS, M. A.; MCGOUGH, K. J.; RIDER-RIOJAS, C. J.; RASTOGI, N.; HAZBÓN, M. H.

Phylogenomic analysis of the species of the Mycobacterium tuberculosis complex demonstrates that Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of Mycob. **International Journal of Systematic and Evolutionary Microbiology**, v. 68, n. 1, p. 324–332, 2017.

RISSMAN, A. I.; MAU, B.; BIEHL, B. S.; DARLING, A. E.; GLASNER, J. D.; PERNA, N. T. Reordering contigs of draft genomes using the Mauve Aligner. **Bioinformatics**, v. 25, n. 16, p. 2071–2073, 2009.

RODRÍGUEZ-CASTILLO, J. G.; PINO, C.; NIÑO, L. F.; ROZO, J. C.; LLERENA-POLO, C.; PARRA-LÓPEZ, C. A.; TAUCH, A.; MURCIA-ARANGUREN, M. I. Comparative genomic analysis of Mycobacterium tuberculosis Beijing-like strains revealed specific genetic variations associated with virulence and drug resistance. **Infection, Genetics and Evolution**, v. 54, p. 314–323, 2017.

RUOPPOLO, V. **Patologia comparada de cetáceos e pinípedes.** [Comparative pathology of cetaceans and pinnipeds]. 2003. 131 f. Dissertação (Mestrado em Patologia Experimental e Comparada) - Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2003.

SABIN, S.; HERBIG, A.; V\R AGENE, \r Ashild J; AHLSTRÖM, T.; BOZOVIC, G.; ARCINI, C.; KÜHNERT, D.; BOS, K. I. A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex. **bioRxiv**, 2019.

SALAMON, H.; KATO-MAEDA, M.; SMALL, P. M.; DRENKOW, J.; GINGERAS, T. R. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. **Genome research**, v. 10, n. 12, p. 2044–54, 2000.

SASSETTI, C. M.; RUBIN, E. J. Genetic requirements for mycobacterial survival during infection. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 22, p. 12989–12994, 2003.

SCHNAPPINGER, D.; EHRT, S.; VOSKUIL, M. I.; LIU, Y.; MANGAN, J. A.; MONAHAN, I. M.; DOLGANOV, G.; EFRON, B.; BUTCHER, P. D.; NATHAN, C.; SCHOOLNIK, G. K. Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: Insights into the phagosomal environment. **Journal of Experimental Medicine**, v. 198, n. 5, p. 693–704, 2003.

SHERMAN, D. R.; VOSKUIL, M.; SCHNAPPINGER, D.; LIAO, R.; HARRELL, M. I.; SCHOOLNIK, G. K. Regulation of the mycobacterium tuberculosis hypoxic response gene

encoding α -crystallin. **Proceedings of the National Academy of Sciences**, v. 98, n. 13, p. 7534–7539, 2001.

SINGH, A.; JAIN, S.; GUPTA, S.; DAS, T.; TYAGI, A. K. mymA operon of Mycobacterium tuberculosis: its regulation and importance in the cell envelope. **FEMS Microbiology Letters**, v. 227, n. 1, p. 53–63, 2003.

SMITH, N. H.; GORDON, S. V.; DE LA RUA-DOMENECH, R.; CLIFTON-HADLEY, R. S.; HEWINSON, R. G. Bottlenecks and broomsticks: The molecular evolution of Mycobacterium bovis. **Nature Reviews Microbiolog**, v. 4, n. 9, p. 670, 2006.

SOUSA, E. H. S.; GONZALEZ, G.; GILLES-GONZALEZ, M.-A. Target DNA stabilizes Mycobacterium tuberculosis DevR/DosR phosphorylation by the full-length oxygen sensors DevS/DosS and DosT. **The FEBS Journal**, v. 284, n. 22, p. 3954–3967, 2017.

STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, v. 30, n. 9, p. 1312–1313, 2014.

SU, H.; ZHANG, Z.; LIU, Z.; PENG, B.; KONG, C.; WANG, H.; ZHANG, Z.; XU, Y. Mycobacterium tuberculosis PPE60 antigen drives Th1/Th17 responses via Toll-like receptor 2-dependent maturation of dendritic cells. **The Journal of biological chemistry**, v. 293, n. 26, p. 10287–10302, 2018.

SUPPLY, P.; ALLIX, C.; LESJEAN, S.; CARDOSO-OELEMANN, M.; RUSCH-GERDES, S.; WILLERY, E.; SAVINE, E.; DE HAAS, P.; VAN DEUTEKOM, H.; RORING, S.; BIFANI, P.; KUREPINA, N.; KREISWIRTH, B.; SOLA, C.; RASTOGI, N.; VATIN, V.; GUTIERREZ, M. C.; FAUVILLE, M.; NIEMANN, S.; SKUCE, R.; KREMER, K.; LOCHT, C.; VAN SOOLINGEN, D. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. **Journal of Clinical Microbiology**, v. 44, n. 12, p. 4498–4510, 2006.

SUPPLY, P.; MARCEAU, M.; MANGENOT, S.; ROCHE, D.; ROUANET, C.; KHANNA, V.; MAJLESSI, L.; CRISCUOLO, A.; TAP, J.; PAWLIK, A.; FIETTE, L.; ORGEUR, M.; FABRE, M.; PARMENTIER, C.; FRIGUI, W.; SIMEONE, R.; BORITSCH, E. C.; DEBRIE, A.-S.; WILLERY, E.; WALKER, D.; QUAIL, M. A.; MA, L.; BOUCHIER, C.; SALVIGNOL, G.; SAYES, F.; CASCIOFERRO, A.; SEEMANN, T.; BARBE, V.; LOCHT, C.; GUTIERREZ, M.-C.; LECLERC, C.; BENTLEY, S. D.; STINEAR, T. P.; BRISSE, S.; MÉDIGUE, C.; PARKHILL, J.; CRUVEILLER, S.; BROSCH, R. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. **Nature Genetics**, v. 45, n. 2, p. 172–179, 2013.

SWARBRICK, C. M. D.; BYTHROW, G. V.; ARAGAO, D.; GERMAIN, G. A.; QUADRI, L. E. N.; FORWOOD, J. K. Mycobacteria Encode Active and Inactive Classes of TesB Fatty-Acyl CoA Thioesterases Revealed through Structural and Functional Analysis. **Biochemistry**, v. 56, n. 10, p. 1460–1472, 2017.

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

THOREL, M. F. Mycobacteria identified in a centre for veterinary research between 1973 and 1979 (author's transl). **Annales de microbiologie**, v. 131, n. 1, p. 61–9, 1980a. THOREL, M. F. Isolation of mycobacterium africanum from monkeys. **Tubercle**, v. 61, n. 2, p. 101–104, 1980b.

THOREL, M. F.; KAROUI, C.; VARNEROT, A.; FLEURY, C.; VINCENT, V. Isolation of Mycobacterium bovis from baboons, leopards and a sea-lion. **Veterinary research**, v. 29, n. 2, p. 207–12, 1998.

TORRELLES, J. B.; KNAUP, R.; KOLARETH, A.; SLEPUSHKINA, T.; KAUFMAN, T. M.; KANG, P.; HILL, P. J.; BRENNAN, P. J.; CHATTERJEE, D.; BELISLE, J. T.; MUSSER, J. M.; SCHLESINGER, L. S. Identification of Mycobacterium tuberculosis clinical isolates with altered phagocytosis by human macrophages due to a truncated lipoarabinomannan. **Journal of Biological Chemistry**, v. 283, n. 46, p. 31417–31428, 2008.

TORRELLES, J. B.; SCHLESINGER, L. S. Diversity in Mycobacterium tuberculosis mannosylated cell wall determinants impacts adaptation to the host. **Tuberculosis**, v. 90, n. 2, p. 84-93, 2010.

TRAORE, B.; DIARRA, B.; DEMBELE, B. P. P.; SOMBORO, A. M.; HAMMOND, A. S.; SIDDIQUI, S.; MAIGA, M.; KONE, B.; SARRO, Y. S.; WASHINGTON, J.; PARTA, M.; COULIBALY, N.; M'BAYE, O.; DIALLO, S.; KOITA, O.; TOUNKARA, A.; POLIS, M. A. Molecular strain typing of Mycobacterium tuberculosis complex in Bamako, Mali. **International Journal of Tuberculosis and Lung Disease**, v. 16, n. 7, p. 911–916, 2012.

UPLEKAR, S.; HEYM, B.; FRIOCOURT, V.; ROUGEMONT, J.; COLE, S. T. Comparative Genomics of esx Genes from Clinical Isolates of Mycobacterium tuberculosis Provides Evidence for Gene Conversion and Epitope Variation †. **Infection and immunity**, v. 79, n. 10, p. 4042–4049, 2011.

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

VANDER BEKEN, S.; AL DULAYYMI, J. R.; NAESSENS, T.; KOZA, G.; MAZA-IGLESIAS, M.; ROWLES, R.; THEUNISSEN, C.; DE MEDTS, J.; LANCKACKER, E.; BAIRD, M. S.; GROOTEN, J. Molecular structure of the Mycobacterium tuberculosis virulence factor, mycolic acid, determines the elicited inflammatory pattern. **European Journal of Immunology**, v. 41, n. 2, p. 450–460, 2011.

VERNIKOS, G.; MEDINI, D.; RILEY, D. R.; TETTELIN, H. Ten years of pan-genome analyses. **Current Opinion in Microbiology**, v. 23, p. 148–154, 2015.

VITA, R.; MAHAJAN, S.; OVERTON, J. A.; DHANDA, S. K.; MARTINI, S.; CANTRELL, J. R.; WHEELER, D. K.; SETTE, A.; PETERS, B. The Immune Epitope Database (IEDB): 2018 update. **Nucleic Acids Research**, v. 47, n. D1, p. D339–D343, 2019.

VOSKUIL, M. I.; SCHNAPPINGER, D.; VISCONTI, K. C.; HARRELL, M. I.; DOLGANOV, G. M.; SHERMAN, D. R.; SCHOOLNIK, G. K. Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program. **The Journal of experimental medicine**, v. 198, n. 5, p. 705–13, 2003.

WALKER, T. M.; IP, C. L.; HARRELL, R. H.; EVANS, J. T.; KAPATAI, G.; DEDICOAT, M. J.; EYRE, D. W.; WILSON, D. J.; HAWKEY, P. M.; CROOK, D. W.; PARKHILL, J.; HARRIS, D.; WALKER, A. S.; BOWDEN, R.; MONK, P.; SMITH, E. G.; PETO, T. E. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. **The Lancet Infectious Diseases**, v. 13, n. 2, p. 137–146, 2013.

WANG, J.; LI, B. X.; GE, P. P.; LI, J.; WANG, Q.; GAO, G. F.; QIU, X. B.; LIU, C. H. Mycobacterium tuberculosis suppresses innate immunity by coopting the host ubiquitin system. **Nature Immunology**, v. 16, n. 3, p. 237–245, 2015.

WARNES, G. R.; BOLKER, B.; BONEBAKKER, L.; GENTLEMAN, R.; LIAW, W. H. A.; LUMLEY, T.; MAECHLER, M.; MAGNUSSON, A.; MOELLER, S.; SCHWARTZ, M.; VENABLES, B. gplots: Various R Programming Tools for Plotting Data. R package version 3.0-1. **The Comprehensive R Archive Network**, 2016.

WATTAM, A. R.; DAVIS, J. J.; ASSAF, R.; BOISVERT, S.; BRETTIN, T.; BUN, C.; CONRAD, N.; DIETRICH, E. M.; DISZ, T.; GABBARD, J. L.; GERDES, S.; HENRY, C. S.; KENYON, R. W.; MACHI, D.; MAO, C.; NORDBERG, E. K.; OLSEN, G. J.; MURPHY-OLSON, D. E.; OLSON, R.; OVERBEEK, R.; PARRELLO, B.; PUSCH, G. D.; SHUKLA, M.; VONSTEIN, V.; WARREN, A.; XIA, F.; YOO, H.; STEVENS, R. L. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. **Nucleic Acids Research**, v. 45, n. D1, p. D535–D542, 2017.

WAYNE, L. G.; HAYES, L. G. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. **Infection and immunity**, v. 64, n. 6, p. 2062–9, 1996.

WELLS, A. Q. Tuberculosis In Wild Voles. The Lancet, v. 229, n. 5934, p. 1221, 1937.

WENIGER, T.; KRAWCZYK, J.; SUPPLY, P.; NIEMANN, S.; HARMSEN, D. MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. **Nucleic Acids Research**, v. 38, n. Web Server, p. W326–W331, 1 jul. 2010. WINGLEE, K.; MANSON MCGUIRE, A.; MAIGA, M.; ABEEL, T.; SHEA, T.; DESJARDINS, C. A.; DIARRA, B.; BAYA, B.; SANOGO, M.; DIALLO, S.; EARL, A. M.; BISHAI, W. R. Whole Genome Sequencing of Mycobacterium africanum Strains from Mali Provides Insights into the Mechanisms of Geographic Restriction. **PLOS Neglected Tropical Diseases**, v. 10, n. 1, p. e0004332, 2016.

WINTHER, K.; TREE, J. J.; TOLLERVEY, D.; GERDES, K. VapCs of Mycobacterium tuberculosis cleave RNAs essential for translation. **Nucleic Acids Research**, v. 44, n. 20, p. 9860–9871, 2016.

WOOLHOUSE, M. E. J.; WEBSTER, J. P.; DOMINGO, E.; CHARLESWORTH, B.; LEVIN, B. R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. **Nature genetics**, v. 32, n. 4, p. 569–577, 2002.

WORBOYS, M. Thomas D. Brock, Robert Koch: a life in medicine and bacteriology, Berlin, Springer, 1988, 8vo, pp. ix, 364, illus., DM 48.00 (N. American distributor: Science Tech Publishers, Madison WI). **Medical History**, v. 34, n. 3, p. 347–348, 1990.

WORLD HEALTH ORGANIZATION. **WHO Tuberculosis Report 2018**. Geneva: World Health Organization: World Health Organization, 2018.

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

XIONG, X.; WANG, R.; DENG, D.; CHEN, Y.; LIU, H.; WANG, T.; WANG, J.; ZHU, X.; ZHU, X.; ZHU, Y.; LU, X.; CHEN, H.; ZHENG, H.; GUO, A. Comparative Genomics of a Bovine Mycobacterium tuberculosis Isolate and Other Strains Reveals Its Potential Mechanism of Bovine Adaptation. **Frontiers in Microbiology**, v. 8, p. 2500, 2017.

YANG, T.; SHENG, Y.; YUE, L.; DING, N.; WANG, G.; ZHONG, J.; XIAO, J.; YU, X.; LI, C.; CHEN, F.; ZHANG, J.; SUN, Y.; LIANG, Q.; JIA, X.; HUANG, H.; MA, G. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes,

Generality/Individuality, and the Interconversion Through Copy Number Variations. **Frontiers in Microbiology**, v. 9, n. August, p. 1–12, 2018.

YAR, A. M.; ZAMAN, G.; HUSSAIN, A.; CHANGHUI, Y.; RASUL, A.; HUSSAIN, A.; BO, Z.; BOKHARI, H.; IBRAHIM, M. Comparative Genome Analysis of 2 Mycobacterium Tuberculosis Strains from Pakistan: Insights Globally Into Drug Resistance, Virulence, and Niche Adaptation. **Evolutionary Bioinformatics**, v. 14, p. 1176934318790252, 2018. ZHANG, F.; XIE, J.-P. Mammalian cell entry gene family of Mycobacterium tuberculosis. **Molecular and Cellular Biochemistry**, v. 352, n. 1–2, p. 1–10, 2011.

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

ZIMPEL, C. K.; BRANDÃO, P. E.; DE SOUZA FILHO, A. F.; DE SOUZA, R. F.; IKUTA, C. Y.; NETO, J. S. F.; SOLER CAMARGO, N. C.; HEINEMANN, M. B.; GUIMARÃES, A. M. S. Complete genome sequencing of Mycobacterium bovis SP38 and comparative genomics of Mycobacterium bovis and M. tuberculosis strains. **Frontiers in Microbiology**, v. 8, n. DEC, p. 2389, dez. 2017.

ZINK, A.; HAAS, C. J.; REISCHL, U.; SZEIMIES, U.; NERLICH, A. G. Molecular analysis of skeletal tuberculosis in an ancient egyptian population. **Journal of Medical Microbiology**, v. 50, n. 4, p. 355–366, 2001.

ZINK, A. R.; GRABNER, W.; REISCHL, U.; WOLF, H.; NERLICH, A. G. Molecular study on human tuberculosis in three geographically distinct and time delineated populations from ancient Egypt. **Epidemiology and Infection**, v. 130, n. 2, p. 239–249, 2003a.

ZINK, A. R.; MOLNÁR, E.; MOTAMEDI, N.; PÁLFY, G.; MARCSIK, A.; NERLICH, A. G. Molecular history of tuberculosis from ancient mummies and skeletons. **International Journal of Osteoarchaeology**, v. 17, n. 4, p. 380-391, 2007.

ZINK, A. R.; SOLA, C.; REISCHL, U.; GRABNER, W.; RASTOGI, N.; WOLF, H.; NERLICH, A. G. Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. **Journal of Clinical Microbiology**, v. 41, n. 1, p. 359–367, 2003b.

APÊNDICES

APÊNDICE A – SNPs que diferenciam as estirpes MP1 e MP2 de M. pinnipedii

				(continua)
Ref Var		Type	Function	COGs
GA	TG	Nonsyn	[Protein-PII] uridylyltransferase (EC 2.7.7.59) / [Protein-PII]-UMP uridylyl-removing enzyme	Е
AGCT	AT	Deletion	13E12 repeat family protein	S
TGGGGGA	GGGGGA	Nonsyn	3-oxoacyl-[acyl-carrier-protein] synthase, KASIII (EC 2.3.1.180)	I
ACC	AA	Deletion	4-alpha-glucanotransferase (amylomaltase) (EC 2.4.1.25)	Q
TG	GC	Nonsyn	5'-methylthioadenosine phosphorylase (EC 2.4.2.28)	E
CGGTG	GGGTG	Nonsyn	ABC transporter, substrate-binding protein (cluster 12, methionine/phosphonates)	V
CTC	TTC	Synon	Aconitate hydratase (EC 4.2.1.3) @ 2-methylisocitrate dehydratase (EC 4.2.1.99)	C, E
TTCGTCAGTGG CGACGTCGTCG T	TT	Deletion	Aconitate hydratase (EC 4.2.1.3) @ 2-methylisocitrate dehydratase (EC 4.2.1.99)	C, E
ACGATCGCGGC	AC	Deletion	Acyl-[acyl-carrier protein] desaturase DesA2 (EC 1.14.19.2)	I
GGGCCGGTGGC CAGGCCCACCC G	GG	Deletion	Adenylate cyclase (EC 4.6.1.1)	T
ATG	GGT	Nonsyn	Alcohol dehydrogenase (EC 1.1.1.1)	E, R
TG	GG	Nonsyn	Aliphatic amidase AmiE (EC 3.5.1.4)	C
CA	AA	Synon	Alkane-1 monooxygenase (EC 1.14.15.3)	I
CA	AC	Nonsyn	Alkyl sulfatase and related hydrolases, MBL-fold metallo-hydrolase superfamily	I
CGGGGGC	CTGGGGC	Synon	Alkyldihydroxyacetonephosphate synthase (EC 2.5.1.26)	L
CGGGGGT	CGGGGGG	Synon	Alpha-glucosidase (EC 3.2.1.20)	G
AT	GG	Nonsyn	Amidohydrolase	Q

				(continua)
Ref	Var	Туре	Function	COGs
TCCC	TC	Deletion	Amidohydrolase EgtC (hercynylcysteine sulfoxide synthase)	Q
CTC	CC	Deletion	Arylamine N-acetyltransferase (EC 2.3.1.5)	K
GGCC	GC	Deletion	ATP-dependent DNA ligase (EC 6.5.1.1) LigC	D
TG	GT	Nonsyn	Branched-chain acyl-CoA dehydrogenase (EC 1.3.99.12) @ Acyl-CoA dehydrogenase (EC 1.3.8.1), Mycobacterial subgroup FadE9	I
AT	CT	Nonsyn	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	I
GT	AC	Nonsyn	Cell division trigger factor (EC 5.2.1.8)	D
AT	TT	Nonsyn	COG1180: Radical SAM, Pyruvate-formate lyase-activating enzyme like	C
CA	AC	AC Nonsyn Cyclopropane-fatty-acyl-phospholipid synthase (EC 2.1.1.79		I
AT	TT	Nonsyn	Cyclopropane-fatty-acyl-phospholipid synthase PcaA (EC 2.1.1.79)	
CT	TG	Nonsyn	Cytochrome P450 143	
GGACCACCGTC GAGGTGCT	GT	Deletion	Cytosol aminopeptidase PepA (EC 3.4.11.1)	C
ATCG	ACC	Deletion	DEAD-box ATP-dependent RNA helicase DeaD (K
TCA	TAA	Synon	Dipeptide transport system permease protein DppB (TC 3.A.1.5.2)	C
CAA	AAA	Nonsyn	Dipeptide transport system permease protein DppC (TC 3.A.1.5.2)	C
TAG	CAG	Nonsyn	DNA polymerase III epsilon subunit-related protein MSMEG4261	L
GGGCGTCGGTC A	GA	Deletion	DNA-binding protein	L
CT	AG	Nonsyn	DNA-dependent DNA polymerase beta chain	L
GCGCGAGTCGC TGGCCGAC	GC	Deletion	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	M
GCT	GT	Deletion	Efflux ABC transporter, permease/ATP-binding protein SCO2464	V
CA	GC	Synon	Efflux transmembrane permease Rv1747 family	

				(continua)
Ref Var		Type	Function	COGs
CA	GC	Nonsyn	Enoyl-CoA hydratase (EC 4.2.1.17)	I
CA	AA	Synon	Esterase LipL	I
GGTCGTCGTCG	GGTCGTTG TCG	Nonsyn	Exodeoxyribonuclease V gamma chain (EC 3.1.11.5)	S
GCGCGATCGGG TC	GC	Deletion	F420-dependent oxidoreductase	Н
TGGGGGC	GGGGGGC	Nonsyn	Ferredoxin reductase	C
ATCGACGCGAT T	AT	Deletion	Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27)	G
GTTG	GTGG	Nonsyn	Glutamine synthetase-like protein	E
CCAACACGTC	CC	Deletion	Glycerol kinase (EC 2.7.1.30)	C
GG	AA	Nonsyn	Glycerol kinase (EC 2.7.1.30)	C
ACCCCCCCG	ACCCCCCC G	Deletion	Glycerol kinase (EC 2.7.1.30)	C
CGGGGGGC	CGGTGGGC	Nonsyn	Heteromeric decaprenylphosphoryl-beta-D-ribose 2'-epimerase (EC 1.1.98.3) subunit DprE1	C
ACCCCCA	ACCCCCG	Nonsyn	Hydrogen peroxide-inducible genes activator	C
AC	TT	Nonsyn	hypothetical protein	S
GAG	GG	Deletion	hypothetical protein	S
TG	CT	Nonsyn	hypothetical protein	S
GT	CG	Nonsyn	hypothetical protein	S
ACCCCCCT	ACCCTCCT	Nonsyn	HYPOTHETICAL PROTEIN	S
TGA	GGA	Nonsyn	hypothetical protein	S
TGC	TC	Deletion	hypothetical protein	
TGGGGGGT	TGGGGGG G	Nonsyn	hypothetical protein	S

				(continua)		
Ref	Var	Type	Function	COGs		
TG	GG	Nonsyn	hypothetical protein	S		
AT	GA	Nonsyn	hypothetical protein	S		
TTTG	CCGA	Nonsyn	hypothetical protein	S		
AGC	AC	Deletion	hypothetical protein	S		
CG	AT	Nonsyn	hypothetical protein	S		
GGTG	GG	Deletion	hypothetical protein	S		
CTT	CCT	Nonsyn	hypothetical protein	S		
TCT	TCC	Nonsyn	hypothetical protein	S		
TG	CA	Nonsyn	hypothetical protein Rv1516c	S		
TGCGCGCG	TGCGTG	Deletion	hypothetical protein Rv2237	S		
CA	AC	Nonsyn	Iron utilization protein	K		
GAGCGCCAAGC GCCAA	GAGCGCCA A	Deletion	Isocitrate lyase (EC 4.1.3.1), group III, Mycobacterial type ICL2	C		
TG	GC	Nonsyn	L-cysteine:1D-myo-inosityl 2-amino-2-deoxy-alpha-D-glucopyranoside ligase MshC (EC 6.3.1.13)	S		
TCGGCGGCGC	GCGGCGGC GGC	Nonsyn	Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5)	S		
CGA	CAA	Synon	LSU ribosomal protein L11p (L12e)	J		
CA	AC	Nonsyn	Lysophospholipase (EC 3.1.1.5)	I		
CC	AC	Synon	Methoxy mycolic acid synthase MmaA2 (EC 2.1.1.79)	Н		
GA	TG	Nonsyn	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit (EC 6.4.1.4)	Q		
GT	CG	Nonsyn	methyltransferase, FkbM family domain protein			
CA	AG	Nonsyn	MFS-type transporter Rv2333c family	P		
TCC	TAC	Nonsyn	Mg/Co/Ni transporter MgtE, CBS domain-containing			

				(continua)		
Ref	Var	Type	Function	COGs		
CGA	CC	Deletion	Mobile element protein	X		
GA	CC	Nonsyn	Mobile element protein	X		
TGCG	GGCG	Nonsyn	Mobile element protein	X		
ACC	CCC	Nonsyn	Mobile element protein	X		
CGG	GGG	Synon	Mobile element protein	X		
GC	CG	Nonsyn	Mobile element protein	X		
GCT	GCC	Synon	Mobile element protein	X		
GTG	GTT	Nonsyn	Mobile element protein	X		
CGC	CGG	Synon	Mobile element protein	X		
AC	CT	Nonsyn	Mobile element protein	X		
TG	GG	Nonsyn	Mobile element protein	X		
AG	CA	Nonsyn	Modular polyketide synthase	Q		
CTC	CC	Deletion	Modular polyketide synthase	Q		
CGGCCGCTGCG AGAGGCCGCTG CGAGAG	CG	Deletion	Molybdopterin-binding oxidoreductase, [4Fe-4S] cluster binding and molybdopterin cofactor binding site containing	C		
ACC	GCC	Nonsyn	Monooxygenase EthA, implicated in activation of thiocarbamide- containing pro-drugs	C		
AC	GG	Nonsyn	NAD kinase (EC 2.7.1.23)	M		
GCCTCGGCGCT G	GA	Deletion	NAD synthetase (EC 6.3.1.5) / Glutamine amidotransferase chain of NAD synthetase	M		
ACCCCCA	CCCCCA	Nonsyn	NAD(P)H quinone reductase LpdA (EC 1.6.5.2)	Н		
TGGGGGA	TGGGGGA	Deletion	patatin-related protein			
TGGGGGGT	GGGGGG T	Synon	Periplasmic serine protease			

				(continua)
Ref	Var	Type	Function	COGs
AT	CC	Nonsyn	Phenolpthiocerol synthesis type-I polyketide synthase PpsA (EC 2.3.1.41)	J
ACC	CCC	Synon	Phenolpthiocerol synthesis type-I polyketide synthase PpsB (EC 2.3.1.41)	J
AGC	CCA	Nonsyn	Phospholipase C (EC 3.1.4.3)	I
AT	CA	Nonsyn	PKS-associated protein PapA, condensation domain-containing	S
CT	TG	Nonsyn	Polyketide synthase / Multifunctional mycocerosic acid synthase membrane-associated mas	Q
CA	TT	Nonsyn	Possible protease IV SppA (endopeptidase IV) (signal peptide peptidase)	S
CA	GA	Nonsyn	Possible secreted alanine rich protein	S
CGG	CGT	Nonsyn	Potassium-transporting ATPase C chain (EC 3.6.3.12) (TC 3.A.3.7.1)	
GGTG	GG	Deletion	Probable acetyl-CoA acetyltransferase FadA6 (EC 2.3.1.9)	
GA	CA	Nonsyn	Probable ATP-dependent helicase lhr (EC 3.6.1)	R
CA	AC	Nonsyn	Probable carboxylesterase LipT (EC 3.1.1)	C
AT	TG	Nonsyn	Probable carboxylesterase LipT (EC 3.1.1)	C
TGGTCGCGTGC AGCGGCGG	TG	Deletion	Probable forkhead-associated protein	E
AGCGCCCCCGG TCGGCTGG	AG	Deletion	PROBABLE PROLINE AND GLYCINE RICH TRANSMEMBRANE PROTEIN	M
CA	AC	Nonsyn	Probable short-chain type dehydrogenase/reductase (EC 1)	R
TGC	TC	Deletion		
GCGAC	GC	Deletion	putative membrane protein	M
CA	GA	Synon	putative membrane protein	M
TG	GG	Synon	putative membrane protein	M

ATTGCCGCGGT AT Deletion putative membrane protein CGGTGGTGGTG GTGGTG GTGGTG CCGCGCGCC C C C					(continua
CGGTGGTGGTG GTGGTG CCGCGCGCCC C C C C C	Ref	Var	Type	Function	COGs
GTGGTG CCGCGCGCGC C C C C C C C C C C C	ATTGCCGCGGT	AT	Deletion	putative membrane protein	M
C GC CA Nonsyn Ribosomal large subunit pseudouridine synthase D (EC 5.4.99.23) TG GT Nonsyn S-adenosyl-L-methionine dependent methyltransferase, similar to cyclopropane-fatty-acyl-phospholipid synthase CT GC Nonsyn Serine/threonine protein kinase CT GC Nonsyn Sphingomyelinase (EC 3.1.4.12) GTC ACT SSU rRNA 16S rRNA, small subunit ribosomal RNA TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC3 @ Toxin 1, PIN domain TC GG Nonsyn Transcription termination factor Rho TGATCCGGCG TG Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)			Deletion	putative membrane protein	M
AC CA Nonsyn Ribosomal large subunit pseudouridine synthase D (EC 5.4.99.23) TG GT Nonsyn S-adenosyl-L-methionine dependent methyltransferase, similar to cyclopropane-fatty-acyl-phospholipid synthase CT GC Nonsyn Serine/threonine protein kinase TGA TTA Nonsyn Sphingomyelinase (EC 3.1.4.12) GTC ACT SSU rRNA 16S rRNA, small subunit ribosomal RNA TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)			Deletion	putative secreted protein	O
TG GT Nonsyn S-adenosyl-L-methionine dependent methyltransferase, similar to cyclopropane-fatty-acyl-phospholipid synthase CT GC Nonsyn Serine/threonine protein kinase TGA TTA Nonsyn Sphingomyelinase (EC 3.1.4.12) GTC ACT SSU rRNA 16S rRNA, small subunit ribosomal RNA TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	CA	AG	Nonsyn	Pyrimidine-nucleoside phosphorylase (EC 2.4.2.2)	L
CT GC Nonsyn Serine/threonine protein kinase TGA TTA Nonsyn Sphingomyelinase (EC 3.1.4.12) GTC ACT SSU rRNA 16S rRNA, small subunit ribosomal RNA TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	AC	CA	Nonsyn	Ribosomal large subunit pseudouridine synthase D (EC 5.4.99.23)	S
TGA TTA Nonsyn Sphingomyelinase (EC 3.1.4.12) GTC ACT SSU rRNA 16S rRNA, small subunit ribosomal RNA TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain STA Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Transcriptional regulator (MarR family) TG CT Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	TG	GT	Nonsyn	1	C
GTC ACT SSU rRNA 16S rRNA, small subunit ribosomal RNA TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TC TA Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	CT	GC	Nonsyn	Serine/threonine protein kinase	T
TCAT CTTA SSU rRNA 16S rRNA, small subunit ribosomal RNA CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain STGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	TGA	TTA	Nonsyn	Sphingomyelinase (EC 3.1.4.12)	S
CG GC Nonsyn toxin VapC12 @ Toxin 1, PIN domain CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TC TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	GTC	ACT		SSU rRNA 16S rRNA, small subunit ribosomal RNA	S
CA AA Nonsyn toxin VapC25 @ Toxin 1, PIN domain CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain STGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	TCAT	CTTA		SSU rRNA 16S rRNA, small subunit ribosomal RNA	S
CT TA Nonsyn toxin VapC29 @ Toxin 1, PIN domain TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	CG	GC	Nonsyn	toxin VapC12 @ Toxin 1, PIN domain	S
TC GG Nonsyn toxin VapC31 @ Toxin 1, PIN domain TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 CA AC Nonsyn Trehalase (EC 3.2.1.28)	CA	AA	Nonsyn	toxin VapC25 @ Toxin 1, PIN domain	S
TGATCCGGCG TG Deletion Transcription termination factor Rho TGA TA Deletion Transcriptional regulator, AraC family TG	CT	TA	Nonsyn	toxin VapC29 @ Toxin 1, PIN domain	S
TGA TA Deletion Transcriptional regulator, AraC family TG CT Nonsyn Transcriptional regulator, MarR family TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 M CA AC Nonsyn Trehalase (EC 3.2.1.28)	TC	GG	Nonsyn	toxin VapC31 @ Toxin 1, PIN domain	S
TG CT Nonsyn Transcriptional regulator, MarR family H TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 M CA AC Nonsyn Trehalase (EC 3.2.1.28)	TGATCCGGCG	TG	Deletion	Transcription termination factor Rho	K
TG CT Nonsyn Translation elongation factor Ts TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 M CA AC Nonsyn Trehalase (EC 3.2.1.28)	TGA	TA	Deletion	Transcriptional regulator, AraC family	K
TA AC Nonsyn Transmembrane transport protein MmpL8/MmpL10/MmpL12 M CA AC Nonsyn Trehalase (EC 3.2.1.28)	TG	CT	Nonsyn		
CA AC Nonsyn Trehalase (EC 3.2.1.28)	TG	CT	Nonsyn	Translation elongation factor Ts	
·	TA	AC	Nonsyn	Transmembrane transport protein MmpL8/MmpL10/MmpL12	
GTG GG Deletion Trehalose-6-phosphate phosphatase (EC 3.1.3.12)	CA	AC	Nonsyn	Trehalase (EC 3.2.1.28)	
	GTG	GG	Deletion	Trehalose-6-phosphate phosphatase (EC 3.1.3.12)	G
TA AC Nonsyn tRNA threonylcarbamoyladenosine biosynthesis protein TsaB .	TA	AC	Nonsyn	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB	J

_				(conclusão)
Ref	Var	Type	Function	COGs
GACGCTGGCTC CGCTG	GG	Deletion	UDP-N-acetylmuramateL-alanine ligase (EC 6.3.2.8)	М
AG	CT	Nonsyn	Uncharacterized protein Q1 colocalized with Q	S
GACCTC	GACTAC	Nonsyn	Uncharacterized protein Q1 colocalized with Q	S
GAT	AAT	Nonsyn	Uncharacterized protein, similar to the N-terminal domain of Lon protease	S
AC	CC	Nonsyn	Vegetative cell wall protein gp1 precursor	M
CA	AC		Intergenic	
CTC	CAC		Intergenic	
TCGC	CCGC		Intergenic	
TGCG	TG	Deletion	Intergenic	
CG	TT		Intergenic	
GGAT	GT	Deletion	Intergenic	
CTGACGACGCG ACAGCGGCGT	СТ	Deletion	Intergenic	
AA	TC		Intergenic	
CACACCATGGT GAA	CA	Deletion	Intergenic	
TGT	TAT		Intergenic	
GTC	GC	Deletion	Intergenic	
ACGGATGC	ACGGATGT		Intergenic	
AGGGGGGGGG GGGGGGGGA	AGGGGGG GGGGGGA	Deletion	Intergenic	
CAA	CCA		Intergenic	
GG	CC		Intergenic	

APÊNDICE B – Padrão de MIRU-VNTR e *Spoligotyping* das estirpes MP1 e MP2 de *Mycobacterium pinnipedii*

ID↓	Species	<u>Lineage</u>	MLVA MtbC15-9	154 MIRU02	424 Mtub04	577 ETRC	580 MIRU04	802 MIRU40	960 MIRU10	1644 MIRU16	1955 Mtub21	2059 MIRU20	2163b OUB11b	2165 ETRA	2347 Mtub29	2401 Mtub30	2461 ETRB	2531 MIRU23	2687 MIRU24	2996 MIRU26	3007 MIRU27	3171 Mtub34	3192 MIRU31	3690 Mtub39	4052 OUB26	4156 OUB4156	4348 MIRU39
M. pinnipedii		P1 (m)	20691-149	2	3	4	7	2	6	4	4	2	9	9	3	4	4	4	2	2	2	3	3	3	7	0	2
M. pinnipedii		P2 (m)	20692-149	2	3	4	5	2	5	4	4	2	9	9	3	4	4	4	2	2	2	3	3	3	7	0	2

<u>Species</u> <u>Lineage</u>	<u>Spoligo pattern</u>	
M. pinnipedii	P1(m)	SB0155
M. pinnipedii	P2(m)	SB2455

APÊNDICE C – Agrupamentos de proteínas de genes ortólogos entre estirpes de *Mycobacterium pinnipedii*

Estirpes de M. pinnipedii/Agrupamentos	(continua) Número de Proteínas
ATCC BAA-688	
G01222	
G01491	8
G01492	
G01498	
MP1	
1 T G G D 1 1 200	
ATCC BAA-688	
G01222	
G01491	4
G01492	
G01498	
MP2	
ATCC BAA-688	
G01222	
G01491	43
G01492	
MP1	
MP2	

	(continua)
Estirpes de M.	Número de
pinnipedii/Agrupamentos	Proteínas
ATCC DAA COO	
ATCC BAA-688	
G01222	12
G01491	12
G01498	
MP1	
MP2	
ATCC BAA-688	
G01222	
G01492	16
	10
G01498	
MP1 MP2	
IVIP 2	
ATCC BAA-688	
G01491	
	0
G01492 G01498	8
MP1	
MP2	
IVII Z	
G01222	
G01491	
G01492	21
G01498	
MP1	
MP2	
ATCC BAA-688	
G01492	
G01492 G01498	1
MP1	1
MP2	
ATCC BAA-688	
G01491	
G01492	1
G01498	
MP1	
ATCC BAA-688	
G01222	
G01491	8
UU1471	O

	(conclusão)
Estirpes de M.	Número de
pinnipedii/Agrupamentos	Proteínas
G01492	
G01498	
G01222	
G01491	
G01492	1
G01498	
MP2	
G01222	
G01491	
G01492	2
MP1	
MP2	
ATCC BAA-688	
G01222	
G01491	3
MP1	
MP2	
ATCC BAA-688	
G01222	
G01498	0
MP1	
MP2	

APÊNDICE D – Lista do conjunto de genomas do Complexo *Mycobacterium tuberculosis* (MTBC) utilizados neste estudo, nome da estirpe, número de acesso do RefSeq do NCBI e linhagem bacteriana

			(continua)
Espécie	Estirpe	Número de Acesso (RefSeq NCBI)	Linhagem
M. tuberculosis	EAI5	NC_021740.1/CP006578.1	L1
M. tuberculosis	96121	NZ_CP009427.1/CP009427.1	L1
M. tuberculosis	HN-024	AP018033.1	L1
M. tuberculosis	W-148	NZ_CP012090.1/CP012090.1	L2
M. tuberculosis	CCDC5180	NC_017522.1/CP001642.1	L2
M. tuberculosis	Beijing/NITR203	NC_021054.1/CP005082.1	L2
M. tuberculosis	CCDC5079	NC_021251.1/CP002884.1	L2
M. tuberculosis	HKBS1	NZ_CP002871.1/CP002871.1	L2
M. tuberculosis	BT2	NZ_CP002882.1/CP002882.1	L2
M. tuberculosis	BT1	NZ_CP002883.1/CP002883.1	L2
M. tuberculosis	CCDC5180	NZ_CP002885.1/CP002885.1	L2
M. tuberculosis	K	NZ_CP007803.1/CP007803.1	L2
M. tuberculosis	KIT87190	NZ_CP007809.1/CP007809.1	L2
M. tuberculosis	ZMC13-264	NZ_CP009100.1/CP009100.1	L2
M. tuberculosis	ZMC13-88	NZ_CP009101.1/CP009101.1	L2
M. tuberculosis	96075	NZ_CP009426.1/CP009426.1	L2
M. tuberculosis	49-02	NZ_HG813240.1/HG813240.1	L2
M. tuberculosis	SCAID 187.0	NZ_CP012506.2/CP012506.2	L2
M. tuberculosis	SCAID 320.0	NZ_CP016794.1/CP016794.1	L2
M. tuberculosis	SCAID 252.0	NZ_CP016888.1/CP016888.1	L2
M. tuberculosis	Beijing	NZ_CP011510.1/CP011510.1	L2
M. tuberculosis	1458	NZ_CP013475.1/CP013475.1	L2
M. tuberculosis	Beijing-like/35049	NZ_CP017593.1/CP017593.1	L2
M. tuberculosis	Beijing-like/36918	NZ_CP017594.1/CP017594.1	L2
M. tuberculosis	Beijing-like/38774	NZ_CP017595.1/CP017595.1	L2
M. tuberculosis	Beijing/391	NZ_CP017596.1/CP017596.1	L2
M. tuberculosis	Beijing-like/50148	NZ_CP017597.1/CP017597.1	L2
M. tuberculosis	Beijing-like/1104	NZ_CP017598.1/CP017598.1	L2
M. tuberculosis	MDRMA2491	NZ_CP023586.1	L2
M. tuberculosis	CSV383	NZ_CP023599.1	L2
M. tuberculosis	LN55	NZ_CP023609.1	L2
M. tuberculosis	MDRMA1565	NZ_CP023626.1	L2
M. tuberculosis	TBV4768	NZ_CP023639.1	L2
M. tuberculosis	CDC1551	NC_002755.2/AE000516.2	L4

			(continua)
Espécie	Estirpe	Número de Acesso (RefSeq NCBI)	Linhagem
M. tuberculosis	KZN 1435	NC_012943.1/CP001658.1	L4
M. tuberculosis	Haarlem	NC_022350.1/CP001664.1	L4
M. tuberculosis	KZN 4207	NC_016768.1/CP001662.1	L4
M. tuberculosis	KZN 605	NC_018078.1/CP001976.1	L4
M. tuberculosis	CTRI-2	NC_017524.1/CP002992.1	L4
M. tuberculosis	H37Rv	NC_018143.2/CP003248.2	L4
M. tuberculosis	7199-99	NC_020089.1/HE663067.1	L4
M. tuberculosis	I0004241-1	NZ_CP018303.1/CP018303.1	L4
M. tuberculosis	M0018684-2	NZ_CP018305.1/CP018305.1	L4
M. tuberculosis	I0004000-1	NZ_CP018302.1/CP018302.1	L4
M. tuberculosis	I0002801-4	NZ_CP018301.1/CP018301.1	L4
M. tuberculosis	I0002353-6	NZ_CP018300.1/CP018300.1	L4
M. tuberculosis	M0002959-6	NZ_CP018304.1/CP018304.1	L4
M. tuberculosis	DK9897	NZ_CP018778.1/CP018778.1	L4
M. tuberculosis	CSV4519	NZ_CP023573.1	L4
M. tuberculosis	CSV4644	NZ_CP023574.1	L4
M. tuberculosis	CSV5769	NZ_CP023575.1	L4
M. tuberculosis	CSV10399	NZ_CP023576.1	L4
M. tuberculosis	CSV11678	NZ_CP023577.1	L4
M. tuberculosis	LE486	NZ_CP023578.1	L4
M. tuberculosis	LE492	NZ_CP023579.1	L4
M. tuberculosis	LN180	NZ_CP023580.1	L4
M. tuberculosis	LN2358	NZ_CP023581.1	L4
M. tuberculosis	LN3756	NZ_CP023582.1	L4
M. tuberculosis	MDRDM260	NZ_CP023583.1	L4
M. tuberculosis	MDRDM627	NZ_CP023584.1	L4
M. tuberculosis	MDRDM1098	NZ_CP023585.1	L4
M. tuberculosis	ME1473	NZ CP023587.1	L4
M. tuberculosis	TBDM425	NZ CP023588.1	L4
M. tuberculosis	TBV5000	NZ_CP023589.1	L4
M. tuberculosis	TBV5362	NZ CP023590.1	L4
M. tuberculosis	TBV5365	NZ_CP023591.1	L4
M. tuberculosis	SLM036	NZ_CP023592.1	L4
M. tuberculosis	SLM040	NZ CP023593.1	L4
M. tuberculosis	SLM056	NZ_CP023594.1	L4
M. tuberculosis	SLM060	NZ_CP023595.1	L4
M. tuberculosis	SLM063	NZ CP023596.1	L4
M. tuberculosis	SLM088	NZ_CP023597.1	L4
M. tuberculosis	SLM100	NZ_CP023598.1	L4
M. tuberculosis	CSV3611	NZ_CP023600.1	L4
M. tuberculosis	CSV9577	NZ_CP023601.1	L4

			(continua)
Espécie	Estirpe	Número de Acesso (RefSeq NCBI)	Linhagem
M. tuberculosis	LE76	NZ_CP023604.1	L4
M. tuberculosis	LE79	NZ_CP023605.1	L4
M. tuberculosis	LE103	NZ_CP023606.1	L4
M. tuberculosis	LE371	NZ_CP023607.1	L4
M. tuberculosis	LE410	NZ_CP023608.1	L4
M. tuberculosis	LN317	NZ_CP023610.1	L4
M. tuberculosis	LN763	NZ_CP023611.1	L4
M. tuberculosis	LN2978	NZ_CP023612.1	L4
M. tuberculosis	LN3584	NZ_CP023613.1	L4
M. tuberculosis	LN3588	NZ_CP023614.1	L4
M. tuberculosis	LN3589	NZ_CP023615.1	L4
M. tuberculosis	LN3668	NZ_CP023616.1	L4
M. tuberculosis	LN3672	NZ_CP023617.1	L4
M. tuberculosis	LN3695	NZ_CP023618.1	L4
M. tuberculosis	LN1100	NZ_CP023619.1	L4
M. tuberculosis	LN1856	NZ_CP023620.1	L4
M. tuberculosis	LN2900	NZ_CP023621.1	L4
M. tuberculosis	MDRDM827	NZ_CP023622.1	L4
M. tuberculosis	MDRMA203	NZ_CP023623.1	L4
M. tuberculosis	MDRMA701	NZ_CP023624.1	L4
M. tuberculosis	MDRMA863	NZ_CP023625.1	L4
M. tuberculosis	MDRMA2019	NZ_CP023627.1	L4
M. tuberculosis	MDRMA2082	NZ_CP023628.1	L4
M. tuberculosis	MDRMA2260	NZ_CP023629.1	L4
M. tuberculosis	MDRMA2441	NZ_CP023630.1	L4
M. tuberculosis	TBDM1506	NZ_CP023631.1	L4
M. tuberculosis	TBDM2189	NZ_CP023632.1	L4
M. tuberculosis	TBDM2444	NZ_CP023633.1	L4
M. tuberculosis	TBDM2487	NZ_CP023634.1	L4
M. tuberculosis	TBDM2489	NZ_CP023635.1	L4
M. tuberculosis	TBDM2699	NZ_CP023636.1	L4
M. tuberculosis	TBDM2717	NZ CP023637.1	L4
M. tuberculosis	TBV4766	NZ_CP023638.1	L4
M. tuberculosis	TBV4952	NZ_CP023640.1	L4
M. bovis	SP38	NZ_CP015773.2/CP015773.2	L1
M. bovis	35	NZ_JQEV0000000.1	L1
M. bovis	D11-00843	NZ_MTZR00000000.1	L1
M. bovis	D11-03130	NZ_MTZS00000000.1	L1
M. bovis	D11-01601	NZ_MTZQ00000000.1	L1
M. bovis	MB1	NZ_CDHF00000000.1	L1
M. bovis	MB3	NZ_CDHH00000000.1	L1

			(continua)
Espécie	Estirpe	Número de Acesso (RefSeq NCBI)	Linhagem
M. bovis	1595	NZ_CP012095.1/CP012095.1	LB3
M. bovis	09-1191	NZ_JPFP00000000.1	LB3
M. bovis	05-567	NZ_JPFQ00000000.1	LB3
M. bovis	05-566	NZ_JPFR00000000.1	LB3
M. bovis	49-09	NZ_JQES00000000.1	LB3
M. bovis	32-08	NZ_JQER00000000.1	LB3
M. bovis	18-08C	NZ_JQEQ00000000.1	LB3
M. bovis	08-08BF2	NZ_JQET00000000.1	LB3
M. bovis	09-1193	NZ_JQEN00000000.1	LB3
M. bovis	534	NZ_JQEM00000000.1	LB3
M. bovis	0822-11	NZ_JQEW00000000.1	LB3
M. bovis	61-09	NZ_JQEX00000000.1	LB3
M. bovis	45-08B	NZ_JQEP00000000.1	LB3
M. bovis	09-1192	NZ_JQEO00000000.1	LB3
M. bovis	50	NZ_JQEU00000000.1	LB3
M. bovis	W-1171	NZ_JXTK00000000.1	LB3
M. bovis	MbURU-001	NZ_LFGY00000000.1	LB3
M. bovis	MB4	NZ_CDHE00000000.1	LB3
M. bovis	B-3222	NZ_LNOF00000000.1	LB3
M. bovis	M1009	NZ_NBZZ00000000.1	LB3
M. bovis	M1010	NZ_NCTD00000000.1	LB3
M. bovis	MbURU-005	NZ_NBAA00000000.1	LB3
M. bovis	MbURU-004	NZ_NAZZ00000000.1	LB3
M. bovis	MbURU-006	NZ_NBAB00000000.1	LB3
M. bovis	MbURU-003	NZ_NAZY00000000.1	LB3
M. bovis	MbURU-002	NZ_NAZX00000000.1	LB3
M. bovis	MbURU-008	NZ_NBAD00000000.1	LB3
M. bovis	MbURU-023	NZ_NBAE00000000.1	LB3
M. bovis	MbURU-007	NZ_NBAC00000000.1	LB3
M. bovis	MbURU-009	NZ_NAZS00000000.1	LB3
M. bovis	MbURU-017	NZ_NAZQ00000000.1	LB3
M. bovis	MbURU-021	NZ_NAZV00000000.1	LB3
M. bovis	MbURU-011	NZ_NAZJ00000000.1	LB3
M. bovis	MbURU-010	NZ_NAZL00000000.1	LB3
M. bovis	MbURU-020	NZ_NAZU00000000.1	LB3
M. bovis	MbURU-013	NZ_NAZN00000000.1	LB3
M. bovis	MbURU-022	NZ_NAZW00000000.1	LB3
M. bovis	MbURU-016	NZ_NAZP00000000.1	LB3
M. bovis	MbURU-019	NZ_NAZT00000000.1	LB3
M. bovis	MbURU-018	NZ_NAZR00000000.1	LB3
M. bovis	MbURU-015	NZ_NAZO00000000.1	LB3

			(conclusão)
Espécie	Estirpe	Número de Acesso (RefSeq NCBI)	Linhagem
M. africanum	MAL010118	NZ_JLAC00000000.1	L5
M. africanum	GM041182	NC_015758.1/FR878060.1	L6
M. africanum	UT307	NZ_CP014617.1/CP014617.1	L6
M. africanum	K85	NZ_JLCY00000000.1	L6
M. africanum	MAL010084	NZ_JKXR00000000.1	L6
M. africanum	MAL020148	NZ_JKXO00000000.1	L6
M. africanum	MAL010070	NZ_JLAZ00000000.1	L6
M. africanum	MAL010071	NZ_JLAY0000000.1	L6
M. africanum	MAL010081	NZ_JLAT00000000.1	L6
M. africanum	MAL010099	NZ_JLAO00000000.1	L6
M. africanum	MAL010100	NZ_JLAN00000000.1	L6
M. africanum	MAL010102	NZ_JLAM00000000.1	L6
M. africanum	MAL010111	NZ_JLAF00000000.1	L6
M. africanum	MAL010112	NZ_JLAE00000000.1	L6
M. africanum	MAL010120	NZ_JLAB00000000.1	L6
M. africanum	MAL010123	NZ_JKZZ00000000.1	L6
M. africanum	MAL010128	NZ_JKZX00000000.1	L6
M. africanum	MAL010129	NZ_JKZW00000000.1	L6
M. africanum	MAL010131	NZ_JKZU00000000.1	L6
M. africanum	MAL010137	NZ_JKZQ00000000.1	L6
M. africanum	MAL020107	NZ_JKZO00000000.1	L6
M. africanum	MAL020135	NZ_JKZI00000000.1	L6
M. africanum	MAL020136	NZ_JKZH00000000.1	L6
M. africanum	MAL020173	NZ_JKYR00000000.1	L6
M. africanum	MAL020176	NZ_JKYP00000000.1	L6
M. africanum	MAL020185	NZ_JKYL00000000.1	L6
M. africanum	MAL020193	NZ_JKYH00000000.1	L6
M. pinnipedii	ATCC BAA-688	NZ_MWXB00000000.1	NA
M. pinnipedii	MP1	NZ_PYQH00000000.2	NA
M. pinnipedii	MP2	NZ_PYQI00000000.2	NA
M. orygis	112400015	NZ_APKD00000000.1	NA
M. caprae	Allgaeu	NZ_CP016401.1	NA
M. caprae	MB2	NZ_CDHG00000000.1	NA
M. caprae	ATCC BAA-824	NZ_MWXD00000000.1	NA
Dassie bacillus	47	NZ_MBDD00000000.1	NA
Dassie bacillus	48	NZ_MBDE00000000.1	NA
Dassie bacillus	49	NZ_MBDF00000000.1	NA
Dassie bacillus	50	NZ_MBDG00000000.1	NA
M. microti	ATCC 19422	NZ_MWXC00000000.1	NA
M. mungi	BM22813	NZ_LXTB00000000.1	NA

APÊNDICE E – Epítopos preditos das proteínas do genoma acessório do Complexo Mycobacterium tuberculosis encontrados no IEDB

		(continua)
ID Epítopo	Epítopo	Proteína
20369	GILTVSVAV	14 kDa antigen
35785	LFAAFPSFA	14 kDa antigen
231704	YVDIYAASV	1-acylglycerol-3-phosphate O-acyltransferase
74761	YLLALRYLA	3-dehydroquinate dehydratase
67077	TVLRFVPPL	3-phosphoshikimate 1-carboxyvinyltransferase
92854	GQFNRYAAM	4,5:9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oate hydrolase
3064	AMASTEGNV	6 kDa early secretory antigenic target
189567	ALQNLARTI	6 KDA early secretory antigenic target
189770	NLARTISEA	6 KDA early secretory antigenic target
3153	AMLQDMAIL	60 kDa chaperonin 2
20859	GLFLTTEAV	60 kDa chaperonin 2
42038	MLQDMAILT	60 kDa chaperonin 2
92380	AVRNAKAAV	60 KDA chaperonin 2 groel2
2372	ALADAVKVT	61 kDa chaperonin 2
31749	KLAGGVAVI	62 kDa chaperonin 2
32103	KLQERLAKL	63 kDa chaperonin 2
261120	AVVADKPEK	64 kDa chaperonin 2
291192	FLTTEAVVA	65 kDa chaperonin 2
21240	GLYRQWALA	Acetyl-/propionyl-coenzyme A carboxylase alpha chain
69514	VLGRLDQKL	Alanine and proline-rich secreted protein apa precursor
226563	AAMASASLV	Alanine and proline-rich secreted protein apa precursor
226573	AMASASLVT	Alanine and proline-rich secreted protein apa precursor
226607	GWVESDAAH	Alanine and proline-rich secreted protein apa precursor
226627	LAESIRPLV	Alanine and proline-rich secreted protein apa precursor
226628	LAIAAMASA	Alanine and proline-rich secreted protein apa precursor
226642	MASASLVTV	Alanine and proline-rich secreted protein apa precursor
226667	SASLVTVAV	Alanine and proline-rich secreted protein apa precursor
226685	TTGDPPFPG	Alanine and proline-rich secreted protein apa precursor
93668 140615	RAWGRRLMI WLYPGAQNL	alpha antigen, antigen 85B, MPB59, antigen 6 amino acid decarboxylase

-		(continua)
ID Epítopo	Epítopo	Proteína
		aminoglycosides/tetracycline-transport integral
151741	TLAPQVEPL	membrane protein
51104	QIMYNYPAM	ANTIGEN 7 ESXH (10 kDa ANTIGEN) (CFP-7)
0110.	Q	(PROTEIN TB10.4)
140649	AMEDLVRAY	ANTIGEN 7 ESXH (10 kDa ANTIGEN) (CFP-7) (PROTEIN TB10.4)
		ANTIGEN 7 ESXH (10 kDa ANTIGEN) (CFP-7)
140652	AMMARDTAE	(PROTEIN TB10.4)
140744	COIMANNINDA	ANTIGEN 7 ESXH (10 kDa ANTIGEN) (CFP-7)
140744	SQIMYNYPA	(PROTEIN TB10.4)
140757	TLQSLGAEI	ANTIGEN 7 ESXH (10 kDa ANTIGEN) (CFP-7)
		(PROTEIN TB10.4)
21078	GLPVEYLQV	Antigen 85-A precursor
2692	ALLDPSQGM	Antigen 85-B precursor (85B)
16332	FIYAGSLSA	Antigen 85-B precursor (85B)
16924	FLTSELPQW	Antigen 85-B precursor (85B)
21275	GMGPSLIGL	Antigen 85-B precursor (85B)
29558	IYAGSLSAL	Antigen 85-B precursor (85B)
32213	KLVANNTRL	Antigen 85-B precursor (85B)
38049	LMIGTAAAV	Antigen 85-B precursor (85B)
54671	RLMIGTAAA	Antigen 85-B precursor (85B)
59627	SMAGSSAMI	Antigen 85-B precursor (85B)
73357	YAGSLSALL	Antigen 85-B precursor (85B)
74768	YLLDGLRAQ	Antigen 85-B precursor (85B)
3118	AMKADIQHV	Antigen 85-C precursor
76221	YTYKWETFL	Antigen 85-C precursor
226562	AAMAAQLQA	ANTIGEN TB8.4 (Hypothetical protein)
226565	AAQFNASPV	ANTIGEN TB8.4 (Hypothetical protein)
226577	APPPQRAAM	ANTIGEN TB8.4 (Hypothetical protein)
226626	LAAPPPQRA	ANTIGEN TB8.4 (Hypothetical protein)
226645	MSLTVGAGV	ANTIGEN TB8.4 (Hypothetical protein)
226665	RLSLTALSA	ANTIGEN TB8.4 (Hypothetical protein)
226676	SLTALSAGV	ANTIGEN TB8.4 (Hypothetical protein)
92301	AIYDTMQYV	ATP-dependent Clp protease proteolytic subunit
93666	RARKRGITL	conserved hypothetical protein
94314	YLYDETQDV	Conserved hypothetical protein
108953	FLADYGWRL	Conserved hypothetical protein
144941	RLANLLPLI	Conserved hypothetical protein
232179	QLLGWYSRV	Conserved hypothetical protein
232337	TLLDYAAGV	Conserved hypothetical protein
232382	VLSGYIASV	Conserved hypothetical protein
93920	SLYPPCLFK	Conserved membrane protein
189894	VLAGSVDEL	cyclopropane-fatty-acyl-phospholipid synthase
93789	RVACRDVEV	cytochrome P450 140 cyp140

		(continua)
ID Epítopo	Epítopo	Proteína
40622	LVVADLSFI	CYTOTOXIN HAEMOLYSIN HOMOLOGUE TLYA
21185	GLVDFVKHI	DNA gyrase subunit B
37296	LLFRFMRPL	DNA gyrase subunit B
38071	LMLPGMNGI	DNA-binding response regulator mtrA
21243	GLYYLTTEV	DNA-directed RNA polymerase subunit beta'
37667	LLPLGYPFV	DNA-directed RNA polymerase subunit beta'
54932	RMPAVTDLV	DnaK
59250	SLLEIGEGV	DnaK
69608	VLLLDVTPL	DnaK
37146	LLDEGKQSL	early secreted antigenic target 6 kDa
189685	IQGNVTSIH	early secreted antigenic target 6 kDa
189695	KLAAAWGGS	early secreted antigenic target 6 kDa
189816	QQWNFAGIE	early secreted antigenic target 6 kDa
189851	SLLDEGKQS	early secreted antigenic target 6 kDa
189854	SLTKLAAAW	early secreted antigenic target 6 kDa
189897	VQQKWDATA	early secreted antigenic target 6 kDa
2733	ALLSDWLPA	Enoyl-[acyl-carrier-protein] reductase [NADH]
21000	GLLHHAPSL	glutamine synthetase GLNA1 (glutamine synthase) (GS-I)
37084	LLAFTNPTV	glutamine synthetase GLNA1 (glutamine synthase) (GS-I)
51535	QLSDVIDRL	glutamine synthetase GLNA1 (glutamine synthase) (GS-I)
140600	SLWKDGAPL	glutamine synthetase GLNA1 (glutamine synthase) (GS-I)
59099	SLFPEFSEL	Heat shock protein hspX
41078	MALLRLPLV	hydrolase
54704	RLPLVLPAV	hydrolase
231881	FLSDYGPQL	Hypothetical alanine and proline rich protein
231761	ALIDFLAAV	hypothetical exported protein
2690	ALLDLGWTI	Hypothetical protein
6303	CGKAGCQTY	hypothetical protein
15769	FEWYYQSGL	hypothetical protein
18698	GANIPAEFL	hypothetical protein
19875	GGHNAVFNF	hypothetical protein
29360	IVMPVGGQS	hypothetical protein
40163	LTSELPQWL	hypothetical protein
41044	MAGSSAMIL	hypothetical protein
51449	QLNAMKGDL	hypothetical protein
52666	QTYKWETFL	hypothetical protein
54878	RMFAANLGV	hypothetical protein
69433	VLDRRLHLV	hypothetical protein

		(continua)
ID Epítopo	Epítopo	Proteína
72084	VYLLDGLRA	hypothetical protein
76581	YYQSGLSIV	hypothetical protein
93270	LLYDGSFAV	hypothetical protein
93802	RVRQAWDTL	Hypothetical protein
93884	SFYGYGFNV	Hypothetical protein
109499	LLWAFAHRQ	Hypothetical protein
109932	SMLPPGYPV	hypothetical protein
140617	YLPDPTVGV	hypothetical protein
161402	LILATMLVA	hypothetical protein
180078	YQSGLSIVM	hypothetical protein
189568	AMKGDLQSS	hypothetical protein
189580	AVFNFPPNG	hypothetical protein
189617	ELGGANIPA	hypothetical protein
189627	FLENFVRSS	hypothetical protein
189635	FYSDWYSPA	hypothetical protein
189649	GLSIVMPVG	hypothetical protein
189650	GLSMAGSSA	hypothetical protein
189720	LLDGLRAQD	hypothetical protein
189744	MGDAGGYKA	hypothetical protein
189782	PAFEWYYQS	hypothetical protein
189828	RLWVYCGNG	hypothetical protein
189833	SAAIGLSMA	hypothetical protein
189853	SLSALLDPS	hypothetical protein
189858	SQGMGPSLI	hypothetical protein
189862	SSFYSDWYS	hypothetical protein
189888	VANNTRLWV	hypothetical protein
189895	VMPVGGQSS	hypothetical protein
189913	YHPQQFIYA	hypothetical protein
189921	YSDWYSPAC	hypothetical protein
231685	YLNDFAQLL	hypothetical protein
232319	SVASGIFNV	hypothetical protein
37140	LLDAHIPQL	Hypothetical protein esxG (PE family protein)
231769	ALSGFCAAV	integral membrane protein
231916	GLAATSFPL	integral membrane protein
151626	FLALALPIV	integral membrane transport protein
151742	TMSTAVLTL	integral membrane transport protein
260013	AVAAVVAPV	iron ABC transporter permease
412721	VVAPVGVLV	iron ABC transporter permease
37075	LLACAVIHA	ISMt1 transposase B
156691	ILVAGLSGC	lipoprotein
156699	KITGTATGV	lipoprotein
156786	VIDGKDQNV	lipoprotein

-		(continua)		
ID Epítopo	Epítopo	Proteína		
226561	AAGNVNIAI	lipoprotein		
226594	GAASGPKVV	lipoprotein		
226597	GLGNVNGVT	lipoprotein		
226608	GYTSGTGQG	lipoprotein		
226640	LTDGNPPEV	lipoprotein		
92331	AMWDWSKSV	lipoprotein lppH		
859671	AAMDGTVLV	lipoprotein LppZ		
859708	ALAADPQSL	lipoprotein LppZ		
859988	GLAALCAAV	lipoprotein LppZ		
860529	KLDDVVFPL	lipoprotein LppZ		
860788	LIFTSPTTL	lipoprotein LppZ		
861034	NLINTKLTV	lipoprotein LppZ		
861349	RLMYAYIST	lipoprotein LppZ		
232057	LMVPGAAPV	lipoprotein lpqB		
5251	AVAGAAILV	Lipoprotein lpqH precursor		
27586	IMYNYPAML	low molecular weight protein antigen 7		
140650	AMLGHAGDM	low molecular weight protein antigen 7		
140599	SLRNWIATL	MCE-family lipoprotein LprL		
231683	YLDAIQQPV	MCE-family protein mce1B		
231762	ALIGGFKPV	MCE-family protein mce1B		
108960	FLQGAKWYL	MCE-family protein Mce3A		
140623	YVYPDNLPV	MCE-family protein MCE3A		
253618	ALLPLPTVV	membrane protein		
140616	YLLADTFTV	membrane-associated phospholipase C		
231976	ILAAIAVGV	metal cation transporter P-type ATPase CtpF		
16506	FLAGGPHAV	MPT51/MPB51 antigen precursor		
34782	LAGKGISVV	MPT51/MPB51 antigen precursor		
64713	TLAGKGISV	MPT51/MPB51 antigen precursor		
489445	AAYPITGKL	MPT63 (ANTIGEN MPT63/MPB63)		
489460	AIATFAAPV	MPT63 (ANTIGEN MPT63/MPB63)		
489471	AMNNGMEDL	MPT63 (ANTIGEN MPT63/MPB63)		
41947	MLGNAPSVV	mtp40 protein - Mycobacterium tuberculosis		
41910	MLASTLTDA	mycocerosic acid synthase		
72750	WLLIDTSNA	mycocerosic acid synthase		
74606	YLDLALMSV	mycocerosic acid synthase		
299369	GIRVPAGQV	NADH dehydrogenase subunit D		
320960	KLVTEGIRV	NADH dehydrogenase subunit D		
231922	GLSEGAYHV	nitrate reductase-related protein		
260778	AVQFAKAAK	oxidoreductase FADB5		
329106	LIHAAAGGV	oxidoreductase FADB5		
333688	LRAGERVLI	oxidoreductase FADB5		
16921	FLTQDLPLL	PE family protein		

		(continua)
ID Epítopo	Epítopo	Proteína
2556	ALGGGATGV	PE_PGRS 33
64412	TIPEALAAV	PE_PGRS 33
20806	GLDPTGVAV	PE-PGRS FAMILY PROTEIN
22962	GVFEGIAVL	PE-PGRS FAMILY PROTEIN
44802	NLLVTGFDT	PE-PGRS family protein
37951	LLWQDPVPA	Peroxidase/catalase T
16739	FLLPDAQSI	Phosphate-binding protein pstS 1 precursor
74914	YLSEGDMAA	Phosphate-binding protein pstS 1 precursor
109451	KVFFVNWFR	Phosphoenolpyruvate carboxykinase
232383	VLTDWAAPV	Phospholipase C 3
108769	ALPGPDGVV	POSSIBLE DEHYDROGENASE
189849	SIIIPTLNV	POSSIBLE GLYCOSYL TRANSFERASE
110075	VVEPPRQLV	POSSIBLE OXIDOREDUCTASE (Hypothetical protein)
231703	YTWNGTQWV	Possible secreted unknown protein
2886	ALSNGIPPV	PPE FAMILY PROTEIN
20863	GLFPFHFTL	PPE FAMILY PROTEIN
37539	LLLPIAPEL	PPE FAMILY PROTEIN
69698	VLPPEINSL	PPE family protein
93255	LLGQNTPAI	PPE FAMILY PROTEIN
174182	LIASNVAGV	PPE family protein
231754	ALAPWQQAV	PPE family protein
418943	ALAMEVYQA	PPE family protein PPE68
418945	ALTEMDYFI	PPE family protein PPE68
418947	AMPPELNTA	PPE family protein PPE68
418964	ELTARLNSL	PPE family protein PPE68
418975	GINTIPIAL	PPE family protein PPE68
418997	KLEPMASIL	PPE family protein PPE68
108768	ALMDYGFRV	Probable 3-isopropylmalate dehydratase (small subunit) leuD
93808	RVYAELAAL	Probable acetolactate synthase
231885	FLWDEAWYA	Probable amino acid decarboxylase
110101	WMYEGKHVL	Probable ATP synthase alpha chain atpA
94323	YMLDMTFPV	Probable conserved secreted protein TB22.2
109979	TLFDWGFAL	probable D-alanyl-D-alanine carboxypeptidase dacB2
231706	YVIEPAAGL	Probable glycyl-tRNA synthetase glyS
108806	ATADLELAY	Probable methylmalonyl-CoA mutase large subunit
109492	LLHEVGITV	Probable Serine hydroxymethyltransferase 1 glyA1
72756	WLLSVLAAV	PROBABLE SERINE PROTEASE PEPA (MTB32A)
92708	FMFDYIPPV	Probable succinate dehydrogenase [iron-sulfur subunit]
92308	ALFPIIWAL	Probable sugar-transport integral membrane protein ABC transporter

		(conclusão)
ID Epítopo	Epítopo	Proteína
109350	IQAVFGFSL	Probable transmembrane alkane 1-monooxygenase alkB
109885	RVMPVFAFK	Probable Ubiquinol-cytochrome C reductase QcrB (cytochrome B subunit)
2968	ALVLLMLPV	PstA-1
18312	FVVALIPLV	PstA-1
21040	GLMTAVYLV	PstA-1
27191	ILLSIARVV	PstA-1
93259	LLLRPFWPA	Putative arabinosyltransferase
58994	SLAGFVRML	Putative virulence-regulating 38 kDa protein
74927	YLSQIAVLL	Putative virulence-regulating 38 kDa protein
231773	AMAAGLSSL	ribonuclease E rne
231917	GLAELLAAL	ribonucleoside-diphosphate reductase (large subunit) nrdZ
42897	MTYAAPLFV	RNA polymerase beta-subunit
92817	GLAGGAATA	secreted antigen 85-B fbpB (85B) (antigen 85 complex B)
31902	KLIANNTRV	Secreted antigen Ag85A
94155	VLMGGVPGV	secreted L-alanine dehydrogenase ald (40 kda antigen)
9414	DMWEHAFYL	Superoxide dismutase
3145	AMLAGTAVV	TRANSMEMBRANE PROTEIN (Hypothetical protein)
93190	KVRGRLLAL	TRANSMEMBRANE PROTEIN (Hypothetical protein)
251793	AIIIAVLLV	transmembrane serine/threonine-protein kinase A
252997	ALAIIIAVL	transmembrane serine/threonine-protein kinase A
253660	ALLWAAGVL	transmembrane serine/threonine-protein kinase A
231686	YLVDCGAEL	trehalose-6-phosphate phosphatase otsB1
54635	RLLDDTPEV	Tuberculin-active protein
37272	LLFLGVVFL	tuberculin-like peptide
13371	EMFGGTPAV	Uncharacterized protein Rv2557/MT2634
74868	YLPRFLFGV	Uncharacterized protein Rv3403c/MT3511
69368	VLAGIFAGV	Uncharacterized transporter Rv1979c/MT2031

APÊNDICE F - Lista dos grupos de fatores de virulência e genes relacionados seguindo a ordem no *heatmap*

				(contin
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
79	Cu.Zn superoxide dismutase	sodC	Rv0432	Stress adaptation
78	AhpC	ahpC	Rv2428	Stress adaptation
19	Carboxylesterase	caeA	Rv2224c	Cell surface components
6	Glutamine synthesis	glnA1	Rv2220	Amino acid and purine metabolism
36	FAS-II	kasB	Rv2246	Lipid and fatty acid metabolism
67	Alpha-crystallin	hspX	Rv2031c	Secreted proteins
71	ESX-3 (T7SS)	eccA3	Rv0282	Secretion system
		eccB3	Rv0283	Secretion system
		eccC3	Rv0284	Secretion system
		eccD3	Rv0290	Secretion system
		eccE3	Rv0292	Secretion system
		espG3	Rv0289	Secretion system
		esxG	Rv0287	Secretion system
		esxH	Rv0288	Secretion system
		mycP3	Rv0291	Secretion system
25	GPL locus	atf	Rv3816c	Cell surface components
		ecf	Rv0182c	Cell surface components
		fadD23	Rv3826	Cell surface components
		fadE5	Rv0244c	Cell surface components
		fmt	Rv1406	Cell surface components
		gap	Rv1436	Cell surface components
		mbtH	Rv2377c	Cell surface components
		mmpL10	Rv1183	Cell surface components

■				(contin
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
		mmpL4	Rv0450c	Cell surface components
		mmpS4	Rv0451c	Cell surface components
		mps1	Rv0403c	Cell surface components
		mps2	Rv0506	Cell surface components
		papA3	Rv1182	Cell surface components
		pe	Rv1184c	Cell surface components
		pks10	Rv1660	Cell surface components
		rmlA	Rv0334	Cell surface components
		rmlB	Rv3464	Cell surface components
15	Lipoprotein	lprG	Rv1411c	Cell surface components
13	FadE29	fadE29	Rv3543c	Catabolism of cholesterol
3	Tryptophan synthesis	trpD	Rv2192c	Amino acid and purine metabolis
12	FadE28	fadE28	Rv3544	Catabolism of cholesterol
11	Cyp125 PDIM (phthiocerol	cyp125	Rv3545c	Catabolism of cholesterol
24	dimycocerosate) and PGL (phenolic glycolipid) biosynthesis and transport	pks15	Rv2947c	Cell surface components
		pks1	Rv2946c	Cell surface components
		ppsA	Rv2931	Cell surface components
		ppsB	Rv2932	Cell surface components
		ppsC	Rv2933	Cell surface components
		ppsD	Rv2934	Cell surface components
		ppsE	Rv2935	Cell surface components

				(continua
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
		tesA	Rv2928	Cell surface components
		drrC	Rv2938	Cell surface components
		fadD22	Rv2948c	Cell surface components
		fadD26	Rv2930	Cell surface components
		fadD28	Rv2941	Cell surface components
		fadD29	Rv2950c	Cell surface components
		lppx	Rv2945c	Cell surface components
		mas	Rv2940c	Cell surface components
		mmpL7	Rv2942	Cell surface components
		papA5	Rv2939	Cell surface components
4	Leucine synthesis	leuD	Rv2987c	Amino acid and purine metabolism
68	Protein kinase G	pknG	Rv0410c	Secreted proteins
75	Accessory secretion factor	secA2	Rv1821	Secretion system
64	(p)ppGpp synthesis and hydrolysis	relA	Rv2583c	Regulation
74	ESX-5 (T7SS)	cyp143	Rv1785c	Secretion system
		eccA5	Rv1798	Secretion system
		eccB5	Rv1782	Secretion system
		eccD5	Rv1795	Secretion system
		eccE5	Rv1797	Secretion system
		esxM	Rv1792	Secretion system
		esxN	Rv1793	Secretion system
		mycP5	Rv1796	Secretion system
40	Mce1	mce1	Rv0166-Rv0174	Mammalian cell entry (mce) operon
54	Sigma A	sigA/rpoV	Rv2703	Regulation
8	Fused nitrate reductase	narX	Rv1736c	Anaerobic respiration

				(continu
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
20	Mycolic acid trans-cyclopropane synthetase	cmaA2	Rv0503c	Cell surface components
9	Nitrate/nitrite transporter	narK2	Rv1737c	Anaerobic respiration
18	Proximal cyclopropane synthase of alpha mycolates	pcaA	Rv0470c	Cell surface components
34	Lipase	lipF	Rv3487c	Lipid and fatty acid metabolism
30	Iron-dependent regulator	ideR	Rv2711	Copper uptake
35	Phospholipases C	plcA	Rv2351c	Lipid and fatty acid metabolism
		plcB	Rv2350c	Lipid and fatty acid metabolism
		plcC	Rv2349c	Lipid and fatty acid metabolism
		plcD	Rv1755c	Lipid and fatty acid metabolism
53	RegX3	espE	Rv3864	Regulation
		espF	Rv3865	Regulation
		espG1	Rv3866	Regulation
		espH	Rv3867	Regulation
		regX3	Rv0491	Regulation
31	Mycobactin	fadD33	Rv1345	Copper uptake
		fadE14	Rv1058	Copper uptake
		mbtA	Rv2384	Copper uptake
		mbtB	Rv2383c	Copper uptake
		mbtC	Rv2382c	Copper uptake
		mbtD	Rv2381c	Copper uptake
		mbtE	Rv2380c	Copper uptake
		mbtF	Rv2379c	Copper uptake
		mbtG	Rv2378c	Copper uptake
		mbtH	Rv2377c	Copper uptake

				(continua
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
		mbtI	Rv2386c	Copper uptake
		mbtJ	Rv2385	Copper uptake
		mbtK	Rv1347c	Copper uptake
23	Trehalose-recycling ABC transporter	chp	Rv1503c-Rv1507c	Cell surface components
		fadD23	Rv3826	Cell surface components
		icl2	Rv1916	Cell surface components
		lpqY	Rv1235	Cell surface components
		sap	Rv3310	Cell surface components
		sugA	Rv1236	Cell surface components
		sugB	Rv1237	Cell surface components
		sugC	Rv1238	Cell surface components
73	ESX-1 (T7SS)	eccA1	Rv3868	Secretion system
		eccB1	Rv3869	Secretion system
		eccCa1	Rv3870	Secretion system
		eccCb1	Rv3871	Secretion system
		eccD1	Rv3877	Secretion system
		eccE1	Rv3882c	Secretion system
		espA	Rv3616c	Secretion system
		espB	Rv3881c	Secretion system
		espC	Rv3615c	Secretion system
		espD	Rv3614c	Secretion system
		espI	Rv3876	Secretion system
		espJ	Rv3878	Secretion system
		espK	Rv3879c	Secretion system
		espL	Rv3880c	Secretion system

				(continu
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
		espR	Rv3849	Secretion system
		esxA	Rv3875	Secretion system
		esxB	Rv3874	Secretion system
		mycP1	Rv3883c	Secretion system
70	ESX-2 (T7SS)	eccA2	Rv3884c	Secretion system
		eccB2	Rv3895c	Secretion system
		eccC2	Rv3894c	Secretion system
		eccD2	Rv3887c	Secretion system
		eccE2	Rv3885c	Secretion system
		espG2	Rv3889c	Secretion system
		esxC	Rv3890c	Secretion system
		esxD	Rv3891c	Secretion system
		mycP2	Rv3886c	Secretion system
21	MymA operon	adhD	Rv3086	Cell surface components
		chp	Rv3404c	Cell surface components
		fadD13	Rv3089	Cell surface components
		lipR	Rv3084	Cell surface components
		tgs4	Rv3088	Cell surface components
63	DevR/S	devS/devR	Rv3132c-Rv3133c	Regulation
		devS	Rv3132c	Regulation
32	ABC transporter	irtA	Rv1348	Copper uptake
		irtB	Rv1349	Copper uptake
60	PhoP/R	phoP	Rv0757	Regulation
		phoR	Rv0758	Regulation

				(continua
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
26	Sulfolipid-1 biosynthesis and transport	mmpL8	Rv3823c	Cell surface components
	-	papA1	Rv3824c	Cell surface components
		papA2	Rv3820c	Cell surface components
		pks2	Rv3825c	Cell surface components
62	MprA/B	mprA	Rv0981	Regulation
		mprB	Rv0982	Regulation
61	PrrA/B	prrA	Rv0902c	Regulation
		prrB	Rv0903c	Regulation
33	Pantothenate synthesis	panC	Rv3602c	Lipid and fatty acid metabolism
		panD	Rv3601c	Lipid and fatty acid metabolism
37	Lipid phosphatase	sapM	Rv3310	Lipid and fatty acid metabolism
1	Purine synthesis	purC	Rv0780	Amino acid and purine metabolism
38	Isocitrate lyase	icl	Rv0467 / Rv1915- Rv1916	Lipid and fatty acid metabolism
52	SenX3	senX3	Rv0490	Regulation
59	WhiB3	whiB3	Rv3416	Regulation
66	Antigen 85 complex	fbpA	Rv3804c	Secreted proteins
		fbpB	Rv1886c	Secreted proteins
		fbpC	Rv0129c	Secreted proteins
69	19-kD protein	lpqH	Rv3763	Secreted proteins
76	Pore-forming protein	ompA	Rv0899	Stress adaptation
45	Nucleoside diphosphate kinase	ndk	Rv2445c	Phagosome arresting
5	Lysine synthesis	lysA	Rv1293	Amino acid and purine metabolism
2	Proline synthesis	proC	Rv0500	Amino acid and purine metabolism
10	NuoG	nuoG	Rv3151	Anti-apoptosis factor

				(continua)
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
14	Exported repetitive protein	erp	Rv3810	Cell surface components
16	Heparin-binding hemagglutinin	hbhA	Rv0475	Cell surface components
17	Potassium/proton antiporter	kefB	Rv3236c	Cell surface components
39	Magnesium transport	mgtC	Rv1811	Magnesium uptake
46	Tyrosine phosphatase	ptpA	Rv2234	Phagosome arresting
49	Sigma D	sigD	Rv3414c	Regulation
50	Sigma L	sigL	Rv0735	Regulation
51	MosR	mosR	Rv0348	Regulation
55	Sigma F	sigF	Rv3286c	Regulation
56	Sigma E	sigE	Rv1221	Regulation
57	Sigma H	sigH	Rv3223c	Regulation
58	Sigma M	sigM	Rv3911	Regulation
77	Catalase-peroxidase	katG	Rv1908c	Stress adaptation
80	Iron-cofactored SOD	sodA	Rv3846	Stress adaptation
29	Heme uptake	mmpL11	Rv0203	Copper uptake
		mmpL3	Rv0206c	Copper uptake
48	Proteasome-associated proteins	mpa	Rv2115c	Protease
		pafA	Rv2097c	Protease
7	Nitrate reductase	narG	Rv1161	Anaerobic respiration
		narH	Rv1162	Anaerobic respiration
		narI	Rv1164	Anaerobic respiration
		narJ	Rv1163	Anaerobic respiration
44	P(1)-type Mn2+ transporting ATPase	ctpC	Rv3270	Protease
28	Copper exporter	ctpV	Rv0969	Copper uptake

				(conclusão)
Numero no Heatmap	Fator de Virulência	Genes Relacionados	Número Rv	Classes
65	Enhanced intracellular survival protein	eis	Rv2416c	Secreted proteins
47	Zn++ metallophrotease	zmp1	Rv0198c	Protease
72	ESX-4 (T7SS)	eccB4	Rv3450c	Secretion system
		eccC4	Rv3447c	Secretion system
		eccD4	Rv3448	Secretion system
		esxT	Rv3444c	Secretion system
		esxU	Rv3445c	Secretion system
		mycP4	Rv3449	Secretion system
41	Mce2	mce2	Rv0586-Rv0594	Mammalian cell entry (mce) operons
43	Mce4	mce4	Rv3501c-Rv3494c	Mammalian cell entry (mce) operons
42	Mce3	mce3	Rv1964-Rv1971	Mammalian cell entry (mce) operons
84	Enolase	eno	Rv1023	Enzyme
83	Receptor/GAPDH	gap	Rv1436	Adherence
86	Pyrimidine biosynthesis	pyrD	Rv2139	Nutritional virulence
87	EF-Tu	tuf	Rv0685	Adherence
81	Urease	ureF	Rv1851	Acid resistance
		ureG	Rv1852	Acid resistance
82	GroEL	groEL	Rv0440	Adherence
85	dTDP-glucose 4.6-dehydratase	Rv3468c	Rv3468c	Nutritional virulence
88	O-antigen	rfbE	Rv3781	Nutritional virulence
27	Laminin-binding protein LBP21/Hlp	Hlp	Rv2986c	Cell surface components
22	Methyltransferase	mmaA4	Rv0642c	Cell surface components