

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

### **Pós-Graduação em Biologia Celular e Molecular**

*LEANDRO SANTIAGO EMMERICK*

#### **Genômica comparativa de micobactérias ambientais da Mata Atlântica**

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Biologia Celular e Molecular

**Orientadores:** Prof. Dr<sup>a</sup>. Leila de Mendonça Lima  
Prof. Dr. Wim Maurits Sylvain Degrave

**RIO DE JANEIRO**

2018

Emmerick, Leandro Santiago.

Genômica comparativa de micobactérias ambientais da Mata Atlântica /  
Leandro Santiago Emmerick. - Rio de Janeiro, 2018.  
192 f.

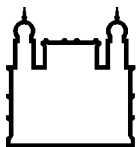
Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia  
Celular e Molecular, 2018.

Orientadora: Leila de Mendonça Lima.

Co-orientador: Wim Maurits Sylvain Degrave.

Bibliografia: f. 160-188

1. Micobactérias ambientais. 2. Filogenia molecular. 3. Vacina BCG. 4.  
Sistema de expressão. 5. Células THP-1. I. Título.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Pós-Graduação em Biologia Celular e Molecular**

***LEANDRO SANTIAGO EMMERICK***

Genômica comparativa de micobactérias ambientais da Mata Atlântica

**ORIENTADORES: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Leila de Mendonça Lima**

**Prof. Dr. Wim Maurits Sylvain Degrave**

Aprovada em: 24/04/2018

### **EXAMINADORES:**

**Prof. Dr.** Harrison Magdinier Gomes (IOC/FIOCRUZ) – Titular, Revisor e Presidente

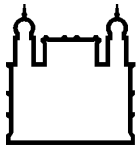
**Prof<sup>ª</sup>. Dr<sup>ª</sup>.** Sylvia Luisa Pincherle Cardoso Leão (DMIP/UNIFESP) – Titular

**Prof. Dr.** Luis Caetano Martha Antunes (ENSP/FIOCRUZ) – Titular

**Prof. Dr.** Philip Noel Suffys (IOC/FIOCRUZ) – Suplente

**Prof<sup>ª</sup>. Dr<sup>ª</sup>.** Ana Maria Mazotto de Almeida (IMPG/UFRJ) – Suplente

Rio de Janeiro, 24 de abril de 2018



Ministério da Saúde

FIOCRUZ

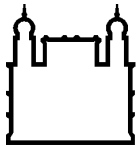
Fundação Oswaldo Cruz

## Genômica comparativa de micobactérias ambientais da Mata Atlântica

### RESUMO

Atualmente com mais de 180 espécies descritas, o gênero *Mycobacterium* está em evidência devido ao aumento no número de casos de tuberculose e de infecções oportunistas causadas por micobactérias ambientais. Embora muito se conheça sobre *M. tuberculosis*, *M. leprae* e *M. bovis*, poucos estudos se concentraram na caracterização detalhada de micobactérias ambientais, as quais encontram-se globalmente distribuídas, sendo encontradas inclusive na Mata Atlântica. Em uma época de acelerado desenvolvimento das tecnologias genômicas, a comparação entre genomas é uma ferramenta fundamental para elucidar diversos aspectos da biologia de quaisquer seres vivos, inclusive das micobactérias. Portanto, o presente estudo tem como foco a caracterização molecular detalhada de quatro micobactérias isoladas da Mata Atlântica, identificando sua capacidade metabólica, antígenos compartilhados com a cepa vacinal BCG Moreau e seu potencial biotecnológico, bem como fornecer subsídios para entender aspectos relacionados a patogenicidade desses isolados. Os isolados foram obtidos da Coleção de Bactérias da Mata Atlântica (CBMA - IOC/Fiocruz) e seus genomas foram sequenciados através das plataformas 454 GS Jr. (Roche) e HiSeq 2500 (Illumina). Após montagem e anotação dos genomas, observou-se que três isolados apresentavam genomas em torno de 6,3 Mb e um isolado com 4,8 Mb, com aproximadamente 6.000 sequências codificantes para cada. Os isolados da CBMA apresentam mais genes envolvidos na degradação de compostos aromáticos e de genes da via de metabolismo de derivados do colesterol quando comparados com micobactérias intracelulares. Através da análise filogenética utilizando os marcadores *16S rRNA*, *hsp65* e *rpoB* foi possível diferenciar os quatro isolados da Mata Atlântica e as espécies de referência e, associado a outras análises realizadas, propor estes isolados como novas espécies. Estes isolados micobacterianos apresentam diversas cutinases, identificadas no genoma e no secretoma, que podem ser atrativas como biocatalizadores de processos industriais envolvendo reações de hidrólise,

esterificação e trans-esterificação. Os quatro isolados são capazes de serem transformáveis, embora apresentem níveis diferentes de eficiência de transformação e estabilidade plasmidial, e apresentam sinais genéticos interessantes para o desenho de novos sistemas de expressão. Foram identificados diversos genes relacionados a virulência, com valores próximos aos observados para *M. tuberculosis* e *M. smegmatis* e também vários antígenos micobacterianos importantes, tanto pela anotação do genoma como pelo secretoma, tais como: Ag85, GlnA1, Mpt51, Mpt53, Mpt63, Mpt64, Mpt70 e MTB12. Ao contrário do observado em *M. smegmatis* mc<sup>2</sup>155, os quatro isolados micobacterianos apresentam diminuição na quantidade de bacilos viáveis durante a infecção em células THP-1. Desse modo, as micobactérias isoladas da Mata Atlântica (Rio de Janeiro) começam a ser caracterizadas, seu potencial biotecnológico explorado e sua infectividade estudada.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## Comparative genomics of environmental mycobacteria from the Brazilian Atlantic Forest

### **ABSTRACT**

Currently with more than 180 species described, the genus *Mycobacterium* is in evidence due to the increase in the number of cases of tuberculosis and opportunistic infections caused by environmental mycobacteria. Although much is known about *M. tuberculosis*, *M. leprae* and *M. bovis*, few studies have focused on the detailed characterization of environmental mycobacteria, which are globally distributed, being found also in the Atlantic Forest. In a time of accelerated development of genomic technologies, comparing genomes is a key tool in elucidating various aspects of the biology of any living being, including mycobacteria. Therefore, the present study focuses on the detailed molecular characterization of four mycobacteria isolated from the Brazilian Atlantic Forest, identifying their metabolic capacity, antigens shared with the BCG Moreau vaccine strain and their biotechnological potential, as well as providing subsidies to understand aspects related to the pathogenicity of these isolates. The isolates were obtained from the *Coleção de Bactérias da Mata Atlântica* (CBMA – IOC/Fiocruz) and their genomes were sequenced using the 454 GS Jr. (Roche) and HiSeq 2500 (Illumina) platforms. After assembly and annotation of the genomes, it was observed that three isolates had genomes around 6.3 Mb and one isolate with 4.8 Mb, with approximately 6,000 coding sequences in each of them. CBMA isolates show an increase in the number of genes involved in the degradation of aromatic compounds and of cholesterol metabolism when compared to intracellular mycobacteria. Through phylogenetic analysis using the *16S rRNA*, *hsp65* and *rpoB* markers it was possible to differentiate the four isolates of the Atlantic Forest and the reference species and, in association with other analyzes, propose these isolates as new species. These mycobacterial isolates have several cutinases, identified in the genome and in the secretome, that can be attractive as biocatalysts of industrial processes involving reactions of

hydrolysis, esterification and trans-esterification. The four isolates are capable of being transformed, although they exhibit different levels of transformation efficiency and plasmid stability, and present interesting genetic signals for the design of new expression systems. Several genes related to virulence were identified, with values close to those observed for *M. tuberculosis* and *M. smegmatis*, as well as many important mycobacterial antigens, both by genome and secretoma annotation, such as: Ag85, GlnA1, Mpt51, Mpt53, Mpt63, Mpt64, Mpt70 and MTB12. Contrary to *M. smegmatis* mc<sup>2</sup>155, the four mycobacterial isolates show a decrease in the amount of viable bacilli during infection in THP-1 cells. Thus, the mycobacteria isolated from the Brazilian Atlantic Forest (Rio de Janeiro) are just now starting to be characterized, their biotechnological potential explored and their infectivity studied.

Trabalho realizado no Laboratório de Genômica Funcional e Bioinformática (LAGFB/IOC), sob orientação da Dr<sup>a</sup>. Leila de Mendonça Lima e do Dr. Wim Maurits Sylvain Degrave.



Dedico este trabalho à minha esposa Carol e ao meu filho Theo que me motivaram a conclusão deste trabalho e sofreram com minha ausência, mesmo quando presente.

“Não é tolo aquele que  
abre mão do que não pode reter  
para ganhar o que não pode perder”  
Jim Elliot

Ao meu Deus que tão graciosamente me amou, conduziu e ajudou em todos os momentos da minha vida, inclusive na realização desta tese.

À minha esposa, Carol, por me amar e incentivar quando estava desanimado, por entender os vários finais de semana e feriados que passei em função da tese. Você é perfeita pra mim, Linda!

Ao nosso filho Theo com quem aprendo a cada dia, por quem desejo ser considerado como seu exemplo e a quem tenho o prazer de cuidar.

Aos meus pais, Osmar e Sueli, por serem meu referencial e exemplo para mim, por serem meus admiradores e por todo apoio e torcida. Amo vocês!

Ao meu irmão, Diego, e sua esposa, Priscilla. A dedicação e persistência de vocês nos estudos sempre será uma inspiração para mim.

Aos meus sogros, Wellington e Geise, minha cunhada, Karine, pela companhia em diversos momentos e por me amarem não como genro, mas como filho.

A todos os meus amigos e irmãos pelas conversas, risos, lágrimas, orações... Obrigado por insistirem em mim apesar de eu sempre estar muito ocupado. Vocês completam minha vida.

A todos os colegas do laboratório de Genômica Funcional e Bioinformática, especialmente ao grupo de micobactérias, por toda ajuda, incentivo, brincadeiras, broncas, ensinamentos... Neste laboratório, mais do que experimentos, eu fiz amizades que vou levar pra sempre. Vocês são demais.

A minha orientadora Dra. Leila de Mendonça Lima e Dr. Wim Degrave por abrirem as portas do laboratório para mim e acreditarem em mim enquanto aluno, por me permitir usar tecnologias tão refinadas no desenvolvimento do projeto e me proporcionar a conquista deste sonho.

A Plataforma de Sequenciamento de Ácidos Nucleicos de Nova Geração e de Espectrometria de Massas por todas as análises.

A toda equipe da Coleção de Bactérias da Mata Atlântica pelas amostras dos isolados e pela assistência que foram fundamentais para o desenvolvimento deste trabalho.

A todos da Escola Municipal Senador Nelson Carneiro pelo apoio e incentivo e por compreender as minhas faltas e dificuldades de estar mais presente.

Ao Instituto Oswaldo Cruz pelo apoio financeiro e pela oportunidade de realizar este trabalho.

## Lista de siglas e abreviaturas

- 16S rRNA – subunidade pequena da molécula de rRNA procariótico
- 2DE – eletroforese bidimensional (*2-dimensional electrophoresis*)
- ACN – acetonitrila
- Ag85 – antígeno 85
- AIC – critério de informação de Akaike (*Akaike Information Criterion*)
- AICc – AIC corrigido (*corrected Akaike Information Criterion*)
- ATP – trifosfato de adenosina (*adenosine triphosphate*)
- BAAR – bacilo álcool-ácido resistente
- BCG – Bacilo de Calmette e Guérin
- BIC – critério de informação Bayesiana (*Bayesian Information Criterion*)
- BLAST – ferramenta de busca de alinhamento local básico de proteínas (*protein Basic Local Alignment Search Tool*)
- BLLF1 – glicoproteína do envelope externo dos vírus da Herpes
- CABOG – montador Celera com a melhor sobreposição de gráficos (*Celera Assembler with Best Overlap Graph*)
- CBAS – Coleção de Bactérias do Ambiente e Saúde
- CBB – *coomassie brilliant blue*
- CBMA – Coleção de Bactérias da Mata Atlântica
- CCD – dispositivo de carga acoplada (*charge-coupled device*)
- CDS – sequência codificante (*coding sequence*)
- CFP – proteína do filtrado de cultura (*culture filtrate protein*)
- CFU – unidades formadoras de colônias (*colony forming units*)
- CGEN – Conselho de Gestão de Patrimônio Genético
- cgMLSA – análise de sequência multilocus do core genome
- CISA – integrador de *contigs* para montagem de sequências (*Contig Integrator for Sequence Assembly*)
- CPEC – clonagem através da extensão circular da polimerase (*Circular Polymerase Extension Cloning*)
- ddNTP – dideoxirribonucleotídeo
- DNA – ácido desoxirribonucleico (*deoxyribonucleic acid*)
- dNTP – deoxinucleotídeos

DO – densidade ótica

DT – teoria de decisão (decision theory)

dTHP-1 – monócitos THP-1 diferenciados em macrófagos

DTT – Ditiotreitól

EC – comissão de enzimas (*Enzyme Commission*)

*eis* – melhora a sobrevivência intracelular (*enhance the intracellular survival*)

ENSP – Escola Nacional de Saúde Pública

ESAT – alvo antigênico secretado precocemente (*early secretory antigenic target*)

ESX – sistema de secreção tipo VII

FAPERJ – Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro

FINEP – Financiadora de Estudos e Projetos

FIOCRUZ – Fundação Oswaldo Cruz

Gb – gigabase

GC – conteúdo de guanina e citosina

GlnA1 – glutamino sintetase

GS – *genome sequencer*

GTR – tempo reversível geral (*General Time Reversible*)

GYT – glicerol, extrato de levedura e triptona (*glycerol, yeast extract and tryptone*).

H37Rv – *Mycobacterium tuberculosis* cepa H37Rv

HIV – Vírus da Imunodeficiência Humana (*Human Immunodeficiency Virus*)

HPLC – cromatografia líquida de alta performance (high performance liquid chromatography)

*hsp* – proteína de choque térmico (*heat shock protein*)

ICMBio – Instituto Chico Mendes de Conservação da Biodiversidade

INDEL – mutação do tipo inserção ou deleção

INPE – Instituto Nacional de Pesquisas Espaciais

IOC – Instituto Oswaldo Cruz

IPG – gradiente imobilizado de pH (*immobilised pH gradient*)

*ITS* – espaçador transcrito interno (*internal transcribed spacer*)

K2P – Kimura 2 parâmetros;

kanR – resistência ao antibiótico canamicina

*lacZ* –  $\beta$ -galactosidade de *E. coli*.

LAGFB – Laboratório de Genômica Funcional e Bioinformática

LB – meio de cultura Luria Bertani

LB/ágar – LB acrescido de 1,5 % (p/v) de bacto-ágar (Difco)

LB/Tw – LB contendo Tween 80 0,05 % (v/v)

LpqH – uma lipoproteína de 19 kDa

MAC – complexo *Mycobacterium avium* (*M. avium complex*)

MAFFT – alinhamento múltiplo usando a transformada rápida de Fourier (*Multiple Alignment using Fast Fourier Transform*)

MALDI – ionização por dessorção a laser assistida por matriz (*matrix assisted laser desorption ionization*)

mc<sup>2</sup>155 – *Mycobacterium smegmatis* mc<sup>2</sup>155

MCS – sítio múltiplo de clonagem (*multiple cloning site*)

MDM – macrófagos derivados de monócitos

ME – evolução mínima (*minimum evolution*)

MEGA – análise genética de evolução molecular (*molecular evolutionary genetics analysis*)

MIRA – montador de *reads* por mimetismo inteligente (*mimicking intelligent read assembly*)

ML – máxima verossimilhança (*maximum likelihood*)

MMA – Ministério do Meio Ambiente

MNT – micobactérias não tuberculosas

MOI – multiplicidade de infecção (*multiplicity of infection*)

MP – máxima parcimônia (*maximum parcimony*)

Mpt – proteína de *M. tuberculosis*

MS/MS – espectrometria de massas *in tandem*

MW – peso molecular (*molecular weight*)

NB – nível de biossegurança

NCBI – Centro Nacional para Informação em Biotecnologia (*National Center for Biotechnology Information*)

NEB – New England Biolabs Inc.

NGS – sequenciamento de nova geração (*next generation sequencing*)

NJ – *Neighbour-Joining*

*oriE* – origem de replicação em *E. coli*

*oriM* – origem de replicação em micobactérias

p/v – peso/volume

PAGE – eletroforese em gel de poliacrilamida (*polyacrylamide gel electrophoresis*)

PAH – hidrocarbonetos policíclicos aromáticos (*polycyclic aromatic hydrocarbon*)

PARNASO – Parque Nacional da Serra dos Órgãos

PATRIC – Centro Integrado de Recursos de Sistemas Patogênicos (*Pathosystems Resource Integration Center*)

pb – pares de base

PCR – reação em cadeia da polimerase (*polymerase chain reaction*)

PEES – penicilina 10.000 U e estreptomicina 10 mg/mL

pH – potencial hidrogeniônico

Pi – fosfato inorgânico

PMA – acetato de forbol-miristila (*phorbol myristate acetate*)

PRA – PCR restriction analysis

QUAST – ferramenta para acesso a qualidade de genomas montados (*quality assessment tool for genome assemblies*)

RAST – anotação rápida usando a tecnologia de subsistemas (*Rapid Annotations using Subsystem Technology*)

RD – região de diferença ou região de deleção

*recA* – sistema de recombinação e reparo do DNA

RNA – ácido ribonucléico (*ribonucleic acid*)

RNS – espécies reativas de nitrogênio (*reactive nitrogen species*)

ROS – espécies reativas de oxigênio (*reactive oxygen species*)

RPMI – *Roswell Park Memorial Institute medium*

RPMI/SFB – meio RPMI acrescido de 10 % de SFB

*rpoB* – subunidade  $\beta$  da RNA polimerase

RPT01J – Plataforma de Sequenciamento de Ácidos Nucléicos de Nova Geração

RPT02A – Plataforma de Espectrometria de Massas

rRNA – RNA ribossomal

SDS – dodecil sulfato de sódio (*sodium dodecyl sulfate*)

Sec – via de secreção geral

SFB – soro fetal bovino

SFF – *standard flowgram format*



SLIC – clonagem independente da sequência e ligação (*sequence and ligation-independent cloning*)

SLiCE – (Seamless Ligation Cloning Extract)

SMRT – molécula única em tempo real (*single molecule real time*)

SNP – polimorfismos de nucleotídeo único (*single nucleotide polymorphism*)

sod – superóxido dismutase

SOLiD – sequenciamento pela ligação e detecção de oligonucleotídeos (sequencing by oligonucleotide ligation and detection)

SPAdes – montador de genomas São Petesburgo (*St. Petersburg genome Assembler*)

Tat – via de translocação twin-arginina

TCA – ácido tricloroacético (*trichloroacetic acid*)

TE – solução de Tris-HCl (100 mM) e EDTA (50 mM), pH 8,0

TetR – proteína repressora de tetraciclina (TetR)

THP-1 – linhagem celular leucêmica monocítica humana

TOF – tempo de voo (time of flight)

Tris – Tris-hidroxi-metil-amino-metano

tRNA – RNA transportador

UFRJ – Universidade Federal do Rio de Janeiro

UPGMA – método de agrupamento em pares ponderados usando médias aritméticas (*unweighted pair group method using arithmetic averages*)

USB – tipo de conexão entre aparelhos (*Universal Serial Bus*)

v/v – volume/volume

WFCC – Federação Mundial para Coleções de Culturas (*World Federation for Culture Collections*)

WHO – Organização Mundial da Saúde (*World Health Organization*)

X-Gal – 5-bromo-4-cloro-3-indoxil- $\beta$ -D-galactopiranosídeo

ZMW – *Zero-Mode-Waveguide*

## LISTA DE QUADROS E TABELAS

### INTRODUÇÃO

1.1	Características físico-químicas do solo do PARNASO	3
-----	--	---

### METODOLOGIA

3.1	Etapas para a focalização isoeétrica	39
3.2	Método de coloração com coomassie G-250 coloidal	40
3.3	Parâmetros utilizados nas buscas do Mascot	42
3.4	Oligonucleotídeos usados na construção I	46
3.5	Ciclagens para a reação de PCR	47
3.6	Oligonucleotídeos usados na construção II	48

### RESULTADOS

4.1	Sequenciamento dos isolados	52
4.2	Montagem dos genomas sequenciados no 454 GS Jr.	56
4.3	Montagem dos genomas sequenciados no HiSeq	56
4.4	Anotação dos genomas sequenciados no 454 GS Jr.	58
4.5	Anotação dos genomas sequenciados no HiSeq	58
4.6	Genes relacionados a virulência e patogenicidade	58
4.7	Mapeamento metabólico	62
4.8	Antígenos compartilhados	66
4.9	Análise do jModelTest2	72
4.10	Identificação através do MALDI TOF-TOF	81
4.11A	Identificação pelo Orbitrap dos <i>spots</i> de 271	81
4.11B	Identificação pelo Orbitrap dos <i>spots</i> de 294	82

<b>4.12</b>	Resultado Q-Exactive	84
<b>4.13</b>	Proteínas compartilhadas entre os isolados	85
<b>4.14</b>	Proteínas hipotéticas	88
<b>4.15</b>	Algumas proteínas hipotéticas identificadas	88
<b>4.16</b>	Quantificação <i>label-free</i>	90
<b>4.17</b>	Capacidade de transformação	92
<b>4.18</b>	Tamanho esperado para os fragmentos da construção I	99
<b>4.19</b>	Tamanho esperado para os fragmentos da construção II	102
<b>4.20</b>	Quantificação dos lotes-semente	103

## **ANEXO**

<b>8.1</b>	Sequências utilizadas para a análise filogenética	164
------------	---	-----

## LISTA DE FIGURAS

### INTRODUÇÃO

1.1	Sítios de coleta das amostras da CBMA	4
1.2	Micobactérias descritas até janeiro de 2018	6
1.3	Filogenia do gênero <i>Mycobacterium</i>	9
1.4	Genoma de <i>M. tuberculosis</i>	16
1.5	Sequenciamento de primeira geração	18
1.6	Sequenciamento de segunda geração	22
1.7	Sequenciamento de terceira geração	24

### METODOLOGIA

3.1	Fluxograma de trabalho	32
-----	------------------------	----

### RESULTADOS

4.1	Extração do DNA genômico	51
4.2	Gráfico de qualidade do sequenciamento pelo 454 GS Jr. gerados pelo FastQC	54
4.3	Gráfico de qualidade do sequenciamento pelo HiSeq 2500 gerados pelo FastQC	55
4.4	Distribuição gênica em subsistemas por isolado após análise no RAST Server dos dados gerados pelo 454 GS Jr.	59
4.5	Distribuição gênica em subsistemas por isolado após análise no RAST Server dos dados gerados pelo 454 GS Jr.	60
4.6	Genoma dos isolados da Mata Atlântica	61
4.7	Via de degradação do benzoato	64
4.8	Teste dos modelos estatísticos para geração das árvores	67
4.9	Árvores filogenéticas	74

<b>4.10</b>	Perfil bidimensional das proteínas secretadas	79
<b>4.11</b>	<i>Spots</i> analisados pelo Orbitrap	80
<b>4.12</b>	Identificação de proteínas por três espectrômetros de massas	84
<b>4.13</b>	Proteínas compartilhadas entre os quatro isolados da Mata Atlântica	87
<b>4.14</b>	Vetor pUS972	91
<b>4.15</b>	Micobactérias transformadas com o plasmídeo pUS989	92
<b>4.16</b>	Análise por digestão com enzimas de restrição	93
<b>4.17</b>	Sintenia 294_01260	95
<b>4.18</b>	Proteína 294_01260	95
<b>4.19</b>	Esquema da Construção I	98
<b>4.20</b>	Fragmentos amplificados por PCR da Construção I	99
<b>4.21</b>	Transformação da Construção I	100
<b>4.22</b>	Esquema da Construção II	101
<b>4.23</b>	Fragmentos amplificados por PCR e transformação da Construção II	102
<b>4.24</b>	Morfologia de colônia e contagem de CFU	103
<b>4.25</b>	Taxa de infecção e viabilidade dos isolados ambientais após infecção em células dTHP-1	105
<b>4.26</b>	Morfologia de colônia das bactérias retiradas dos macrófagos no tempo de 48 h após a infecção	106

## ÍNDICE

Conteúdo	Página
<b>RESUMO</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>v</b>
<b>LISTA DE SIGLAS E ABREVIATURAS</b>	<b>xii</b>
<b>LISTA DE QUADROS E TABELAS</b>	<b>xvii</b>
<b>LISTA DE FIGURAS</b>	<b>xix</b>
<b>1. INTRODUÇÃO</b>	<b>1</b>
1.1. Micobactérias	4
1.1.1 Identificação	6
1.1.2 Classificação	9
1.1.3 Habitat	11
1.1.4 Micobactérias da Mata Atlântica	13
1.2 Genomas	14
1.2.1 Sequenciamento de primeira geração	17
1.2.2 Sequenciamento de segunda geração	19
1.2.3 Sequenciamento de terceira geração	23
1.2.4 Genômica comparativa	25
1.3 Biotecnologia	26
1.4 Sistemas de expressão	27
<b>2. OBJETIVOS</b>	<b>31</b>
2.1 Objetivo geral	31
2.2 Objetivos específicos	31

<b>3.</b>	<b>METODOLOGIA</b>	<b>32</b>
3.1	Fluxograma geral do trabalho	32
3.2	Escolha dos isolados micobacterianos	32
3.3	Cultivo de micobactérias	33
3.4	Coloração de Ziehl-Neelsen	33
3.5	Extração de DNA genômico	33
3.6	Sequenciamento dos genomas	34
3.7	Montagem e anotação 454 GS Jr.	34
3.8	Montagem e anotação HiSeq	36
3.9	Mapeamento metabólico	37
3.10	Análise filogenética	37
3.11	Antígenos compartilhados	38
3.12	Obtenção de proteínas secretadas	38
3.13	Eletroforese bidimensional	38
3.14	Excisão e processamento dos <i>spots</i>	40
3.15	Análise no MALDI e no Orbitrap	41
3.16	Secretoma	42
3.17	Proteínas hipotéticas	42
3.18	Quantificação <i>label-free</i>	42
3.19	Capacidade de transformação	42
3.20	Análise de sequências reguladoras, promotoras e terminadoras	43
3.21	Cultivo de <i>E. coli</i>	44
3.22	Extração de DNA plasmidial	44
3.23	Preparo de células eletrocompetentes	44
3.24	Digestão do vetor pUS989	45
3.25	Construção do vetor I	45

3.25.1	PCRs	45
3.25.2	Digestão	46
3.25.3	Reação de Gibson e transformação	46
3.25.4	PCR de colônia	46
3.26	Construção do vetor II	48
3.27	Cultura de células THP-1	48
3.28	Diferenciação dos monócitos THP-1 em macrófagos	49
3.29	Quantificação dos lotes-semente	49
3.30	Viabilidade intracelular dos isolados	49
<b>4.</b>	<b>RESULTADOS</b>	<b>51</b>
4.1	Análise genômica	51
4.1.1	Sequenciamento	51
4.1.2	Qualidade das sequências	52
4.1.3	Montagem	53
4.1.4	Anotação	56
4.1.5	Mapeamento metabólico	62
4.1.6	Antígenos compartilhados com BCG	65
4.1.7	Análise filogenética	65
4.2	Análise do secretoma	78
4.2.1	Perfil de secreção de proteínas	78
4.2.2	Identificação de proteínas	78
4.3	Construção de um novo sistema de expressão	91
4.3.1	Capacidade de transformação e estabilidade plasmidial	91
4.3.2	Análise de sequências reguladoras, promotoras e terminadoras	93
4.3.3	Construção de um novo vetor	96



4.3.3.1	Construção I – mCherry	97
4.3.3.2	Construção II – Mpt64	100
4.4	Sobrevivência intracelular dos isolados ambientais	102
<b>5</b>	<b>DISCUSSÃO</b>	<b>107</b>
5.1	Análise genômica	108
5.2	Análise do secretoma	123
5.3	Construção de um novo sistema de expressão	126
5.4	Sobrevivência intracelular dos isolados ambientais	129
<b>6.</b>	<b>CONCLUSÕES</b>	<b>133</b>
<b>7.</b>	<b>BIBLIOGRAFIA</b>	<b>135</b>
<b>8.</b>	<b>ANEXO</b>	<b>164</b>

## 1. Introdução

O Brasil é considerado o país mais megadiverso do mundo, principalmente devido ao número de espécies e níveis de endemismo (Mittermeier *et al.*, 1997; Brandon *et al.*, 2005), e a Mata Atlântica, um dos biomas mais ricos do planeta (Brandon *et al.*, 2005). Apesar de haver apenas 11,4 % de cobertura florestal nativa (SOS Mata Atlântica & INPE, 2010), a Mata Atlântica é um dos 25 *hotspots* mundiais de biodiversidade, a segunda maior floresta ombrófila<sup>1</sup> densa do continente americano, e abriga mais de 8.000 espécies endêmicas de plantas vasculares e 560 de vertebrados (Myers *et al.*, 2000). É extremamente heterogênea, se estendendo em uma ampla faixa latitudinal (4° a 32° S) e longitudinal (35° a 60° W) com diferentes zonas climáticas, altitude (nível do mar a 2.900 m), tipo e profundidade dos solos, temperatura média do ar, pluviosidade (1.000 mm a 4.000 mm/ano) (Barbosa & Thomas, 2002; Mantovani, 2003).

Atualmente, a preservação da Mata Atlântica ocorre através das Unidades de Conservação, com destaque para o Parque Nacional da Serra dos Órgãos (PARNASO), terceiro parque nacional mais antigo do país (criado em 30 de novembro de 1939). O PARNASO é uma Unidade de Conservação Federal de Proteção Integral<sup>2</sup>, subordinada ao Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), e com aproximadamente 200 km<sup>2</sup>, estando inserido nos municípios de Magé, Petrópolis, Teresópolis e Guapimirim, ocupando posição central no Corredor da Serra do Mar. O clima no Parque é tropical superúmido (com 80 a 90 % de umidade relativa do ar), com média anual de temperatura que varia de 13 °C a 23 °C e variação pluviométrica de 1.700 a 3.600 mm. Apresenta relevo bastante acidentado com grande variação de altitude (de 80 m a 2.263 m) e pelo menos cinco tipos de solos e diversas associações entre eles. Os mananciais das duas principais bacias hidrográficas fluminenses, a do Paraíba do Sul e o conjunto de bacias hidrográficas que drenam em direção a Baía de Guanabara, estão situados no Parque. O PARNASO abriga cerca de 2.800 espécies vegetais, mais de 750 de vertebrados e mais de 500 de invertebrados (<http://www.icmbio.gov.br/parnaserradosorgaos>).

---

<sup>1</sup> Antigamente chamadas de Floresta Pluvial, caracterizada por vegetação de folhas largas e perenes e chuva abundante

<sup>2</sup> Unidades de Proteção Integral são dedicadas exclusivamente à preservação do ambiente natural, sendo proibida a presença de populações permanentes e vedadas as atividades econômicas.

O PARNASO é a unidade de conservação federal com o maior número de pesquisas autorizadas no país, segundo o Sistema de Autorização e Informação em Biodiversidade. A sua proximidade à cidade do Rio de Janeiro (90 km), fácil acesso, boa infraestrutura de apoio à pesquisa, possibilidade de acesso ao banco de dados de gestão do conhecimento científico produzido no Parque e dados climáticos obtidos em sua rede de estações meteorológicas, e, principalmente, a alta biodiversidade da região são alguns dos atrativos para o desenvolvimento de pesquisas no Parque (<http://www.icmbio.gov.br/parnaserradosorgaos>; MMA, 2007).

Dentre os diversos projetos de pesquisa aprovados que incluem essa região, destaca-se o projeto “Biodiversidade da Mata Atlântica do Rio de Janeiro” (2007), iniciativa da FAPERJ/FINEP e FIOCRUZ, que culminou na criação da “Coleção de Bactérias da Mata Atlântica” (CBMA). Essa Coleção foi incorporada à Coleção de Bactérias do Ambiente e Saúde (CBAS) e está localizada no Laboratório Interdisciplinar de Pesquisas Médicas do Instituto Oswaldo Cruz, FIOCRUZ. A CBAS é filiada à *World Federation for Culture Collections*, WFCC, sob o registro WDCM 958 e foi credenciada como Fiel Depositária de Amostras de Patrimônio Genético pelo Conselho de Gestão de Patrimônio Genético (CGEN/ Ministério do Meio Ambiente), de acordo com o Aviso de Credenciamento nº 097/2013/SECEX/CGEN, publicado no Diário Oficial da União em 12 de setembro de 2013 (<http://cbas.fiocruz.br/index>).

Amostras de solo contendo 1 kg dos 10 cm da camada superior após retirada da cobertura vegetal foram extraídas ao longo de oito sítios de coleta na trilha de aproximadamente 40 km que liga Petrópolis a Teresópolis (figura 1.1). As bactérias depositadas na CBMA foram isoladas após sucessivas diluições em soluções salinas e plaqueamento em seis tipos de meios de cultura: ágar nutriente, ágar triptona de soja, ágar MacConkey e ágar actinomiceto (Difco), enriquecido de celulose e enriquecido de lignina (*in house*) (<http://cbas.fiocruz.br/index?history>; Bruce *et al.*, 2010).

De acordo dados do site do PARNASO (<http://www.icmbio.gov.br/parnaserradosorgaos/atributos-naturais.html?id=47:solos&catid=3:artigos>) os cinco tipos de solo principais do Parque são: os latossolos Amarelo, Vermelho e Vermelho Amarelo (solos profundos e presentes na maior parte do Parque), litossolo (característico de áreas de maiores altitudes) e solos hidromórficos (encontrados nas

regiões mais baixas, próximas as margens dos rios). No entanto, em estudo realizado por Martins e colaboradores (2007) que analisaram a relação do solo com as formas de relevo no PARNASO após coleta de 80 amostras de solo de 40 pontos diferentes na trilha Petrópolis-Teresópolis, foram identificadas sete classes de solo diferentes das descritas acima (tabela 1.1). É importante destacar que houve grande mudança no Sistema Brasileiro de Classificação de Solos<sup>3</sup> (EMBRAPA, 2006) e que as informações contidas no site do PARNASO encontram-se desatualizadas.

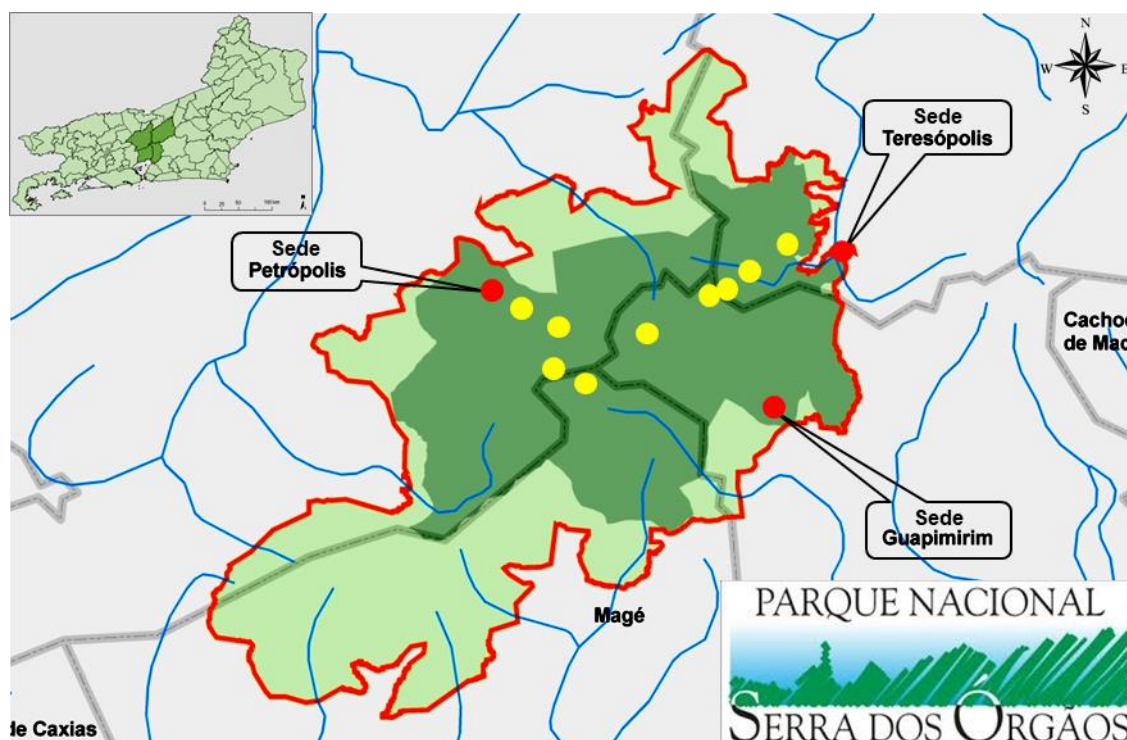
**Tabela 1.1: Características físico-químicas do solo do PARNASO.** Análises realizadas a partir de amostras dos 20 cm da camada superior do solo. Valores de alumínio (Al), cálcio (Ca), magnésio (Mg), potássio (K) e hidrogênio + alumínio (H + Al) em cmol/dm<sup>3</sup> e matéria orgânica (MO) em g/Kg

Classe de solo	Silte/Argila	pH H <sub>2</sub> O	Al	Ca	Mg	K	H + Al	MO
Neossolo Litólico distro-úmbrico	0,60	4,22	2,88	0,08	0,09	0,10	16,30	78,7
Neossolo Litólico	0,80	4,95	3,27	0,22	0,18	0,21	23,02	123,7
Cambissolo Háptico Ta distrófico	0,36	4,09	1,39	0,09	0,08	0,05	9,00	33,4
Gleissolo Melânico alítico	1,00	4,28	1,39	0,09	0,08	0,16	19,34	157,6
Argissolo Bruno-Acinzentado alítico	0,86	4,70	4,05	0,10	0,13	0,18	19,20	133,6
Argissolo Amarelo distrófico	0,44	4,40	4,05	1,39	0,03	0,11	8,72	34,44
Latosolo Amarelo distrófico	0,47	4,78	0,81	1,52	0,20	0,002	5,42	14,3

Um primeiro estudo realizado com as amostras que deram origem à CBMA analisou a microbiota de seis dos oito pontos de coleta: bonfim amarelo, bonfim vermelho, ajax, torto, campo úmido e cavalinho, que apresentam solos típicos, representativos e abundantes da Mata Atlântica. Em análise independente de cultura, foram identificadas quase 900 sequências diferentes do gene *16S rRNA* e mais de 20 gêneros bacterianos (Bruce *et al.*, 2010; EMBRAPA Cerrados, 2008), dentre os quais destaco o gênero *Mycobacterium*, cujas espécies têm sido

<sup>3</sup> Os solos são classificados e organizados em 6 níveis categóricos, mas o 5<sup>o</sup> e 6<sup>o</sup> nível categórico ainda se encontram em discussão. Os quatro primeiros níveis são denominados de ordens, subordens, grandes grupos e subgrupos. A classificação pedológica nacional consiste numa evolução do antigo sistema norte-americano, formulado na primeira metade do século XX. Desde a década de 1950, os estudos do solo brasileiro foram realizando modificações e acréscimos nesse sistema, resultando na publicação do Sistema Brasileiro de Classificação de Solos (EMBRAPA, 1998 e 1999). Uma segunda edição foi publicada em 2006 (EMBRAPA), com diversas modificações importantes, desde o primeiro nível categórico. A última edição (EMBRAPA, 2013) foi publicada com a incorporação de algumas classes de solos.

identificadas e estudadas desde o isolamento de *M. leprae* em 1873 (Rastogi *et al.*, 2001). Atualmente a CBMA possui mais de 800 isolados em sua coleção.



**Figura 1.1: Sítios de coleta das amostras da CBMA.** Mapa PARNASO mostrando o limite dos municípios onde o Parque está inserido. A área do Parque está delimitada por uma linha em vermelho e as sedes marcadas por pontos vermelhos. Os sítios de coleta representados por pontos amarelos. Fonte: Adaptado do site do Ministério do Meio Ambiente < [http://www.mma.gov.br/img/ascom/fotos/serradosorgaos\\_mapa.jpg](http://www.mma.gov.br/img/ascom/fotos/serradosorgaos_mapa.jpg) >.

## 1.1. Micobactérias

O gênero *Mycobacterium* compreende mais de 180 espécies e subespécies oficialmente reconhecidas<sup>4</sup>, sendo *M. tuberculosis*, o agente etiológico da tuberculose em humanos, considerada a espécie de referência do gênero, de acordo com a *List of Prokaryotic Names with Standing in Nomenclature* (<http://www.bacterio.net/Mycobacterium.html>). Taxonomicamente, é o único gênero dentro da família *Mycobacteriaceae*, pertencente à subordem *Corynebacterineae* da ordem dos *Actinomycetales* (Whitman, 2012).

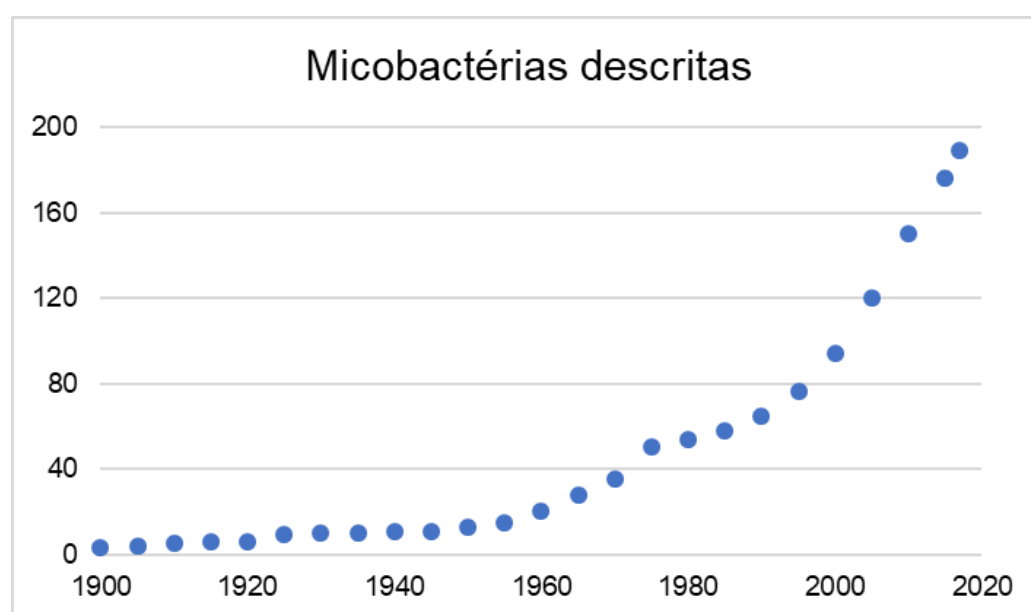
<sup>4</sup> O site *List of Prokaryotic Names with Standing in Nomenclature* considera como oficialmente reconhecidos os nomes citados em *Approved Lists of Bacterial Names* ou publicados no *International Journal of Systematic Bacteriology* ou *International Journal of Systematic and Evolutionary Microbiology*. Este site é atualizado no início de cada mês desde 2006. Além dessas espécies, existem diversas outras já identificadas, como *M. indicus pranii* e *M. mungi* (<http://www.bacterio.net/>; Tortoli, 2014;).

Caracteriza-se principalmente por apresentar um distinto envelope celular constituído por três segmentos: I) cápsula (camada mais externa), constituída principalmente de proteínas, com pequenas quantidades de carboidratos e uma quantidade mínima de lipídeos; II) uma complexa parede celular, constituída pela micomembrana (rica em ácidos micólicos), arabinogalactana e peptideoglicana; e III) a tradicional membrana plasmática. A parede celular das micobactérias contém mais de 60 % de lipídeos, comparado com apenas 20 % da parede celular rica em lipídeos das Gram-negativas. Essa abundância de lipídeos, que chega a 40 % do peso seco de *M. tuberculosis*, é a principal característica das micobactérias (Chiaradia *et al.*, 2017). Outra característica chave deste gênero é o elevado conteúdo de guanina e citosina (GC) no genoma (entre 60 e 70 %, com exceção para *M. leprae*, 58 %). *M. tuberculosis*, por exemplo, apresenta cerca de 65,6 % de conteúdo GC, acarretando em uma alta frequência com que o códon de iniciação GTG é encontrado no genoma (35 %), além de uma composição diferenciada do proteoma, onde há prevalência de aminoácidos codificados por códons ricos em GC, como alanina, glicina, prolina e arginina (Cole *et al.*, 1998). Embora sejam estruturalmente relacionadas às bactérias Gram-positivas, as micobactérias não se classificam como tal, devido à peculiaridade da sua parede celular rica em lipídeos, que frequentemente não retém cristal violeta e não aparecem após coloração de Gram (Barrera, 2007). São caracterizadas como bacilos álcool-ácido resistentes (BAAR), pois resistem à ação descorante de uma solução álcool-ácida durante a coloração de Ziehl-Neelsen (Daffé & Draper, 1998).

Além disso, as micobactérias são microaerofílicas, imóveis (sem flagelo) e não encapsuladas, apresentam em sua maioria morfologia bacilar ou coco bacilar ou, raramente, filamentosa, onde os bacilos são pequenos, finos, retos ou ligeiramente encurvados, apresentando-se isolados, aos pares ou em pequenos aglomerados (Ventura *et al.* 2007). Embora seja descrito como não esporulado (Tortora, 2005; Uhía *et al.*, 2015), análises de bioinformática indicam que o genoma de *M. tuberculosis* possui genes semelhantes aos envolvidos na esporulação de *B. subtilis* (Piggot & Hilbert, 2004). Evidências de dados microscópicos e de coloração de esporos indicam uma provável esporulação em *M. marinum* e *M. bovis* BCG (Ghosh *et al.*, 2009). No entanto, a ocorrência da esporulação em micobactérias ainda é controversa, visto a necessidade de compreensão das condições que originam o processo (Uhía *et al.*, 2015).

### 1.1.1. Identificação

Durante 100 anos, os estudos taxonômicos em micobactérias foram realizados baseados unicamente em aspectos fenotípicos, como álcool-ácido resistência, tempo e condições de crescimento, aspectos da colônia, morfologia microscópica, resistência a antibióticos e testes bioquímicos (Rogal *et al.*, 1990 A; Wayne & Sramek, 1992; Tortoli, 2003). A partir da década de 1990, o foco dos estudos foi redirecionado para aspectos genotípicos, por terem uma maior precisão e serem mais rápidos, tendo como alvo regiões altamente conservadas do genoma que contém sequências hipervariáveis nas quais deleções, inserções ou substituições de nucleotídeos seriam específicas para cada espécie (Rogal *et al.*, 1990 A; Dobner *et al.*, 1996; Tortoli, 2003). Atualmente, o sequenciamento desses genes conservados é o método de referência para identificação de micobactérias. Os principais genes escolhidos para essa tarefa são: *16S rRNA*, *hsp65*, *sodA*, *recA*, a região intergênica *ITS* e outros. Essa nova metodologia para identificação de micobactérias associada ao crescente interesse nas micobactérias não-tuberculosas (MNT, que não pertencem ao complexo *M. tuberculosis* e *M. leprae*) resultou em um aumento exponencial no número de novas espécies descritas nos últimos 20 anos (Tortoli, 2014) (figura 1.2)



**Figura 1.2: Micobactérias descritas até janeiro de 2018.** Cada ponto representa o total cumulativo de espécies e subespécies descritas a cada cinco anos. Gráfico criado com base nos dados da *List of Prokaryotic Names with Standing in Nomenclature* (<http://www.bacterio.net/Mycobacterium.html>).

O gene que codifica a subunidade pequena da molécula de RNA do ribossomo procariótico, *16S rRNA*, possui aproximadamente 1.500 pares de base e é o gene mais amplamente usado para análises filogenéticas em bactérias, principalmente por estar presente em quase todas as bactérias, por manter sua função ao longo do tempo, sugerindo que as variações no DNA sejam indicativas de evolução, e por ser grande o suficiente para análises estatísticas. No entanto o uso deste gene tem suas limitações. Estudos sugerem que, após sequenciamento do gene *16S rRNA* para isolados desconhecidos, é possível identificar o gênero em mais de 90 % dos casos, porém, a identificação de espécie varia entre 65 e 83 % e cerca de 1 a 14 % dos isolados permanecem sem identificação. Essas dificuldades ocorrem devido ao reconhecimento de novas taxa, pouca informação em banco de dados, alta (ou mesmo total) identidade com a sequência de outras espécies ou mesmo sequências variadas identificadas como a mesma espécie ou complexo (Tortoli, 2003; Janda & Abbott, 2007). Nas micobactérias, o gene *16S rRNA* (chamado *rrs* para *M. tuberculosis* H37Rv) começou a ser utilizado em 1990 (Rogal *et al.*, 1990 A e B), quando foi avaliada a relação filogenética entre aproximadamente 20 espécies micobacterianas a partir de sequências geradas por PCR. Foi observado a presença de duas regiões principais (posições 130 a 210 e 430 a 500 do gene de *Escherichia coli*) que são conservadas em todas as espécies, mas também hipervariáveis, sendo específicas para cada espécie. Novos estudos reforçaram os resultados obtidos por Rogal e colaboradores. Além disso, não existe um número mínimo de nucleotídeos diferentes para propor uma nova espécie em micobactérias: espécies diferentes podem apresentar poucas bases diferentes ou, diante de claras diferenças fenotípicas, não diferir em base alguma. É possível também que uma mesma espécie apresente diferenças de até sete pares de base (Tortoli, 2003). No entanto, aceita-se como regra geral que valores de identidade menores que 98,7 % deste gene, entre isolados, seria indicativo de diferentes espécies, enquanto que valores menores que 95,0 % indicariam um novo gênero (Stackebrandt & Ebers, 2006). Já foi descrito também que as espécies de crescimento lento tendem a apresentar apenas uma cópia deste gene, enquanto que as de crescimento rápido apresentam duas (Tortoli, 2003).

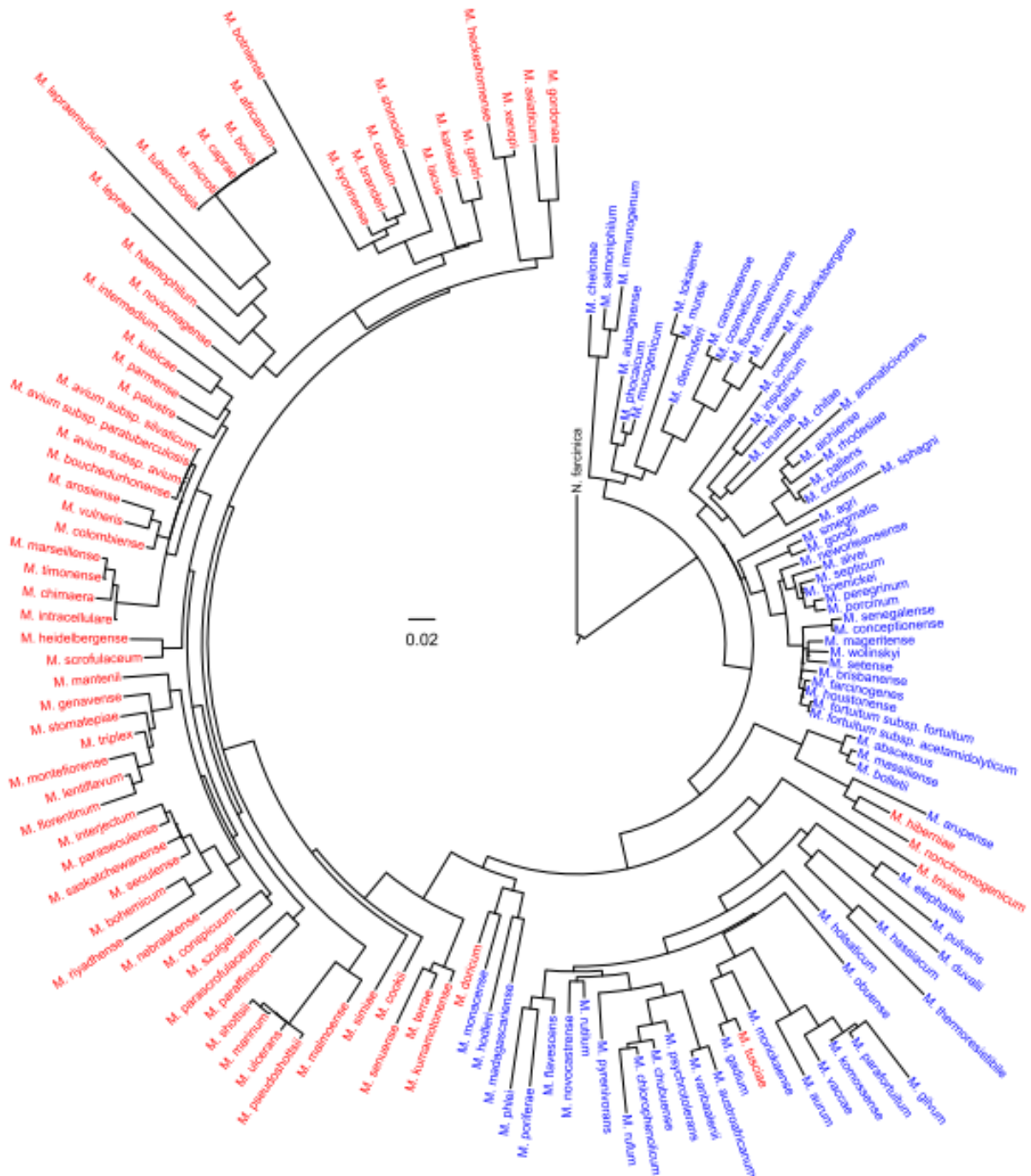
O gene *hsp65*, codifica a proteína de choque térmico de 65 kDa, é composto por aproximadamente 1.600 pares de bases (Shinnick, 1987) e tem sido proposto



para identificação de micobactérias em nível de espécie desde a década de 1990 (Telenti *et al.*, 1993). É mais conservado que os demais *loci* utilizados para identificação, exceto o gene *16S rRNA*, mas também não é capaz de diferenciar os membros do complexo *M. tuberculosis* (com exceção do *M. africanum*) nem subespécies dentro de *M. avium* e *M. fortuitum* (Dai *et al.*, 2011A). Apresenta duas regiões hipervariáveis (posições 624 a 664 e 683 a 725 do gene de H37Rv), que fazem parte do chamado fragmento Telenti (formado por 439 pares de base da posição 398 a 836), usado na grande maioria das identificações e propostas filogenéticas (figura 1.3) (Tortoli, 2003; Dai *et al.*, 2011B).

O gene *rpoB* codifica para a subunidade  $\beta$  da RNA polimerase, possui apenas uma cópia e começou a ser usado como ferramenta para identificar micobactérias no final da década de 1990 (Gingeras *et al.*, 1998). Novos estudos se seguiram utilizando a sequência parcial do gene (~ 400 pb), os quais mostraram-se eficazes na diferenciação de diversas espécies e isolados clínicos (Kim *et al.*, 1999; 2001). Desde então, as novas espécies descritas trazem informações deste gene. No entanto, a região da sequência parcial do gene escolhida não se mostrou a mais adequada. Os novos estudos passaram a utilizar a sequência de outra região ou ainda a sequência total do gene (Adékambi *et al.*, 2003; Tortoli, 2014).

Entre os outros marcadores já utilizados podemos citar o *sodA*, que codifica a superóxido dismutase; *recA*, que codifica parte do sistema de recombinação e reparo do DNA; *ITS*, que representa o espaçador interno entre as subunidades ribossomais 16S e 23S; *gyrA* e *gyrB*, que codificam as duas subunidades da DNA girase; *dnaJ* ou *hsp40*, que codifica uma proteína de choque térmico de 40 kDa. Outras dezenas de marcadores também já foram propostos, porém não obtiveram grande aceitação para fins taxonômicos (Dai *et al.*, 2011; Kim *et al.*, 2016).



**Figura 1.3: Filogenia do gênero *Mycobacterium*.** Marcador filogenético, *hsp65*; grupo externo, *Nocardia farcinica*; modelo estatístico, máxima verossimilhança; programa de construção de árvores, PhyML 3.0 (configuração padrão) e iTOL. Micobactérias de crescimento lento em vermelho e de crescimento rápido em azul. Escala equivalente a 0,02 substituições/sítio. Fonte: Dai *et al.*, 2011B.

### 1.1.2. Classificação

Diversas classificações têm sido usadas para as micobactérias baseadas em diferentes características como genótipo, tempo de crescimento, patogenicidade,

origem, relação agente-hospedeiro, habitat, relação com o complexo *M. tuberculosis* (Runyon 1959; Davidson, 1989; Tortoli, 2003; Dai *et al.*, 2011). Dentre elas, destaco a classificação proposta por Runyon (1959) que separa as micobactérias pertencentes ao complexo *M. tuberculosis* e *M. leprae* das demais e divide as não pertencentes a este complexo em quatro grupos baseado no tempo de crescimento e produção de pigmentos carotenóides: fotocromógenas, escotocromógenas, acromógenas e de crescimento rápido (produtoras ou não de pigmento) (Tortoli, 2003). Essa classificação foi a mais utilizada até o final do século passado, pois baseia-se principalmente em características fenotípicas.

No entanto, nos últimos anos tem se utilizado uma classificação baseada apenas no tempo de crescimento, de acordo com a capacidade de formar colônias em meio sólido – uma simplificação da classificação proposta por Runyon. Micobactérias de crescimento rápido (colônias em até 7 dias) são geralmente oportunistas ou não-patogênicas, como *M. smegmatis*, e micobactérias de crescimento lento (colônias em mais de 7 dias) são geralmente patogênicas ao homem, como *M. tuberculosis*. (Runyon, 1959; Broset *et al.*, 2015). Alguns estudos descrevem que o uso de marcadores filogenéticos, em especial o gene *16S rRNA*, *hsp65* e *rpoB*, são capazes de separar a árvore nesses dois grupos (Kim *et al.*, 1999; Saviola & Bishai, 2006; Dai *et al.*, 2011B; Tortoli, 2014).

Além dessa classificação geral do gênero, diversas espécies são agrupadas em complexos com base em um elevado grau de similaridade genética associado à patogenicidade. O complexo *M. tuberculosis* é o mais amplamente estudado, cujos membros são altamente relacionados, abrangendo oito espécies e subespécies que incluem *M. tuberculosis*, *M. canettii*, *M. africanum*, *M. bovis*, *M. bovis* BCG, *M. microti*, *M. pinnipedii*, *M. mungi* e *M. caprae* (van Soolingen *et al.*, 1997; Niemann *et al.*, 2002; Cousins *et al.*, 2003; Ernst *et al.*, 2007; Alexander *et al.*, 2010; Jia *et al.*, 2017). São estritamente patogênicos, de crescimento lento e podem desencadear a tuberculose no homem e em animais (Tortoli, 2006; Alexander *et al.*, 2010). Apresentam 99,9 % de similaridade em relação ao genoma e sequências idênticas do gene *16S rRNA*, mas diferem quanto a hospedeiros, fenótipos e patogenicidade (Brosch *et al.*, 2000; Rodriguez-Campos *et al.*, 2014; Galagan, 2014).

O Complexo *M. avium* (MAC) era composto até a década de 1980 por apenas duas espécies (*M. avium* e *M. intracellulare*), porém novas espécies têm sido

introduzidas ao grupo ao longo dos anos: *M. arosiense*, *M. bouchedurhonense*, *M. chimera*, *M. marseilense*, *M. paraintracellulare*, *M. timonense*, *M. vulneris* e *M. yongonense*, além de três subespécies de *M. avium* (*hominissuis*, *paratuberculosis* e *silvaticum*). Entre as MNT, as pertencentes ao MAC são as mais amplamente distribuídas, sendo isoladas geralmente de água e solo – porém, no Brasil, só foram encontradas em casos clínicos. Causam doenças pulmonares crônicas com avanço lento; no entanto, poucas pessoas infectadas desenvolvem a doença. Um dos maiores problemas é o aumento de infecções oportunistas especialmente em pessoas com HIV (Daley, 2017). O Complexo *M. abscessus* abriga três subespécies: *M. abscessus abscessus*, *M. abscessus massiliense* e *M. abscessus bolletii*. É um grupo de micobactérias de crescimento rápido, responsáveis por grande parte das infecções por MNT, que tendem a ser crônicas e recorrentes devido a multirresistência aos antibióticos usuais (Tortoli *et al.*, 2016; Adékambi *et al.*, 2017). Há ainda quem considere a existência de outros, como o complexo *M. fortuitum*, *M. simiae* e *M. terrae* (Tortoli, 2014).

### 1.1.3. Habitat

Visto que a tuberculose ocupa o primeiro lugar no ranking mundial de doenças infectocontagiosas desde 2015 (reemergindo após anos atrás da AIDS), é esperada uma ampla distribuição das micobactérias (WHO, 2016). Ao considerarmos apenas as MNT, observa-se que essas espécies também estão globalmente distribuídas, sendo encontradas em todos os continentes, inclusive na Antártida (Learn-Han *et al.*, 2012), e nos mais diversos habitats: reservatórios naturais de água (inclusive pântanos e águas ácidas), sistemas de distribuição de água potável e estações de tratamento de água e de esgoto, encanamentos, banheiras de hidromassagem e spas, piscinas públicas, máquinas de gelo, edifícios danificados pela água, aerossóis (poeiras ou gotículas) naturais, diferentes tipos de solos, poluentes ambientais, fluidos para remoção de partículas metálicas, fazendas de gado, alimentos, animais silvestres e domésticos, vegetais, insetos, protozoários e humanos. Essa diversidade de habitats é resultado, principalmente: 1) da hidrofobicidade do envelope celular rico em ácidos micólicos, que confere às micobactérias resistência a antibióticos e desinfetantes, aderência a superfícies e formação de biofilmes; 2)

sua baixa taxa de síntese de proteínas e uma taxa metabólica normal, que determina seu lento crescimento, permitindo uma melhor adaptação às variações nas condições ambientais; e 3) sua resistência a diferentes tipos de estresse (variações de pH, nutrientes, temperatura, oxigênio), que permite sua sobrevivência em diversos ambientes antrópicos e naturais (Primm *et al.*, 2004; Falkinham III, 2009; Falkinham III, 2015).

No Brasil, há poucos registros descrevendo habitats para as micobactérias que não pertencem ao complexo *M. tuberculosis* nem a *M. leprae*. Falcão e colaboradores (1993) verificaram que as MNT foram os únicos organismos potencialmente patogênicos encontrados em amostras de água tratada e de nascentes. Em 2004, houve um surto de infecções causadas por micobactérias de crescimento rápido em pacientes submetidos a procedimentos invasivos como laparoscopia, artroscopia, cirurgias plásticas e procedimentos cosméticos. Foram mais de 2.000 casos oficialmente reportados em quatro anos, causados principalmente por membros do complexo *M. abscessus* (Leão *et al.*, 2009). Restrepo e colaboradores (2009) verificaram a presença de *M. mucogenicum* nas águas das torneiras do centro cirúrgico, indicando que a água do ambiente hospitalar seria a mais provável fonte de contaminação para os casos dos surtos pós-cirúrgicos. Outro habitat indicado como possível fonte de infecção foi o leite puro de búfala, onde foram identificadas cinco espécies de MNT (*M. flavescens*, *M. gordonae*, *M. simiae*, *M. kansasii* e *M. lentiflavum*), sendo as últimas três caracterizadas como patógenos oportunistas (Jordão Júnior *et al.*, 2009).

*M. fragae* foi a primeira espécie de micobactéria descrita primariamente no Brasil. Foi isolada de um paciente cearense com infecções pulmonares e identificada no Centro de Referência Professor Hélio Fraga, a instituição nacional de referência para tuberculose pertencente à Escola Nacional de Saúde Pública (ENSP/Fiocruz). É uma espécie de crescimento lento próxima filogeneticamente a *M. celatum*, *M. branderi* e *M. kyorinense* (Ramos *et al.*, 2013). Dois anos depois foi descrita uma nova espécie, *M. paraense*, a partir de 14 cepas isoladas de 16 pacientes do estado do Pará com sintomas de problemas pulmonares, no Instituto Evandro Chagas. *M. paraense* é de crescimento lento, pertencente ao complexo *M. simiae* e mais próxima a *M. interjectum* (da Costa *et al.*, 2015). A terceira e última nova espécie micobacteriana descrita até o momento no Brasil, foi a *M. saopaulense*: espécie de

crescimento rápido próxima ao grupo *M. abscessus*-*M. chelonae*. Ela foi identificada a partir de dois isolados obtidos de pacientes com infecções no cristalino (São Paulo em 1999), um com abscesso cervical (São Paulo em 2007), e outros dois isolados de *Danio rerio* (peixe paulistinha ou *zebrafish* em 2003) (Nogueira *et al.*, 2015).

#### 1.1.4. Micobactérias da Mata Atlântica

Micobactérias também foram isoladas de amostras de solo do Parque Nacional da Serra dos Órgãos (Bruce *et al.*, 2010), como dito anteriormente, indicando possíveis novas fontes de contato com as MNT. Alguns estudos já têm sido realizados especificamente com algumas destas micobactérias, como o trabalho anterior do nosso grupo, que realizou caracterizações específicas de nove micobactérias, visando a escolha de um candidato a possível hospedeiro de um futuro sistema de expressão (Emmerick, 2013). Neste trabalho, verificou-se que essas micobactérias 1) crescem melhor a 22 °C em meio sólido ou líquido; 2) podem ser reunidas em três grupos filogenéticos diferentes; 3) expressam a mesma celulase característica de *M. bovis* BCG Moreau; 4) apresentam alguns antígenos compartilhados com a cepa vacinal brasileira BCG Moreau; e 5) são transformáveis. Como a análise filogenética realizada foi preliminar, não foi possível afirmar que as micobactérias estudadas tratavam-se de novas espécies, sendo definidas apenas como isolados micobacterianos. Entre elas, destaco os isolados 226, 271, 293 e 294, representantes dos três grupos filogenéticos estudados.

O isolado 226 e 293 apresentam-se como bacilos longos e bem definidos, 271 como bacilos menores e mais grumados e 294 com formato coco-bacilar, menores, mais grumados e de difícil visualização. Quanto ao perfil de crescimento, os isolados 293 e 294 crescem rápido, atingindo a fase estacionária entre 50 h e 80 h; 271 cresce mais lentamente e com baixa densidade bacteriana e 226 apresentou um perfil atípico, com longa fase *lag* seguida por rápido crescimento. Bioquimicamente, esses quatro isolados são capazes de utilizar citrato como única fonte de carbono, crescem em meio Sabouraud e não realizam fermentação butilienoglicólica nem hidrólise de amido. Os isolados 226, 293 e 294 são capazes de reduzir nitrato (226 e 293 com intensidade muito similar), 293 apresenta gelatinases e 294 é capaz de crescer em elevadas concentrações de NaCl. Quanto ao perfil de secreção de

proteínas, os isolados 226, 293 e 294 apresentam um perfil complexo, com um aumento gradual na quantidade de proteínas secretadas e a presença de bandas proteicas características ao longo de toda a curva, e o isolado 271 apresenta poucas proteínas secretadas ao longo de toda a curva. Os perfis de proteínas totais e secretadas de 226 e 293 são muito semelhantes. Após análise filogenética com sequências parciais dos genes *16S rRNA* e *hsp65*, verificou-se que 226 e 293 pertencem ao mesmo grupo. Após estudo da capacidade de transformação, verificou-se que 226, 293 e 294 reconhecem um promotor exógeno, ao contrário de 271 (Emmerick, 2013).

Outro importante trabalho foi a avaliação da diversidade, taxonomia genômica e resistoma de 14 destas micobactérias através de análises genômicas (Morgado, 2017). A análise por taxonomia genômica realizada sugere a divisão das micobactérias estudadas em três clusters, concordando com os dados descritos anteriormente (Emmerick, 2013). Utilizando diversos critérios de distâncias moleculares, foi sugerido que entre os isolados estudados haveria cinco novas espécies dentro do gênero *Mycobacterium*, mas não foi realizada a descrição delas. Foram preditos também cinco plasmídeos putativos, quatro bacteriófagos e resistência a diversos antibióticos. Em outro trabalho do grupo (Morgado *et al.*, 2017) foi identificado e caracterizado um plasmídeo do isolado 213.

## 1.2. Genomas

O primeiro genoma micobacteriano completamente sequenciado e disponível para a comunidade científica foi o de H37Rv<sup>5</sup> (figura 1.4). Este estudo foi realizado com o objetivo de gerar informação e conhecimento que permitissem elucidar diversos aspectos da biologia do organismo e o desenvolvimento de novas e mais eficientes ferramentas de diagnóstico, além de intervenções terapêuticas capazes de detectar, controlar e eliminar a tuberculose. Para tal, o genoma inteiro foi dividido em insertos grandes (selecionados) e pequenos (aleatórios) que foram clonados e sequenciados pelo método de Sanger. Após a montagem, foi obtida uma sequência única de 4.411.529 pares de base (pb) com um conteúdo GC de 65,6 %, 50 genes

---

<sup>5</sup> Os genes identificados de H37Rv foram anotados pelas letras Rv seguida por um número; por exemplo, Rv0440, também chamado de *hsp65*. Por esse motivo, grande parte dos genes identificados em outras micobactérias também recebem essa anotação.

que codificam para RNAs funcionais e 3.924 preditos como codificantes para proteínas, das quais 16 % não tinham função conhecida e mais de 6 % estavam envolvidas no metabolismo de lipídeos (Cole *et al.*, 1998). Em seguida, o genoma de *M. leprae* foi sequenciado, também pela metodologia de Sanger, com uma cobertura<sup>6</sup> de sete vezes, apresentando diversas características peculiares em relação às demais espécies do gênero: apenas 3.268.203 pb com um conteúdo GC de 57,8 % e apenas 1.604 genes preditos como codificantes para proteínas (menos de 50 %) (Cole *et al.*, 2001).

Embora os primeiros genomas micobacterianos sequenciados tenham sido das espécies patogênicas (*M. tuberculosis*, *M. leprae*, *M. bovis*, *M. abscessus* e *M. avium*), as saprófitas rapidamente entraram nessa lista. Em 2006, *M. smegmatis*, uma espécie de crescimento rápido amplamente usada como modelo no estudo das micobactérias, foi a primeira saprófita a ter seu genoma completamente sequenciado: 6.988.209 pares de base com um conteúdo GC de 67,4 % e 6.717 genes que codificam proteínas. Um mês depois, foi publicado a sequência de *M. vanbaalenii*, outra espécie de crescimento rápido e saprófita que tem a capacidade de degradar hidrocarbonetos aromáticos policíclicos: 6.491.865 pares de base com um conteúdo GC de 67,8 % e 5.999 genes que codificam proteínas e 59 que codificam para RNAs (<https://www.ncbi.nlm.nih.gov/genome>).

Todos os genomas citados até o momento foram sequenciados e montados em um único *contig*<sup>7</sup>. No entanto, entre as 167 espécies de micobactérias com o genoma completamente sequenciado atualmente, muitas apresentam apenas a versão *scaffold*<sup>8</sup>. Além disso, várias espécies apresentam o genoma de diversas cepas sequenciadas. *M. tuberculosis*, por exemplo, apresenta 157 sequências completas em um único *contig* disponíveis, além de mais de 3.900 em *scaffold* (<https://www.ncbi.nlm.nih.gov/genome>). Esse aumento exponencial no número de

---

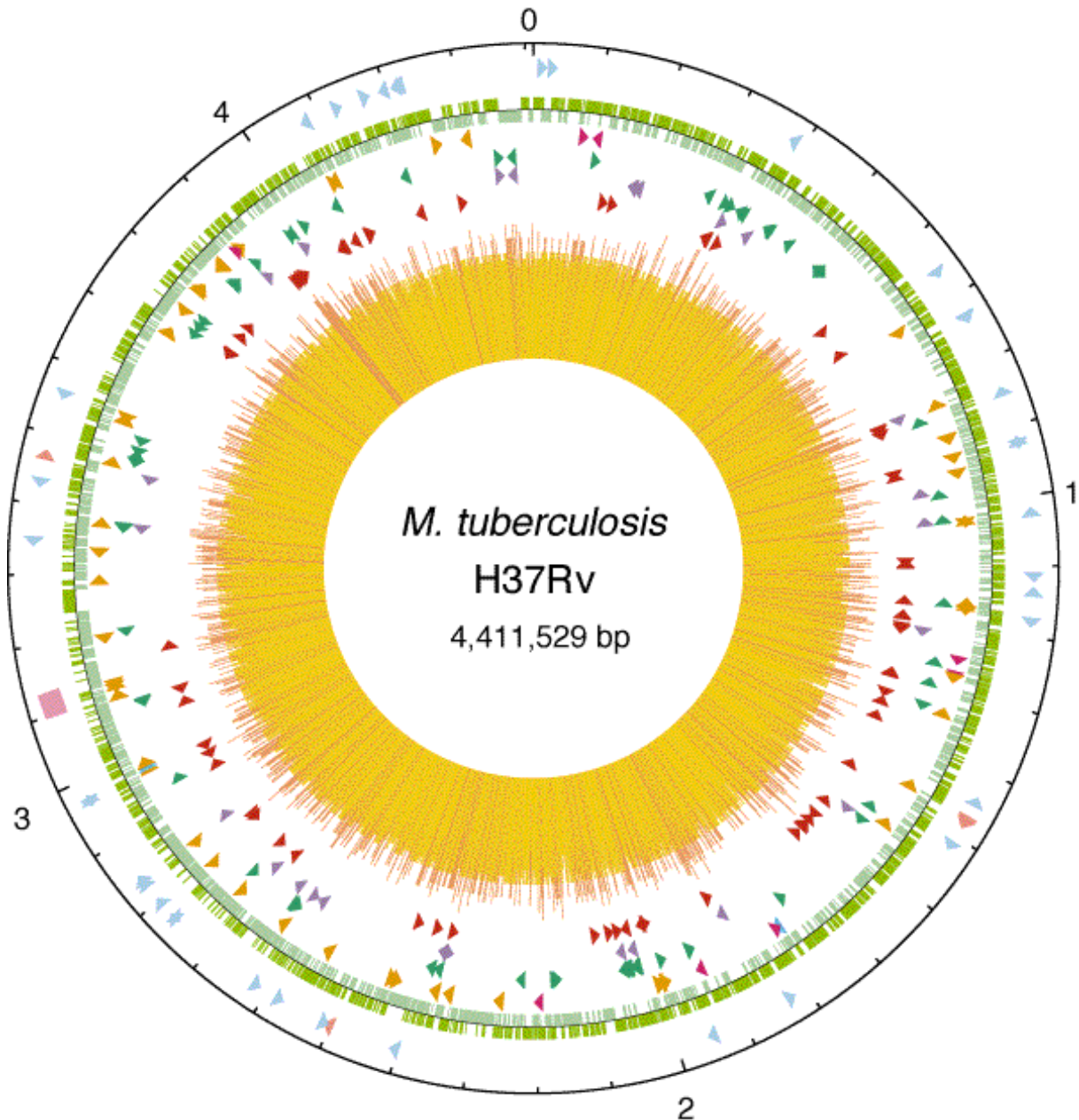
<sup>6</sup> A cobertura pode ser definida como o número de vezes que uma determinada sequência foi obtida. A maneira mais comum de calcular a cobertura de um genoma é dividir o número de bases obtidas pelo sequenciamento pelo tamanho total do genoma. Esse valor reflete uma cobertura média, onde existem regiões com uma cobertura bem maior e outras que talvez nem tenham sido sequenciadas. Fato é que quanto maior a cobertura, maior a chance de ocorrer sobreposição, possibilitando a correção de erros e maior extensão da sequência (Verli, 2104).

<sup>7</sup> *Contigs* ou fragmentos contíguos podem ser definidos como a união de duas ou mais sequências (*reads*) formadas por sobreposição de elementos comuns a pelo menos duas sequências (Verti, 2014).

<sup>8</sup> *Scaffold* é a porção do genoma constituído por *contigs* e *gaps*, onde os *contigs* estão ordenados e orientados (Verti, 2014).



genomas de micobactérias deve-se em parte ao aumento no número de espécies descritas e ao crescente interesse nas MNT. Entretanto, a principal razão é o avanço nas metodologias de sequenciamento.



**Figura 1.4: Genoma de *M. tuberculosis*.** Mapa circular do cromossomo de *M. tuberculosis* H37Rv. O círculo mais externo mostra a escala em Mb, com o 0 representando a origem de replicação. O primeiro anel de fora para dentro mostra a posição dos genes de RNA (RNAat em azul e outros em rosa) e região de repetição direta (quadrado rosa), e o segundo, as sequências codificantes por fita (sentido horário, verde escuro; anti-horário, verde claro). Do terceiro ao sexto anel, são representados, respectivamente, o DNA repetitivo (Inserções, laranja; família REP 12E13, rosa escuro; profagos, azul), membros da família PPE (verde), membros da família PE (roxo) e sequências polimórficas ricas em GC (vermelho escuro). O último anel é representado por um histograma (< 65 % GC, amarelo; > 65 % GC, vermelho). Fonte: Adaptado de Cole *et al.*, 1998.

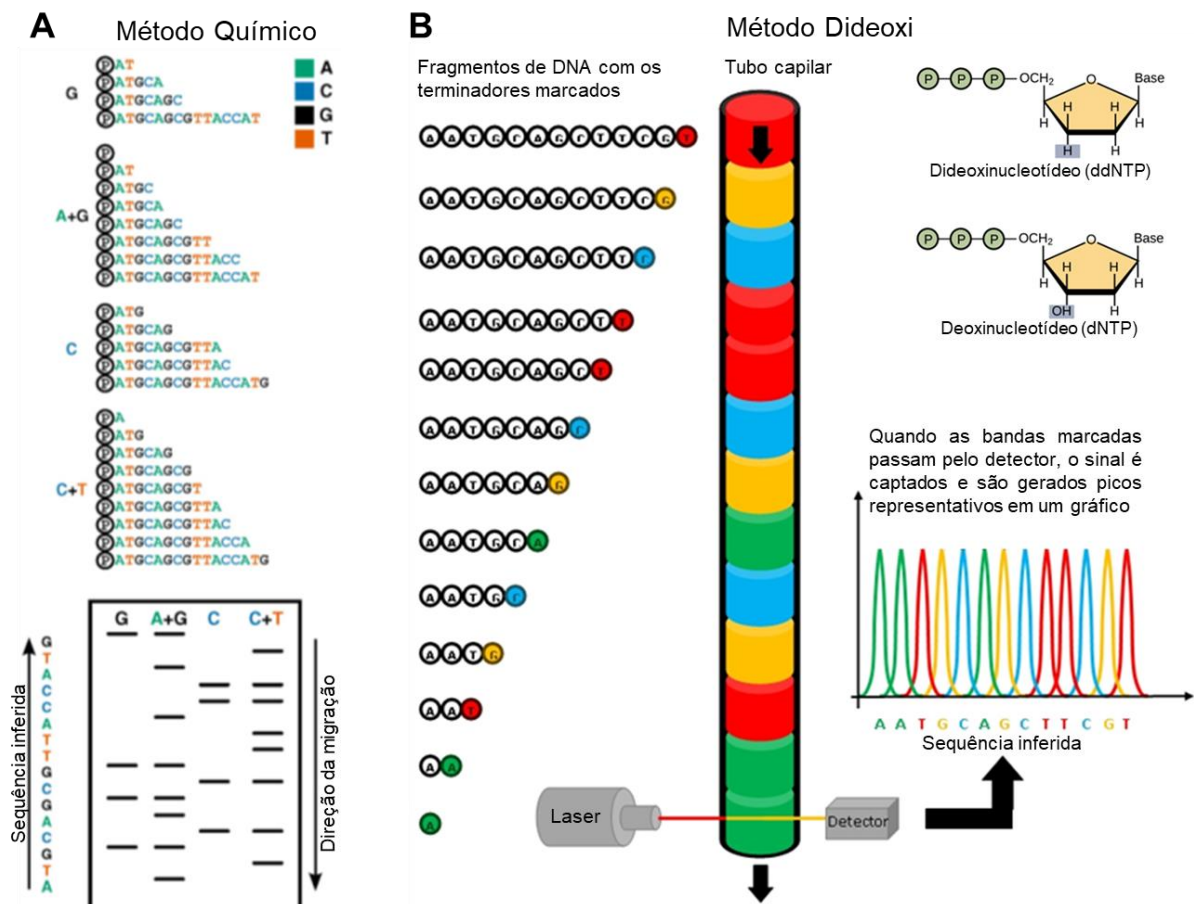
### 1.2.1. Sequenciamento de primeira geração

Desde a descoberta de que o DNA carrega a informação genética por Avery, MacLeod e McCarty em 1944 e da caracterização da sua estrutura por Watson e Crick em 1953, tem se buscado maneiras de determinar a sequência de nucleotídeos dessa molécula. Frederick Sanger (1980) afirmou que o que motivou seu engajamento na resolução desse problema foi sua convicção de que o conhecimento das sequências poderia contribuir muito para nosso entendimento da matéria viva.

Em 1977, o primeiro DNA genômico (do bacteriófago  $\phi$ X174) foi sequenciado por Sanger e colaboradores através da técnica de “mais e menos”, desenvolvida dois anos antes, que usava uma DNA polimerase para incorporar nucleotídeos radioativamente marcados através de duas reações de polimerização e visualizados em um gel de poliacrilamida, determinando a sequência. Na primeira reação (“mais”), apenas um tipo de nucleotídeo estava presente e a extensão sempre terminava nesta base. Na segunda reação (“menos”), os outros três nucleotídeos eram usados e produzindo sequências até o próximo nucleotídeo ausente. Essas reações eram realizadas marcando separadamente cada base para a reação tipo “mais” (Sanger *et al.*, 1977). Paralelamente, Maxam e Gilbert desenvolveram a técnica de “sequenciamento químico” que fragmentava o DNA marcado radioativamente em bases específicas e, depois, inferiam a sequência nucleotídica a partir da posição dos fragmentos no gel de poliacrilamida (figura 1.5 A). Esta metodologia foi amplamente usada, sendo considerada o marco do sequenciamento de DNA de primeira geração (Maxam & Gilbert, 1977).

Ainda em 1977, Sanger desenvolve a técnica de “terminação da cadeia”, também chamada de técnica do dideoxi, pois utiliza o dideoxinucleotídeo (ddNTP) – um análogo químico dos deoxinucleotídeos (dNTP, monômeros da cadeia de DNA) que não apresenta o grupamento hidroxila na extremidade 3', necessária para a extensão do DNA (figura 1.5 B). Nessa metodologia, ddNTPs radiomarcados e dNTPs são misturados a uma reação de extensão do DNA, resultando na síntese de fitas de diversos tamanhos onde os ddNTPs foram aleatoriamente incorporados. A sequência é determinada após corrida em gel poliacrilamida (Sanger & Nicklen, 1977). Esta foi a principal inovação na área de sequenciamento de DNA, sendo

amplamente usada até os dias de hoje, e ficou conhecida como “sequenciamento de Sanger”. Ao longo dos anos, diversas modificações foram realizadas nesta tecnologia, especialmente a troca da marcação radioativa por fluoróforos e a eletroforese capilar ao invés de géis montados entre placas de vidro, que permitiram o desenvolvimento de sequenciadores automáticos (Heather & Chain, 2016).



**Figura 1.5: Sequenciamento de primeira geração.** (A) Método químico desenvolvido por Maxam & Gilbert. Primeiro o DNA é marcado radioativamente ( $P^{32}$ ) na extremidade 5' e depois diferentes tratamentos químicos são feitos para seletivamente remover as bases de DNA: hidrazina remove pirimidinas e hidrazina + grandes concentrações de sal remove apenas citosina. Ácidos removem purinas e dimetilsulfato remove apenas guaninas. Os fragmentos são visualizados por eletroforese em gel de poliacrilamida. (B) Método do dideoxi desenvolvido por Sanger e Nicklen adaptado. São adicionados em uma reação de extensão de DNA dNTP e ddNTP marcados com fluorescência. Os ddNTPs são adicionados aleatoriamente, resultando em fragmentos marcados em todas as posições do DNA a ser sequenciado. Esses fragmentos são encaminhados para uma eletroforese capilar e ao final da corrida são excitados por um laser e a fluorescência liberada é captada pelo detector e representado por picos de cores específicas em um gráfico (cromatograma). Fonte: adaptado de Heather *et al.*, 2016 e <https://di.uq.edu.au/community-and-alumni/sparq-ed/sparq-ed-services/sequencing-dna-using-dye-terminators>.

Assim, os sequenciadores de DNA de primeira geração geram *reads*<sup>9</sup> grandes (aproximadamente 1.000 pares de base) e de alta qualidade. No entanto, a quantidade de bases geradas por corrida é pequena, resultando no elevado custo por base. Para sequenciamento de genomas inteiros, é necessário o uso da técnica de *shotgun*, onde fragmentos de DNA sobrepostos são clonados e sequenciados separadamente – processo demorado e altamente trabalhoso – e depois montados em uma longa e ininterrupta sequência (ou *contig*) através de ferramentas de bioinformática (Heather & Chain, 2016). Um marco desta metodologia foi o sequenciamento do genoma da bactéria de vida livre *Haemophilus influenzae*, o primeiro ser vivo a ter seu genoma completamente sequenciado (Fleischmann *et al.*, 1995). Outro marco importante foi o sequenciamento da bactéria fitopatogênica *Xylella fastidiosa*, primeiro genoma sequenciado no Brasil (Simpson *et al.*, 2000). Diversos projetos se desenvolveram desde então, com destaque para o sequenciamento da cepa vacinal brasileira contra a tuberculose, *M. bovis* BCG Moreau, realizado na FIOCRUZ (Gomes *et al.*, 2011).

### 1.2.2. Sequenciamento de segunda geração

As tecnologias de sequenciamento de nova geração podem ser divididas em segunda e terceira geração; todavia, muito se discute sobre quais parâmetros definem o que é segunda e terceira geração. Aqui utilizou-se os parâmetros adotados na maioria dos trabalhos e na revisão sobre o assunto de Heather & Chain (2016). A capacidade de realizar diversas reações de sequenciamento em paralelo, aumentando exponencialmente a geração de dados, foi a característica adotada para definir as plataformas de segunda geração: 454, Illumina, SOLiD e Ion Torrent. Sequenciamento de terceira geração foi definido como as plataformas que possuem a capacidade de sequenciar moléculas únicas sem a necessidade de amplificar o DNA (Helicos, PacBio e Nanopore). Vale ressaltar que existiram e existem outras plataformas de sequenciamento, mas que não tiveram grande relevância na atual revolução genômica.

Na década de 1980, foi desenvolvida uma nova metodologia de sequenciamento sem o uso de moléculas marcadas e sem eletroforese, mas que

---

<sup>9</sup> *Reads* são pequenos fragmentos de DNA sequenciado.

permite realizar milhares de reações de sequenciamento em uma mesma corrida sem a necessidade de clonagem: o pirosequenciamento. Esta técnica baseia-se na capacidade de sintetizar pirofosfato durante a síntese da nova molécula de DNA e conta com a ação de quatro enzimas: DNA polimerase, ATP sulfurilase, luciferase e apirase (Nyrén & Lundin, 1985; Nyrén, 1987; Hyman, 1988). Essa tecnologia foi licenciada à 454 Life Sciences, que a disponibilizou através dos sequenciadores GS20 (referência aos 20 milhões de pares de base gerados por corrida) em 2005 e depois vendida à Roche Diagnostics, que atualizou os sequenciadores e mudou o nome para 454 GS FLX e, sua versão compacta, 454 GS Junior em 2007 (Heather & Chain, 2016). Apesar de todas as contribuições para a nova geração de sequenciadores, essa plataforma foi descontinuada em 2016.

O sequenciamento utilizando a plataforma 454 GS é dividido em três etapas: preparo de biblioteca de DNA, PCR em emulsão e pirosequenciamento. Para o preparo da biblioteca, o DNA genômico é fragmentado por nebulização em pedaços pequenos (300 a 800 pb, em média) e as extremidades são tratadas para receber os adaptadores (A e B). Em seguida, são ligados adaptadores em ambas as extremidades do DNA dupla fita que, então, é desnaturado. O DNA se liga a uma microesfera por complementariedade entre a sequência da esfera e do adaptador B, sendo encaminhada para a segunda etapa. Todos os reagentes necessários para uma reação de PCR e as microesferas são emulsionados em uma mistura de água e óleo e os fragmentos de DNA são amplificados clonalmente (30 milhões de cópias). É realizado o enriquecimento, onde as esferas contendo o DNA amplificado (geralmente 10 %) são separadas das demais e encaminhadas para um poço dentro de uma placa PicoTiter com 3,2 milhões de poços. Em cada poço é adicionado um mix de enzimas, os iniciadores (complementar ao adaptador A) e a placa segue para a etapa de pirosequenciamento, que realiza o processo automaticamente. No sequenciador, os nucleotídeos são adicionados separada e sequencialmente à placa e, havendo complementariedade entre o(s) nucleotídeo(s) e o DNA molde ocorre a emissão de um sinal luminoso que é capturado por uma câmera CCD. Essa emissão do sinal se dá pela reação do fosfato inorgânico (Pi), liberado na incorporação do nucleotídeo pela DNA polimerase, com a ATP sulfurilase, que converte o Pi em ATP. Esse ATP age como substrato da luciferase que transforma luciferina em oxiluciferina gerando luz visível na intensidade da quantidade de ATP. Os

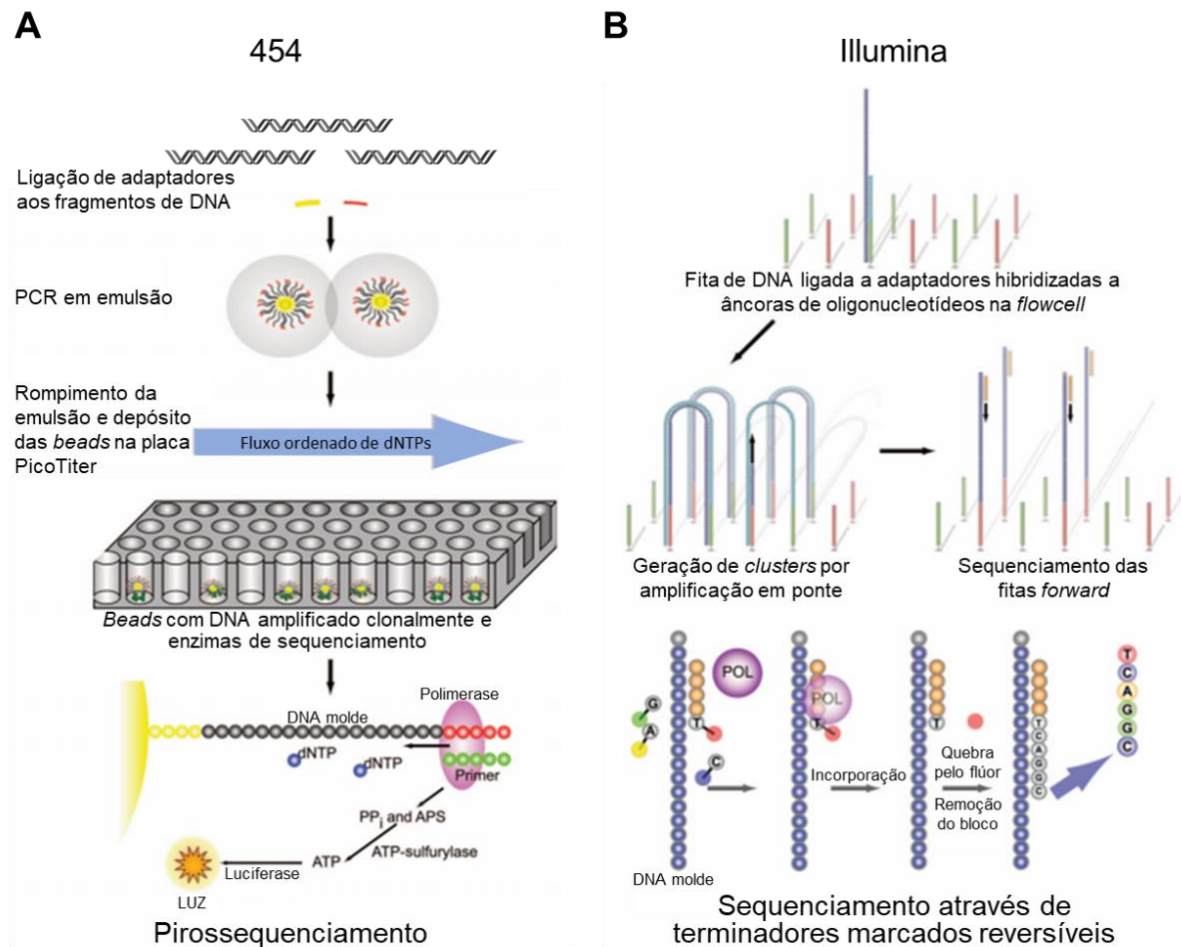
nucleotídeos e ATP não incorporados são degradados pela enzima apirase e a reação recomeça com outro nucleotídeo. Os dados são armazenados em arquivos do tipo *standard flowgram format* (SFF) para posteriores análises (Roche, 2012; Heather & Chain, 2016).

Em 2006, foi lançado o sequenciador Solexa Genome Analyzer pela empresa Solexa, sendo vendido à Illumina no mesmo ano. Esta tecnologia baseia-se no sequenciamento por síntese de *reads* pequenas através de terminadores marcados reversivelmente. Após a fragmentação enzimática do DNA genômico, são adicionadas adeninas à extremidade 3' preparando os fragmentos para receberem os adaptadores que se ligarão aos oligonucleotídeos fixados na *flowcell*<sup>10</sup>. Então, é realizado a amplificação em ponte, que consiste em subseqüente PCRs em fase sólida que produz populações clonais para cada fita de DNA ligada à *flowcell*. O sequenciamento ocorre pela adição de nucleotídeos marcados com fluoróforos (diferentes para cada tipo de nucleotídeo) localizados na região 3', impedindo a incorporação de novos nucleotídeos. Essa fluorescência é captada por uma câmera CCD após a sua excitação com laser e, então, o fluoróforo é removido e um novo ciclo começa. Este primeiro sequenciador fazia a leitura dos fragmentos em apenas um sentido e as *reads* eram muito pequenas (~35 pb), chegando a 1 Gb por corrida. Posteriormente, foram lançados os equipamentos HiSeq (2010) e MiSeq (2011), que são capazes de fazer leituras *paired-end*<sup>11</sup> e com fragmentos de 100 pb, gerando até 600 Gb por corrida. Assim, a Illumina ocupou a maior parte do mercado de NGS (Voelkerding *et al.*, 2009; Heather & Chain, 2016; Besser *et al.*, 2017).

---

<sup>10</sup> Compartimento que permite a passagem de amostras e/ou reagentes e a detecção, medição ou contagem de substâncias específicas por sinais químicos ou físicos. Na tecnologia Illumina, por exemplo, a *flowcell* mantém a amostra, permite o fluxo dos reagentes e a detecção da fluorescência liberada pelos nucleotídeos incorporados.

<sup>11</sup> O sequenciamento *paired-end* realiza a leitura da *read* em ambas as direções, gerando um par de *reads* com o tamanho do inserto da biblioteca, facilitando a montagem.



**Figura 1.6: Sequenciamento de segunda geração.** (A) Pirossequenciamento – plataforma 454 (Roche). O DNA é fragmentado, ligado a adaptadores e amplificado clonalmente via PCR em emulsão. Em seguida as beads juntamente com as enzimas são depositadas em placa PicoTiter, onde ocorre a reação de sequenciamento. Um evento de incorporação de nucleotídeo resulta na liberação do pirofosfato e subsequentemente na emissão de luz. APS, adenosina 5'-fosfosulfato. (B) Sequenciamento através de terminadores reversíveis – plataforma Illumina. Fragmentos de DNA se ligam a moléculas de DNA simples fita imobilizadas por hibridização a flowcell. Amplificação em ponte gera clusters de fragmentos de DNA amplificados clonalmente. Os clusters são desnaturados e o sequenciamento inicia com a adição do iniciador, DNA polimerase (POL) e 4 terminadores reversivelmente marcados. O flúor e o bloco são removidos antes do próximo ciclo de síntese. Fonte: adaptado de Voelkerding *et al.*, 2009.

Também em 2006, a Applied Biosystem (após fusão com a Invitrogen em 2008, tornou-se Life Technologies) lançou a plataforma SOLiD (*Sequencing by Oligonucleotide Ligation and Detection* – sequenciamento pela ligação e detecção de oligonucleotídeos), que, ao contrário das tecnologias descritas até o momento, baseia-se na ação da DNA ligase e não da DNA polimerase, reduzindo a taxa de erro. No entanto, utiliza *reads* muito pequenas e a quantidade de dados gerados é inferior à plataforma Illumina, sendo descontinuada no final de 2017. Em 2010, foi lançada também pela Life Technologies o Ion Torrent. Embora seja muito similar ao

454, a incorporação de nucleotídeos não é medida por fluorescência ou luminescência, mas pela mudança de pH resultante da liberação de prótons durante a polimerização, a qual é captada por um chip semicondutor, permitindo uma detecção muito mais rápida que as demais tecnologias (Mckernan *et al.*, 2009; Rothberg *et al.*, 2011; Heather & Chain, 2016).

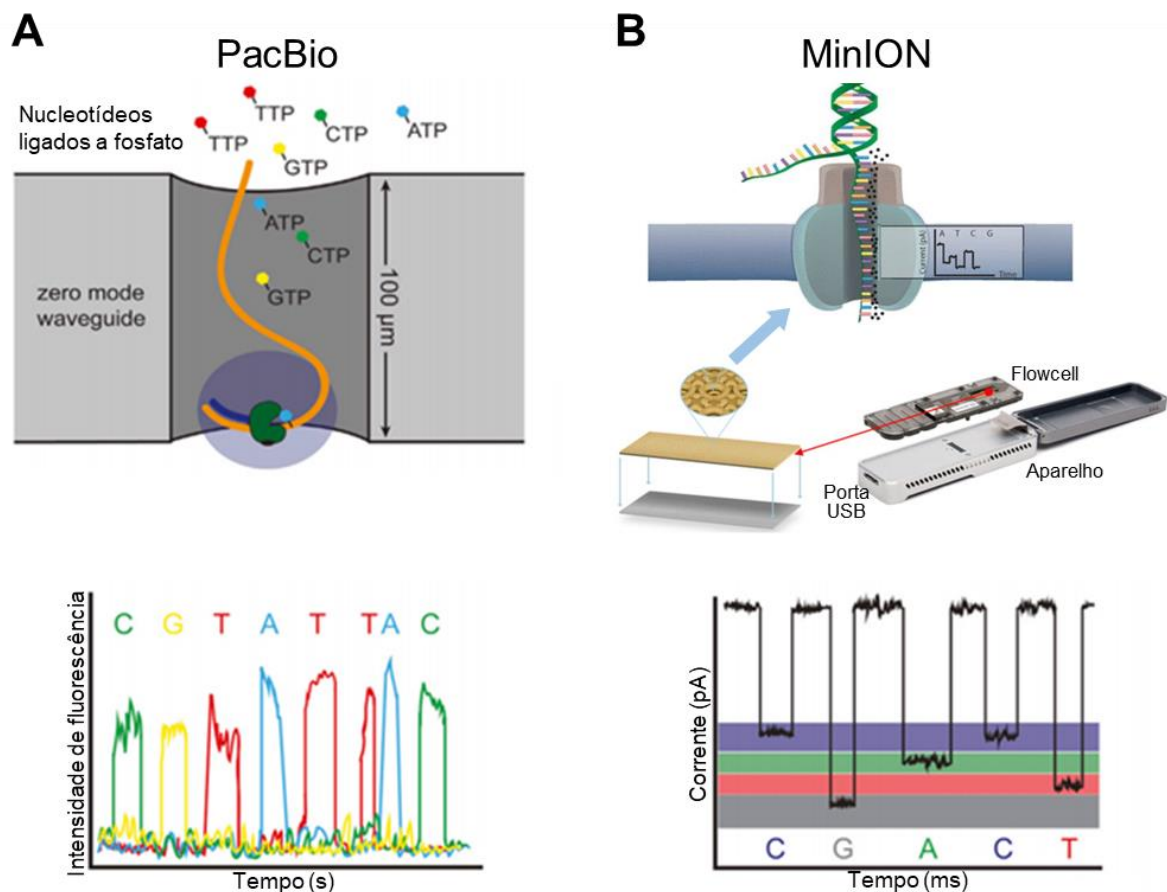
### 1.2.3. Sequenciamento de terceira geração

Em 2009, a Helicos BioSciences foi a primeira a desenvolver uma tecnologia de sequenciamento de moléculas únicas, evitando os erros associados à amplificação do DNA. A metodologia era muito similar à da Illumina, através de terminadores reversíveis fluorescentes; no entanto, por ser um processo longo, caro e com *reads* curtas, foi descontinuado em 2012 (Harris *et al.*, 2008). Em 2011, foi lançado pela Pacific Bioscience, o PacBio: sequenciamento por síntese de moléculas únicas em tempo real (SMRT – *Single Molecule Real Time*). As reações ocorrem dentro de câmaras chamada *Zero-Mode-Waveguides* (ZMWs) que permitem a visualização da fluorescência gerada pela incorporação do nucleotídeo marcado. Embora o processo seja rápido e permita o sequenciamento de *reads* muito longas (de 3 a 20 kb), a taxa de erro é alta e o equipamento é muito caro, grande e necessita de uma boa infraestrutura (Eid *et al.*, 2009).

A Oxford Nanopore Technologies lançou em 2016 uma nova plataforma – embora uma versão teste tenha sido disponibilizada em 2014 – baseado em uma tecnologia descrita em 1996: o sequenciamento através de nanoporos, que permite o sequenciamento do DNA quando este passa através de um poro microscópico em uma membrana, alterando o fluxo de íons. Primeiramente, uma proteína cria um poro em uma bicamada lipídica e mantém um “adaptador” em seu interior e outra proteína realiza a desnaturação da molécula de DNA, encaminhando uma fita para o interior do poro. Existe uma diferença no gradiente de íons entre os lados da membrana, criando um fluxo de íons quando o poro é aberto. Este fluxo cria uma correnteza que carrega a molécula de DNA pelo poro; no entanto, as bases bloqueiam parcialmente esse fluxo (cada base com um grau diferente). O “adaptador” mantém as bases no lugar até serem identificadas eletronicamente (Clarke *et al.*, 2009). Existem 4 principais equipamentos disponibilizados pela Oxford



Nanopore que diferem principalmente na quantidade de dados gerados: SmidgION, MinION, GridION X5 e PromethION. Sem dúvida o principal é o MinION, que além de oferecer o sequenciamento de longas *reads*, em tempo real, com alta quantidade de dados gerados, de maneira rápida, com fácil preparação do material e para diversas finalidades, o MinION é portátil. Com aproximadamente 90 g e do tamanho de um telefone celular, funciona através de um cabo USB 3.0 ligado a um computador (desktop ou laptop), podendo ser usado tanto no laboratório quanto no ártico, nas montanhas, selvas ou estações espaciais. Esta versatilidade permite não apenas uma nova revolução na composição dos dados gerados, como também no local, tempo e por quem esses dados serão gerados (Heather & Chain, 2016; Jain *et al.*, 2016).



**Figura 1.7: Sequenciamento de terceira geração.** (A) Sequenciamento SMRT – PacBio. Uma única polimerase é posicionada na base de uma ZMW. Os quatro tipos de nucleotídeos marcados com fosfato são disponibilizados, permitindo a polimerização contínua da fita de DNA molde. A incorporação do nucleotídeo aumenta o tempo de permanência do nucleotídeo no interior da ZMW, resultando em um sinal de fluorescência que é captado. (B) Sequenciamento por nanoporos – MinION. O DNA molde é ligado a um adaptador que o introduz no poro formado na membrana. Mudanças na corrente de íons ocorrem no momento que os nucleotídeos passam pelo poro, sendo usadas para diferenciar as bases. Fonte: adaptado de <http://biochemistri.es/post/119865709426/of-nanopores-and-isoforms>

Entre as micobactérias, considerando apenas a primeira vez em que o genoma de cada espécie foi sequenciado, 44 foram realizados exclusivamente pelo método de Sanger, 3 exclusivamente por 454, 94 exclusivamente por Illumina, 8 exclusivamente pelo Ion Torrent, 7 exclusivamente pelo PacBio, 3 através de uma abordagem híbrida entre 454 e Illumina, 1 entre 454 e PacBio, 6 entre Illumina e PacBio e 1 entre Sanger, 454 e PacBio. Mas se considerarmos os quase 8.000 genomas micobacterianos sequenciados e disponibilizados no GeneBank, aqueles sequenciados pela tecnologia de Sanger representam menos de 1 %, e outras tecnologias como Solexa, SOLiD e MinION entram nessa lista (<https://www.ncbi.nlm.nih.gov/genome>).

#### 1.2.4. Genômica comparativa

A compreensão da biologia dos diversos seres vivos sempre teve suas respostas na análise comparativa e evolutiva das informações dispostas, a exemplo da comparação entre os bicos dos tentilhões de Galápagos descritos por Darwin. Na análise dos genomas não é diferente: desde que o sequenciamento dos primeiros genomas, ficou evidente que a informação de um único genoma era difícil de ser analisada separadamente (Moreira, 2015). Nesse contexto, análises de genômica comparativa tem fornecido uma visão geral das características compartilhadas entre diversos genomas a exemplo de um estudo realizado com mais de 100 espécies da ordem *Actinomycetales* que revelaram uma ampla diversidade entre os diferentes genes biossintéticos do gênero *Mycobacterium* (Doroghazi & Metcalf, 2013). Saha e colaboradores (2018), em um estudo genômico amplo com ênfase nas espécies do complexo *M. avium*, mostraram que 1) os efeitos ambientais também influenciaram na variação do uso de códons em micobactérias não patogênicas, já que eles tinham que se adaptar mais frequentemente a mudanças ambientais do que as oportunistas para determinados hospedeiros; 2) o complexo *M. avium* está mais intimamente relacionado a espécies não patogênicas e 3) a alta taxa evolutiva do complexo *M. avium* mostrou sua natureza em constante evolução

Prasanna e Mehra (2013) realizaram um estudo comparativo entre 10 genomas micobacterianos (seis espécies patogênicas e quatro não-patogênicas), mostrando que o tamanho do genoma é inversamente proporcional ao tempo de duplicação,

variando de 3,27 Mb (*M. leprae* – tempo de duplicação de 12 a 14 dias) a 6,99 Mb (*M. smegmatis* – tempo de duplicação de 2 a 3 horas). Como esperado, a comparação destes genomas aponta diferenças, dentre elas um número bem superior de genes codificantes para proteínas transportadoras, proteínas regulatórias e aquelas preditas como participantes do metabolismo de energia e vias de metabolismo intermediário central nas espécies saprófitas. Neste mesmo trabalho foi observado que a maioria dos quase 400 genes específicos de espécies não patogênicas tem função desconhecida, evidenciando um grande potencial biotecnológico entre as micobactérias. Como esperado, espécies ambientais expostas a condições de crescimento muito variáveis, apresentam um maior repertório metabólico; espécies patogênicas, por outro lado, são mais restritas quanto às suas necessidades nutricionais.

### 1.3. Biotecnologia

O alto potencial biotecnológico das micobactérias ambientais começou a ser explorado neste século – com exceção para o desenvolvimento da vacina BCG, que protege crianças contra a tuberculose e a hanseníase, desenvolvida no início do século XX. Micobactérias já foram descritas atuando em diversos papéis biotecnologicamente interessantes: *M. marinum* na conversão de ácidos graxos em precursores de combustíveis e outros *commodities* químicos (Akhtar *et al.*, 2013), *M. gilvum* e *Mycobacterium sp.* KMS na biodegradação de hidrocarbonetos aromáticos policíclicos (Zang & Anderson, 2013; Badejo *et al.*, 2013), *M. austroafricanum* e *M. vaccae* na biodegradação de aditivos de gasolina (Hyman, 2013), e na biodesulfurização (Kibane, 2006), *M. cosmeticum* na degradação de componentes de gasolina e outros combustíveis (Zang *et al.*, 2013), *M. vanbaalenii* na biotransformação de acridina (Sutherland *et al.*, 2009), diversas espécies na degradação de sais biliares (Philipp, 2011) e de fitosteróis (Josefsen *et al.*, 2017), *Mycobacterium sp.* djl-10 na degradação do fungicida agrícola carbendazim (Zang *et al.*, 2017). Além disso, *M. neoaurum* já foi usado na produção de esteroides (Yao *et al.*, 2013) e *M. smegmatis* foi engenheirado para a produção de precursores de esteroides (Galán *et al.*, 2016); *M. indicus pranii*, uma espécie saprófita, é hoje um dos principais candidatos ao desenvolvimento de imunoterápicos contra a

tuberculose (Gupta *et al.*, 2012); e *M. aurum* foi proposto como modelo para investigação de drogas contra a tuberculose (Phelan *et al.*, 2015).

### 1.3.1. Sistemas de expressão

Micobactérias são também hospedeiros interessantes para o desenvolvimento de sistemas de expressão de proteínas heterólogas, apresentando boa capacidade secretora e sendo inclusive capazes de realizar alguns tipos de modificações pós-traducionais como glicosilação, ADP-ribosilação e acetilação de proteínas (revisto em Connell, 2001; Parish & Brown, 2008). Embora, atualmente, o uso da tecnologia do DNA recombinante para produzir grandes quantidades de proteínas para estudos funcionais e estruturais seja praticamente uma rotina nos laboratórios de pesquisa, é preciso destacar a importância desse processo em quase todas as áreas das ciências da vida: análises de proteínas, melhoramento genético, produção de proteínas, enzimas e fármacos para fins médicos e industriais e terapia gênica (Gellissen, 2005; Brondyk, 2009). A escolha de um sistema de expressão apropriado é uma variável fundamental na produção de proteínas em grandes quantidades e alta qualidade. Por isso, tanto a melhora das estratégias de expressão existentes quanto o desenvolvimento de novos sistemas de expressão são considerados um dos maiores desafios em diversos projetos, especialmente na área de genômica estrutural (Bashiri & Baker, 2014).

Diversos sistemas de expressão têm sido desenvolvidos utilizando diferentes organismos, desde procariotos simples a eucariotos complexos. Apesar dessa grande diversidade de sistemas de expressão, existem proteínas cuja expressão ainda é difícil devida à baixa estabilidade, toxicidade para o hospedeiro, formação de corpos de inclusão, carência de modificações pós-traducionais, elevados custos e complexidade do processo (Miyake *et al.*, 2007; Fernandez & Veja, 2013). *Escherichia coli* é um dos hospedeiros mais utilizados para a produção de proteínas heterólogas por ser bem caracterizado geneticamente, simples de manusear e por apresentar crescimento rápido e barato em cultura (Arya *et al.*, 2008). Diversas modificações têm sido desenvolvidas de modo a melhorar a expressão neste sistema, tais como o uso de novas cepas (Sahdev *et al.*, 2008), novos vetores (Hunt, 2005), novos promotores (Baneyx, 1999) e ligação de proteínas que facilitem a

purificação da proteína recombinante (Terpe, 2002). No entanto, diversos genes permanecem incapazes de serem expressos nesse sistema ou apresentam graves desvantagens como a não realização de modificações pós-traducionais, a expressão em corpos de inclusão e o uso de códons muito diferentes daqueles usados em proteínas eucarióticas (Jonasson *et al.*, 2002; Arya *et al.*, 2008, Fernandez & Vega, 2013).

Neste cenário, destaca-se o uso das micobactérias como hospedeiros de novos sistemas de expressão. *M. smegmatis* foi o primeiro e tem sido o hospedeiro mais comum entre as micobactérias (Snapper *et al.*, 1990; Delogu *et al.*, 2004; Gutierrez-Lugo *et al.*, 2006; Goldstone *et al.*, 2008). Isolada no final do século XIX, não patogênica, de crescimento rápido, com tempo de duplicação de aproximadamente três horas, e genoma sequenciado, tem sido usada como modelo em diversos estudos micobacterianos (Alvarez & Tavel, 1885; Goldstone *et al.*, 2008; He & Buck, 2010). Seu uso como hospedeiro é principalmente para expressar proteínas de *M. tuberculosis*, pois, além do crescimento rápido, fornece chaperonas específicas das micobactérias que poderiam ajudar a corrigir o enovelamento das proteínas recombinantes (Goldstone *et al.*, 2008; Noens *et al.*, 2011). A cepa *M. smegmatis* mc<sup>2</sup>155 é a mais utilizada, uma vez que adquiriu mudanças na replicação plasmidial e mudanças na estrutura e composição do envelope celular que permitiram uma alta propensão à transformação (Snapper *et al.*, 1990; Bashiri & Baker, 2014). Além de *M. smegmatis*, outras micobactérias têm sido usadas para o desenvolvimento de novos sistemas de expressão, como *M. aurum* e *M. vaccae* (Medeiros *et al.*, 2002; Parish & Brown, 2008). *M. aurum* tem sido usado principalmente em estudos de sensibilidade e resistência a antibióticos usados no tratamento de tuberculose e hanseníase (Lazraq *et al.*, 1991). *M. vaccae*, na expressão de diversas proteínas, como superóxido dismutase e lipoproteína de 19 kDa de *M. tuberculosis* (Garbe *et al.*, 1994; Abou-Zeid *et al.*, 1997).

Existem diversos sistemas induzíveis em micobactérias, tais como o promotor de acetamidase, o promotor T7, o promotor  $\beta$ -lactamase, a indução por tetraciclina e sistemas alternativos como a indução por arabinose e pristinamicina, e responsivos a nitrila e tioestreptona (Bashiri & Baker, 2014). Vetores utilizando o promotor de acetamidase foram construídos no final da década de 1990 (Triccas *et al.*, 1998) a partir de um fragmento de 1.500 pb do promotor da acetamidase de *M. smegmatis* e,

também, proporcionando a fusão com 6 histidinas na porção C-terminal da proteína, permitindo a expressão regulada de proteínas recombinantes em altos níveis e um método simples de purificação. No entanto, estes vetores são muitas vezes instáveis para expressão em *M. tuberculosis* devido a deleção da região promotora (Brown & Parish, 2006). Os vetores que utilizam o promotor T7 baseiam-se na ação da RNA polimerase do bacteriófago T7 em transcrever os genes clonados a jusante do promotor (Studier & Moffatt, 1986; Wang *et al.*, 2010). A principal vantagem é que a expressão e os testes de solubilidade podem ser realizados em paralelo entre *E. coli* e *M. smegmatis*. A expressão de proteínas recombinantes sob o controle do promotor da  $\beta$ -lactamase de *M. fortuitum* tem sido usada principalmente para expressão de antígenos micobacterianos de *M. smegmatis* (Trimm *et al.*, 1994). Já o sistema de expressão indutível por tetraciclina é regulado através da expressão constitutiva da proteína repressora de tetraciclina (TetR), onde TetR se liga ao operador evitando a transcrição na ausência do indutor. Porém, quando o indutor se liga a TetR, acaba a repressão e a transcrição ocorre. Esse sistema tem sido usado especialmente na expressão de genes essenciais em micobactérias patogênicas (Blokpoel *et al.*, 2005).

Na maioria das vezes, sistemas de expressão micobacterianos têm sido usados para expressar proteínas de micobactérias, principalmente devido à dificuldade de se expressar essas proteínas em *E. coli*. No entanto, já foi demonstrado o uso de *M. vaccae* – bactéria isolada originalmente de esterco bovino – para expressar a subunidade S1 da toxina pertussis de *Bordetella pertussis* (Medeiros *et al.*, 2002). Outra característica pouco explorada nos sistemas de expressão micobacterianos é a capacidade secretora dessas bactérias, que favorece a expressão de proteínas recombinantes na fração solúvel (Feltcher *et al.*, 2010). Assim, acredita-se ser possível (1) o desenvolvimento de novos sistemas de expressão a partir de micobactérias ambientais, utilizando-as como hospedeiros, bem como (2) a construção de novos vetores a partir dos sinais de regulação e secreção de proteínas encontradas nessas bactérias, contribuindo para a expressão de proteínas recombinantes mais facilmente purificadas e parecidas com a nativa.

São inúmeras as possibilidades de aplicações do conhecimento oriundo do estudo da biodiversidade nacional, em especial dos microorganismos. Até o início deste século apenas 800 espécies de microorganismos tinham sido identificadas em

todo o território nacional (Lewinsohn & Prado, 2002), não existindo nenhum estudo atualizado sobre o assunto. Diante disso, o estudo das micobactérias ambientais brasileiras contribuiria para a melhor caracterização da biodiversidade nacional e do gênero *Mycobacterium*, para a identificação de possíveis genes relacionados à vacina BCG Moreau, para verificar potencial patogenicidade e virulência de isolados ambientais, para a análise da capacidade metabólicas dos isolados da Mata Atlântica frente ao seu ecossistema, para a exploração do potencial biotecnológico e de engenharia genética desses isolados ambientais.

## 2. Objetivos

### Objetivo geral

Caracterizar, utilizando ferramentas moleculares, isolados de micobactérias de crescimento rápido provenientes de solo da Mata Atlântica (Rio de Janeiro), contribuindo para ampliação do conhecimento sobre o gênero *Mycobacterium*.

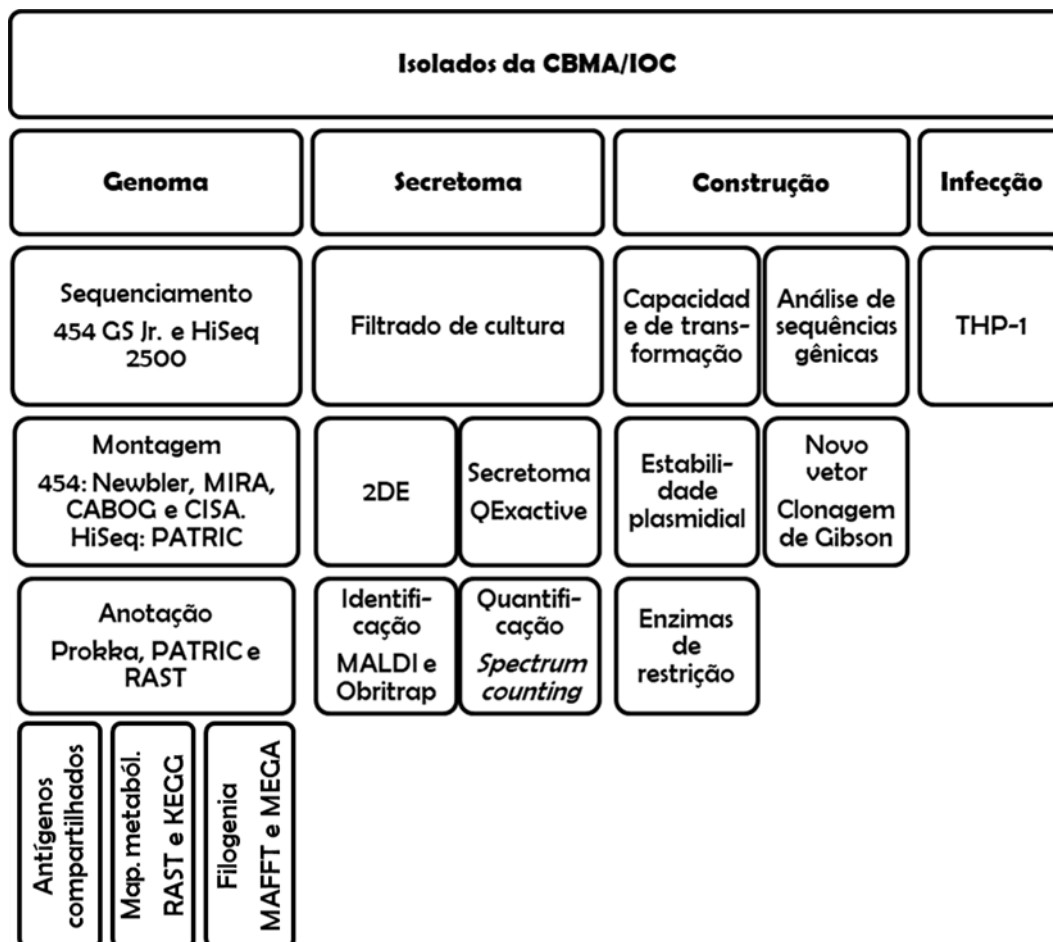
### Objetivos específicos

- ✓ Sequenciar, montar e anotar os genomas de quatro diferentes micobactérias isoladas da Mata Atlântica;
- ✓ Comparar a distribuição gênica em categorias funcionais entre os isolados da Mata Atlântica e espécies de referência;
- ✓ Mapear a capacidade metabólica dos quatro isolados micobacterianos e identificar vias com potencial biotecnológico;
- ✓ Identificar antígenos compartilhados entre esses isolados e a vacina BCG Moreau;
- ✓ Inferir a relação filogenética destes isolados micobacterianos e as demais espécies do gênero;
- ✓ Avaliar o perfil de secreção de proteínas (secretoma) desses isolados micobacterianos;
- ✓ Avaliar a estabilidade plasmidial após transformação com vetores replicativos e integrativos;
- ✓ Determinar a capacidade de sobrevivência dos isolados ambientais em macrófagos, através de infecção *in vitro*



### 3. Metodologia

#### 3.1. Fluxograma geral do trabalho



**Figura 3.1: Fluxograma de trabalho.** Os isolados micobacterianos foram adquiridos da CBMA e os experimentos foram agrupados em quatro áreas principais: análise do genoma, análise do secretoma, construção de um novo sistema de expressão e infecção em macrófagos.

#### 3.2. Escolha dos isolados micobacterianos

Com base em estudos preliminares (Emmerick, 2013), foram escolhidos quatro isolados micobacterianos obtidos da Coleção de Bactérias da Mata Atlântica (CBMA/IOC) para seguir no estudo: 226, 271, 293 e 294. A análise filogenética a partir de fragmentos do gene *16S rRNA* e *hsp65* indicou que estes isolados representam grupos filogenéticos distintos. O isolado 226 foi adicionado por apresentar características de crescimento peculiares, embora tenha sido indicado como pertencente ao mesmo grupo do isolado 293.

### 3.3. Cultivo de micobactérias

Culturas provenientes de lotes-semente dos isolados ambientais (Emmerick, 2013) foram reativadas em meio Luria-Bertani contendo Tween 80 0,05 % (v/v) (LB/Tw) e cultivadas em erlenmeyers de 250 mL contendo 50 mL do mesmo meio. O acompanhamento do crescimento bacteriano foi realizado através de leituras da densidade ótica a 600 nm ( $DO_{600nm}$ ) no espectrofotômetro Ultrospec 1100 pro (Amershan Biosystem), com início aproximado igual a 0,1  $DO_{600nm}$ . Os isolados da CBMA foram crescidos sob agitação a 22 °C e *M. smegmatis* mc<sup>2</sup>155 (mc<sup>2</sup>155) sob agitação a 37 °C. Para crescimento em meio sólido, foram utilizadas placa de petri contendo 25 mL de meio LB acrescido de 1,5 % (p/v) de bacto-ágar (Difco) (LB/ágar).

### 3.4. Coloração de Ziehl-Neelsen

Realizada de acordo com o Manual de Bacteriologia da Tuberculose (Ministério da Saúde, 2005). Esfregaço preparado a partir de 5 µL de cultura, fixado ao fogo, coberto com a solução de fucsina fenicada (0,3 % fucsina básica; 5,5 % fenol aquoso; e 10 % etanol), aquecido até a emissão de vapores (sem ferver) por três vezes ao longo de 5 minutos. Retirado o corante, o material foi descorado com solução álcool-ácido (97 % etanol e 3 % ácido clorídrico) por 2 minutos, lavado em água corrente e corado com solução de azul de metileno (0,1 % azul de metileno e 10 % etanol) por 30 s, seguido de lavagem em água corrente, secagem ao ar e observação em microscópio óptico Eclipse E100 (Nikon).

### 3.5. Extração de DNA genômico

Os isolados micobacterianos foram crescidos em 10 mL de meio LB/Tw por 48 h a 22 °C. As bactérias foram coletadas por centrifugação a 10.000 *g* por 10 minutos. O sedimento bacteriano foi ressuspense em 4 mL de tampão de lise e o material lisado mecanicamente no equipamento *Bead Beater* (Biospec Products INC.) através de 5 ciclos de um minuto com intervalo no gelo. O DNA dos isolados foi extraído utilizando-se o kit *Blood & Cell Culture DNA Mini* (Qiagen), segundo as recomendações do fabricante. Ao final, o material foi ressuspendido em água MilliQ, de acordo com a recomendação da Plataforma de Sequenciamento de Ácidos

Nucléicos de Nova Geração (RPT01J), alocada no Laboratório de Genômica Funcional e Bioinformática/IOC – RJ.

### 3.6. Sequenciamento dos genomas

2 µg de DNA genômico de cada isolado foram entregues à RPT01J. Foi avaliada a pureza do DNA com o equipamento NanoDrop 2000 spectrophotometer (Thermo Scientific), usando como parâmetro principal a razão das absorvâncias a 260/280 nm e, secundário, 260/230 nm. A seguir, o DNA foi quantificado no equipamento Qubit 2.0 fluorometer com o kit *Qubit ds DNA BR* (Invitrogen, Life Technologies).

A partir de 500 ng do DNA genômico entregue à plataforma, foi realizado todo o procedimento conforme as orientações do fabricante pela RPT01J e as amostras foram analisadas no Sequenciador 454 GS Junior<sup>12</sup> (Roche Diagnostics).

A partir de 1 ng do DNA genômico entregue a plataforma, foi realizado o preparo da biblioteca do tipo *paired-end* com o kit de preparação de biblioteca de DNA Nextera XT (Illumina). Todos os procedimentos foram realizados conforme as orientações do fabricante pela RPT01J e as amostras foram analisadas no Sequenciador HiSeq 2500<sup>13</sup> (Illumina).

### 3.7. Montagem e anotação 454 GS Jr.

Foram usados os arquivos sff para a montagem *de novo* dos genomas utilizando-se o programa Newbler<sup>14</sup> versão 2.7 (<http://454.com/contact-us/software-request.asp>). Foi realizado o corte dos adaptadores, definido 500 pb como tamanho

---

<sup>12</sup> O sequenciamento pelo 454 GS Jr. é dividido em preparo de biblioteca de DNA, PCR em emulsão e pirosequenciamento. O DNA genômico é fragmentado, são ligados adaptadores e o DNA é desnaturado. Cada fragmento se liga a uma microesfera que são emulsionadas e os fragmentos de DNA são amplificados clonalmente. As microesferas são encaminhadas para uma placa e inseridas no sequenciador, onde os nucleotídeos são adicionados separada e sequencialmente e, havendo complementariedade entre o nucleotídeo e o DNA molde ocorre a emissão de um sinal luminoso. Os dados são armazenados em arquivos do tipo SFF para posteriores análises (Roche, 2012).

<sup>13</sup> Para o sequenciamento na plataforma HiSeq 2500, o DNA genômico foi fragmentado enzimaticamente (~100 pb), marcado e adaptadores foram adicionados as extremidades. Após controle de qualidade e quantificação, a amostra foi colocada na *flowcell* em dois canais e cada fragmento foi clonalmente amplificado em ponte. Os fragmentos *reverse* são retirados e segue-se para a etapa de sequenciamento através de terminadores marcados reversivelmente apenas com os fragmentos *forward*. Ao final, a sequência sintetizada é retirada e repete-se o processo com os fragmentos *reverse* e os dados são disponibilizados no BaseSpace® Sequence Hub.

<sup>14</sup> Newbler é um software que foi desenvolvido para montagem *de novo* de sequências de DNA geradas por sequenciadores da série 454 GS (Roche Diagnostics).

mínimo para grandes *contigs*, 100 pb como tamanho mínimo para todos os *contigs*, 50 pb como tamanho mínimo para as *reads*, 40 pb como tamanho mínimo de sobreposição entre duas *reads*. Os arquivos sff foram usados novamente para uma nova montagem dos genomas utilizando-se o programa MIRA<sup>15</sup> (<http://sourceforge.net/projects/mira-assembler/>). Mais uma montagem foi realizada, utilizando-se o programa CABOG<sup>16</sup> (<http://www.jcvi.org/cms/research/projects/cabog/>).

As diversas montagens foram integradas com o auxílio do programa CISA<sup>17</sup> (<http://sb.nhri.org.tw/CISA/en/CISA>), e gerado um arquivo FASTA<sup>18</sup> contendo os grandes *contigs* (maiores que 500 pb). Os dados das montagens foram avaliados através da ferramenta QUAST<sup>19</sup> (<http://cab.spbu.ru/software/quast/>). Essas análises foram realizadas em colaboração com o MSc. Ricardo Junqueira da RPT01J.

Foi utilizado o programa Prokka<sup>20</sup> (<http://www.vicbioinformatics.com/software/prokka.shtml>) para a anotação funcional dos genes. O arquivo de saída da montagem em formato FASTA foi o único parâmetro necessário. Ao final, foram gerados 10 arquivos com características específicas (Seemann, 2014). Outra

---

<sup>15</sup> MIRA (*Mimicking Intelligent Read Assembly*) é recomendado para a reconstrução de transcriptomas e na detecção e classificação de polimorfismos de nucleotídeo único (SNPs) (Chevreux *et al.*, 2004).

<sup>16</sup> CABOG (*Celera Assembler with Best Overlap Graph*) é uma extensão do software Celera, usado no primeiro rascunho do genoma humano. Ele é capaz de montar desde genomas bacterianos a genomas de mamíferos. Foi o primeiro programa a montar dados de organismos multicelulares, de sequências ambientais dos oceanos e o primeiro a combinar *reads* de sequenciamento de Sanger e pirosequenciamento (Miller *et al.*, 2008).

<sup>17</sup> CISA (*Contig Integrator for Sequence Assembly*) integra grupos de *contigs* de diferentes montadores e avalia a qualidade dessa montagem integrada. Este software oferece uma montagem melhor em termos de contiguidade e acurácia, comparados com cada montador individualmente (Lin & Liao, 2013).

<sup>18</sup> Arquivos FASTA são simples e permitem fácil manipulação tanto para usuários quanto para algoritmos; por isso, são amplamente utilizados como padrão para armazenamento de sequências. Esse formato é caracterizado pelo sinal de maior (>) no início do cabeçalho da sequência (Pearson & Lipman, 1988)

<sup>19</sup> A ferramenta para acesso à qualidade de genomas montados (QUAST, *Quality Assessment Tool for Genome Assemblies*) tem a finalidade de gerar relatórios para análises de montagens de sequências de genomas, com ou sem a sequência de um organismo de referência (Gurevich *et al.*, 2013).

<sup>20</sup> Prokka é um software gratuito em linha de comando para anotação de genomas de procaríotos. Esta ferramenta pode ser instalada em qualquer sistema Unix e permite a anotação de um genoma bacteriano a partir de sequências completas ou de *scaffolds* em formato FASTA, em menos de uma hora e com ótima qualidade. Este software depende de outros softwares para realizar a anotação: *Prodigal*, predição de sequências codificantes (CDS); *RNAmmer*, predição de genes de RNA ribossomal; *Aragorn*, predição de genes de RNA de transferência; SignalP, predição de peptídeos sinais; *Infernal*, predição de RNA não codificantes. O *Prodigal* faz a identificação das CDS mas não descreve sua função. A anotação funcional é dada de acordo com o melhor resultado após a comparação de sequências proteicas com banco de dados confiáveis e conhecidos (Seemann, 2014).

anotação funcional foi realizada através do RAST Server<sup>21</sup> (<http://rast.nmpdr.org>), utilizando como entrada o arquivo FASTA gerado pelo Prokka contendo os *contigs* originais (em nucleotídeos) usados na anotação pelo próprio Prokka (.fna). Também foi realizada a anotação de espécies de referência a partir das sequências disponíveis no NCBI para fins de comparação. Os dados da anotação realizada pelo RAST foram visualizados no Seed Viewer ([http://www.theseed.org/wiki/Main\\_Page](http://www.theseed.org/wiki/Main_Page)).

### 3.8. Montagem e anotação HiSeq

A biblioteca genômica de cada isolado foi aplicada em dois canais diferentes, dando origem a dois arquivos, *forward* e *reverse*, de cada. Os arquivos de saída de cada canal foram concatenados, gerando apenas um arquivo *forward* e um *reverse* para cada isolado. Foi verificada a qualidade do sequenciamento de cada arquivo, antes e depois de serem concatenados, através do programa FastQC<sup>22</sup> (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

A partir dos arquivos concatenados foi realizada a montagem e anotação através do software *on line* PATRIC<sup>23</sup> (<https://www.patricbrc.org/>), utilizando a configuração “auto” para a montagem. Os dados das montagens foram avaliados através da ferramenta QUAST. Também foi realizada a anotação funcional diretamente através RAST Server, utilizando como entrada o arquivo FASTA gerado pelo PATRIC.

---

<sup>21</sup> O RAST (*Rapid Annotations using Subsystem Technology*) oferece um serviço gratuito de anotação automatizada para genomas bacterianos completos ou quase completos. A partir do carregamento do genoma em formato FASTA, os usuários têm acesso à anotação dentro de um período de 12-24h, com elevada acurácia, abrangência e consistência. Essa qualidade é baseada em uma crescente biblioteca de subsistemas curados manualmente e em amplas famílias proteicas derivadas desses subsistemas (Aziz *et al.*, 2008).

<sup>22</sup> FastQC é uma ferramenta de controle de qualidade para dados de sequenciamento de alta vazão. Escrito em Java, ele importa arquivos do tipo BAM, SAM ou FastQ e fornece uma visão geral rápida dos dados, resumindo isso em gráficos e tabelas. Embora os sequenciadores já forneçam um boletim de qualidade, ele é restrito a problemas gerados pelo sequenciador. O boletim do FastQC provê informações de problemas tanto do sequenciamento como do início do preparo da biblioteca.

<sup>23</sup> Centro Integrado de Recursos de Sistemas Patogênicos (PATRIC, *Pathosystems Resource Integration Center*) disponibiliza uma fonte *on line* que integra uma variedade de tipos de dados: genomas, transcriptomas, interações proteína-proteína, estruturas proteicas 3D e metadados associados. A interface gráfica permite de maneira simples a montagem através do pré-processamento usando o BayesHammer, montagem através dos algoritmos Velvet, SPAdes e IBDA, pontuação através do ARAST, gerando três montagens e a indicação da melhor. Para a anotação de genomas o PATRIC utiliza uma versão customizada do RAST (RASTtk, RAST *tool kit*). Essa versão oferece também a anotação de genes e proteínas especializados através de banco de dados curados como o ARDB (*Antibiotic Resistance Database*), o CARD (*Comprehensive Antibiotic Resistance Database*), o VICTORS, o VFDB (*Virulence Factor Database*) e um banco de dados próprio de fatores de virulência do PATRIC (Wattam *et al.*, 2014 e 2017).

### 3.9. Mapeamento metabólico

A ferramenta comparativa do RAST de análise metabólica através do KEGG<sup>24</sup>, que se baseia nos dados de *EC number*<sup>25</sup> fornecidos pela anotação, foi utilizada para observar diferenças no perfil metabólico dos isolados ambientais quando comparados entre si e contra espécies de referência. Essa análise foi realizada a partir das sequências geradas pelo 454 GS Jr. Alguns mapas metabólicos foram gerados diretamente do KEGG Pathway (<http://www.genome.jp/kegg/pathway.html>).

### 3.10. Análise filogenética

As sequências parciais dos genes *16S rRNA* e *hsp65* e total do *rpoB* foram obtidas do GenBank<sup>26</sup> (<https://www.ncbi.nlm.nih.gov/genbank/>) para as 185 espécies e subespécies tipo de micobactérias descritas até novembro de 2017 (Anexo). A sequência dos isolados foram extraídas dos genomas sequenciados pelo 454 GS Jr. a partir do arquivo da anotação realizada pelo Prokka que contém a anotação e a sequência nucleotídica em formato FASTA (.ffn). Foi realizado o alinhamento utilizando o programa MAFFT<sup>27</sup> (<https://mafft.cbrc.jp/alignment/server/>), com as configurações padrão. As sequências de cada gene foram submetidas a análise no

---

<sup>24</sup> A Enciclopédia Kyoto de genes e genomas (KEGG – *Kyoto Encyclopedia of genes and genomes*) é uma coleção de banco de dados idealmente com informações de genomas, vias biológicas, substâncias químicas e outros. O *KEGG Pathway* é uma coleção de mapas manualmente desenhados representando o conhecimento experimental em metabolismo e várias outras funções da célula e do organismo. Cada mapa contém uma rede de reações e interações moleculares e é desenhado para ligar genes do genoma ao produto gênico (proteínas) nas vias. Esta ferramenta permite comparar o conteúdo gênico com o banco de dados KEGG Pathway para examinar quais vias e funções associadas podem ser codificadas pelo genoma (<http://www.genome.jp/kegg/>).

<sup>25</sup> EC (*Enzyme Commission number*) consiste em uma classificação numérica das enzimas, instituído pelo Comitê de Nomenclatura Bioquímica, baseada nas reações químicas que elas catalisam. As enzimas recebem um código que consiste nas letras EC seguido por quatro números separados por pontos, onde cada número representa uma parte da função enzimática. O primeiro número representa a classe enzimática, o segundo e terceiro número dizem respeito ao composto, grupo, ligação ou produto envolvido na reação e o último número representa metabólitos específicos e co-fatores envolvidos.

<sup>26</sup> GenBank é um banco de dados de livre acesso de sequências genéticas do *National Institute of Health* (NIH, Bethesda, Maryland, EUA), que contém uma coleção anotada de todas as sequências de DNA publicamente disponíveis. É de acesso livre, atualizado a cada dois meses e mantido pelo *National Center for Biotechnology Information* (NCBI). Sua finalidade é prover a comunidade científica das mais atualizadas e abrangentes informações das sequências de DNA.

<sup>27</sup> MAFFT (*Multiple Alignment using Fast Fourier Transform*) é um software gratuito para alinhamento múltiplo de sequências de nucleotídeos ou aminoácidos, baseado na transformada rápida de Fourier e em um sistema de pontuação simplificado. Apresenta acurácia equivalente ao CLUSTALW, com drástica redução do tempo computacional (Katoh *et al.*, 2002).

jModelTest<sup>28</sup> (<http://darwin.uvigo.es/our-software/>) para a escolha do melhor modelo de substituição de nucleotídeos. Foi realizada a inferência das árvores através do programa MEGA<sup>29</sup> (<http://www.megasoftware.net/home>) versão 7.0.26, utilizando como modelo de substituição de nucleotídeos o proposto pelo jModelTest e Máxima Verossimilhança como modelo estatístico, e os valores de *bootstrap* para 1.000 replicatas. Para uma análise conjunta, os diferentes genes foram concatenados, alinhados e submetidos à construção filogenética com os mesmos parâmetros anteriores.

### 3.11. Antígenos compartilhados

Foi realizada uma busca por diversos antígenos e proteínas já descritos na literatura como caracteristicamente imunogênicos e estes foram escolhidos para uma busca ativa contra os dados da anotação do genoma dos isolados micobacterianos sequenciados pelo 454 GS Jr. gerados pelo Prokka.

### 3.12. Obtenção de proteínas secretadas

50 mL de cultura em meio LB/Tw foram crescidas por 48 h a 22 °C. As bactérias foram descartadas após centrifugação a 5.000 g por 15 minutos e o sobrenadante foi filtrado em membrana de PVDF com poro de 0,22 µm (MilliPore). As proteínas do filtrado de cultura foram precipitadas após adição de sulfato de amônio na concentração de 85 % (m/v) e centrifugação a 5.000 g por 15 minutos. O sobrenadante foi descartado e o precipitado ressuspenso em tampão Tris-HCl 10 mM pH 7,2. O material foi adicionado em membrana para diálise e dialisado três vezes em recipiente contendo 2 L do mesmo tampão por 4 h, 18 h e 24 h, sequencialmente. O tampão foi trocado nos intervalos.

### 3.13. Eletroforese bidimensional

A eletroforese bidimensional (2DE) é dividida em duas etapas principais: primeira e segunda dimensão. Para a focalização isoelétrica (primeira dimensão) foi

---

<sup>28</sup> jModelTest é uma versão aprimorada do ModelTest, programa que seleciona o melhor modelo de substituição de nucleotídeos para um grupo de sequências alinhadas. O modelo selecionado pode ser usado para análises filogenéticas (Posada, 2008)

<sup>29</sup> O MEGA (*Molecular Evolutionary Genetics Analysis*) é um software para análises comparativas de sequências de DNA e proteínas cujo objetivo é a inferência de padrões evolutivos moleculares de genes, genomas e espécies.

utilizada fita de gradiente imobilizado de pH (IPG) de 7 cm em diferentes faixas de pH (Biorad) e foram aplicadas aproximadamente 200 µg de proteínas, obtidas do filtrado de cultura, solubilizadas em tampão de focalização isoeletrica suplementado com 20 mM de DTT e 1 % de anfólitos pH 3-10. A rampa de focalização isoeletrica foi iniciada após a reidratação ativa da fita de pH realizada por 11 h a 50 V, conforme o quadro 3.1.

**Quadro 3.1: Etapas para a focalização isoeletrica.**

Géis de 7 centímetros			
1	250 V	Corrente constante	20 minutos
2	4000 V	Corrente constante	2 horas
3	4000 V	12000 VH	Rampagem Rápida
4	500 V	Rampagem Rápida	5 horas

Em seguida, o processo denominado equilíbrio dos strips foi realizado, o qual é fundamental para manter a redução das pontes dissulfeto através da alquilação dos grupos sulfidrilas resultantes dos resíduos de cisteína. Assim, as fitas de IPG foram incubadas por 15 minutos, sob leve agitação à temperatura ambiente no tampão de equilíbrio I (6 M uréia; 0,2 % SDS; 375 mM Tris-HCl pH 8,8; 20 % glicerol; 130 mM DTT). Após o seu descarte, as fitas IPG foram incubadas com o tampão de equilíbrio II (6 M uréia; 0,2 % SDS; 375 mM Tris-HCl pH 8,8; 20 % glicerol; 135 mM iodoacetamida) por 20 minutos à temperatura ambiente, sob leve agitação.

Depois, as tiras de IPG foram lavadas em tampão de eletroforese (250 mM glicina, 25 mM Tris-base e 0,1 % de SDS) e aplicadas cuidadosamente na superfície de um gel 14 % de poliacrilamida. As proteínas foram separadas através de eletroforese em gel de poliacrilamida com dodecil sulfato de sódio (SDS-PAGE, segunda dimensão), perpendicularmente à primeira dimensão, de acordo com sua massa molecular. Usou-se uma voltagem constante de 150 V. Ao final, os géis foram corados por coomassie G-250 coloidal (quadro 3.2), que possui uma sensibilidade de detecção de até 10 ng (Neuhoff *et al.*, 1988) e possibilita a retirada dos *spots* de interesse, os quais serão futuramente analisados por espectrometria de massas.



**Quadro 3.2: Método de coloração com coomassie G-250 coloidal.**

	Solução	Tempo
1	Etanol 30 %, ácido fosfórico 2 % (v/v)	3 x 30 minutos
2	Ácido fosfórico 2 % (v/v)	3 x 20 minutos
3	Ácido fosfórico 2 % (v/v), etanol 18 %, sulfato de amônio 15 % (p/v)	30 minutos
4	Adicionar 2 % do volume da solução 3 contendo Coomassie Blue G-250 a 20 mg/mL	24 a 72 h

### 3.14. Excisão e processamento dos spots

Após visualização do perfil bidimensional das proteínas secretadas, os *spots* proteicos de interesse foram excisados manualmente do gel. O processamento iniciou-se pela descoloração dos *spots* com acetonitrila (ACN) 50 % (v/v) / bicarbonato de amônio 25 mM, pH 8,0. Após o tempo necessário o descorante foi então removido e foi adicionado a cada tubo a solução de ACN 100 % (p/v) (solução de desidratação) por 5 minutos sob agitação contínua. Os *spots* foram colocados no evaporador a vácuo (Speed Vac – Savant) por 15 minutos para evaporação de toda a solução de acetonitrila. Em seguida, foi realizada a etapa de digestão tríptica, adicionando-se 10 µL da solução a 20 ng/µL de tripsina (Tripsin Gold, Mass Spectrometry Grade – PROMEGA). Os tubos foram mantidos no gelo durante este procedimento de digestão enzimática, garantindo que a mesma não sofra degradação. Após o período de 1 h, todo o excesso de tripsina foi removido e foi adicionado 20 µL de solução de bicarbonato de amônio 50 mM, disparando-se a digestão. Os tubos foram incubados por 16 h a 37 °C.

Aos tubos contendo as amostras foi aplicado ultrassom por 10 minutos e, em seguida, foram agitados no vortex. Os 20 µL da solução acima foram removidos para um tubo limpo de 500 µL e o material restante re-extraído com 30 µL de solução de ácido fórmico 5 % (v/v) / ACN 50 % (v/v). Os tubos foram agitados em vortex por 20 s e a solução permaneceu no tubo durante 15 minutos à temperatura ambiente. Após esse período, foi novamente aplicado ultrassom por 2 minutos, seguido por agitação no vortex. A solução contendo os peptídeos foi removida e unida aos 20 µL de solução reservada anteriormente em tubo de 500 µL. A extração foi repetida mais uma vez e todos os extratos foram, ao final do processo, combinados em um mesmo tubo, com um volume final de aproximadamente 60 µL. A amostra foi então concentrada no evaporador a vácuo até atingir o volume aproximado de 5-10 µL. Os

peptídeos foram armazenados a -20 °C para posterior dessalinização e concentração utilizando para isto ponteiras contendo uma resina de fase reversa (Zip Tip C18 – Millipore) seguindo o protocolo do fabricante.

### 3.15. Análise no MALDI e no Orbitrap

As amostras contendo os peptídeos trípticos foram misturadas à matriz de ácido  $\alpha$ -ciano-4-hidroxicinâmico e aplicadas em placa de aço inoxidável para MALDI e entregue à Plataforma de Espectrometria de Massas (RPT02A), alocada no Laboratório de Toxinologia/IOC – RJ, onde foram analisadas em espectrômetro de massas do tipo MALDI TOF-TOF 4700 Proteomics Analyser<sup>30</sup> (Applied Biosystems). Os espectros obtidos foram analisados utilizando o programa Mascot<sup>31</sup> ([http://www.matrixscience.com/cgi/search\\_form.pl?FORMVER=2&SEARCH=MIS](http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS)) versão 2.2. contra o banco de dados NCBIprot (<http://www.ncbi.nlm.nih.gov/protein>) sem restrição taxonômica e contra os dados da anotação dos genomas dos isolados, que foram introduzidos no banco de dados. Para identificação dos peptídeos foram utilizados os parâmetros descritos no quadro 3.3. Esses peptídeos também foram analisados no equipamento LTQ Orbitrap XL<sup>32</sup> (Thermo Scientific). Os dados foram analisados no software Peaks<sup>33</sup>.

---

<sup>30</sup> Ionização por dessorção a laser assistida por matriz (*Matrix Assisted Laser Desorption Ionization*, MALDI) é uma técnica de ionização de espectrometria de massas que utiliza uma matriz que absorve energia laser para criar íons a partir de grandes moléculas. O analisador mais amplamente utilizado em associação ao MALDI é o do tipo tempo de voo (*Time Of Flight*, TOF). Nesse analisador os íons passam dentro de um tubo a vácuo com um campo elétrico onde são separados de acordo com sua razão massa/carga, chegando ao detector em tempos diferentes. Esse processo pode ser otimizado ao acoplar dois analisadores (TOF-TOF).

<sup>31</sup> MASCOT (*Matrix Science*) é uma ferramenta de busca padrão que usa dados de espectrometria de massas para identificar, caracterizar e quantificar proteínas a partir de banco de dados de sequências peptídicas, tais como SwissProt, NCBItr e EST (EMBL).

<sup>32</sup> LTQ Orbitrap XL é um espectrômetro de massas que combina a tecnologia Linear Ion Trap com a tecnologia Orbitrap e utiliza a transformada de Fourier na detecção e conversão do sinal.

<sup>33</sup> PEAKS é um software de análise de dados proteômicos gerados por espectrômetros de massas que auxilia na identificação e quantificação de proteínas, análise de modificações pós-traducionais e mutações, e sequenciamento proteico *de novo*.

**Quadro 3.3: Parâmetros utilizados nas buscas do Mascot.**

Database: NCBI nr	Peptide tol.: $\pm 1.2$ Da
Taxonomy: All entries	MS/MS tol.: $\pm 0.6$ Da
Enzyme: Trypsin	Peptide charge: +1 Da
Allow up to: 2 missed cleavages	Monoisotopic
Fixed modifications: none	Instrument : MALDI TOF-TOF
Variable modifications: Acetyl (protein N-term), Carbamidomethyl (C), Deamidated (NQ), Gln $\rightarrow$ Pyro-Glu (N-termQ), Glu $\rightarrow$ Pyro-Glu (N-termE), Oxidation (HW), Oxidation (M), Phospho (ST), Phospho (Y).	

**3.16. Secretoma**

500 ng de proteínas secretadas foram analisadas na RPT02A em colaboração com o Dr. Richard Hemmi Valente (Laboratório de Toxinologia/IOC) no equipamento Q Exactive Plus Orbitrap<sup>34</sup> (Thermo Scientific). Os dados foram analisados no software Peaks.

**3.17. Proteínas hipotéticas**

A partir da sequência de aminoácidos das proteínas identificadas pelo Q-Exactive, foi realizado uma busca por similaridade através da ferramenta *protein Basic Local Alignment Search Tool* (BLASTp) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Essa busca foi realizada contra o banco de dados de sequências proteicas não redundantes do gênero *Mycobacterium* e outros que não *Mycobacterium*.

**3.18. Quantificação *label-free***

Foi realizada uma quantificação *label-free* para a identificação das proteínas secretadas mais abundantes de cada isolado a partir da contagem do número de espectros (*spectrum counting*) de peptídeos únicos de cada proteína.

**3.19. Capacidade de transformação**

Após adição de 1  $\mu$ g do DNA plasmidial do pUS989 a 200  $\mu$ L de micobactérias eletrocompetentes, foi realizada a transformação por eletroporação no equipamento

<sup>34</sup> QExactive apresenta duas colunas cromatográficas (troca iônica e fase reversa) acopladas ao espectrômetro de massas com ionização por *electrospray* e analisadores quadrupolo e orbitrap *in tandem*.

Gene Pulser™ (BIO-RAD) nas seguintes condições: 2,5 kV; 25 µF e 1.000 Ω durante 15 – 25 ms em cubetas de 2 mm. As bactérias foram imediatamente transferidas para tubo Falcon de 50 mL contendo 5 mL de meio LB/Tw e incubadas a 22 °C (isolados ambientais) ou 37 °C (mc<sup>2</sup>155) nos tempos determinados. Em seguida, 10 µL da suspensão bacteriana foram plaqueados em LB/ágar contendo canamicina (25 µg/mL) (Difco™, Invitrogen) e X-Gal (5 µg/mL) (Promega), incubado a 22 °C (isolados ambientais) ou 37 °C (mc<sup>2</sup>155) e as colônias foram contadas após sete dias.

### 3.20. Análise de sequências reguladoras, promotoras e terminadoras

Diversas análises foram realizadas com os genes *294\_01260* e *294\_04986*, e suas respectivas regiões flanqueadoras, identificadas no genoma do isolado 294 e cujas proteínas foram identificadas a partir dos dados de proteômica.

A sintenia foi analisada através da ferramenta de comparação de regiões do RAST e comparada com mc<sup>2</sup>155. Os genes *294\_01260* e *294\_04986* e suas regiões flanqueadoras foram analisados através do programa SnapGene (<http://www.snapgene.com/>) versão 3.3, buscando identificar outras CDS e sítios para enzimas de restrição.

Foram utilizados quatro diferentes programas para a busca por peptídeo sinal, sendo todos disponibilizados pelo Centro para Análise de Sequências Biológicas (<http://www.cbs.dtu.dk/services/software.php>). O LipoP prediz sítios de clivagem de peptidase sinal do tipo I e II em bactérias Gram negativas. SecretomeP é utilizado para identificar secreção de proteínas sem a sequência líder e não clássicas. TatP prediz peptídeos sinal a partir da via Twin-arginina. SignalP, identifica possíveis sítios de clivagem e de peptídeo sinal em Gram positivas e negativas e em eucariotos a partir da sequência de aminoácidos. Para este programa foi utilizada a busca tanto para Gram positivas como Gram negativas.

O preditor de promotores bacterianos BROM (<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>) foi utilizado para identificar as possíveis regiões -10 e -35 dos genes estudados. O programa mfold (<http://unafold.rna.albany.edu/?q=mfold>), que prediz a formação de estruturas secundárias no DNA que resulta no desligamento da RNA polimerase, foi utilizado para indicar possíveis terminadores de transcrição (Zuker, 2003).

### 3.21. Cultivo de *E. coli*

Culturas de *E. coli* foram crescidas em erlenmeyer de 50 mL contendo 10 mL de meio LB sob agitação a 200 rpm. Cepas de *E. coli* foram reativadas do estoque em nitrogênio líquido por estriamento em placas LB/ágar usando alças bacteriológicas. Para obter colônias isoladas, realizou-se o plaqueamento por espalhamento usando pequenas esferas de vidro. Todas as culturas de *E. coli* foram crescidas a 37 °C.

### 3.22. Extração de DNA plasmidial

A extração do DNA plasmidial foi realizada de acordo com o método de lise enzimática proposto por Birnboim (1983). As células de *E. coli* foram recuperadas por centrifugação a partir de 3 mL de cultura e rompidas através de detergente aniônico e solução alcalina. O DNA cromossômico foi desnaturado durante a lise e precipitado após adição do acetato de potássio, sendo eliminado juntamente com os fragmentos celulares. O DNA plasmidial foi renaturado após retorno das condições normais, precipitado com etanol e ressuspenso em 50 µL de tampão TE.

### 3.23. Preparo de células eletrocompetentes

Diferentes cepas de *E. coli* foram preparadas a partir de cultura em fase exponencial de crescimento,  $DO_{600nm}$  entre 0,350 e 0,400. Após 30 minutos no gelo, a cultura foi centrifugada a 4.000 g por 20 minutos a 4 °C. O sobrenadante foi descartado e o sedimento bacteriano foi lavado com água estéril e gelada em um volume igual a 50 % do original. Repetiu-se mais duas vezes a centrifugação e lavagem com solução aquosa de 10 % glicerol estéril e gelado com 25 % e 1 % do volume original, respectivamente. Centrifugou-se uma última vez e o sedimento foi ressuspenso em 0,1 % do volume original da cultura em meio GYT (glicerol, extrato de levedura e triptona). A suspensão bacteriana foi aliquoteada em microtubos com 40 µL cada e armazenada a -70 °C.

Para o preparo de micobactérias eletrocompetentes, as bactérias foram crescidas em 200 mL de cultura em meio LB/Tw sob agitação a 22 °C até atingir uma  $DO_{600}$  entre 0,8 e 1,0, medido por espectrofotometria mediante leitura no equipamento Ultrospec 1100 Pro® (Amersham Biosciences). Após 90 minutos no gelo, a cultura foi centrifugada a 4.000 g por 20 minutos a 4 °C. O sobrenadante foi

descartado e o sedimento foi lavado com solução aquosa de 10 % glicerol estéril e gelado em um volume igual a 25 % do original. Repetiu-se duas vezes mais a centrifugação e as lavagens com 10 % e 5 % do volume original. Centrifugou-se uma última vez e o sedimento foi ressuspensado em 10 % glicerol com 1 % do volume original da cultura. A suspensão bacteriana foi aliquoteada em microtubos com 200  $\mu$ L e armazenada a -70 °C.

### **3.24. Digestão do vetor pUS989**

Foi realizada a digestão com duas das enzimas de restrição que reconhecem sítios no DNA plasmidial pUS989: *Hind*III (incubadas por duas horas a 37 °C com o tampão NEB 2.1) e *Bam*HI (tampão NEB 3.1). Em seguida, foi realizada a visualização do DNA digerido em gel de agarose 1 %, em TBE 0,5 X.

### **3.25. Construção do vetor I:**

Com base nas informações da capacidade de transformação, das proteínas majoritariamente expressas e secretadas e das sequências nucleotídicas dessas proteínas, dois testes foram realizados para a construção de um novo vetor para expressão de proteínas heterólogas em micobactérias. Ambos os testes foram realizados baseados na técnica de clonagem de Gibson. Para esta primeira construção, utilizou-se o como gene repórter o *mCherry*.

#### **3.25.1 PCRs**

As reações em cadeia da polimerase (*Polymerase Chain Reaction*, PCR) foram realizadas em volume final de 25  $\mu$ L, utilizando o kit Master Mix PCR (Promega), 200 nM de cada iniciador específico (F e R) e 2  $\mu$ L de DNA genômico como DNA molde. Iniciadores específicos foram desenhados para cada região (quadro 3.4) de acordo com as recomendações do método de Clonagem de Gibson (New England Biolabs Inc., NEB) utilizando-se a ferramenta NEBuilder Assembly (NEB). A ciclagem, descrita no quadro 3.5, foi realizada no termociclador GeneAmp PCR System 9700 (Thermo Scientific).

**Quadro 3.4: Oligonucleotídeos usados na construção I.** Sublinhado a parte de sobreposição dos fragmentos.

Iniciador	Sequência
P1 Fwd	<u>TCC TCT AGA GTC GAC</u> CTG CAT GCA TCG AAA TAT TTG TCA GAG CCG CCC
P1 Rev	TCG AGA CCA TGG CAA CAA GCG CGC CGG C
C Fwd	<u>GCT TGT TGC</u> CAT GGT CTC GAA GGG GCA G
C Rev	CAG TCA GTC <u>ACT</u> TGT ACA GCT CGT CCA TGC
T1 Fwd	<u>GCT GTA CAA</u> <u>GTG</u> ACT GAC TGA CCC CTG GTA TTC ACT CAA CGG
T1 Rev	AGC TAG CTT GCA TGC CTG <u>CAG</u> GCG ATA GAT ATT GGC GCC

### 3.25.2 Digestão

O DNA do vetor pUS972 (Medeiros *et al.*, 2002) foi extraído de *E. coli* conforme descrito acima (item 3.21), digerido com a enzima *Pst*I por duas horas a 37 °C em tampão NEB 3.1 e a enzima foi inativada a 65 °C por 30 minutos.

### 3.25.3 Reação de Gibson e transformação

Todo o procedimento foi realizado de acordo com as recomendações do fabricante. Em um microtubo foram adicionados 25 µL de MasterMix de Gibson (NEB), 0,5 ng do DNA plasmidial e 0,5 pM de cada fragmento amplificado como descrito acima e purificado, totalizando 50 µL de reação. Após homogeneizar, a reação foi incubada por uma hora a 50 °C. Em seguida, o material foi diluído com uma fração da reação de Gibson mais duas frações de água MilliQ estéril.

*E. coli* Top10 eletrocompetentes foram transformadas por eletroporação com 1 µL da reação de Gibson diluída, nas seguintes condições: 2,5 kV; 25 µF; 200 ohms. O material foi imediatamente transferido para tubo Falcon de 50 mL com 1 mL de meio LB e incubado por 1 hora sob agitação a 37 °C. Os transformantes foram selecionados em meio LB/ágar contendo canamicina (25 µg/mL) por 18-24 h a 37 °C.

### 3.25.4 PCR de colônia

A presença do inserto foi verificada por PCR de colônia. Foram preparados microtubos contendo kit Master Mix PCR (Promega) e 200 nM de cada iniciador específico (F e R) em um volume final de 20 µL e tubos Falcon contendo 5 mL de meio LB contendo canamicina na concentração 25 µg/mL (LB/Kan). Em seguida, palitos estéreis foram encostados sobre as colônias, encostados no fundo dos

microtubos contendo o mix para a reação de PCR, e depositados no Falcon contendo LB/Kan. Foi realizada a reação de PCR com a ciclagem descrita no quadro 3.5 para o fragmento 294\_01260 I. Foi realizada a extração plasmidial dos clones positivos a partir das culturas em tubo Falcon, conforme item 3.21.

**Quadro 3.5: Ciclagens para a reação de PCR**

Construção I				Construção II			
Fragmento	Temp (°C)	Tempo	Ciclos	Fragmento	Temp (°C)	Tempo	Ciclos
P1 (região promotora 294_1260)	94	5 min	1	P2 (região promotora 294_1260)	94	5 min	1
	94	30 s	5		94	30 s	
	58	30 s			50	30 s	
	72	30 s			72	30 s	
	94	30 s	30		94	30 s	
	72	30 s			55	30 s	
72	30 s	72		30 s			
	72	7 min	1		72	7 min	1
C (mCherry de pCherry3)	94	5 min	1	M (Mpt64 de <i>M. bovis</i> BCG)	94	5 min	1
	94	30 s	5		94	30 s	
	50	30 s			55	30 s	
	72	1 min			72	45 s	
	94	30 s	30		94	30 s	
	72	30 s			60	30 s	
72	1 min	72		45 s			
	72	7 min	1		72	7 min	1
T1 (região terminadora 294_1260)	94	5 min	1	T2 (região terminadora 294_1260)	94	5 min	1
	94	30 s	5		94	30 s	
	55	30 s			50	30 s	
	72	30 s			72	30 s	
	94	30 s	30		94	30 s	
	72	30 s			60	30 s	
72	30 s	72		30 s			
	72	7 min	1		72	7 min	1
294_1260 I	94	5 min	1	294_1260 II	94	5 min	1
	94	30 s	5		94	30 s	
	55	30 s			50	30 s	
	72	2 min			72	2 min	
	94	30 s	30		94	30 s	
	72	30 s			62	30 s	
72	2 min	72		2 min			
	72	7 min	1		72	7 min	1



### 3.26. Construção do vetor II

Nesta construção foi utilizado o *Mpt64* com gene repórter. As reações de PCR foram realizadas conforme item 3.24.1 com novos iniciadores (quadro 3.6) e ciclagem (quadro 3.5). O vetor foi digerido primeiramente com a enzima *NdeI* por duas horas a 37 °C em tampão NEB 2.1, inativada por 30 min a 65 °C, e, em seguida, pela enzima *SphI* nas mesmas condições. A reação de Gibson, transformação e PCR de colônia foi realizado conforme item 3.24.3 e 3.24.4.

**Quadro 3.6: Oligonucleotídeos usados na construção II.** Sublinhado a parte de sobreposição dos fragmentos.

Iniciador	Sequência
P2 Fwd	<u>AAG TAC TGG TCA AGC TAG CAT</u> CTG CAT CGA AAG ATT TGT CAG A
P2 Rev	<u>TAG GTC TTG GGT GCG</u> GCA ACA AGC GCT CCG
M Fwd	<u>TTG TTG CCG</u> CGC CTA AGA CCT ACT GC
M Rev	<u>GTT AGT CAG CTA</u> GGC CAG CAT CGA ATC G
T2 Fwd	<u>GAT GCT GGC ATA</u> GCT GAC TAA CCC CTG GTA TTC ACT CAA
T2 Rev	<u>AGT CGA TCT GCA GGC ATG</u> CGT TTG GGA CCA TAC GGC TGA

### 3.27. Cultura de células THP-1

Células THP-1 foram gentilmente cedidas pelo Dr. Milton Ozório Moraes (Laboratório de Hanseníase, IOC/Fiocruz – RJ) e foram mantidas em garrafas de cultura de células em meio RPMI-1640 contendo L-Glutamina (Sigma); suplementado com 10 % de soro fetal bovino (SFB) (Sigma), e L-Glutamina 200 mM, penicilina 10.000 U e estreptomicina 10 mg/mL (PESS) (Sigma) (RPMI completo), a 37 °C em atmosfera de 5 % de CO<sub>2</sub>.

As células foram mantidas até a terceira passagem com troca de meio a cada três dias. Na terceira passagem, as células foram centrifugadas a 450 x g por 10 minutos a 4 °C. Em seguida, as células foram contadas em câmara de Neubauer e plaqueadas e diferenciadas em macrófagos (protocolo descrito no item a seguir) em placas de 24 poços (Nunc). Foram plaqueadas 2 x 10<sup>5</sup> células/poço, conforme padronizado previamente no laboratório. As células foram observadas em microscópio invertido Eclipse TS100 (Nikon).

### 3.28. Diferenciação dos monócitos THP-1 em macrófagos

Para os ensaios de infecção, as células foram quantificadas em câmara de Neubauer e a viabilidade foi avaliada pelo método de exclusão por coloração com 0,4 % de azul de Tripán. Posteriormente, as células foram estimuladas com 200 nM de acetato de forbol-miristila (PMA) (Sigma-Aldrich) em meio RPMI acrescido de 10 % de SFB (RPMI/SFB) pré-aquecido para a diferenciação em macrófagos (dTHP-1). Após 48 h de estímulo, as células foram lavadas para a retirada do PMA e então adicionado meio RPMI/SFB pré-aquecido por 48 h, com troca diária.

### 3.29. Quantificação dos lotes-semente

Os isolados micobacterianos foram crescidos por 48 h em meio LB/Tw, foi adicionado 15 % de glicerol (v/v), alíquotado em microtubos com 1 mL cada e estocados à -70 °C. Após 7 dias, foi realizado a quantificação através da determinação de unidades formadoras de colônias (*colony forming units*, CFU). Alíquotas dos lotes-semente foram descongeladas por 10 minutos a 22 °C, centrifugadas por 10 minutos a 16.000 x g e o sobrenadante descartado. O sedimento bacteriano foi ressuspensão em 1 mL de meio LB/Tw e os grumos de bactérias foram mecanicamente dissociados através de sucessivas passagens em seringa de 27,5 G X ½". A partir dessa suspensão bacteriana foram feitas diluições seriadas. O equivalente a 100 µL de cada diluição foi plaqueado por espalhamento com pequenas esferas de vidro em meio LB/ágar e as placas mantidas em estufa a 22 °C. As colônias foram contadas sete dias após o plaqueamento. O mesmo procedimento foi realizado para mc<sup>2</sup>155 variando apenas a temperatura, 37 °C.

### 3.30. Viabilidade intracelular dos isolados

Alíquotas dos lotes-semente foram descongeladas e centrifugadas e o sedimento bacteriano foi ressuspensão e dissociado conforme o item anterior. Os isolados foram utilizados para infectar as células diferenciadas usando multiplicidade de infecção de 10 bactérias para cada célula (*multiplicity of infection*, MOI 10:1), de acordo com a quantificação dos lotes-semente. Ou seja, 2 x 10<sup>6</sup> bactérias por poço da placa de 24 poços.

Ao final do tempo de infecção, 4 h, o meio de cultura foi removido e as células lavadas três vezes com meio RPMI, a fim de retirar as bactérias que não foram

fagocitadas, e foi adicionado meio RPMI/SFB pré-aquecido. Foi realizada a troca do meio a cada 24 h. A viabilidade intracelular dos isolados foi avaliada através de uma cinética de tempo, que compreendia os tempos de 4, 24, 48 e 72 h após a infecção. Nos tempos selecionados, os mdTHP-1 infectados foram lisados com 0,05 % SDS (Venkataswamy *et al.*, 2012) e as bactérias foram recolhidas e centrifugadas por 10 minutos a 16.000 x *g*. O sobrenadante foi descartado e o sedimento bacteriano foi ressuspenso em 1 mL de meio LB/Tw. A partir dessa suspensão bacteriana foram feitas diluições seriadas e plaqueamento conforme o descrito no item anterior.

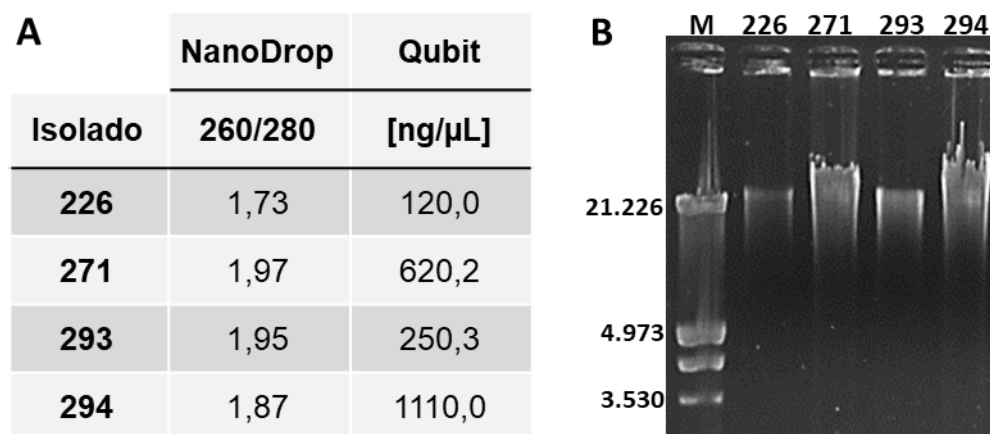
## 4. Resultados

Os resultados foram agrupados em quatro áreas diferentes: análise genômica, análise do secretoma, construção de um novo vetor e infecção em células THP-1.

### 4.1. Análise genômica

#### 4.1.1. Sequenciamento

A partir do sedimento bacteriano de culturas no final da fase log de crescimento, foi extraído o DNA genômico dos quatro isolados, a concentração e pureza foram determinadas utilizando-se o Qubit e NanoDrop e o material foi observado em gel de agarose 0,8 % (figura 4.1). Os isolados 271, 293 e 294 apresentaram uma alta concentração e pureza do material, e o isolado 226 apresentou uma pureza pouco abaixo do recomendado. Apesar da etapa de lise mecânica no processo de extração, a pequena fragmentação do DNA resultante desse processo não afetou consideravelmente a integridade do material.



**Figura 4.1. Extração do DNA genômico.** DNA genômico dos isolados micobacterianos extraído com o kit *Blood & Cell Culture DNA Mini* (Qiagen). (A) Pureza e concentração das amostras. (B) Gel de agarose 0,8 % corado com brometo de etídeo. Foi usado como marcador de tamanho molecular (M) o fago  $\lambda$  digerido com as endonucleases *EcoRI* e *HindIII* ( $\lambda$ E/H). Tamanho molecular indicado à esquerda, em pares de base (pb).

O DNA genômico foi encaminhado para a RPT01J para sequenciamento nos equipamentos 454 GS Jr. e HiSeq. Foi observado um aumento médio de 25 e 50 vezes na quantidade de bases e *reads*, respectivamente, geradas pelo HiSeq em

relação ao 454 GS Jr. (tabela 4.1). Analisando os dados do 454 GS Jr., o sequenciamento do isolado 271 gerou uma quantidade de dados muito maior que os demais isolados (entre 22 e 56 % de aumento). O oposto foi observado nos dados do HiSeq, onde 271 obteve a menor quantidade de base e *reads* entre os isolados.

**Tabela 4.1: Sequenciamento dos isolados.** O DNA genômico de quatro micobactérias isoladas da Mata Atlântica foi sequenciado através de duas plataformas de sequenciamento de nova geração: 454 GS Jr. e HiSeq

Isolado	454 GS Jr.		HiSeq	
	Bases	Reads	Bases	Reads
226	85,68 Mb	184.056	518,04 MB	8.672.292
271	108,43 Mb	225.404	475,25 MB	8.038.252
293	76,63 Mb	161.175	629,16 MB	10.442.120
294	65,76 Mb	144.165	617,30 MB	10.255.668

#### 4.1.2. Qualidade das sequências

O arquivo contendo as *reads* em formato FASTA e o arquivo contendo os valores de qualidade em formato *qual* foram convertidos em um arquivo FASTQ<sup>35</sup>, com o auxílio do programa Phred33 conversion (<http://www.mrdnalab.com/16freesoftware/fasta-qual-fastq-conversion.html>) versão 17.03.22. A qualidade dos dados foi analisada pelo programa FastQC. O gráfico de distribuição de comprimento de *reads* (figura 4.2) mostra que os isolados 226, 271 e 294 apresentam *reads* com tamanho entre 40 e 1195 bases e o isolado 293, entre 40 e 875 bases. No entanto, o tamanho da grande maioria está entre 400 e 550 bases para os 4 isolados. De acordo com o gráfico de qualidade da sequência por base (figura 4.2), onde verde representa uma boa qualidade, amarelo média e vermelho ruim e a curva azul é a média, percebemos que a qualidade reduz a partir da 320<sup>o</sup> base nos *reads* dos isolados 226 e 271 e da 240<sup>o</sup> base nos *reads* dos isolados 293 e 294.

Os dados do HiSeq também foram analisados pelo FastQC. O gráfico de distribuição de comprimento de *reads* (figura 4.3) mostra que os 4 isolados apresentam sequências de tamanho entre 35 e 101 bases com a maciça maioria

<sup>35</sup> Arquivos FASTQ são utilizados para armazenar sequências biológicas com seus valores de qualidade. É a saída padrão de diversos sequenciadores e têm o objetivo de controlar a precisão com que a base foi identificada. Caracterizado pelo símbolo @ no início do cabeçalho seguido da sequência na próxima linha. Os valores de qualidade estão na escala PHRED, mas são codificados com caracteres ASCII para ter sua representação compactada (Cock *et al.*, 2010)

tendo 100 bases. De acordo com os gráficos de qualidade da sequência por base (figura 4.3) e por coordenada gerados pelo FastQC, as *reads* apresentaram uma alta qualidade, com valor médio de PHRED<sup>36</sup> entre 39 e 40, e todas as bases com valores de qualidade acima da média. A qualidade dos dados não reduziu após concatenar as sequências.

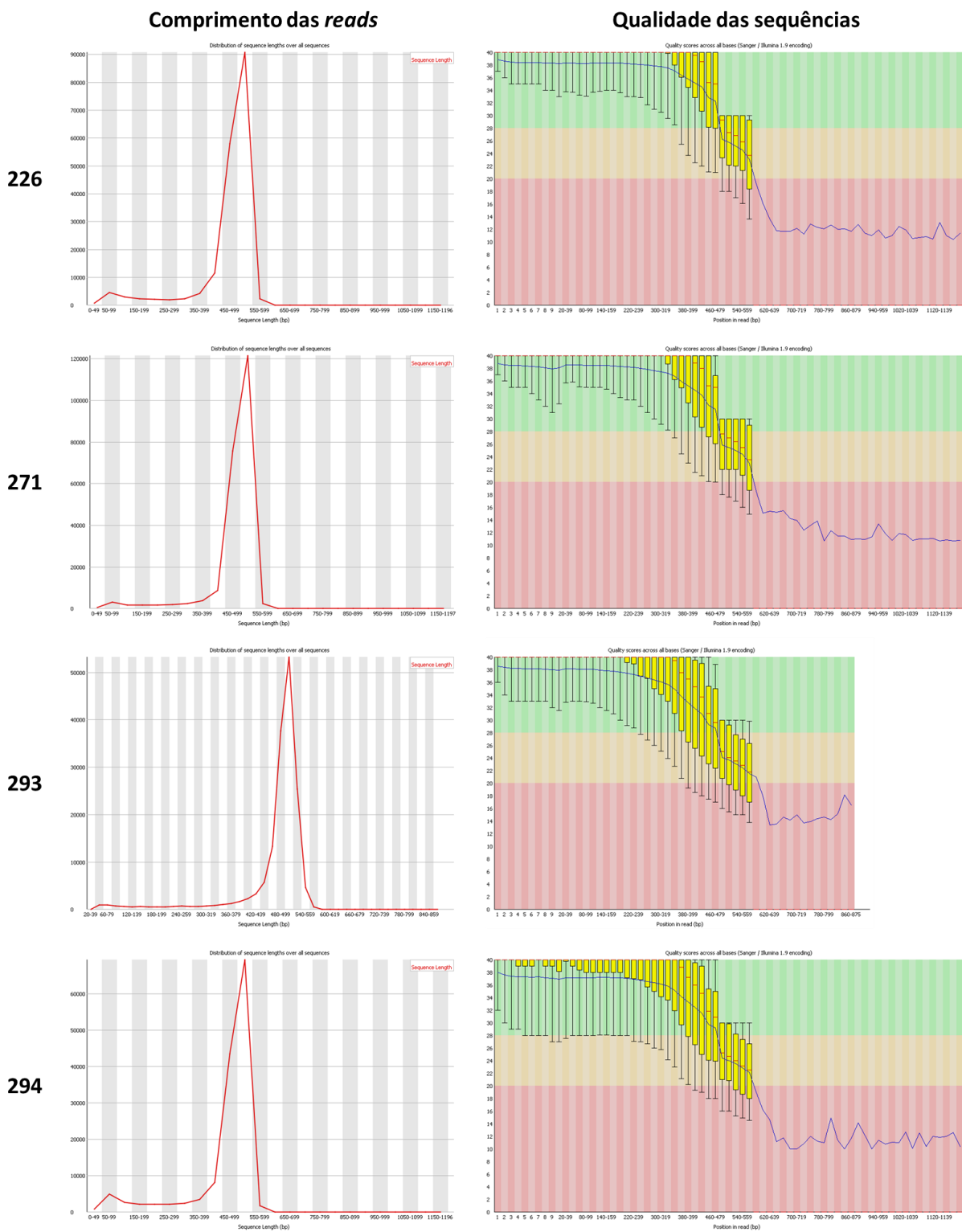
#### 4.1.3. Montagem

Após o sequenciamento pelo 454 GS Jr. foi realizada uma pré-montagem automática com o software Newbler com o objetivo de visualizar os dados obtidos do sequenciamento. Uma nova montagem das *reads* foi realizada posteriormente em colaboração com MSc. Ricardo Junqueira (RPT01J), usando três softwares distintos, Newbler, MIRA e CABOG, e integrando-as com o CISA (tabela 4.2). Os isolados tiveram mais de 99 % do genoma montado. A variação da cobertura, do conteúdo GC e do tamanho do genoma após o refinamento da montagem foi muito pequena ou nula. No entanto, houve uma redução significativa no número de *contigs* e, conseqüentemente, um aumento no N50<sup>37</sup> de todos os isolados estudados. Observando as diferenças entre os isolados, notou-se que o isolado 271 apresenta uma cobertura maior e tamanho menor, resultando em um número de *contigs* muito menor e, conseqüentemente, maior valor de N50. Os isolados 226, 293 e 294 apresentaram valores similares nos parâmetros avaliados. A montagem final estimou a quantidade de genes, encontrando valores próximos a 5.500 para os isolados 226, 293 e 294 e redução de mais de mil genes para o isolado 271.

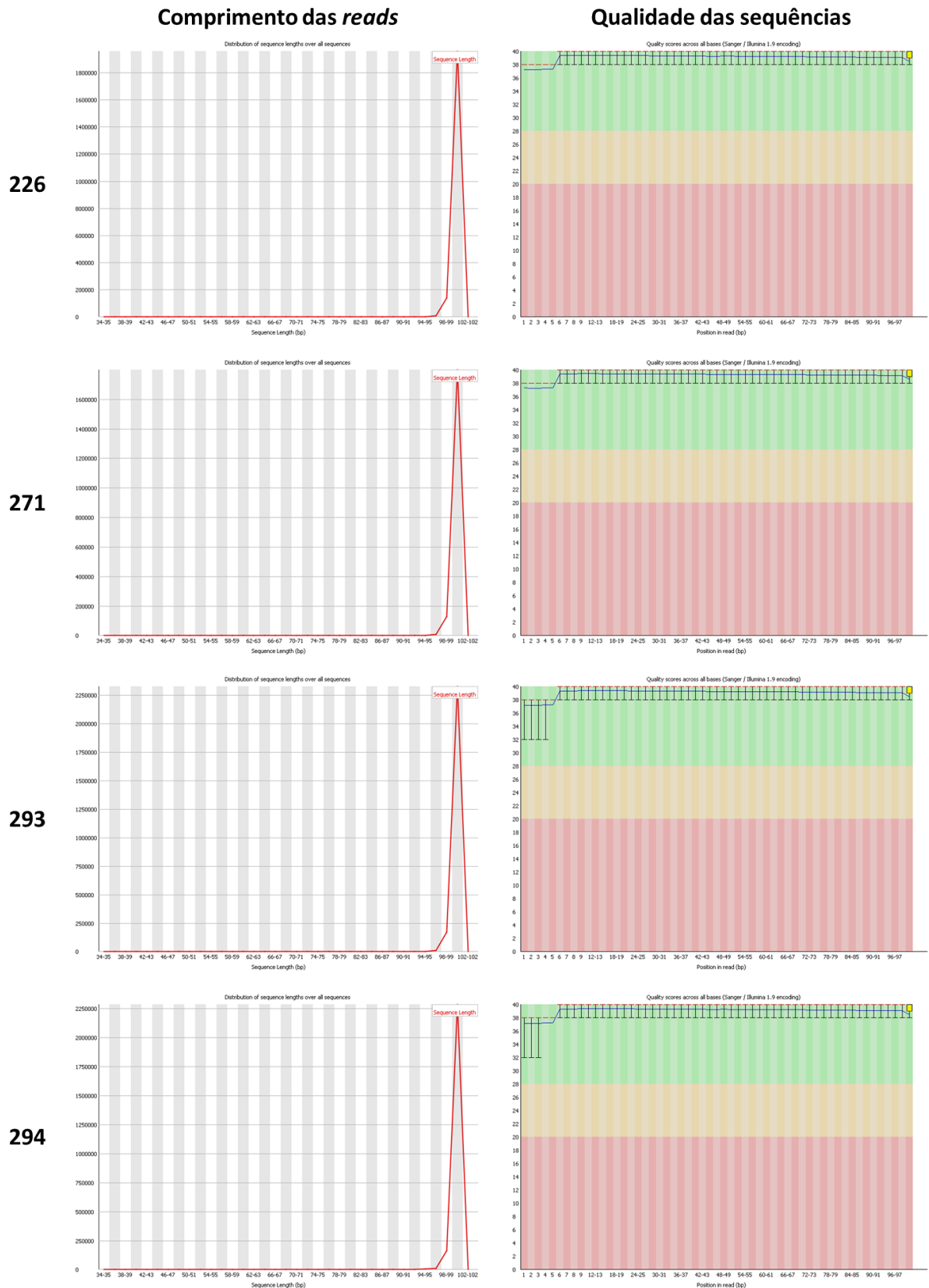
---

<sup>36</sup> PHRED scores indica a probabilidade de uma base estar incorreta. O valor do PHRED score de "X" corresponde a uma probabilidade de erro de  $10^{-x/10}$ . Por exemplo, um score de 30 corresponde a uma acurácia de 99,9 % ( $10^{-3}$  de erro) (Ewing *et al.*, 1998).

<sup>37</sup> N50 é uma medida estatística muito utilizada para avaliar a qualidade da montagem. O valor de N50 indica que 50 % das *reads* estão montadas em *contigs* deste tamanho ou maior (Verli, 2014).



**Figura 4.2: Gráficos de qualidade do sequenciamento pelo 454 GS Jr. gerados pelo FastQC.** A coluna da esquerda mostra os gráficos de distribuição de comprimento de *reads* e a coluna da direita mostra os gráficos de qualidade da sequência por base.



**Figura 4.3: Gráficos de qualidade do sequenciamento pelo HiSeq 2500 gerados pelo FastQC.** A coluna da esquerda mostra os gráficos de distribuição de comprimento de *reads* e a coluna da direita mostra os gráficos de qualidade da sequência por base.



**Tabela 4.2: Montagem dos genomas sequenciados no 454 GS Jr.**

Isolado	Pré-Montagem					Montagem					
	Cobert. (x)	Tam. (Mb)	Contigs	GC (%)	N50 (kb)	Cobert. (x)	Tam. (Mb)	Contigs	GC (%)	N50 (kb)	Genes preditos
226	13,9	6,18	69	66,44	165,0	13,7	6,24	44	66,44	199,8	5.440
271	22,6	4,80	17	64,03	537,8	22,6	4,79	5	64,04	1.083,6	4.224
293	12,1	6,34	92	65,48	176,8	12,0	6,39	63	65,47	206,1	5.572
294	10,4	6,31	89	66,24	124,0	10,4	6,32	67	66,23	140,5	5.413

A partir das sequências do HiSeq concatenadas, foi realizada a montagem com o auxílio do programa PATRIC (<https://www.patricbrc.org/>) (tabela 4.3). Os isolados 226, 293 e 294 apresentaram genoma com tamanho em torno de 6,3 Mb e 271, em torno de 4,8 Mb, similar aos dados obtidos com o 454 GS Jr. O conteúdo GC variou muito pouco. Observou-se uma cobertura de sequenciamento próximo a 100 vezes para os isolados 271, 293 e 294 e próximo a 80 vezes para 226, representando um aumento de 5 a 10 vezes em comparação com a cobertura obtida no 454 GS Jr. No entanto, a montagem resultou em número de *contigs* maior que a montagem do 454 GS Jr., exceto 294 que apresentou redução de 30 %. E, apesar de um maior número de *contigs*, os isolados 226 e 293 apresentaram N50 maior em relação aos dados do 454 GS Jr.

**Tabela 4.3. Montagem dos genomas sequenciados no HiSeq.**

Isolado	Cobertura (x)	Tamanho (Mb)	Contigs	GC (%)	N50 (kb)
226	83,8	6,18	48	66,45	376,9
271	99,0	4,80	65	64,03	142,4
293	99,2	6,34	69	65,48	288,2
294	97,7	6,32	43	66,23	321,8

#### 4.1.4. Anotação

Os arquivos em formato FASTA contendo os grandes *contigs* do 454 GS Jr. foram anotados pelos programas Prokka e RAST (tabela 4.4). Ambos anotadores encontraram valores muito próximos relativos à quantidade de sequências codificantes (*coding sequence*, CDS) e ao número de tRNAs. Observou-se que o número de CDS preditos para os genomas dos isolados 226, 293 e 294 foi de cerca

de 6.000 e para o do isolado 271, próximo a 4.700. O isolado 294 apresenta uma quantidade significativamente maior de tRNAs. Através do Prokka foi observada a ausência da subunidade 23S do rRNA no isolado 226 e a presença de três cópias extras da subunidade 5S do rRNA em 293 e 294. A análise pelo RAST mostrou que o isolado 271 apresenta 7 % menos subsistemas e cerca de 24 % menos genes dentro de subsistemas. A distribuição gênica dos isolados em subsistemas foi comparada com mc<sup>2</sup>155 e H37Rv (figura 4.4), onde observou-se que H37Rv é a que apresenta mais diferenças em relação às demais.

Os dados do HiSeq foram anotados através do programa RAST e do PATRIC a partir do arquivo contendo todos os *contigs* da montagem (tabela 4.5). A quantidade de CDS observada nos isolados é parecida entre ambas anotações e similar à observada com os dados do 454 GS Jr. De acordo os dados do PATRIC, os quatro isolados apresentam apenas três rRNAs. Tal como observado na anotação anterior, 294 apresenta mais tRNAs que os demais isolados. Visto o interesse em identificar elementos que auxiliem classificação destes isolados como patogênicos ou saprófitos, olhamos mais atentamente para os genes relacionados a doenças infecciosas, conforme definido no PATRIC (tabela 4.6). Os isolados 226 e 293 apresentaram quantidades semelhantes dos genes observados nesta categoria. O isolado 294 apresenta o maior número de fatores de virulência e o menor número de genes de resistência a antibióticos. Comparativamente, mc<sup>2</sup>155 e H37Rv apresentam respectivamente 30 % e 300 % mais genes nessas categorias, que os isolados da Mata Atlântica. O RAST também classifica os genes em diversas categorias funcionais (ou subsistemas). Entre esses subsistemas, destaco o de “virulência, doença e defesa”, na qual os isolados apresentaram uma quantidade de genes muito similar a mc<sup>2</sup>155 e H37Rv; com exceção para o isolado 294, que apresentou cerca de 25 % mais genes nessa categoria que os demais (tabela 4.6). A análise pelo RAST mostrou diferenças mínimas na quantidade de subsistemas e na distribuição gênica dos isolados em subsistemas resultantes da análise comparativa entre ambas metodologias (figura 4.5). Através dos dados da anotação do HiSeq foram gerados os mapas dos genomas dos quatro isolados com o PATRIC (figura 4.6).

**Tabela 4.4. Anotação dos genomas sequenciados no 454 GS Jr.**

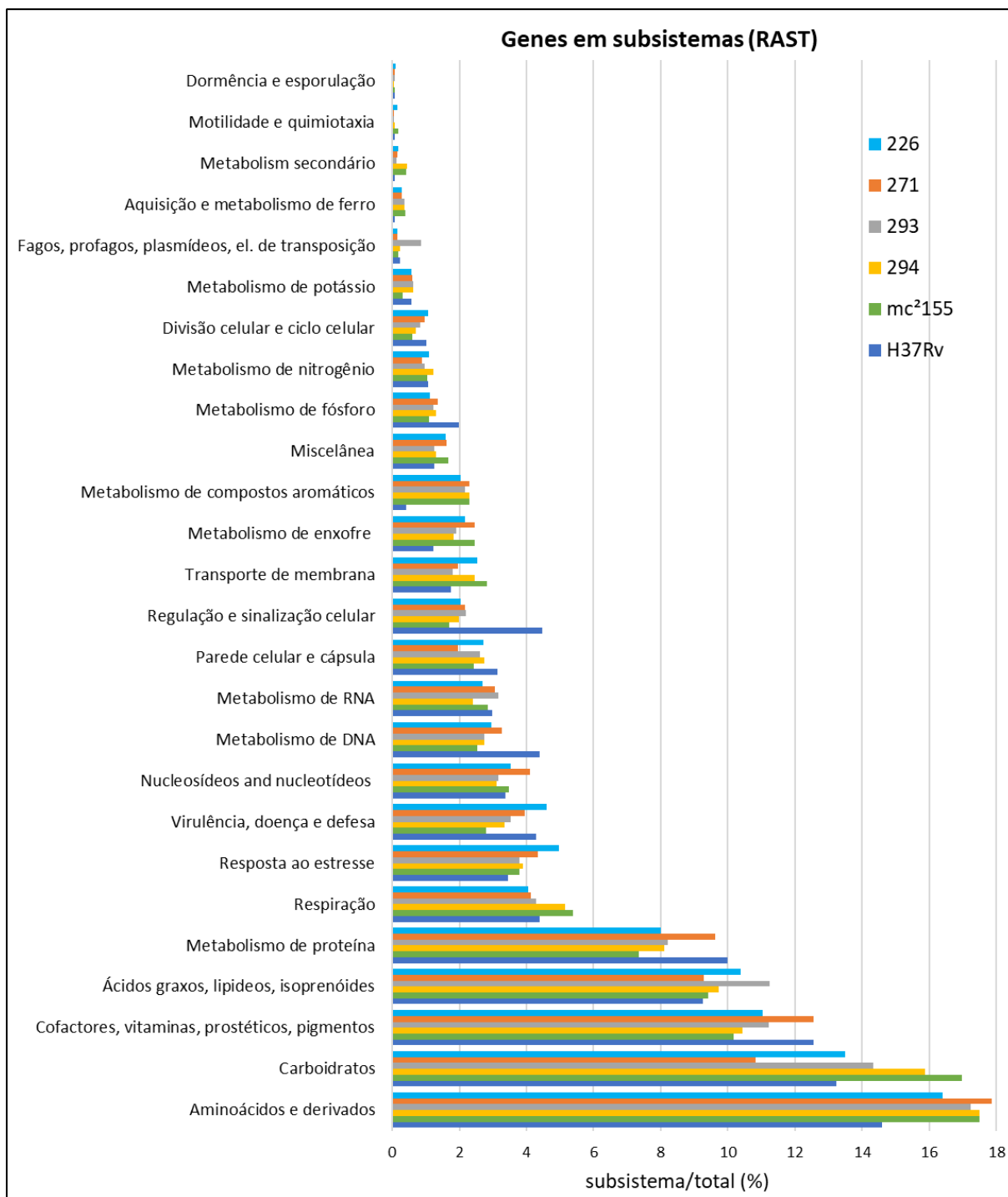
Isolado	Prokka				RAST					
	CDS	rRNA	tRNA	ncRNA	CDS	Genes perdidos	RNA	Subsistemas	Genes dentro de subsistemas	Genes fora de subsistemas
<b>226</b>	5.968	2	54	26	6.002	54	53	422	2.101	3.901
<b>271</b>	4.654	3	48	20	4.734	14	48	391	1.611	3.123
<b>293</b>	6.117	6	51	24	6.265	35	60	417	2.061	4.204
<b>294</b>	5.978	6	76	26	6.032	40	73	433	2.184	3.848

**Tabela 4.5. Anotação dos genomas sequenciados no HiSeq.**

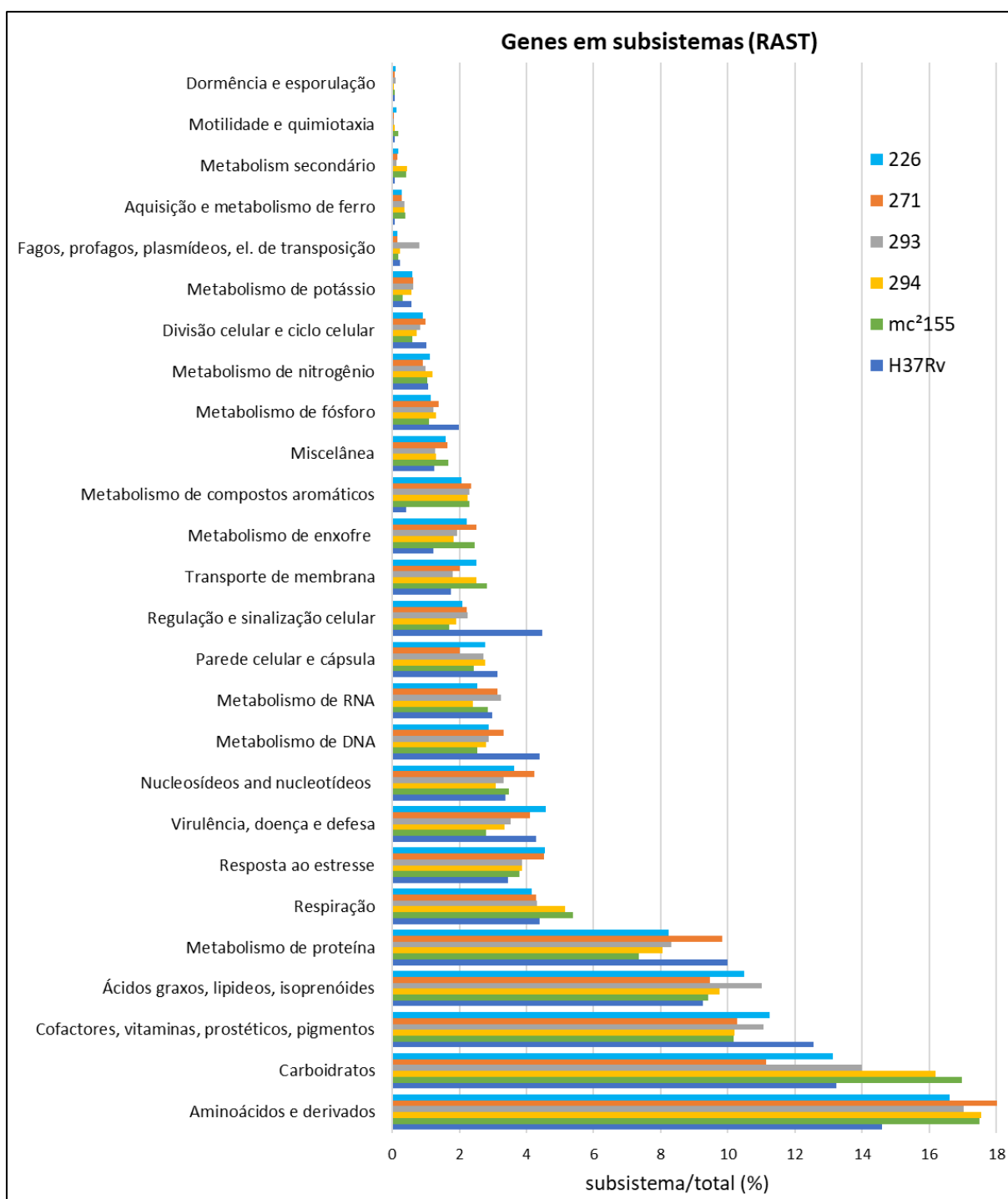
Isolado	PATRIC			RAST					
	CDS	rRNA	tRNA	CDS	Genes perdidos	RNAs	Subsistemas	Genes dentro de subsistemas	Genes fora de subsistemas
<b>226</b>	6.029	3	46	5.919	49	49	421	2.045	3.874
<b>271</b>	4.784	3	46	4.743	36	49	389	1.616	3.127
<b>293</b>	6.277	3	48	6.185	32	51	417	2.039	4.146
<b>294</b>	6.128	3	63	6.004	30	66	434	2.195	3.809

**Tabela 4.6. Genes relacionados a virulência e patogenicidade.** Dados gerados pelo PATRIC, na anotação especial de genes relacionados a doenças infecciosas, e pelo RAST na classificação no subsistema virulência, doença e defesa.

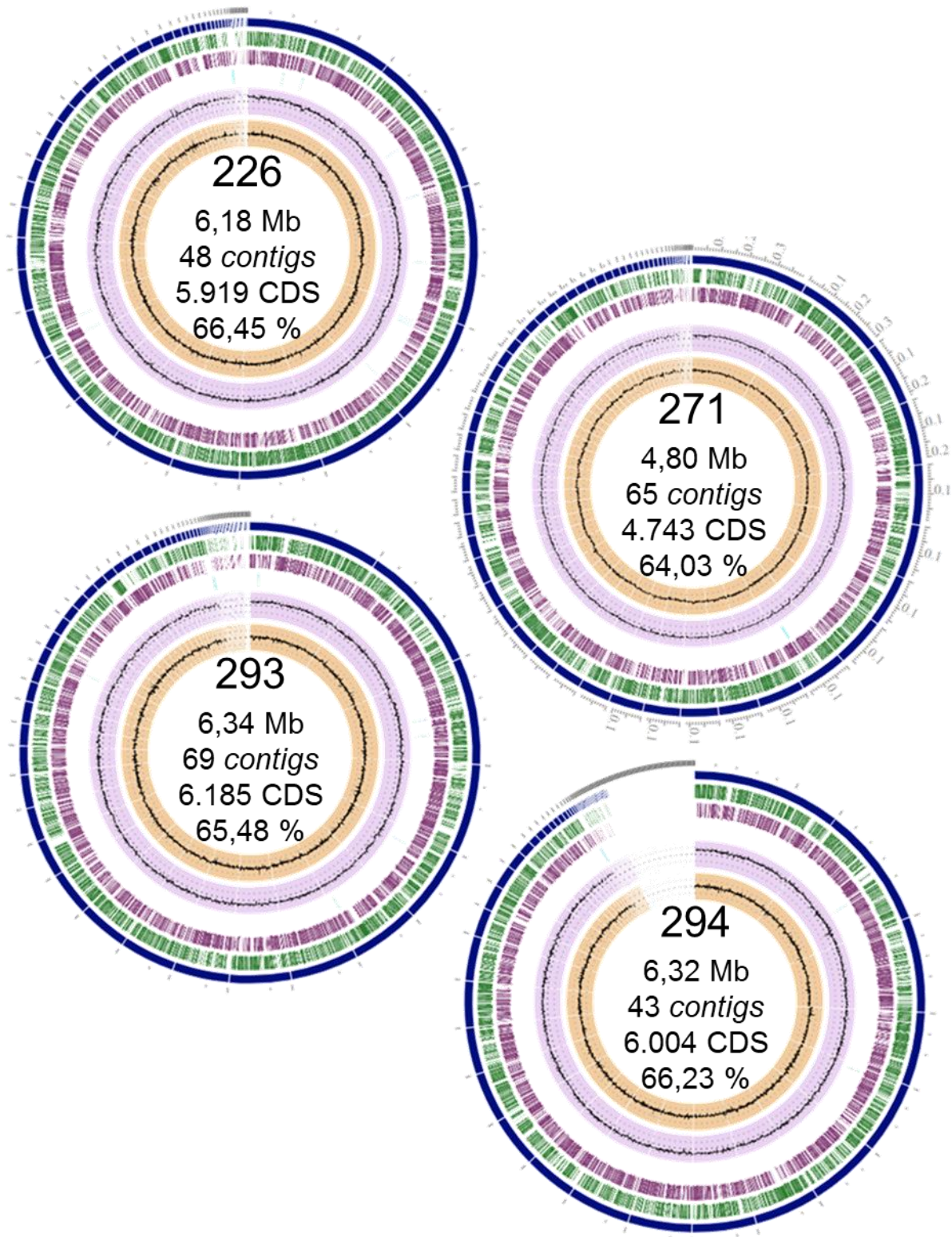
Isolado	PATRIC				Total	RAST	
	Fatores de virulência	Transportador	Alvo de drogas	Resistência a antibiótico		454	HiSeq
<b>226</b>	104	15	25	33	177	151	146
<b>271</b>	59	12	20	24	115	98	100
<b>293</b>	115	12	24	34	185	119	117
<b>294</b>	152	27	30	45	254	118	119
<b>mc<sup>2</sup>155</b>	140	45	33	45	263	105	
<b>H37Rv</b>	806	118	100	60	1.084	116	



**Figura 4.4. Distribuição gênica em subsistemas por isolado após análise no RAST Server dos dados gerados pelo 454 GS Jr.** Após montagem e anotação no RAST, foi comparada a distribuição gênica em subsistemas entre os isolados e duas espécies de referência, mc<sup>2</sup>155 e H37Rv. Esses dados foram observados no SEED Viewer, sendo agrupados em 26 subsistemas. Os dados foram normalizados com base no número total de subsistemas de cada micobactéria. O eixo vertical mostra os diferentes subsistemas e o horizontal, o percentual de genes preditos em cada um.



**Figura 4.5. Distribuição gênica em subsistemas por isolado após análise no RAST Server dos dados gerados pelo HiSeq.** Após montagem e anotação no RAST, foi comparada a distribuição gênica em subsistemas entre os isolados e duas espécies de referência, mc<sup>2</sup>155 e H37Rv. Esses dados foram observados no SEED Viewer, sendo agrupados em 26 subsistemas. Os dados foram normalizados com base no número total de subsistemas de cada micobactéria. O eixo vertical mostra os diferentes subsistemas e o horizontal, o percentual de genes preditos em cada um.



**Figura 4.6: Genoma dos isolados da Mata Atlântica.** Mapa circular da anotação gerados pelo PATRIC através dos dados do HiSeq. Os traços mais externos (cinza) mostram a escala em Mb. O primeiro anel de fora para dentro (azul escuro) mostra os *contigs*. O segundo e o terceiro anel representam as sequências codificantes por fita (sentido horário, verde; anti-horário, roxo). O quarto anel (azul claro) mostra as sequências não codificantes. O quinto e sexto anel representam o conteúdo GC: a quantidade absoluta é mostrada em lilás e a variação em relação a média em laranja. No interior de cada mapa estão contidas as informações sobre o tamanho estimado do genoma, o número de contigs, de CDSs e o conteúdo GC.

#### 4.1.5. Mapeamento metabólico

A análise comparativa através do KEGG forneceu mais de 160 mapas das diferentes vias metabólicas. Para cada via, foi descrito o total de diferentes enzimas presentes com base no *EC number* e a quantidade de genes codificantes para essas enzimas em cada isolado estudado, em três espécies ambientais (mc<sup>2</sup>155, *M. vaccae* e *M. sp. JLS*), uma patogênica (H37Rv) e a vacina BCG (*M. bovis* BCG Pasteur). *M. sp. JLS* é uma espécie de crescimento rápido, que foi isolada de uma unidade de tratamento de terra onde foi realizada a remediação de hidrocarbonetos policíclicos aromáticos (PAH), sendo caracterizado por mineralizar esses hidrocarbonetos com poucos ou muitos anéis (tabela 4.7). A título de exemplo, podemos observar a via de degradação do benzoato com as enzimas identificadas pela anotação do genoma de isolado 226 destacadas (figura 4.7). Neste exemplo, observa-se que a partir do catecol, anel benzênico com duas hidroxilas, este isolado é capaz (com exceção de uma enzima não identificada) de sintetizar piruvato ou acetil CoA, podendo utilizar esse composto como fonte de carbono tanto em anaerobiose quanto em aerobiose.

**Tabela 4.7: Mapeamento metabólico.**

Mapas do KEGG	ECs Totais	226	271	293	294	mc <sup>2</sup> 155	<i>M. vaccae</i>	<i>M. sp. JLS</i>	BCG	H37Rv
<i>Androgen and estrogen metabolism</i>	24	4	1	4	4	4	4	3	2	2
<i>Arginine and proline metabolism</i>	97	32	28	32	35	40	37	31	29	29
<i>Benzoate degradation via CoA ligation</i>	44	19	15	19	16	22	20	20	17	17
<i>Benzoate degradation via hydroxylation</i>	50	14	12	16	15	21	19	21	8	8
<i>Beta-Lactam resistance</i>	1	1	1	1	1	1	1	1	1	1
<i>Biosynthesis of alkaloids derived from histidine and purine</i>	62	47	47	47	49	47	47	46	46	46
<i>Biosynthesis of alkaloids derived from ornithine, lysine, nicotinic acid</i>	71	34	36	35	39	36	36	35	33	33
<i>Biosynthesis of alkaloids derived from shikimate pathway</i>	101	43	42	43	43	42	42	41	38	38
<i>Biosynthesis of unsaturated fatty acids</i>	15	8	9	9	9	9	8	9	7	7
<i>Biosynthesis of vancomycin group antibiotics</i>	1	1	1	1	1	1	1	1	1	1
<i>Bisphenol A degradation</i>	9	2	2	2	2	2	2	2	3	3
<i>Brassinosteroid biosynthesis</i>	3	3	3	3	3	3	3	3	3	3
<i>Butanoate metabolism</i>	52	23	18	24	22	27	26	24	21	21
<i>Carbazole degradation</i>	11	5	3	4	4	4	4	5	2	2

<i>Citrate cycle (TCA cycle)</i>	22	14	15	14	16	14	15	14	15	15
<i>Cysteine and methionine metabolism</i>	64	20	19	21	18	21	18	20	24	24
<i>Dichlorobenzene degradation</i>	22	9	6	9	9	10	10	11	5	5
<i>Dichlorobenzoate degradation</i>	29	4	2	1	3	5	1	3	1	5
<i>Fluorobenzoate degradation</i>	10	2	1	4	2	4	5	4	1	1
<i>Glycerolipid metabolism</i>	36	11	7	11	11	12	11	10	9	9
<i>Glycine, serine, threonine metabolism</i>	57	27	23	26	28	30	27	26	24	24
<i>Glycolysis / Gluconeogenesis</i>	41	20	19	21	21	22	20	20	19	19
<i>Glycosphingolipid biosynthesis - globo series</i>	14	4	3	4	4	4	3	3	3	3
<i>Indole alkaloid biosynthesis</i>	17	1	1	1	1	1	1	0	0	0
<i>Isoflavonoid biosynthesis</i>	18	1	1	1	1	1	1	1	2	2
<i>Limonene and pinene degradation</i>	22	8	8	10	9	11	11	10	9	9
<i>Lysine degradation</i>	54	13	11	14	16	12	14	12	10	10
<i>Metabolism of xenobiotics by cytochrome P450</i>	7	3	3	3	3	3	3	3	2	2
<i>Methane metabolism</i>	33	6	6	6	7	10	8	8	6	6
<i>Methylnaphthalene degradation</i>	17	7	7	8	10	8	8	8	8	8
<i>Novobiocin biosynthesis</i>	6	4	5	4	4	4	4	4	4	4
<i>Other glycan degradation</i>	9	4	3	4	3	2	3	2	2	2
<i>Penicillin/cephalosporin biosynthesis</i>	14	1	1	1	1	1	1	1	2	2
<i>Pentose/glucuronate interconversions</i>	56	3	3	4	6	15	11	7	5	5
<i>Primary bile acid biosynthesis</i>	19	4	3	4	4	4	4	4	3	3
<i>Purine metabolism</i>	104	49	47	45	45	49	45	45	44	44
<i>Retinol metabolism</i>	18	4	4	4	4	4	4	4	5	5
<i>Starch and sucrose metabolism</i>	71	21	15	20	23	23	23	21	19	19
<i>Steroid hormone metabolism</i>	15	4	0	4	4	4	4	2	2	2
<i>Streptomycin biosynthesis</i>	18	7	7	7	9	9	9	9	7	7
<i>Styrene degradation</i>	21	10	5	9	9	9	10	10	5	5
<i>Sulfur metabolism</i>	30	9	10	9	8	8	8	9	9	9
<i>Taurine and hypotaurine metabolism</i>	17	6	6	6	6	6	6	6	6	6
<i>Tetracycline biosynthesis</i>	3	1	0	1	1	1	1	1	1	1
<i>Toluene and xylene degradation</i>	23	7	4	7	6	8	8	8	3	3
<i>Trinitrotoluene degradation</i>	7	3	2	3	4	4	3	4	3	3
<i>Tropane, piperidine and pyridine alkaloid biosynthesis</i>	15	2	3	2	2	3	3	3	3	3
<i>Tryptophan metabolism</i>	67	19	14	20	17	19	19	18	17	18
<i>Vancomycin group antibiotics</i>	1	1	1	1	1	1	1	1	1	1



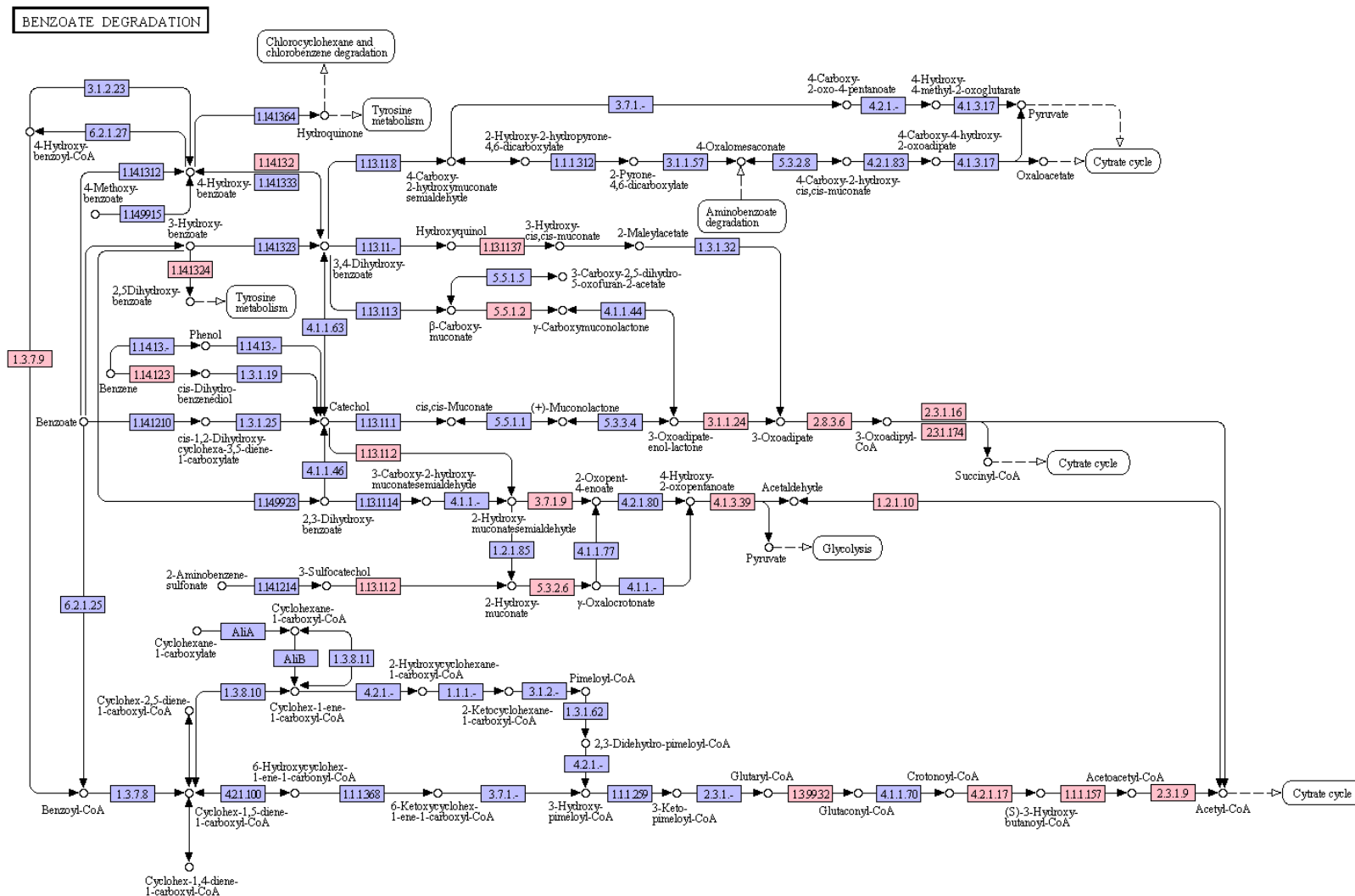


Figura 4.7: Via de degradação do benzoato. Enzimas identificadas pela anotação do genoma do isolado 226 destacadas em rosa e as não identificadas, em roxo.

#### 4.1.6. Antígenos compartilhados com BCG

Tem sido proposto que a exposição à micobactérias ambientais pode mascarar o efeito protetor da vacina BCG ou causar um efeito protetor cruzado, devido a antígenos compartilhados entre BCG e MNT; pode ainda causar um efeito antagonista, devido à presença de antígenos no BCG, mas não nas MNT (Palmer & Long, 1966; Brandt *et al.*, 2002; Checkley *et al.*, 2011). Alguns desses antígenos foram identificados em estudo anterior através de ensaios de *western blot* (Emmerick, 2013) e, diante dos dados dos genomas, buscou-se a identificação destes e outros antígenos descritos na literatura como importantes (tabela 4.8). Essa busca foi realizada com base nos dados da anotação realizada pelo Prokka a partir das sequências geradas pelo 454 GS Jr. Os dados foram comparados com as espécies de referência, mc<sup>2</sup>155, H37Rv, *M. leprae*, *M. bovis* BCG Moreau. As informações das espécies de referência foram obtidas diretamente da anotação disponibilizada pelo NCBI.

#### 4.1.7. Análise filogenética

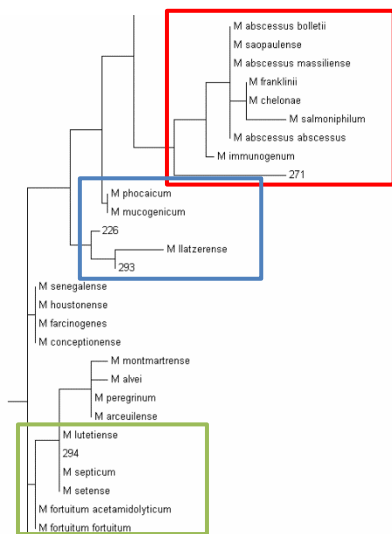
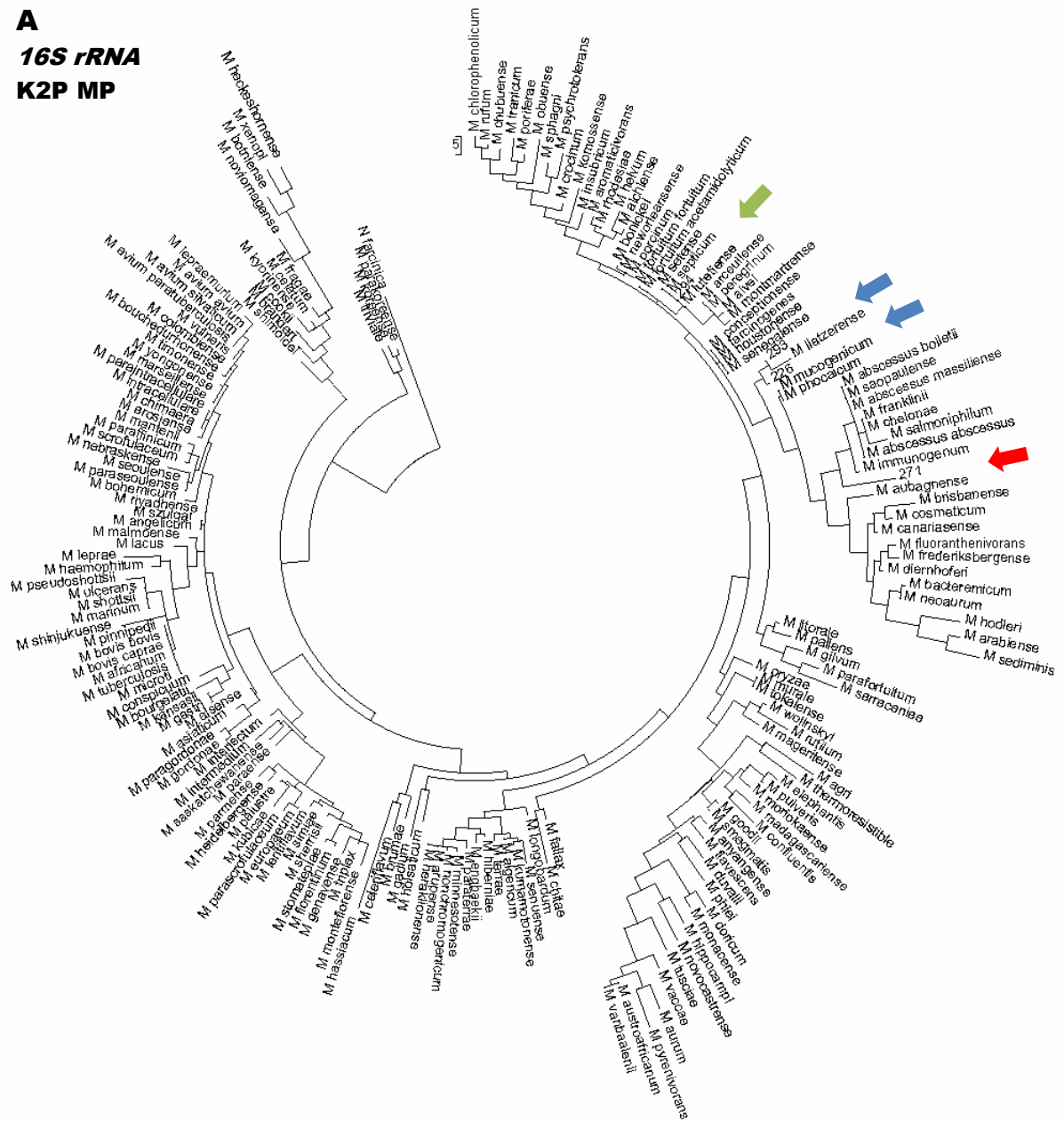
Foram escolhidos os genes *16S rRNA*, *hsp65* e *rpoB* para a análise filogenética molecular dos isolados em comparação com todas as espécies descritas até novembro de 2017 e *Nocardia farcinica* como grupo externo, totalizando 190 espécies diferentes. Apesar de existirem diversas sequências disponíveis de cada espécie, foi utilizada apenas uma sequência de cada espécie. Após o alinhamento das sequências no MAFFT, foi possível observar regiões hipervariáveis e regiões conservadas (dado não mostrado). Para a escolha do modelo estatístico, foi realizado um teste com os cinco modelos disponíveis no MEGA7 usando o gene *16S rRNA* como marcador filogenético e Kimura 2 parâmetros (K2P) como modelo de substituição de nucleotídeos (figura 4.8). Em todas as árvores avaliadas os isolados apresentaram uma distribuição filogenética muito semelhante: 271 é próximo de *M. immunogenum*, *M. chelonae* e *M. abscessus*; 294, de *M. septicum*, *M. setense* e *M. fortuitum*; e 226 e 293, de *M. mucogenicum*, *M. phocaicum* e *M. llatzerense*. Seguiu-se com as análises no jModelTest2 para escolha do modelo de substituição de nucleotídeos, sendo realizado quatro testes: critério de informação Bayesiana, Akaike e Akaike corrigido (BIC, AIC e AICc) e o método de teoria de decisão (DT)

(tabela 4.9). Com base nesses testes foi escolhido o modelo *General Time Reversible* (GTR) para os três marcadores utilizados.

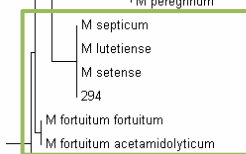
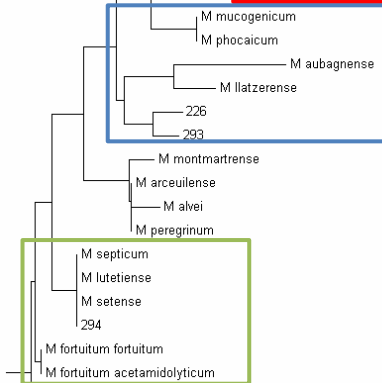
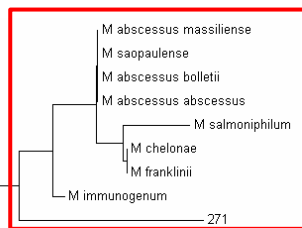
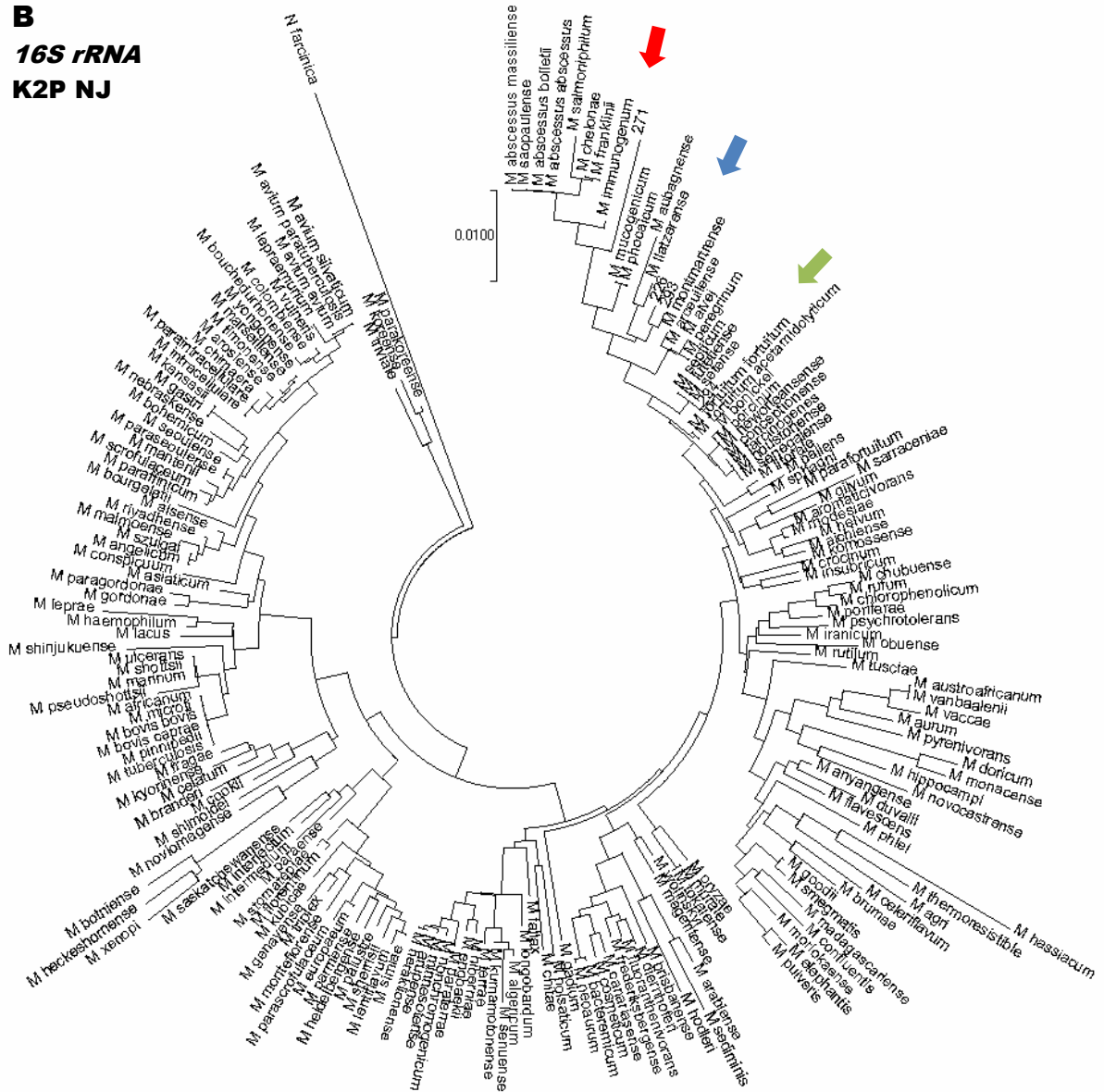
**Tabela 4.8: Antígenos compartilhados.**

Anotação	226	271	293	294	mc <sup>2</sup> 155	H37Rv	<i>M. leprae</i>	BCG
<i>MPT51/MPB51 antigen precursor (fbpD)</i>	2	-	2	1	-	1	-	1
<i>Soluble secreted antigen MPT53 precursor</i>	1	2	1	1	-	1	1	1
<i>Immunogenic protein MPT63 precursor</i>	3	1	3	4	3	1	-	1
<i>Immunogenic protein MPT64 precursor</i>	2	2	2	1	-	1	1	1
<i>Immunogenic protein MPT70 precursor</i>	1	-	1	1	-	1	-	1
<i>Mycolyltransferase Ag85A precursor (fbpA)</i>	-	2	-	1	-	1	1	1
<i>Mycolyltransferase Ag85B precursor (fbpB)</i>	2	-	2	2	-	1	1	1
<i>Mycolyltransferase Ag85C precursor (fbpC)</i>	1	1	2	4	-	1	1	1
<i>26 kDa periplasmic immunogenic protein precursor</i>	1	2	1	1	-	-	-	-
<i>Low molecular weight antigen MTB12 precursor</i>	1	1	1	1	-	1	1	1
<i>Low molecular weight protein antigen 6 (CFP-6)</i>	1	1	1	1	-	1	-	1
<i>Mycobacterium 19 kDa lipoprotein antigen (lpqH)</i>	4	6	5	3	-	1	2	1
<i>ESAT-6-like protein EsxH</i>	1	2	1	2	1	1	2	1
<i>6 kDa early secretory antigenic target (ESAT-6)</i>	-	-	2	1	-	1	-	-
<i>ESX-1 secretion-associated protein EspA</i>	1	-	2	3	9	1	-	-
<i>ESX-1 secretion-associated protein EspB</i>	-	-	-	1	.	1	-	-
<i>ESX-1 secretion-associated protein EspD</i>	-	-	2	4	.	1	-	-
<i>ESX-1 secretion-associated protein EspG1</i>	2	2	2	2	.	1	-	-
<i>ESX-1 secretion-associated protein EspK</i>	1	-	1	1	.	1	-	-
<i>ESX-1 secretion system protein EccA1</i>	4	1	2	2	2	1	-	-
<i>ESX-1 secretion system protein eccB1</i>	4	2	3	3	3	1	-	-
<i>ESX-1 secretion system protein EccCa1</i>	4	2	3	2	1	1	-	-
<i>ESX-1 secretion system protein EccCb1</i>	-	-	-	2	1	1	-	-
<i>ESX-1 secretion system protein eccD1</i>	2	1	2	1	3	1	-	-
<i>ESX-1 secretion system protein EccE1</i>	1	1	1	1	2	1	-	-

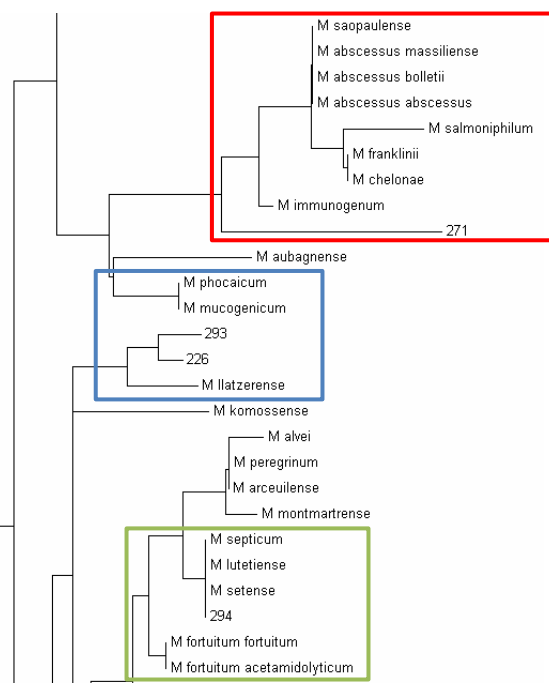
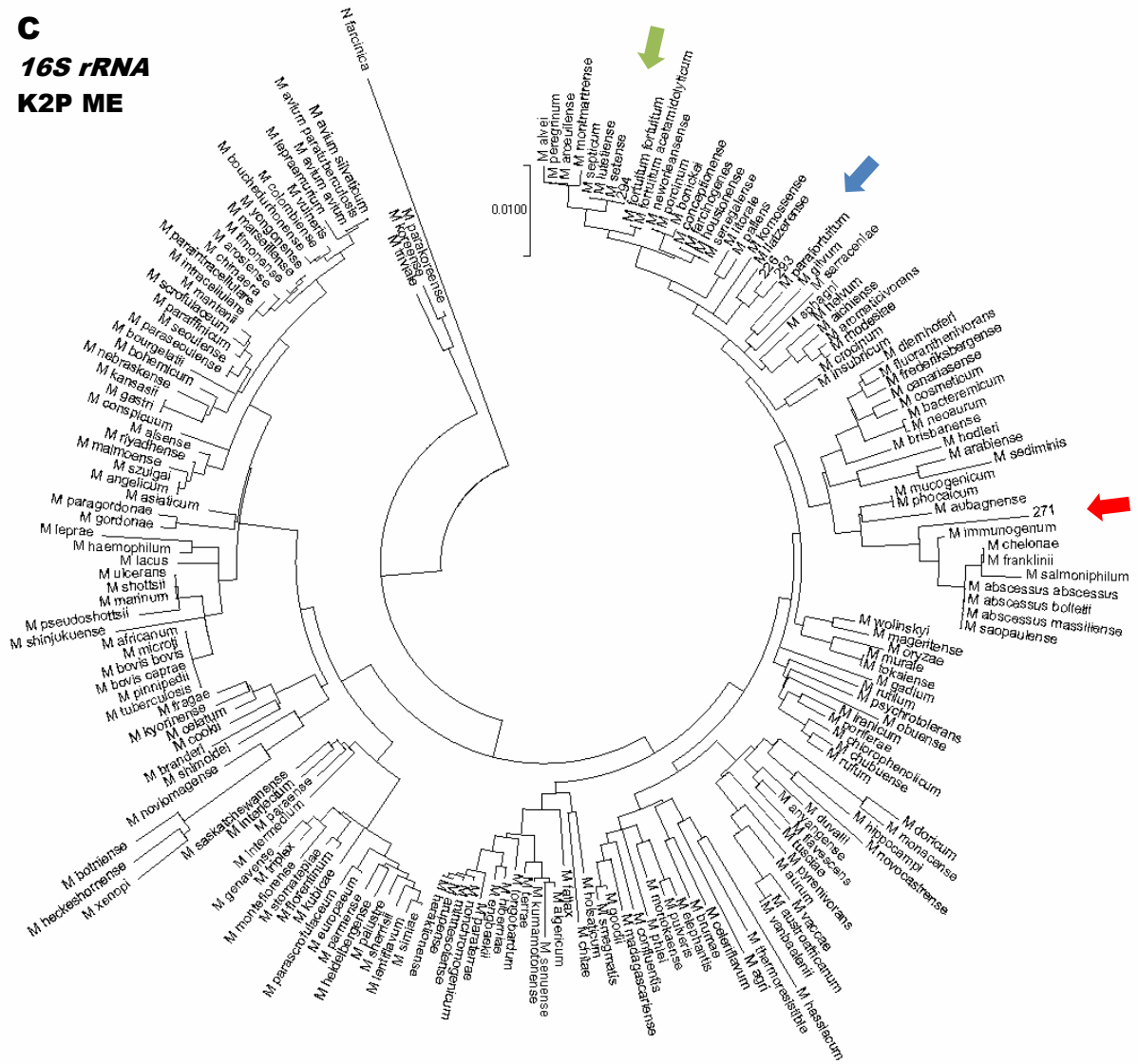
**A**  
**16S rRNA**  
**K2P MP**



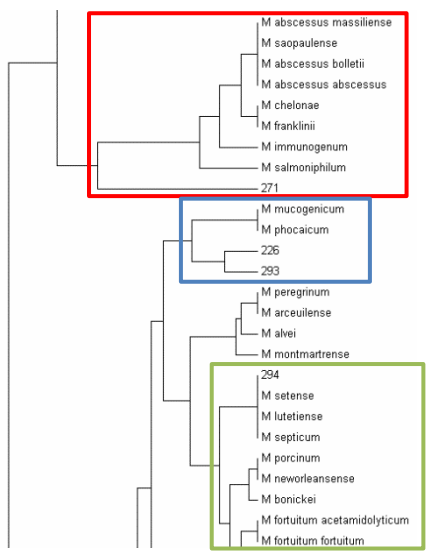
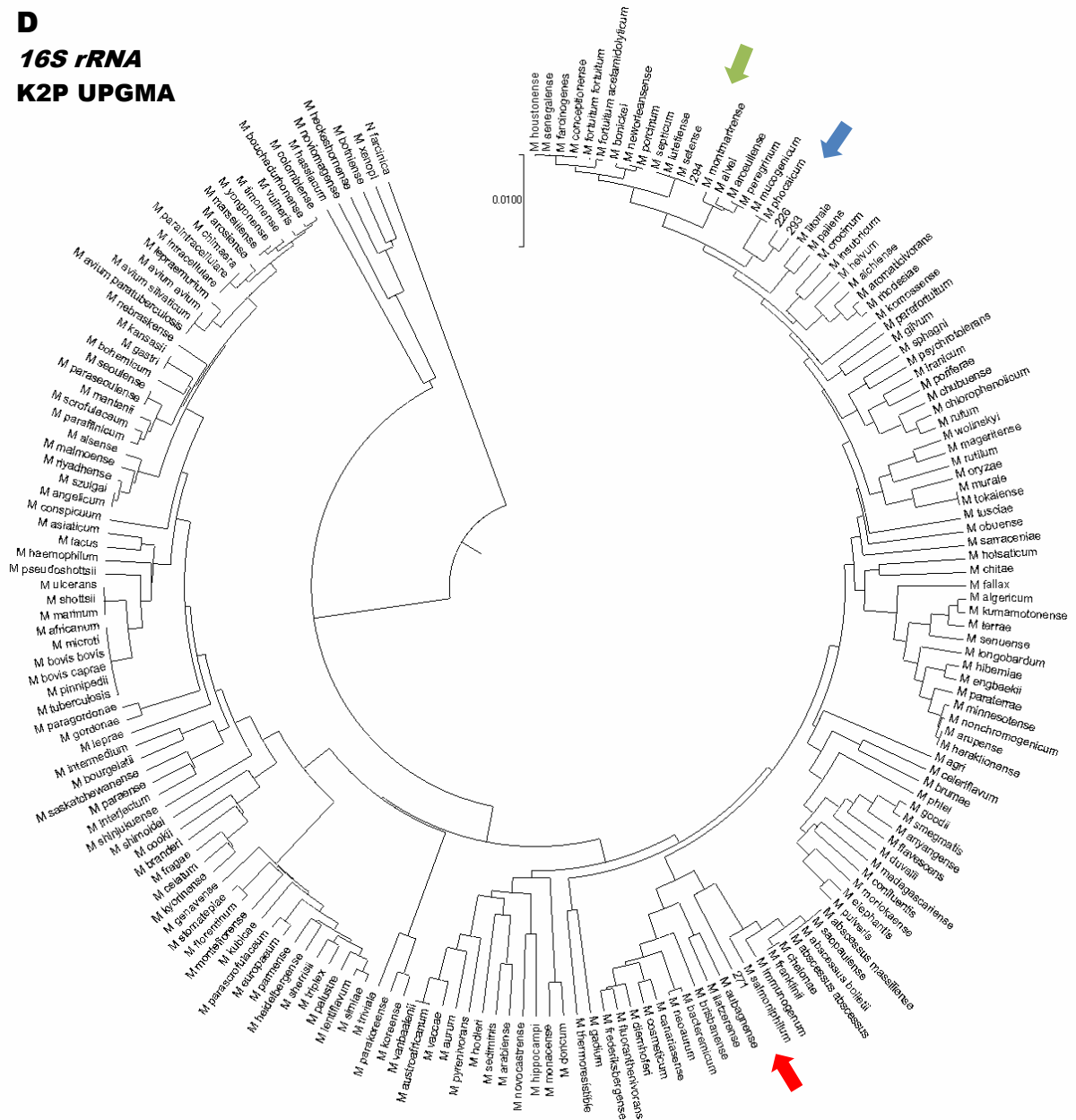
**B**  
**16S rRNA**  
**K2P NJ**



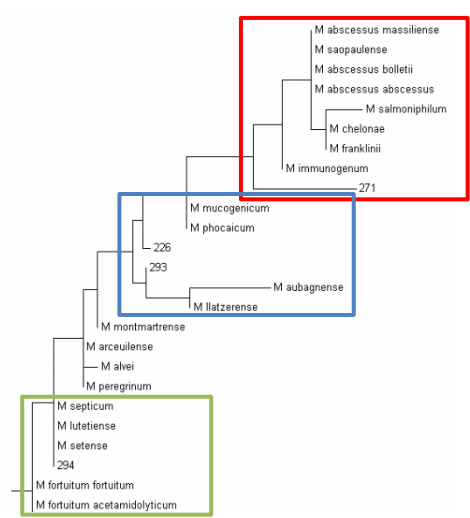
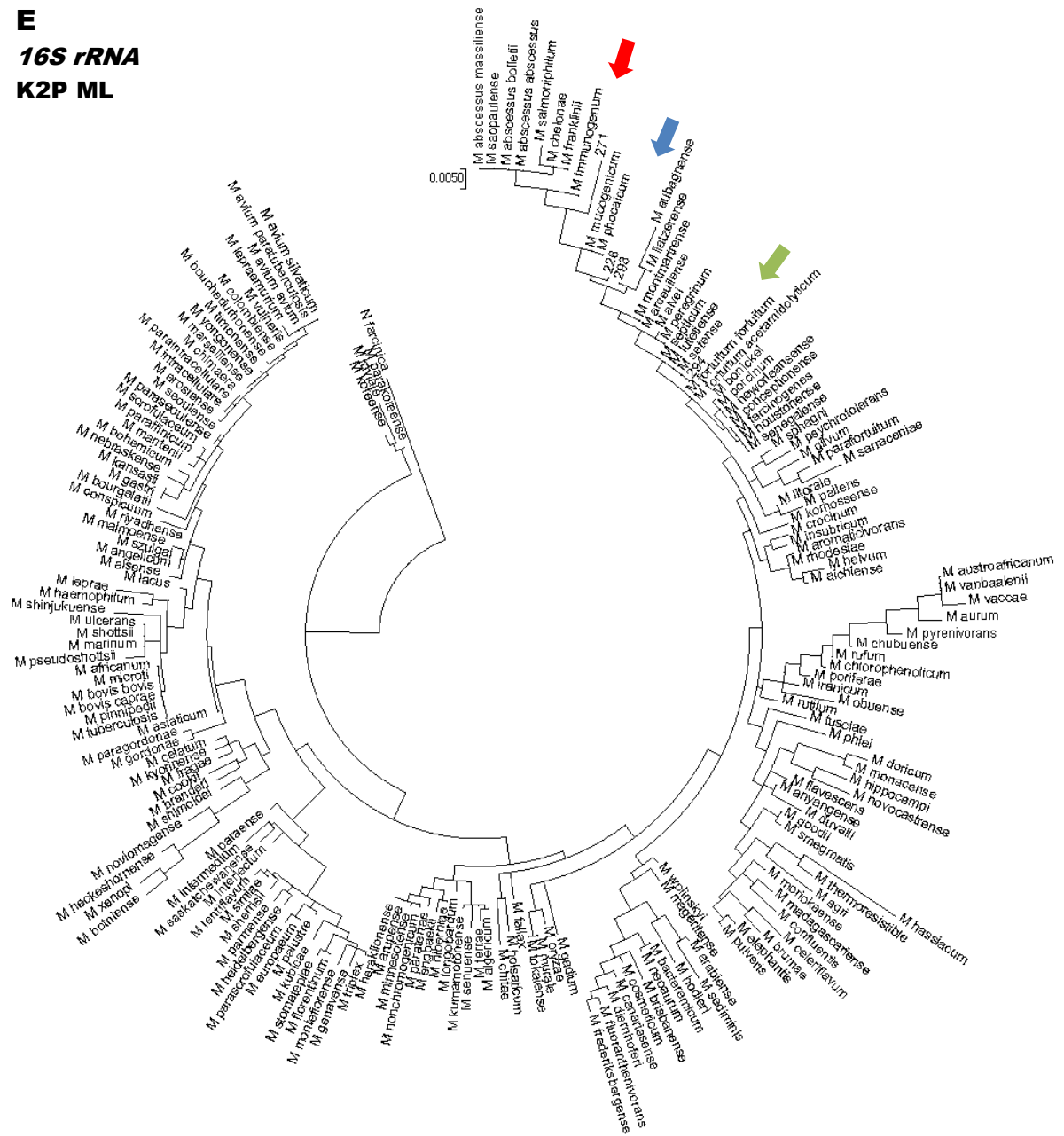
**C**  
**16S rRNA**  
**K2P ME**



**D**  
**16S rRNA**  
**K2P UPGMA**



**E**  
**16S rRNA**  
**K2P ML**





**Figura 4.8: Teste dos modelos estatísticos para geração das árvores.** *16S rRNA*, gene que codifica a subunidade 16S do RNA ribossomal; K2P, Kimura 2 parâmetros; (A) MP, Máxima Parcimônia; (B) NJ, *Neighbor-Joining*; (C) ME, *Minimum Evolution* (Evolução Mínima); (D) UPGMA, *Unweighted Pair Group Method using Arithmetic averages* (Método de agrupamento em pares ponderados usando médias aritméticas); (E) ML, *Maximum Likelihood* (Máxima Verossimilhança). O Isolado 271 está inserido no quadrado vermelho, 226 e 293, no quadrado azul e 294, no quadrado verde.

**Tabela 4.9: Análise do jModelTest2.** Foram realizados os testes AICc, AIC, BIC e DT para seleção do melhor modelo de substituição de nucleotídeos. Os valores de delta representam o resultado do modelo em questão menos o resultado do modelo com o menor valor.

Gene	AICc	Delta	AIC	Delta	BIC	Delta	DT	Delta
<b>16S <i>rRNA</i></b>	TrN+I+G	0,00	TrN+I+G	0,00	TrN+I+G	0,00	TrN+I+G	0,0000
	TPM2uf+I+G	5,01	TPM2uf+I+G	5,01	TPM2uf+I+G	5,01	TPM1uf	0,2106
	TVM+I+G	16,15	TVM+I+G	12,84	HKY+I+G	22,18	TPM2+I	0,2192
	TIM2+I+G	24,14	TIM2+I+G	22,49	TVM+I+G	23,45	TPM2uf+I	0,2195
	HKY+I+G	25,83	HKY+I+G	27,47	TIM2+I+G	27,79	TIM2	0,2226
	TIM3+I+G	31,43	TIM3+I+G	29,78	TIM3+I+G	35,08	TrNef	0,2234
	TPM3uf+I+G	38,41	TPM3uf+I+G	38,41	TPM3uf+I+G	38,41	TIM3+I	0,2261
	TIM1+I+G	55,04	TIM1+I+G	53,39	TIM1+I+G	58,69	JC	0,2268
	TPM1uf+I+G	63,40	TPM1uf+I+G	63,40	TPM1uf+I+G	63,40	<b>GTR</b>	0,2280
	<b>GTR+I+G</b>	90,84	<b>GTR+I+G</b>	85,87	<b>GTR+I+G</b>	101,78	TPM2uf+I+G	0,2821
<b><i>hsp65</i></b>	JC+G	0,00	<b>GTR+I+G</b>	0,00	<b>GTR+I+G</b>	0,00	<b>GTR+I+G</b>	0,0000
	JC+I+G	448,70	TIM2+I+G	29,89	TIM2+I+G	21,90	TVMef	0,3816
	K80+G	486,11	TPM1uf+I+G	47,21	TPM1uf+I+G	35,23	TrN+I	0,3864
	TPM1+G	890,78	TrN+I+G	61,21	TrN+I+G	49,23	TVM	0,3879
	JC+I	952,61	TPM2uf+I+G	69,90	TPM2uf+I+G	57,92	F81	0,3969
	TPM2+G	959,73	TVM+I+G	92,87	TIM3+I+G	86,35	TPM1	0,4002
	K80+I+G	1002,84	TIM3+I+G	94,34	TVM+I+G	88,88	TrNef	0,4017
	TrNef+G	1112,44	TPM3uf+I+G	104,46	TPM3uf+I+G	92,48	TIM2ef	0,4041
	TIM2+I+G	5206,89	HKY+I+G	118,89	HKY+I+G	102,91	TIM2	0,4043
	<b>GTR+I+G</b>	8037,29	TIM2+G	135,88	TIM2+G	123,89	TIM2+I+G	0,4771
<b><i>rpoB</i></b>	<b>GTR+I+G</b>	0,00	<b>GTR+I+G</b>	0,00	<b>GTR+I+G</b>	0,00	<b>GTR+I+G</b>	0,0000
	TIM1+I+G	56,09	TIM1+I+G	56,84	TIM1+I+G	44,44	TPM1+I	0,4122
	TVM+I+G	115,57	TVM+I+G	115,94	TVM+I+G	109,74	TPM2	0,4360
	TPM1uf+I+G	223,43	TPM1uf+I+G	224,54	TPM1uf+I+G	205,95	TPM2+I	0,4500
	TIM2+I+G	433,63	TIM2+I+G	434,37	TIM2+I+G	421,98	TVMef+I	0,4526
	TIM3+I+G	491,03	TIM3+I+G	491,78	TIM3+I+G	479,38	JC+I	0,4534
	TrN+I+G	507,53	TrN+I+G	508,64	TrN+I+G	490,05	K80	0,4609
	TPM3uf+I+G	565,72	TPM3uf+I+G	566,84	TPM3uf+I+G	548,24	SYM	0,4609
	TPM2uf+I+G	582,70	TPM2uf+I+G	583,82	TPM2uf+I+G	565,22	TIM2+I	0,4643
	HKY+I+G	592,82	HKY+I+G	594,30	HKY+I+G	569,51	TIM1+I+G	0,6023

Assim, foram escolhidos os modelos GTR e ML e valores de *bootstrap* para 1.000 replicatas para a construção final das árvores (figura 4.9). O padrão geral observado na análise anterior se repetiu: os isolados formando três grupos distintos. Através da análise da árvore gerada a partir do gene *16S rRNA*, com aproximadamente

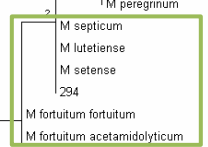
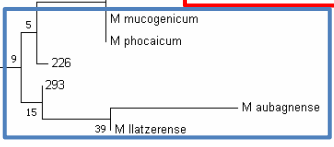
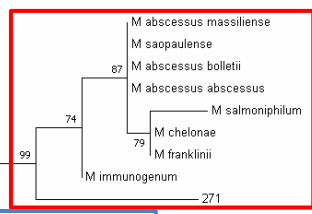
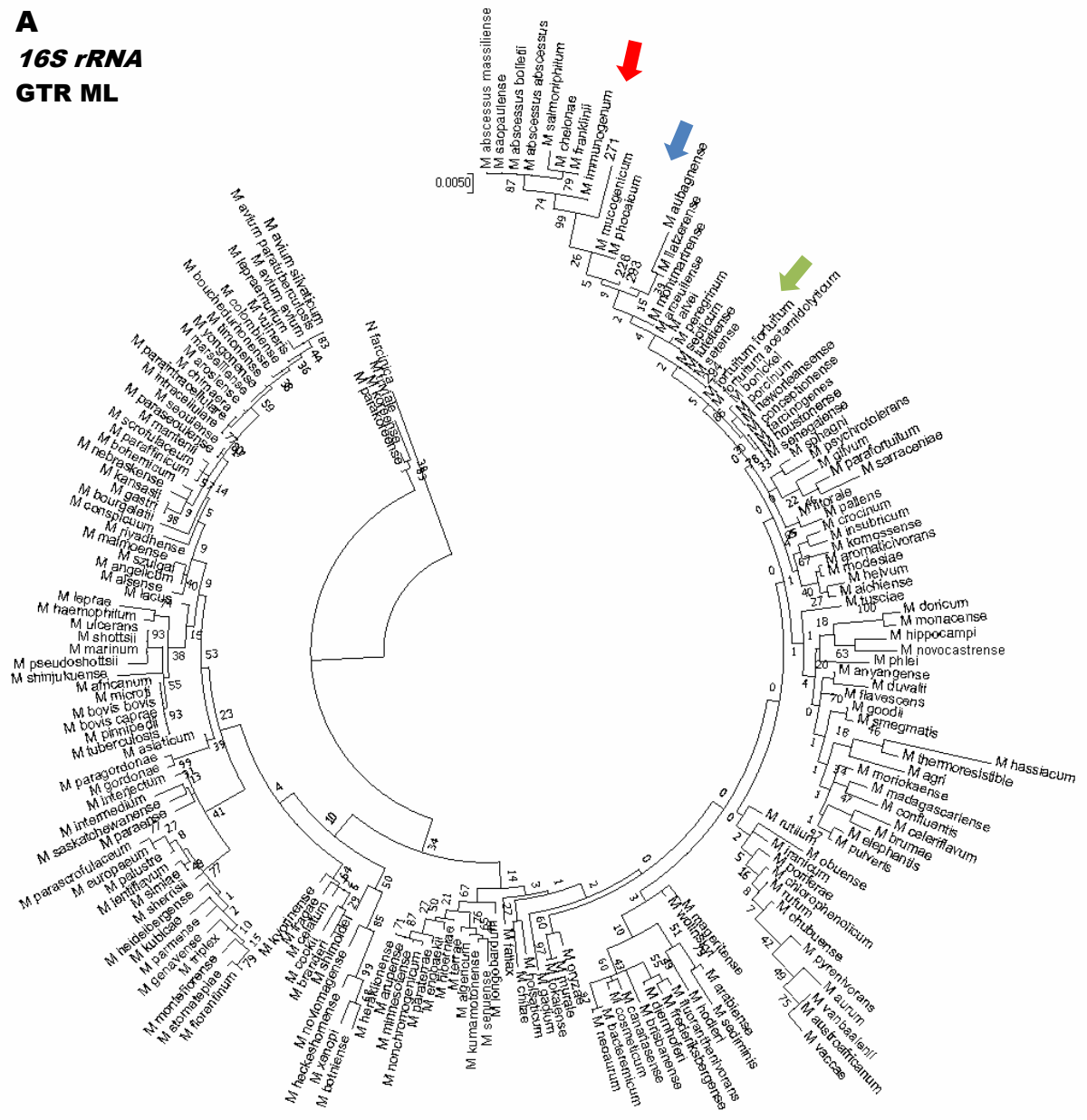
1.300 nucleotídeos, observou-se que esse marcador não é capaz de diferenciar diversas espécies e subespécies, como o complexo *M. tuberculosis*, composto por seis espécies e subespécies; o complexo *M. avium*, composto por três subespécies; o grupo *M. conceptionense*, *M. farcinogenes*, *M. houstonense* e *M. senegalense*; e outros. Assim, embora 294 apresente a mesma sequência deste gene que *M. septicum*, *M. lutetiense* e *M. setense*, não é possível afirmar que este isolado pertença a alguma destas espécies ou seja uma subespécie. Por outro lado, todas as subespécies incluídas na análise mostraram-se no mesmo ramo filogenético. Então, isso sugere que os isolados 226, 293 e 271 sejam provavelmente espécies distintas das descritas até o momento e distintas entre si, apesar da proximidade entre 226 e 293.

A árvore construída a partir da sequência parcial do gene *hsp65*, com 401 nucleotídeos, foi capaz de distinguir entre as subespécies de *M. abscessus* e entre *M. pinnipedii* e o restante do complexo *M. tuberculosis*. No entanto, não mostrou diferenças entre as subespécies de *M. bovis*, *M. fortuitum* e *M. avium*. As espécies *M. septicum*, *M. lutetiense* e *M. setense*, que agruparam junto com 294 na análise anterior, mostraram nítidas diferenças nesta árvore, indicando que o isolado 294 também represente provavelmente uma nova espécie. *M. hippocampi* não foi utilizada nesta análise por não apresentar sequência disponível no GenBank.

Com 3.500 nucleotídeos em média, o gene *rpoB* mostrou-se um ótimo marcador filogenético ao avaliarmos os valores de *bootstrap*. Foi capaz de distinguir as subespécies *M. avium* e *M. abscessus*, mas não entre os membros do complexo *M. tuberculosis* e as subespécies de *M. fortuitum*. Os isolados foram novamente distribuídos em ramos filogenéticos diferentes, reforçando a hipótese de serem espécies distintas. 45 espécies foram excluídas dessa análise por não apresentarem a sequência completa disponível no GenBank.

Além das árvores para cada gene, também foi construída uma árvore com os três marcadores concatenados. Para isso, foram utilizadas 140 espécies de micobactérias que apresentavam as sequências dos três marcadores no GenBank, além do grupo externo e os quatro isolados. A árvore se mostrou muito similar às demais, apresentando ótimos valores de *bootstrap* e sendo capaz de distinguir as subespécies de *M. abscessus*, *M. avium* e, ainda que sutilmente, *M. fortuitum*, mas não o complexo *M. tuberculosis*. Ao contrário das outras árvores, o grupo externo apresentou-se completamente separado das demais espécies estudadas.

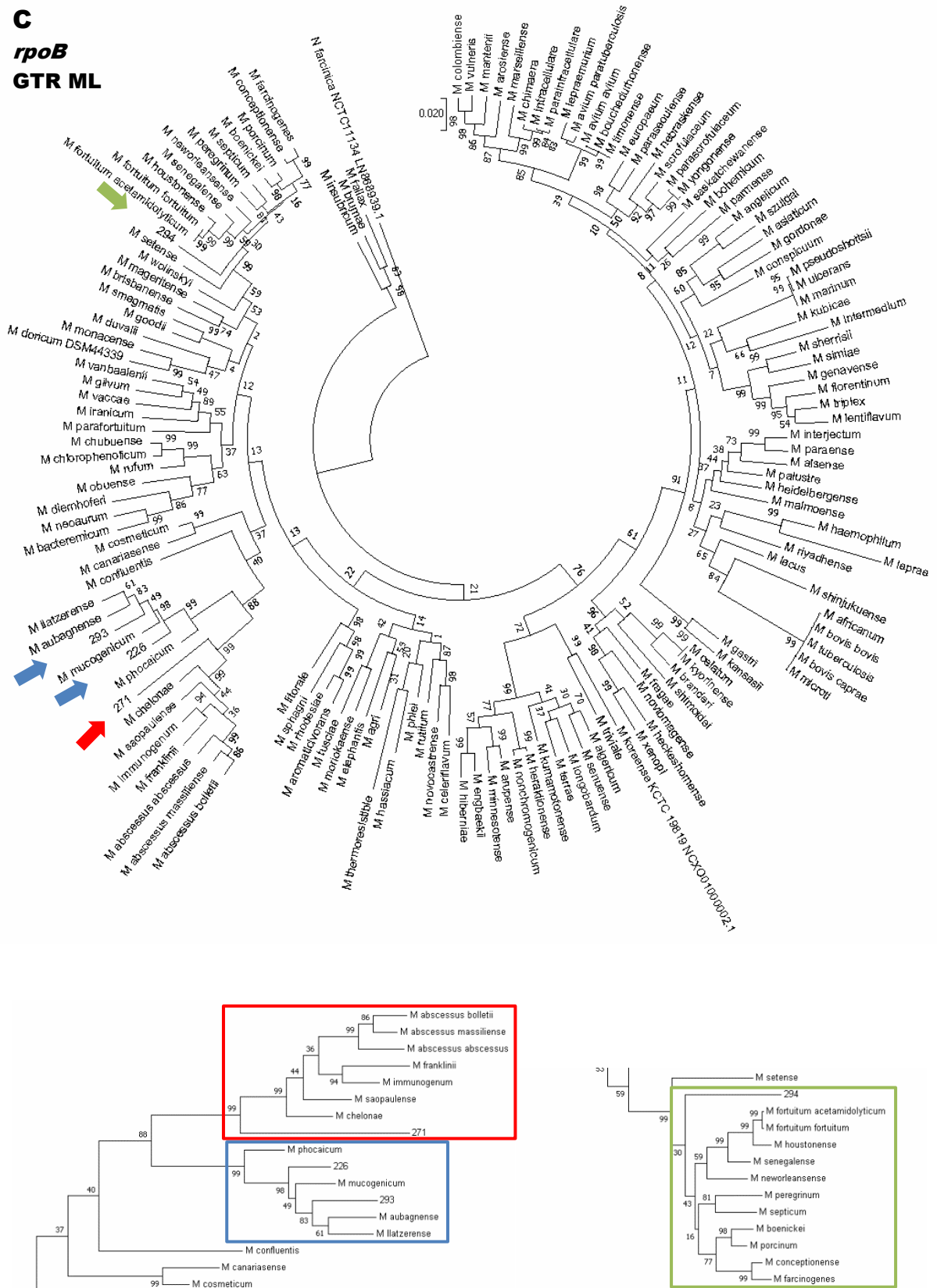
**A**  
**16S rRNA**  
**GTR ML**



**B**  
**hsp65**  
**GTR ML**



**C**  
**rpoB**  
**GTR ML**



**D**  
Concatenada  
GTR ML



**Figura 4.9: Árvores filogenéticas.** Genes (A) *16S rRNA*, (B) *hsp65* (codificante para a proteína de choque térmico de 65 kDa), (C) *rpoB* (codificante para a subunidade  $\beta$  da RNA polimerase) e (D) os três genes concatenados (Concat). Foi utilizado como modelo de substituição de nucleotídeos o *General Time Reversible* (GTR) e Máxima Verossimilhança (ML) como modelo estatístico. Os valores de *bootstrap* para 1.000 replicatas são mostrados nos nós. *N. farcinica* foi usado como grupo externo e a escala representa o número de substituições por sítio. O isolado 271 está inserido no quadrado vermelho, 226 e 293, no quadrado azul e 294, no quadrado verde.

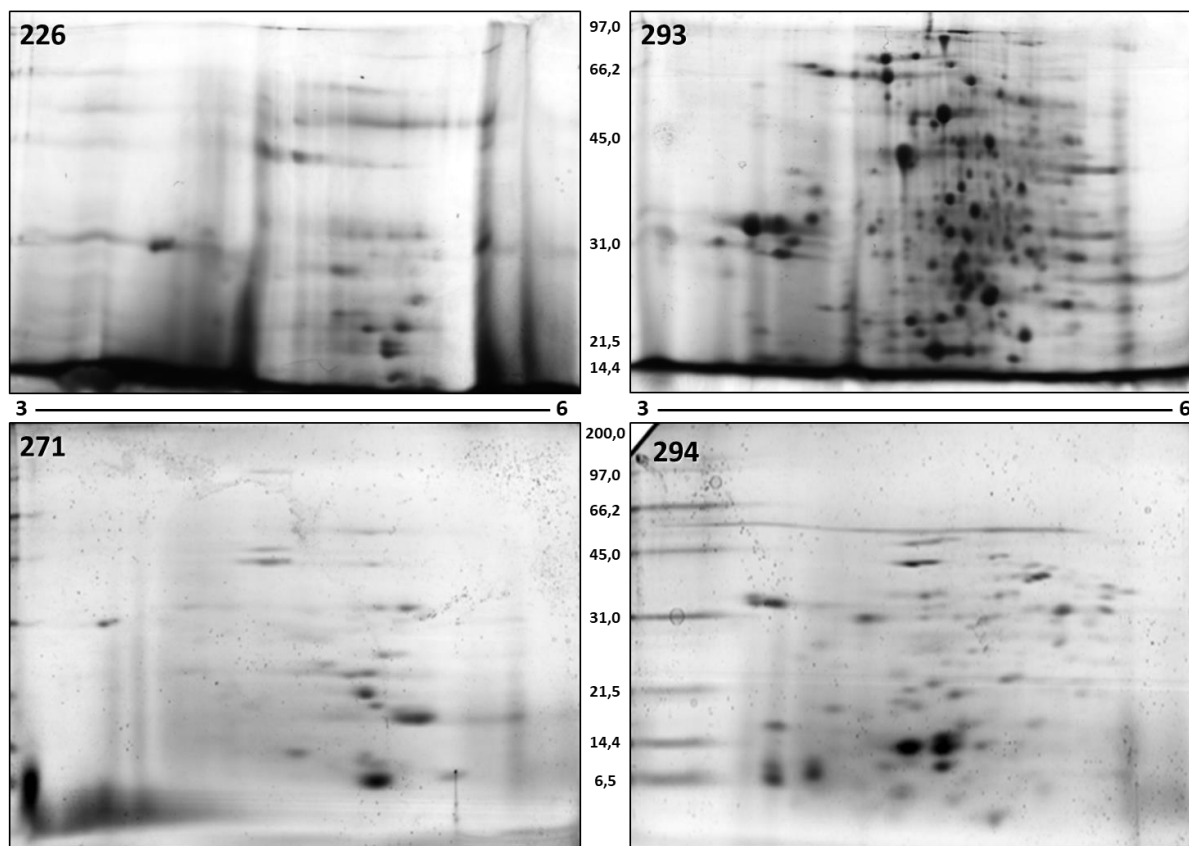
## 4.2. Análise do secretoma

### 4.2.1. Perfil de secreção de proteínas

Foi avaliado previamente o perfil de secreção de proteínas ao longo do crescimento para cada um dos isolados cultivados em meio LB/Tw, através de SDS-PAGE (Emmerick, 2013). Visando obter uma melhor resolução do perfil de proteínas secretadas pelos quatro isolados, utilizamos a técnica de 2DE, com amostras obtidas de culturas em final de fase *log* de crescimento (figura 4.10). Foi avaliado o perfil bidimensional em diferentes faixas de pH (resultado não mostrado), e observou-se proteínas majoritariamente na faixa de pH ácido (pH 3-6) e maior abundância de proteínas de baixo peso molecular. Uma análise comparativa entre os isolados mostrou perfis muito diferentes, com maior abundância de proteínas secretadas por 293. Entretanto, o isolado 294 apresentou um perfil menos complexo, com algumas proteínas secretadas em grande quantidade. Esta é uma das características desejáveis para o desenvolvimento de um novo sistema de expressão de proteínas heterólogas, visando a expressão de proteínas secretadas.

### 4.2.2. Identificação de proteínas

Com o intuito de identificar as proteínas secretadas em grandes quantidades, os *spots* mais abundantes foram excisados manualmente do gel (segunda dimensão), descorados com acetonitrila/bicarbonato de amônio e desidratados com acetonitrila. Posteriormente, os *spots* foram digeridos com tripsina e os peptídeos extraídos do gel (figura 4.11). As amostras foram concentradas, dessalinizadas e analisadas por espectrometria de massas em equipamento MALDI TOF-TOF (tabela 4.10) e Orbitrap (tabela 4.11). Para a análise dos dados gerados foi utilizado um banco de dados contendo a anotação genômica desses isolados, obtida a partir do sequenciamento genômico. Essas análises foram realizadas apenas para os isolados 271 e 294.



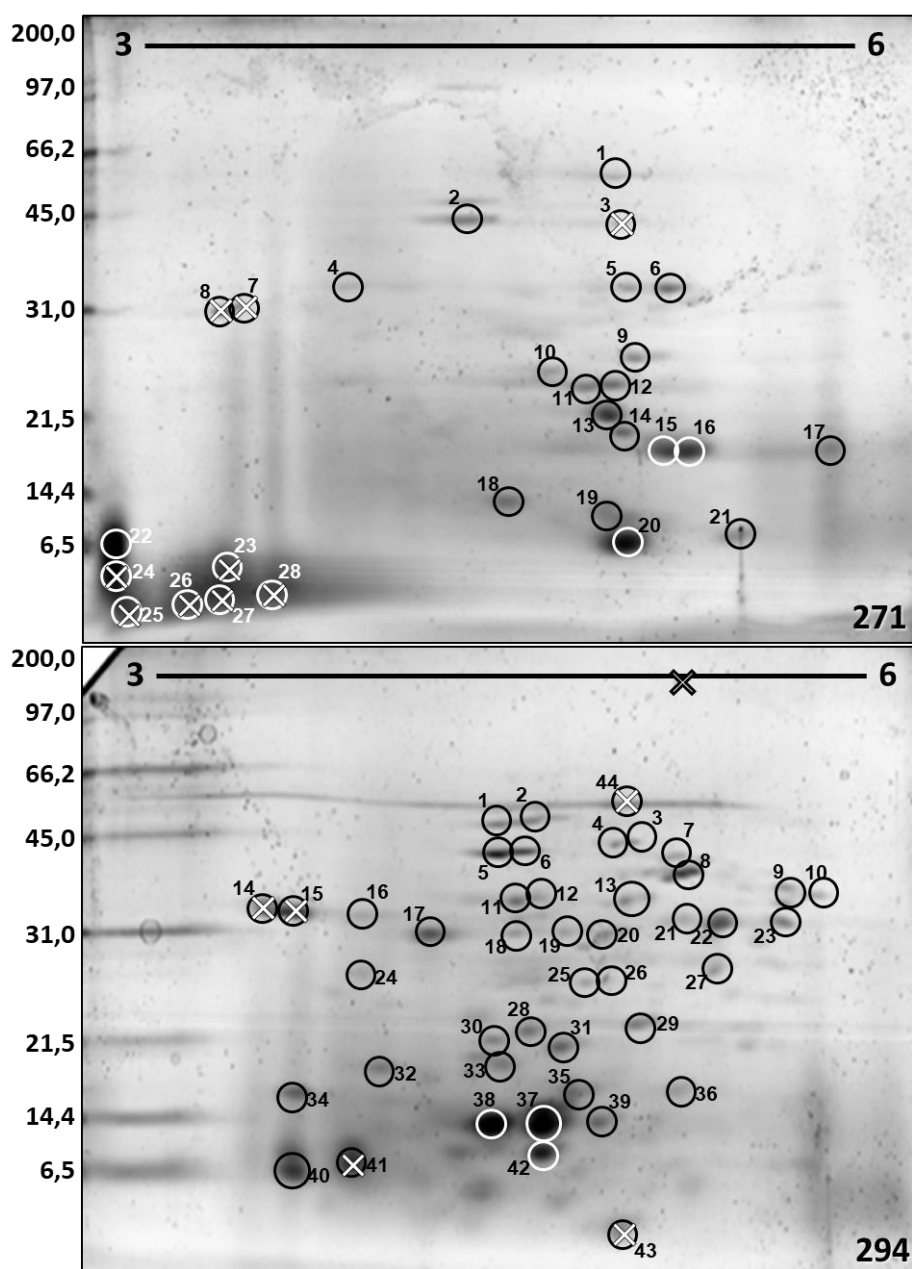
**Figura 4.10: Perfil bidimensional das proteínas secretadas.** As proteínas foram focalizadas em tiras de IPG de pH 3-6, com posterior separação em géis SDS-PAGE 14 %. As proteínas foram coradas com Coomassie coloidal (CBB-G250), as imagens digitalizadas no densitômetro GS-800. Foram aplicadas 200  $\mu$ g de proteínas para 271 e 294 e 500  $\mu$ g para 226 e 293. Marcador de peso molecular indicado ao centro, em kDa.

A análise através do MALDI resultou na identificação com baixa qualidade de poucos peptídeos e, conseqüentemente, de poucas proteínas. Esses resultados tiveram uma melhora após análise pelo Orbitrap. Do total de 72 *spots* excisados, 58 foram identificados, com uma média de 4 a 5 possíveis proteínas por *spot*. Dessas identificações obteve-se espectros relativos a poucos peptídeos (aproximadamente dois) tendo cobertura de cerca de 10 % da proteína respectiva.

Para uma análise mais abrangente, o secretoma dos quatro isolados foi avaliado através da abordagem *shotgun* em equipamento QExactive, em colaboração com o Dr. Richard Valente. Esta análise resultou em um maior número de proteínas identificadas (mais de 1.000 proteínas ao todo) e com uma qualidade muito superior quando comparado ao resultado obtido da análise dos *spots* pelo Orbitrap (tabela 4.12, figura 4.12). Esta análise é compatível com o perfil observado no gel unidimensional, visto que o isolado 226 apresentou o dobro da quantidade de



proteínas identificadas quando comparado aos demais. O isolado 294 continua apresentando um perfil menos complexo e com ótima qualidade. Considerando apenas as proteínas que receberam anotação diferente, foram encontradas ao todo 460. Destas, apenas sete proteínas mostraram-se presentes nos quatro isolados, mostrando a diversidade entre eles. No entanto, 40 proteínas mostram-se presentes exclusivamente nos isolados 226 e 293, reforçando uma provável maior proximidade filogenética entre eles (tabela 4.13 e figura 4.13).



**Figura 4.11: Spots analisados pelo Orbitrap.** Foram excisados manualmente do gel 28 spots do isolado 226 e 52 spots de CBMA294, tratados e analisados em duplicata por espectrometria de massas em equipamento Orbitrap. Os "X" representam os spots que não obtiveram resultado.

**Tabela 4.10: Identificações através do MALDI TOF-TOF**

Proteína identificada	Código	Massa (kDa)	Spots 271 (peptídeos únicos)			Spots 294 (peptídeos únicos)
Ag85A	271_01301	35	4 (1)	5 (1)	6 (1)	
Cutinase	271_03853	23	10 (3)			
Isomerase tipo FKBP 22 kDa	271_02036	20	10 (1)			
Lipoproteína LpqN	271_01499	32	2 (1)			
Proteína hipotética	271_02826	12	15 (2)	16 (1)		
Proteína hipotética	271_04598	23	11 (1)	12 (3)		
Proteína hipotética	271_04600	23	10 (2)			
Proteína hipotética	271_04151	20	11 (1)			
Proteína hipotética	271_02761	18	15 (1)			
Proteína hipotética	271_02066	12	20 (1)			
Proteína hipotética	271_02261	18	17 (1)			
Proteína ligadora de manose	271_00547	29	5 (1)			
Cutinase	294_04582	22				28 (2)
ESAT-6	294_03850	10				41 (1)
MTB12	294_06034	16				42 (1)
Proteína hipotética	294_05676	12				43 (1)
Proteína secretada Apa	294_02949	33				6 (1)

**Tabela 4.11A: Identificações dos spots de 271 através do Orbitrap.**

Proteína identificada	Código	m (kDa)	Spot (nº de peptídeos únicos)						
Ag85A	271_01301	35	4 (4)	5 (5)	6 (5)	13 (2)	15 (3)		
Ag85A	271_01302	35	4 (2)	5 (2)	6 (4)				
Aminopeptidase S	271_02937	53	14 (2)						
Carboxilesterase A	271_01639	54	1 (2)						
Carboxypeptidase DacB	271_04412	30	4 (1)	5 (1)					
Cutinase	271_03853	23	10 (3)	11 (1)	12 (1)				
Cutinase	271_03851	19	14 (1)						
Fator A de ligação a ribossomo	271_04531	17	13 (1)						
Fator de alongação Ts	271_04459	29	5 (1)						
Isomerase tipo FKBP 22	271_02036	20	10 (1)						
L,D-transpeptidase LppS	271_04490	26	13 (1)	14 (2)					
Lipase 1	271_04282	29	6 (2)						
Lipoproteína LpqN	271_01499	32	2 (3)	6 (1)					
MPT53	271_04402	19	14 (1)	16 (1)					
MspA	271_02230	21	13 (1)						
Proteína hipotética	271_02826	12	14 (1)	15 (3)	16 (3)	17 (1)	18 (1)	22 (1)	
Proteína hipotética	271_02261	18	13 (1)	14 (2)	16 (1)	17 (1)			
Proteína hipotética	271_03816	18	13 (5)	14 (7)					
Proteína hipotética	271_04598	23	9 (1)	10 (2)	11 (2)	12 (2)	13 (3)	14 (2)	15 (1)
Proteína hipotética	271_04602	20	14 (3)	15 (1)	17 (3)				

Proteína hipotética	271_00670	20	13 (5)	14 (2)	15 (1)				
Proteína hipotética	271_04600	23	10 (4)	11 (3)	12 (2)	13 (3)			
Proteína hipotética	271_02761	18	14 (2)	15 (2)					
Proteína hipotética	271_03385	12	19 (1)	20 (2)					
Proteína hipotética	271_02564	14	18 (1)						
Proteína hipotética	271_00608	22	9 (3)	13 (1)					
Proteína hipotética	271_02066	12	20 (1)						
Proteína hipotética	271_02415	19	13 (2)	14 (1)					
Proteína hipotética	271_02648	19	13 (1)	17 (1)					
Proteína hipotética	271_00520	12	21 (1)						
Proteína hipotética	271_00615	24	14 (1)						
Proteína hipotética	271_00817	25	13 (1)						
Proteína hipotética	271_04571	24	9 (1)						
Proteína hipotética	271_00668	25	17 (1)						
Proteína hipotética	271_00847	19	13 (1)						
Proteína regulatória RecX	271_02734	22	14 (1)	15 (1)					
Proteína secretada Apa	271_02124	?	2 (1)						
Succinato-semialdeído desidrogenase [NADP(+)]	271_03982	55	1 (1)						

**Tabela 4.11B: Identificações dos spots de 294 através do Orbitrap.**

Proteína identificada	Código	m (kDa)	Spot (nº de peptídeos únicos)						
50S Proteína ribossomal L7/L12	294_02379	13	37 (4)						
ACP extensão de meromicolato	294_02988	11	40 (3)						
Ag85B	294_02578	36	9 (3)	12 (2)	22 (2)	23 (4)			
Ag85C	294_02694	44	3 (3)	4 (3)	6 (3)	7 (3)	8 (3)		
Ag85C	294_02921	37	20 (2)						
Beta-lactamase	294_01481	31	19 (3)	20 (5)					
Carboxipeptidase	294_03998	19	32 (2)						
Carboxipeptidase DacF	294_02214	31	20 (5)						
Carboxylesterase A	294_04954	55	2 (2)						
Chaperonina 10	294_02452	11	42 (2)						
Cutinase	294_04582	22	28 (4)	29 (2)	30 (4)	31 (4)	33 (3)		
Cutinase	294_05805	33	12 (3)	13 (7)					
Cutinase	294_01304	22	29 (3)						
Cutinase	294_04581	22	29 (3)						
Cutinase	294_02356	29	20 (2)						
Desidrogenase/reductase	294_03876	30	20 (2)						
EccB1 (ESX-1)	294_03844	51	9 (2)	13 (3)					
EfeO	294_00417	?	12 (2)	18(10)					
Endo-1,3-beta-glicosidase A1	294_00474	30	17 (2)	19 (4)	21 (3)	27 (3)			
Endoglucanase A	294_01686	33	19 (4)	20 (4)					
Endopeptidase RipA	294_05345	47	24 (3)	29 (2)					
Endoprotease DegQ	294_00233	?	11 (2)	12 (2)					
EsxB (tipo ESAT-6)	294_03849	?	32 (3)	37 (4)	38 (5)	39 (4)	42 (4)		

Fator de elongação Ts	294_02149	29	12 (3)	13 (2)					
Fator de elongação Tu	294_02357	33	3 (2)	8 (4)					
FlavoProteína	271_04275	28	19 (2)	20 (4)					
Frutose-bisfosfato aldolase	294_02868	32	16 (8)						
Isomerase tipo FKBP 22	294_05509	19	30 (2)	31 (4)	33 (2)	35 (2)			
L,D-transpeptidase LppS	294_03351	36	6 (2)	7 (3)	8 (8)	9 (2)	13 (8)		
L,D-transpeptidase LppS	294_02887	30	26 (2)	27 (3)	32 (2)	33 (2)			
Lipoprotein LpqN	294_04246	29	13 (3)	20 (2)	27 (3)	28 (2)			
Lipoprotein YbbD	294_04094	39	12 (6)						
MPT63	294_00424	?	1 (3)	2 (3)	3 (2)	4 (2)	5 (2)	6 (3)	8 (2)
MPT64	294_01976	?	25 (3)	26 (3)	29 (2)				
MPT70	294_00611	?	17 (6)	18 (2)	19 (2)				
MTB12	294_06034	16	37 (2)	38 (2)	42 (4)				
O-methyltransferase	294_00738	23	25 (5)						
O-methyltransferase	271_00266	17	24 (2)	25 (2)	26 (2)	27 (2)	31 (2)		
Oxidoreductase YvdP	294_02465	49	6 (3)						
Peptidyl-prolyl isomerase B	294_05011	30	18 (2)						
PPE68	294_03848	45	7 (2)	8 (4)	13 (3)	18 (2)			
Proteína com domínio MaoC	294_01842	30	13 (2)						
Proteína da superfamília BON	294_03328	21	17 (2)						
Proteína de formação de S-S	294_01697	26	26 (2)						
Proteína hipotética	294_01260	50	1 (5)	2 (9)	3 (10)	4 (12)	6 (5)	7 (13)	8 (13)
			9 (11)	10 (9)	13 (6)	20 (7)	22 (6)	23 (5)	25 (3)
Proteína hipotética	294_01077	17	31 (2)	32 (3)	33 (3)	35 (4)	36 (2)	37 (3)	38 (3)
Proteína hipotética	294_04986	12	24 (3)	32 (3)	35 (3)	37 (4)	38 (3)	39 (3)	
Proteína hipotética	294_01685	21	33 (3)						
Proteína hipotética	271_02071	17	26 (2)						
Proteína hipotética	294_05308	15	35 (2)						
Proteína hipotética	294_03030	19	33 (3)						
Proteína hipotética	294_03390	37	12 (2)						
Proteína hipotética	294_02794	19	29 (2)						
Proteína hipotética	294_04617	20	37 (2)						
Proteína hipotética	294_01139	17	36 (2)						
Proteína hipotética	294_03689	30	42 (3)						
Proteína hipotética	294_05019	15	39 (2)						
Proteína hipotética	294_02662	18	29 (3)						
Proteína hipotética	294_05920	38	27 (2)						
Proteína hipotética	294_02580	16	37 (2)						
Proteína hipotética	294_02734	27	28 (2)						
Proteína hipotética	294_01788	19	28 (2)						
Proteína hipotética	294_04374	23	26 (2)						
Proteína hipotética	294_04238	87	33 (2)						
Proteína induzida por calor	294_01477	26	19 (3)	20 (3)					
Proteína secretada Apa	294_02949	33	5 (4)	6 (5)					
Serine protease HtrA	294_04404	34	11 (3)	12 (3)					
Serine/threonine kinase PknE	294_01690	25	29 (3)						

Tabela 4.12: Resultado Q-Exactive.

Isolado	226	271	293	294
MS scan	26.151	24.246	25.513	23.897
MS/MS scan	24.049	30.145	25.435	30.570
Peptídeos	3.232	3.473	2.742	3.596
Sequência de peptídeos	1.130	623	629	671
Proteínas	463	175	233	204
Grupos de proteínas	369	149	185	174
Proteínas com mais de 2 peptídeos únicos	116	22	30	25
Peptide-spectrum matches	2.045	993	1.027	1.124
<i>De novo</i> only spectra	2.112	3.061	2.303	3.141

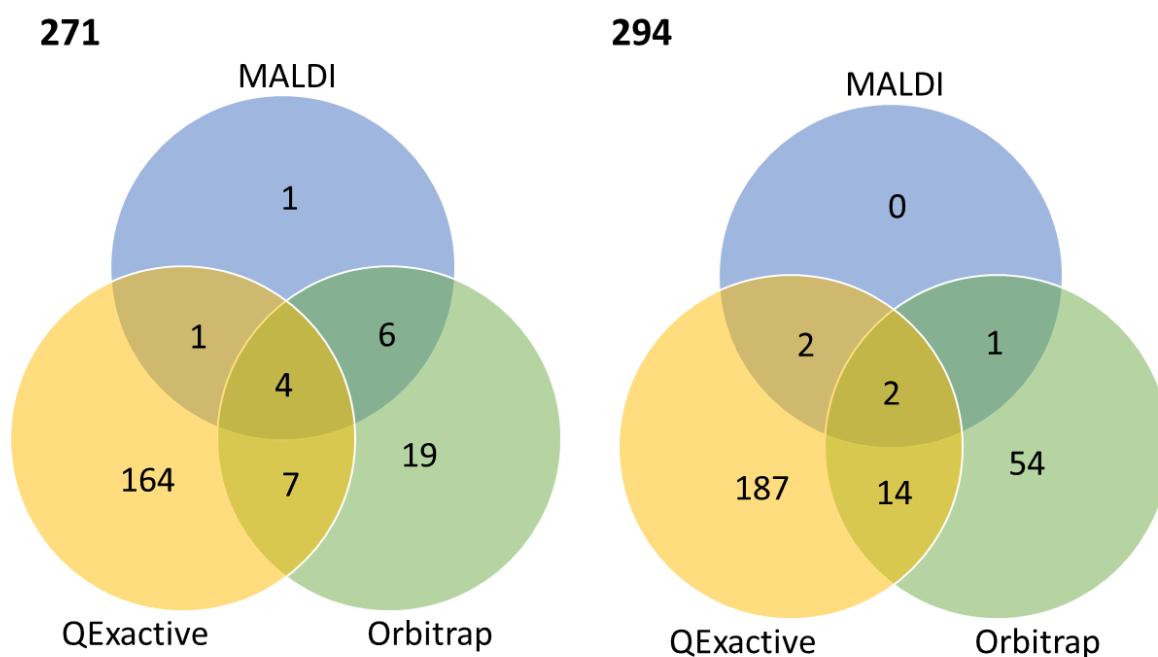


Figura 4.12: Identificação de proteínas por três espectrômetros de massas.

Tabela 4.13: Proteínas compartilhadas entre os isolados.

Isolado	Código	Proteína Identifica	m (kDa)		
226	293	226_03820	10 kDa chaperonin	10,6	
226	293	293_02609	10 kDa chaperonin	10,6	
226	293	226_04866	30S ribosomal protein S1	53,2	
226	293	293_04227	30S ribosomal protein S1	53,2	
226	293	226_01598	30S ribosomal protein S9	18,0	
226	293	293_02646	30S ribosomal protein S9	20,6	
271	293	226_01527	50S ribosomal protein L14	13,3	
271	293	271_03809	50S ribosomal protein L14	13,3	
271	293	294_02717	50S ribosomal protein L14	13,3	
226	293	226_01460	50S ribosomal protein L7/L12	13,2	
226	293	293_05751	50S ribosomal protein L7/L12	13,2	
271	293	294_05544	60 kDa chaperonin 1	56,2	
226	271	294_00232	60 kDa chaperonin 1	56,4	
271	294	294_03511	60 kDa chaperonin 1	56,5	
226	271	293	294_03821	60 kDa chaperonin 2	55,6
271	293	294_02608	60 kDa chaperonin 2	55,8	
271	294	271_03889	60 kDa chaperonin 2	56,0	
226	294	226_05546	Acetyl-coenzyme A synthetase	70,4	
293	294	293_04177	Acyl-CoA dehydrogenase	45,1	
271	293	294_03199	Adenosine tRNA methyltransferase MiaB	56,2	
271	293	293_03175	Adenosine tRNA methyltransferase MiaB	62,0	
226	293	271_01302	Ag85A precursor	35,2	
226	293	226_05049	Ag85B precursor	35,5	
226	293	226_01298	Ag85B precursor	35,6	
226	293	293_05360	Ag85B precursor	35,7	
226	293	294_02141	Ag85C precursor	40,0	
226	293	293_02787	Alanine and proline-rich secreted protein Apa precursor	34,6	
226	271	293	293_01921	Alpha-(1->3)-arabinofuranosyltransferase	149,2
271	294	294_02112	Amidohydrolase	47,3	
271	294	226_00717	Carboxylesterase NlhH	33,3	
226	293	226_02321	Chaperone protein DnaK	66,4	
226	293	294_03608	CobQ/CobB/MinD/ParA nucleotide binding domain protein	54,3	
226	293	226_03449	Cold shock protein	7,2	
226	293	293_05106	Cold shock protein	7,2	
271	294	294_04174	Copper resistance protein A precursor	50,6	
271	294	271_00760	Dimodular nonribosomal peptide synthase	141,1	
271	293	293_01033	Dimodular nonribosomal peptide synthase	184,3	
271	294	293_00770	DNA topoisomerase 1	102,9	
226	293	294_01574	DNA-directed RNA polymerase subunit alpha	37,9	
226	293	293_02672	DNA-directed RNA polymerase subunit alpha	37,9	
226	293	226_01462	DNA-directed RNA polymerase subunit beta	128,8	
226	294	226_00849	Doxorubicin resistance ATP-binding protein DrrA	33,3	
226	293	226_01481	Elongation factor G	76,7	
226	293	226_01593	EPTC-inducible aldehyde dehydrogenase	55,0	
226	293	226_02797	FK506-binding protein	12,8	
226	293	293_00157	FK506-binding protein	13,1	
226	271	294_04501	Flagellum site-determining protein YlxH	39,1	
226	271	293_06116	Flagellum site-determining protein YlxH	39,2	
271	294	293_00829	Formate dehydrogenase H	77,3	

271	294	226_05490	<i>Formate dehydrogenase H</i>	77,8
226	271	226_00411	<i>Glutamate--tRNA ligase</i>	53,2
226	293	226_05030	<i>Glutamine synthetase 1</i>	53,5
271	294	294_00097	<i>Glutamine-dependent NAD(+) synthetase</i>	79,1
	293	294_04155	<i>Glutathione transport system permease protein GsiD</i>	31,0
271	293	294_02100	<i>Glyceraldehyde-3-phosphate dehydrogenase</i>	35,7
226	293	226_05467	<i>Homocysteine methyltransferase</i>	81,0
271	294	271_03295	<i>Homocysteine methyltransferase</i>	81,8
226	293	226_00640	<i>Hypothetical protein</i>	9,1
226	293	226_00228	<i>Hypothetical protein</i>	20,4
226	293	293_03113	<i>Hypothetical protein</i>	21,5
226	293	226_03266	<i>Hypothetical protein</i>	21,6
226		294_226_05528	<i>Hypothetical protein</i>	45,5
	293	294_271_04668	<i>Hypothetical protein</i>	70,2
226	271	271_03563	<i>Hypothetical protein</i>	71,2
226	293	226_04071	<i>Isocitrate dehydrogenase [NADP]</i>	82,1
226		294_226_01959	<i>Linear gramicidin synthase subunit D</i>	447,4
	293	294_294_05210	<i>Linear gramicidin synthase subunit D</i>	865,7
	271	294_271_03374	<i>Linear gramicidin synthase subunit D</i>	868,1
	271	293_294_293_02471	<i>Long-chain-fatty-acid--CoA ligase</i>	101,8
226	293	226_02134	<i>Malate synthase G</i>	80,0
226	293	293_02769	<i>Malate synthase G</i>	81,0
226	293	271_04031	<i>Mannose-6-phosphate isomerase</i>	43,4
226	293	226_02208	<i>mce related protein</i>	45,9
	271	294_271_03386	<i>Membrane transport protein mmpL8</i>	103,1
226	271	293_294_226_01362	<i>Membrane transport protein mmpL8</i>	105,2
226	271	271_01079	<i>Membrane transport protein mmpL8</i>	110,7
226	271	293_226_01561	<i>Methylmalonate-semialdehyde dehydrogenase [acylating]</i>	54,3
	271	294_271_02131	<i>NADH dehydrogenase-like protein</i>	53,1
		293_294_226_02786	<i>NADPH-ferredoxin reductase FprA</i>	59,7
226	271	294_02264	<i>Peptidase family S58</i>	38,5
226	293	226_02548	<i>Phosphoenolpyruvate carboxykinase [GTP]</i>	67,4
		293_294_226_00929	<i>Phosphoenolpyruvate carboxylase</i>	102,6
		293_294_293_00948	<i>Phosphoglycerate mutase</i>	48,7
226		294_271_02788	<i>Polyribonucleotide nucleotidyltransferase</i>	79,3
	271	293_294_293_03128	<i>Polyribonucleotide nucleotidyltransferase</i>	80,1
226	271	294_226_03256	<i>Polyribonucleotide nucleotidyltransferase</i>	80,3
226		294_271_04051	<i>Protein translocase subunit SecA</i>	104,5
	271	294_271_02939	<i>Putative ABC transporter ATP-binding protein YxIF</i>	36,4
	271	293_294_04955	<i>Putative glutamine synthetase 2</i>	49,5
226	271	226_05213	<i>Putative HTH-type transcriptional regulator/MT0914</i>	98,7
226		294_293_03832	<i>Putative HTH-type transcriptional regulator/MT0914</i>	116,3
	271	294_226_00589	<i>Putative L D-transpeptidase LppS precursor</i>	26,4
226	293	293_04809	<i>Putative lipoprotein LpqN</i>	33,5
	271	294_293_05128	<i>Putative multidrug export ATP-binding/permease protein</i>	63,9
	271	294_226_01756	<i>Putative multidrug export ATP-binding/permease protein</i>	64,0
	271	294_293_01602	<i>Putative PPE family protein PPE42</i>	82,9
226	293	226_02644	<i>Putative short-chain type dehydrogenase/reductase</i>	29,8
		293_294_271_00833	<i>Putative TrmH family tRNA/rRNA methyltransferase</i>	32,9
226	271	294_226_01874	<i>Quercetin 2 3-dioxygenase</i>	47,7
	271	294_293_05714	<i>Quercetin 2 3-dioxygenase</i>	47,9

271	293	294	271_02444	<i>Quinone oxidoreductase 1</i>	46,2
271		294	271_03303	<i>Short-chain-fatty-acid--CoA ligase</i>	56,8
226		293	226_02916	<i>Succinyl-CoA ligase [ADP-forming] subunit beta</i>	40,9
	271		293_04169	<i>Succinyl-CoA:(R)-benzylsuccinate CoA-transferase</i>	42,8
226		293	226_01204	<i>Superoxide dismutase [Mn]</i>	22,9
226		293	294_05132	<i>Taurine--pyruvate aminotransferase</i>	43,7
226		293	293_03884	<i>Trigger factor</i>	53,9

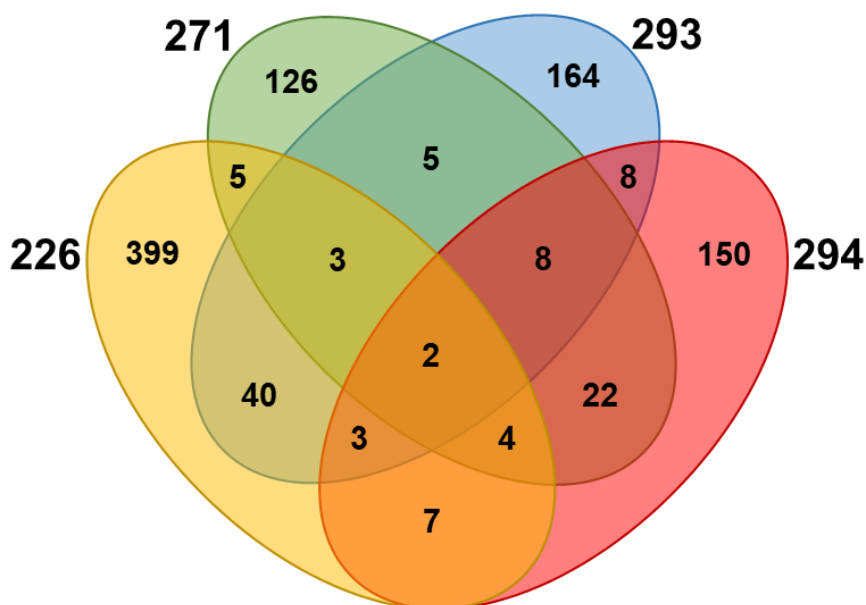


Figura 4.13: Proteínas compartilhadas entre os quatro isolados da Mata Atlântica.

No conjunto de proteínas identificadas pelo QExactive, 145 eram de proteínas hipotéticas. Ou seja, sequências nucleotídicas com características de sequência codificante, porém sem dados que comprovassem essa informação, sendo anotadas como hipotéticas. Ao serem encontradas no secretoma, essas proteínas passam a ser consideradas como “proteínas de função desconhecida”, visto que a proteína é de fato expressa, porém ainda não se conhece sua função. Após busca por similaridade em banco de dados de sequências proteicas, foi possível inferir função a quase 40 % dessas proteínas hipotéticas (tabela 4.14 e 4.15). Essas proteínas pertenciam a diversos grupos proteicos, como transferases, quinases, ATPases, proteínas de membrana e outros. O isolado 226 apresentou o maior número absoluto de proteínas hipotéticas e, também, o maior número de identificações após a busca por similaridades. Isto era esperado, visto o fato de ter apresentado o maior número de proteínas identificadas pelo QExactive.



Tabela 4.14: Proteínas hipotéticas identificadas pelo Q-Exactive

Isolado	226	271	293	294	Total
Proteínas hipotéticas	53	27	43	30	145
Com identificação	20	7	17	13	54
% identificação	38	26	39	43	37

Tabela 4.15: Algumas proteínas hipotéticas identificadas

Isolado	CÓD	m (kDa)	BLASTp	Ident (%)
226	226_04312	9,4	Glutathione peroxidase ( <i>M. goodii</i> )	93
	226_02661	11,2	N-acetyltransferase ( <i>Mycobacterium</i> )	94
	226_03266	21,7	Membrane protein ( <i>M. llatzerense</i> )	93
	226_01442	29,2	Transcription termination/antitermination protein NusG ( <i>M. mucogenicum</i> )	96
	226_03197	48,2	DNA repair ATPase ( <i>M. sp.</i> )	93
271	271_01427	10,1	Transposase ( <i>Mycobacterium</i> )	90
	271_03385	11,9	Low molecular weight T-cell antigen ( <i>M. abscessus</i> )	83
	271_01565	29,6	thioesterase family protein ( <i>M. boenickei</i> )	94
	271_05895	29,8	cyclodehydratase ( <i>Mycobacterium</i> )	80
	271_01476	71,0	ATP-dependent DNA helicase ( <i>Mycobacterium</i> )	93
293	293_00301	12,6	Potassium transporter ( <i>M. mucogenicum</i> )	94
	293_01376	27,5	Twin-arginine translocation pathway ( <i>M. llatzerense</i> )	96
	293_00302	35,9	Minor capsid protein E ( <i>M. mucogenicum</i> )	85
	293_01126	64,9	Glycosyl transferase ( <i>M. chelonae</i> group)	84
	293_01208	70,3	FAD-binding protein ( <i>M. salmoniphilum</i> )	91
294	294_01817	18,9	Putative lipoprotein LppK precursor ( <i>M. abscessus</i> )	80
	294_03560	27,7	Diacylglycerol kinase ( <i>M. abscessus</i> )	89
	294_04301	42,7	winged helix DNA-binding domain-containing protein ( <i>M. llatzerense</i> )	83
	294_03612	47,0	Sulfotransferase ( <i>M. abscessus</i> )	83
	294_06147	59,4	AAA family ATPase ( <i>Mycobacterium</i> )	73

Foi realizada uma quantificação *label-free* das proteínas secretadas mais abundantes de cada isolado baseada na contagem de espectros MS/MS de um dado peptídeo (tabela 4.16), visto que quanto mais abundante for a proteína na amostra, provavelmente será maior o número de espectros gerados. Embora seja uma quantificação relativa, essa metodologia permite identificar as proteínas majoritárias em cada amostra analisada separadamente pelo espectrômetro de massas. Essa informação será importante para a escolha de alvos cujos sinais regulatórios (de expressão e secreção) serão utilizados no desenho do sistema de expressão heteróloga. De acordo com essa quantificação, dentre as 40 proteínas mais secretadas (10 por isolado), observa-se que 35 delas são de baixo peso molecular (abaixo de 60 kDa), similar ao resultado observado no gel.

Tabela 4.16: Quantificação *label-free*.

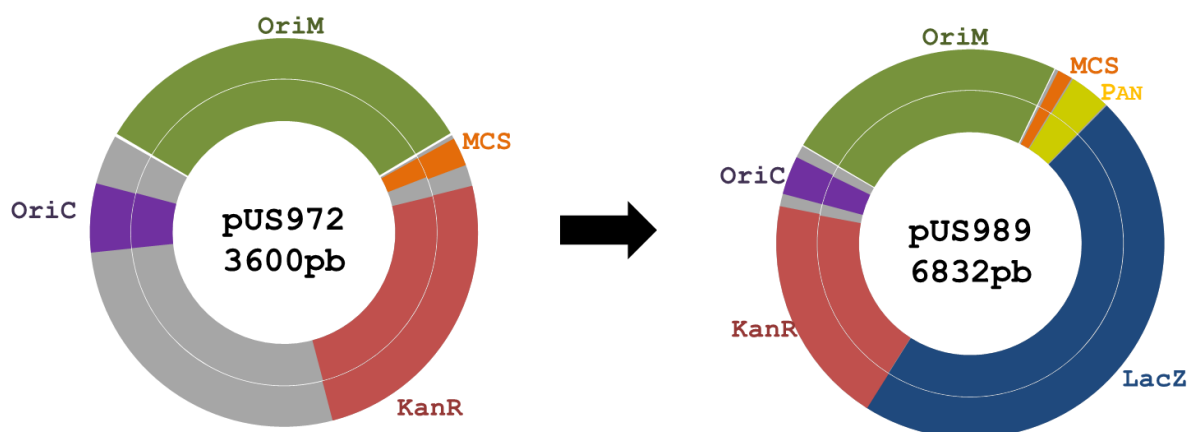
CBMA226				CBMA271			
Proteína	Código	m (kDa)	MS/MS	Proteína	Código	m (kDa)	MS/MS
DNA-directed RNA polymerase subunit alpha	226_01574	38,0	39	Hypotetical protein	271_00520	12,2	15
Catalase-peroxidase	226_00623	80,4	38	Hypotetical protein	271_02261	17,6	11
Elongation factor Tu	226_01482	43,6	32	Polyribonucleotide nucleotidyl-transferase	226_03256	80,2	10
Phosphoenolpyruvate carboxylkinase [GTP]	271_03179	67,1	27	Hypotetical Protein	271_02564	14,4	8
60 kDa chaperonin 1	226_00232	56,4	27	Low molecular weight T-cell antigen	271_03385	11,9	8
Superoxide dismutase	226_01204	23,0	26	PAP2 superfamily protein	226_02380	23,3	8
Methylmalonate-semialdehyde dehidrogenase	226_01561	54,3	25	Linear gramicidin synthase subunit D	271_03374	868,2	8
60 kDa chaperonin 2	226_03821	55,6	25	Hypotetical protein	271_04598	22,5	6
Glutamine synthetase 1	226_05030	53,5	24	N-acetylmuramoyl-l-alanine amidase I	271_00573	32,4	6
Chaperone protein DnaK	226_02321	66,5	23	Alanine and proline-rich secreted protein Apa precursor	271_02124	39,3	6
CBMA293				CBMA294			
Proteína	Código	m (kDa)	MS/MS	Proteína	Código	m (kDa)	MS/MS
5-hydroxyisourate hydrolase precursor	226_00297	11,6	24	6 kDa early secretory antigenic target	294_03850	9,7	21
DNA-directed RNA polymerase subunit alpha	294_02251	37,9	15	ESAT-6-like protein EsxB	294_03849	10,4	19
Hypotetical protein	293_00872	1,6	14	Hypotetical protein	294_01260	50,4	18
Hypotetical protein	293_00619	17,3	13	Hypotetical protein	294_05019	15,3	14
60 kDa chaperonin 1	293_05544	56,3	13	Putative TrmH family tRNA/rRNA methyltransferase	271_00833	33,0	12
Hypotetical protein	293_03317	14,9	12	Phosphogluco-mutase	226_00008	57,7	10
Immunogenic protein MPT63 precursor	293_00984	35,5	12	Periplasmatic pH-dependent serine endo protease DegQ precursor	294_00233	23,2	9
Electron transfer flavoprotein subunit beta	293_04652	27,6	11	Hypotetical protein	294_00942	18,9	9
Elongation factor Ts	293_03051	28,9	11	Hypotetical protein	294_01766	12,6	8
Electron transfer flavoprotein subunit alpha	293_04651	31,8	9	Imunogenic protein MPT63 precursor	294_00424	33,5	8

### 4.3. Construção de um novo sistema de expressão

Para a construção de um novo sistema de expressão, avaliou-se 1) a capacidade de transformação e estabilidade plasmidial dos quatro isolados; 2) sequências reguladoras, promotoras e terminadoras de duas proteínas majoritariamente secretadas pelo isolado 294; e 3) buscou-se a construção de um novo vetor de expressão.

#### 4.3.1. Capacidade de transformação e estabilidade plasmidial

A capacidade de transformação de nove isolados foi avaliada preliminarmente (Emmerick, 2013). Um novo estudo mais aprofundado com os isolados 226, 271, 293 e 294 foi realizado usando o mesmo vetor utilizado anteriormente (pUS989), que foi construído a partir do pUS972 (figura 4.14) (Medeiros *et al.*, 2002). Este plasmídeo confere resistência ao antibiótico canamicina (para seleção dos transformantes) e possui a sequência codificante para a enzima beta-galactosidase de *E. coli* (gene *lacZ*) sob controle do promotor *PAN*. Assim, espera-se que bactérias transformadas produzam colônias de coloração azul em meio contendo canamicina e X-Gal.



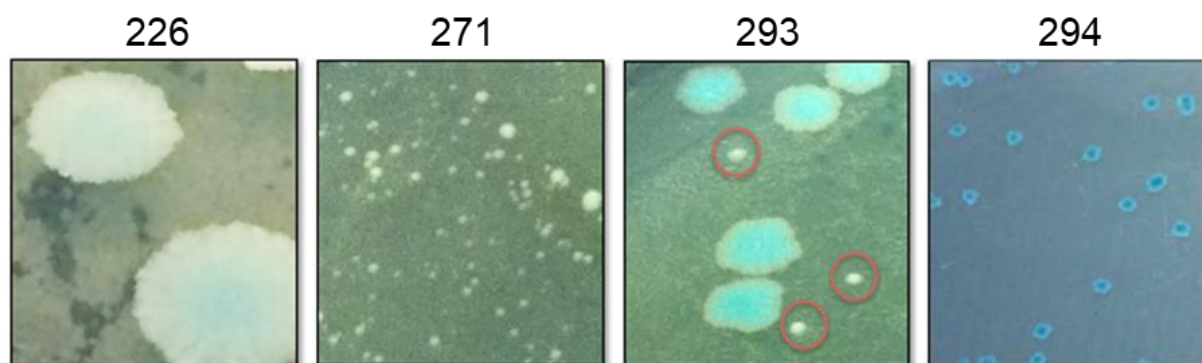
**Figura 4.14: Vetores pUS972 e pUS989.** O vetor pUS972 foi construído a partir do plasmídeo pUC18 com adição da origem de replicação em micobactérias a partir do plasmídeo pAL5000 e a substituição do gene de resistência à ampicilina por pelo gene de resistência à canamicina. O vetor pUS989 é resultado da adição do gene repórter *lacZ* e do promotor *PAN* isolado de *M. paratuberculosis* (Adaptado de Medeiros *et al.*, 2002).

Primeiramente, foi avaliado o tempo de duplicação destas bactérias (ou tempo de descanso pós eletroporação) ao longo de 6 horas após transformação em

triplicata, de forma a não superestimar a eficiência da transformação. Foram escolhidos os tempos de 2 h para o isolado 226 e 3 h para os isolados 271, 293 e 294 (dados não mostrados). Verificou-se que todos os isolados são sensíveis a canamicina (não crescem naturalmente em meio contendo o antibiótico) e que todos são capazes de serem transformados (crescem em meio contendo antibiótico após transformação). Porém, a eficiência de transformação e a integridade total do vetor é variada, visto que o isolado 271 apresentou apenas colônias brancas, 293, colônias azuis e brancas e 293 e 294 apresentam apenas colônias azuis (tabela 4.17, figura 4.15).

**Tabela 4.17: Capacidade de transformação.** Os valores de CFU representam a média e o desvio padrão de três experimentos independentes.

Isolado	Incubação (h)	CFU		% colônias azuis	Eficiência (azul) (CFU/ $\mu$ g DNA plasm.)
		Azul	Branco		
mc <sup>2</sup> 155	2	280	0	100	$1,40 \times 10^5$
226	2	$37,0 \pm 23,5$	0	100	$2,02 \times 10^4$
271	3	0	$1912 \pm 311,8$	0	0
293	3	$13,3 \pm 11,5$	$28,7 \pm 7,6$	31,7	$2,10 \times 10^4$
294	3	$29,3 \pm 22,5$	0	100	$1,47 \times 10^4$

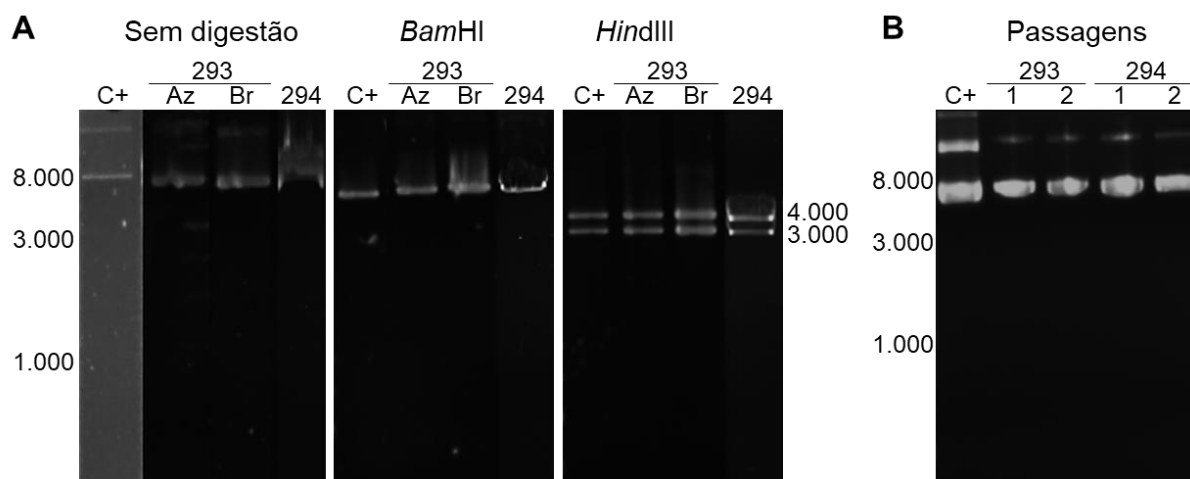


**Figura 4.15: Micobactérias transformadas com o plasmídeo pUS989.** Os quatro isolados micobacterianos foram eletroporados com 1  $\mu$ g do plasmídeo pUS989 e incubados a 22 °C por 7 dias. Os isolados 226 e 294 apresentaram apenas colônias azuis, 293, colônias brancas e azuis e 271, apenas colônias brancas.

Para avaliar a estabilidade plasmidial, o DNA plasmidial dos isolados 293 e 294 transformados foi extraído, introduzido em *E. coli* para aumentar a quantidade de material, extraído de *E. coli* e, novamente introduzido nos isolados ambientais. Em seguida, foram realizadas algumas passagens sucessivas em meio LB/ágar

contendo canamicina. O DNA plasmidial extraído de *E. coli* foi analisado por digestão com enzimas de restrição, visando encontrar grandes mutações que poderiam resultar na variação de tamanho desta molécula. Foram utilizadas as enzimas *Bam*HI e *Hind*III que cortam o vetor no sítio múltiplo de clonagem e antes e depois do gene *lacZ*, respectivamente. Não foram observadas grandes variações no tamanho do DNA plasmidial. E o DNA plasmidial extraído ao final das passagens foi visualizado diretamente em gel de agarose, também não sendo encontradas diferenças em relação ao plasmídeo original (figura 4.16).

A partir dos plasmídeos extraídos após a transformação dos isolados e de *E. coli* foi realizado sequenciamento do gene *lacZ* usando quatro conjuntos de iniciadores, que anelam em pontos diferentes do gene. Com isso, espera-se identificar a causa das colônias brancas e azuis, especialmente no isolado 293. No entanto, ainda não obtivemos esse resultado.



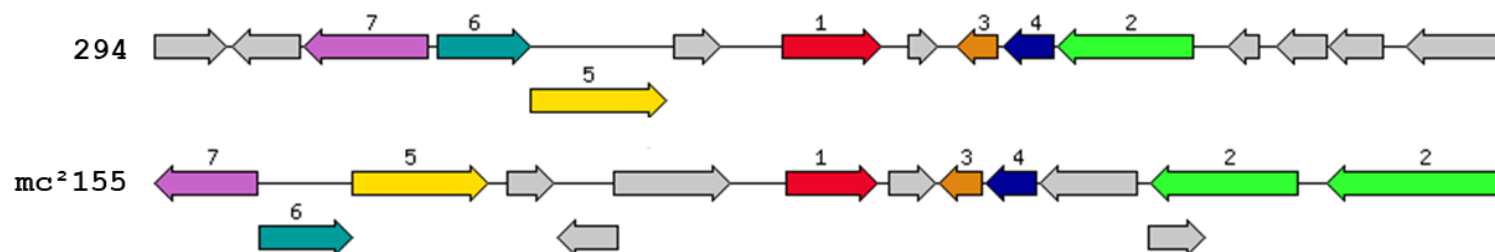
**Figura 4.16: Análise por digestão com enzima de restrição.** (A) O DNA plasmidial dos isolados 293 (colônia azul e branca) e 294 (azul) foram digeridos com as enzimas *Bam*HI e *Hind*III. (B) o DNA plasmidial extraído de colônias azuis dos isolados 293 e 294 após duas passagens (1 e 2). Os resultados foram visualizados em gel de agarose 1 %. O vetor pUS989 extraído de *E. coli* DH5 $\alpha$  foi usado com controle positivo.

#### 4.3.2. Análise de sequências reguladoras, promotoras e terminadoras

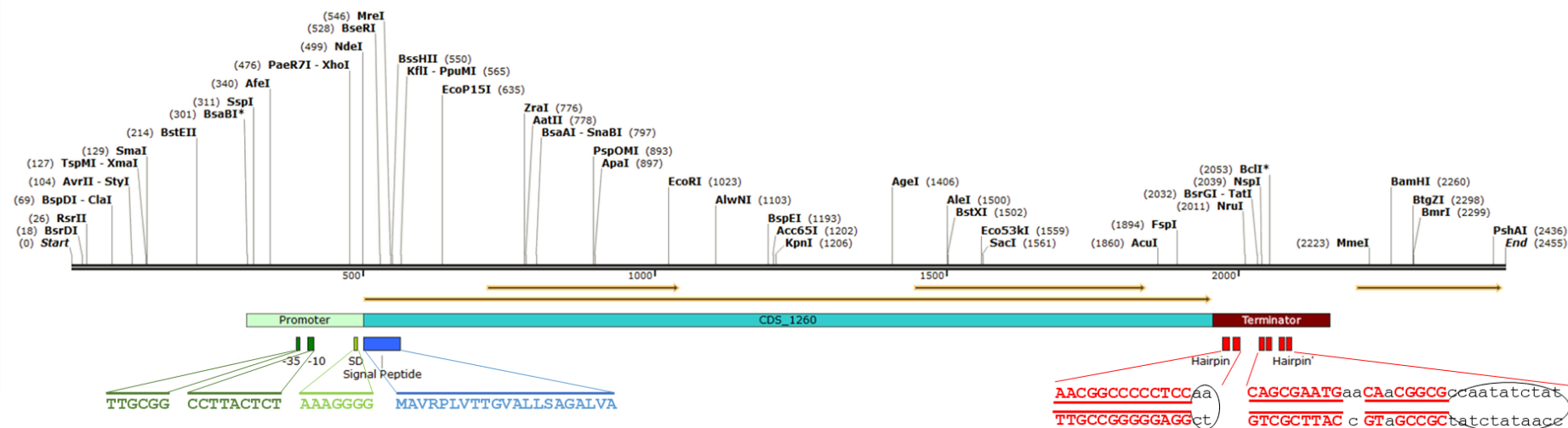
A proteína 294\_01260 foi descrita como hipotética na anotação do genoma do isolado 294 com base no sequenciamento pelo 454 GS Jr. Foi uma das proteínas identificadas pela análise dos *spots* da 2DE através do orbitrap (tabela 4.11B), sendo identificada também pelo QExactive e predita pela quantificação *label-free* como a terceira proteína mais secretada pelo isolado 294, apresentando

aproximadamente 50 kDa (tabela 4.16). Foi verificada a anotação dessa proteína a partir de uma busca direcionada através da ferramenta BLASTp, confirmando a anotação como proteína hipotética, apresentando entre 50 e 55 % de identidade com proteínas de diversas espécies de micobactérias. Parte dessa proteína apresenta similaridade com o domínio proteico conservado da família da glicoproteína mais abundantemente expressa no envelope externo dos vírus da Herpes (BLLF1). Essa glicoproteína, também chamada de gp350/220, é responsável por estimular a produção de anticorpos neutralizantes *in vivo* (Janz *et al.*, 2000).

Através da análise de sintenia desse gene, foi observado que este está situado em uma região contendo diversos genes sem função conhecida e com grande similaridade com a distribuição gênica de mc<sup>2</sup>155 (figura 4.17). A sequência nucleotídica do gene e da região flangeadora foi recuperada a partir do arquivo contendo os *contigs* da montagem realizada pelo Prokka e, após análise mais detalhada, foi observada a presença de outras duas CDS na mesma região e direção do gene 294\_01260, além de diversos sítios para enzimas de restrição (figura 4.18). Em seguida, a partir da sequência de aminoácidos, foi realizada a busca por peptídeo sinal através dos programas LipoP, SecretomeP, TatP e SignalP. Foi predita a região contendo os 21 primeiros aminoácidos através do SignalP (Gram positiva), que realiza a predição com base nos sistemas de secreção tipo Sec. Entretanto, os demais programas, que realizam essa busca por vias alternativas de secreção, não foram capazes de identificar o peptídeo sinal, indicando que essa proteína pode ser secretada pelo sistema tipo Sec. Através do programa BPRM foram identificadas as regiões -10 e -35 da região promotora. A análise da região 3' não-traduzida com o programa "mfold" levou à identificação de prováveis estruturas secundárias no DNA, indicando possíveis terminadores de transcrição (figura 4.18). Foram utilizados aproximadamente 200 nucleotídeos antes e depois da CDS para garantir a presença da região promotora e terminadora, entendendo que podem existir diversos outros sinais regulatórios importantes, mas ainda não caracterizados, para essas regiões.



**Figura 4.17: Sintenia 294\_01260.** (1) 294\_01260, (2) proteína hipotética, (3) proteína hipotética, (4) proteína hipotética, (5) Long form Mg-chelase associated protein with vWA domain, (6) Mg-chelatase, subunidade ChII, (7) serina/treonina quinase. Outras possíveis CDS estão representadas em cinza, mas sem anotação definida



**Figura 4.18: Proteína 294\_01260.** Linha preta – gene 294\_01260 mais 500 nucleotídeos nas regiões flangeadoras, com destaque para os sítios de enzimas de restrição; setas amarelas – CDSs presentes nessa região; verde – região promotora, com destaque para a região -35, -10 e sequência de Shine-Dalgarno; azul – gene 294\_01260, com destaque para o peptídeo sinal; vermelho – região terminadora, com destaque para dois *hairpins*.



Realizou-se também a análise da proteína 294\_04986: proteína hipotética aparentemente secretada em grandes quantidades de acordo resultado do Orbitrap e observação do gel bidimensional, embora esse resultado não se reflita na quantificação *label-free* realizada com os dados do QExactive. Apresenta um domínio conservado de hemóforo, ou seja, um polipeptídeo que liga grupamentos heme e o entrega a um sistema de transporte para importação. Possui ortólogo com *M. smegmatis*, estando próximo a transferases e proteínas transportadoras, e peptídeo sinal formado por 22 aminoácidos. Região promotora, sequência de Shine-Dalgarno e possível terminador de transcrição também foram preditos (dados não mostrados).

#### 4.3.3. Construção de um novo vetor

Um vetor de expressão precisa de algumas características básicas: 1) origem de replicação no hospedeiro desejado; 2) marcador de seleção das bactérias transformadas, geralmente resistência a antibiótico; 3) sítio múltiplo de clonagem, que contém regiões reconhecidas por diferentes enzimas de restrição; 4) promotor forte e regulável que dirija a expressão em grandes quantidades da proteína de interesse.

Para a construção de um novo vetor, foi utilizado como base o vetor pUS972 (figura 4.14). Também foram usadas as sequências reguladoras, região promotora e região terminadora do gene codificante para a proteína 01260 do isolado 294, identificada como predominante no filtrado de cultura. Outro elemento importante a destacar foi o desenho de uma região de parada de tradução em três fases, substituindo o códon de parada da proteína, de modo a garantir a final da transdução nessa região mesmo que tenham ocorrido eventos de mutação do tipo inserção/deleção (INDEL) durante a replicação bacteriana. Foram utilizadas duas abordagens principais, onde a principal diferença entre essas construções foi o gene repórter utilizado (I *mCherry* e II *Mpt64*). Ambas foram baseadas na técnica de clonagem de Gibson: um método de clonagem molecular que permite unir, em uma única reação a uma mesma temperatura, múltiplos fragmentos de DNA, não havendo restrição quanto ao tamanho do fragmento e a compatibilidade das extremidades. Essa técnica também reduz a formação de estruturas secundárias e

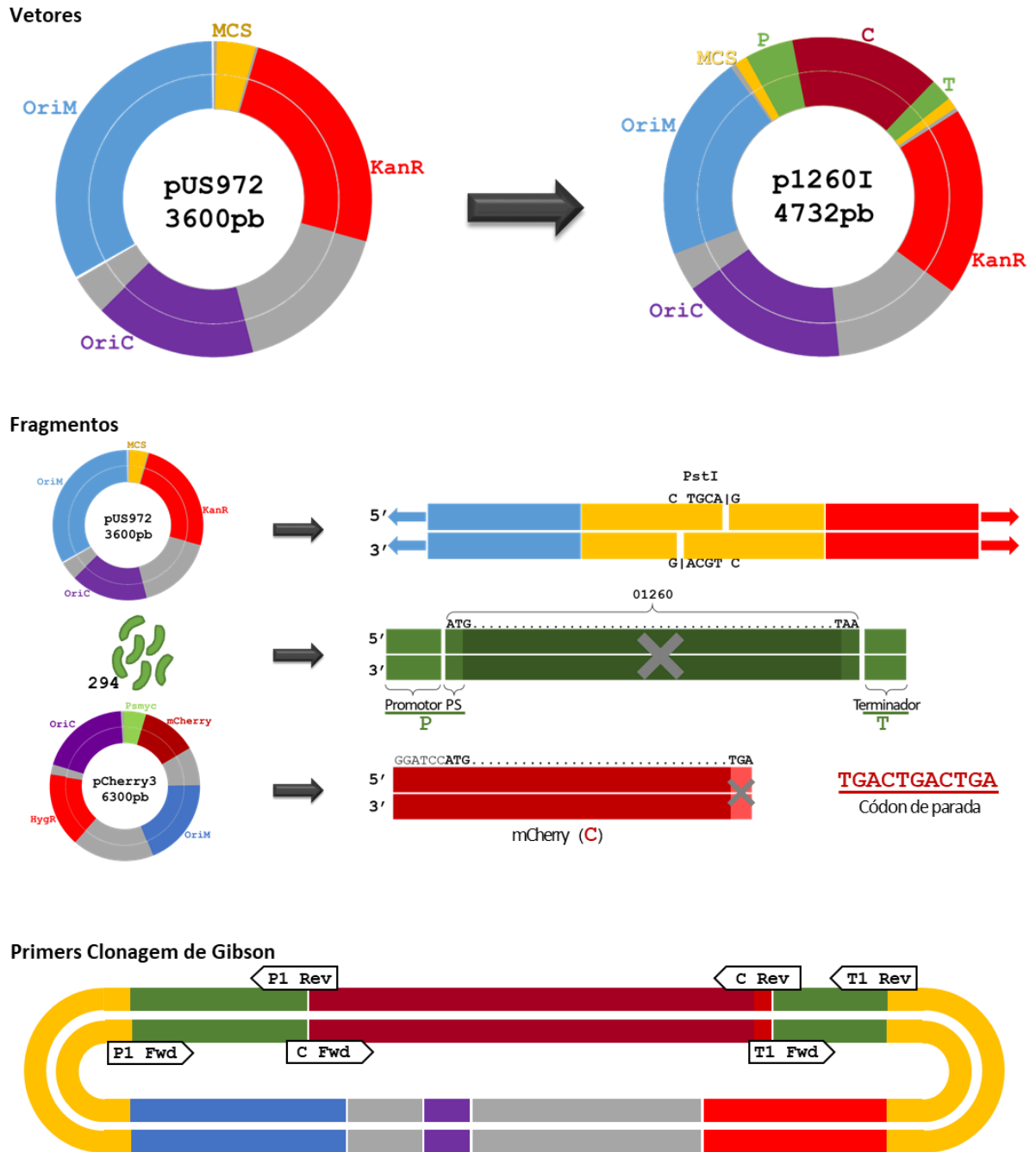
permite a clonagem sem cicatriz, ou seja, sem o corte do DNA em diversos sítios para enzimas de restrição para a inserção de fragmentos (Gibson *et al.*, 2009).

#### 4.3.3.1. Construção I – mCherry

Nesta construção foi utilizado o gene codificante para a proteína mCherry como gene repórter. mCherry é uma proteína fluorescente de tonalidade avermelhada derivada da proteína drFP583 (ou DsRed) isolada da anêmona *Discosoma sea*; que é uma proteína tetramérica e de maturação lenta. mCherry é resultado de sucessivas mutações na DsRed, sendo monomérica, de maturação rápida, com melhor fotoestabilidade e com excitação ótima a 587 nm e emissão máxima a 610 nm (Shaner, 2004).

Foi desenhado o esquema da construção do novo vetor e dos iniciadores para amplificação dos fragmentos (figura 4.19). Os iniciadores possuíam em média 25 nucleotídeos específicos do fragmento a ser amplificado e 15 relativos ao fragmento adjacente ao qual irá se ligar. Foi desenhado um códon de parada em três fases, substituindo o original da proteína mCherry, sendo adicionado à sequência do iniciador. Desse modo é possível garantir o final da tradução mesmo que haja alguma mutação do tipo inserção/deleção (INDEL).

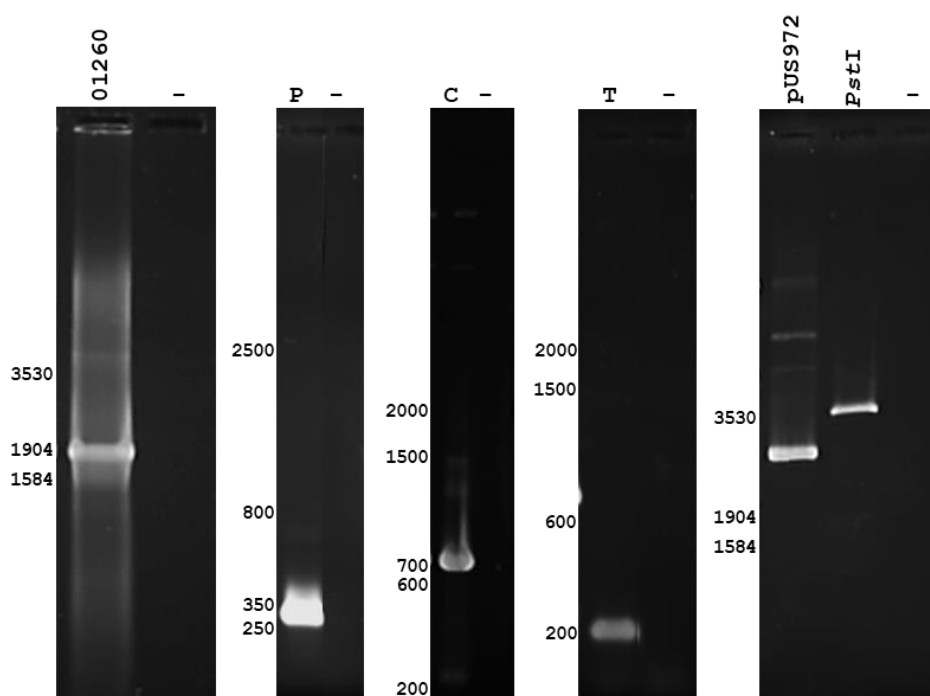
Em seguida, foi realizada a técnica de PCR para amplificação do fragmento de DNA contendo a região promotora e peptídeo sinal (P1) e do fragmento contendo o códon de parada e a região terminadora (T1) da proteína 294\_01260 usando como DNA molde o DNA genômico do isolado 294. Para amplificação do fragmento de DNA contendo o gene mCherry (C), foi extraído o plasmídeo pCherry3 de uma *E. coli* DH5 $\alpha$ , gentilmente cedida pelo Dra. Teca Calcagno Galvão (Laboratório de Genômica Funcional e Bioinformática/IOC), o qual foi utilizado como DNA molde. A tabela 4.18 mostra o tamanho esperado para os fragmentos antes e depois da PCR, quando a extremidade complementar à sequência adjacente é adicionada. Em paralelo foi realizada a digestão do pUS972 com a enzima de restrição *Pst*I, que corta no sítio CTGCAG criando extremidades coesivas (5' *overhang*). Todos os fragmentos e o vetor apresentaram o tamanho esperado, após a amplificação e digestão, respectivamente (figura 4.20).



**Figura 4.19: Esquema Construção I.** (Vetores) representação do vetor base (pUS972) e o vetor a ser construído (p1260I); (Fragmentos) indicação da origem dos fragmentos utilizados na construção; (Primers) esquema do vetor a ser construído com destaque para as regiões que deram origem aos iniciadores.

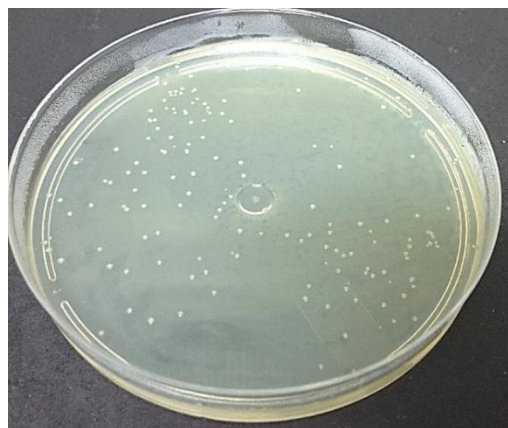
Tabela 4.18: Tamanho esperado para os fragmentos da construção I

Fragmento	Antes PCR	Após PCR
<b>01260</b>	1455 nt	1841 nt
<b>P1</b>	263 nt	292 nt
<b>C</b>	708 nt	728 nt
<b>T1</b>	146 nt	187 nt



**Figura 4.20. Fragmentos amplificados por PCR Construção I.** 1260, proteína 294\_01260 amplificada com P1 Fwd e T1 Rev; P, região promotora; C, mChery; T, região terminadora; pUS972, plasmídeo extraído de *E. coli*; *PstI*, vetor digerido com esta enzima de restrição; (-) controle negativo. Marcador a esquerda em pb.

Quantidades equimolares dos fragmentos foram adicionados a microtubo contendo o reagente de Gibson e o plasmídeo purificado linearizado (na proporção de três vezes mais fragmentos que plasmídeos), incubados por uma hora, transformados em *E. coli* Top10 e selecionados em meio contendo canamicina. Foi observado o crescimento de diversas colônias (figura 4.21), as quais foram analisadas por PCR de colônia e depois por extração do DNA plasmidial e visualização em gel de agarose. No entanto, nenhuma das colônias apresentou o plasmídeo com o inserto.

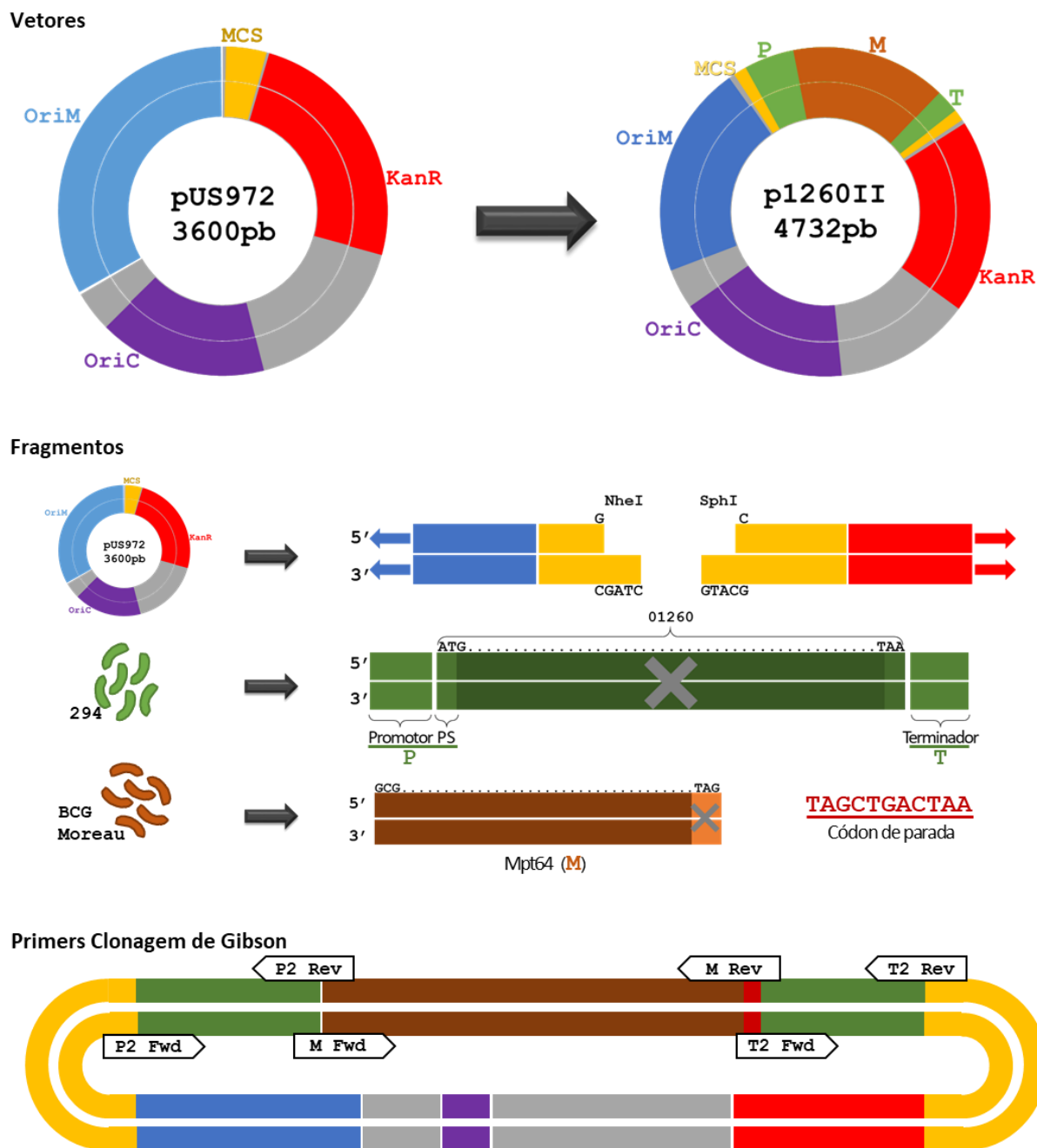


**Figura 4.21: Transformação Construção I.** Placa LB/ágar contendo colônias de *E. coli* TOP10 após transformação com 1  $\mu$ L diluído 1:2 da reação de Gibson e mantidas em estufa à 37 °C por 24 h.

#### 4.3.3.2. Construção II – Mpt64

Para uma nova construção, algumas mudanças em relação à tentativa anterior foram implementadas. Primeiramente, foi alterado o gene repórter. Utilizou-se o gene Mpt64, visto termos anticorpo contra a proteína por ele expressa, sendo capaz de detectar nas análises posteriores. O vetor foi digerido com duas enzimas de digestão diferentes, evitando que este se recircularize sem a presença do inserto. Por último, novos iniciadores foram mais cuidadosamente desenhados, seguindo as orientações para o desenho de iniciadores para clonagens não tradicionais: sem formação de *hairpins* e dímeros; extremidade 5' com homologia de 15 nucleotídeos; extremidade 3' específica ao gene, com tamanho de 18 a 25 bases, conteúdo GC entre 40 e 60 %, temperatura de *melting* entre 58 e 65 °C, e no máximo 3 G ou C nas últimas 5 bases.

Um novo esquema foi desenhado (figura 4.22), e seguiu-se com a amplificação dos fragmentos de DNA: P2, região promotora e peptídeo sinal; T2, códon de parada e a região terminadora; e M, gene Mpt64. Foi utilizado o DNA genômico de 294 como molde para amplificação dos fragmentos P2 e T2 e o DNA genômico de *M. bovis* BCG Moreau para amplificação do fragmento M. A tabela 4.19 mostra o tamanho esperado para os fragmentos antes e depois da PCR. O plasmídeo pUS972 foi digerido com as enzimas de restrição *NheI*, que corta no sítio GCTAGC (3' *overhang*), e *SphI*, que corta no sítio GCATGC (5' *overhang*). Ambas possuem a mesma temperatura de reação, mesmo tampão e mesma temperatura de inativação. Todos os fragmentos e o vetor apresentaram o tamanho esperado, após a amplificação e digestão, respectivamente (figura 4.23 A).

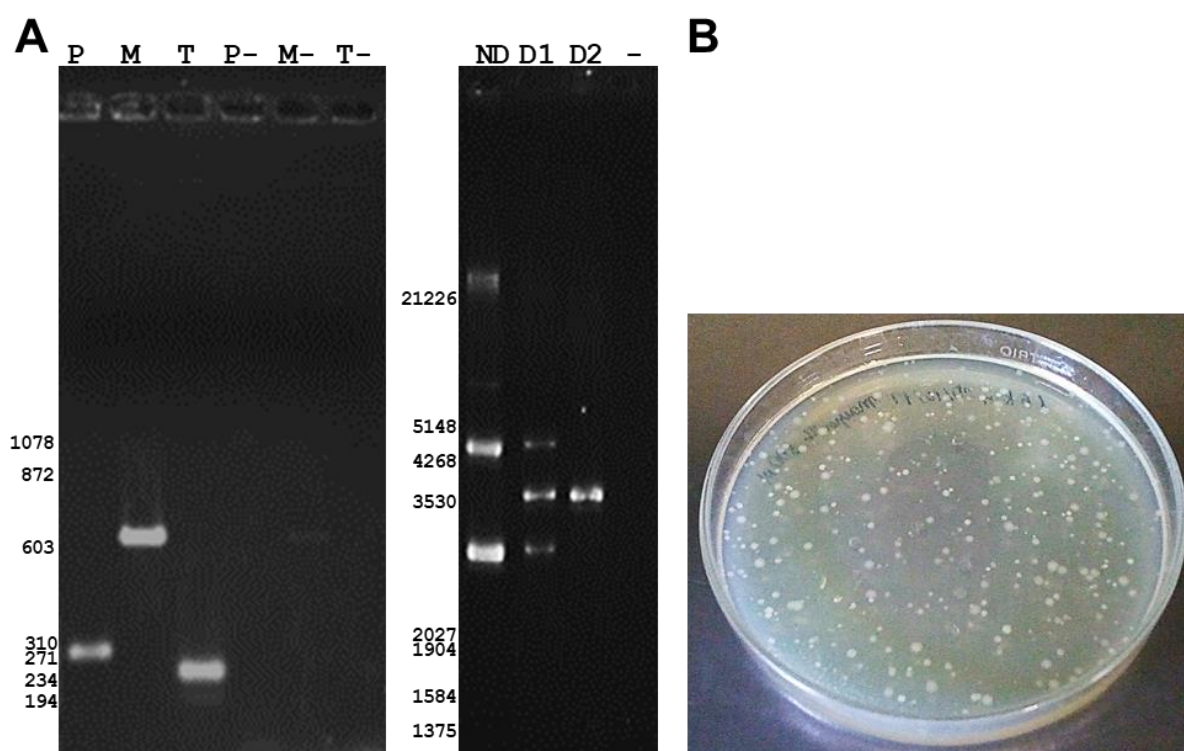


**Figura 4.22: Esquema Construção II.** Vetores – representação do vetor base (pUS972) e o vetor a ser construído (p1260II); Fragmentos – indicação da origem dos fragmentos utilizados na construção; *Primers* – esquema do vetor a ser construído com destaque para as regiões que deram origem aos iniciadores.

Então, foi realizada a reação de Gibson, a eletroporação de *E. coli* Top10 e seleção em meio com antibiótico. Foi observado o crescimento de diversas colônias (figura 4.23 B), as quais foram analisadas por PCR de colônia e posterior extração de DNA plasmidial e visualização em gel de agarose. Todavia, nenhuma das colônias apresentou o plasmídeo com o inserto.

Tabela 4.19: Tamanho esperado para os fragmentos da Construção II

Fragmento	Antes PCR	Após PCR
01260	1455	1869
P2	265	298
M	614	634
T2	209	248



**Figura 4.23: Fragmentos amplificados por PCR e transformação da Construção II.** (A) P, região promotora; M, Mpt64; T, região terminadora; ND, plasmídeo extraído de *E. coli*; D1, vetor digerido com as enzimas de restrição *NheI* e *SphI* simultaneamente; D2, vetor digerido com as mesmas enzimas de restrição sequencialmente; (-) controle negativo. Marcador a esquerda em pb. (B) Placa LB/ágar contendo colônias de *E. coli* TOP10 após transformação com 1  $\mu$ L diluído 1:2 da reação de Gibson e mantidas em estufa à 37 °C por 24 h.

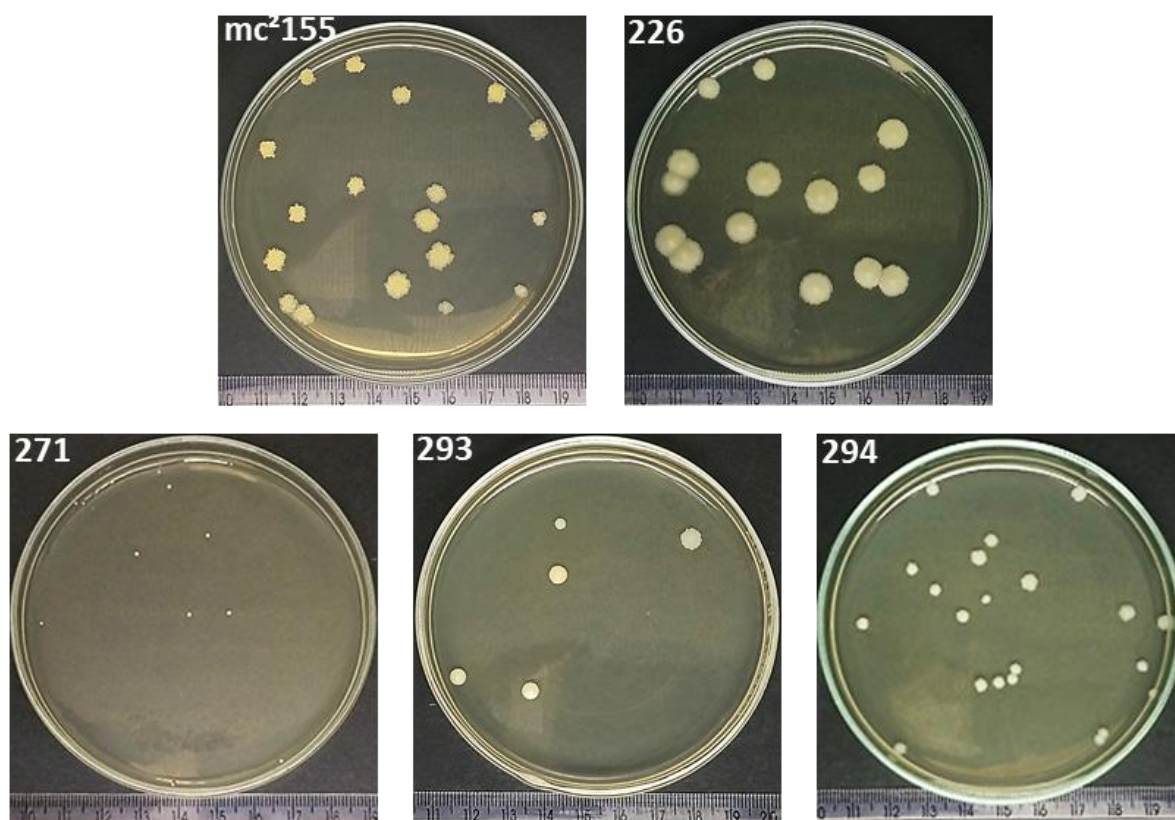
#### 4.4. Sobrevivência intracelular dos isolados ambientais

Para avaliar a capacidade de sobrevivência intracelular dos isolados ambientais, um indicador de sua possível patogenicidade, foi utilizado o modelo de infecção em células da linhagem monocítica humana THP-1 (após sua diferenciação para macrófagos) e avaliada a viabilidade dos isolados no ambiente intracelular. mc<sup>2</sup>155 foi incluído neste estudo para fins de comparação.

A quantificação dos lotes-semente dos quatro isolados micobacterianos e de *M. smegmatis* foi realizada através da determinação de CFU, onde todos apresentaram uma alta concentração bacteriana (tabela 4.20, figura 4.24). Os quatro isolados apresentam colônias espalhadas, os isolados 226 e 294 apresentam colônias rugosas e os isolados 271 e 293, colônias lisas. O tamanho varia entre 1 e 10 mm. Vale ressaltar que o isolado 226 apresenta colônias visíveis em três dias, enquanto que 271, em 7 dias. Os isolados 226 e 293 apresentam colônias ligeiramente amareladas, enquanto que 271 e 294, colônias brancas.

**Tabela 4.20: Quantificação dos lotes-semente.** O diâmetro das colônias foi determinado com base na média  $\pm$  desvio padrão de 5 colônias

Isolado	CFU	Diâmetro	Aspecto
mc <sup>2</sup>	$1,8 \times 10^8$	$6,4 \pm 0,99$ mm	Rugoso e espalhado
226	$1,4 \times 10^8$	$8,9 \pm 0,48$ mm	Rugoso e espalhado
271	$1,2 \times 10^8$	$1,5 \pm 0,11$ mm	Liso e espalhado
293	$1,3 \times 10^8$	$4,8 \pm 1,03$ mm	Liso e espalhado
294	$2,0 \times 10^8$	$4,0 \pm 0,58$ mm	Rugoso e espalhado



**Figura 4.24: Morfologia de colônia e contagem de CFU.**



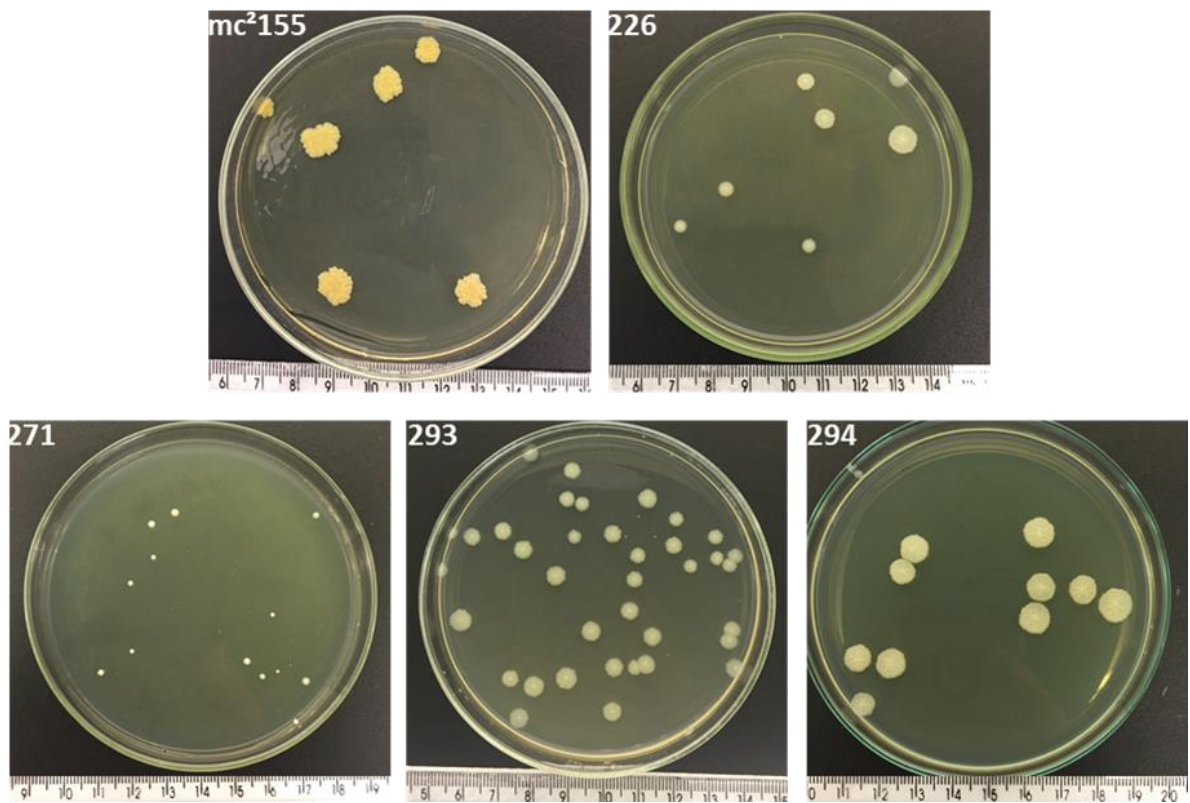
Após a diferenciação de THP-1, os macrófagos foram infectados com os isolados e mc<sup>2</sup>155 por 4 horas com uma multiplicidade de infecção (MOI) de 10:1. Após esse tempo, as células foram lavadas, a fim de retirar as bactérias que não foram internalizadas, e a quantidade de bactérias avaliada por CFU em diferentes tempos após infecção. Assim, foi possível observar a taxa de internalização comparada ao quantitativo inicial de bactéria utilizado para a infecção,  $2 \times 10^6$  bactérias (MOI 10:1). Todos apresentaram uma baixa taxa de internalização: mc<sup>2</sup>155, 293 e 271 entre 5,5 e 1,5 % das bactérias totais internalizadas e 226 e 294 com menos de 0,2 % (figura 4.25). Essa variação pode estar refletindo no quantitativo final observado ao longo da cinética.

A viabilidade dos isolados ambientais e de mc<sup>2</sup>155 foi avaliada através da contagem de CFU em LB/ágar nos tempos de 4, 24, 48 e 72 horas após infecção (figura 4.25). O perfil de viabilidade bacteriana se manteve ao longo do tempo avaliado para 226. Os isolados 293 e 294 apresentaram um leve declínio, em torno de 10 vezes, no número de bactérias e 271 apresentou uma redução da ordem de 100 vezes. Diferentemente dos isolados, mc<sup>2</sup>155 teve um aumento médio de 10 vezes a cada 24 h.

O aspecto morfológico das colônias dos isolados internalizados pelos macrófagos foi analisado, mas não foram encontradas diferenças significativas quando comparadas com as colônias antes da infecção (figura 4.26).



**Figura 4.25: Taxa de infecção e viabilidade dos isolados ambientais após infecção em células dTHP-1.** A taxa de infecção foi avaliada no tempo de quatro horas com base na quantidade de bactérias adicionadas na infecção ( $2 \times 10^6$ ). O número de bactérias recuperadas do interior dos macrófagos após as 4 h foi contado por CFU e plotado nas barras do gráfico (absoluto) com os valores a esquerda. O percentual dessas bactérias em relação ao número de bactérias adicionadas durante a infecção foi plotado na linha (relativo) com os valores à direita. A sobrevivência dos isolados e de mc<sup>2</sup>155 no interior de macrófagos THP-1 foi avaliada em triplicata a cada 24 h por 72 h. Todos os gráficos estão em escala logarítmica.



**Figura 4.26: Morfologia de colônia das bactérias retiradas dos macrófagos no tempo de 48 h após infecção**

## 5. Discussão

Embora Brasil seja mundialmente conhecido por sua biodiversidade, muito poucos estudos se concentram na exploração da microbiota nacional. Após a Conferência das Nações Unidas sobre Meio Ambiente e o Desenvolvimento, realizada no Rio de Janeiro em 1992, também chamada de Rio-92 ou Eco-92, o Brasil se comprometeu a conservar a diversidade biológica, promover seu uso sustentável e compartilhar os resultados benéficos do uso de fontes genéticas. Até o momento, nenhum levantamento da biodiversidade nacional tinha sido realizado. Em 1998 foi criado o Plano Estratégico da Biodiversidade Nacional, cujas ações resultaram em um primeiro levantamento da biodiversidade nacional e do potencial de cada região para expandir esses estudos (Lewinsohn & Prado, 2002). Uma versão parcialmente atualizada e expandida foi gerada no ano seguinte, sendo estes os únicos documentos oficiais com os dados da biodiversidade nacional gerados até o momento (Lewinsohn, 2003).

Nestes trabalhos foram identificadas apenas dezoito coleções nacionais de microorganismos. O grau de coleta e de conhecimento sobre os microorganismos é classificado na média nacional entre “nenhum” e “ruim” e, especificamente sobre a Mata Atlântica, como ruim – apesar da Região Sudeste e Sul, localidade de maior predominância deste bioma, conterem juntas cerca de 80 % do número de profissionais e instituições para o estudo dos microorganismos do país. Olhando especificamente para o reino *Monera* (*Bacteria* e *Archaea*, não há informações separadas), este levantamento menciona que no mundo são descritas 4.310 espécies e no Brasil, entre 800 e 900 espécies. Estima-se que menos de 1 % do total de microorganismos seja conhecido em todo o mundo. De acordo com a *List of Prokaryotic Names with Standing in Nomenclature* (<http://www.bacterio.net/Mycobacterium.html>) existem cerca de 15.000 espécies distribuídas em aproximadamente 2.500 gêneros descritas até março deste ano. Não há nenhuma informação disponível sobre a quantidade de bactérias identificadas até este momento na Mata Atlântica e/ou um levantamento nacional sobre as micobactérias identificadas. O próprio site do PARNASO menciona dados da flora e fauna, mas não dispõe de informações sobre a microbiota da região.

Assim, a Coleção de Bactérias do Ambiente e Saúde é de fundamental importância neste contexto: é a primeira coleção contendo microorganismos ambientais da Fiocruz, com destaque para as subcoleções “Bactérias da Amazônia Azul” (CBAA) e “Bactérias da Mata Atlântica” (CBMA). A CBMA é responsável pelo primeiro estudo da microbiota da Mata Atlântica do Rio de Janeiro e, conseqüentemente, do PARNASO. Esta coleção possui em seu acervo mais de 800 isolados bacterianos e mais de 20 gêneros diferentes, apresentando a mesma quantidade de isolados bacterianos (não existe ainda informações precisas sobre o número de espécies diferentes) que o total de bactérias descritas para todo o território nacional em 2002.

Entre os isolados da Mata Atlântica foram identificadas 19 bactérias classificadas como micobactérias. Os trabalhos realizados com algumas destas micobactérias, como o trabalho anterior do grupo (Emmerick, 2013) e os trabalhos do grupo da Dra. Ana Carolina Paulo Vicente (Morgado, 2017 e Morgado *et al.*, 2017), tem contribuído grandemente para a exploração da biodiversidade microbiana da Mata Atlântica e, em especial, do gênero *Mycobacterium*.

### **5.1. Análise genômica**

Para permitir a análise mais aprofundada de representantes destes clusters, foram escolhidos quatro isolados para o presente trabalho, os quais já foram objeto das nossas pesquisas anteriores. Os respectivos DNAs genômicos foram extraídos usando o kit *Blood & Cell Culture DNA Mini* (Qiagen), com uma etapa inicial de lise mecânica. Essa etapa é essencial para melhorar o rendimento desse processo, devido ao complexo envelope celular rico em ácidos micólicos que reveste a membrana plasmática das micobactérias e dificulta uma lise simples através de processos enzimáticos, com diferentes concentrações salinas e com variação da temperatura (Odumeru *et al.*, 2001; Liang *et al.*, 2006). Embora a lise mecânica não fosse ideal para a preparação de DNA genômico para fins de sequenciamento de nova geração, pois promove a quebra do DNA (Illumina e 454 Guide), essa etapa não é tão prejudicial, visto que o material será fragmentado para o preparo das bibliotecas. Após eletroforese do DNA obtido nesta lise, observa-se um leve grau de degradação no DNA genômico extraído dos isolados micobacterianos; no entanto, o

rendimento do processo é consideravelmente mais alto que outros métodos, e a alta qualidade das *reads* sequenciadas comprova a eficácia desta etapa. Entretanto, vale a pena destacar a variação na concentração de DNA genômico obtido de cada amostra. Visto que quantidades similares de células foram lisadas e o DNA dos quatro isolados eluídos no mesmo volume de tampão, esperava-se concentrações similares. Os isolados 226 e 293, pertencentes ao mesmo cluster, apresentaram concentrações próximas e um perfil similar no gel. Entretanto, a lise dos isolados 271 e 294 redenderam bem menos DNA genômico, o que já era esperado, já que a caracterização microbiológica realizada anteriormente para esses isolados mostrava que 271 e 294 tendem a formar grumos em meio de cultura líquido mais facilmente que 226 e 293 (Emmerick, 2013), dificultando a lise e a recuperação de DNA. Isso pode ser um indício de diferenças na composição lipídica do envelope celular destes isolados.

O sequenciamento pelo 454 GS Junior resultou, em média, em 180.000 *reads* e 80 milhões de bases por corrida. Apesar de superar consideravelmente o descrito para essa plataforma (100.000 *reads*, 40 Mb) (Levy & Myers, 2016), outros trabalhos obtiveram resultados semelhantes (Tun *et al.*, 2012). A corrida de sequenciamento realizada pela plataforma HiSeq foi compartilhada com diversas amostras de outros grupos; por isso, cada isolado apresenta uma quantidade de dados muito menor (560 MB em média) que aquela prometida pelo fabricante (até 1.000 GB) (MacCannell, 2016). Após a montagem das sequências genômicas observou-se que o isolado 271 apresenta um genoma com 1,5 Mb a menos que os demais. Provavelmente essa redução de tamanho tenha contribuído para uma maior geração de dados pelo 454. Quando o genoma foi sequenciado pelo equipamento HiSeq, foi informado previamente à RPT01J o tamanho esperado da sequência, com o intuito de dividir homoganeamente a corrida entre todas as amostras. Por isso, observamos uma redução na quantidade de *reads*, mas uma cobertura similar aos demais isolados. Segundo dados dos fabricantes, o 454 GS Jr. gera *reads* de 400 bases com 99 % de acurácia (Q20) enquanto que o HiSeq gera *reads* de 100 bases com mais de 99,9 % de acurácia (Q30) (Illumina e 454 Guide). Isso reflete os dados de qualidade encontrados neste estudo, com destaque para a alta qualidade das sequências geradas pela plataforma Illumina, que atingiu uma acurácia próxima de 99,99 % (Q40) em quase todas as *reads*.

A partir dos dados da montagem dos genomas, observou-se que o conteúdo GC e o tamanho do genoma está dentro do esperado para micobactérias (Rastogi *et al.*, 2001; <https://www.ncbi.nlm.nih.gov/genome>). Morgado (2017) encontrou resultados muito próximos aos deste estudo. O isolado 226, único com o genoma sequenciado em ambos os trabalhos, apresentou o mesmo conteúdo GC. É relevante analisar a diferença no tamanho do genoma encontrado no isolado 271. Prasanna e Mehra (2013) mostraram que o tamanho do genoma é inversamente proporcional ao tempo de duplicação, onde as espécies patogênicas tendem a ter um genoma menor e serem de crescimento lento. O crescimento lento confere às micobactérias uma maior capacidade de evasão dos mecanismos de defesa do hospedeiro e de resistência a antibióticos (Abuhammad, 2017). O isolado 326, pertencente ao mesmo cluster que 271, também apresentou um genoma menor (5,05 Mb) (Morgado, 2017). Analisando a curva de crescimento de 271 e 326, observa-se que apresentam crescimento mais lento em relação aos demais isolados, com tempo de duplicação próximo a 24 h (Emmerick, 2013). Esses indícios sugerem 271 como possivelmente patogênica. De qualquer forma todas as micobactérias são classificadas, no Brasil, como nível de biossegurança NB2 (exceto pelas espécies manifestamente patogênicas para o homem ou para animais, de nível NB3, como *M. bovis* e *M. tuberculosis*), e são manipuladas em condições de contenção apropriadas.

Entre as 166 espécies de micobactérias com o genoma sequenciado, considerando apenas a primeira vez que cada espécie foi sequenciada, 8 o foram utilizando a plataforma 454 (sendo três complementados com tecnologia Illumina e dois com PacBio), dos quais sete obtiveram a montagem em apenas um *contig*; 102 utilizaram a plataforma Illumina (três complementados com tecnologia 454 e seis com PacBio), dos quais apenas 13 apresentam a montagem em apenas um *contig* (seis gerados exclusivamente pelo Illumina). No entanto, não existe uma relação matemática direta que consiga determinar a cobertura necessária para que um genoma seja completamente montado. Por exemplo, *M. goodii* apresenta o quarto maior genoma (7,11 Mb) entre as micobactérias e foi completamente montado apesar de ter apenas 17 x de cobertura. Por outro lado, *M. parascrofulaceum*, com cobertura de 53 x e 6,56 Mb, foi montado em 405 *contigs*. Ambos foram sequenciados pelo 454 e montados pelo Newbler (<https://www.ncbi.nlm.nih.gov/>

genome). Embora hajam poucos genomas sequenciados pelo 454, esta plataforma é muito eficaz na montagem completa de genomas micobacterianos (em apenas um *contig*), tendo em vista o tamanho dos *reads* (~ 400 nucleotídeos) e o tamanho limitado dos genomas, facilitando o processo. Assim, obteve-se um pequeno número de *contigs* entre os isolados da CBMA com a montagem pelo 454, apesar da cobertura ter sido pequena. A montagem completa *de novo* de um genoma através da plataforma Illumina é mais complicada, por ser uma metodologia baseada em *reads* curtas. Existem diversos programas para montagem de dados gerados pelo Illumina. A escolha do programa e das configurações mais adequadas são essenciais para a obtenção de resultados satisfatórios durante a montagem. O número de *contigs* gerados pela montagem dos dados do Illumina não foi muito inferior, embora a cobertura tenha chegado a quase 100 x. A montagem dos dados do 454 foi realizada em colaboração com o MSc. Ricardo Junqueira, ao contrário dos dados do Illumina, que foram montados utilizando a interface *on-line* do PATRIC ([www.patricbrc.org](http://www.patricbrc.org)). Através de estudo complementar de sentenças e o uso de *reads* de menor qualidade de ambas as tecnologias de sequenciamento, para fins de mapeamento, espera-se conseguir ainda minimizar o número de *contigs* e melhorar a montagem final.

Existem diversas metodologias e programas para anotação funcional de seqüências, sendo os principais o RAST e o Prokka. O RAST oferece um *pipeline* gratuito de anotação para genomas bacterianos completos ou quase completos. A partir do *upload* do genoma em formato FASTA, os usuários têm acesso à anotação dentro de um período de 12-24 h, com elevada acurácia, abrangência e consistência. Essa qualidade é baseada em uma crescente biblioteca de subsistemas curados manualmente e na disponibilidade de anotações de amplas famílias proteicas derivadas desses subsistemas (Aziz *et al.*, 2008). O Prokka (<http://www.vicbioinformatics.com/software.prokka.shtml>) é um software gratuito, que utiliza linha de comando, para anotação de genomas de procariotos. Esta ferramenta pode ser instalada em sistema Unix/Linux e permite a anotação bastante precisa de um genoma bacteriano, a partir de seqüências completas ou de *scaffolds* em formato FASTA, em pouco tempo (geralmente menos de uma hora) (Seemann, 2014). Foi escolhido o Prokka por gerar os dados mais rapidamente, com melhor qualidade e por disponibilizar os arquivos resultantes da anotação em vários



formatos diferentes, permitindo diversas outras análises. O servidor RAST, com interface amigável, foi utilizado para obtermos as informações dos subsistemas e realizarmos diversas análises adicionais. Embora ambas metodologias possuam alta acurácia, sempre há discrepâncias, incertezas e lacunas entre seus resultados, reforçando a necessidade de comprovação experimental *in vitro* ou *in vivo* para muitas das hipóteses oriundas das análises *in silico*.

Segundo dados genômicos de outras 10 micobactérias, observa-se que a quantidade de CDS é diretamente proporcional ao tamanho do genoma (aproximadamente 1:1000), com exceção para *M. leprae* (1.604 CDS para 3,27 Mb de genoma). As micobactérias apresentam em média 47 genes codificantes para tRNAs (<https://www.ncbi.nlm.nih.gov/genome>; Cole *et al.*, 1998; Cole *et al.*, 2001; Garnier *et al.*, 2003; Ho *et al.*, 2012; Saini *et al.*, 2012; Tobias *et al.*, 2012; Das *et al.*, 2016). Os isolados estudados neste trabalho também apresentam em torno de 1:1000 CDS preditas em relação ao tamanho do genoma. O número de genes codificantes para tRNAs apresentado pelos isolados 226, 271 e 293 é muito similar aos outros genomas micobacterianos. No entanto, 294 apresenta 76 tRNAs, de acordo com a montagem pelo Prokka (a partir de dados do 454), ou 63, de acordo com a montagem pelo PATRIC (a partir dos dados do Illumina). Nenhuma micobactéria sequenciada até o momento e com o genoma disponível no GenBank apresenta um número tão expressivo de genes para tRNAs quanto 294, nem mesmo os hospedeiros micobacterianos mais comumente usados para expressão de proteínas heterólogas como *M. smegmatis* (47), *M. vaccae* (49) e *M. aurum* (46). Ainda é necessário analisar as características deste elevado número de tRNAs, mas este dado pode indicar uma maior flexibilidade no uso de códons, sugerindo a escolha deste isolado como um bom hospedeiro para o desenvolvimento de um novo sistema de expressão. Todavia, no trabalho de Morgado (2017), os isolados 247 (mesmo cluster que 294) e 312 (mesmo cluster que 226 e 293) foram sequenciados pela plataforma Illumina, anotados pelo Prokka e descritos como apresentando 75 e 85 tRNAs, respectivamente. Como comparação, *E. coli* apresenta um número de genes para tRNAs ainda maior (89).

A maioria das micobactérias de crescimento lento tendem a apresentar apenas uma cópia de cada um dos genes de rRNA (5S, 16S e 23S), enquanto que as de crescimento rápido apresentam esses genes duplicados (Tortoli, 2003). Os isolados

293 e 294 apresentaram esses genes duplicados segundo a anotação realizada pelo Prokka com os dados do 454 GS Jr. Esse fato pode explicar um perfil de crescimento mais rápido para esses isolados em relação aos isolados 226 e 271 (Emmerick, 2013). No entanto, através da anotação dos dados gerados pelo HiSeq foram encontradas apenas uma cópia de cada um desses genes levantando dúvidas sobre o real número de rRNAs nos quatro isolados. Uma análise mais direcionada se faz necessário para elucidar essa questão.

Os subsistemas usados pela anotação com o RAST são um conjunto de papéis funcionais que apresentam alguma relação entre si, sendo definidos através de análises no laboratório (Aziz *et al.*, 2008). Foi observado que isolados ambientais apresentam mais genes envolvidos no metabolismo de compostos aromáticos, de derivados de enxofre e de ferro, visto que esses são constituintes comuns encontrados no solo (Martins *et al.*, 2007). Por outro lado, H37Rv apresenta aproximadamente o dobro da quantidade de genes envolvidos em regulação e sinalização celular e no metabolismo de fósforo que os isolados ambientais e mc<sup>2</sup>155. H37Rv, por ser um patógeno intracelular, apresenta diversos mecanismos para evadir do sistema imune e sobreviver no ambiente intracelular. O fósforo encontra-se em maior abundância no corpo humano, estando envolvido em diversos processos metabólicos, que disponível no solo

Hidrocarbonetos aromáticos policíclicos são poluentes ambientais resultantes da combustão incompleta de combustíveis fósseis ou da incineração de resíduos, sendo considerados tóxicos, mutagênicos e carcinogênicos. Esse material particulado se deposita e impregna no solo, devido à sua propriedade hidrofóbica, sendo de difícil eliminação, diminuindo processos de biodisponibilidade e biorremediação. No entanto, diversas micobactérias já foram identificadas como capazes de degradar esses compostos (Miller *et al.*, 2004). Outra substância rica em hidrocarbonetos aromáticos policíclicos é a lignina: uma macromolécula tridimensional amorfa encontrada nas plantas terrestres, associada à celulose na parede celular cuja função é de conferir rigidez, impermeabilidade e resistência a ataques microbiológicos e mecânicos aos tecidos vegetais. A lignina compõe entre 18 % e 35 % da massa seca da parede celular dos vegetais (Azadi *et al.*, 2013). É mais provável que as micobactérias ambientais aqui estudadas estejam adaptadas a degradação desse tipo de material (oriundo da decomposição da matéria orgânica)

do que de poluentes ambientais, visto que foram isoladas de uma Unidade de Proteção Integral – PARNASO.

O isolado 293 apresenta um alto número de genes relacionados com a presença de fagos, profagos, elementos transponíveis e plasmídeos; no entanto, não foi realizada uma busca direcionada para a identificação deste tipo de elementos replicativos entre os *reads* obtidos dos quatro isolados no estudo. A observação do DNA obtido nas preparações antes do sequenciamento não permitiu suspeitar da presença de plasmídeos. Existem poucos plasmídeos micobacterianos descritos (Shintani *et al.* 2015), mas destacamos o pCBMA213\_2, um plasmídeo circular com pouco mais de 160 kb, 161 CDS e 65,4 % de conteúdo GC que foi isolado de uma micobactéria da CBMA (Morgado *et al.*, 2017). Em outro trabalho do mesmo grupo foram descritos também indícios da presença de quatro bacteriófagos e cinco plasmídeos em outros isolados da CBMA (Morgado, 2017). É importante destacar que os plasmídeos são importantes veículos para a comunicação de informação genética entre bactérias, que podem contribuir com a troca de características relativas à patogenicidade e interação com o meio ambiente, promovendo uma rápida evolução e adaptação aos mais diferentes habitats. Plasmídeos também são importantes ferramentas genéticas usadas para manipular e analisar microorganismos, facilitando a introdução, modificação e remoção de genes alvos e, também, para direcionar a produção de grandes quantidades de proteínas recombinantes para os mais diversos fins (Shintani *et al.* 2015). Neste contexto, destaca-se o pAL500 (Labidi *et al.*, 1985), um plasmídeo de baixo número de cópias isolado de *M. fortuitum* que tem sido usado no desenvolvimento de novos vetores micobacterianos (Lazraq *et al.*, 1991; Medeiros *et al.*, 2002).

Embora o subsistema do RAST relacionado a “virulência, doença e defesa” forneça informações importantes sobre os genes envolvidos nestes processos, é impossível relacionar o número destes genes ao grau de patogenicidade de uma espécie. Por exemplo, mc<sup>2</sup>155, uma espécie notoriamente saprófita e usada largamente na pesquisa, apresenta mais genes nesse subsistema que *M. tuberculosis*, agente etiológico da tuberculose – principal doença infectocontagiosa desde 2015 (WHO, 2016). Assim, mesmo que 271 apresente um baixo número de genes relacionados a doenças infecciosas (dados do PATRIC e do RAST) e os outros isolados apresentem mais genes nessas categorias que H37Rv, não é

possível fazer qualquer tipo de inferência, sendo necessários outras análises para esclarecer essa questão.

Comparar um grande volume de dados é sempre árduo, porém é comum recorrer a comparações de blocos de dados para se completar uma tarefa, reduzindo a especificidade da informação, mas aumentando a abrangência da análise. Assim, o perfil metabólico dos quatro isolados micobacterianos da CBMA foram comparados entre si e a outras espécies de referência (ambientais e patógenos intracelulares) genericamente, através das vias metabólicas, sem olhar especificamente para as enzimas. Por isso, é possível que bactérias apresentem o mesmo número de enzimas em uma determinada via, mas que sejam enzimas totalmente diferentes. No entanto, o fato de um microorganismo apresentar mais genes codificantes para enzimas em uma determinada via que outro sugere que esta via tenha uma importância maior para aquele microorganismo.

Na grande maioria das vias, 271 apresenta uma quantidade muito similar de genes aos patógenos intracelulares, provavelmente devido ao fato de também apresentarem genomas de tamanhos similares. Na via de metabolismo de andrógenos e estrógenos (derivados do colesterol) observamos que 226, 293 e 294 apresentam quatro vezes mais genes nesta via que 271. A habilidade de realizar etapas na via de síntese destes hormônios já foi descrita em micobactérias desde a década de 1960 e recentemente tem se proposto o engenheiramento de *M. smegmatis* para a produção de testosterona, substituindo a síntese química (Dlugovitzky *et al.*, 2015; Sood *et al.*, 2017). Foram identificadas onze vias envolvidas na degradação de compostos aromáticos, atividade característica de diversas micobactérias ambientais (Kweon *et al.*, 2015; Hennessee & Li, 2016): os isolados 226, 293 e 294 apresentaram 30 % de aumento no número de genes em relação a *M. bovis* BCG Pasteur e *M. tuberculosis* H37Rv e 15 % a menos que as espécies ambientais de referência (mc<sup>2</sup>155, *M. vaccae*, *M. sp.* JLS). As nove bactérias, os quatro isolados da Mata Atlântica e cinco espécies de referência, apresentam genes de resistência a penicilina, cefalosporina, tetraciclina, beta-lactâmicos e antibióticos do grupo vancomicina, reforçando a característica de resistência a antibióticos das micobactérias (Al-Saeedi & Al-Hajoj, 2017). Entre as 14 vias envolvendo o metabolismo de aminoácidos, apenas seis apresentaram alguma diferença expressiva: cinco vias apresentaram 15 % mais genes para os isolados da

CBMA e as espécies ambientais de referência em relação aos patógenos intracelulares. No entanto, na via de metabolismo dos aminoácidos cisteína e metionina, que apresentam enxofre em sua cadeia lateral, os isolados ambientais apresentam 20 % menos genes que os patógenos intracelulares. Essa menor quantidade de genes envolvidos no metabolismo de aminoácidos sulfurosos pode estar relacionada à maior disponibilidade de enxofre no solo que em hospedeiros vertebrados.

Durante a caracterização geral destes isolados (Emmerick, 2013), foi realizada a identificação de alguns antígenos compartilhados (Mpt64, Mpt70, Mpt83, HspX, GlnA1, Ag85) com o complexo *M. tuberculosis* através de western blot. A proteína Mpt64 não foi identificada na caracterização preliminar (Emmerick, 2013), mas o gene foi encontrado nos quatro isolados, e duplicado em 226, 271 e 293. Essa proteína também foi identificada em três spots de 294 através do orbitrap e no secretoma de 226. Mpt64 é um antígeno secretado codificado pela região RD2, presente nas cepas BCG Moreau (cepa vacinal brasileira), Japão, Rússia e Suécia, mas deletada da cepa BCG original na década de 1920, sendo um candidato para diagnóstico tanto em infecções em gado como em humanos (Roche *et al.*, 1996). Mpt70 é a principal proteína secretada de *M. bovis* BCG em meio Sauton (Harboe & Nagai, 1984) e estimula a resposta imune humoral e celular durante infecções com *M. tuberculosis* ou *M. bovis* (Al-Attayah *et al.*, 2003). Mpt70 foi identificada no filtrado de cultura de todos os isolados da CBMA (Emmerick, 2013) e, no atual estudo, em três spots de 294 e no secretoma de 226, 293 e 294. O fato do gene codificante não ter sido identificado no genoma de 271 pode indicar deficiências no processo de sequenciamento, na montagem genômica ou na anotação gênica. GlnA1 é a principal e essencial glutamina sintetase em *M. tuberculosis*, e parece ter um papel na sua sobrevivência durante infecção e do hospedeiro humano (Hayward *et al.*, 2009). Esta proteína foi identificada no lisado total de 226, 293 e 294 (Emmerick, 2013) e quatro cópias em 294 pelo QExactive, além de uma cópia deste gene em 226.

O complexo do antígeno 85 (Ag85) é composto por três distintas trealose dimicolil transferases, Ag85A, Ag85B e Ag85C (codificada pelos genes *fbpA*, *fbpB* e *fbpC*), que estão envolvidas na deposição de micolato e na biosíntese tanto dos ácidos micólicos quanto da parede celular, sendo críticas para a construção da

parede celular micobacteriana normal e o crescimento celular (Belisle *et al.*, 1997). Esses antígenos foram identificados nos quatro isolados micobacterianos ambientais por western blot (Emmerick, 2013) e pelo sequenciamento e anotação do genoma; em 271 e 294 pela análise de seus *spots* no orbitrap; e em 226, 293 e 294 através da análise do secretoma pelo QExactive. Todavia, nem todos os genes/proteínas pertencentes ao complexo foram encontrados nessas análises. É possível que tenha ocorrido algum erro de anotação no isolado 226, identificando como Ag85B o que seria Ag85A, fazendo com que o mesmo tivesse dois Ag85B e nenhum Ag85A; ou identificando como MPT51 o que seria Ag85A. Algo semelhante pode ter ocorrido com 271 e 293 e nas amostras proteicas também. Mpt83 e HspX não foram identificados na análise inicial (Emmerick, 2013) nem no genoma ou secretoma destes isolados.

Também foi realizada uma busca contra outros antígenos micobacterianos importantes, tanto no genoma quanto no secretoma destes isolados. O antígeno Mpt51, codificado pelo gene *fbpD* (Rv3803c), age na ligação de fibronectina à matriz extracelular e é um importante fator de virulência que induz a resposta imune Th1, promovendo a proteção em camundongos infectados com *M. tuberculosis*, sendo um candidato para o desenvolvimento de vacinas BCG recombinantes (Silva *et al.*, 2009; da Costa *et al.*, 2014). Foi identificado no genoma de 226, 293 e 294 e no secretoma de 226. Mpt53, codificada pelo *rv2878c*, é uma proteína secretada, antigênica, de 15 kDa, que induz uma resposta de anticorpos na tuberculose bovina, sendo encontrada em diversas MNT (Johnson *et al.*, 2001); está presente no genoma dos 4 isolados (duas cópias em 271) e em dois *spots* excisados de 271. Mpt63 é a principal proteína secretada por *M. tuberculosis*, com propriedades imunogênicas e relacionada à virulência, estando presente em níveis elevados na resposta humoral de paciente infectados com tuberculose (Kundu *et al.*, 2016). Está presente no genoma de 226, 293, 294 e *M. smegmatis* mc<sup>2</sup>155 em múltiplas cópias e apenas em uma cópia em H37Rv e BCG. A proteína foi identificada em sete *spots* excisados de 294 e pelo QExactive nos isolados 293 e 294. MTB12, ou CFP-2, é uma das principais proteínas secretadas de *M. tuberculosis* com um importante papel na resposta pró-inflamatória nos estágios iniciais da tuberculose em humanos, sendo um candidato para aumentar a acurácia dos diagnósticos contra a tuberculose (Lee *et al.*, 2006). Cada um dos quatro isolados apresentam um gene deste no

genoma; 294 apresenta esta proteína em um *spot* analisado pelo MALDI e em três *spots* analisados pelo Orbitrap; e 271, 293, 294 apresentam esta proteína no secretoma. A lipoproteína de 19 kDa (LpqH), identificada em múltiplas cópias no genoma dos isolados, está envolvida na apoptose de macrófagos infectados com *M. tuberculosis* e induz a resposta imune mediada por linfócitos T (Sánchez *et al.*, 2012). CFP-6, identificado no filtrado de cultura de *M. tuberculosis*, desencadeia uma intensa resposta imune (*in vitro* e *in vivo*) (Bhaskar *et al.*, 2000), estando no genoma de todos os isolados estudados.

Identificar micobactérias não é uma tarefa trivial, a exemplo de *M. abscessus*. Inicialmente *M. abscessus* foi descrita como uma espécie independente (Moore & Frerichs, 1953); porém, quase 20 anos depois, foi reclassificado como subespécie de *M. chelonae*, visto apresentarem sequências idênticas da região 54 a 510 do gene 16S *rRNA* (Kubica *et al.*, 1972). Mais 20 anos se passaram e foram observadas diferenças no conteúdo genético e em testes bioquímicos entre essas “subespécies”, e *M. abscessus* novamente foi elevada a espécie (Kusunoki & Ezaki, 1992). Em seguida, foram identificadas as espécies *M. bolletti* (Adékambi *et al.*, 2006 A) e *M. massiliense* (Adékambi *et al.*, 2004 e 2006 B). Elas foram agrupadas como a mesma espécie e propostas como subespécie de *M. abscessus* (*M. abscessus abscessus* e *M. abscessus massiliense*), devido à alta identidade de seus marcadores filogenéticos (Leão *et al.*, 2009). Por descumprir algumas normas de classificação, *M. abscessus massiliense* foi reclassificada como *M. abscessus bolletti* (Leão *et al.*, 2011). Depois observou-se que algumas cepas dentro de *M. abscessus* apresentavam diferenças nos perfis de susceptibilidade a macrólidos (grupo de antibiótico) e, após estudo com 41 cepas, esta espécie foi dividida em três subespécies: *M. abscessus abscessus*, *M. abscessus bolletti* e *M. abscessus massiliense* (Tortoli *et al.*, 2016). No ano seguinte, Adékambi e colaboradores (2017) propuseram elevar à espécie *M. bolletti* e *M. massiliense*, com base em estudos genômicos. Por último, Tortoli e colaboradores (2018) escreveram uma carta ao *International Journal of Systematic and Evolutionary Microbiology* rejeitando a proposta de Adékambi e colaboradores (2017) esclarecendo os diversos motivos.

Assim, diversas metodologias têm sido propostas e exploradas ao longo dos anos: testes bioquímicos, condições de crescimento, análise por digestão de fragmentos de PCR (*PCR restriction analysis*, PRA), sondas de DNA, cromatografia

(camada delgada, gás-líquido e líquida de alta performance (HPLC)) e espectrometria de massas por MALDI-TOF. Diversas destas metodologias têm sido amplamente usadas em laboratórios clínicos; no entanto, são testes lentos, não são 100 % específicas, apresentam reação cruzada, necessitam de etapas específicas de preparação da amostra e muitas vezes são inconclusivos. Por isso, a metodologia de referência é a amplificação por PCR seguido de sequenciamento de (fragmentos de) genes conservados – especialmente *16S rRNA*, *hsp65*, *rpoB* e *sodA* – por ser mais rápida, mais simples, mais específica e com menos erros e não identificações (Tortoli, 2014). Quando a análise do *16S rDNA* não é suficiente para identificar espécies, a análise multi-local traz resolução suficiente, e foi também chamada de análise de *barcodes*. Embora ainda seja a mais utilizada atualmente na descrição de novas espécies (Nouioui *et al.*, 2017; Paniz-Mondolfi *et al.*, 2017; Lee & Whang, 2017; Fukano *et al.*, 2017; Shahraki *et al.*, 2017; Gcebe *et al.*, 2017), essa metodologia aos poucos tem sido substituída pela taxonomia genômica, que compara o genoma total das espécies ao invés de um conjunto de genes pontuais (Fedrizzi *et al.*, 2017; Tortoli *et al.*, 2017; Gupta *et al.*, 2018). De fato, o sequenciamento genômico e análise comparativa através de bioinformática é na prática tão rápido quanto análise de barcodes, acaba tendo um custo comparável, e traz uma riqueza adicional incomparável de dados sobre os organismos sob análise, seja como espécies novas, variantes relevantes, ou SNPs que possam ser relacionados com dados bioquímicos, antigênicos, patogênicos, epidemiológicos ou taxonômicos.

Para a análise filogenética dos isolados, diversos parâmetros foram analisados. Primeiramente, optou-se por fazer cladogramas a partir dos genes *16S rRNA*, *hsp65* e *rpoB*, usando como grupo externo *Nocardia farcinica* (uma bactéria pertencente a outra família da ordem *Actinomycetales*, a mesma das micobactérias), com valores de bootstrap para 1.000 replicatas, parâmetros usados na maioria dos estudos taxonômicos micobacterianos. Em seguida, foi avaliado qual seria o melhor modelo de substituição de nucleotídeos a ser usado. Lembrando que toda análise de distância genética é uma questão de inferência estatística, escolher o melhor modelo de divergência (ou modelo de substituição de nucleotídeos) que explique as mutações nos genes analisados é fundamental (Posada, 2009). A maioria dos estudos taxonômicos utiliza “Kimura dois parâmetros” (K2P; Kimura, 1980), que,



apesar de considerar probabilidades diferentes para transições e transversões (dois parâmetros), ainda é um modelo muito simplificado. Então, foi escolhido com base no jModelTest, o modelo GTR (Tavaré, 1986), que leva em conta uma frequência diferencial entre as bases e diferentes probabilidades de substituição para cada nucleotídeo, sendo um modelo mais robusto e preciso. Para a escolha do modelo estatístico para a construção das árvores, testamos todas as opções disponíveis no MEGA. Esses modelos podem ser divididos em três grupos: métodos baseados em parcimônia (Máxima Parcimônia), distância (*Neighbor-joining*, Evolução mínima e UPGMA) ou probabilidade (Máxima verossimilhança). Comumente utiliza-se o *Neighbor-Joining* (Saitou & Nei, 1987), método baseado no modelo de evolução mínima que identifica os vizinhos que sequencialmente minimizam o tamanho da árvore. É um método simples, eficiente e muito rápido, podendo ser aplicado a um extenso volume de dados; no entanto, não considera os dados originais, apenas as distâncias. Aqui foi utilizado o método de Máxima Verossimilhança, que tenta reconstruir a filogenia a partir de um modelo explícito de evolução que pode ser testado, procurando obter o modelo mais provável de gerar os dados obtidos a partir do cálculo de várias verossimilhanças. Esse modelo utiliza todos os sítios, considera o modelo evolutivo específico, estimando de maneira acurada o tamanho dos ramos, sendo considerado o de maior consistência e robustez; porém, não é simples nem intuitivo, e é altamente dependente do modelo e exige um alto esforço computacional (Aldrich, 1997).

Confirmando os dados preliminares (Emmerick, 2013), 226 e 293 pertencem ao mesmo grupo filogenético, estando próximos de *M. aubagnense*, *M. Ilatzerense*, *M. mucogenicum* e *M. phocaicum*. Dados de análise de sequência *multilocus* do *core genome* (cgMLSA) (Morgado, 2017) confirmam essa proximidade e sugerem que os isolados da CBMA pertencentes ao cluster 1 (226 e 293 fazem parte) sejam divididos em dois grupos: 226 e outros. Embora os dados indiquem que estes isolados micobacterianos sejam espécies diferentes, é possível também que sejam subespécies, sendo necessários estudos específicos para esclarecer este ponto. *M. aubagnense* e *M. phocaicum* foram inicialmente identificados em pacientes com pneumonia crônica (Adékambi *et al.*, 2006), *M. mucogenicum* é comumente encontrado em casos clínicos e em surtos resultantes de infecções de instrumentos e das fontes de água de hospitais (Adékambi, 2009). *M. Ilatzerense* foi identificada

em diversas amostras de água de torneira e é, junto com *M. chelonae*, a espécie mais prevalente no sistema de distribuição de água de Paris (França) (Dubrou *et al.*, 2013), sendo também identificada em diversas infecções, inclusive do sistema nervoso central (Greninger *et al.*, 2015).

O isolado 271 mostrou-se presente no mesmo ramo filogenético que *M. abscessus*, *M. chelonae*, *M. franklinii*, *M. immunogenum*, *M. salmoniphilum* e *M. saopaulense* em todas as árvores geradas neste estudo (*M. salmoniphilum* não aparece na árvore concatenada nem na do gene *rpoB* porque não havia sequência disponível no GenBank para este gene), indicando sua associação a essas espécies. O mesmo agrupamento foi observado no trabalho de Morgado (2017) para o isolado 326. Esses dados, associados aos testes bioquímicos (Emmerick, 2013) e às demais análises realizadas, sugerem fortemente que seja uma nova espécie micobacteriana. Uma análise comparativa mais detalhada entre as espécies deste grupo através de ensaios *in silico* de hibridização DNA-DNA, outros testes bioquímicos e análise do perfil de ácidos micólicos, seriam capazes de confirmar esta hipótese. Este grupo já foi descrito em diversos estudos como patógenos oportunistas em humanos, desencadeando infecções na pele, pulmões, linfonodos, olhos e até no sistema esquelético (Simmon *et al.*, 2011; Lage *et al.*, 2015; Nogueira *et al.*, 2015); com destaque para *M. abscessus*, que é responsável por 90 % dos casos de doenças respiratórias causadas por micobactérias de crescimento rápido (Brown-Elliott *et al.*, 2012). Uma exceção é *M. salmoniphilum*, que é um patógeno característico de peixes (Whipps *et al.*, 2007), mas que já foi descrito como capaz de crescer em macrófagos humanos (Harriff *et al.*, 2008).

O isolado 294 foi identificado em um grupo que abriga diversas espécies. A sequência do gene *16S rRNA* não é suficiente para discriminar as espécies desse grupo dividindo-o em quatro subgrupos, onde 294 apresenta a mesma sequência que *M. septicum*, *M. lutetiense* e *M. setense* e próxima a *M. fortuitum*. No entanto, os demais genes e, conseqüentemente, a árvore concatenada são capazes de distinguir os membros deste grupo – destacando a importância do uso de vários marcadores filogenéticos – separando 294 dos demais membros deste grupo. Este mesmo agrupamento foi observado no trabalho de Morgado (2017) para o isolado 247. *M. septicum* foi isolado originalmente de culturas de sangue de um menino de dois anos com hepatoblastoma metastático na Austrália (Hogg *et al.*, 2000). *M.*

*setense* foi isolada de pacientes com infecção nos ossos e tecidos conjuntivos na França (Lammy *et al.*, 2008; Toro *et al.*, 2008). *M. lutetiense* foi isolada do sistema de distribuição de água de Paris (França) (Konjek *et al.*, 2016). *M. fortuitum* é tido como o responsável por 60 – 80 % dos casos de infecções pós-cirúrgicas causadas por micobactérias (Brown-Elliott *et al.*, 2012).

As árvores geradas neste trabalho parecem concordar com os agrupamentos propostos na literatura. O fato dos isolados da CBMA geralmente agruparem próximos a patógenos oportunistas, pode estar associado ao fato de que grande parte das micobactérias identificadas seja resultante de amostras clínicas responsáveis pelos mais diversos tipos de infecções. Porém, pode realmente indicar certo grau de patogenicidade associados a esses isolados. Novos estudos nessa área são necessários para elucidar essa questão.

Gupta e colaboradores (2018), após análise filogenética e genômica e identificação de INDELS e proteínas com assinaturas conservadas, observaram cinco distintos grupos monofiléticos dentro do gênero *Mycobacterium*. As diferenças entre esses grupos são tão fortes e consistentes que, segundo os autores, este gênero deveria ser dividido em cinco: 1) *Mycobacterium*, espécies de crescimento lento contendo o clado “*Tuberculosis-Simiae*” com as principais espécies patogênicas em humanos e animais e tendo *Mycobacterium tuberculosis* como a espécie tipo; 2) *Mycolicibacterium*, 90 espécies de crescimento rápido contendo o clado “*Fortuitum-Vaccae*” e *Mycolicibacterium fortuitum* sendo a espécie tipo; 3) *Mycolicibacter*, 13 espécies de crescimento lento contendo o clado “*Terrae*” e *Mycolicibacter terrae* como a espécie tipo; 4) *Mycolicibacillus*, três espécies de crescimento lento contendo o clado “*Triviale*” e *Mycolicibacillus trivialis* como a espécie tipo e 5) *Mycobacteroides*, nove espécies de crescimento rápido contendo o clado “*Abscessus-Cheloniae*” e *Mycobacteroides abscessus* como a espécie tipo. *Mycolicibacterium*, *Mycolicibacter* e *Mycolicibacillus* significa gênero de bacilos que contém ácido micólicos e *Mycobacteroides*, gênero semelhante a *Mycobacterium*. Segunda esta nova classificação da família *Mycobacteriaceae*, os isolados 226, 293 e 294 provavelmente pertenceriam ao gênero *Mycolicibacterium* e o isolado 271, ao gênero *Mycobacteroides*; no entanto, são necessários estudos das assinaturas de INDELS e proteínas específicas nesses isolados para comprovar essa classificação.

## 5.2. Análise do secretoma

Todas as bactérias, incluindo as micobactérias, possuem sistemas de exportação para transportar proteínas específicas, sintetizadas no citoplasma, para fora da membrana citoplasmática. Essas proteínas podem permanecer no envelope celular ou ainda serem secretadas para além da parede celular bacteriana. Essas vias de exportação de proteínas estão geralmente envolvidas em processos fisiológicos essenciais. Bactérias patogênicas costumam depender dos sistemas de exportação de proteínas especializados, essenciais na interação com o hospedeiro (Felther *et al.*, 2010; Cléon *et al.*, 2015).

Micobactérias apresentam um envelope celular com uma estrutura única que protege a bactéria do ambiente extracelular, limitando o acesso a moléculas extracelulares nocivas. Essa barreira espessa e extremamente hidrofóbica gera um problema para a exportação de produtos pelas micobactérias (Champion & Cox, 2007; Jankute *et al.*, 2015). Para contornar essa situação, as micobactérias, além de possuírem as vias de exportação de proteínas clássicas – a via de secreção geral (Sec) e a via de translocação twin-arginina (Tat) – que são essenciais e realizam a maior parte deste trabalho, possuem também diversos outros sistemas, como a via SecA2 e o sistema de secreção tipo VII (ESX), que são importantes para diversos processos como virulência, aquisição de ferro ou mesmo para alguns aspectos de crescimento *in vivo* ou *in vitro*. Esse conjunto de sistemas confere às micobactérias uma ótima capacidade secretora (Feltcher *et al.*, 2010; Houben *et al.*, 2013). O perfil complexo de proteínas secretadas pelos isolados em estudo comprova a capacidade secretora característica destas micobactérias, a qual foi analisada através da proteômica baseada em 2DE e *shotgun*.

Proteômica baseada em eletroforese bidimensional seguido de visualização dos *spots*, excisão, digestão com tripsina, e análise proteômica através de MALDI TOF-TOF e comparação de perfis de peptídeos (massa/carga) com um banco de dados através de software como Mascott apresenta diversas vantagens quando comparada à abordagem *shotgun*: menor tempo de máquina, uso de espectrômetros mais simples e baratos, maior reprodutibilidade e robustez, análises quantitativas mais simples, amplamente usado em amostras bacterianas. Suas principais limitações são as dificuldades no estudo de proteínas de membrana (devido à baixa

solubilidade das mesmas) e de amostras muito complexas, além de ser uma técnica bastante trabalhosa, e com baixa identificação de proteínas com ponto isoelétrico e massa molecular extremas (Rabillouda & Lelong, 2011). Embora não seja mais o padrão ouro no estudo proteômico, 2DE continua sendo uma técnica útil no fracionamento e identificação de amostras proteicas (Sandim *et al.*, 2016; González-López *et al.*, 2017; Gaspar *et al.*, 2017). Inicialmente, foi analisado o secretoma a partir dessa abordagem, visto que o perfil de amostras não era tão complexo e o interesse principal era a identificação de proteínas majoritariamente expressas, que poderiam servir de base para a construção de novos vetores para expressão e secreção de proteínas recombinantes.

A análise de misturas mais complexas através de espectrometria de massas não é um processo simples, sendo difícil de analisar em equipamentos “comuns”, visto que milhares de peptídeos eluem em intervalos muito curtos de tempo e com enormes diferenças na magnitude do sinal. Então, espectrômetros cada vez mais sensíveis, rápidos e com ótima resolução são necessários para uma abordagem *shotgun*, estando disponíveis principalmente quatro tipos de analisadores: *Quadrupole*, *Time of Flight*, *Linear Ion Trap* e *Orbitrap*. A combinação do quadrupolo, como filtro de massas, com o analisador Orbitrap resultou em um equipamento de alta eficiência e com pouca necessidade de manipulação, o QExactive. Este tem mostrado resultados com elevada sensibilidade, rapidez e resolução comparado ao QQQ, Q-TOF ou LTQ Orbitrap (Michalski *et al.*, 2011; Willianson *et al.*, 2016). Diante da possibilidade de realizar uma análise *shotgun*, as mesmas amostras utilizadas na 2DE foram encaminhadas para análise no QExactive, resultando em um número muito maior de proteínas identificadas, abrindo diversas perspectivas para novas análises.

Apesar de algumas proteínas terem sido identificadas exclusivamente pela abordagem baseada em 2DE (MALDI ou Orbitrap), o número e a qualidade das identificações realizadas pela abordagem *shotgun* foi muito maior.

O sistema ESX-1 é responsável pela secreção de diversas proteínas, entre elas duas pequenas proteínas que estão entre os componentes principais do filtrado de cultura de *M. tuberculosis*: EsxA e EsxB (ou ESAT-6 e CFP-10) (Stanley *et al.*, 2003). ESAT-6 tem sido descrita com uma proteína antigênica que estimula fortemente os linfócitos T e, portanto, tem sido o foco de diversos estudos vacinais contra a tuberculose (Samten *et al.*, 2009). Essa proteína já foi identificada em

micobactérias ambientais como *M. marinum* e *M. simiae*, isolados originalmente de peixes e símios, respectivamente. Posteriormente verificou-se que essas espécies são patógenos oportunistas, geralmente em pessoas imunocomprometidas (Hashishi *et al.*, 2018; Steffani-Vallejo *et al.*, 2018). Genes codificantes para proteínas com similaridade à ESAT-6 de *M. tuberculosis* foram identificadas nos quatro isolados da CBMA e a própria ESAT-6 foi identificada nos isolados 293 (duas cópias) e 294. Através do secretoma, foi possível identificar apenas uma proteína tipo ESAT-6 e a própria ESAT-6 de 294. A presença dessas proteínas pode indicar que não apenas micobactérias patogênicas apresentam essas vias, ou podem sugerir um possível grau de patogenicidade para esses isolados.

Uma das grandes dificuldades no estudo do secretoma é a contaminação das proteínas secretadas com proteínas intracelulares, devido a algum grau de lise celular durante o crescimento ou preparação das amostras, com a liberação de conteúdo intracelular no meio de cultura. Para evitar esse processo, costuma-se coletar o sobrenadante ainda na fase exponencial de crescimento; no entanto, existe uma taxa de morte – ainda que pequena – durante todo o crescimento bacteriano. São, então, utilizados marcadores de lise de modo a mensurar a quantidade de células mortas durante a coleta da amostra. Esse teste é geralmente utilizado para avaliar o efeito de drogas bactericidas, de modo a verificar se a morte foi, de fato, resultante da droga. Um marcador que tem sido usado no estudo micobacteriano é adenilato quinase (EC 2.7.4.3), uma enzima intracelular ubíqua que é liberada para o meio extracelular durante a lise celular. Esta proteína é uma fosfotransferase que catalisa a interconversão de nucleotídeos adenina, e desempenha um papel importante na homeostase de energia celular (Jacobs *et al.*, 2012; Forbes *et al.*, 2015). Este marcador foi anotado nos quatro genomas (com duas cópias em 226, 293 e 294) e encontrado tanto no secretoma de 226 quanto de 294, indicando a presença de proteínas intracelulares.

Na busca por enzimas biotecnologicamente interessantes, foram identificadas diversas cutinases, tanto na anotação do genoma como nas três abordagens proteômicas. Na anotação do genoma foram identificadas 15 CDS que codificam para cutinases no isolado 226, 7 em 271, 11 em 293 e 15 em 294. Pelo MALDI, estas foram identificadas em 271 (1) e em 294 (1); pelo Orbitrap, em 271 (2) e 294 (5); e pelo QExactive, em 226 (3), 293 (1) e 294 (1). Cutinases (EC 3.1.1.74) são

enzimas descobertas originalmente em fungos fitopatogênicos que utilizam cutina como única fonte de carbono. Cutina é um biopolímero complexo composto de ácidos graxos que forma a cutícula das plantas superiores. Estas enzimas compartilham propriedades catalíticas com lipases e esterases, mas tem como particularidade a capacidade de ser ativa independente da presença de uma interface água-óleo, tornando-as atrativas como biocatalisadores em diversos processos industriais envolvendo reações de hidrólise, esterificação e transesterificação. Além disso, apresentam uma alta estabilidade em solventes orgânicos e líquidos iônicos, permitindo sua aplicação em áreas como a indústria alimentícia, cosméticos, química fina, degradação de pesticidas e inseticidas, tratamento de resíduos da indústria têxtil (Pio & Macedo, 2009). Essa característica não é apenas teórica, cutinases já têm sido usadas na degradação de poliésteres alifáticos e aromáticos (Gamerith *et al.*, 2017), polibutileno succinato (Pan *et al.*, 2018) e outros, gerando inclusive matéria-prima para etanol de segunda geração (Duan *et al.*, 2017).

O elevado número de proteínas sem função conhecida identificadas pelo secretoma evidencia o quanto ainda há para se estudar das micobactérias e todo um potencial biotecnológico ainda desconhecido. O número de proteínas sem função conhecida para as quais obteve-se uma anotação após uma busca específica, provavelmente é resultado de uma atualização dos bancos de dados entre o período da anotação do genoma e dessa busca específica. Essa quantidade de proteínas hipotéticas não é particular aos isolados da CBMA, sendo também característica de todo o gênero (Prasanna & Mehra, 2013).

A quantificação *label-free* por contagem de espectros, embora não permita uma quantificação absoluta, é simples de ser realizada. Essa metodologia tem um viés para as proteínas mais abundantes, sendo algo positivo neste estudo, visto que o principal objetivo era a identificação de proteínas majoritárias do secretoma para o desenvolvimento de um novo sistema de expressão heteróloga.

### **5.3. Construção de um novo sistema de expressão**

Pouco mais de uma década após o desenvolvimento da tecnologia do DNA recombinante por Berg, Cohen e Boyer (Jackson *et al.*, 1972, Cohen *et al.*, 1973), as micobactérias já estavam sendo usadas como sistemas de expressão de proteínas

(Jacobs *et al.*, 1987). Embora apresentem algumas desvantagens quando comparadas com o hospedeiro mais utilizado (*E. coli*), tais como maior tempo de duplicação e parede celular rica em lipídeos, as micobactérias destacam-se pela sua capacidade de expressar proteínas e secretá-las e por realizar algumas modificações pós-traducionais, obtendo proteínas recombinantes mais parecidas com a nativa (Bashiri & Baker, 2014). As desvantagens apresentadas pelas micobactérias, como maior tempo de crescimento, meios de cultura específicos e patogenicidade associada, podem ser atenuadas adaptando protocolos de crescimento (Singh & Reyrat, 2009) ou selecionando espécies/cepas mais capacitadas para este fim (Snapper *et al.*, 1990; Medeiros *et al.*, 2012).

As micobactérias são o quinto hospedeiro bacteriano mais utilizado para expressão de proteínas para análises em biologia estrutural. Isso devido a apenas duas espécies, *M. smegmatis* e *M. vaccae*, que têm sido exploradas basicamente para a expressão de proteínas de *M. tuberculosis* (Fernández & Vega, 2013). O uso de micobactérias para expressar proteínas de *M. tuberculosis* é compreensível, visto que a tuberculose permanece sendo um grave problema de saúde global (WHO, 2016) e expressar proteínas micobacterianas em *E. coli* de maneira solúvel oferece bastante dificuldades (Fernández & Vega, 2013). Embora diversos sistemas de expressão tenham sido desenvolvidos nos últimos 40 anos, diversas proteínas ainda não foram expressas com sucesso devido a limitações nesses sistemas (Bashiri & Baker, 2014).

Um sistema de expressão está baseado em quatro elementos principais: 1) um hospedeiro de crescimento rápido, seguro e bem conhecido para a clonagem, expressão, eventual secreção e eventual modificação translacional; 2) facilidade de transformação desse hospedeiro e padronização das condições de crescimento em meio simples e barato; 3) vetor de expressão com os elementos genéticos necessários para expressão episomal ou integrativo, marcador de seleção, promotor adequado e controlável e demais elementos permitindo estabilidade e versatilidade; e 4) padronização das condições de clonagem, seleção, crescimento e expressão. Com o objetivo de desenvolver um novo sistema de expressão que atenda aos requisitos mencionados, e possa aproveitar as vantagens específicas de micobactérias como sistema, diferentes isolados micobacterianos foram estudados previamente, tendo sido padronizados suas condições de crescimento e de



transformação (Emmerick, 2013). Para a construção de um novo vetor, foi utilizado como base o vetor pUS972 (Medeiros *et al.*, 2002): vetor-ponte pequeno que possui origem de replicação em *E. coli* (oriE) e em micobactérias (oriM), possui o gene *aph* que codifica resistência ao antibiótico canamicina (kanR) e apresenta um sítio múltiplo de clonagem (MCS) para 10 diferentes enzimas de restrição. Apresenta um baixo número de cópias, mas uma boa eficiência de transformação.

Atualmente, diversas tecnologias de clonagem estão disponíveis. A clonagem tradicional (Jackson *et al.*, 1972, Cohen *et al.*, 1973), através da digestão com enzimas de restrição e ligação com a T4 DNA ligase, apresenta diversas desvantagens: técnica não padronizada, com cicatriz, dependente da sequência e permite a clonagem de apenas um fragmento por vez. Ainda pode haver problemas com o vetor vazio ou ligação na direção oposta. A clonagem por PCR, ou clonagem TA, não precisa da adição de sítios de restrição ao fragmento, mas de uma deoxiadenosina (A) a extremidade 3' que pareia com a timina (T) no vetor. Embora apresente uma elevada eficiência e seja uma metodologia padronizada e rápida, não apresenta muitos vetores disponíveis, o custo é alto, não permite a clonagem de múltiplos fragmentos e a ligação pode ocorrer na direção oposta. Uma variação desta técnica, TOPO TA<sup>®</sup> (Life Technologies), utiliza a topoisomerase para unir o inserto ao vetor ao invés da DNA ligase (<https://www.neb.com/applications/cloning-and-synthetic-biology/pcr-cloning>).

Gateway<sup>®</sup> (Life Technologies) foi uma metodologia desenvolvida para transferir sequências de DNA para múltiplos vetores, baseada na capacidade de recombinação do fago lambda, permitindo uma rápida e eficiente análise funcional e expressão de proteínas. Este método mantém a orientação correta do inserto, porém necessita de vetores específicos e enzimas muito caras (Hartley *et al.*, 2000). A chamada clonagem Electra<sup>™</sup> (DNA 2.0), baseada na ação da enzima de restrição *SapI*, permite que um fragmento de um vetor seja transferido para outro de maneira rápida sem que haja qualquer tipo de amplificação por PCRs, não gerando cicatriz. No entanto, depende de vetores específicos. Gibson<sup>®</sup> (New England Biolabs; Gibson *et al.*, 2009), In-Fusion<sup>®</sup> (Clontech), SLIC (*Sequence and Ligation-Independent Cloning*; Li & Elledge, 2007), CPEC (*Circular Polymerase Extension Cloning*; Quan & Tian, 2009), SLiCE (*Seamless Ligation Cloning Extract*; Zang *et al.*, 2012) e GeneArt<sup>®</sup> (Life Technologies), são técnicas de clonagem que iniciam com o vetor

linearizado e o inserto amplificado por PCR, e realizam a clonagem sem cicatriz. O princípio de todas estas técnicas é a complementariedade entre a extremidade dos fragmentos, permitindo que mais de um fragmento seja ligado rapidamente e na direção correta. A limitação é a sequência dessas extremidades, que não devem conter sequências repetitivas, hairpins ou loops, nem serem complementares a mais de um fragmento.

Para a construção do vetor foi escolhida a clonagem de Gibson, visto que proporciona a clonagem de três fragmentos mais o vetor de maneira rápida e sem cicatriz, o que tem sido usado satisfatoriamente há quase 10 anos (Gibson *et al.*, 2009; Ramon & Smith, 2011). Durante a clonagem, três enzimas atuam: exonuclease 5' (que remove parte dos nucleotídeos da extremidade 5' dos fragmentos), DNA polimerase (estende a parte da extremidade 5' degradada até onde as fitas se complementaram) e DNA ligase (realiza a ligação covalente entre os fragmentos). Durante a primeira tentativa da construção utilizou-se os iniciadores desenhados automaticamente pelo NEBuilder. Após o insucesso dessa construção, foi realizada uma nova abordagem, sendo desenhados cuidadosamente novos iniciadores, verificando o tamanho adequado, a formação de hairpins e dímeros, elevado conteúdo GC, sequências repetitivas e temperatura de melting. Ainda assim, a tentativa de montagem simultânea de três fragmentos utilizando o sistema de Gibson não resultou na construção do novo vetor conforme esperado. Assim, ainda não foi possível avaliar os sinais regulatórios da proteína 294\_01260, nem os comparar com outros já descritos. Também não foi possível comparar este sistema com sistemas já estabelecidos em *M. smegmatis* e *M. vaccae*.

#### **5.4. Sobrevivência intracelular dos isolados ambientais**

Diante de alguns indícios de possível patogenicidade por parte dos isolados da CBMA, realizou-se um ensaio para avaliar sua capacidade de sobrevivência no ambiente intracelular de macrófagos (dTHP-1). É descrito que *M. tuberculosis*, após a inalação, é fagocitada por macrófagos residentes no pulmão e sobrevive aos diversos mecanismos microbicidas desencadeados; dentre eles mencionamos a produção de espécies reativas de oxigênio (ROS) e intermediários reativos de nitrogênio (RNS), a acidificação fagossomal e a secreção de citocinas e quimiocinas

(Russel, 2001; Awuh & Flo, 2017). Para o estudo da interação entre micobactérias e macrófagos, vários sistemas vêm sendo empregados como modelo, de modo a simular a associação de *M. tuberculosis* com macrófagos humanos. Dentre os modelos usados, cabe destacar (1) os isolados de células primárias, como macrófagos peritoneais de murino (Stokes *et al.*, 1993), macrófagos alveolares de murino (Schlesinger *et al.*, 1990) e macrófagos alveolares humanos (Aston *et al.*, 1998); (2) as células diferenciadas *in vitro*, tais como macrófagos de medula óssea de murino (Xu *et al.*, 1994) e macrófagos derivados de monócitos (MDM) humanos (Silver *et al.*, 1998); e (3) linhagens celulares semelhantes a macrófagos de murino (Mehta *et al.*, 1996) e humanos (Kremlev & Phelps, 1997). Esses diferentes modelos têm propriedades específicas e interagem com as micobactérias de forma diferente. Estudos que avaliam a função de monócitos/macrófagos são restritivos, uma vez que estas células não podem ser expandidas *ex vivo*. Linhagens de células monocíticas apresentam poucos receptores e são menos eficientes em fagocitar patógenos. No entanto, macrófagos diferenciados exibem um grande número de receptores, com capacidade fagocítica aumentada. Além disso, macrófagos de linhagem constituem uma população homogênea, conferindo uma vantagem em relação aos monócitos humanos nativos (Auwerx, 1991; Daigneault *et al.*, 2010). Apesar do modelo com células de linhagem ser considerado um modelo de interface simplificada da relação patógeno-hospedeiro, este vem sendo usado para definir interações bacterianas com tipos celulares específicos, fornecendo uma visão geral da progressão da doença bacteriana (Waddell *et al.*, 2007).

A partir da purificação de monócitos do sangue periférico de uma criança com leucemia monocítica aguda obteve-se a linhagem celular leucêmica monocítica humana, THP-1. Esses monócitos adquirem características de macrófagos mediante a utilização de citocinas e agentes farmacológicos, como acetato de forbol-miristila (PMA) e 1,25-dihidroxitamina d3 para promover sua diferenciação. O PMA atua através da ativação da proteína quinase C, induzindo a diferenciação dos monócitos THP-1 em macrófagos. Esse processo de diferenciação acarreta na diminuição da replicação de DNA, adesão ao substrato (plástico) e morfologia similar a macrófagos primários (Schwende *et al.*, 1996).

O modelo de células THP-1 é considerado um bom modelo para substituir MDM humano quando usado para investigar a relação de micobactéria/macrófago,

apresentando como vantagem a menor variabilidade entre doadores comumente encontrados em MDM e a quantidade de células que podem ser facilmente cultivadas e diferenciadas para estudos bioquímicos ou de expressão gênica. Além disso, a linhagem celular tende a ser mais favorável à manutenção a longo prazo após a infecção com *M. tuberculosis* (Stokes & Doxsee, 1999). Essa linhagem já foi empregada para análise do crescimento e persistência bacteriana (Theus *et al.*, 2004), avaliação do perfil de expressão gênica de *M. bovis* BCG no contexto intracelular macrofágico (Rienksma *et al.*, 2015), do perfil de expressão proteica de isolados clínicos de *M. tuberculosis* (Singhal *et al.*, 2012) e sobrevivência intracelular de *M. smegmatis* (Ganaie *et al.*, 2016).

O protocolo para diferenciação dos macrófagos a partir de monócitos de linhagem THP-1 usando PMA é bastante diverso na literatura, com uso de concentrações variando entre 30 nM (Sousa-Vasconcelos *et al.*, 2015), 80 nM (Guerreiro *et al.*, 2013), 100 nM (Mendoza-Coronel & Castañón-Arreola, 2016) e 200 nM (Daigneault *et al.*, 2010). Além da concentração de PMA, o tempo de exposição e de repouso pós-diferenciação também é bastante diverso. A falta de um protocolo uniforme impacta fortemente na interpretação dos resultados e na comparação entre os estudos. Corrêa (2018) realizou a padronização da diferenciação de THP-1 através do teste de vários protocolos de diferenciação das células THP-1 usando diferentes concentrações de PMA (30 nM, 80 nM e 200 nM) com e sem tempo de descanso sem a droga após o estímulo. Ao final desses experimentos que analisaram parâmetros como morfologia e viabilidade celulares, optamos por estimular as células com 200 nM de PMA por 48 horas e as células diferenciadas foram mantidas em meio sem PMA pelo mesmo tempo para assegurar que as células se recuperem do estímulo antes da infecção.

Os isolados da CBMA apresentam uma baixa taxa de internalização e um declínio na viabilidade intracelular ao longo do tempo analisado, com exceção para o isolado 226, que mantém estável o número de bactérias no interior do macrófago. Essa é a característica esperada para espécies saprófitas. No entanto, mc<sup>2</sup>155, espécie caracteristicamente saprófita, aumentou exponencialmente ao longo de todo o período avaliado. Esses dados diferem do observado por Ganaie e colaboradores (2011), que analisaram a influência do gene *eis* (*enhance the intracellular survival*) de *M. tuberculosis* na sobrevivência intracelular de *M. smegmatis* em macrófagos

---

THP-1, verificando que ausência desse gene leva a uma redução na viabilidade intracelular de *M. smegmatis* e a presença de eis mantém os níveis de bactéria no interior do macrófago. Em estudo similar do mesmo grupo, foi avaliada a influência da interação da proteína Erp com Rv2212 (ambos de *M. tuberculosis*), resultando na redução da viabilidade na ausência desses genes e na redução em menor escala na presença desses genes (Ganaie *et al.*, 2016).

Apesar das limitações intrínsecas das análises deste estudo terem sido realizada com células de linhagem, o uso de células THP-1 tem vantagens como a facilidade de expansão e diferenciação em macrófagos. Esses dados fornecem subsídios iniciais para elucidar uma possível patogenicidade associada a esses isolados da CBMA.

## 6. Conclusões

- ✓ Os isolados micobacterianos da CBMA parecem agrupar em três grupos muito distintos.
- ✓ O sequenciamento genômico pelas tecnologias 454 GS Jr. (Roche) e HiSeq 2500 (Illumina), separadamente, não foi suficiente para montar o genoma dos isolados micobacterianos da CBMA em um único *contig*.
- ✓ Apesar das sequências do 454 serem maiores, os dados do HiSeq são mais uniformes e com melhor qualidade.
- ✓ Os isolados 226, 293 e 294 apresentam genoma com mais de 6 Mb e 271, com menos de 5 Mb.
- ✓ Os isolados micobacterianos ambientais apresentam diversos antígenos caracteristicamente imunogênicos expressos, tais como: Mpt51, Mpt53, Mpt63, Mpt64, Mpt70, GlnA1, Ag85 e MTB12.
- ✓ Os genes *16S rRNA*, *hsp65* e *rpoB* são capazes de diferenciar as diversas espécies de micobactérias.
- ✓ Os isolados 226 e 293 pertencem ao mesmo grupo filogenético, estando próximos de *M. aubagnense*, *M. llatzerense*, *M. mucogenicum* e *M. phocaicum*. O isolado 271 mostrou-se presente no mesmo ramo filogenético que *M. abscessus*, *M. chelonae*, *M. franklinii*, *M. immunogenum*, *M. salmoniphilum* e *M. saopaulense*. O isolado 294 está próximo de *M. septicum*, *M. lutetiense*, *M. setense* e *M. fortuitum*.
- ✓ As proteínas secretadas dos isolados micobacterianos encontram-se majoritariamente na faixa de pH ácido (pH 3-6), e há maior abundância de proteínas de baixo peso molecular.
- ✓ A abordagem proteômica shotgun permite um maior número de proteínas identificadas que a abordagem via 2DE.
- ✓ Os isolados micobacterianos são transformáveis por eletroporação.

- ✓ Não foi possível realizar a construção de novos vetores utilizando as abordagens descritas neste estudo.
- ✓ Os isolados da CBMA apresentam baixa taxa de internalização por macrófagos THP-1 em relação ao descrito para outras micobactérias.
- ✓ Os isolados da CBMA sobrevivem em macrófagos, nas condições estudadas.
- ✓ Os isolados 271, 293 e 294 apresentam diminuição na quantidade de bacilos viáveis durante a infecção em células THP-1, nos tempos estudados, ao contrário de *M. smegmatis* mc<sup>2</sup>155 cujos bacilos viáveis aumentaram exponencialmente.

## 7. Bibliografia

- Abou-Zeid C, Gares MP, Inwald J, Janssen R, Zhang Y, Young DB, Hetzel C, Lamb JR, Baldwin SL, Orme IM, Yeremeev V, Nikonenko BV, Apt AS. Induction of a type-1 immune response to a recombinant antigen from *Mycobacterium tuberculosis* expressed in *Mycobacterium vaccae*. *Infect Immun* 1997, 65:1856-1862;
- Abuhammad A. Cholesterol metabolism: a potential therapeutic target in mycobacteria. *Br. J. Pharmacol* 2017; 174(14):2194-2208.
- Adékambi T, Berger P, Raoult D, Drancourt M. *rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. *Int J Syst Evol Microbiol* 2006; 56:133-43.
- Adékambi T, Berger P, Raoult D, Drancourt M. *rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. *Int J Syst Evol Microbiol* 2006A; 56:133–143.
- Adékambi T, Colson P, Drancourt M. *rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. *J Clin Microbiol* 2003; 41:5699-5708.
- Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, Raoult D, Drancourt M. Amoebal coculture of “*Mycobacterium massiliense*” sp. nov. from the sputum of a patient with hemoptoic pneumonia. *J Clin Microbiol* 2004; 42:5493–5501.
- Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, Raoult D, Drancourt M. *Mycobacterium massiliense* sp. nov. In *List of New Names and New Combinations Previously Effectively, but not Validly, Published*, Validation List no. 111. *Int J Syst Evol Microbiol* 2006B; 56:2025–2027.



- Adekambi T, Sassi M, van Ingen J, Drancourt M. Reinstating *Mycobacterium massiliense* and *Mycobacterium bolletii* as species of the *Mycobacterium abscessus* complex. *Int J Syst Evol Microbiol* 2017; 67:2726–2730.
- Adékambi T. *Mycobacterium mucogenicum* group infections: a review. *Clin Microbiol Infect* 2009; 15(10):911-8. Review.
- Akhtara *et al.* Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *PNAS* 2013; 110:87–92.
- Al-Attayah R, Shaban FA, Wiker HG, Oftung F, Mustafa AS. Synthetic peptides identify promiscuous human Th1 cell epitopes of the secreted mycobacterial antigen MPB70. *Infect Immun.* 2003 Apr;71(4):1953-60;
- Aldrich J. R. A. Fisher and the making of maximum likelihood 1912–1922. *Statistical Science* 1997; 12(3):162–176.
- Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, Gey van Pittius NC. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis* 2010; 16(8):1296-9.
- Al-Saeedi M & Al-Hajoj S. Diversity and evolution of drug resistance mechanisms in *Mycobacterium tuberculosis*. *Infect Drug Resist.* 2017; 13;10:333-342.
- Alvarez E, Tavel E. Recherches sur le bacille de Lustgarten. *Arch Physiol Norm Pathol* 1885, 6:303-321;
- Arya R, Bhattacharya A, Saini KS. *Dictyostelium discoideum* – a promising expression system for the production of eukaryotic proteins. *FASEB J.* 2008 Dec;22(12):4055-66. Review;
- Aston C, Rom WN, Talbot AT, Reibman J. Early inhibition of mycobacterial growth by human alveolar macrophages is not due to nitric oxide. *Am J Respir Crit Care Med* 1998; 157(6 Pt 1):1943-50.
- Auwerx, J. The human leukemia cell line, THP-1: A multifaceted model for the study of monocyte-macrophage differentiation. *Experientia* 1991; 47(1):22-29.
- Awuh JA, Flo TH. Molecular basis of mycobacterial survival in macrophages. *Cell Mol Life Sci* 2017; 74(9):1625-1648 Review.

- Azadi P, Inderwildi OR, Farnood R, Kinga DA. Liquid fuels, hydrogen and chemicals from lignin: A critical review. *Renewable and Sustainable Energy Reviews* 2013; 21: 506-523
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008; 8(9):75.
- Badejo AC, Badejo AO, Shin KH, Chai YG. A Gene Expression Study of the Activities of Aromatic Ring-Cleavage Dioxygenases in *Mycobacterium gilvum* PYR-GCK to Changes in Salinity and pH during Pyrene Degradation. *PLoS ONE* 2013; 8(2): e58066.
- Baneyx F. Recombinant protein expression in *E. coli*. *Curr. Opin. Biotechnol.* 1999; 10: 411–421;
- Barbosa MRV & Thomas WW. Biodiversidade, conservação e uso sustentável da Mata Atlântica no Nordeste. In: Araújo EL, Moura NA, Sampaio ESB, Gestinari LMS, Carneiro JMT. Biodiversidade, conservação e uso sustentável da flora do Brasil. Recife, Imprensa Universitária, p.19-22. 2002;
- Barrera L. The basis of clinical bacteriology. In: Palomino JC, Leão SC, Ritacco V. Tuberculosis 2007. From basic science to patient care. Tuberculosis Textbook.com. 2007; p.93-112.
- Bashiri G & Baker EN. Production of recombinant proteins in *Mycobacterium smegmatis* for structural and functional studies. *Protein Science* 2015; 24:1-10.
- Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ, Besra GS. Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* 1997; 276:1420–2;
- Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect* 2017; S1198-743X(17)30575-X. Review.

- Bhaskar S, Khanna SP, Mukherjee R. Isolation, purification and immunological characterization of novel low molecular weight protein antigen CFP 6 from culture filtrate of *M. tuberculosis*. *Vaccine*. 2000. 15;18(25):2856-66.
- Birnboim HC. A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol*. 1983;100:243-55.
- Blokpoel MC, Murphy HN, O'Toole R, Wiles S, Runn ES, Stewart GR, Young DB, Robertson BD. Tetracycline-inducible gene regulation in mycobacteria. *Nucleic Acids Res* 2005; 33:e22.
- Brandon K, da Fonseca GAB, Rylands AB, da Silva JMC. Conservação brasileira: desafios e oportunidades. *Megadiversidade*. 2005 Jul; Vol 1, N° 1;
- Brandt L, Feino Cunha J, Weinreich Olsen A, Chilima B, Hirsch P, Appelberg R, Andersen P. Failure of the *Mycobacterium bovis* BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis. *Infect Immun* 2002; 70(2):672-8.
- Brondyk WH. Selecting an appropriate method for expressing a recombinant protein. *Methods Enzymol* 2009; 463:131–147.
- Brosch R, Gordon SV, Pym A, Eiglmeier K, Garnier T, Cole ST. Comparative genomics of the mycobacteria. *Int J Med Microbiol* 2000; 290:143–152.
- Broset E, Martín C, Gonzalo-Asensio J. Evolutionary landscape of the *Mycobacterium tuberculosis* complex from the viewpoint of PhoPR: Implications for virulence regulation and application to vaccine development. *MBio* 2015; 6(5):1-15.
- Brown AC & Parish T. Instability of the acetamide-inducible expression vector pJAM2 in *Mycobacterium tuberculosis*. *Plasmid* 2006; 55:81–86.
- Brown-Elliott BA, Nash KA, Wallace RJ Jr. Antimicrobial susceptibility testing, drug resistance mechanisms, and therapy of infections with nontuberculous mycobacteria. *Clin Microbiol Rev* 2012;25:545-82.
- Bruce T, Martinez IB, Neto OM, Vicente ACP, Kruger RH, Thompson FL. Bacterial community diversity in the brazilian Atlantic Forest soils. *Microb Ecol* 2010; 60:840–849.

- Champion PAD & Cox JS. Protein secretion systems in Mycobacteria. *Cellular Microbiology* 2007; 9(6):1376–1384.
- Checkley AM, Wyllie DH, Scriba TJ, Golubchik T, Hill AV, Hanekom WA, McShane H. Identification of antigens specific to non-tuberculous mycobacteria: the Mce family of proteins as a target of T cell immune responses. *PLoS One* 2011; 6(10):26434.
- Chen S, Larsson M, Robinson RC, Chen SL. Direct and convenient measurement of plasmid stability in lab and clinical isolates of *E. coli*. *Sci Rep* 2017; 7(1):4788.
- Chevreur *et al.* Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. *Genome Research* 2004, 14:1147-1159.
- Chiaradia L, Lefebvre C, Parra J, Marcoux J, Burlet-Schiltz O, Etienne G, Tropis M, Daffé M. *Sci Rep.* 2017; 7(1):12807.
- Clarke J, Wu H, Jayasinghe L, Patel A. Continuous base identification for single molecule nanopore DNA sequencing. *Nat Nanotechnol* 2009; 4:265–270.
- Cléon F, Habersetzer J, Alcock F, Kneuper H, Stansfeld PJ, Basit H, Wallace MI, Berks BC, Palmer T. The TatC component of the twin-arginine protein translocase functions as an obligate oligomer. *Molecular Microbiology* 2015; 98(1):111–129.
- Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci USA* 1973; 70(11):3240-4.
- Cole *et al.* Massive gene decay in the leprosy bacillus. *Nature* 2001; 409:1007.
- Cole St, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S, Eiglmeier K, Gas S, Barry Ce 3rd, 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, Rajandream M, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston J, Taylor K, Whitehead S, Barrell B. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998; 393(6685):537-44.

- Connell ND. Expression systems for use in actinomycetes and related organisms. *Current Opinion in Biotechnology*, 2001; 12:446–449.
- Corrêa PR. Caracterização do comportamento de *Mycobacterium bovis* BCG Moreau frente ao ambiente intracelular de macrófagos THP-1. Rio de Janeiro, Brasil 2018
- Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, Duignan P, Murray A, Dupont C, Ahmed N, Collins DM, Butler WR, Dawson D, Rodríguez D, Loureiro J, Romano MI, Alito A, Zumarraga M, Bernardelli A. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol* 2003; 53(Pt 5):1305-14.
- da Costa AC, Costa-Júnior Ade O, de Oliveira FM, Nogueira SV, Rosa JD, Resende DP, Kipnis A, Junqueira-Kipnis AP. A new recombinant BCG vaccine induces specific Th17 and Th1 effector cells with higher protective efficacy against tuberculosis. *PLoS One*. 2014; 9(12):e116033.
- Daffé M & Draper P. The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol* 1998; 39:131-203. Review.
- Dai J, Chen Y, Dean S, Morris JG, Salfinger M, Johnson JA. Multiple-genome comparison reveals new loci for *Mycobacterium* species identification. *Journal Clin Microb* 2011(A); 49(1):144–153.
- Dai J, Chen Y, Lauzardo M. Web-Accessible Database of *hsp65* sequences from *Mycobacterium* reference strains. *Journal Clin Microb* 2011 (B); 49(6):2296-2303.
- Daigneault M, Preston JA, Marriott HM, Whyte MK, Dockrell DH. The identification of markers of macrophage differentiation in PMA-stimulated THP-1 cells and monocyte-derived macrophages. *PLoS One* 2010; 5(1):e8668.
- Daley CL. *Mycobacterium avium* complex disease. *Microbiol Spectr* 2017; 5(2). Review.

- Das S, Pettersson BM, Behra PR, Ramesh M, Dasgupta S, Bhattacharya A, Kirsebom LA. The *Mycobacterium phlei* Genome: Expectations and Surprises. *Genome Biol Evol* 2016; 8(4):975-85.
- Davidson PT. The diagnosis and management of disease caused by *M. avium* complex, *M. kansasii*, and other mycobacteria. *Clin Chest Med* 1989; 10(3):431-43. Review.
- Delogu G, Bua A, Pusceddu C, Parra M, Fadda G, Brennan MJ, Zanetti S. Expression and purification of recombinant methylated HBHA in *Mycobacterium smegmatis*. *FEMS Microbiol Lett.* 2004 Oct 1;239(1):33-9;
- Dlugovitzky DG, Fontela MS, Martinel Lamas DJ, Valdez RA, Romano MC. *Mycobacterium smegmatis* synthesizes in vitro androgens and estrogens from different steroid precursors. *Can J Microbiol.* 2015; 61(7):451-5.
- Dobner P, Feldmann K, Rifai M, Löscher T, Rinder H. Journal rapid identification of mycobacterial species by PCR amplification of hypervariable 16S rRNA gene promoter region. *Of Clinical Microbiology*, Apr. 1996, p. 866–869 Vol. 34, No. 4;
- Doroghazi JR & Metcalf WW. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* 2013, 14:611
- Duan X, Liu Y, You X, Jiang Z, Yang S, Yang S. High-level expression and characterization of a novel cutinase from *Malbranchea cinnamomea* suitable for butyl butyrate production. *Biotechnol Biofuels.* 2017; 10:223.
- Dubrou S, Konjek J, Macheras E, Welté B, Guidicelli L, Chignon E, Joyeux M, Gaillard JL, Heym B, Tully T, Sapriel G. Diversity, community composition, and dynamics of nonpigmented and late-pigmenting rapidly growing mycobacteria in an urban tap water production and distribution system. *Appl Environ Microbiol* 2013; 79(18):5498-508.
- Eid J, *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* 2009; 323:133-138.
- EMBRAPA. Centro Nacional de Pesquisa de Solos. Sistema Brasileiro de Classificação de Solos. 2 ed. Rio de Janeiro: Embrapa Solos, 2006.

- EMBRAPA Cerrados, 2008. Levantamento de reconhecimento de média intensidade dos solos do Parque Nacional da Serra dos Órgãos – Parnaso, Rio de Janeiro, Escala 1:100.000 / Éder de Souza Martins et al. Planaltina, DF; p. 37 (Boletim de pesquisa e desenvolvimento / Embrapa Cerrados, ISSN 1676-918X; 209)
- Emmerick LS. Micobactérias não-tuberculosas isoladas da Mata Atlântica: aspectos genéticos, bioquímicos e identificação de antígenos compartilhados com a vacina BCG. Rio de Janeiro, Brasil, 2013.
- Ernst JD, Trevejo-Nuñez G, Banaiee N. Genomics and the evolution, pathogenesis, and diagnosis of tuberculosis. *J Clin Invest* 2007; 117(7):1738-45 Review.
- Falcão DP, Valentini SR, Leite CQF. Pathogenic or potentially pathogenic bacteria as contaminants of fresh water from different sources in Araraquara, Brazil. *Water Resources* 1993; 27(12):1737-1741.
- Falkinham III JO. Environmental sources of nontuberculous mycobacteria. *Clin Chest Med* 2015; 36(1):35-41
- Falkinham III JO. Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *J Appl Microbiol* 2009; 107(2):356-67.
- Fedrizzi T, Meehan CJ, Grottola A, Giacobazzi E, Fregni-Serpini G, Tagliazucchi S, Fabio A, Bettua C, Bertorelli R, De Sanctis V, Rumpianesi F, Pecorari M, Jousson O, Tortoli E, Segata N. Genomic characterization of nontuberculous mycobacteria. *Sci. Rep.*, 2017; 7,45258.
- Feltcher ME, Sullivan JT, Braunstein M. Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development? *Future Microbiol* 2010; 5:1581–1597.
- Fernandez FJ & Vega MC. Technologies to keep an eye on: alternative hosts for protein production in structural biology. *Curr Opin Struct Biol* 2013; 23:365-373.
- Fleischmann RD, Adams MD, White O, *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269: 496–512.
- Forbes L, Ebsworth-Mojica K, DiDone L, Li S-G, Freundlich JS, Connell N, *et al.* A High Throughput Screening Assay for Anti-Mycobacterial Small Molecules

Based on Adenylate Kinase Release as a Reporter of Cell Lysis. PLoS ONE 2015; 10(6):e0129234

Friehs K. Plasmid Copy Number and Plasmid Stability. Scheper T. (eds) New Trends and Developments in Biochemical Engineering. Advances in Biochemical Engineering 2004; vol86. Springer, Berlin, Heidelberg

Fukano H, Wada S, Kurata O, Katayama K, Fujiwara N, Hoshino Y. *Mycobacterium stephanolepidis* sp. nov., a rapidly growing species related to *Mycobacterium chelonae*, isolated from marine teleost fish, *Stephanolepis cirrhifer*. Int J Syst Evol Microbiol. 2017 Aug;67(8):2811-2817.

Fundação SOS Mata Atlântica & INPE. Atlas dos remanescentes florestais da Mata Atlântica, período de 2005 a 2008. 2010 <<http://www.sosmatatlantica.org.br>>;

Fusco da Costa AR, Fedrizzi T, Lopes ML, Pecorari M, Oliveira da Costa WL, Giacobazzi E, da Costa Bahia JR, De Sanctis V, Batista Lima KV, Bertorelli R, Grottola A, Fabio A, Mariottini A, Ferretti P, Di Leva F, Fregni Serpini G, Tagliazucchi S, Rumpianesi F, Jousson O, Segata N, Tortoli E. Characterization of 17 strains belonging to the *Mycobacterium simiae* complex and description of *Mycobacterium paraense* sp. nov. Int J Syst Evol Microbiol 2015; 65(Pt 2):656-62.

Galagan JE. Genomic insights into tuberculosis. Nat Rev Genet 2014; 15:307-320.

Galán B, Uhía I, García-Fernández E, Martínez I, Bahillo E, de la Fuente JL, Barredo JL, Fernández-Cabezón L, García JL. *Mycobacterium smegmatis* is a suitable cell factory for the production of steroidal synthons. Microb Biotechnol 2017; 10(1):138-150.

Gamerith C, Vastano M, Ghorbanpour SM, Zitzenbacher S, Ribitsch D, Zumstein MT, Sander M, Herrero Acero E, Pellis A, Guebitz GM. Enzymatic degradation of aromatic and aliphatic polyesters by *P. pastoris* expressed cutinase 1 from *Thermobifida cellulolytica*. Front Microbiol. 2017; 8:938.

Ganaie AA, Lella RK, Solanki R, Sharma C. Thermostable hexameric form of *Eis* (Rv2416c) protein of *M. tuberculosis* plays an important role for enhanced intracellular survival within macrophages. PLoS One 2011; 6:e27590



- Ganaie AA, Trivedi G, Kaur A, Jha SS, Anand S, Rana V, Singh A, Kumar S, Sharma C. Interaction of Erp protein of *Mycobacterium tuberculosis* with Rv2212 enhances intracellular survival of *Mycobacterium smegmatis*. *J Bacteriol* 2016; 198(20):2841-52.
- Garbe TR, Barathi J, Barnini S, Zhang Y, Abou-Zeid C, Tang D, Mukherjee R, Young DB. Transformation of mycobacterial species using hygromycin resistance as selectable marker. *Microbiology* 1994; 140:133–138;
- Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG. The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci USA* 2003; 100(13):7877-82.
- Gaspar LM, Santos FM, Albuquerque T, Castro-de-Sousa JP, Passarinha LA, Tomaz CT. Proteome analysis of vitreous humor in retinal detachment using two different flow-charts for protein fractionation. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2017; 1061-1062:334-341.
- Gcebe N, Rutten V, Pittius NGV, Naicker B, Michel A. *Mycobacterium malmesburyense* sp. nov., a non-tuberculous species of the genus *Mycobacterium* revealed by multiple gene sequence characterization. *Int J Syst Evol Microbiol.* 2017; 67(4):832-838.
- Gellissen, G. (ed). Front Matter, in *Production of recombinant proteins: novel microbial and eukaryotic expression systems*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG; 2005.
- Ghosh J, Larsson P, Singh B, Pettersson BM, Islam NM, Sarkar SN, Dasgupta S, Kirsebom LA. Sporulation in mycobacteria. *Proc Natl Acad Sci USA* 2009; 30;106(26):10781-6.
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison III CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 2009; vol6, p343–345.
- Gingeras TR, Ghandour G, Wang E, Berno A, Small PM, Drobniowski F, Alland D, Desmond E, Holodniy M, Drenkow J. Simultaneous genotyping and species

- identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res* 1998; 8(5):435-48.
- Goldstone RM, Moreland NJ, Bashiri G, Baker EN, Shaun Lott J. A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*. *Protein Expr Purif.* 2008 Jan;57(1):81-7. Epub 2007 Sep 14;
- Gomes LH, Otto TD, Vasconcellos EA, Ferrão PM, Maia RM, Moreira AS, Ferreira MA, Castello-Branco LR, Degraive WM, Mendonça-Lima L. Genome sequence of *Mycobacterium bovis* BCG Moreau, the Brazilian vaccine strain against tuberculosis. *J Bacteriol* 2011; 193(19):5600-1.
- González-López L, Carballar-Lejarazú R, Arrevillaga Boni G, Cortés-Martínez L, Cázares-Raga FE, Trujillo-Ocampo A, Rodríguez MH, James AA, Hernández-Hernández FC. Lys48 ubiquitination during the intraerythrocytic cycle of the rodent malaria parasite, *Plasmodium chabaudi*. *PLoS One* 2017; 12(6):e0176533.
- Greninger AL, Langelier C, Cunningham G, Keh C, Melgar M, Chiu CY, Miller S. Two rapidly growing mycobacterial species isolated from a brain abscess: first whole-genome sequences of *Mycobacterium immunogenum* and *Mycobacterium llatzerense*. *J Clin Microbiol* 2015; 53(7):2374-7.
- Gupta A, Ahmad FJ, Ahmad F, Gupta UD, Natarajan M, Katoch V, Bhaskar S. Efficacy of *Mycobacterium indicus pranii* immunotherapy as an adjunct to chemotherapy for tuberculosis and underlying immune responses in the lung. *PLoS One* 2012; 7(7):e39215.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 2013; 29(8):1072-5.
- Gutierrez-Lugo MT, Newton GL, Fahey RC, Bewley CA. Cloning, expression and rapid purification of active recombinant mycothiol ligase as B1 immunoglobulin binding domain of streptococcal protein G, glutathione-S-transferase and maltose binding protein fusion proteins in *Mycobacterium smegmatis*. *Protein Expr Purif.* 2006 Nov;50(1):128-36;

- Harboe M & Nagai S. MPB70, a unique antigen of *Mycobacterium bovis* BCG. *Am Rev Respir Dis.* 1984 Mar;129(3):444-52;
- Harriff MJ, Wu M, Kent ML, Bermudez LE. Species of environmental mycobacteria differ in their abilities to grow in human, mouse, and carp macrophages and with regard to the presence of mycobacterial virulence genes, as observed by DNA microarray hybridization. *Appl Environ Microbiol* 2008; 74(1):275-85.
- Harris TD, *et al.* Single-molecule DNA sequencing of a viral genome, *Science* (New York, N.Y.) 2008; 320:106–109
- Hartley JL, Temple GF, Brasch MA. DNA cloning using *in vitro* site-specific recombination. *Genome Research* 2000; 10,1788-1795.
- Hashish E, Merwad A, Elgaml S, Amer A, Kamal H, Elsadek A, Marei A, Sitohy M. *Mycobacterium marinum* infection in fish and man: epidemiology, pathophysiology and management: a review. *Vet Q.* 2018; 1:1-34.
- Hayward D, van Helden PD, Wiid IJ. Glutamine synthetase sequence evolution in the mycobacteria and their use as molecular markers for Actinobacteria speciation. *BMC Evol Biol.* 2009 Feb 26;9:48;
- He Z, De Buck J. Cell wall proteome analysis of *Mycobacterium smegmatis* strain mc<sup>2</sup>155. *BMC Microbiol.* 2010 Apr 22;10:121;
- Heather JM & Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics* 2016; 107(1):1-8. Review.
- Hennessee CT & Li QX. Effects of polycyclic aromatic hydrocarbon mixtures on degradation, gene expression, and metabolite production in four *Mycobacterium* species. *Appl Environ Microbiol.* 2016; 16;82(11):3357-69.
- Ho YS, Adroub SA, Abadi M, Al Alwan B, Alkhateeb R, Gao G, Ragab A, Ali S, van Soolingen D, Bitter W, Pain A, Abdallah AM. Complete genome sequence of *Mycobacterium vaccae* type strain ATCC 25954. *J Bacteriol* 2012; 194(22):6339-40.
- Hogg GG, Schinsky MF, McNeil MM, Lasker BA, Silcox VA, Brown JM. Central line sepsis in a child due to a previously unidentified *Mycobacterium*. *J Clin Microbiol.* 1999 Apr;37(4):1193-6;

- Houben ENG, Korotkov KV, Bitter W. Take five – Type VII secretion systems of *Mycobacteria*. *Biochimica et Biophysica Acta* 2014; 1843:1707–1716.
- Hunt I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.* 2005; 40, 1–22;
- Hyman ED. A new method for sequencing DNA. *Anal Biochem* 1988; 174:423–436
- Hyman M. Biodegradation of gasoline ether oxygenates *Current Opinion in Biotechnology* 2013; 24:443–450.
- Ingeras TR, Ghandour G, Wang E, Berno A, Small PM, Drobniewski F, Alland D, Desmond E, Holodniy M, Drenkow J. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res* 1998; 8:435–448.
- Jackson DA, Symons RH, Berg P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc Natl Acad Sci USA* 1972; 69(10):2904-9.
- Jacobs A, DiDone L, Jobson J, Sofia M, Krysan DJ, Dunman PM. Adenylate kinase release as a high throughput screening compatible reporter of bacterial lysis for the identification of antibacterial agents. *Antimicrob Agents Chemother.* 2013; 57:26–36.
- Jacobs WR Jr, Tuckman M, Bloom BR. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature* 1987; 327:532–535.
- Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 2016; 17:239
- Janda JM & Abbott SL. Minireview. *16S rRNA* gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology* 2007; 45(9):2761–2764.
- Jankute M, Cox JAG, Harrison J, Besra GS. Assembly of the Mycobacterial Cell Wall; *Annu. Rev. Microbiol* 2015; 69:405–423.

- Janz A, Oezel M, Kurzeder C, Mautner J, Pich D, Kost M, Hammerschmidt W, Delecluse HJ. Infectious Epstein-Barr Virus lacking major glycoprotein BLLF1 (gp350/220) demonstrates the existence of additional viral ligands. *J Virol* 2000; 74(21):10142–10152.
- 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. *Front Cell Infect Microbiol*. 2017; 7:88.
- Johnson S, Brusasca P, Lyashchenko K, Spencer JS, Wiker HG, Bifani P, Shashkina E, Kreiswirth B, Harboe M, Schluger N, Gomez M, Gennaro ML. Characterization of the secreted MPT53 antigen of *Mycobacterium tuberculosis*. *Infect Immun*. 2001; 69(9):5936-9.
- Jonasson P, Liljeqvist S, Nygren PA, Stahl S. Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. *Biotechnol. Appl. Biochem*. 2002; 35, 91–105;
- Jordão Junior CM, Lopes FC, David S, Farache Filho A, Leite CQ. Detection of nontuberculous mycobacteria from water buffalo raw milk in Brazil. *Food Microbiol*. 2009 Sep;26(6):658-61. 2009 May 3;
- Josefsen KD, Nordborg A, Sletta H. Bioconversion of phytosterols into androstenedione by *Mycobacterium*. *Methods Mol Biol* 2017; 1645:177-197.
- Kibane JJ. Microbial biocatalyst developments to upgrade fossil fuels. *Current Opinion in Biotechnology* 2006; 17:305–314
- Kim BJ, Kim BR, Lee SY, Kim GN, Kook YH, Kim BJ. Molecular Taxonomic Evidence for Two Distinct Genotypes of *Mycobacterium yongonense* via Genome-Based Phylogenetic Analysis. *PLoS One*. 2016; 11(3):e0152703.
- Kim BJ, Lee KH, Park BN, Kim SJ, Bai GH, Kim SJ, Kook YH. Differentiation of mycobacterial species by PCR-restriction analysis of DNA (342 base pairs) of the RNA polymerase gene (*rpoB*). *J Clin Microbiol* 2001; 39:2102–2109.

- Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, Chae GT, Kim EC, Cha CY, Kook YH. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J Clin Microbiol* 1999; 37(6):1714-20.
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980; 16, 111–120.
- Konjek J, Souded S, Guerardel Y, Trivelli X, Bernut A, Kremer L, Welte B, Joyeux M, Dubrou S, Euzeby JP, Gaillard JL, Sapriel G, Heym B. *Mycobacterium lutetiense* sp. nov., *Mycobacterium montmartrense* sp. nov. and *Mycobacterium arcueilense* sp. nov., members of a novel group of non-pigmented rapidly growing mycobacteria recovered from a water distribution system. *Int J Syst Evol Microbiol*. 2016; 66(9):3694-3702.
- Kremlev SG, Phelps DS. Effect of SP-A and surfactant lipids on expression of cell surface markers in the THP-1 monocytic cell line. *Am J Physiol*. 1997; 272(6 Pt 1):1070-7.
- Kubica GP, Baess I, Gordon RE, Jenkins PA, Kwapinski JBG, McDermont C, Pattyn SR, Saito H, Silcox V *et al.* A cooperative numerical analysis of the rapidly growing mycobacteria. *J Gen Microbiol* 1972; 73:55–70.
- Kundu A, Kundu S, Chattopadhyay K. The presence of non-native helical structure in the unfolding of a beta-sheet protein MPT63. *Protein Sci*. 2017; 26(3):536-549.
- Kusunoki S & Ezaki T. Proposal of *Mycobacterium peregrinum* sp. nov., nom. rev., and elevation of *Mycobacterium chelonae* subsp. *abscessus* (Kubica *et al.*) to species status: *Mycobacterium abscessus* comb. nov. *Int J Syst Bacteriol* 1992; 42:240–245.
- Kweon O, Kim SJ, Blom J, Kim SK, Kim BS, Baek DH, Park SI, Sutherland JB, Cerniglia CE. Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus *Mycobacterium*. *BMC Evol Biol.*, 2015; 15,21.

- Labidi A, David HL, Roulland-Dussoix D. Restriction endonuclease mapping and cloning of *Mycobacterium fortuitum* var. *fortuitum* plasmid pAL5000. Ann. Inst. Pasteur Microbiol. 1985. 136B, 209–215.
- Lage R, Biccigo DG, Santos FB, Chimara E, Pereira ES, Costa Ad. *Mycobacterium chelonae* cutaneous infection in a patient with mixed connective tissue disease. An Bras Dermatol. 2015; 90(1):104-7.
- Lamy B, Marchandin H, Hamitouche K, Laurent F. *Mycobacterium setense* sp. nov., a *Mycobacterium fortuitum*-group organism isolated from a patient with soft tissue infection and osteitis. Int J Syst Evol Microbiol 2008; 58:486-490.
- Lazraq R, Clavel-Seres S, David HL. Transformation of distinct mycobacterial species by shuttle vectors derived from the *Mycobacterium fortuitum* pAL5000 plasmid. Curr Microbiol 1991. 22, 9-13.
- Leão SC, Tortoli E, Viana-Niero C, Ueki SM, Batista Lima K, Lopes ML, Yubero J, Menendez MC, Garcia MJ. Characterization of mycobacteria from a major Brazilian outbreak suggests that revision of the taxonomic status of members of the *Mycobacterium chelonae*-*M. abscessus* group is needed. J Clin Microbiol 2009; 47,2691-2698.
- Learn-Han L, Yoke-Kqueen C, Shiran MS, Vui-Ling CM, Nurul-Syakima AM, Son R, Andrade HM. Identification of actinomycete communities in Antarctic soil from Barrientos Island using PCR-denaturing gradient gel electrophoresis. Genet Mol Res. 2012 Feb 8;11(1):277-91;
- Lee JC & Whang KS. *Mycobacterium aquiterrae* sp. nov., a rapidly growing bacterium isolated from groundwater. Int J Syst Evol Microbiol 2017; Oct;67(10):4104-4110.
- Lee JS, Son JW, Jung SB, Kwon YM, Yang CS, Oh JH, Song CH, Kim HJ, Park JK, Paik TH, Jo EK. Ex vivo responses for interferon-gamma and proinflammatory cytokine secretion to low-molecular-weight antigen MTB12 of *Mycobacterium tuberculosis* during human tuberculosis. Scand J Immunol. 2006; 64(2):145-54.
- Levy SE & Myers RM. Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet. 2016 31;17:95-115. Review.

- Lewinsohn TM & Prado PI. Biodiversidade brasileira: síntese do estado atual do conhecimento. São Paulo: Contexto, 176 p.; 2002.
- Li MZ & Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* 2007; 4(3):251-6.
- Liang Y, Gardner DR, Miller CD, Chen D, Anderson AJ, Weimer BC, Sims RC. Study of biochemical pathways and enzymes involved in pyrene degradation by *Mycobacterium sp.* strain KMS. *AEM* 2006; 72(12):7821-7828.
- Lin, SH, Liao YC. CISA: contig integrator for sequence assembly of bacterial genomes. *PloS one* 2013, 8(3):e60843
- MacCannell D. Next generation sequencing in clinical and public health microbiology. *Clinical Microbiology Newsletter*, 2016. Vol. 38, Issue 21, 169-176.
- Mantovani W. A degradação dos biomas brasileiros. Patrimônio ambiental brasileiro. Universidade de São Paulo. São Paulo, pp. 367-439, 2003;
- Martins ES, Carvalho Júnior OA, Souza VV, Couto Júnior AF, Oliveira SN, Gomes RA, Trancoso Reatto A. Relação solo-relevo em vertentes assimétricas no Parque Nacional da Serra dos Órgãos, RJ. *Revista Brasileira de Geomorfologia* 2007; 8(1):45-62.
- Maxam AM & Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977; 74:560–564
- Mckernan KJ, *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding, *Genome Res* 2009; 19:1527–1541.
- Medeiros MA, Dellagostin OA, Armôa GR, Degrave WM, De Mendonça-Lima L, Lopes MQ, Costa JF, McFadden J, McIntosh D. Comparative evaluation of *Mycobacterium vaccae* as a surrogate cloning host for use in the study of mycobacterial genetics. *Microbiology* 2002; 148(Pt 7):1999-2009.
- Mehta PK, King CH, White EH, Murtagh JJ Jr, Quinn FD. Comparison of in vitro models for the study of *Mycobacterium tuberculosis* invasion and intracellular replication. *Infect Immun.* 1996 Jul;64(7):2673-9.



- Mendoza-Coronel E & Castañón-Arreola M. Comparative evaluation of in vitro human macrophage models for mycobacterial infection study. *Pathog Dis.*, 2016; 74(6).
- Michalski A, Eugen Damoc, Jan-Peter Hauschild, Oliver Lange, Andreas Wiegand, Alexander Makarov, Nagarjuna Nagaraj, Juergen Cox, Matthias Mann, Stevan Horning. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics* 2011; 10(9).
- Miller CD, Hall K, Liang YN, Nieman K, Sorensen D, Issa B, Anderson AJ, Sims RC. Isolation and characterization of polycyclic aromatic hydrocarbon-degrading *Mycobacterium* isolates from soil. *Microb Ecol* 2004; 48(2):230-8.
- Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, ... Sutton G. Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 2008, 24(24):2818–24
- Ministério da Saúde. Secretaria de Vigilância em Saúde. Centro de Referência Hélio Fraga. Manual de bacteriologia da tuberculose. Rio de Janeiro, 3ª ed, 240 p., 2005.
- Mittermeier RA, Robles-Gil P, Mittermeier CG (eds.). Megadiversity: Earth's biologically wealthiest nations. CEMEX, Agrupación Serra Madre, S.C., Mexico. 1997;
- Miyake R, Kawamoto J, Wei YL, Kitagawa M, Kato I, Kurihara T, Esaki N. Construction of a low-temperature protein expression system using a cold-adapted bacterium, *Shewanella* sp. strain Ac10, as the host. *Appl Environ Microbiol* 2007; 73(15):4849-56.
- MMA (Ministério do Meio Ambiente). Cartilha do Pesquisador do Parque Nacional Serra dos Órgãos. Setor de Pesquisa Parque Nacional Serra dos Órgãos. 2007 Fev;
- Moore M & Frerichs JB. An unusual acid-fast infection of the knee with subcutaneous, abscess-like lesions of the gluteal region; report of a case with a study of the organism, *Mycobacterium abscessus*, n. sp. *J Invest Dermatol* 1953; 20:133–169.

- Moreira LM (organizador). Ciências genômicas: fundamentos e aplicações. Ribeirão Preto: Sociedade Brasileira de Genética, 2015; 403 p: il.
- Morgado SM, Marín MA, Freitas FS, Fonseca EL, Vicente ACP. Complete plasmid sequence carrying type IV-like and type VII secretion systems from an atypical mycobacteria strain. Mem Inst Oswaldo Cruz 2017; 112(7):514-516.
- Morgado SM. Diversidade, taxonomia genômica e resistoma de micobactérias da Mata Atlântica. Rio de Janeiro, Brasil 2017.
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. Nature 2000, 403:853-845;
- Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. 1988. Electrophoresis 1988; 9(6):255-62.
- Niemann S, Rüscher-Gerdes S, Joloba ML, Whalen CC, Guwatudde D, Ellner JJ, Eisenach K, Fumokong N, Johnson JL, Aisu T, Mugerwa RD, Okwera A, Schwander SK. *Mycobacterium africanum* subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda. J Clin Microbiol. 2002 Sep;40(9):3398-405.
- Noens EE, Williams C, Anandhakrishnan M, Poulsen C, Ehebauer MT, Wilmanns M. Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1ΔC expression strain. BMC Biotechnol. 2011; 11:27;
- Nogueira CL, Whipps CM, Matsumoto CK, Chimara E, Droz S, Tortoli E, de Freitas D, Cnockaert M, Palomino JC, Martin A, Vandamme P, Leão SC. *Mycobacterium saopaulense* sp. nov., a rapidly growing *Mycobacterium* closely related to members of the *Mycobacterium chelonae-Mycobacterium abscessus* group. Int J Syst Evol Microbiol. 2015 Dec;65(12):4403-9.
- Nouioui I, Sangal V, Carro L, Teramoto K, Jando M, Montero-Calasanz MDC, Igual JM, Sutcliffe I, Goodfellow M, Klenk HP. Two novel species of rapidly growing mycobacteria: *Mycobacterium lehmannii* sp. nov. and *Mycobacterium neumannii* sp. nov. Int J Syst Evol Microbiol. 2017 Dec;67(12):4948-4955.

- Nyrén PI & Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal Biochem* 1985; 509:504–509.
- Nyrén PI. Enzymatic method for continuous monitoring of DNA polymerase activity, *Anal Biochem*; 1987 238:235–238.
- Odumeru J, Gao A, Chen S, Raymond M, Mutharia L. Use of the bead beater for preparation of *Mycobacterium paratuberculosis* template DNA in milk. *Can J Vet Res*. 2001 Oct;65(4):201-5.
- Palmer CE & Long MW. Effects of infection with atypical mycobacteria on BCG vaccination and tuberculosis. *Am Rev Respir Dis* 1966; 94(4):553-68.
- Pan W, Bai Z, Su T, Wang Z. Enzymatic degradation of poly(butylene succinate) with different molecular weights by cutinase. *Int J Biol Macromol* 2018; 111:1040-1046.
- Paniz-Mondolfi AE, Greninger AL, Ladutko L, Brown-Elliott BA, Vasireddy R, Jakubiec W, Vasireddy S, Wallace RJ Jr, Simmon KE, Dunn BE, Jackoway G, Vora SB, Quinn KK, Qin X, Campbell S. *Mycobacterium grossiae* sp. nov., a rapidly growing, scotochromogenic species isolated from human clinical respiratory and blood culture specimens. *Int J Syst Evol Microbiol*. 2017 Nov;67(11):4345-4351.
- Parish T & Brown AC. *Mycobacteria protocols*. Second edition. *Methods in molecular Biology* 465. Humana Press, 2008;
- Phelan J, Maitra A, McNerney R, Nair M, Gupta A, Coll F, Pain A, Bhakta S, Clark TG. The draft genome of *Mycobacterium aurum*, a potential model organism for investigating drugs against *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *Int J Mycobacteriol* 2015; 4(3):207-16.
- Philipp B. Bacterial degradation of bile salts. *Appl Microbiol Biotechnol* 2011; 89(4):903-15.
- Piggot PJ & Hilbert DW. Sporulation of *Bacillus subtilis*. *Curr Opin Microbiol*. 2004; 7(6):579-86. Review.
- Pio TF & Macedo GA. Cutinases: properties and industrial applications. *Adv Appl Microbiol*. 2009;66:77-95.

- Posada D. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Lemey P, Salemi M, Vandamme AM (eds.) Section IV Testing models and trees, Unit 10 Selecting models of evolution, pp 345-361. Cambridge University Press 2009.
- Prasanna AN & Mehra S. Comparative Phylogenomics of Pathogenic and Non-Pathogenic *Mycobacterium*. PLoS ONE 2013; 8(8):e71248.
- Primm TP, Lucero CA, Falkinham III JO. Health Impacts of Environmental Mycobacteria. Clin Microb Reviews 2004; 17(1):98–106.
- Quan J & Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS ONE 2009; 4(7):e6441.
- Rabilloud T & Lelong C. Two-dimensional gel electrophoresis in proteomics: a tutorial. J Proteomics. 2011; 74(10):1829-41.
- Ramon A & Smith HO. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. Biotechnology Letters 2011; 33(3):549–555.
- Ramos JP, Campos CE, Caldas PC, Ferreira NV, da Silva MV, Redner P, Campelo CL, Vale SF, Barroso EC, Medeiros RF, Montes FC, Galvão TC, Tortoli E. *Mycobacterium fragae* sp. nov., a non-chromogenic species isolated from human respiratory specimens. Int J Syst Evol Microbiol 2013; 63(Pt 7):2583-7.
- Rastogi N, Legrand E, Sola C. The mycobacteria: an introduction to nomenclature and pathogenesis. Rev Sci Tech. 2001; 20(1):21-54 Review.
- Restrepo AV, Salem JI, Ogusku MM, Gomes LF, Fraiji NA. Pesquisa de Micobactérias Ambientais em água de torneira, luvas e soluções utilizadas em procedimentos cirúrgicos no Hospital Universitário Getúlio Vargas - Manaus/AM. Acta Amazonica 2009; 39(4):889 – 900.
- Rienksma RA, Suarez-Diez M, Mollenkopf HJ, Dolganov GM, Dorhoi A, Schoolnik GK, Martins Dos Santos VA, Kaufmann SH, Schaap PJ, Gengenbacher M. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC Genomics. 2015; 5(16):34.

- Roche PW, Feng CG, Britton WJ. Human T-cell epitopes on the *Mycobacterium tuberculosis* secreted protein MPT64, *Scand. J Immunol* 1996; 43:662–670.
- Rodriguez-Campos S, Smith NH, Boniotti MB, Aranaz A. Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics and legislation of bovine tuberculosis. *Res Vet Sci* 2014; 97,S5–S19.
- Rogall T, Flohr T, Böttger EC. Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. *J Gen Microbiol* 1990 (B); 136(9):1915-20.
- Rogall T, Wolters J, Flohr T, Bottger EC. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *Intern Journal of System Bacteriol*, 1990 (A), p. 323-330. Vol. 40, No. 4.
- Rothberg JM, *et al.* An integrated semiconductor device enabling non-optical genome sequencing, *Nature* 2011; 475:348–352.
- Runyon EH. Anonymous mycobacteria in pulmonary disease. *Med. Clin. North Am.* 1959 43:273–290.
- Russell, DG. *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat Rev Mol Cell Biol* 2001; 2(8):569-77.
- Saha MS, Pal S, Sarkar I, Roy A, Das Mohapatra PK, Sen A. Comparative genomics of *Mycobacterium* reveals evolutionary trends of *M. avium* complex. *Genomics*. 2018 Feb 28. S0888-7543(18)30129-0.
- Saini V, Raghuvanshi S, Khurana JP, Ahmed N, Hasnain SE, Tyagi AK, Tyagi AK. Massive gene acquisitions in *Mycobacterium indicus pranii* provide a perspective on mycobacterial evolution. *Nucleic Acids Res* 2012; 40(21):10832-50.
- Saitou N & Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; 4(4):406-25.
- Samten B, Wang X, Barnes PF. *Mycobacterium tuberculosis* ESX-1 system-secreted protein ESAT-6 but not CFP10 inhibits human T-cell immune responses. *Tuberculosis* 2009; 89,S74-S76
- Sánchez A, Espinosa P, García T, Mancilla R. The 19 kDa *Mycobacterium tuberculosis* lipoprotein (LpqH) induces macrophage apoptosis through extrinsic

- and intrinsic pathways: a role for the mitochondrial apoptosis-inducing factor. *Clin Dev Immunol* 2012; 2012:950503.
- Sandim V, Pereira Dde A, Kalume DE, Oliveira-Carvalho AL, Ornellas AA, Soares MR, Alves G, Zingali RB. Proteomic analysis reveals differentially secreted proteins in the urine from patients with clear cell renal cell carcinoma. *Urol Oncol*. 2016; 34(1):5.e11-25.
- Sanger F, *et al.*, Nucleotide sequence of bacteriophage phi X174 DNA, *Nature* 1977; 265:687–695
- Sanger F. Frederick Sanger — Biographical, 1980 (URL [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1980/sanger-bio.html](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-bio.html)).
- Sanger SF & Nicklen ARC. DNA sequencing with chain-terminating, *Proc Natl Acad Sci* 1977; 74:5463–5467.
- Saviola B & Bishai W. The Genus *Mycobacterium* – Medical. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds) *The Prokaryotes*. Springer, New York, NY; 2006.
- Schwende H, Fitzke E, Ambs P, Dieter P (1996) Differences in the state of differentiation of THP-1 cells induced by phorbol ester and 1,25-dihydroxyvitamin D3. *J Leukoc Biol* 1996; 59:555–561.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014. 30(14):2068–2069.
- Shahraki AH, Trovato A, Mirsaeidi M, Borroni E, Heidarieh P, Hashemzadeh M, Shahbazi N, Cirillo DM, Tortoli E. *Mycobacterium persicum* sp. nov., a novel species closely related to *Mycobacterium kansasii* and *Mycobacterium gastri*. *Int J Syst Evol Microbiol*. 2017; 67(6):1766-1770.
- Shinnick TM. The 65-Kilodalton antigen of *Mycobacterium tuberculosis*. *Journal of Bacteriology* 1987; 169(3):1080-1088.
- Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology*. 2015; 6:242.

- Silva BD, da Silva EB, do Nascimento IP, Dos Reis MC, Kipnis A *et al.* MPT-51/CpG DNA vaccine protects mice against *Mycobacterium tuberculosis*. *Vaccine* 2009 27: 4402–4407.
- Silver RF, Li Q, Ellner JJ. Expression of virulence of *Mycobacterium tuberculosis* within human monocytes: virulence correlates with intracellular growth and induction of tumor necrosis factor alpha but not with evasion of lymphocyte-dependent monocyte effector functions. *Infect Immun* 1998; 66(3):1190-9.
- Simmon KE, Brown-Elliott BA, Ridge PG, Durtschi JD, Mann LB, Slechta ES, Steigerwalt AG, Moser BD, Whitney AM, Brown JM, Voelkerding KV, McGowan KL, Reilly AF, Kirn TJ, Butler WR, Edelstein PH, Wallace RJ Jr, Petti CA. *Mycobacterium chelonae-abscessus* complex associated with sinopulmonary disease, Northeastern USA. *Emerg Infect Dis.* 2011; 17(9):1692-700.
- Simpson AJ, Reinach FC, Arruda P, Abreu FA, ..., Setubal JC (The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 2000; 406(6792):151-9.
- Singh AK, Reyrat JM. Laboratory maintenance of *Mycobacterium smegmatis*. *Curr Protoc Microbiol* 2009. Chapter 10:Unit10C1.
- Singhal N, Sharma P, Kumar M, Joshi B, Bisht D. Analysis of intracellular expressed proteins of *Mycobacterium tuberculosis* clinical isolates. *Proteome Science* 2012, 10:14;
- Snapper SB, Melton RE, Mustafa S, Kieser T, Jacobs JWR. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* 1990; 4:1911–1919.
- Sood U, Singh Y, Shakarad M, Lal R. Highlight on Engineering *Mycobacterium smegmatis* for testosterone production. *Microb Biotechnol.* 2017; 10(1):73-75.
- Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; 33:152–155.

- Stanley SA, Raghavan S, Hwang WW, Cox JS. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci USA* 2003; 100(22):13001-6.
- Steffani-Vallejo JL, Brunck ME, Acosta-Cruz EY, Montiel R, Barona-Gómez F. Genomic insights into *Mycobacterium simiae* human colonization. *Stand Genomic Sci.* 2018; 13:1.
- Stokes RW, Doxsee D. The receptor-mediated uptake, survival, replication, and drug sensitivity of *Mycobacterium tuberculosis* within the macrophage-like cell line THP-1: a comparison with human monocyte-derived macrophages. *Cell Immunol* 1999; 10;197(1):1-9.
- Stokes RW, Haidl ID, Jefferies WA, Speert DP. Mycobacteria-macrophage interactions. Macrophage phenotype determines the nonopsonic binding of *Mycobacterium tuberculosis* to murine macrophages. *J Immunol.* 1993; 151(12):7067-76.
- Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 1986; 189:113-130.
- Summers DK & Sherratt DJ. Bacterial plasmid stability. *Bioessays* 1985; 2:209–211.
- Sutherland JB, Heinze TM, Pearce MG, Deck J, Williams AJ, Freeman JP. Biotransformation of acridine by *Mycobacterium vanbaalenii*. *Environ Toxicol Chem* 2009 Jan;28(1):61-4.
- Tavaré S. Some probabilistic and statistical problems in the analysis of dna sequences. *Some mathematical questions in biology - DNA sequence analysis.* Amer. Math. Soc., Providence, RI, pages 57–86; 1986.
- Telenti A, Marchesi F, Balz M, Bally F, Böttger EC, Bodmer T. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *Journal Clin Microb* 1993; 31(2):175-178.
- Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 2002; 60, 523–533;



- Theus SA, Cave MD, Eisenach KD. Activated THP-1 cells: an attractive model for the assessment of intracellular growth rates of *Mycobacterium tuberculosis* isolates. *Infection and immunity* 2004; 72(2):1169–1173.
- Timm J, Perilli MG, Duez C, Trias J, Orefici G, Fattorini L, Amicosante G, Oratore A, Joris B, Frere JM, Pugsley AP, Gicquel B. Transcription and expression analysis, using lacZ and phoA gene fusions, of *Mycobacterium fortuitum* beta-lactamase genes cloned from a natural isolate and a high-level beta-lactamase producer. *Mol Microbiol* 1994; 12:491–504.
- Tobias NJ, Doig KD, Medema MH, Chen H, Haring V, Moore R, Seemann T, Stinear TP. Complete genome sequence of the frog pathogen *Mycobacterium ulcerans* ecovar Liflandii. *J Bacteriol* 2013; 195(3):556-64.
- Toro A, Adekambi T, Cheynet F, Fournier PE, Drancourt M. *Mycobacterium setense* infection in humans. *Emerg Infect Dis* 2008; 14:1330-332.
- Tortoli E, Fedrizzi T, Meehan CJ, Trovato A, Grottola A, Giacobazzi E, Serpini GF, Tagliazucchi S, Fabio A, Bettua C, Bertorelli R, Frascaro F, De Sanctis V, Pecorari M, Jousson O, Segata N, Cirillo DM. The new phylogeny of the genus *Mycobacterium*: The old and the news. *Infect Genet Evol* 2017; 56:19-25.
- Tortoli E, Kohl TA, Brown-Elliott BA, Trovato A, Le~ao SC *et al.* Emended description of *Mycobacterium abscessus*, *Mycobacterium abscessus* subsp. *abscessus* and *Mycobacterium abscessus* subsp. *bolletii* and designation of *Mycobacterium abscessus* subsp. *massiliense* comb. nov. *Int J Syst Evol Microbiol* 2016; 66:4471–4479.
- Tortoli E, Kohl TA, Brown-Elliott BA, Trovato A, Leão SC, Garcia MJ, Vasireddy S, Turenne CY, Griffith DE, Philley JV, Niemann S, Wallace Jr RJ, Cirillo DM. *Mycobacterium abscessus*, a taxonomic puzzle - Letter to the editor. *Int J Syst Evol Microbiol* 2018; 68:467–469.
- Tortoli E. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin Microb Reviews*, Apr. 2003, p. 319–354 Vol. 16, No. 2;
- Tortoli E. Microbiological features and clinical relevance of new species of the genus *Mycobacterium*. *Clin Microbiol Rev* 2014; 27(4):727-52.

- Tortoli E. The new mycobacteria: an update. *FEMS Immunol Med Microbiol* 2006, 48:159–178;
- Tortora GJ, Funke BR, Case CL. *Microbiology*; 8.ed. Porto Alegre: Editora Artmed, 2005. 894p.
- Triccas JA, Parish T, Britton WJ, Gicquel B. An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*. *FEMS Microbiol Lett* 1998; 167:151-156.
- Tun HM, Brar MS, Khin N, Jun L, Hui RK, Dowd SE, Leung FC. Gene-centric metagenomics analysis of feline intestinal microbiome using 454 junior pyrosequencing. *J Microbiol Methods*. 2012. 88(3):369-76.
- Uhía I, Williams KJ, Shahrezaei V, Robertson BD. Mycobacterial growth. *Cold Spring Harb Perspect Med*. 2015; 8;5(10). Review.
- van Soolingen D, Hoogenboezem T, de Haas PE, Hermans PW, Koedam MA, Teppema KS, Brennan PJ, Besra GS, Portaels F, Top J, Schouls LM, van Embden JD. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* 1997; 47(4):1236-45.
- Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen W. Genomics of actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* 2007; 71(3):495–548.
- Verli H. *Bioinformática: da biologia à flexibilidade molecular*. Ed Sociedade Brasileira de Bioquímica e Biologia Molecular. 2014
- Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009; 55:641–658.
- Waddell SJ, Butcher PD, Stoker NG. RNA profiling in host-pathogen interactions. *Curr Opin Microbiol* 2007; 10(3):297-302.
- Wang F, Jain P, Gulten G, Liu Z, Feng Y, Ganesula K, Motiwala AS, Ioerger TR, Alland D, Vilcheze C, Jacobs WR, Jr., Sacchettini JC. *Mycobacterium tuberculosis* dihydrofolate reductase is not a target relevant to the antitubercular activity of isoniazid. *Antimicrob Agents Chemother* 2010; 54:3776–3782.

- Wayne LG & Sramek HA. Agents of newly recognized or infrequently encountered mycobacterial diseases. *Clinical Microbiology Reviews*, Jan. 1992, p. 1-25 Vol. 5, N<sup>o</sup>. 1;
- Wheeler PR & Ratledge C. In *Tuberculosis: pathogenesis, protection, and control* (ed. Bloom, B. R.) 353-385 (Am Soc Microbiol, Washington DC, 1994).
- Whipps CM, Butler WR, Pourahmad F, Watral VG, Kent ML. Molecular systematics support the revival of *Mycobacterium salmoniphilum* (ex Ross 1960) sp. nov., nom. rev., a species closely related to *Mycobacterium chelonae*. *Int J Syst Evol Microbiol*. 2007; 57:2525-31.
- Whitman WB. *Bergey's Manual of Systematic Bacteriology*. Springer New York. Vol5: The Actinobacteria, p312-375. 2012.
- Williamson JC, Edwards AV, Verano-Braga T, Schwämmle V, Kjeldsen F, Jensen ON, Larsen MR. High-performance hybrid Orbitrap mass spectrometers for quantitative proteome analysis: observations and implications. *Proteomics*. 2016.
- World Health Organization (WHO). Global tuberculosis control 2016. Disponível em: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/). Acessado em: janeiro, 2017.
- Xu S, Cooper A, Sturgill-Koszycki S, van Heyningen T, Chatterjee D, Orme I, Allen P, Russell DG. Intracellular trafficking in *Mycobacterium tuberculosis* and *Mycobacterium avium*-infected macrophages. *J Immunol* 1994; 153(6):2568-78.
- Yao K, Wang FQ, Zhang HC, Wei DZ. Identification and engineering of cholesterol oxidases involved in the initial step of sterols catabolism in *Mycobacterium neoaurum*. *Metab Eng* 2013; 15:75-87.
- Zhang C, Anderson AJ. The gluconeogenic pathway in a soil *Mycobacterium* isolate with bioremediation ability. *Curr Microbiol* 2013; 66(2):122-31.
- Zhang J, Yuan Q, Yang W, Wang X. Complete genome sequence of carbendazim-degrading *Mycobacterium* sp. strain djl-10. *Genome Announc* 2017; 5(8):e01683-16.

Zhang L, Zhang C, Cheng Z, Yao Y, Chen J. Biodegradation of benzene, toluene, ethylbenzene, and o-xylene by the bacterium *Mycobacterium cosmeticum* byf-4. Chemosphere 2013; 90(4):1340-7.

Zhang Y, Werling U, Edelman W. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Research 2012; 40(8):e55.

## 8. Anexo

Tabela 8.1: Sequências para análise filogenética

Espécie	16S rRNA	<i>hsp65</i>	<i>rpoB</i>
1 <i>M. abscessus abscessus</i>	AY457071.1	AY458075.1	AY147164.1
2 <i>M. abscessus bolletii</i>	AY859681.1	AY859675.1	AY859692.1
3 <i>M. abscessus massiliense</i>	AY593980.1	EU266578.1	AY593981.2
4 <i>M. africanum</i>	AF480605.1	AF547803.1	FR878060.1
5 <i>M. agri</i>	AJ429045.1	AY438080.1	PDCP01000008.1
6 <i>M. aichiense</i>	X55598.1	AF547804.1	
7 <i>M. algericum</i>	GU564404.1	GU564405.1	MVHC01000005.1
8 <i>M. alsense</i>	AJ938169.1	DQ381733.1	MVHD01000001.1
9 <i>M. alvei</i>	AF023664.1	AF547805.1	
10 <i>M. angelicum</i>	AM884328.2	AM902962.2	MVHE01000010.1
11 <i>M. anyangense</i>	KJ855063.1	KF910182.1	
12 <i>M. arabiense</i>	KC010491.1	KC010487.1	
13 <i>M. arceuilense</i>	KP644745.1	KP644751.1	
14 <i>M. aromaticivorans</i>	JALN02000001.1	DQ841182.1	JALN02000001.1
15 <i>M. arosiense</i>	EF054881.2	EU370531.1	MVHG01000006.1
16 <i>M. arupense</i>	DQ157760.2	DQ168662.1	MVHH01000051.1
17 <i>M. asiaticum</i>	NR_041901.1	GU362517.1	MVHI01000003.1
18 <i>M. aubagnense</i>	AY859683.1	AY859677.1	AY859694.1
19 <i>M. aurum</i>	NR_116555.1	FJ172326.1	
20 <i>M. austroafricanum</i>	NR_026284.1	AF547807.1	
21 <i>M. avium avium</i>	GQ153272.1	JF491291.1	CP000479.1
22 <i>M. avium paratuberculosis</i>	NR_116041.1	AF547809.1	NZ_AGAR01000293.1
23 <i>M. avium silvaticum</i>	EF521891.1	EU239781.1	
24 <i>M. bacteremicum</i>	FJ172308.1	FJ172314.1	MVHJ01000002.1
25 <i>M. boenickei</i>	AY012573.2	AY943195	NZ_FUWC01000002.1
26 <i>M. bohemicum</i>	NR_026054.1	AF547811.1	LQOK01000013.1
27 <i>M. botniense</i>	NR_028878.1	AF547812	
28 <i>M. bouchedurhonense</i>	EF591053.1	HM602039	MVHL01000001.1
29 <i>M. bourgelatii</i>	NR_135729.1	HF566385.2	
30 <i>M. bovis bovis</i>	NZ_CP009449.1	AF547813	LT708304.1
31 <i>M. bovis caprae</i>	NR_028879.1	AF547884	CP016401.1
32 <i>M. branderi</i>	NR_041897.1	AF547815	JQ599259.2
33 <i>M. brisbanense</i>	AY012577.2	JF491333	BCSX01000024.1
34 <i>M. brumae</i>	NR_025233.1	AF547816	PDCN02000001.1
35 <i>M. canariasense</i>	NR_025737.1	JF491316	BCSY01000070.1
36 <i>M. celatum</i>	NR_029143.1	JF491292	AB824285.1
37 <i>M. celeriflavum</i>	KJ607136.1	KJ586615.1	MVHN01000005.1
38 <i>M. chelonae</i>	AY457072.1	JF491293	AY147163.1
39 <i>M. chimaera</i>	NR_029003.1	GQ153296	CP015278.1
40 <i>M. chitae</i>	NR_029220.1	AF547819	
41 <i>M. chlorophenicum</i>	X79292.1	AF547820	JYNL01000051.1
42 <i>M. chubuense</i>	NR_041902.1	AF547821	MVHO01000016.1
43 <i>M. colombiense</i>	AM062764.1	GQ153298	AFVW02000001.1
44 <i>M. conceptionense</i>	AY859684.1	AM902957	AY859695.1
45 <i>M. confluentis</i>	AJ634379.1	AF547822	LQOQ01000001.1
46 <i>M. conspicuum</i>	NZ_LQOR01000030.1	AF547823	LQOR01000031.1
47 <i>M. cookii</i>	NR_114661.1	AF547824	AY544904.1
48 <i>M. cosmeticum</i>	AY449728.1	AY449730	CCBB010000003.1
49 <i>M. crocinum</i>	DQ534008.1	DQ533998	DQ534002.1
50 <i>M. diernhoferi</i>	NR_041903.1	AF547825	PDCR01000058.1
51 <i>M. doricum</i>	NR_025099.1	AF547826	LQOS01000049.1

52	<i>M. duvalii</i>	NR_026073.1	AF547827	PDCQ01000008.1
53	<i>M. elephantis</i>	NR_025296.1	AF547828	MVHP01000002.1
54	<i>M. engbaekii</i>	NR_114655.1	JN571196.1	LQOT01000019.1
55	<i>M. europaeum</i>	NR_125568.1	HM022220.1	LQOU01000021.1
56	<i>M. fallax</i>	NR_041904.1	JF491294.1	LQOJ01000040.1
57	<i>M. farcinogenes</i>	AY457084.1	AY458073.1	HG964481.1
58	<i>M. flavescens</i>	NR_044815.1	GU362519.1	
59	<i>M. florentinum</i>	NR_042223.1	JF491317.1	LQOV01000015.1
60	<i>M. fluoranthenivorans</i>	AJ617741.1	JF491318.1	DQ987721.1
61	<i>M. fortuitum acetamidolyticum</i>	FR733720.1	JF491314.1	BCSZ01000064.1
62	<i>M. fortuitum fortuitum</i>	AY457066.1	JF491295.1	AY147165.1
63	<i>M. fragae</i>	JQ898451.1	JQ902012.1	LQOW01000024.1
64	<i>M. franklinii</i>	NZ_MVHQ01000003.1	KM392060.1	MVHQ01000008.1
65	<i>M. frederiksbergense</i>	AJ276274.1	AF547834.1	
66	<i>M. gadium</i>	DQ473310.1	AF547835.1	
67	<i>M. gastris</i>	NR_041905.1	JF491315.1	LQOX01000010.1
68	<i>M. genavense</i>	NR_029223.1	AF547837.1	NZ_JAGZ01000017.1
69	<i>M. gilvum</i>	NR_118915.1	AF547838.1	CP000656.1
70	<i>M. goodii</i>	AY457079.1	AY458071.1	CP012150.1
71	<i>M. gordonae</i>	NR_044812.1	AF434734.1	LQOY01000155.1
72	<i>M. haemophilum</i>	U06638.1	JF491296.1	CP011883.2
73	<i>M. hassiacum</i>	NR_026011.1	AF547842.1	AMRA01000027.1
74	<i>M. heckeshornense</i>	NR_028759.1	AF547843.2	MPJF02000001.1
75	<i>M. heidelbergense</i>	NR_025268.1	AF547844.1	MVHR01000012.1
76	<i>M. helvum</i>	KM216314.3	KM216312.1	
77	<i>M. heraklionense</i>	GU084182.2	JN571192.1	LZKB01000106.1
78	<i>M. hiberniae</i>	NZ_LQOZ01000008.1	JF491297.1	LQOZ01000020.1
79	<i>M. hippocampi</i>	FN430736.1		
80	<i>M. hodleri</i>	NR_026286.1	AF547845.1	
81	<i>M. holsaticum</i>	AJ310467.1	AY438084.1	
82	<i>M. houstonense</i>	AY457067.1	AY458077.1	
83	<i>M. immunogenum</i>	NR_104876.1	AY458081.1	AY262739.1
84	<i>M. insubricum</i>	NR_125525.1	JF491319.1	MVHS01000036.1
85	<i>M. interjectum</i>	HM037998.1	JF491298.1	LQPB01000090.1
86	<i>M. intermedium</i>	NR_044840.1	AF547847.1	MVHT01000034.1
87	<i>M. intracellulare</i>	AJ536036.1	AF126035.1	JF271805.1
88	<i>M. iranicum</i>	HQ009482.1	HQ009485.1	LQPC01000020.1
89	<i>M. kansasii</i>	NR_121712.2	JF491300.1	CP006835.1
90	<i>M. komossense</i>	NR_026086.1	AY438649.1	AY544936.1
91	<i>M. koreense</i>	JF271826.1	JF271827.1	NCXO01000002.1
92	<i>M. kubicae</i>	NR_025000.1	AF547850.1	LQPD01000019.1
93	<i>M. kumamotoense</i>	AB239925.1	JF491323.1	MVHU01000008.1
94	<i>M. kyorinense</i>	AB370111.2	AB370171.1	JQ717032.1
95	<i>M. lacus</i>	HM022200.1	AY438090.1	LQPF01000022.1
96	<i>M. lentiflavum</i>	NR_041898.1	AF547851.1	CTEE01000001.1
97	<i>M. leprae</i>	X53999.1	NC_002677.1	Z14314.1
98	<i>M. lepraemurium</i>	CP021238.1	AY550232.1	CP021238.1
99	<i>M. litorale</i>	CP019882.1	JF431428.1	CP019882.1
100	<i>M. llutzerense</i>	AJ746070.2	JF491330.1	JXST01000003.1
101	<i>M. longobardum</i>	JN571166.1	JN571199.1	LQPG01000033.1
102	<i>M. lutetiense</i>	KP676904.1	KP676906.1	
103	<i>M. madagascariense</i>	NR_104690.1	AF547852.1	
104	<i>M. mageritense</i>	AJ699399.1	AY458070.1	AY147169.1
105	<i>M. malmoense</i>	NR_044814.1	JF491301.1	MVHV01000002.1
106	<i>M. mantonii</i>	FJ042897.1	FJ232523.1	MVHW01000003.1
107	<i>M. marinum</i>	AF456240.1	AY299134.1	CP000854.1
108	<i>M. marseillense</i>	EU266632.1	EU239787.1	MVHX01000018.1
109	<i>M. microti</i>	NR_025234.1	AY299135.1	CP010333.1

110	<i>M. minnesotense</i>	NZ_MVHZ01000040.1	JN546612.1	NZ_MVHZ01000002.1
111	<i>M. monacense</i>	AF107039.2	JF491320.1	MVIA01000030.1
112	<i>M. montefiorensis</i>	NR_028808.1	AY943204.1	
113	<i>M. montmartrensis</i>	KP676919.1	KP676921.1	
114	<i>M. moriokaense</i>	AJ429044.1	AF547857.1	MVIB01000024.1
115	<i>M. mucogenicum</i>	AY457074.1	AY458079.1	AY147170.1
116	<i>M. murale</i>	NR_104691.1	AF547859.1	
117	<i>M. nebraskense</i>	AY368456.1	AY368457.1	LQPH01000173.1
118	<i>M. neoaurum</i>	FJ172306.1	JF491302.1	LK021342.1
119	<i>M. neworleansense</i>	AY457068.1	AY458076.1	CWKH01000001.1
120	<i>M. nonchromogenicum</i>	NR_044813.1	JF491303.1	LQPI01000033.1
121	<i>M. noviomagensis</i>	EU239955.1	EU600390.1	MVIC01000024.1
122	<i>M. novocastrensis</i>	NR_029208.1	AF547862.1	BCTA01000056.1
123	<i>M. obuense</i>	NR_029218.1	AF547863.1	JYNU01000057.1
124	<i>M. oryzae</i>	LN846834.1	LN886510.1	
125	<i>M. pallens</i>	DQ370008.1	DQ533997.1	
126	<i>M. palustre</i>	NR_028940.1	AY943200.1	LQPJ01000147.1
127	<i>M. paraense</i>	KJ948996.1	HM056137.1	LQPM01000030.1
128	<i>M. paraffinicum</i>	GQ153270.1	GM153287.1	
129	<i>M. parafortuitum</i>	NR_026285.1	AF547864.1	MVID01000028.1
130	<i>M. paragordoniae</i>	KC525204.1	KC525205.1	
131	<i>M. paraintracellulare</i>	NR_146658.1	KP670342.1	NCXN01000038.1
132	<i>M. parakorensis</i>	JF271823.1	JF271824.1	
133	<i>M. parascrofulaceum</i>	AY337273.1	AY337274.1	JF271799.1
134	<i>M. paraseoulense</i>	DQ536404.1	JF491324.1	MVIE01000010.1
135	<i>M. paraterrae</i>	EU919229.1	EU919228.1	
136	<i>M. parmense</i>	NR_025225.1	HM022199.1	LQPO01000084.1
137	<i>M. peregrinum</i>	AY457069.1	AY458069.1	AY147166.1
138	<i>M. phlei</i>	NR_041906.1	AY299158.1	CP014475.1
139	<i>M. phocaicum</i>	AY859682.1	AY859676.1	AY859693.1
140	<i>M. pinnipedii</i>	NR_025249.1	EF656462.1	
141	<i>M. porcinum</i>	AY457077.1	JF491326.1	AY262737.1
142	<i>M. poriferae</i>	NR_025235.1	AF547868.1	AY544956.1
143	<i>M. pseudoshottsii</i>	NR_113136.1	AM902956.1	AP018410.1
144	<i>M. psychrotolerans</i>	AJ534886.1	HM602035.1	
145	<i>M. pulveris</i>	AJ429046.1	AF547869.1	
146	<i>M. pyrenivorans</i>	AJ431371.1	JF510463.1	
147	<i>M. rhodesiae</i>	AJ429047.1	AF547870.1	MVIH01000008.1
148	<i>M. riyadhense</i>	EU274642.1	EU921671.1	LQPQ01000029.1
149	<i>M. rufum</i>	JROA01000001.1	DQ841181.1	JROA01000001.1
150	<i>M. rutilum</i>	DQ370011.1	DQ841180.1	LT629971.1
151	<i>M. salmoniphilum</i>	NR_043989.1	DQ866777.1	
152	<i>M. saopaulense</i>	NZ_MVH01000006.1	KM973026.1	MVII01000002.1
153	<i>M. sarraceniae</i>	KM216315.3	KM216313.1	
154	<i>M. saskatchewanense</i>	NR_042793.1	JF491331.1	LQPR01000076.1
155	<i>M. scrofulaceum</i>	NR_025237.1	GQ153288.1	MVIJ01000008.1
156	<i>M. sediminis</i>	KC010490.1	KC010486.1	
157	<i>M. senegalense</i>	AY457081.1	JF491327.1	AY262738.1
158	<i>M. senuense</i>	DQ536408.1	JF491328.1	LQPS01000051.1
159	<i>M. seoulense</i>	DQ536403.2	JF491322.1	
160	<i>M. septicum</i>	AY457070.1	JF491329.1	AY147167.1
161	<i>M. setense</i>	EU371504.1	EU371505.1	JTJW01000005.1
162	<i>M. sherrisii</i>	NR_115209.1	AY365190.1	LQPT01000069.1
163	<i>M. shimoidei</i>	NR_041945.1	JF491305.1	LQPU01000062.1
164	<i>M. shinjukuense</i>	AB268503.2	AB268505.2	MVIK01000002.1
165	<i>M. shottsii</i>	AY005147.3	AY550225.1	
166	<i>M. simiae</i>	NR_117227.1	JF491306.1	CP010996.1
167	<i>M. smegmatis</i>	AJ131761.1	JF491307.1	AY262735.1

---

168	<i>M. sphagni</i>	FR733719.1	AF547877.1	NOZR01000019.1
169	<i>M. stomatepiae</i>	AM884331.1	AM902968.1	
170	<i>M. szulgai</i>	NR_118584.1	JF491308.1	LQPW01000031.1
171	<i>M. terrae</i>	NR_118881.1	AF434736.1	LQPX01000002.1
172	<i>M. thermoresistibile</i>	X55602.1	AF547880.1	AGVE01000046.1
173	<i>M. timonense</i>	NR_104956.1	HM602038.1	MVIL01000032.1
174	<i>M. tokaiense</i>	NR_025236.1	JF491309.1	AY544969.1
175	<i>M. triplex</i>	U57632.1	GQ153291.1	HG964446.1
176	<i>M. triviale</i>	NR_119192.1	JF491310.1	LQPZ01000021.1
177	<i>M. tuberculosis</i>	NR_102810.2	NC_000962	AL123456.3
178	<i>M. tusciae</i>	AF058299.1	AF547887.1	MVIM01000003.1
179	<i>M. ulcerans</i>	AB548725.1	AY299145.1	CP000325.1
180	<i>M. vaccae</i>	NR_041899.1	JF491312.1	JH814687.1
181	<i>M. vanbaalenii</i>	NR_074572.1	AY438091.1	CP000511.1
182	<i>M. vulneris</i>	EU834055.1	EU834054.1	NCXM01000019.1
183	<i>M. wolinskyi</i>	AY457083.1	AY299164.1	LQQA01000030.1
184	<i>M. xenopi</i>	NR_118882.1	AF434738.1	LQQB01000145.1
185	<i>M. yongonense</i>	JF738056.1	JN605801.1	CP003347.1