

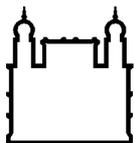
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Mestrado em Programa de Pós-Graduação Biologia Computacional e Sistemas

**DIVERSIDADE, TAXONOMIA GENÔMICA E RESISTOMA DE MICOBACTÉRIAS
DA MATA ATLÂNTICA**

SERGIO MASCARENHAS MORGADO

Rio de Janeiro
Fevereiro de 2017



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

Sergio Mascarenhas Morgado

Diversidade, taxonomia genômica e resistoma de micobactérias da Mata Atlântica

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Biologia Computacional e Sistemas

Orientadores: Prof. Dra. Ana Carolina Paulo Vicente
Prof. Dr. Michel Francisco Abanto Marín

RIO DE JANEIRO

Fevereiro de 2017

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

M847 Morgado, Sergio Mascarenhas

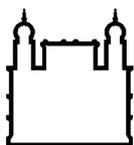
Diversidade, taxonomia genômica e resistoma de micobactérias da Mata Atlântica / Sergio Mascarenhas Morgado. – Rio de Janeiro, 2017. xviii, 133 f. : il. ; 30 cm.

Dissertação (Mestrado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2017.

Bibliografia: f. 62-75

1. Micobactéria. 2. Taxonomia genômica. 3. Mobiloma. 4. Resistoma. 5. Diversidade de espécies. 6. Mata Atlântica. I. Título.

CDD 579.5



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: SERGIO MASCARENHAS MORGADO

Diversidade, taxonomia genômica e resistoma de micobactérias da Mata Atlântica

**ORIENTADORES: Prof. Dra. Ana Carolina Paulo Vicente
Prof. Dr. Michel Francisco Abanto Marín**

Aprovada em: ____/____/____

EXAMINADORES:

Prof. Dr. Wim Maurits Sylvain Degrave	(IOC)
Prof. Dra. Cristiane Carneiro Thompson	(UFRJ)
Prof. Dr. Rodolpho Mattos Albano	(UERJ)
Prof. Dra. Adriana Machado Fróes	(UFRJ)
Prof. Dra. Verônica Viana Vieira	(IOC)

Rio de Janeiro, 21 de Fevereiro de 2017

AGRADECIMENTOS

Aos meus orientadores Dra. Ana Carolina e Dr. Michel Abanto por todo auxílio, paciência e motivação no decorrer do estudo.

À Dra. Érica Lourenço da Fonseca pelo auxílio e participação em valiosas discussões. Minha terceira orientadora.

À todos os membros e ex-membros do LGMM, que conheci durante o estudo, pela amizade e auxílios diversos.

Aos membros da banca por aceitarem o convite.

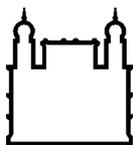
À minha família e amigos por toda ajuda e compreensão pelas minhas ausências, devido aos estudos.

Aos professores do Instituto Oswaldo Cruz pelo conhecimento compartilhado.

À pós-graduação pelo auxílio na participação de congressos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo auxílio financeiro.

**"A ciência nunca resolve um problema
sem criar pelos menos outros dez"
George Bernard Shaw**



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Diversidade, taxonomia genômica e resistoma de micobactérias da Mata Atlântica

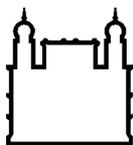
RESUMO

DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Sergio Mascarenhas Morgado

O gênero *Mycobacterium* é diverso e ubíquo no ambiente, com quase 200 espécies descritas. As espécies são classificadas como de crescimento rápido ou lento e, em ambos os grupos, existem espécies patogênicas e patógenos oportunistas. O solo da Mata Atlântica apresenta uma significativa diversidade bacteriana e uma pequena parte dessa diversidade está preservada na Coleção de Bactérias da Mata Atlântica - CBMA/FIOCRUZ, incluindo isolados do gênero *Mycobacterium*. Neste estudo caracterizamos a diversidade de micobactérias da coleção CBMA utilizando a informação genômica global, tendo como objetivo sua taxonomia e caracterização do resistoma. Utilizando a taxonomia genômica foram caracterizadas cinco novas espécies de micobactérias. Estas novas espécies são filogeneticamente relacionadas com as espécies *M. septicum*, *M. llatzerense* e *M. abscessus*. A análise das sequências genômicas permitiu também a caracterização do mobiloma, que compreende dois micobacteriófagos completos e cinco plasmídeos. A maioria desses elementos está nos genomas dos isolados CBMA 311/312/360, pertencentes a uma das novas espécies. Estes elementos são únicos considerando o repertório de mobilomas das micobactérias até agora identificado. Este estudo corrobora o solo da Mata Atlântica como um reservatório da diversidade de micobactérias, uma vez que, pelo menos cinco novas espécies e micobacteriófagos deste gênero, foram caracterizados. As análises *in silico* do resistoma identificaram de dezenas a centenas de genes/mecanismos relacionados à resistência aos antibióticos, dependendo da base de dados utilizada. Para associar genótipo com fenótipo, genes relacionados a três mecanismos de resistência foram analisados: proteção de alvo (genes *mfpA/B*), modificação de antibióticos (gene *arr*) e degradação (gene de beta-lactamase). As análises *in vitro* destes genes, em

sistema heterólogo, mostraram que estes alelos *mfpA/B* conferiam uma redução da sensibilidade às quinolonas; e o gene da beta-lactamase (*bla326*) apresentou um espectro de atividade restrito aos beta-lactâmicos; e o gene *arr* não alterou o perfil de sensibilidade à rifampicina. Portanto, para associar um gene/mecanismo à um perfil de resistência a antibióticos, é importante que a predição pela bioinformática, mesmo baseada em análises de genoma completo, seja validada com ensaios *in vitro*.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Diversity, genomic taxonomy and resistome of Atlantic Forest mycobacteria

ABSTRACT

MASTER DISSERTATION IN COMPUTATIONAL AND SYSTEMS BIOLOGY

Sergio Mascarenhas Morgado

The genus *Mycobacterium*, ubiquitous in the environment, presents a high diversity, with almost 200 species described. These species are classified into fast- and slow-growing organisms and, in both groups, there are pathogens and opportunistic species. The Brazilian Atlantic Forest soil, a hot spot of bacteria diversity, has part of this diversity preserved in the CBMA/FIOCRUZ collection which includes isolates from *Mycobacterium* genus. Here, we characterized the diversity of mycobacteria isolates from CBMA Collection with focus on the taxonomy and the resistome using the whole genome information. Applying genomic taxonomy approach five new mycobacteria species were characterized. These new species are phylogenetically related to *M. septicum*, *M. llatzerense* and *M. abscessus*. The whole genomic sequences analyses allowed also the mobilome identification. It was characterized by two complete phages and four plasmids, most of these elements harbored by CBMA311/312/360 that belong to one of the new species. Considering the mycobacteria mobilome repertoire so far identified, these elements are unique. These findings corroborated the Brazilian Atlantic Forest soil as a hot spot of bacteria diversity whereas at least five new *Mycobacterium* species as well as mycobacteriophages were characterized. The resistome *in silico* analyses identified from tens to hundreds genes/mechanisms related with antibiotics resistance, depending on the database used. In order to associate genotype with phenotype, we analyzed genes related to three resistance mechanisms: target protection (*mfpA/B* genes), antibiotic modification (*arr* gene) and degradation (beta-lactamase gene). *in vitro* analyses of these genes using heterologous systems showed that these *mfpA/B* alleles conferred a reduction of sensibility to quinolones; and the beta-lactamase (*bla326*) gene presented a restricted spectrum of activity to beta-lactams and the *arr* gene did not alter the sensibility profile to rifampicin. Therefore, in order to address the role of a gene/mechanism to an antibiotic resistance profile it is important to associate the bioinformatic prediction, based on whole genome analyses, with *in vitro* assays.

ÍNDICE

1	INTRODUÇÃO	1
1.1	Micobactéria – Características gerais	1
1.2	Diversidade do gênero <i>Mycobacterium</i>	2
1.3	Genomas das micobactérias	4
1.4	Taxonomia de bactérias.....	6
1.4.1	<i>Taxonomia genômica</i>	7
1.5	Resistoma	8
1.5.1	<i>Caracterização genômica do resistoma</i>	10
2	OBJETIVOS	11
2.1	Objetivo Geral	11
2.2	Objetivos Específicos	11
3	MATERIAL E MÉTODOS	12
3.1	Isolados e condições de crescimento	12
3.2	Extração de DNA total	12
3.3	PCR e Sequenciamento de genes específicos	12
3.4	Análise de multilocus (MLSA)	13
3.5	Sequenciamento genômico por alto desempenho	14
3.6	Montagem e anotação dos genomas	14
3.7	Genomas públicos de micobactérias utilizados	16
3.8	Taxonomia genômica.....	20
3.9	Identificação do resistoma	20
3.10	Identificação de elementos extra-cromossômicos/mobiloma.....	21
3.11	Clonagem e expressão heteróloga	21
3.12	Testes de susceptibilidade aos antimicrobianos	22
4	RESULTADOS	24
4.1	Diversidade genética das micobactérias da Mata Atlântica	24
4.2	Montagem e anotação dos genomas	29
4.3	Identificação do mobiloma	36
4.3.1	<i>Bacteriófagos</i>	36
4.3.2	<i>Plasmídeos</i>	37

4.4	Taxonomia genômica das micobactérias da Mata Atlântica	38
4.5	RESISTOMA	47
4.5.1	<i>Predição in silico do resistoma</i>	47
4.5.2	<i>Perfil de sensibilidade das micobactérias CBMA.....</i>	47
4.5.3	<i>Associação resistoma (predição in silico a partir do genótipo) versus perfil de sensibilidade (fenótipo in vitro).....</i>	47
4.5.4	<i>Funcionalidade dos genes associados às resistências em sistema heterólogo.....</i>	52
4.5.4.1	<i>Resistência a rifampicina</i>	52
4.5.4.2	<i>Resistência a quinolonas</i>	53
4.5.4.3	<i>Resistência a beta-lactâmicos.....</i>	56
5	DISCUSSÃO	57
6	CONCLUSÕES	61
7	REFERÊNCIAS BIBLIOGRÁFICAS	62

LISTA DE FIGURAS

Figura 1.1. Imagens de microscopia eletrônica de <i>M. tuberculosis</i> (A) e <i>M. smegmatis</i> (B).....	2
Figura 1.2. Árvore filogenética baseada no gene 16S rRNA de espécies de micobactérias.....	3
Figura 1.3. Atlas genômico de diferentes espécies de micobactérias.....	5
Figura 1.4. Principais mecanismos de resistência bacteriana.....	9
Figura 4.1. Árvore de máxima verossimilhança de sequências 16S (1.185 pb) dos isolados CBMA e do banco de dados RDP.....	26
Figura 4.2. Árvore de máxima verossimilhança da abordagem MLSA dos isolados de micobactéria da Mata Atlântica.....	27
Figura 4.3. Categorização funcional dos genes do isolado CBMA 226.....	32
Figura 4.4. Categorização funcional dos genes do isolado CBMA 234.....	33
Figura 4.5. Categorização funcional dos genes do isolado CBMA 247.....	33
Figura 4.6. Categorização funcional dos genes do isolado CBMA 311.....	34
Figura 4.7. Categorização funcional dos genes do isolado CBMA 312.....	34
Figura 4.8. Categorização funcional dos genes do isolado CBMA 326.....	35
Figura 4.9. Categorização funcional dos genes do isolado CBMA 360.....	35
Figura 4.10. Mapa genômico de P4.....	38

Figura 4.11. Árvore de máxima verossimilhança da abordagem MLSA.....	39
Figura 4.12. Árvore de máxima verossimilhança da abordagem cgMLSA.....	40
Figura 4.13. Resolução taxonômica entre isolados CBMA.....	45
Figura 4.14. Resolução taxonômica entre isolados CBMA e espécies filogeneticamente próximas.....	46
Figura 4.15. Heatmap da associação do resistoma versus fenótipo de resistência das micobactérias CBMA.....	50
Figura 4.16. Alinhamento de aminoácidos codificados pelo gene arr de <i>M. smegmatis</i> e isolados CBMA.	53
Figura 4.17. Ambientes genéticos do gene arr nos isolados CBMA 226, 234 e 326.....	53
Figura 4.18. Alinhamento de aminoácidos da região QRDR de GyrA dos isolados CBMA.....	54
Figura 4.19. Alinhamento de aminoácidos da região QRDR de GyrB dos isolados CBMA.....	54
Figura 4.20. Árvore de máxima verossimilhança da proteína MfpA de micobactérias.....	55
Figura 4.21. Árvore de máxima verossimilhança da proteína MfpB de micobactérias.....	55

LISTA DE TABELAS

Tabela 3.1. Iniciadores utilizados no estudo.....	13
Tabela 3.2. Genomas utilizados no estudo obtidos a partir do Genbank.....	17
Tabela 3.3. Antibióticos usados nos testes de susceptibilidade.....	22
Tabela 4.1. Matriz de identidade das sequências 16S dos isolados CBMA e de sequências do banco de dados RDP próximas aos dos isolados CBMA.....	25
Tabela 4.2. Matriz de identidade das sequências concatenadas de cinco genes parciais (2801 pb) dos isolados CBMA e de bancos de dados mais próximas aos dos isolados CBMA.....	28
Tabela 4.3. Comparação das métricas obtidas nas montagens.....	30
Tabela 4.4. Métricas finais das montagens dos genomas CBMA.....	30
Tabela 4.5. Anotação geral a partir do Prokka.....	31
Tabela 4.6. Número de genes de duas categorias funcionais versus genomas CBMA.....	32
Tabela 4.7. Número de genes relacionados ao mobiloma versus genomas CBMA.....	36
Tabela 4.8. Plasmídeos putativos preditos nos genomas CBMA.....	37
Tabela 4.9. Comparação de valores DDH entre genomas CBMA e espécies próximas, considerando cgMLSA.....	41
Tabela 4.10. Comparação dos valores de AF e gANI.....	42

Tabela 4.11. Número de genes associados à resistência nas diferentes bases de dados.....	47
Tabela 4.12. E-test realizado nos isolados CBMA.....	49
Tabela 4.13. Fenótipo e genes de resistência identificados.....	51
Tabela 4.14. Perfil de resistência a quinolonas das transformantes carregando os diferentes alelos mfpA/B.....	56

LISTA DE SIGLAS E ABREVIATURAS

AAI: identidade média de aminoácido
AF: fração de alinhamento
AK: amicacina
AMO: Amoxicilina
ANI: identidade média de nucleotídeo
arr: ADP-ribosiltransferase
ATM: aztreonam
AZM: azitromicina
B: bacitracina
bla: beta-lactamase
BLAST: Basic Local Alignment Search Tool
BLASTn: Basic Local Alignment Search Tool nucleotide
CARD: Comprehensive Antibiotic Resistance Database
CAZ: ceftazidime
CBAS: Coleção de Bactérias do Ambiente e Saúde
CBMA: Coleção de Bactérias da Mata Atlântica
C: cloranfenicol
CDS: Coding DNA Sequence
CFL: cefalotina
CFO: cefoxitina
cgMLSA: core genome multilocus sequence analysis
CIP: ciprofloxacina
CLR: claritromicina
CN: gentamicina
CPM: cefepime
CTAB: cetyltrimethylammonium bromide
CT: colistina
CXM: cefuroxima
DA: clindamicina
DB: database (base de dados)
DDH: hibridação DNA-DNA
DNA: ácido desoxirribonucleico
dNTP: desoxirribonucleotídeos trifosfatados

ERI: eritromicina
E-test: epsilon test
FOS: fosfomicina
gANI: identidade média de nucleotídeo global
GTR: Generalised time reversible
gyrA: girase subunidade A
gyrB: girase subunidade B
HGT: transferência horizontal de genes
HSP: high-scoring segment pairs
hsp65: heat shock protein 65 kDa
IPM: imipinem
isDDH: hibridação DNA-DNA *in silico*
kb: kilobase
K: canamicina
LVX: levofloxacina
Mb: Megabase
MEM: meropenem
MET: metilina
mfpA: *Mycobacterium* fluoroquinolone resistance protein A
mfpB: *Mycobacterium* fluoroquinolone resistance protein B
MgCl₂: Cloreto de magnésio
MHA: Ágar Müller Hinton
MIC: Concentração Mínima Inibitória
MiSI: Identificador de Espécies Microbianas
MLSA: multilocus sequence analysis
MLST: multilocus sequence typing
mob: genes de mobilização
MY: lincomicina
NAL: Ácido Nalidíxico
N: neomicina
NOR: norfloxacina
pb: pares de bases
PCR: Polymerase Chain Reaction (Reação em cadeia da polimerase)
PRL: piperacilina
QRDR: quinolone resistance determining region

RDP: Ribosomal Database Project
repA: replication protein A
RIF: rifampicina
rpoB: RNA polimerase subunidade β
RRDR: rifampicin resistance determining region
rRNA: ácido ribonucleico ribossomal
S: estreptomicina
T7SS: type VII secretion system
TE: tetraciclina
TGC: tigeciclina
TOB: tobramicina
tRNA: ácido ribonucleico transportador
U: unidades
VA: vancomicina
Vs: versus
WGS: whole genome sequence

1 INTRODUÇÃO

1.1 Micobactéria – Características gerais

As micobactérias são bacilos aeróbios obrigatórios, Gram-positivos, desprovidos de flagelo, e que não formam esporos. Sua parede celular é constituída principalmente por peptidoglicanos, que contêm ácido N-glicolil-murâmico, polissacarídeos e ácido micólico, que caracteriza as micobactérias (Ryan & Ray 2003). Esses componentes conferem um caráter hidrofóbico à parede celular e portanto uma baixa permeabilidade, que dificulta a absorção de nutrientes e limitando o crescimento do bacilo. A baixa permeabilidade da membrana resulta em uma resistência natural a vários antimicrobianos, principalmente os hidrofílicos (Rastogi et al. 1981).

As micobactérias são classificadas em dois grupos: as de crescimento rápido e as de crescimento lento. As espécies de crescimento rápido produzem colônias visíveis em até sete dias, após o inóculo, enquanto que as espécies de crescimento lento levam mais de sete dias para produzirem colônias visíveis. A maioria das micobactérias patogênicas, entre as quais *Mycobacterium tuberculosis* (Figura 1.1A) e *Mycobacterium leprae*, são do grupo de crescimento lento, sendo denominadas típicas. A maioria das espécies ambientais pertencem ao grupo de crescimento rápido, como a *Mycobacterium smegmatis* (Figura 1.1B) e são denominadas atípicas (Stahl & Urbance 1990).

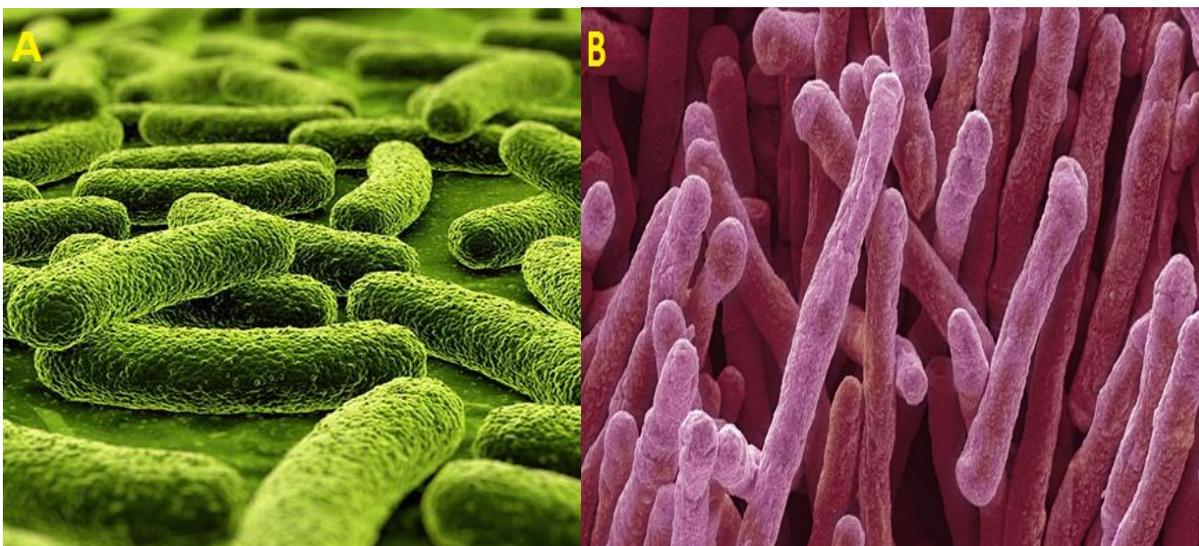


Figura 1.1. Imagens de microscopia eletrônica de *M. tuberculosis* (A) e *M. smegmatis* (B). Fontes: <http://williamrjacobs.org>; <http://media.gettyimages.com/videos/mycobacterium-smegmatis-bacteria-sem-video-id618596807?s=640x640>

As micobactérias atípicas são classificadas como saprófitas, comensais ou simbiotes; e possuem uma diversidade metabólica que possibilita sua sobrevivência e crescimento em diferentes condições como variação de temperatura, pH e depleção de nutrientes (Nyka 1974, Hall-Stoodley & Lappin-Scott 1998, Primm et al. 2004). Dentre as espécies deste grupo, algumas são eventuais patógenos, como por exemplo *Mycobacterium chelonae* (Umrao et al. 2016), *Mycobacterium llatzerense* (Cárdenas et al. 2014) e *Mycobacterium smegmatis* (Saffo & Ognjan 2016).

1.2 Diversidade do gênero *Mycobacterium*

Estudos de diversidade genética de bactérias, consideram inicialmente, a análise da sequência do gene 16S rRNA. Valores de identidade < 98.7% deste gene, entre isolados, seria indicativo de diferentes espécies, enquanto que valores < 95% indicariam um novo gênero (Stackebrandt & Ebers 2006).

Em estudos de diversidade e evolução de bactérias, particularmente isolados de uma mesma espécie, a abordagem aplicada é a análise de sequência multilocus (MLSA), que se baseia em vários genes intrínsecos codificantes de proteínas (*housekeeping genes*) (Konstantinidis & Tiedje 2005, Glaeser & Kämpfer 2015).

Muitos estudos realizavam análises filogenéticas a partir de um único gene, como por exemplo, o gene *hsp65*, que era o único utilizado em análises filogenéticas e até na definição de espécies em micobactéria (Telenti et al. 1993, Steingrube et al. 1995). Contudo, as filogenias baseadas em sequências de vários genes concatenados refletiriam com mais confiança a relação entre os organismos. Ainda assim, a definição dos genes para a análise multilocus deve avaliar eventos de recombinação, para que genes que estejam envolvidos nestes eventos não sejam incorporados à análise (Glaeser & Kämpfer 2015)

O gênero *Mycobacterium* é diverso (Figura 1.2), até 2016, mais de 175 espécies tinham sido descritas (<http://www.bacterio.net/>), sendo que destas, pouco mais de 100 tiveram seu genoma completo sequenciado e disponibilizado no Genbank.

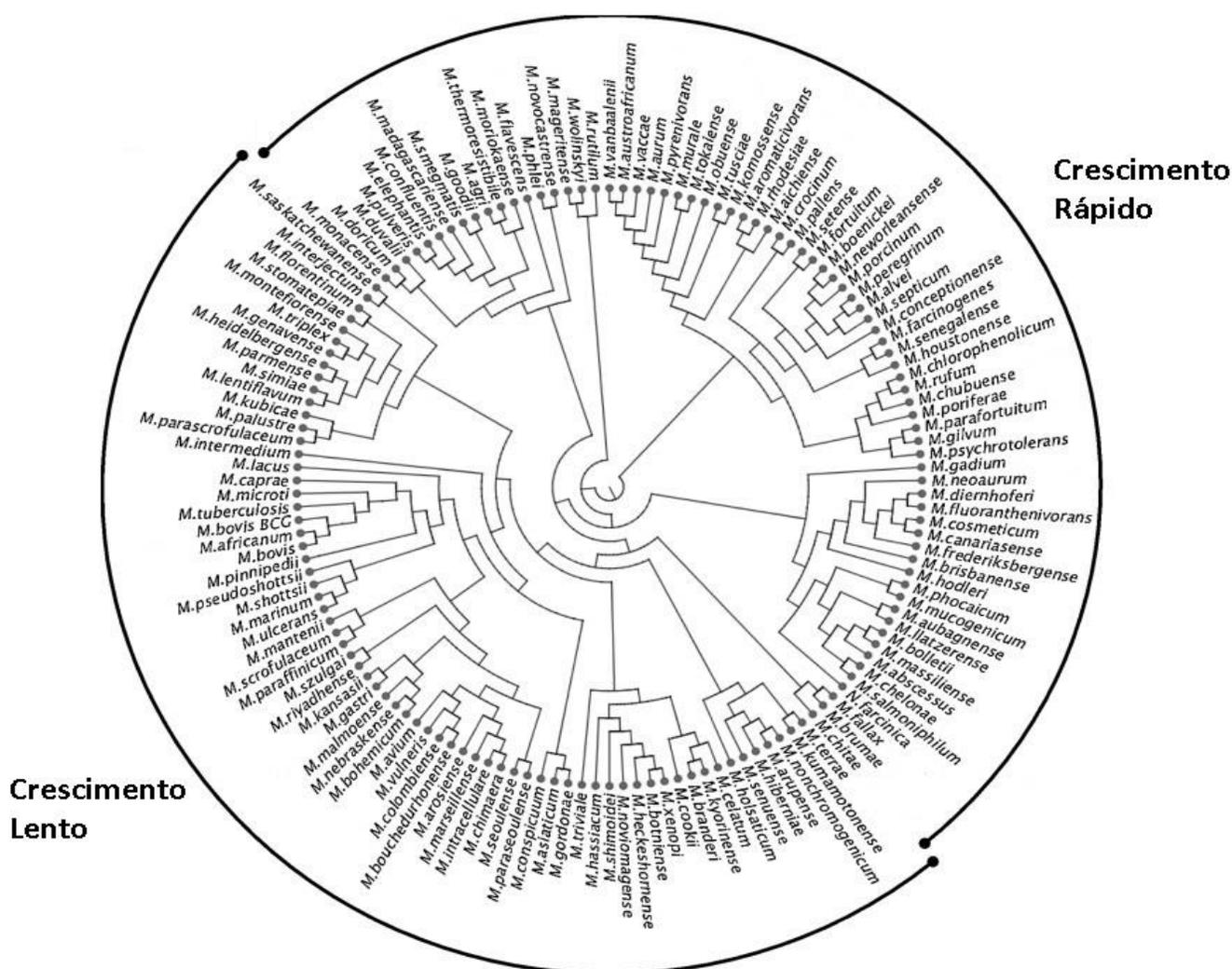


Figura 1.2. Árvore filogenética baseada no gene 16S rRNA de espécies de micobactérias (Noguera-Ortega et al. 2016-modificado pelo autor).

A diversidade das espécies do gênero se reflete na ampla variedade de nichos, incluindo solos, ambientes aquáticos e organismos (animais, plantas e humanos), que são explorados por esses organismos. Isto pode estar relacionado tanto a uma diversidade metabólica, quanto a especialização de algumas das espécies (Honda et al. 2016; Zhou et al. 2016; Chin'ombe et al. 2016; Ahmed et al. 2017; Azadi et al. 2016). Além da diversidade entre as espécies, isolados de uma mesma espécie também apresentam variabilidade. Por exemplo, o complexo abscessus, que compreende as espécies *Mycobacterium abscessus subsp. abscessus*, *Mycobacterium abscessus subsp. bolletii* e *Mycobacterium abscessus subsp. massiliense*, apresentam diferentes perfis de resistência a antibióticos (Nessar et al. 2012). As subespécies do complexo são caracterizadas pelo esquema MLST (*Multilocus Sequence Typing*), que considera alelos de um conjunto de genes ubíquos na espécie (Pérez-Losada et al. 2013).

A diversidade de espécies também se reflete nas associações a infecções, por exemplo, subespécies do complexo *abscessus* estão relacionadas a infecções em humanos; enquanto que *Mycobacterium marinum*, infecta peixes (Velu et al. 2016); *Mycobacterium avium/canariense*, aves (Parvandar et al. 2016); e *Mycobacterium bovis/caprae*, ruminantes (Pesciaroli et al. 2014).

A diversidade metabólica e de espécies impacta também na biotecnologia, já que muitas espécies, como *M. smegmatis*, *M. neoaurum*, *M. gilvum* e *M. vanbaalenii*, tem a capacidade de biorremediação de ambientes impactados por hidrocarbonetos, além de serem utilizadas na produção de biomoléculas (Lee et al. 2015, Fernández-Cabezón et al. 2016, Galán et al. 2016, Ma et al. 2016, Muangchinda et al. 2016).

1.3 Genomas das micobactérias

Considerando as quase duas centenas de espécies do gênero *Mycobacterium*, a maioria dos mais de 6000 genomas deste gênero já disponibilizados pertence a isolados de *M. tuberculosis*. Esta é a espécie do gênero mais estudada em função do seu impacto na saúde pública mundial, e em contraste com *M. leprae*, que também tem enorme impacto na saúde pública, é cultivável (Figura 1.3).

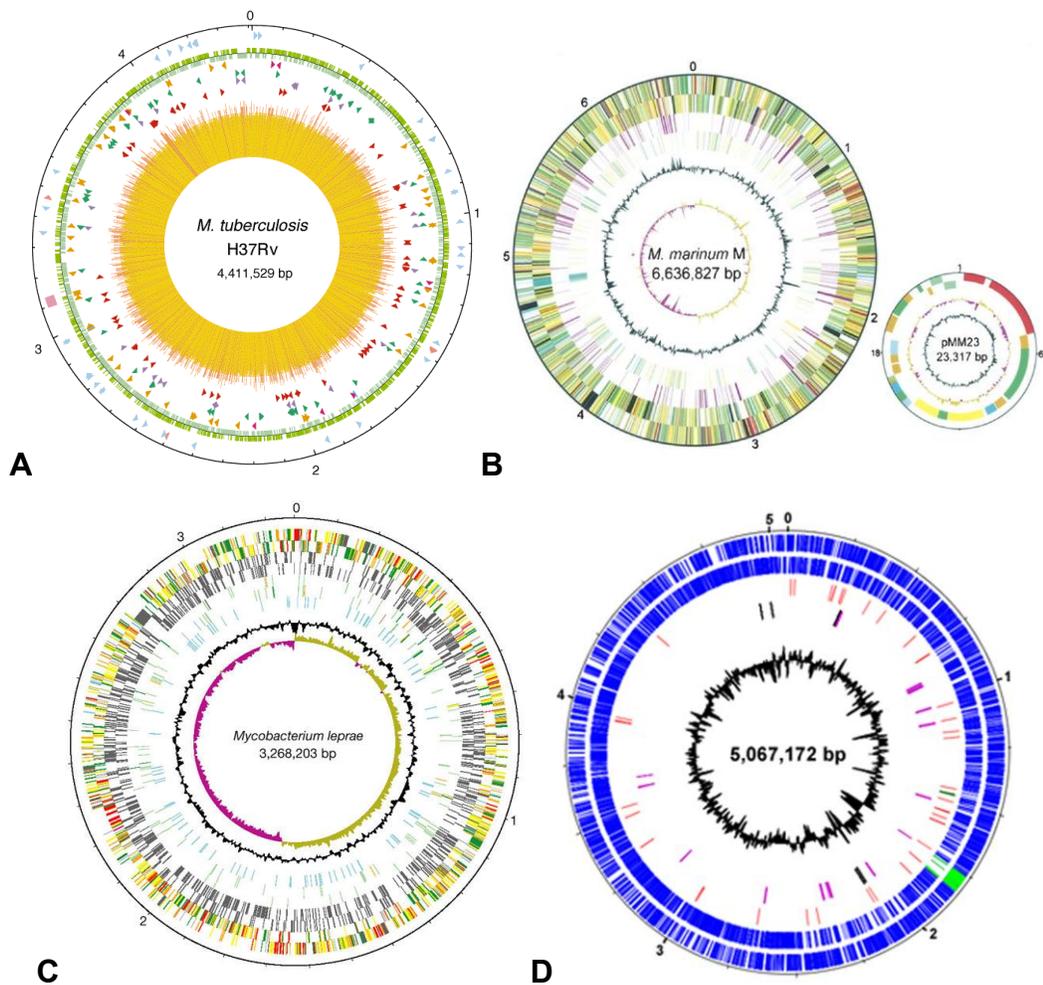


Figura 1.3. **Atlas genômico de diferentes espécies de micobactérias.** A, *M. tuberculosis* (Cole et al. 1998); B, *M. marinum*, com seu plasmídeo (Stinear et al. 2008 - modificado pelo autor); C, *M. leprae*, (Cole et al. 2001); D, *M. abscessus* (Ripoll et al. 2009 - modificado pelo autor).

Além dos genomas sequenciados de espécies já descritas, há ~162 genomas completos sequenciados de isolados do gênero *Mycobacterium* sem designação taxonômica a nível de espécie. Isso é um indicativo tanto da diversidade do gênero quanto da dificuldade da aplicação da abordagem de taxonômica tradicional para caracterização de isolados deste gênero.

Os genomas das espécies deste gênero apresentam tamanhos que variam de ~3.20 - 8 Mb, com uma média de 6.6/6.7 Mb, e conteúdo GC variando de 57 - 69%, com uma média de 66/67%. Os menores genomas ocorrem em *M. leprae* e espécies relacionadas, enquanto que os maiores em espécies atípicas. Os genomas deste

gênero compreendem um único cromossomo circular, que codificam ~1600 - 7000 genes; a maioria apresenta 1 a 2 operons ribossomais; e uma média de 50 tRNAs.

Plasmídeos, que fazem parte do genoma acessório, foram descritos em poucas espécies/linhagens do gênero. Plasmídeos lineares e circulares, com tamanho de ~7 - 615 kb; com conteúdo GC, ~64%; e raras cópias de tRNAs; foram caracterizados (<https://www.ncbi.nlm.nih.gov/genome/>).

1.4 Taxonomia de bactérias

A taxonomia é um ramo da ciência que envolve a identificação, classificação e a nomenclatura de organismos. A classificação em hierarquias foi proposta por Carolus Linnaeus, sendo "espécie" a de menor grau. Diferentes definições de espécie procariota têm sido propostos no decorrer dos anos, isso vem ocorrendo em função do desenvolvimento tecnológico. Hoje os taxonomistas de procariotos concordam que uma classificação confiável deve explorar a diversidade dos táxons através de várias abordagens, e isso é conhecido como a abordagem polifásica (Rosselló-Mora & Amann 2001). Esta abordagem integra informações bioquímicas, genotípicas, fenotípicas e quimiotaxonômicas de um organismo para, a partir de um consenso dos dados, delinear a espécie microbiana. Análises como hibridização DNA-DNA (DDH); conteúdo GC; comparação da sequência nucleotídica de marcadores taxonômicos, como 16S rRNA; identificação e caracterização de ácidos graxos; lipídeos polares; composição da parede celular; morfologia; vias metabólicas; permitem a definição de espécies (Colwell 1968, Vandamme et al. 1996)

Com o desenvolvimento e disponibilidade das tecnologias de sequenciamento genômico, surgiram novas possibilidades para a taxonomia de bactérias. A taxonomia genômica compara a informação genômica global de organismos, e já foi aplicada para descrever e reclassificar novas espécies dos gêneros *Burkholderia*, *Geobacter*, *Vibrio* e *Prochlorococcus*; além de auxiliar na caracterização de novas subespécies de *Francisella*; um novo gênero de *Sphaerochaeta*; e uma nova classe de *Dehalococcoidetes* (Mikalsen et al. 2007, Thompson et al. 2009, Vanlaere et al. 2009, Prakash et al. 2010, Löffler et al. 2012, Ritalahti et al. 2012, Thompson et al. 2013a).

1.4.1 Taxonomia genômica

A taxonomia baseada em genomas caracteriza as espécies a partir de métodos e parâmetros, tais como: *in silico* DDH (Auch et al. 2010a), identidade média de aminoácido (AAI) (Rohwer, Edwards 2002), identidade média de nucleotídeo (ANI) (Konstantinidis & Tiedje 2005), assinatura genômica de Karlin (Karlin & Burge 1995), análises de super árvores (Brown et al. 2001), viés de uso de códon (Wright 1990), conteúdos de vias metabólicas, análises do genoma *core* e árvores de famílias do pangenoma (Snipen & Ussery 2010), análise *in silico* do proteoma e características metabólicas derivadas de predições *in silico* (*genotype-to-phenotype*) (Dutilh et al. 2013, 2014, Amaral et al. 2014).

- ❖ A metodologia *in silico* DDH realiza comparações entre as sequências nucleotídicas de pares de genomas, através de buscas por BLAST, identificando segmentos com alta taxa de similaridade (HSPs). Os valores desses segmentos são então convertidos em um análogo ao da DDH *in vitro* (Auch et al. 2010a). Essa ferramenta apresentou resultados com alta correlação com os resultados *in vitro* (Auch et al. 2010b). A plataforma web para esta ferramenta se encontra no endereço: <http://ggdc.dsmz.de/>.
- ❖ A metodologia ANI é definida pela porcentagem média da identidade de sequências nucleotídicas de genes ortólogos compartilhados por dois genomas, e reproduz os resultados de DDH com acurácia. Valores de ANI $\geq 95\%$ corresponderiam a um valor de DDH $\geq 70\%$ (Goris et al. 2007). Por sua eficácia, já foi sugerido que o ANI poderia substituir análises DDH no futuro (Tindall et al. 2010). Recentemente, ANI foi aprimorado, agora como gANI (média de identidade nucleotídica genômica). Em conjunto com outro parâmetro, essa metodologia foi nomeada MiSI (Identificador de Espécies Microbianas) (Varghese et al. 2015). O método MiSI utiliza a identidade média nucleotídica com base no genoma total como medida de similaridade entre dois genomas, utilizando dois parâmetros, gANI e AF (*Alignment Fraction*). O gANI é calculado, para um par de genomas, pela média de identidade nucleotídica de genes ortólogos identificados

de maneira bidirecional. E como medida complementar, a fração de genes ortólogos é considerada. O AF mede a fração de genes ortólogos entre o par de genomas, e é aplicado como um filtro inicial ao gANI, variando de 0 a 1, onde 0 significa ausência de *hits* bidirecionais, e 1 corresponde a completa conservação dos genes codificantes de proteínas. Após esse filtro, a identificação final se dá pelo gANI, onde são computados as médias de identidade de ortólogos identificados por *hits* bidirecionais derivados de alinhamentos nucleotídeos que apresentam $\geq 70\%$ de identidade, e pelo menos 70% de cobertura em relação a menor sequência do par de ortólogos. Esse parâmetro avalia o grau de conservação entre os ortólogos, e varia de 0 a 100. A determinação de um valor de corte para designação de espécie foi avaliada realizando comparações entre 1.1 milhões de pares de genomas da mesma espécie, onde para 99.7% dos casos, pelo menos 60% do conteúdo gênico foi identificado em pares de ortólogos. Portanto, o valor de AF escolhido foi ≥ 0.6 . Os genomas de 31 espécies apresentaram $AF < 0.6$, e em alguns casos, apresentando 23% de pares de genes homólogos, como para *Clostridium botulinum*. Considerando o gANI, realizando essas comparações entre genomas da mesma espécie que apresentaram $AF \geq 0.6$, para 95% dos casos a identidade nucleotídica foi ≥ 96.5 . Portanto, um par de genomas que apresente um $AF \geq 0.6$ e $gANI \geq 96.5$ é considerado da mesma espécie. A plataforma web para esta ferramenta se encontra no endereço: <https://ani.jgi-psf.org/html/home.php> .

1.5 Resistoma

O resistoma compreende o conjunto de genes/mecanismos de resistência aos antibióticos presentes em um ambiente ou organismo. Seu estudo busca compreender a origem, evolução e emergência destas resistências (D'Costa et al. 2006).

A definição de resistência a antibiótico depende do objetivo de cada estudo, e pode ser dividida em três tipos: clínica, epidemiológica e operacional. A clínica é baseada em valores de corte da concentração mínima inibitória (MIC) de crescimento bacteriano, que são avaliadas quando há falha terapêutica em pacientes humanos. A epidemiológica estabelece resistência a antibióticos baseada no maior valor de MIC

obtido dentre uma população selvagem, de uma espécie em particular. Nessa definição estão quaisquer combinações de microrganismos e antibióticos, incluindo aqueles que não utilizados na clínica. A definição operacional é baseada na atividade de algum gene, já que compara uma linhagem referência com uma linhagem mutante ou contendo um gene adquirido por HGT (transferência lateral de genes). A linhagem é considerada resistente se apresentar, para um antibiótico específico, um valor de MIC superior ao da linhagem selvagem (Martínez et al 2015).

Vários mecanismos estão envolvidos com fenótipos de resistência aos antibióticos (Figura 1.4). Os principais seriam (i) alteração na permeabilidade da parede celular, impedindo o influxo de compostos; (ii) ação enzimática, que degrada compostos; (iii) bombas de efluxo, que transferem composto do meio intracelular para o extracelular; e (iv) alteração do sítio de ação, impedindo que o composto atue na molécula (Allen et al. 2010)

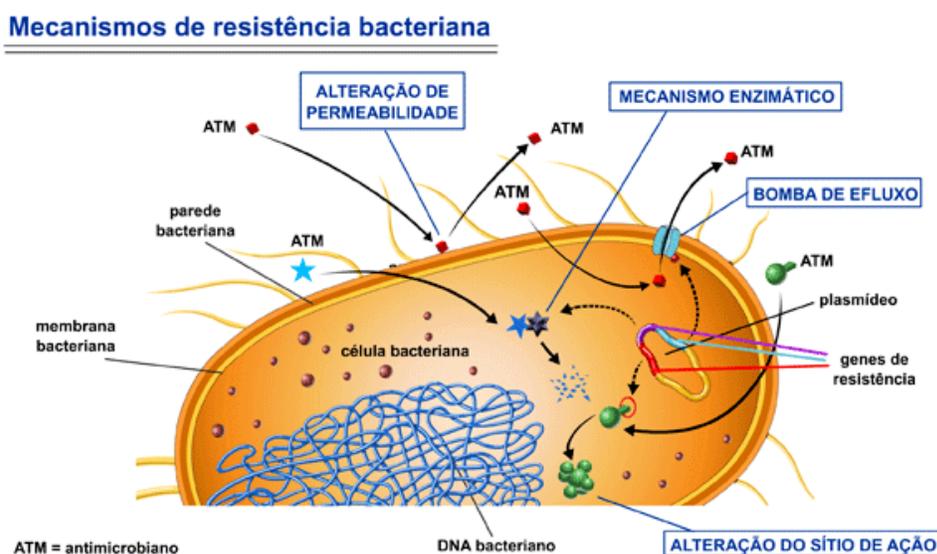


Figura 1.4. **Principais mecanismos de resistência bacteriana.** Fonte: http://www.anvisa.gov.br/servicosade/controlere/rede_rm/cursos/atm_racional/modulo1/res_principais2.htm .

Os genes/mecanismos de resistência presentes em um organismo podem ser (i) aqueles que foram selecionados e evoluem em decorrência da presença de antimicrobianos naturais; (ii) aqueles que são adquiridos por transferência lateral de genes. Estes geralmente codificam enzimas que degradam um determinado antibiótico.

Outros genes/mecanismos que caracterizariam o resistoma contribuem com o fenótipo de resistência de uma forma indireta. Por exemplo, bombas de efluxo de multidrogas são ubíquas e possuem uma vasta gama de substratos, incluindo antibióticos. Genes/mecanismos que são ubíquos, geralmente estão envolvidos em processos fisiológicos do organismo e a resistência aos antibióticos poderia ou não estar diretamente relacionada à sua presença.

1.5.1 Caracterização genômica do resistoma

A informação de genomas e metagenomas têm sido explorada no sentido de caracterizar a presença de genes/mecanismos relacionados à resistência a antibióticos, definido como resistoma. A caracterização dos resistomas, além de acessar a diversidade, busca caracterizar reservatórios ambientais destes elementos que podem vir eventualmente a emergir na clínica (Duranti et al. 2016, Martínez et al. 2016, Pawlowski et al. 2016, Boolchandani et al. 2017). Análises computacionais para predição de resistomas são baseadas em comparações por similaridade de sequência utilizando bases de dados especializadas (Liu & Pop 2009, McArthur et al. 2013, Gupta et al. 2014, Gibson et al. 2015, Wallace et al. 2017). Porém, a acurácia destas bases de dados ainda é comprometida devido a diversidade de mecanismos (intrínsecos/adquiridos), genes, alelos, regulação, entre outros relacionados a fenótipos de resistência (Moran et al. 2016). Assim, para evitar uma superestimação do resistoma de um genoma/metagenoma, análises funcionais devem ser implementadas considerando as informações recuperadas *in silico* (Martínez et al 2015).

2 OBJETIVOS

2.1 Objetivo Geral

Avaliar a diversidade de espécies e linhagens de micobactérias da Mata Atlântica, assim como seu resistoma, a partir de análises genômicas.

2.2 Objetivos Específicos

- Avaliar a diversidade das micobactérias pela abordagem de MLSA;
- Gerar genomas completos de isolados representantes desta diversidade;
- Aplicar a taxonomia genômica nestes isolados;
- Identificar *in silico* determinantes de resistência aos antimicrobianos;
- Determinar *in vitro* a funcionalidade destes genes.

3 MATERIAL E MÉTODOS

3.1 Isolados e condições de crescimento

As micobactérias (n=14) analisadas neste estudo encontram-se depositadas na Coleção de Bactérias do Ambiente e Saúde (CBAS) – Fiocruz, especificamente na subcoleção CBMA (Coleção de Bactérias da Mata Atlântica), e foram obtidas a partir de amostras do solo da Mata Atlântica. Os isolados criopreservados foram inoculados em meio de cultura líquido TSB à temperatura ambiente por cinco dias.

3.2 Extração de DNA total

Para o sequenciamento de alta vazão, o DNA total foi extraído pelo método CTAB (Petersen & Scheie 2000), com a incubação dos extratos bacterianos com 50 µl de lisozima (10 mg/ml) por 4h à 37 °C. Para a realização das reações de PCR (*Polymerase Chain Reaction*), a extração de DNA se deu por choque térmico. A massa celular foi ressuspensa em 300 µl de água Mili-Q estéril e incubada em banho-maria à aproximadamente 100 °C por 20 min, seguido de congelamento à -20 °C. Além do método CTAB, também utilizamos o kit QIAmp DNA Mini Kit - QIAGEN, de acordo com as instruções do fabricante.

3.3 PCR e Sequenciamento de genes específicos

Ensaio de PCR e sequenciamento foram realizados com o objetivo de recuperar sequências dos genes aplicados na análise de sequências multilocus (MLSA); e genes relacionados a resistência aos antimicrobianos. Os iniciadores utilizados estão listados na Tabela 3.1. Os iniciadores relativos aos genes de resistência antimicrobiana foram desenhados baseados nas sequências obtidas pelo sequenciamento do genoma completo. As reações foram compostas por tampão da PCR 1X (Promega), 3 mM MgCl₂ (Promega), 0.2 mM dNTP, 200 ng de cada iniciador, 1,5 U de Taq polimerase (Promega), 5 µl do DNA extraído por choque térmico, em um volume final de 50 µl. As

reações foram submetidas a um passo de desnaturação inicial a 94 °C por 5 min, e seguida de 40 ciclos de amplificação (desnaturação a 95 °C por 30 seg, anelamento a 55 °C por 30 seg, e extensão a 72 °C de 45 seg a 150 seg, dependendo do tamanho do fragmento); seguido por um passo de extensão final a 72 °C por 10 min. Os produtos da PCR foram submetidos a uma eletroforese em gel de agarose 1,5% e visualizados em transiluminador (KODAK Gel Logic Imaging System GL112). Os amplicons foram purificados utilizando o kit comercial Gel Band da GE Healthcare Life Sciences, e submetidos a reações de sequenciamento. Os amplicons tiveram as duas fitas sequenciadas utilizando o kit de sequenciamento BigDye (Life Technologies) e o sequenciador 3730XL - DNA Analyser Applied Biosystems (Plataforma de sequenciamento Sanger-FIOCRUZ).

Tabela 3.1. Iniciadores utilizados no estudo.

Iniciadores	Sequência (5' → 3')	Tamanho do produto PCR (pb)
16S-Pa F	AGAGTTTGATCCTGGCTCAG	1.500
16S-PH R	AAGGAGGTGATCCAGCCGCA	
<i>gyrB</i> F	GAGTTGGTGCGGCGTAAGAGC	218
<i>gyrB</i> R	CGGCCATCCARCACGATCTTG	
<i>rpoB</i> F	GGCAAGGTCACCCCGAAGGG	764
<i>rpoB</i> R	AGCGGCTGCTGGGTGATCATC	
<i>sodI</i> F	GAAGGAATCTCGTGGCTGAATAC	514
<i>sodI</i> R	AGTCGGCCTTGACGTTCTTGTAC	
<i>hsp65</i> -21M13F	TACCAACGATGGTGTGTCCAT	477
<i>hsp65</i> -M13R	CCCTTGTCGAACCGCATACCCT	
<i>mfpA</i> -MT360 F	CTGTGAGGAAACCACGTCTC	789
<i>mfpA</i> -MT360 R	CTGCCAATGACAAGTTCTG	
<i>mfpA</i> -MT247 F	TACTCGTTGAGCACGCTCA	750
<i>mfpA</i> -MT247 R	GCTGTCCGGTGTCTACGTCA	
<i>mfpB</i> -MT360 F	TCACGCAGGGTTATCTTCAA	788
<i>mfpB</i> -MT360 R	ATGGCCGTGCTGTATCAGT	
<i>mfpB</i> -MT247 F	AGGGTTATCTGAGGGTGGAA	812
<i>mfpB</i> -MT247 R	CTGAAATCGCACTCTGTGAAC	
<i>arr234</i> F	AGCATCGCTGAGTTCAAGG	630
<i>arr234</i> R	TTAGCTGTTTGACCCTGCTG	
<i>arr326</i> F	GAGATTTGTAGCGGCATGAG	650
<i>arr326</i> R	TGGTGATCTTCGTTGGACTC	
<i>bla326</i> F	GATCGACAAGACCGGAATG	1.220
<i>bla326</i> R	TAAGCATCTGGGTCACGTTT	

3.4 Análise de multilocus (MLSA)

As sequências parciais dos cinco genes (16S rRNA, *hsp65*, *sodA*, *rpoB*, *gyrB*), que fazem parte do esquema MLSA do gênero *Mycobacterium*, obtidas por PCR, foram

sequenciadas, editadas no Seqman v.7.0.0, alinhadas e concatenadas no Seaview v.4.6 (Gouy et al. 2010), e submetidas a análises genéticas com sequências homólogas de várias espécies de micobactérias. As análises do gene ribossomal 16S e MLSA foram realizadas usando o método de máxima verossimilhança, modelo GTR, com 100 replicatas de *bootstrap*. As matrizes de identidade foram geradas pelo Bioedit v.7.2.5.

A partir dessas análises, selecionamos os isolados que tiveram seu genoma completo sequenciado.

3.5 Sequenciamento genômico por alto desempenho

Um total de sete isolados representando a diversidade de Micobactérias da CBMA/2014 tiveram seus genomas sequenciados pelo método *shotgun* na plataforma Illumina Hiseq 2500, utilizando bibliotecas genômicas Nextera (*paired-end*) de 100 e 250 pb; e na plataforma de pirosequenciamento 454 GS Junior, gerando *reads* de 450 pb *single-end* (Plataforma de Sequenciamento de Alto Desempenho do Instituto Oswaldo Cruz – FIOCRUZ).

3.6 Montagem e anotação dos genomas

As *reads* provenientes dos sequenciamentos foram pré-processadas para (i) seleção de *reads* sem erros de sequenciamento utilizando o Quake v.0.3 (Kelley et al. 2010); (ii) filtragem de *reads* com Phred Score < Q20 e remoção de adaptadores utilizando NGSQCToolkit v.2.3.3 (Patel, Jain 2012).

O Quake possui duas funções: selecionar *reads* sem erros de sequenciamento e corrigir *reads* com erros de sequenciamento. Para isso, o software fragmenta o genoma em k-mers (oligonucleotídeos de tamanho k), e baseado na cobertura de sequenciamento destes, seleciona aqueles com maior probabilidade de estarem corretos ou errados. Se um k-mer possuir algum erro, este não irá alinhar perfeitamente com outros que não possuem o erro, logo, com um nível de cobertura mínima (15x), os k-mers com erros (*untrusted*) apresentariam uma baixa cobertura, enquanto que os k-mers corretos (*trusted*) apresentariam uma alta cobertura. Para determinar o valor de

corte da cobertura que separa os k-mers com erros e sem erros, são utilizados os valores de qualidade atribuídos a cada base.

As montagens de genomas são realizadas por programas (montadores) que funcionam através da fragmentação das *reads* em k-mers; alinhamento e sobreposição destas; e formação de sequências consenso (*contigs*).

As montagens genômicas podem seguir duas estratégias: baseado em um genoma referência ou sem referência (montagem *de novo*). As montagens que utilizam referências geralmente são de organismos já descritos na literatura e muito próximos filogeneticamente ao genoma referência. Assim, a montagem é feita de modo a preservar todos os elementos do genoma referência, o que reduz o nível de complexidade do processo. O resultado disso, geralmente, são montagens com poucos e ordenados *contigs* representando os genomas. Esse tipo de estratégia envia a montagem ao genoma referência, informações relativas à regiões de plasticidade ou elementos móveis, que possam estar presentes no genoma sequenciado, mas ausentes do genoma referência podem não ser utilizadas pelo montador.

As montagens *de novo* não utilizam uma sequência referência, o que aumenta o grau de complexidade do processo, principalmente por causa de regiões genômicas repetitivas e de baixa complexidade. Além disso, se não tratados, erros de sequenciamento podem ser incorporados na montagem. Por causa dessas questões, montagens *de novo* tendem a gerar mais *contigs*.

A estratégia de montagem aplicada aos genomas deste estudo foi a *de novo* já que, a princípio, seriam genomas de espécies ainda não descritas de micobactéria. As montagens foram realizadas com vários montadores: A5 pipeline v.20150522 (Tritt et al. 2012), SPAdes v.3.7 (Bankevich et al. 2012), Edena v.3.13.1028 (Hernandez et al. 2008), Mira v.4.0.2 (Miller et al. 2010), MaSuRCA v.3.1.3 (Zimin et al. 2013), Velvet v.1.2.10 (Zerbino et al. 2008), Abyss v.1.9.0 (Simpson et al. 2009); utilizando subconjuntos das *reads*, que geraram múltiplas montagens parciais. A partir destas, realizamos sua integração para obter uma montagem consenso utilizando CISA v.1.3 (Lin & Liao 2013).

O CISA (*Contig Integrator for Sequence Assembly*) foi desenvolvido com o objetivo de integrar múltiplas montagens de um mesmo genoma, realizadas por diferentes montadores. Inicialmente, o programa identifica *contigs* de maior tamanho, que são considerados representativos, e tenta estendê-los alinhando *contigs* menores

nas suas pontas; em seguida, compara *contigs* que representariam uma mesma sequência e remove ou fragmenta aqueles montados erroneamente; o próximo passo envolve a fusão de *contigs* que tenham sobreposição nas pontas, estimando o tamanho de regiões repetitivas.

Nesse estudo, utilizamos o CISA para integrar múltiplas montagens parciais, de um mesmo montador. Assim, a montagem final para tal montador englobaria a totalidade de *reads* geradas para aquele genoma. Essa montagem consenso, foi submetida ao Pilon v.1.14 (Walker et al. 2014), que realiza o mapeamento da montagem em busca de erros, corrigindo-os; e ao término, obtivemos a montagem final.

A anotação dos genomas foi realizada pelo Prokka v.1.11 (Seemann 2014) e no servidor RAST (<http://rast.nmpdr.org/>) (Aziz et al. 2008), que identificaram e associaram sequências à funções biológicas, a partir de suas base de dados.

3.7 Genomas públicos de micobactérias utilizados

Foram utilizados 94 genomas depositados no Genbank, completos ou *drafts*, de diferentes espécies e subespécies do gênero *Mycobacterium* para as análises genéticas com os nossos genomas (Tabela 3.2). Os genomas das micobactérias que apresentaram relações mais próximas com os nossos foram considerados para nossas análises filogenéticas.

Tabela 3.2. Genomas utilizados no estudo obtidos a partir do Genbank.

Genoma	Tamanho (Mb)	GC %	Número de acesso	Crescimento
<i>M. abscessus</i> ATCC 19977*	5,09	64,12	NC_010397.1	Rápido
<i>M. abscessus</i> subsp. <i>bolletii</i> 50594	5,27	64,18	NC_021282.1	Rápido
<i>M. africanum</i> GM041182	4,38	65,60	NC_015758.1	Lento
<i>M. aromaticivorans</i> JS19b1	6,30	66,40	NZ_JALN00000000.2	Rápido
<i>M. arupense</i> GUC1	4,44	67,30	NZ_LASW00000000.2	Lento
<i>M. asiaticum</i> DSM 44297*	5,93	66,20	CCBD01 (WGS)	Lento
<i>M. austroafricanum</i> DSM 44191*	6,77	67,70	NZ_CCAW00000000.1	Rápido
<i>M. avium</i> ATCC 25291*	4,85	69,20	ACFI01 (WGS)	Lento
<i>M. avium</i> subsp. <i>hominissuis</i> A5	4,87	69,40	AUZQ01 (WGS)	Lento
<i>M. avium</i> subsp. <i>paratuberculosis</i> ATCC 19698*	4,73	69,30	AGAR01 (WGS)	Lento
<i>M. avium</i> subsp. <i>silvaticum</i> ATCC 49884*	4,71	69,20	AYOC01 (WGS)	Lento
<i>M. bohemicum</i> DSM 4427*	5,10	68,70	CSTD00000000.1	Lento
<i>M. bovis</i> AF2122/97*	4,35	65,60	NC_002945.3	Lento
<i>M. canettii</i> CIPT 140010059*	4,48	65,60	NC_015848.1	Lento
<i>M. caprae</i> MB2	4,29	65,50	NZ_CDHG00000000.1	Lento
<i>M. celatum</i> ATCC 51131*	4,66	66,80	NZ_BBUN00000000.1	Lento
<i>M. chelonae</i> ATCC 35752*	4,89	63,90	NZ_CP010946.1	Rápido
<i>M. chimaera</i> MCIMRL4	6,02	67,70	LJHM01 (WGS)	Lento
<i>M. chlorophenicum</i> DSM 43826*	7,37	68,40	JYNL01 (WGS)	Rápido
<i>M. chubuense</i> DSM 44219*	5,94	69,20	JYNX01 (WGS)	Rápido
<i>M. colombiense</i> CECT 3035*	5,57	68,10	AFVW02 (WGS)	Lento
<i>M. conceptionense</i> MLE	7,09	66,40	LFOD01 (WGS)	Rápido
<i>M. cosmeticum</i> DSM 44829*	6,46	68,20	NZ_CCBB00000000.1	Rápido
<i>M. crocinum</i> JCM 16369*	5,19	65,80	BBHD00000000.1	Rápido
<i>M. elephantis</i> Lipa	5,19	67,80	NZ_LBNO00000000.1	Rápido
<i>M. europaeum</i> CSUR P1344*	6,15	68,20	CTEC00000000.1	Rápido
<i>M. farcinogenes</i> DSM 43637*	6,14	66,60	NZ_CCAY00000000.1	Rápido
<i>M. fluoranthenvivorans</i> JCM 14741*	5,51	66,30	BBFT00000000.1	Lento
<i>M. fortuitum</i> DSM 46621*	6,34	66,00	ALQB01 (WGS)	Rápido
<i>M. gastris</i> Wayne	6,00	66,10	NZ_AZYN00000000.1	Lento
<i>M. genavense</i> ATCC 51234*	4,94	66,90	NZ_JAGZ00000000.1	Lento
<i>M. gilvum</i> Spyr1	5,55	67,90	NC_014814.1	Rápido
<i>M. goodii</i> X7B	7,11	67,60	NZ_CP012150.1	Rápido
<i>M. haemophilum</i> ATCC 29548*	4,23	63,90	NZ_CP011883.2	Lento
<i>M. hassiacum</i> DSM 44199*	5,08	69,40	ARBU01 (WGS)	Rápido
<i>M. heckeshornense</i> RLE	5,01	65,90	LFOF01 (WGS)	Lento
<i>M. heraklionense</i> Davo	5,10	67,90	LDPO01 (WGS)	Lento
<i>M. hodleri</i> JCM 12141*	3,87	66,30	BBGO00000000.1	Lento
<i>M. immunogenum</i> SMUC14	5,56	64,30	JXUU01 (WGS)	Rápido
<i>M. indicus pranii</i> MTCC 9506*	5,59	68,00	NC_018612.1	Lento

Tabela 3.2. Genomas utilizados no estudo obtidos a partir do Genbank (cont.).

<i>M. intracellulare</i> ATCC 13950*	5,40	68,10	NC_016946.1	Lento
<i>M. iranicum</i> UM_TJL	6,13	66,10	AUWT01 (WGS)	Rápido
<i>M. kansasii</i> ATCC 12478*	6,43	66,20	NC_022663.1	Lento
<i>M. kyoriense</i> KUM060204T	5,30	66,90	BBKA01 (WGS)	Lento
<i>M. lentiflavum</i> CSUR P1491*	6,82	65,80	CTEE00000000.1	Rápido
<i>M. leprae</i> Br4923	3,27	57,80	NC_011896.1	Lento
<i>M. lepromatosis</i> Mx1-22A	3,20	57,90	JRPY01 (WGS)	Lento
<i>M. liflandii</i> 128FXT	4,82	65,60	NC_020133.1	Lento
<i>M. llatzerense</i> CLUC14	6,09	66,60	JXST01 (WGS)	Rápido
<i>M. mageritense</i> DSM 44476*	7,97	67,00	NZ_CCBF000000000.1	Rápido
<i>M. marinum</i> M*	6,64	65,70	NC_010612.1	Lento
<i>M. abscessus</i> subsp. <i>massiliense</i> GO 06	4,68	64,30	NC_018150.2	Rápido
<i>M. mucogenicum</i> CSUR P2099*	6,21	67,20	CYSI01 (WGS)	Rápido
<i>M. nebraskense</i> AKUC1	6,46	66,60	LASX01 (WGS)	Lento
<i>M. neoaurum</i> ATCC 25795*	5,46	66,70	JMDW01 (WGS)	Rápido
<i>M. neworleansense</i> ATCC 49404*	6,29	66,90	CWKH00000000.1	Rápido
<i>M. nonchromogenicum</i> NCK8460	6,35	66,30	CVTC00000000.1	Rápido
<i>M. obuense</i> DSM 44075*	5,71	68,00	JYNU01 (WGS)	Rápido
<i>M. orygis</i> 112400015	4,28	65,60	NZ_APKD00000000.1	Lento
<i>M. pallens</i> JCM 16370*	5,31	65,70	BBHE00000000.1	Lento
<i>M. parascrofulaceum</i> ATCC BAA-614*	6,56	68,50	NZ_ADNV00000000.1	Lento
<i>M. peregrinum</i> CSUR P2098*	7,10	66,20	CYSH01 (WGS)	Rápido
<i>M. phlei</i> DSM 43239*	5,32	69,40	ANBO01 (WGS)	Rápido
<i>M. pseudoshottsii</i> L15	5,94	65,70	BCND00000000.1	Lento
<i>M. pyrenivorans</i> JCM 15927*	5,13	66,30	BBHB00000000.1	Lento
<i>M. rhodesiae</i> NBB3	6,42	65,50	NC_016604.1	Rápido
<i>M. rufum</i> JS14*	6,17	69,25	JROA00000000.1	Rápido
<i>M. rutilum</i> DSM 45405*	5,99	68,40	LT629971.1	Lento
<i>M. senegalense</i> CK1	6,73	66,40	LDCC01 (WGS)	Rápido
<i>M. septicum</i> DSM 44393*	6,91	66,70	NZ_CBMO000000000.1	Rápido
<i>M. setense</i> DSM 45070*	6,26	66,40	JTJW01 (WGS)	Rápido
<i>M. simiae</i> ATCC 25275*	5,78	65,80	CBMJ02 (WGS)	Lento
<i>M. sinense</i> JDM601	4,64	68,40	NC_015576.1	Lento
<i>M. smegmatis</i> MC2 155*	6,99	67,40	NC_008596.1	Rápido
<i>M. sp.</i> 360MFTsu5	6,05	67,20	ARTK01 (WGS)	Rápido
<i>M. sp.</i> EPa45	6,17	66,40	NZ_CP011773.1	Rápido
<i>M. sp.</i> HXXIII	5,72	63,90	LJYT01 (WGS)	Rápido
<i>M. sp.</i> JLS	6,04	68,40	NC_009077.1	Rápido
<i>M. sp.</i> KMS	6,25	68,19	NC_008705.1	Rápido
<i>M. sp.</i> MCS	5,92	68,43	NC_008146.1	Rápido
<i>M. sp.</i> MOTT36Y	5,61	67,90	NC_017904.1	Lento

Tabela 3.2. Genomas utilizados no estudo obtidos a partir do Genbank (cont.).

<i>M. sp. UNC280MFTsu5</i>	5,69	66,70	JQKU01 (WGS)	Rápido
<i>M. sp. UNC410</i>	6,05	67,10	JMLM01 (WGS)	Rápido
<i>M. sp. VKM ac-1817D</i>	6,32	66,20	NZ_CP009914.1	Rápido
<i>M. thermoresistibile ATCC 19527*</i>	4,87	69,00	AGVE01 (WGS)	Rápido
<i>M. triplex DSM 44626*</i>	6,38	66,60	NZ_CCAU000000000.1	Lento
<i>M. tuberculosis H37Rv*</i>	4,41	65,6	NC_000962.3	Lento
<i>M. tusciae JS617</i>	7,31	65,4	NZ_AGJJ00000000.2	Lento
<i>M. ulcerans Agy99</i>	5,80	65,50	CP000325.1	Lento
<i>M. vaccae ATCC 25954*</i>	6,24	68,50	ALQA01 (WGS)	Rápido
<i>M. vanbaalenii PYR-1</i>	6,49	67,80	NC_008726.1	Rápido
<i>M. vulneris DSM 45247*</i>	6,98	66,70	CCBG01 (WGS)	Rápido
<i>M. xenopi RIVM700367</i>	4,43	66,1	NZ_AJFI00000000.1	Lento
<i>M. yongonense DSM 45126*</i>	5,52	68,00	NC_021715.1	Lento

*, genomas de isolados tipo

3.8 Taxonomia genômica

Para realizar as análises de taxonomia genômica foram utilizadas diversas abordagens: MLSA com genes completos, *core genome* MLSA (*cgMLSA*), *in silico* hibridização DNA-DNA (*isDDH*) e *Microbial Species Identifier* (MiSI).

Sequências nucleotídicas completas dos setes genes do esquema de MLSA estendido (16S, *rpoB*, *gyrB*, *hsp65*, *fusA*, *secA* e *tuf*) foram extraídas dos genomas sequenciados assim como sequências homólogas destes genes de várias espécies de micobactérias. Estas foram concatenadas, alinhadas e submetidas à reconstrução filogenética por máxima verossimilhança (100 replicatas de *bootstrap*).

A abordagem *cgMLSA*, similar ao MLSA, mas utilizando o conjunto de genes ortólogos comuns a todos os genomas analisados, foi realizada pelo GET_HOMOLOGUES (Contreras-Moreira & Vinuesa 2013). Os ortólogos foram identificados utilizando o algoritmo orthoMCL (Li et al. 2003), considerando uma cobertura de alinhamento $\geq 70\%$ e identidade $\geq 40\%$. Pós identificação, os ortólogos foram alinhados com MAFFT v.7.305b (Kato et al. 2002) e concatenados por *script* Perl. A reconstrução filogenética foi realizada pelo método de máxima verossimilhança com 100 replicatas de *bootstrap*.

A abordagem *isDDH* foi aplicada utilizando o programa de distância genoma-a-genoma (Auch et al. 2010a) por interface web (<http://ggdc.dsmz.de/>). E a abordagem MiSI foi aplicada utilizando o programa ANIcalculator (Varghese et al. 2015). Como este programa recomenda o fornecimento de um arquivo com as sequências de tRNAs e rRNAs, para evitar que os valores do resultado não sejam inflados, estas sequências foram obtidas, de todos os genomas analisados, pelos softwares ARAGORN v.1.2.37 (Laslett & Canback 2004) e barrnap v.0.7.

3.9 Identificação do resistoma

A busca por homólogos de genes associados à resistência foi realizada por perfis *Hidden Markov Models* (HMM) (*e-value* $1e-30$) utilizando a base de dados curada Resfams (Gibson et al. 2015), que foi desenvolvida usando estudos metagenômicos

que investigavam o resistoma de microbiota do solo (assim como os isolados CBMA), e apresentou uma maior sensibilidade comparado a buscas por BLAST. Além deste, também utilizamos outra base de dados curada, CARD (The Comprehensive Antibiotic Resistance Database) (McArthur et al. 2013) de interface web (<https://card.mcmaster.ca/home>).

3.10 Identificação de elementos extra-cromossômicos/mobiloma

A busca por elementos móveis presentes nos genomas foi realizada utilizando o PHASTER (<http://phaster.ca/>) (Arndt et al. 2016), para bacteriófagos; e plasmidSPAdes (Antipov et al. 2016), juntamente com análises por BLASTn, para plasmídeos. As análises por BLASTn se basearam em genes marcadores de plasmídeos em micobactérias, como relaxase e *repA* (Ummels et al. 2014; Uchiya et al. 2015). Além disso, a topologia dos plasmídeos putativos foi definida baseada na informação *paired-end* das *reads*.

3.11 Clonagem e expressão heteróloga

Experimentos de clonagem em sistema heterólogo foram realizados para verificar a atividade de genes preditos associados a resistência a antibióticos. Os genes considerados foram: (i) *bla326*, com atividade enzimática para cefalosporinas, penicilinas e carbapenemas (beta-lactâmicos); (ii) *arr*, com atividade de modificação do antibiótico rifampicina; e (iii) *mfpA/mfpB*, com atividade de proteção ao alvo das quinolonas. Como o sinergismo do *mfpA/B* levaria a uma resistência aumentada às quinolonas (Tao et al. 2013), um único fragmento de DNA contendo os dois genes foi clonado.

Os produtos de PCR correspondendo a toda região codificadora, assim como a região reguladora, dos genes acima foram clonados no vetor comercial pGEM-T-Easy Vector System (Promega), de acordo com as recomendações do fabricante, e posteriormente transformados em células quimicamente competentes da linhagem DH5 α (*arr* e *bla326*) e MC1061 (*mfpA/mfpB*) de *Escherichia coli*.

Para confirmar o sucesso da clonagem, o DNA total de alguns clones foi extraído e utilizado como molde em reações de PCR, juntamente com os iniciadores M13F e M13R, disponibilizados no kit de clonagem, nas mesmas condições experimentais já citadas anteriormente. Os *amplicons* foram purificados e sequenciados para verificar a orientação do inserto em relação ao promotor presente no vetor.

3.12 Testes de susceptibilidade aos antimicrobianos

O perfil de susceptibilidade dos isolados selvagens e das transformantes às fluoroquinolonas, rifampicina e beta-lactâmicos, foi determinado através da técnica de discos de difusão e E-test, que define a concentração mínima inibitória (MIC) destes antibióticos em relação às amostras. Os antibióticos utilizados nos testes com os isolados CBMA e transformantes estão listados na Tabela 3.3.

Os isolados de micobactéria da Mata Atlântica e as transformantes foram crescidos até uma escala 0.5 de McFarland e espalhados com *swab* em placas contendo Ágar Müller Hinton (MHA) na qual são aplicados os E-Test/Oxoid e discos contendo antibióticos.

Tabela 3.3. Antibióticos usados nos testes de susceptibilidade.

Classe - Antibiótico	Método	Isolado testado
Beta-lactâmicos		
Amoxicilina (AMO)	Disco de difusão/E-test	CBMA/Transformante
Aztreonam (ATM)	Disco de difusão	CBMA/Transformante
Bacitracina (B)	Disco de difusão	CBMA
Cefalotina (CFL)	Disco de difusão/E-test	CBMA/Transformante
Cefepime (CPM)	Disco de difusão/E-test	CBMA/Transformante
Cefoxitina (CFO)	Disco de difusão	Transformante
Ceftazidime (CAZ)	Disco de difusão/E-test	CBMA/Transformante
Cefuroxima (CXM)	Disco de difusão	Transformante
Imipinem (IPM)	Disco de difusão/E-test	CBMA/Transformante
Meropenem (MEM)	Disco de difusão/E-test	CBMA/Transformante
Meticilina (MET)	Disco de difusão	CBMA
Piperacilina (PRL)	Disco de difusão	CBMA/Transformante
Quinolonas		
Ácido Nalidíxico (NAL)	Disco de difusão	CBMA/Transformante
Ciprofloxacina (CIP)	Disco de difusão/E-test	CBMA/Transformante
Levofloxacina (LVX)	E-test	CBMA

Tabela 3.3. Antibióticos usados nos testes de susceptibilidade (cont.).

Norfloxacina (NOR)	Disco de difusão	CBMA/Transformante
Aminoglicosídeo		
Amicacina (AK)	Disco de difusão/E-test	CBMA
Canamicina (K)	Disco de difusão	CBMA
Estreptomicina (S)	Disco de difusão	CBMA
Gentamicina (CN)	Disco de difusão	CBMA
Neomicina (N)	Disco de difusão	CBMA
Tobramicina (TOB)	Disco de difusão	CBMA
Lincosamidas		
Clindamicina (DA)	Disco de difusão	CBMA
Lincomicina (MY)	Disco de difusão	CBMA
Glicopéptido		
Vancomicina (VA)	Disco de difusão	CBMA
Fosfônicos		
Fosfomicina (FOS)	Disco de difusão	CBMA
Tetraciclina		
Tetraciclina (TE)	Disco de difusão	CBMA
Tigeciclina (TGC)	Disco de difusão	CBMA
Macrolídeo		
Azitromicina (AZM)	Disco de difusão	CBMA
Claritromicina (CLR)	Disco de difusão	CBMA
Eritromicina (ERI)	Disco de difusão	CBMA
Anfenicóis		
Cloranfenicol (C)	Disco de difusão	CBMA
Ansamícinas		
Rifampicina (RIF)	Disco de difusão/E-test	CBMA/Transformante
Polimixinas		
Colistina (CT)	Disco de difusão	CBMA

4 RESULTADOS

4.1 Diversidade genética das micobactérias da Mata Atlântica

Aplicamos a análise do gene 16S rRNA para classificar os isolados da CBMA. Esta análise definiu 14 isolados da CBMA como *Mycobacterium*, já que todos eles apresentaram no mínimo 98,7% de identidade com sequências de bactérias do gênero *Mycobacterium* depositadas no *Ribosomal Database Project* (RDP), que corresponde a um banco de dados curado contendo espécies tipo e seus genes 16S rRNA. Nossa classificação se baseou nos limites definidos para os diferentes níveis taxonômicos, que no caso de gênero, abaixo de 95% de identidade, excluiria a possibilidade de ser do mesmo gênero.

Comparando as sequências do gene ribossomal 16S dos isolados CBMA entre si encontramos cinco sequências distintas: sequência 1 nos isolados 230/293/311/312/335/360; sequência 2 nos isolados 247/294/295/324/329; sequência 3 no isolado 226; sequência 4 no isolado 234 e sequência 5 no isolado 326 (Tabela 4.1).

Considerando espécies já descritas, o grupo de isolados com as sequências 1, 3 e 4 estão mais próximos da *Mycobacterium anthracenicum*; enquanto que os demais, que apresentam a sequência tipo 2, estão próximos a *Mycobacterium septicum*; e o isolado 326 (sequência tipo 5) com *M. chelonae* e *M. salmoniphilum*.

Com essas sequências realizamos uma análise genética através de uma análise de máxima verossimilhança, que apresentou os seguintes grupamentos (Figura 4.1).

Tabela 4.1. Matriz de identidade das sequências 16S dos isolados CBMA e de sequências do banco de dados RDP próximas aos dos isolados CBMA. Isolados CBMA estão marcados em vermelho.

Seq->	M.anthracenicum	M.sp.Site4-B31	CBMA.226	CBMA.234	CBMA.312	CBMA.360	CBMA.311	CBMA.335	CBMA.293	CBMA.230	M.lutzerense.MG18	M.septicum.D13	CBMA.247	CBMA.294	CBMA.295	CBMA.329	CBMA.334	M.chelonae.QIA-55	M.abscessus	CBMA.326
M.anthracenicum	ID	0,987	0,987	0,99	0,991	0,991	0,991	0,991	0,991	0,991	0,989	0,978	0,977	0,977	0,977	0,977	0,977	0,961	0,962	0,955
M.sp.Site4-B31	0,987	ID	0,987	0,991	0,992	0,992	0,992	0,992	0,992	0,992	0,985	0,988	0,988	0,988	0,988	0,988	0,988	0,966	0,966	0,959
CBMA.226	0,987	0,987	ID	0,993	0,992	0,992	0,992	0,992	0,992	0,992	0,988	0,977	0,978	0,978	0,978	0,978	0,978	0,963	0,964	0,957
CBMA.234	0,99	0,991	0,993	ID	0,999	0,999	0,999	0,999	0,999	0,999	0,99	0,982	0,983	0,983	0,983	0,983	0,983	0,963	0,964	0,957
CBMA.312	0,991	0,992	0,992	0,999	ID	1	1	1	1	1	0,991	0,983	0,983	0,983	0,983	0,983	0,983	0,964	0,965	0,958
CBMA.360	0,991	0,992	0,992	0,999	1	ID	1	1	1	1	0,991	0,983	0,983	0,983	0,983	0,983	0,983	0,964	0,965	0,958
CBMA.311	0,991	0,992	0,992	0,999	1	1	ID	1	1	1	0,991	0,983	0,983	0,983	0,983	0,983	0,983	0,964	0,965	0,958
CBMA.335	0,991	0,992	0,992	0,999	1	1	1	ID	1	1	0,991	0,983	0,983	0,983	0,983	0,983	0,983	0,964	0,965	0,958
CBMA.293	0,991	0,992	0,992	0,999	1	1	1	1	ID	1	0,991	0,983	0,983	0,983	0,983	0,983	0,983	0,964	0,965	0,958
CBMA.230	0,991	0,992	0,992	0,999	1	1	1	1	1	ID	0,991	0,983	0,983	0,983	0,983	0,983	0,983	0,964	0,965	0,958
M.lutzerense.MG18	0,989	0,985	0,988	0,99	0,991	0,991	0,991	0,991	0,991	0,991	ID	0,974	0,975	0,975	0,975	0,975	0,975	0,963	0,962	0,957
M.septicum.D13	0,978	0,988	0,977	0,982	0,983	0,983	0,983	0,983	0,983	0,983	0,974	ID	0,999	0,999	0,999	0,999	0,999	0,965	0,966	0,957
CBMA.247	0,977	0,988	0,978	0,983	0,983	0,983	0,983	0,983	0,983	0,983	0,975	0,999	ID	1	1	1	1	0,966	0,966	0,958
CBMA.294	0,977	0,988	0,978	0,983	0,983	0,983	0,983	0,983	0,983	0,983	0,975	0,999	1	ID	1	1	1	0,966	0,966	0,958
CBMA.295	0,977	0,988	0,978	0,983	0,983	0,983	0,983	0,983	0,983	0,983	0,975	0,999	1	1	ID	1	1	0,966	0,966	0,958
CBMA.329	0,977	0,988	0,978	0,983	0,983	0,983	0,983	0,983	0,983	0,983	0,975	0,999	1	1	1	ID	1	0,966	0,966	0,958
CBMA.334	0,977	0,988	0,978	0,983	0,983	0,983	0,983	0,983	0,983	0,983	0,975	0,999	1	1	1	1	ID	0,966	0,966	0,958
M.chelonae.QIA-55	0,961	0,966	0,963	0,963	0,964	0,964	0,964	0,964	0,964	0,964	0,963	0,965	0,966	0,966	0,966	0,966	0,966	ID	0,999	0,987
M.abscessus	0,962	0,966	0,964	0,964	0,965	0,965	0,965	0,965	0,965	0,965	0,962	0,966	0,966	0,966	0,966	0,966	0,966	0,999	ID	0,986
CBMA.326	0,955	0,959	0,957	0,957	0,958	0,958	0,958	0,958	0,958	0,958	0,957	0,957	0,958	0,958	0,958	0,958	0,958	0,987	0,986	ID

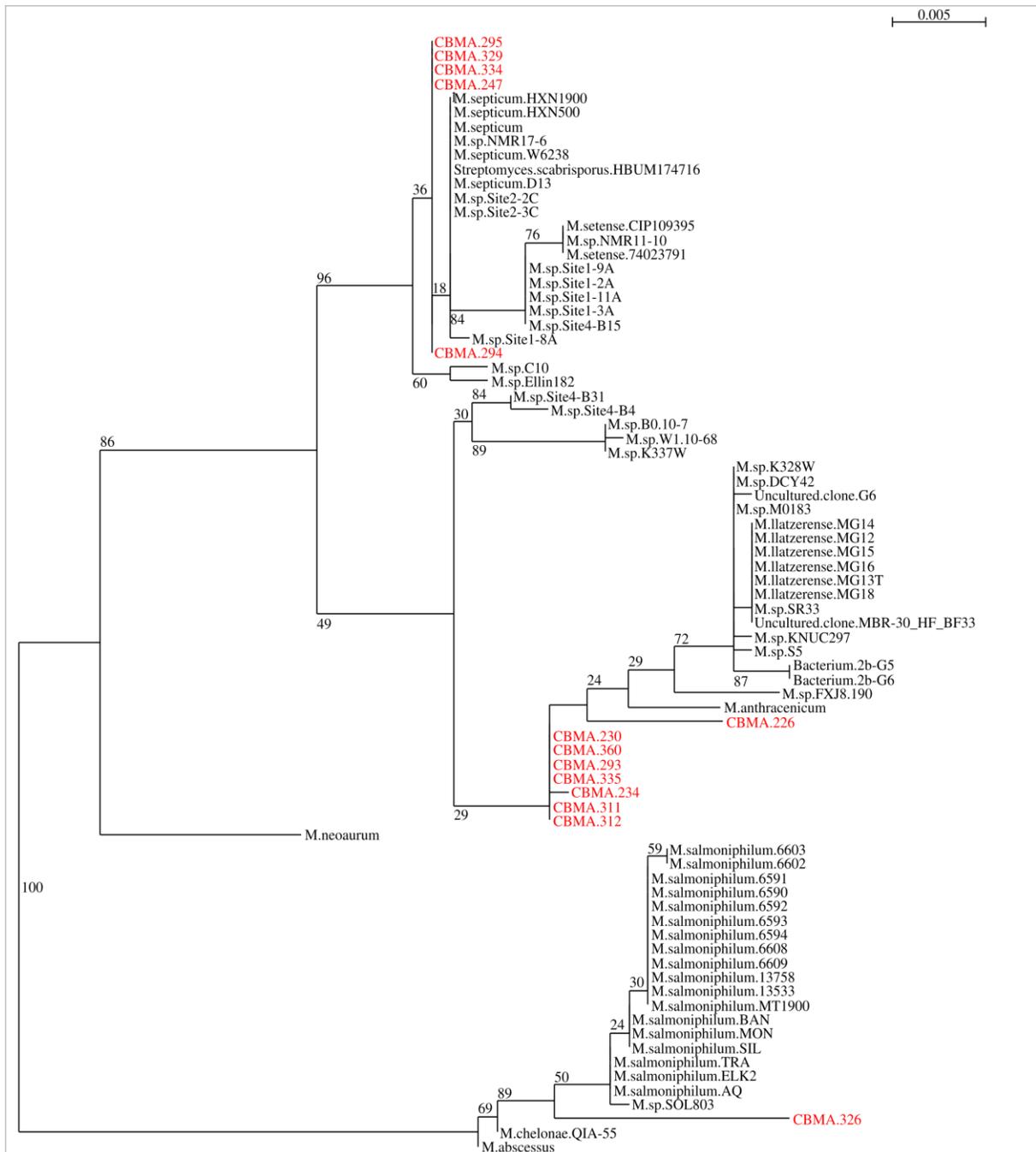


Figura 4.1. Árvore de máxima verossimilhança de sequências 16S (1.185 pb) dos isolados CBMA e do banco de dados RDP. Sequências dos isolados CBMA estão marcadas em vermelho. 100 replicatas de *bootstrap*.

Assim, aplicamos o esquema de MLSA, considerando cinco genes concatenados (16S rRNA, *hsp65*, *sodA*, *rpoB* e *gyrB* - 2.801 pb), para aferir a diversidade de espécies/linhagens dentre estes isolados (Figura 4.2). Os isolados deste estudo foram agrupados em três *clusters*. Todos esses *clusters* estavam associados a espécies atípicas/crescimento rápido. Sequências CBMA do cluster 1 estão mais relacionadas

com *M. llatzerense*; do cluster 2, com *M. abscessus*; e cluster 3, com *M. septicum*. Os valores de identidade entre as sequências concatenadas dentre os isolados CBMA e espécies próximas filogeneticamente estão representados na Tabela 4.2.

Isolados representando os distintos *clusters* foram então selecionados para o sequenciamento genômico: CBMA 226, 234, 247, 312 e 326. Além destes, foram sequenciados também CBMA 311 e 360, com finalidade de realizarmos análises intra espécie/linhagem.

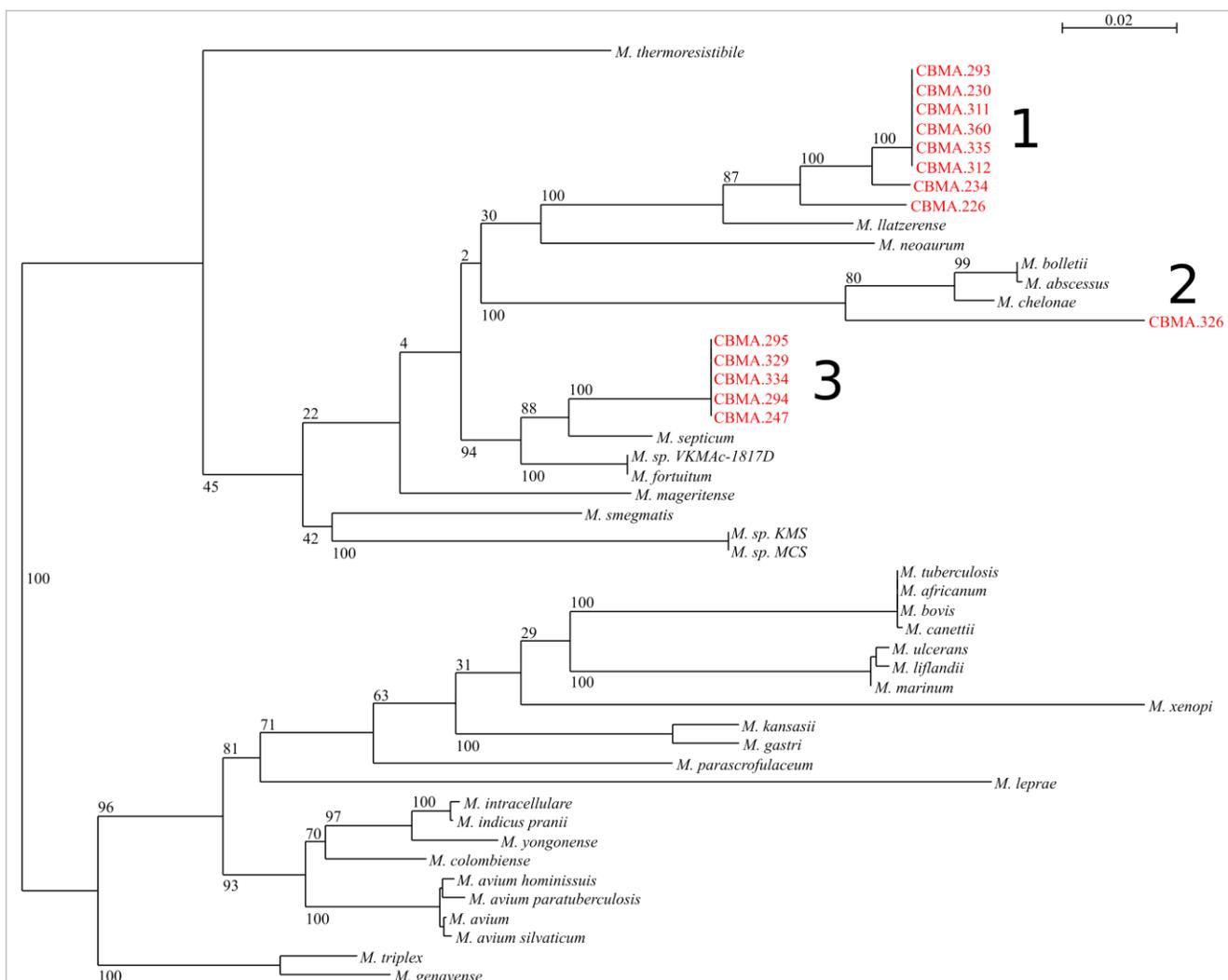


Figura 4.2. **Árvore de máxima verossimilhança da abordagem MLSA dos isolados de micobactéria da Mata Atlântica.** Clusters 1, 2 e 3 estão representados em vermelho. 100 replicatas de *bootstrap*.

Tabela 4.2. Matriz de identidade das sequências concatenadas de cinco genes parciais (2801 pb) dos isolados CBMA e de bancos de dados mais próximas aos dos isolados CBMA. Isolados CBMA estão marcados em vermelho.

Seq->	M.septicum	CBMA247	CBMA294	CBMA295	CBMA329	CBMA334	CBMA226	CBMA230	CBMA293	CBMA312	CBMA335	CBMA360	CBMA311	CBMA234	M.llatzerense	M.abscessus	M.chelonae	CBMA326	
M.septicum	ID	0,971	0,971	0,971	0,971	0,971	0,939	0,938	0,938	0,938	0,938	0,938	0,938	0,937	0,938	0,926	0,926	0,916	
CBMA247	0,971	ID	1	1	1	1	0,941	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,922	0,923	0,912	
CBMA294	0,971	1	ID	1	1	1	0,941	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,922	0,923	0,912	
CBMA295	0,971	1	1	ID	1	1	0,941	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,922	0,923	0,912	
CBMA329	0,971	1	1	1	ID	1	0,941	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,922	0,923	0,912	
CBMA334	0,971	1	1	1	1	ID	0,941	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,939	0,922	0,923	0,912	
CBMA226	0,939	0,941	0,941	0,941	0,941	0,941	ID	0,972	0,972	0,972	0,972	0,972	0,972	0,972	0,973	0,963	0,919	0,918	0,913
CBMA230	0,938	0,939	0,939	0,939	0,939	0,939	0,972	ID	1	1	1	1	1	0,989	0,964	0,918	0,917	0,911	
CBMA293	0,938	0,939	0,939	0,939	0,939	0,939	0,972	1	ID	1	1	1	1	0,989	0,964	0,918	0,917	0,911	
CBMA312	0,938	0,939	0,939	0,939	0,939	0,939	0,972	1	1	ID	1	1	1	0,989	0,964	0,918	0,917	0,911	
CBMA335	0,938	0,939	0,939	0,939	0,939	0,939	0,972	1	1	1	ID	1	1	0,989	0,964	0,918	0,917	0,911	
CBMA360	0,938	0,939	0,939	0,939	0,939	0,939	0,972	1	1	1	1	ID	1	0,989	0,964	0,918	0,917	0,911	
CBMA311	0,938	0,939	0,939	0,939	0,939	0,939	0,972	1	1	1	1	1	ID	0,989	0,964	0,918	0,917	0,911	
CBMA234	0,937	0,939	0,939	0,939	0,939	0,939	0,973	0,989	0,989	0,989	0,989	0,989	0,989	ID	0,965	0,916	0,916	0,91	
M.llatzerense	0,938	0,939	0,939	0,939	0,939	0,939	0,963	0,964	0,964	0,964	0,964	0,964	0,964	0,965	ID	0,916	0,917	0,911	
M.abscessus	0,926	0,922	0,922	0,922	0,922	0,922	0,919	0,918	0,918	0,918	0,918	0,918	0,918	0,916	0,916	ID	0,986	0,945	
M.chelonae	0,926	0,923	0,923	0,923	0,923	0,923	0,918	0,917	0,917	0,917	0,917	0,917	0,917	0,916	0,917	0,986	ID	0,949	
CBMA326	0,916	0,912	0,912	0,912	0,912	0,912	0,913	0,911	0,911	0,911	0,911	0,911	0,911	0,91	0,911	0,945	0,949	ID	

4.2 Montagem e anotação dos genomas

Após sequenciamento na plataforma Illumina 2500 dos sete isolados CBMA selecionados, os genomas foram montados em vários programas, utilizando diferentes parâmetros a fim de obter montagens otimizadas. Foram aplicadas dois tipos de estratégia de montagem, (I) utilizando a totalidade de *reads* obtidas pelo sequenciamento, e (II) dependendo da quantidade de *reads* obtidos por genoma, foram realizadas duas ou três montagens parciais, utilizando frações (em torno de 1 milhão de *reads*) do total de *reads* sequenciados. Essas montagens parciais foram posteriormente integradas, gerando a montagem final. Comparativamente, os resultados obtidos pela estratégia I apresentaram métricas inferiores à estratégia II. Portanto, apenas consideramos as montagens realizadas parcialmente e integradas no final.

As montagens com melhores métricas (menor número de *contigs*, maior tamanho de *contig*, maior N50, maior tamanho de genoma final) foram obtidas pelos montadores A5 pipeline e SPAdes, portanto apenas esses dois foram considerados no estudo. Inicialmente, como ambos os montadores incluíam módulos de pré-processamento, as *reads* foram submetidas diretamente à montagem.

Na maioria das montagens, observamos que as métricas do A5 pipeline foram inferiores as do SPAdes, apresentando discrepâncias em relação ao tamanho do genoma e elevado número de *contigs*. Todas essas informações estão no módulo 1 da tabela 4.3. As montagens realizadas a partir do pré-processamento das *reads* apresentaram resultados melhores para os dois montadores. Por exemplo, para o genoma CBMA 234, os resultados dos dois montadores foram similares; para o genoma CBMA 312, o SPAdes apresentou os melhores resultados; enquanto que para o genoma CBMA 326, o A5 pipeline foi superior (módulo 2 da tabela 4.3). As montagens parciais de cada genoma foram submetidas ao CISA, que as integrou em uma montagem única, através da identificação e sobreposição de *contigs* representativos, e remoção de *contigs* montados erroneamente (módulo 3 da tabela 4.3).

Na Tabela 4.3 encontram-se as métricas relativas aos sequenciamentos da CBMA 234, 312 e 326. As métricas dos demais sequenciamentos (CBMA 226, 247, 311, 360) foram comparáveis a estas.

Tabela 4.3. Comparação das métricas obtidas nas montagens, utilizando *reads* (1) parciais não pré-processadas; (2) parciais pré-processadas; (3) parciais pré-processadas e unificadas pelo CISA, de três genomas exemplos.

	234_001		312_001		326_001				
	A5	SPAdes	A5	SPAdes	A5	SPAdes			
1	# contigs	1658	392	# contigs	594	164	# contigs	4608	2414
	Maior contig	613071	738552	Maior contig	494912	666768	Maior contig	693164	693216
	Tamanho	6896265	6262537	Tamanho	6824927	6652518	Tamanho	10071004	5744176
	N50	342023	392180	N50	185320	288368	N50	6857	372399
	N's	2744	0	N's	0	0	N's	24	0
	234_001_PP		312_001_PP		326_001_PP				
	A5	SPAdes	A5	SPAdes	A5	SPAdes			
2	# contigs	55	56	# contigs	130	105	# contigs	43	74
	Maior contig	612612	403626	Maior contig	494765	516318	Maior contig	692824	397799
	Tamanho	6138847	6142073	Tamanho	6621303	6626901	Tamanho	4800537	4804806
	N50	355493	247213	N50	177195	197329	N50	371836	138044
	N's	50	0	N's	149	0	N's	70	0
	234_CISA		312_CISA		326_CISA				
	A5	SPAdes	A5	SPAdes	A5	SPAdes			
3	# contigs	46	38	# contigs	77	57	# contigs	34	52
	Maior contig	613056	550620	Maior contig	494851	1165555	Maior contig	692824	400574
	Tamanho	6132866	6139450	Tamanho	6902936	6619361	Tamanho	5122613	5046008
	N50	355518	254439	N50	177260	231204	N50	330797	203668
	N's	42	0	N's	154	0	N's	56	0

As montagens passaram por uma última etapa de correção de erros utilizando o Pilon, que analisa o alinhamento das *reads* com o genoma montado a procura de inconsistências. Considerando os objetivos deste projeto, que são o estudo da diversidade, taxonomia e análise de resistoma, a partir de informação genômica global, após a aplicação do Pilon, chegamos, de um modo geral, a valores das métricas adequados para nossas análises (Tabela 4.4).

Tabela 4.4. Métricas finais das montagens dos genomas CBMA.

Genoma	#Contigs	Tamanho	Maior Contig	N50	N's	Cobertura	Sequenciador
226	35	6.211.632	683.952	412.807	0	96x	Illumina
234	38	6.139.886	550.620	254.481	30	193x	Illumina
247	12	6.321.839	2.332.024	1.151.177	0	80x	Illumina+454
311	32	6.426.473	1.132.116	526.172	10	118x	Illumina
312	57	6.634.577	1.168.401	231.204	180	168x	Illumina
326	52	5.046.155	400.574	203.668	100	290x	Illumina
360	81	6.349.136	437.938	236.574	187	11x	454

O tamanho dos genomas variou entre 5.0 - 6.6 Mb, sendo a maioria superior a 6.0 Mb. As coberturas do sequenciamento na plataforma Illumina variaram de 80 a 290x, que são valores considerados suficientes para genomas bacterianos (Chun &

Rainey 2014). A cobertura de 11x obtida pela plataforma 454 também é considerado suficiente, uma vez que as *reads* têm em média ~450 pb, ou seja de duas a quatro vezes maiores que as obtidas na plataforma Illumina.

Os genomas montados foram anotados no Prokka e no RAST. De um modo geral, a anotação pelos dois programas foi equivalente. Aqui apresentamos os resultados gerados pelo Prokka (Tabela 4.5). O conteúdo GC variou entre 64.05 - 66.44%. Todos os genomas apresentaram um único operon ribossomal, e um número variável de cópias do gene rRNA 5S). O número de tRNAs variou entre 48 - 53, com exceção do CBMA 247, com 75 tRNAs, e CBMA 312, com 85 tRNAs.

Tabela 4.5. Anotação geral a partir do Prokka.

CBMA	# rRNA	#CDS	#Genes	#tRNA	GC (%)
226	5	5.874	5.950	49	66.44
234	3	5.764	5.839	50	65.63
247	5	5.923	6.022	75	66.23
311	6	6.087	6.166	51	65.47
312	3	6.374	6.477	85	65.38
326	3	4.929	4.998	48	64.05
360	6	6.086	6.164	53	65.48

Como a anotação pelo RAST fornece também informações relativas às categorias funcionais dos genes preditos, as Figuras 5.3 - 5.9 representam quantitativamente e qualitativamente as categorias definidas para os sete genomas. De um modo geral, o número de genes relacionados a algumas das categorias funcionais é similar entre os genomas. Uma das exceções foi a categoria carboidrato, que para o genoma 326, apresentou o menor número de genes, quase metade em relação aos demais genomas.

Apresentamos de forma estendida as categorias: virulência/doença/defesa e fagos/profagos/elementos transponíveis/plasmídeos por estarem relacionados a objetivos específicos de nosso estudo. Nestas categorias foram observados conteúdos com variações, tanto quantitativas (Tabela 4.6), quanto qualitativas (Figuras 4.3 a 4.9), consideráveis entre os genomas.

Tabela 4.6. Número de genes de duas categorias funcionais versus genomas CBMA.

Categoria funcional/CBMA	226	234	247	311	312	326	360
Virulência, doença e defesa	62	136	118	65	123	105	65
Fagos, profagos, elementos transponíveis e plasmídeos	3	0	8	22	22	1	23

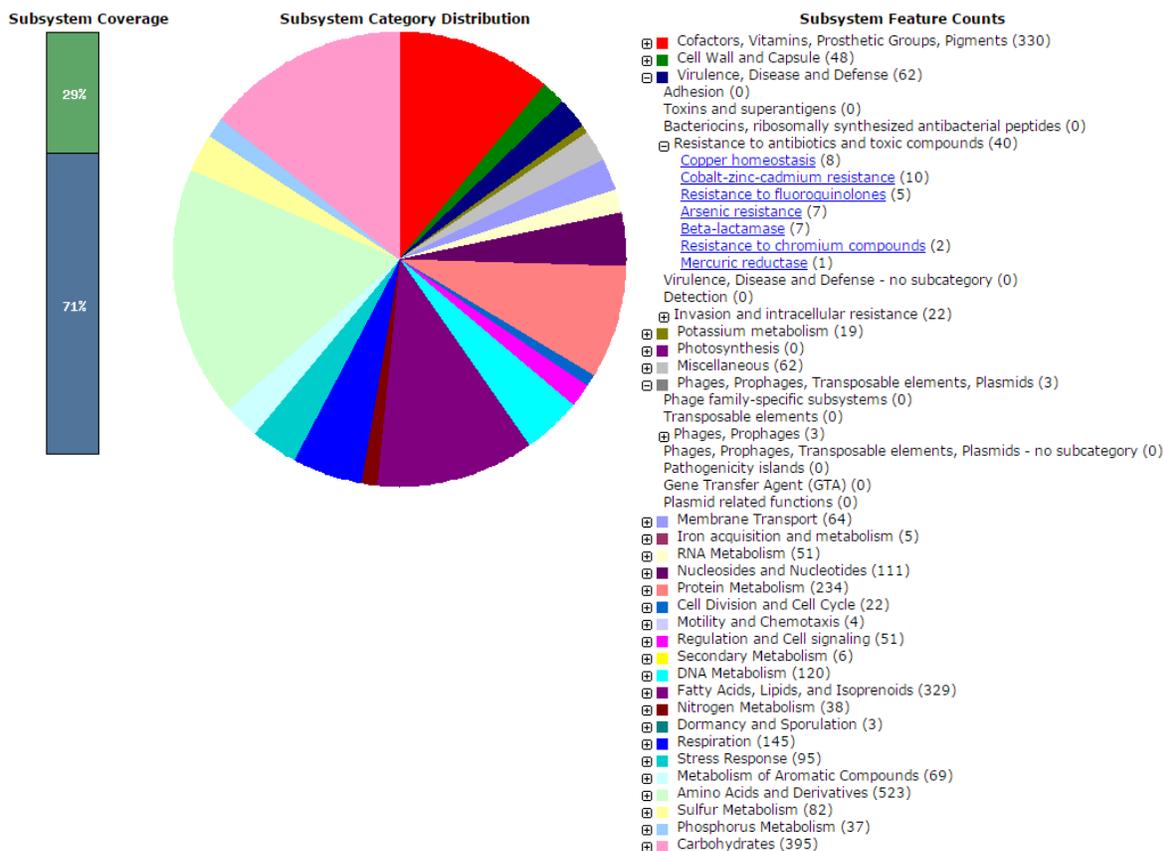


Figura 4.3. Categorização funcional dos genes do isolado CBMA 226.

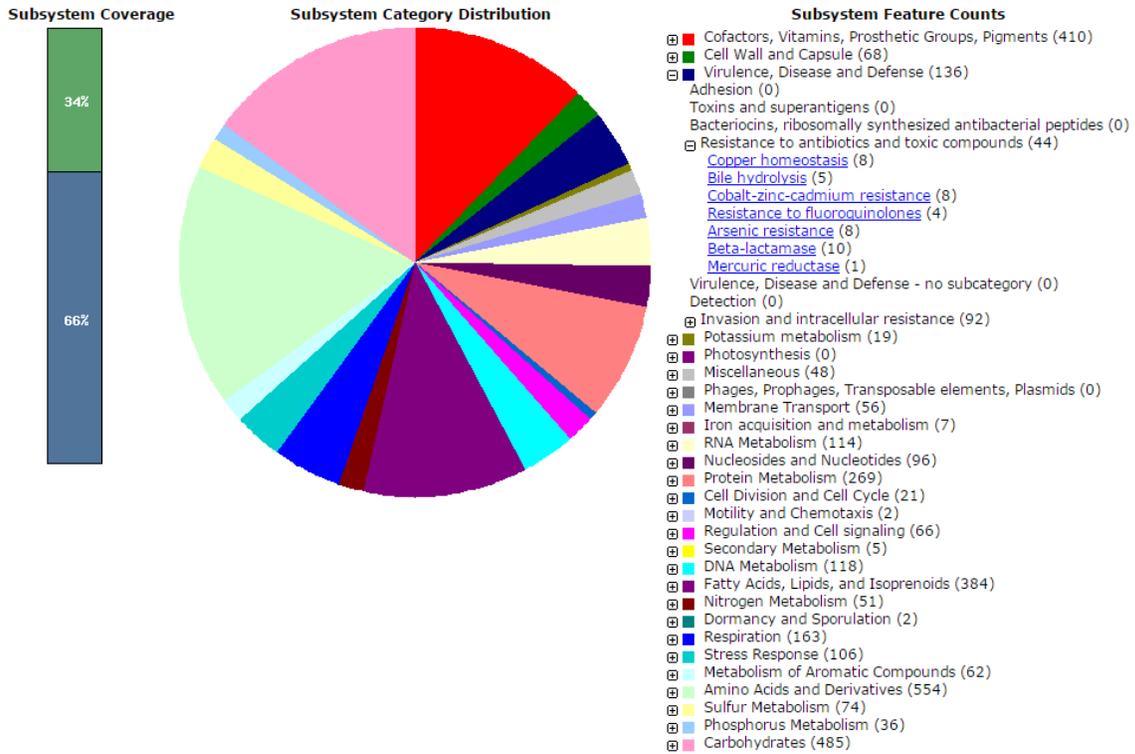


Figura 4.4. Categorização funcional dos genes do isolado CBMA 234.

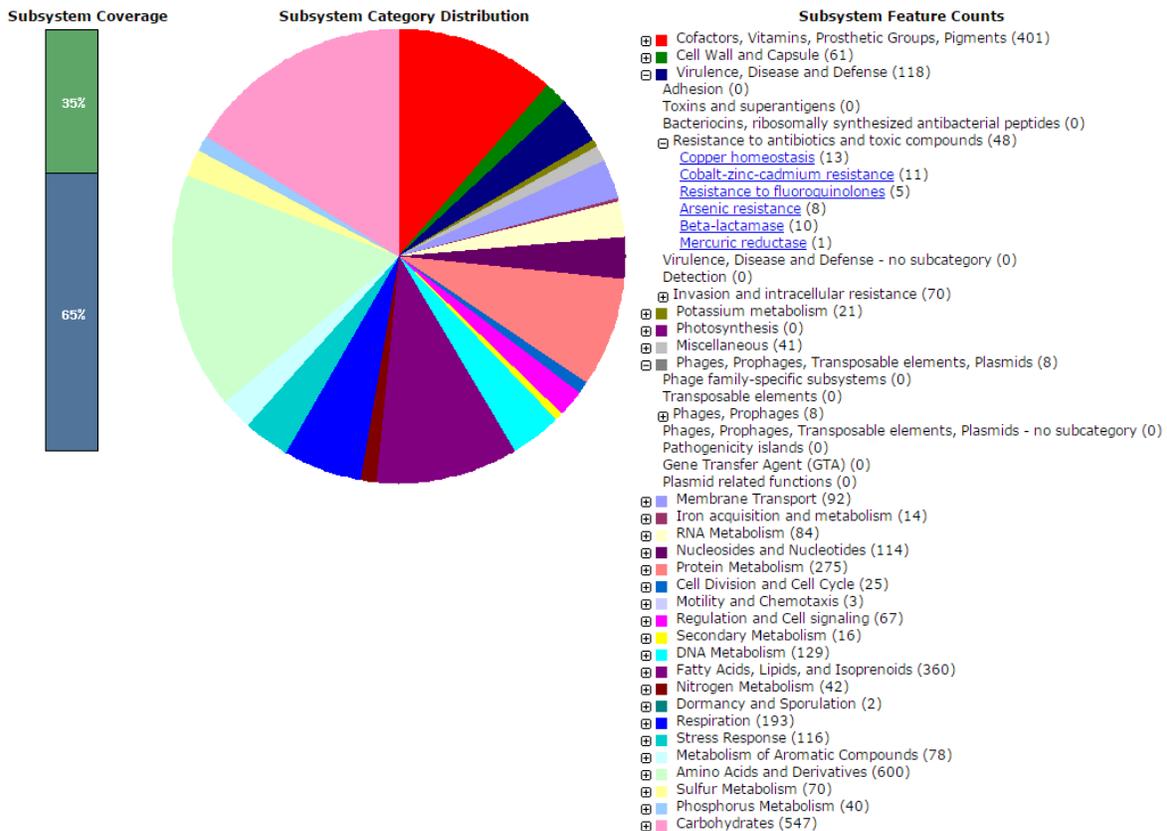


Figura 4.5. Categorização funcional dos genes do isolado CBMA 247.

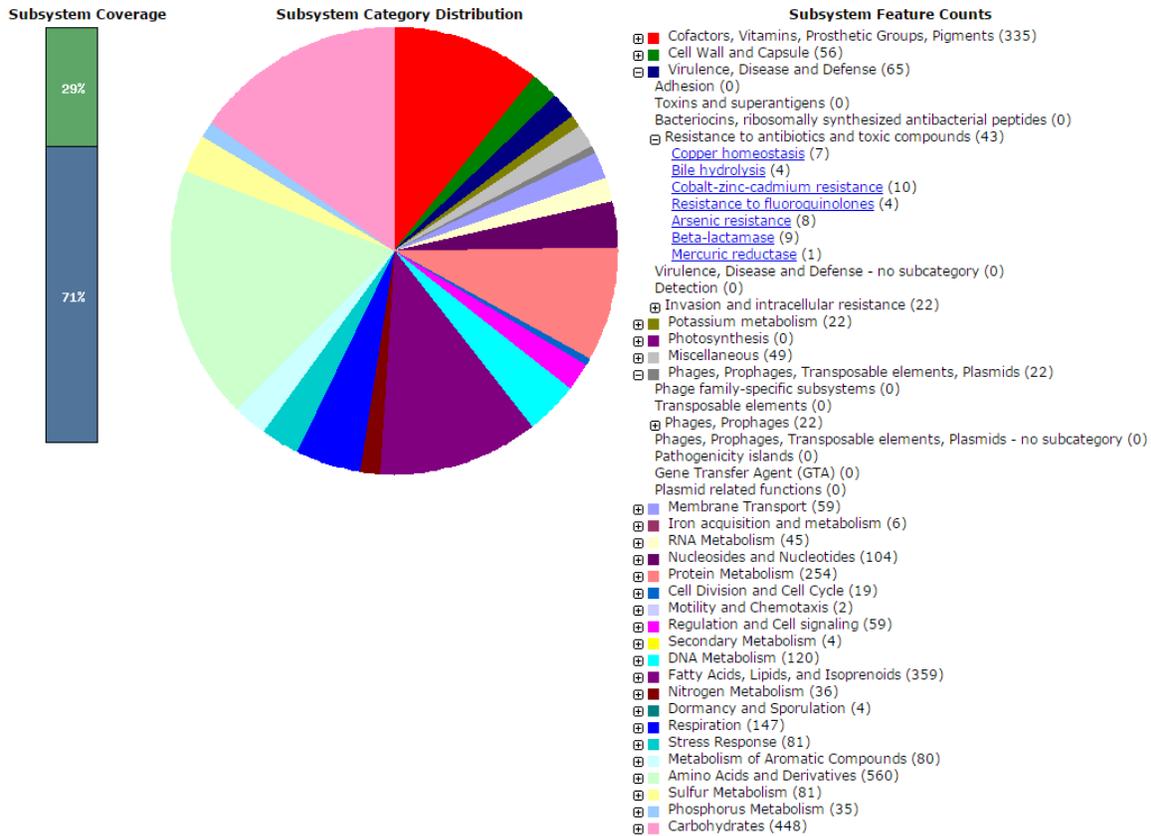


Figura 4.6. Categorização funcional dos genes do isolado CBMA 311.

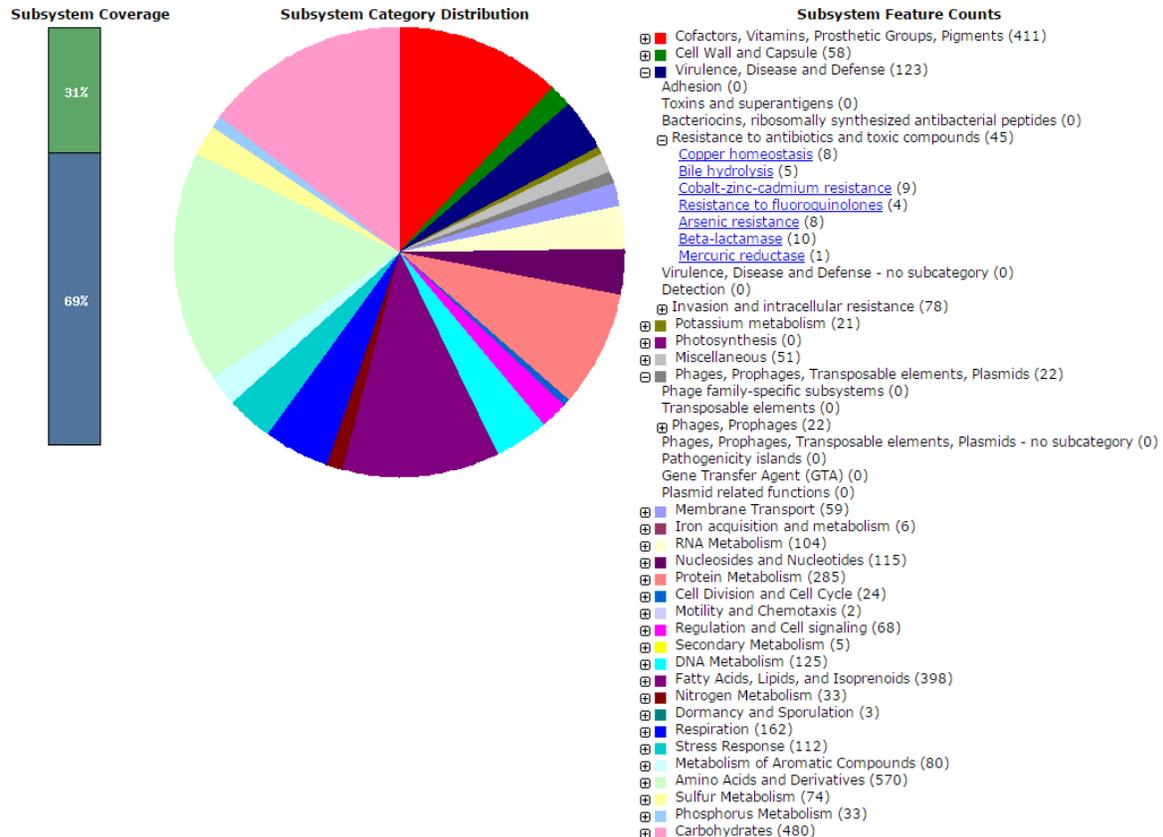


Figura 4.7. Categorização funcional dos genes do isolado CBMA 312.

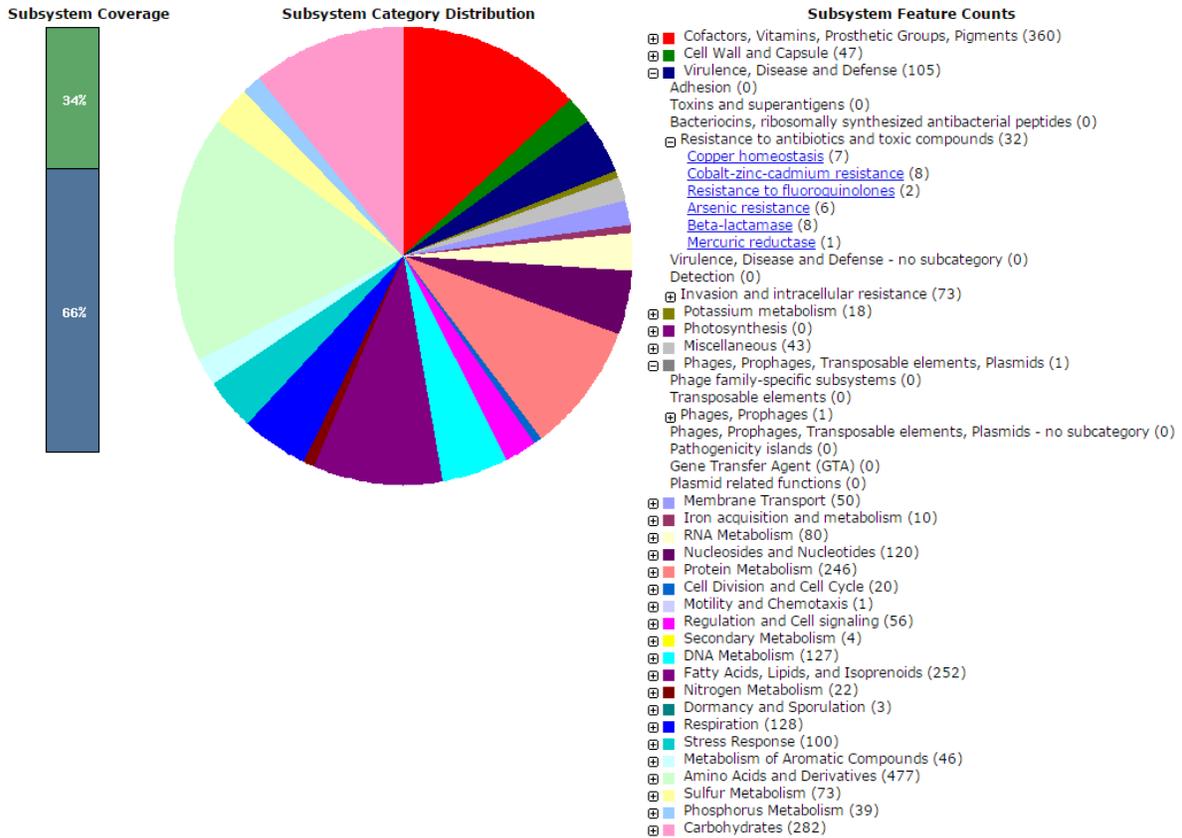


Figura 4.8. Categorização funcional dos genes do isolado CBMA 326.

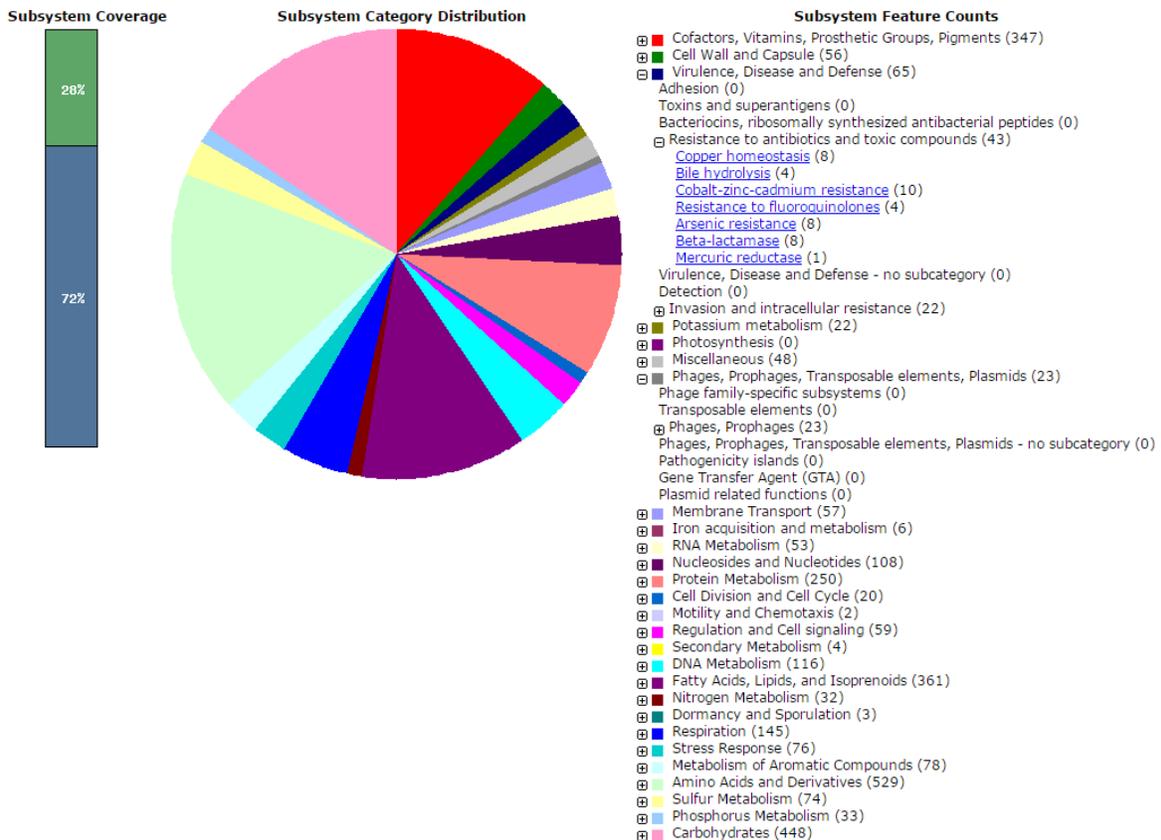


Figura 4.9. Categorização funcional dos genes do isolado CBMA 360.

Uma categoria de genes que pode interferir na anotação e montagem de genomas é a de fagos, profagos, elementos transponíveis e plasmídeos. Na anotação pelo RAST esta categoria apresentou a seguinte distribuição de genes nos distintos genomas (Tabela 4.7):

Tabela 4.7. Número de genes relacionados ao mobiloma versus genomas CBMA.

Categoria funcional/CBMA	226	234	247	311	312	326	360
fagos, profagos, elementos transponíveis e plasmídeos	3	0	8	22	22	1	23

4.3 Identificação do mobiloma

4.3.1 Bacteriófagos

Uma vez que alguns genes foram categorizados na subcategoria fagos e profagos, utilizamos uma ferramenta específica para identificação e anotação destes elementos, PHASTER. Nos isolados 226, 234 e 326 o PHASTER não identificou nenhum gene com associação com fago/profago. Em relação ao isolado 247, o PHASTER caracterizou um genoma incompleto contendo 39,9 kb, codificando 19 proteínas, e com conteúdo GC de 64,62%; com parte dos elementos que constituem um micobacteriófago. Da mesma forma, um genoma incompleto de micobacteriófago foi caracterizado no isolado 311. Este genoma seria de 8 kb, codificador de 10 proteínas, com conteúdo GC de 62,66%. Neste mesmo isolado (CBMA 311), dois bacteriófagos completos, compartilhados com os isolados CBMA 312 e 360, também foram caracterizados pelo PHASTER. Um deles com genoma de 27 kb, que codifica 31 proteínas, e possui conteúdo GC de 63,01%. Enquanto o outro bacteriófago apresentou um tamanho de 49,3 kb, codificador de 71 CDS, e conteúdo GC de 64,40%.

Todos esses bacteriófagos foram caracterizados por similaridade com genes estruturais dos actinobacteriófagos. Além disso, estes actinobacteriófagos apresentaram conteúdo gênico único, considerando o conteúdo gênico dos actinobacteriófagos que estão depositados no Genbank e/ou Actinobacteriophage DB (<http://phagesdb.org>).

4.3.2 Plasmídeos

Como a anotação do RAST não identificou nenhum plasmídeo entre os isolados CBMA, realizamos outras análises visando confirmar esta predição. Buscas por marcadores de plasmídeos de micobactéria, *repA* e relaxase, nos genomas CBMA, utilizando BLASTn, identificaram a presença destes marcadores em dois *contigs* (nomeados P2 e P4). Já buscas utilizando BLASTn versus Genbank, identificaram sequências similares à de plasmídeos identificados em micobactérias, em outros dois *contigs* (nomeados P1 e P3). Além disso, utilizando uma ferramenta do montador SPAdes, o plasmidSPAdes, um outro *contig* foi identificado como plasmídeo (nomeado P5). O plasmidSPAdes também identificou os P2 e P3 previamente caracterizados pela busca dos marcadores e BLAST geral, respectivamente (Tabela 4.8).

Tabela 4.8. Plasmídeos putativos preditos nos genomas CBMA.

CBMA	P1	P2	P3	P4	P5
226	-	-	-	X	-
311	-	-	X	-	X
312	X	X	X	-	X*
360	-	X	X	-	X*
Tamanho	273,030 pb	160,489 pb	21,616 pb	389,088 pb	50,173 pb
GC %	62,90	65,90	64,45	64,50	64,50
Topologia	L	C	L	C	C

*, integrado no genoma; X, presença; -, não identificado; L, linear; C, circular

Para verificar se esses *contigs* não eram artefatos de montagem, suas sequências foram mapeadas contra as *reads* dos respectivos genomas e remontados, e apresentaram 100% de cobertura. Utilizando a informação *paired-end* das *reads*, foi possível determinar a topologia desses plasmídeos, sendo P1 e P3 lineares, enquanto P2, P4 e P5 circulares.

Considerando o conteúdo gênico geral, observamos que segmentos do plasmídeo P2, totalizando ~112 kb estão distribuídos ao longo do plasmídeo P4 (~389 kb) (Figura 4.10), e apresentam ~91% de identidade. A anotação desses segmentos mostrou que estão relacionados a sistema de secreção do tipo VII (T7SS), relaxase, genes relacionados com conjugação (*virB4*, *virD4*), transportadores.

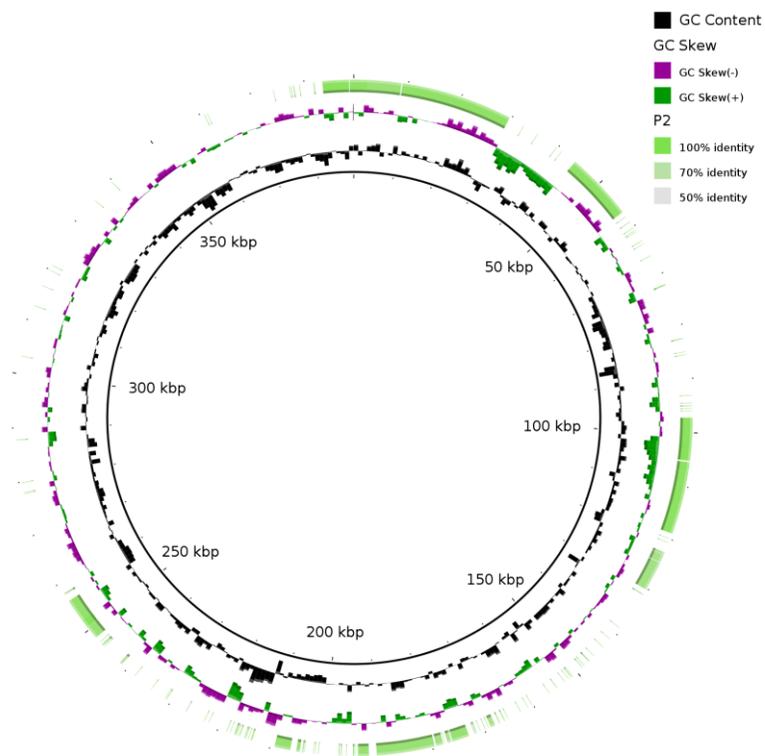


Figura 4.10. **Mapa genômico de P4.** Círculo interno preto representa a sequência nucleotídica de P4. Em comparação, segmentos em verde, no círculo externo, correspondem a ~70% do genoma de P2. Imagem gerada pelo programa BRIG v0.95 (Alikhan et al. 2011).

O genoma de 50,173 bp do P5, que foi identificado pelo plasmidSPAdes, contém um dos bacteriófagos identificados pelo PHASTER, o de ~49 kb, que é compartilhado pelos isolados CBMA 311/312/360. Este elemento estava representado como um único *contig* no genoma 311, que as análises das informações *paired-end* caracterizaram como um elemento circular. Nos genomas 312 e 360, este mesmo elemento estaria contido em *contigs* com tamanhos superiores a 200 kb, indicando que P5 estaria integrado nesses genomas, enquanto que no genoma 311 estaria em forma livre.

4.4 Taxonomia genômica das micobactérias da Mata Atlântica

Para definirmos as espécies deste conjunto de micobactérias, aplicamos a abordagem de taxonomia genômica, considerando as seguintes análises: MLSA estendido, *cgMLSA*, *in silico* hibridação DNA-DNA e MiSI (AF e gANI) e conteúdo GC.

A abordagem MLSA considerou os genes completos: 16S, *rpoB*, *gyrB*, *hsp65*, *fusA*, *secA* e *tuf*. Estes genes concatenados totalizaram 14.196 pb. Estas sequências foram submetidas a uma análise filogenética com outras 46 espécies de micobactéria. A árvore do MLSA definiu que os isolados da CBMA estariam em três clusters sendo que um deles compreendia dois sub-clusters (Figura 4.11).

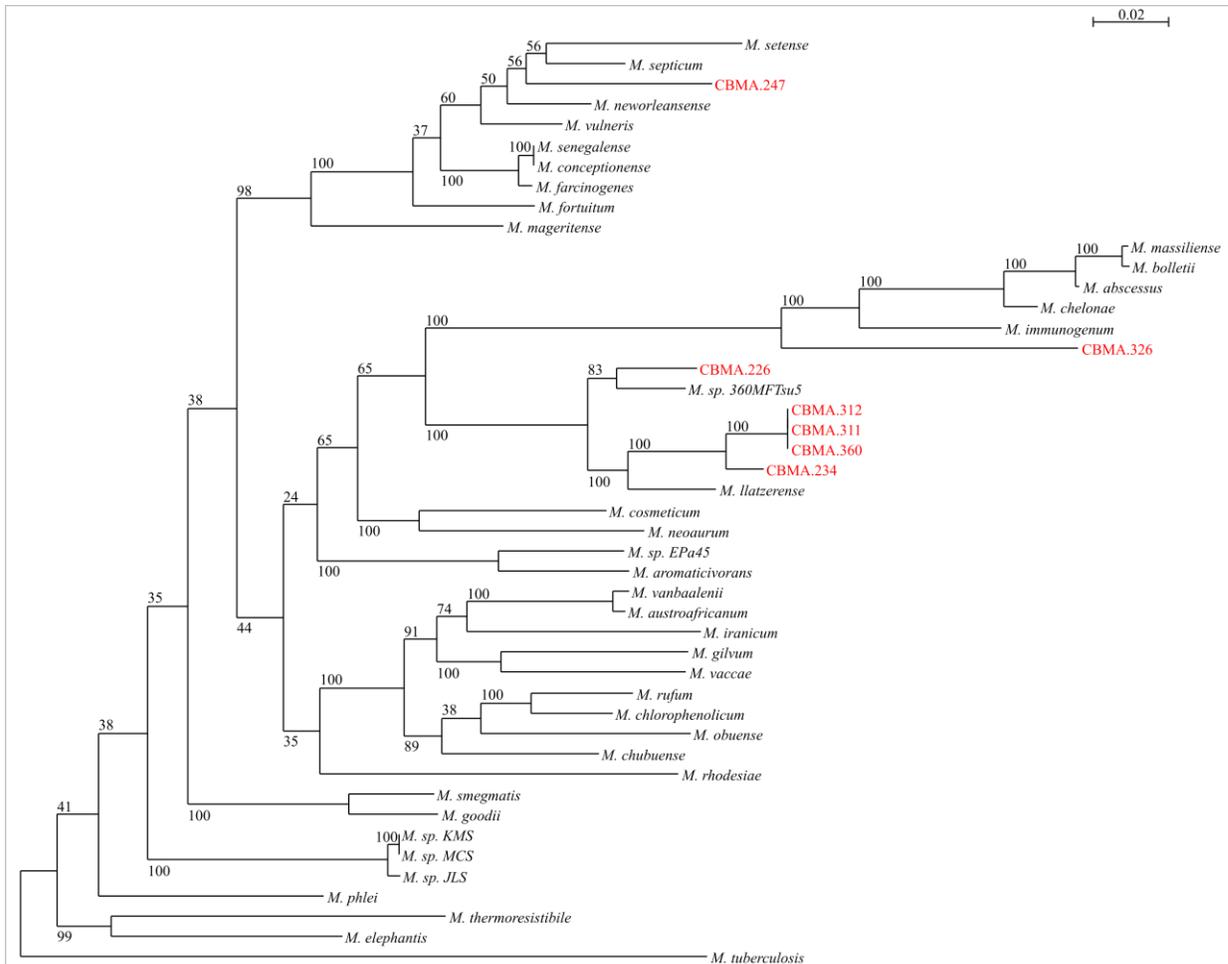


Figura 4.11. **Árvore de máxima verossimilhança da abordagem MLSA** (14.196 pb). Sequências dos isolados CBMA estão marcadas em vermelho. 100 réplicas de *bootstrap*.

A análise do *cgMLSA*, que aplica o genoma *core*, se baseou 239 genes (~160 kb), que correspondem aos genes comuns a todos os genomas das espécies consideradas. Assim como no MLSA, as sequências do genoma *core* foram concatenadas, alinhadas e submetidas a análises genéticas.

A topologia da árvore *cg*MLSA foi similar à da MLSA. Os agrupamentos onde se encontram os isolados CBMA se mantêm, assim como a distância genética entre as espécies próximas (Figura 4.12).

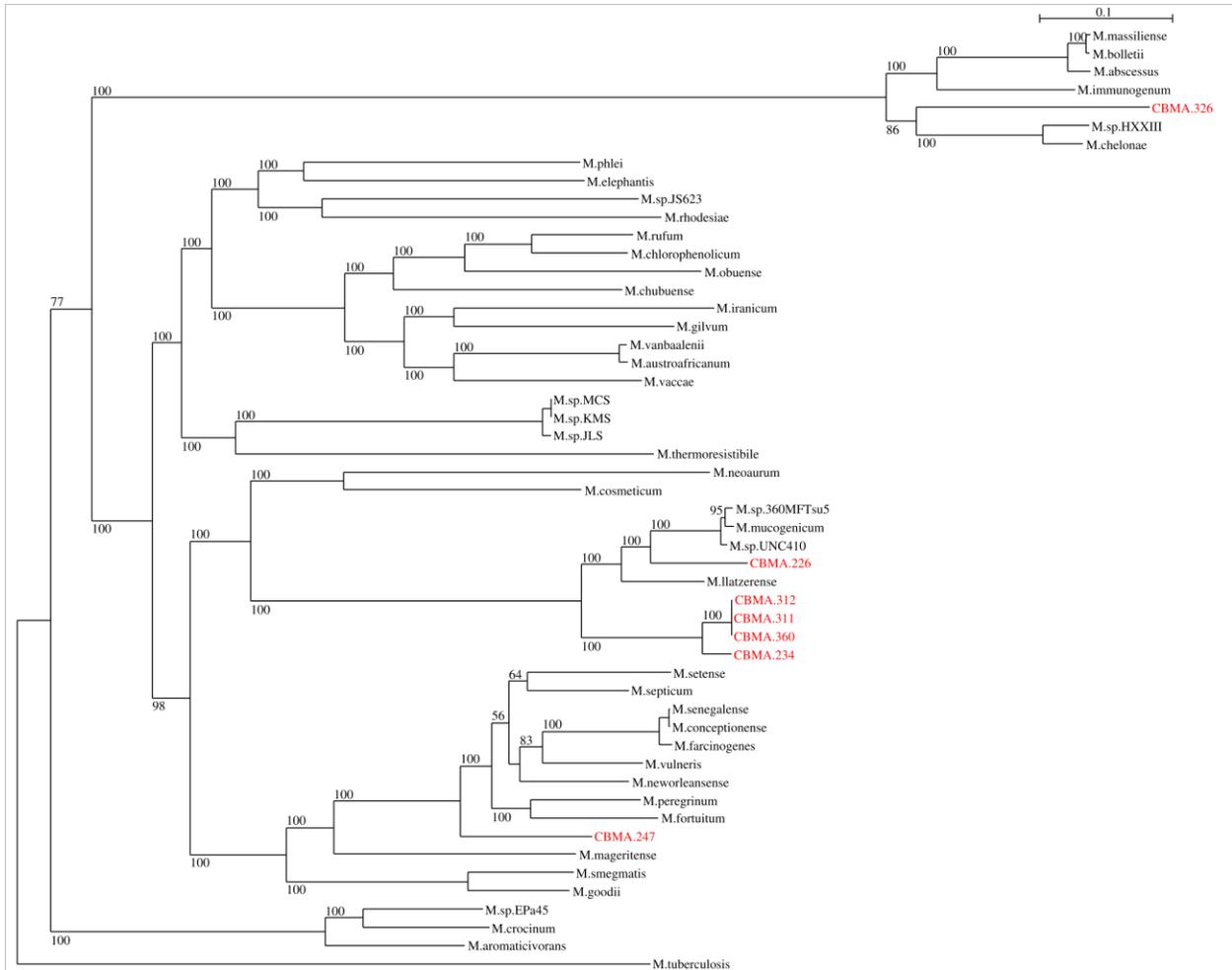


Figura 4.12. **Árvore de máxima verossimilhança da abordagem *cg*MLSA.** Sequências dos isolados CBMA estão marcadas em vermelho. 100 réplicas de *bootstrap*.

As sequências genômicas foram submetidas a análise DDH *in silico*. DDH <70% é indicativo de distintas espécies. Nossos resultados, apresentados na Tabela 4.9, indicam que os isolados dos três clusters e dos dois subclusters corresponderiam a distintas espécies. Já que os valores de DDH foram inferiores a 70% quando a partir de comparações com genomas completos disponíveis de espécies já descritas, e isto se refletiu também nas diferenças entre os conteúdos GC.

Tabela 4.9. Comparação de valores DDH entre genomas CBMA e espécies próximas, considerando cgMLSA.

Query	Vs	Referência	DDH (%)	Δ GC (%)
CBMA 234	x	CBMA 226	29,70 +/- 2,50	0,81
CBMA 234	x	CBMA 247	20,50 +/- 2,40	0,60
CBMA 234	x	CBMA 312	66,80 +/- 2,90	0,25
CBMA 234	x	CBMA 311	66,80 +/- 2,80	0,15
CBMA 234	x	CBMA 360	66,70 +/- 2,90	0,15
CBMA 234	x	CBMA 326	19,60 +/- 2,40	1,58
CBMA 234	x	<i>M. llatzerense</i>	30,70 +/- 2,45	0,93
CBMA 234	x	<i>M. sp.360MFTsu5</i>	29,90 +/- 2,45	1,55
CBMA 234	x	<i>M. sp.UNC410</i>	29,80 +/- 2,45	1,50
CBMA 234	x	<i>M. mucogenicum</i>	29,80 +/- 2,40	1,57
CBMA 312	x	CBMA 226	29,70 +/- 2,50	1,07
CBMA 312	x	CBMA 234	66,80 +/- 2,90	0,25
CBMA 312	x	CBMA 247	20,60 +/- 2,40	0,85
CBMA 312	x	CBMA 311	99,90 +/- 0,1	0,10
CBMA 312	x	CBMA 360	99,70 +/- 0,1	0,10
CBMA 312	x	CBMA 326	19,50 +/- 2,40	1,33
CBMA 312	x	<i>M. llatzerense</i>	30,50 +/- 2,45	1,18
CBMA 312	x	<i>M. sp.360MFTsu5</i>	29,70 +/- 2,45	1,80
CBMA 312	x	<i>M. sp.UNC410</i>	29,60 +/- 2,44	1,76
CBMA 312	x	<i>M. mucogenicum</i>	29,60 +/- 2,50	1,83
CBMA 311	x	CBMA 312	99,90 +/- 0,1	0,1
CBMA 311	x	CBMA 360	99,90 +/- 0,1	0,0
CBMA 360	x	CBMA 311	99,90 +/- 0,1	0,0
CBMA 360	x	CBMA 312	99,70 +/- 0,1	0,10
CBMA 247	x	CBMA 226	20,30 +/- 2,40	0,21
CBMA 247	x	CBMA 234	20,50 +/- 2,50	0,61
CBMA 247	x	CBMA 312	20,60 +/- 2,40	0,86
CBMA 247	x	CBMA 326	19,90 +/- 2,40	2,19
CBMA 247	x	<i>M. fortuitum</i>	29,80 +/- 2,45	0,02
CBMA 247	x	<i>M. nonchromogenicum</i>	29,60 +/- 2,44	0,09
CBMA 247	x	<i>M. septicum</i>	33,20 +/- 2,47	0,49
CBMA 247	x	<i>M. setense</i>	30,70 +/- 2,45	0,19
CBMA 247	x	<i>M. sp.VKM</i>	29,70 +/- 2,45	0,04
CBMA 247	x	<i>M. peregrinum</i>	31,30 +/- 2,50	0,0
CBMA 247	x	<i>M. mageritense</i>	23,40 +/- 2,50	0,72
CBMA 326	x	CBMA 226	19,40 +/- 2,40	2,40
CBMA 326	x	CBMA 234	19,60 +/- 2,40	1,58
CBMA 326	x	CBMA 247	19,90 +/- 2,40	2,19
CBMA 326	x	CBMA 312	19,50 +/- 2,40	1,33
CBMA 326	x	<i>M. abscessus</i>	27,70 +/- 2,40	0,09

Tabela 4.9. Comparação de valores DDH entre genomas CBMA (cont.).

CBMA 326	x	<i>M. abscessus.subsp.bolletii</i>	24,80 +/- 2,40	0,17
CBMA 326	x	<i>M. chelonae</i>	24,70 +/- 2,40	0,02
CBMA 326	x	<i>M. immunogenum</i>	25,20 +/- 2,40	0,22
CBMA 326	x	<i>M. massiliense</i>	24,80 +/- 2,40	0,28
CBMA 226	x	CBMA 234	29,70 +/- 2,50	0,81
CBMA 226	x	CBMA 247	20,30 +/- 2,40	0,21
CBMA 226	x	CBMA 312	29,70 +/- 2,50	1,07
CBMA 226	x	CBMA 326	19,40 +/- 2,40	2,40
CBMA 226	x	<i>M. llatzerense</i>	35,80 +/- 2,50	0,11
CBMA 226	x	<i>M. sp.360MFTsu5</i>	37,40 +/- 2,50	0,73
CBMA 226	x	<i>M. sp.UNC410</i>	37,40 +/- 2,50	0,69
CBMA 226	x	<i>M. mucogenicum</i>	37,30 +/- 2,50	0,76

A outra abordagem foi MiSI, na qual valores $\geq 0,6$ para AF (*Alignment Fraction*) indicam a possibilidade de ser a mesma espécie, o que é confirmado pelo teste final utilizando o parâmetro de gANI. O valor $\geq 96,5$ seria indicativo de uma mesma espécie. Os resultados dessa abordagem estão na Tabela 4.10

Tabela 4.10. Comparação dos valores de AF e gANI entre genomas CBAS e espécies próximas, considerando cgMLSA.

Query	Vs	Referência	AF	gANI
CBMA 234	x	CBMA 226	0,72	86,39
CBMA 234	x	CBMA 247	0,52	77,84
CBMA 234	x	CBMA 312	0,80	96,25
CBMA 234	x	CBMA 311	0,80	96,27
CBMA 234	x	CBMA 360	0,80	96,23
CBMA 234	x	CBMA 326	0,33	74,37
CBMA 234	x	<i>M. llatzerense</i>	0,73	86,86
CBMA 234	x	<i>M. sp.360MFTsu5</i>	0,73	86,36
CBMA 234	x	<i>M. sp.UNC410</i>	0,73	86,37
CBMA 234	x	<i>M. mucogenicum</i>	0,73	86,32
CBMA 312	x	CBMA 226	0,68	86,34
CBMA 312	x	CBMA 234	0,76	96,24
CBMA 312	x	CBMA 247	0,47	77,88
CBMA 312	x	CBMA 311	0,94	99,99
CBMA 312	x	CBMA 360	0,95	99,99
CBMA 312	x	CBMA 326	0,31	74,49
CBMA 312	x	<i>M. llatzerense</i>	0,66	86,78
CBMA 312	x	<i>M. sp.360MFTsu5</i>	0,67	86,24
CBMA 312	x	<i>M. sp.UNC410</i>	0,67	86,24
CBMA 312	x	<i>M. mucogenicum</i>	0,72	86,17
CBMA 311	x	CBMA 312	0,96	99,99

Tabela 4.10. Comparação dos valores de AF e gANI (cont.)

CBMA 311	x	CBMA 360	0,99	99,99
CBMA 360	x	CBMA 311	0,97	99,99
CBMA 360	x	CBMA 312	0,99	99,99
CBMA 247	x	CBMA 226	0,49	78,07
CBMA 247	x	CBMA 234	0,50	77,84
CBMA 247	x	CBMA 312	0,48	77,93
CBMA 247	x	CBMA 326	0,33	74,89
CBMA 247	x	<i>M. fortuitum</i>	0,69	86,47
CBMA 247	x	<i>M. nonchromogenicum</i>	0,62	86,37
CBMA 247	x	<i>M. septicum</i>	0,73	88,17
CBMA 247	x	<i>M. setense</i>	0,70	87,07
CBMA 247	x	<i>M. sp.VKM</i>	0,72	86,43
CBMA 247	x	<i>M. peregrinum</i>	0,73	87,23
CBMA 247	x	<i>M. mageritense</i>	0,67	81,94
CBMA 326	x	CBMA 226	0,42	74,73
CBMA 326	x	CBMA 234	0,43	74,38
CBMA 326	x	CBMA 247	0,45	74,94
CBMA 326	x	CBMA 312	0,42	74,49
CBMA 326	x	<i>M. abscessus</i>	0,79	83,01
CBMA 326	x	<i>M. abscessus.subsp.bolletii</i>	0,78	83,06
CBMA 326	x	<i>M. chelonae</i>	0,80	84,14
CBMA 326	x	<i>M. immunogenum</i>	0,79	83,10
CBMA 326	x	<i>M. massiliense</i>	0,78	83,06
CBMA 226	x	CBMA 234	0,71	86,40
CBMA 226	x	CBMA 247	0,51	78,03
CBMA 226	x	CBMA 312	0,72	86,34
CBMA 226	x	CBMA 326	0,33	74,66
CBMA 226	x	<i>M. llatzerense</i>	0,73	89,28
CBMA 226	x	<i>M. sp.360MFTsu5</i>	0,77	89,83
CBMA 226	x	<i>M. sp.UNC410</i>	0,77	89,71
CBMA 226	x	<i>M. mucogenicum</i>	0,77	89,83

Portanto, no conjunto de 14 isolados de micobactérias da Mata Atlântica haveria cinco novas espécies deste gênero que não apresentaram métricas compatíveis com nenhuma das espécies de micobactéria até agora descritas e sequenciadas. Em relação aos sete genomas avaliados, os genomas CBMA 311, 312 e 360 seriam de uma mesma espécie, enquanto que os demais corresponderiam a de espécies distintas. As métricas da CBMA 234 se encontram nos limites de definição de espécie quando comparadas com os genomas CBMA 311, 312 e 360, portanto poderiam ser definidas como espécies irmãs.

As Figuras 4.13 e 4.14 (geradas no Microsoft Office Excel 2007) resumem os resultados obtidos pelas ferramentas de taxonomia genômica, entre os genomas CBMA e com os genomas próximos, respectivamente. O par de genomas que apresenta valores abaixo das linhas de corte para AF (60%) ou gANI (96,5%), e DDH (70%), corresponderia a diferentes espécies; enquanto que para Δ GC% (5%), a mesma espécie.

Figura 4.13. Resolução taxonômica entre isolados CBMA.

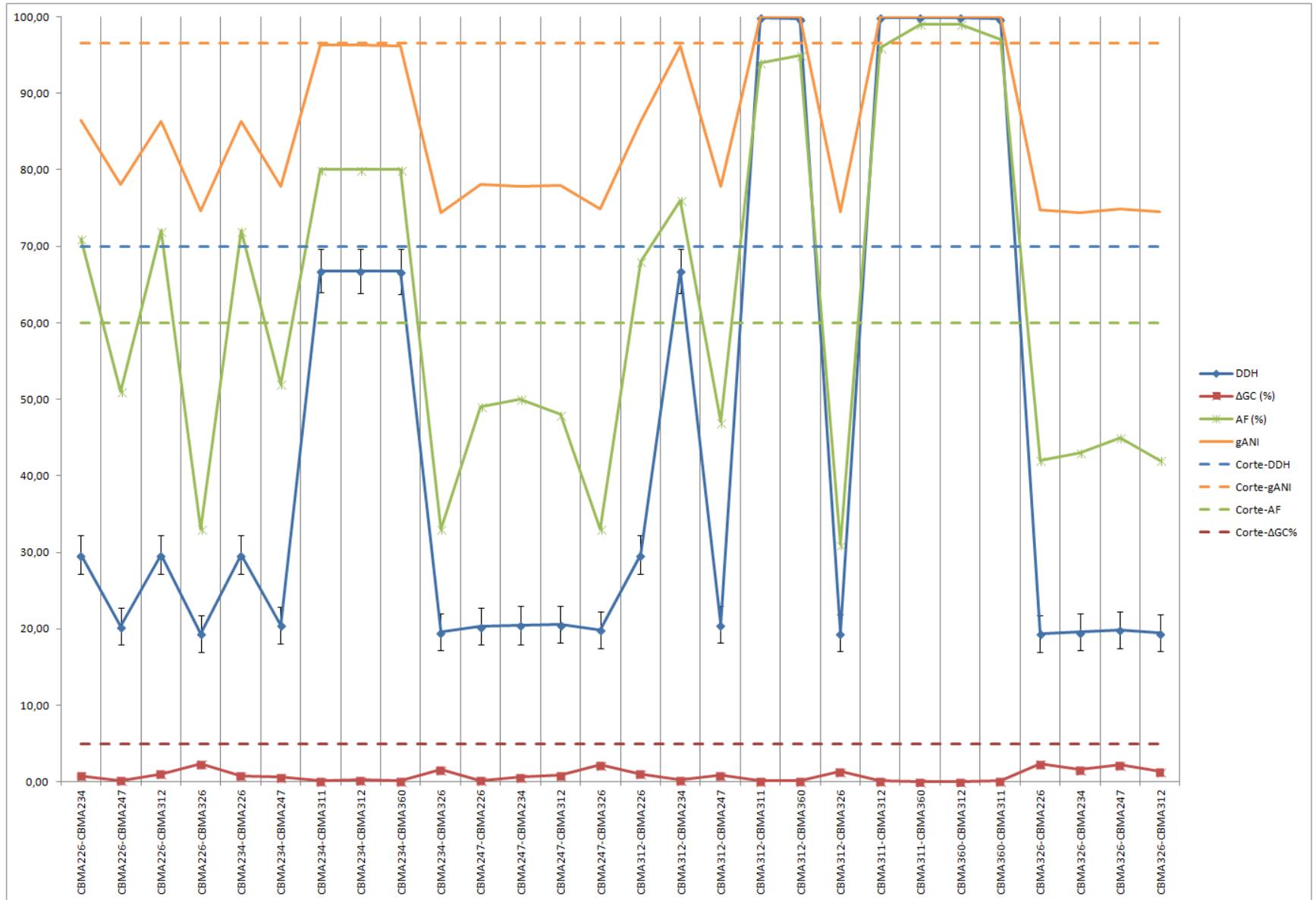
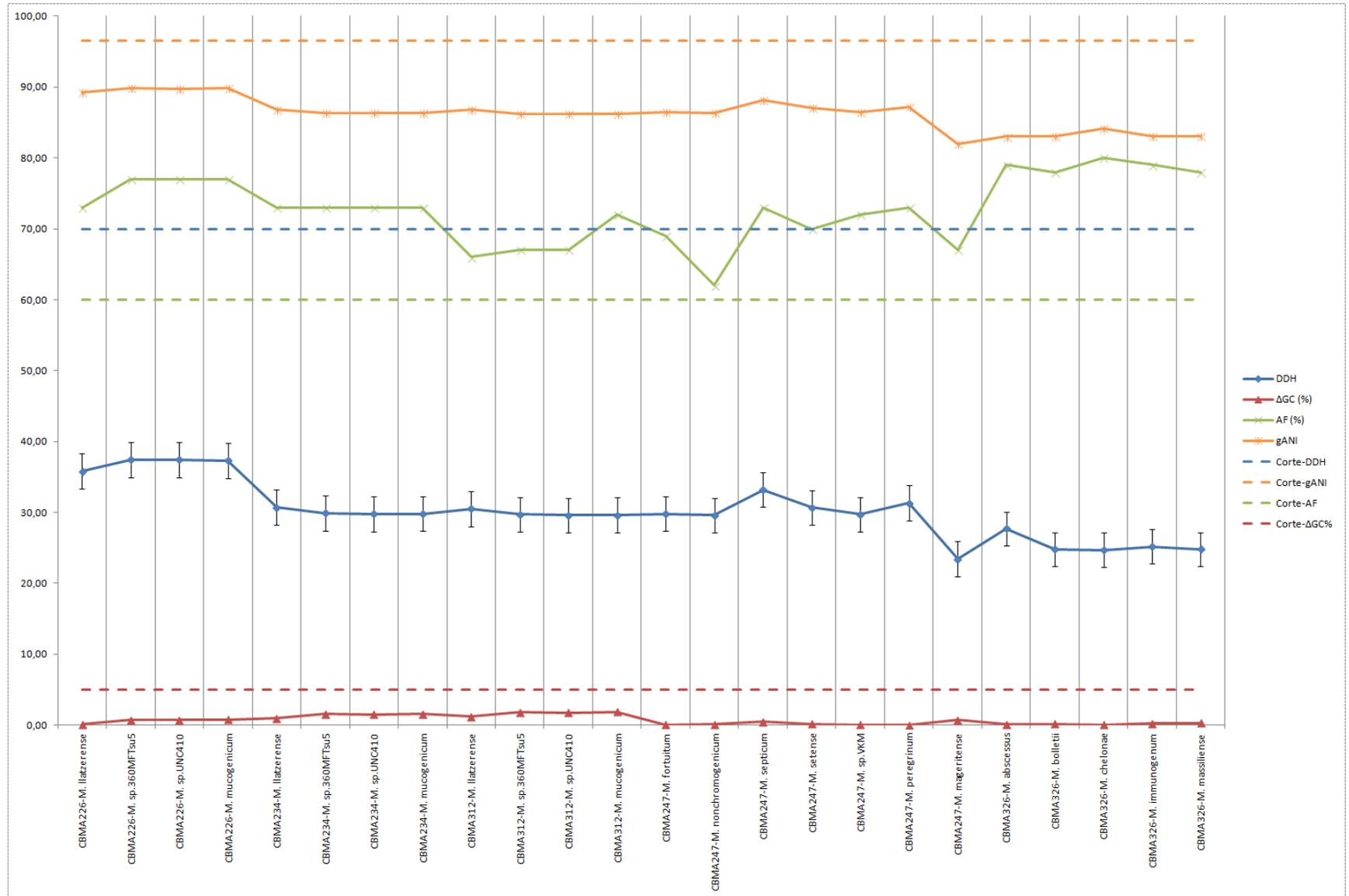


Figura 4.14. Resolução taxonômica entre isolados CBMA e espécies filogeneticamente próximas.



4.5 RESISTOMA

4.5.1 Predição *in silico* do resistoma

A anotação dos genomas em categorias funcionais, utilizando o RAST, identificou um conjunto de genes/mecanismos que estaria relacionado a resistência aos antibióticos de um modo geral. Além disto, existem bases de dados específicas para buscas de genes/mecanismos de resistência, como CARD e Resfams.

Em contraste com os resultados do RAST, que identificou de 32 a 46 genes nos sete genomas analisados (Figuras 4.3 a 4.9), o Resfams identificou centenas de genes associados à resistência nestes mesmos genomas, enquanto que a análise com o CARD revelou o menor número de genes dos três preditores. Realizamos a mesma busca em genomas da espécie *M. tuberculosis*, incluída como referência de espécie patogênica de crescimento lento, e de espécies relacionadas aos isolados CBMA: *M. setense* (CBMA 247), *M. abscessus* (CBMA 326) e *M. llatzerense* (CBMA 226, 234, 311, 312 e 360) (Tabela 4.11).

Tabela 4.11. Número de genes associados à resistência nas diferentes bases de dados.

Genoma	Resfams	CARD	RAST
CBMA 226	160	16	40
CBMA 234	167	20	44
CBMA 247	208	23	46
CBMA 311	169	22	43
CBMA 312	166	20	45
CBMA 326	155	13	32
CBMA 360	166	20	43
<i>M. tuberculosis</i>	128	16	22
<i>M. setense</i>	198	25	54
<i>M. abscessus</i>	162	15	37
<i>M. llatzerense</i>	158	14	52

As funções associadas aos genes preditos pertencem a vários dos mecanismos de resistência, incluindo tanto os mecanismos não-específicos, como bombas de efluxo, quanto os específicos, como por exemplo enzimas que modificam (fosfotransferases e ribosiltransferases) e hidrolisam (beta-lactamases) o antibiótico e

proteínas que protegem o alvo (pentapeptídeos) (lista completa de todos os genes/mecanismos preditos pelo ResFams e CARD relacionados à resistência encontram-se nos anexos 1 e 2, respectivamente).

Além disso, as predições também incluem genes e enzimas que são alvos de antibióticos, como a *gyrA*.

4.5.2 Perfil de sensibilidade das micobactérias CBMA

Realizamos análises *in vitro* para avaliação da sensibilidade destes isolados frente aos antimicrobianos para os quais as predições realizadas *in silico* revelaram a presença de um resistoma. Desta forma, os isolados foram testados para quatro classes de antimicrobianos: beta-lactâmicos (8 antibióticos), quinolonas (3 antibióticos), ansamicinas (1 antibiótico) e derivados de ácido fosfônico (1 antibiótico).

Inicialmente realizamos os experimentos por disco difusão, que geram resultados qualitativos, ou seja, informa se o isolado apresenta resistência ou sensibilidade a um antibiótico em função do tamanho do halo. Destacamos aqui que o isolado que apresentou o maior conjunto de resistências foi a CBMA 326, já que foi resistente a 9/13 antibióticos, contemplando todas as quatro classes de antimicrobianos.

O resistoma predito dos isolados CBMA 311, 312 e 360, que pertencem a mesma linhagem, foi o mesmo. Contudo, o isolado CBMA 311 foi resistente a apenas 4/13 antibióticos testados, enquanto que as CBMA 312 e 360 apresentaram duas resistências adicionais aos beta-lactâmicos piperacilina e cefepime, sendo o último uma cefalosporina de 4ª geração. Os resultados completos referente aos testes de disco-difusão estão detalhados na Figura 4.15.

Para alguns antibióticos realizamos também a análise quantitativa da sensibilidade através da determinação da concentração mínima inibitória dos antibióticos das classes de beta-lactâmicos, quinolonas e ansamicinas pelo método de E-Test. Neste ensaio verificamos que a CBMA 326, que apresenta o maior conjunto de resistências, também apresentou os maiores valores de MIC. Um outro resultado que destacamos é em relação à rifampicina, já que os três isolados que possuem o gene *arr* apresentaram valores de MIC entre 1 - 32 µg/ml (Tabela 4.12).

Tabela 4.12. E-test realizado nos isolados CBMA.

Antibiótico/CBMA	226	234	247	326
Beta-lactâmicos				
Amicacina (AK)	> 0,5	> 0,5	0,5	1
Amoxicilina (AMO)	0,12	0,03	> 8	256
Cefepime (CPM)	2	0,38	256	4
Ceftazidime (CAZ)	24	4	256	256
Imipinem (IPM)	0,12	> 0,06	2	0,5
Meropenem (MEM)	0,25	> 0,06	2	32
Quinolonas				
Ciprofloxacina (CIP)	0,15	0,008	0,01	0,3
Levofloxacina (LVX)	0,06	0,06	0,02	32
Ansamícinas				
Rifampicina (RIF)	1	12	-	32

Valores em µg/ml; Negrito: concentração máxima da fita; -, não realizado.

4.5.3 Associação resistoma (predição *in silico* a partir do genótipo) versus perfil de sensibilidade (fenótipo *in vitro*)

Considerando os fenótipos de resistência, realizamos a associação destes com o resistoma predito para cada um dos isolados CBMA. Obtivemos uma correlação direta para quatro classes de antibióticos: beta-lactâmicos (8 antibióticos), quinolonas (3 antibióticos), ansamícinas (1 antibiótico) e derivados de ácido fosfônico (1 antibiótico) (Figura 4.15). A associação detalhada do fenótipo dos isolados com os genes identificados no resistoma está na Tabela 4.13.

Classe	Antibiótico / Predito	CBMA							
		226	234	247	311	312	360	326	
Beta-lactâmicos	Fenótipo	Meropenem (MEM)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Cefalotina (CFL)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Imipinem (IPM)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Amoxicilina (AMO)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Piperacilina (PRL)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Ceftazidime (CAZ)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Aztreonam (ATM)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Cefepime (CPM)	Verde	Verde	Verde	Verde	Verde	Verde	Verde
	Genótipo	<i>typA</i>	Verde	Verde	Verde	Verde	Verde	Verde	Verde
		Beta-lactamase precursor Beta-lactamase metalo Beta-lactamase	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Quinolonas	Fenótipo	Ciprofloxacina (CIP)	Verde	Verde	Verde	Verde	Verde	Verde	
		Norfloxacina (NOR)	Verde	Verde	Verde	Verde	Verde	Verde	
		Ácido Nalidíxico (NAL)	Verde	Verde	Verde	Verde	Verde	Verde	
	Genótipo	<i>lrfA</i>	Verde	Verde	Verde	Verde	Verde	Verde	
		<i>mfd</i>	Verde	Verde	Verde	Verde	Verde	Verde	
		<i>mdtH</i>	Verde	Verde	Verde	Verde	Verde	Verde	
		<i>mfpA</i> <i>mfpB</i>	Verde	Verde	Verde	Verde	Verde	Verde	
Ansamícinas	Fenótipo	Rifampicina (RIF)	Verde	Verde	Verde	Verde	Verde		
	Genótipo	<i>rbpA</i>	Verde	Verde	Verde	Verde	Verde		
		P55 efflux pump <i>arr</i> <i>lri</i>	Verde	Verde	Verde	Verde	Verde		
		Fosfônicos	Fenótipo	Fosfomicina (FOS)	Verde	Verde	Verde	Verde	
Genótipo	<i>murA</i>	Verde	Verde	Verde	Verde	Verde			

Figura 4.15. Heatmap da associação do resistoma versus fenótipo de resistência das micobactérias CBMA. Os genes identificados *in silico* através do CARD, ResFams e RAST foram correlacionados ao perfil de sensibilidade obtidos pelo teste de disco-difusão. Na categoria fenótipo, encontram-se os antibióticos testados por disco de difusão; e na categoria genótipo, os genes identificados *in silico* em relação àquela classe. Verde, resistente ou presente; amarelo, sensível ou ausente.

Tabela 4.13. Fenótipo e genes de resistência identificados.

CBMA	Fenótipo de resistência	<i>typA</i>	Beta-lactamase precursor	Beta-lactamase	Metallo Beta-lactamase	<i>lrfA</i>	<i>mfd</i>	<i>mdtH</i>	<i>mfpA</i>	<i>mfpB</i>	<i>rbpA</i>	P55 efflux pump	<i>arr</i>	<i>iri</i>	<i>murA</i>
226	PRL ^a CAZ ^a ATM ^a NAL ^b RIF ^c FOS ^d	+	+	+	-	+	+	-	+	+	+	+	+	-	+
234	CFL ^a PRL ^a CAZ ^a ATM ^a NAL ^b RIF ^c FOS ^d	+	+	+	+	+	+	+	+	+	+	+	+	-	+
247	MEM ^a CFL ^a PRL ^a CAZ ^a ATM ^a CPM ^a NAL ^b FOS ^d	+	+	+	+	-	+	-	+	+	+	+	-	+	+
311	CAZ ^a ATM ^a NAL ^b FOS ^d	+	+	-	+	-	+	+	+	+	+	+	-	-	+
312	PRL ^a CAZ ^a ATM ^a CPM ^a NAL ^b FOS ^d	+	+	-	+	-	+	+	+	+	+	+	-	-	+
360	PRL ^a CAZ ^a ATM ^a CPM ^a NAL ^b FOS ^d	+	+	-	+	-	+	+	+	+	+	+	-	-	+
326	MEM ^a CFL ^a AMO ^a PRL ^a CAZ ^a ATM ^a CPM ^a NAL ^b RIF ^c FOS ^d	+	+	+	-	-	+	+	+	-	+	+	+	-	+

^a, beta-lactâmico; ^b, quinolona; ^c, ansamicina; ^d, fosfônico.

Em particular, uma correspondência entre o resistoma (gene *murA*) e o fenótipo (resistência a fosfomicina) foi observada em todos os isolados. Por outro lado, a resistência a quinolonas foi predita para todos os isolados, mas os mesmos foram resistentes a apenas um único antibiótico da classe das quinolonas (ácido nalidíxico). Similarmente, várias beta-lactamases foram identificadas como parte do resistoma de isolados resistentes a beta-lactâmicos, indicando uma provável influência destas beta-lactamases no fenótipo de resistência observado. Interessantemente, para a rifampicina, que é o antibiótico de primeira linha utilizado no tratamento de tuberculose, foi observada uma associação direta entre a presença do gene *arr* e o fenótipo de resistência a este antibiótico em 3/7 isolados CBMA.

4.5.4 Funcionalidade dos genes associados às resistências em sistema heterólogo

A funcionalidade de alguns dos genes de resistência identificados, cuja associação com o fenótipo de resistência foi observada, foi acessada através de ensaios de expressão em sistema heterólogo.

4.5.4.1 Resistência a rifampicina

A resistência a rifampicina pode emergir através da alteração do alvo do antibiótico - RNA polimerase - como consequência de mutações no gene codificante *rpoB*, e através da modificação do antibiótico pela ADP-ribosiltransferase, codificada pelo gene *arr*. Os isolados CBMA 226, 234 e 326, que apresentam MICs de rifampicina distintas, carregam o gene *rpoB* canônico, ou seja, com nenhum polimorfismo nos sítios RRDR (*Rifampin Resistance Determining Region*) relacionados à resistência. Contudo, estes três isolados carregam distintos alelos do gene *arr* (Figura 4.16), que se encontram em ambientes genéticos diferentes (Figura 4.17). Desta forma, o papel destes alelos *arr* no fenótipo de resistência a rifampicina foi avaliado em sistema heterólogo.

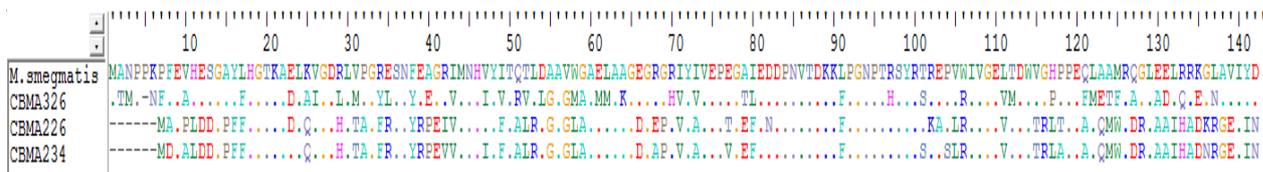


Figura 4.16. Alinhamento de aminoácidos codificados pelo gene *arr* de *M. smegmatis* e isolados CBMA.

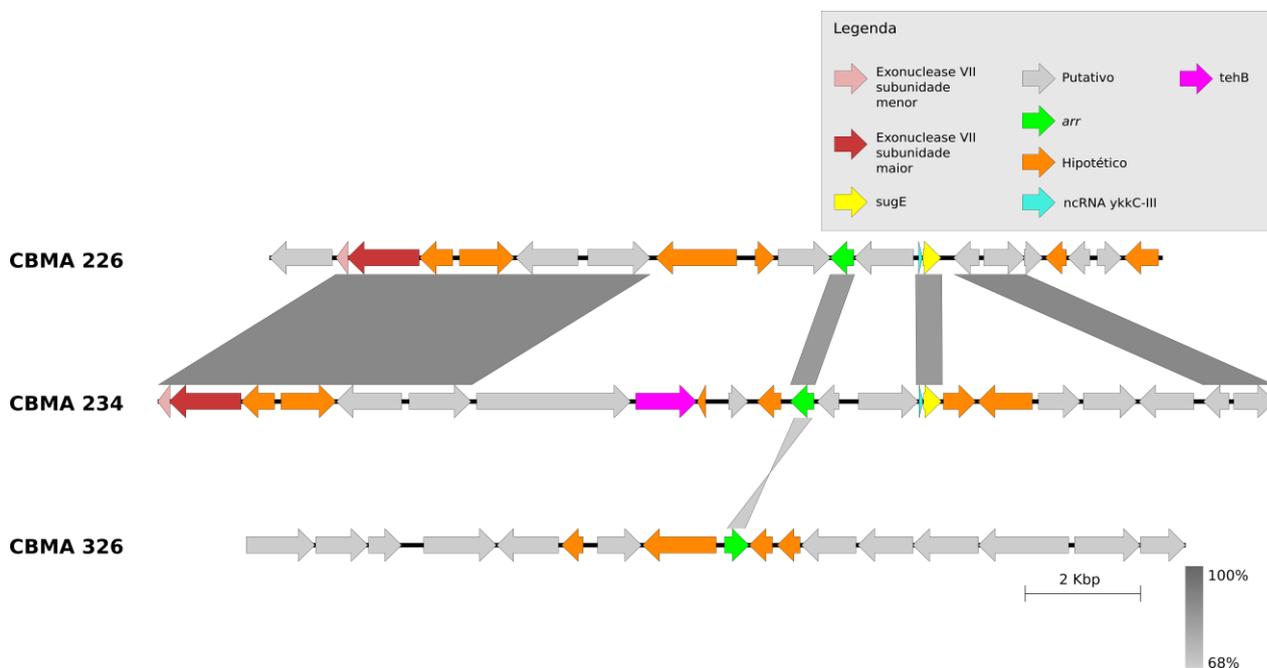


Figura 4.17. Ambientes genéticos do gene *arr* nos isolados CBMA 226, 234 e 326.

Os alelos *arr* 234 e *arr* 326 não foram determinantes para alteração do perfil de sensibilidade a rifampicina no sistema heterólogo. Portanto estes alelos de *arr* não estariam associados ao fenótipo de resistência das CBMA 234 e 326.

4.5.4.2 Resistência a quinolonas

Mantendo esta lógica, fomos avaliar a associação fenótipo/genótipo relativo às quinolonas. A resistência a quinolonas pode emergir através de diversos mecanismos, sendo o principal deles a modificação do alvo do antibiótico (*GyrA/B*). As quinolonas tem como alvo os produtos dos genes intrínsecos *gyrA/B*. A ocorrência de mutações não-sinônimas na região QRDR (*Quinolone Resistance Determining Region*) destes genes levam à resistência aos antibióticos quinolonas. Analisamos a sequência QRDR dos genes *gyrA/B* de todos os isolados CBMA e verificamos que elas eram idênticas e não apresentavam as mutações associadas a resistência. Dois polimorfismos, A92S e S97T (posições 90 e 95, respectivamente,

na numeração de GyrA de *M. tuberculosis*), foram observados na sequência aminoacídica de GyrA em todos os isolados, contudo, nenhuma associação entre estas alterações e resistência a quinolonas foi reportada até o momento (Avalos et al. 2015) (Figuras 4.18 e 4.19).

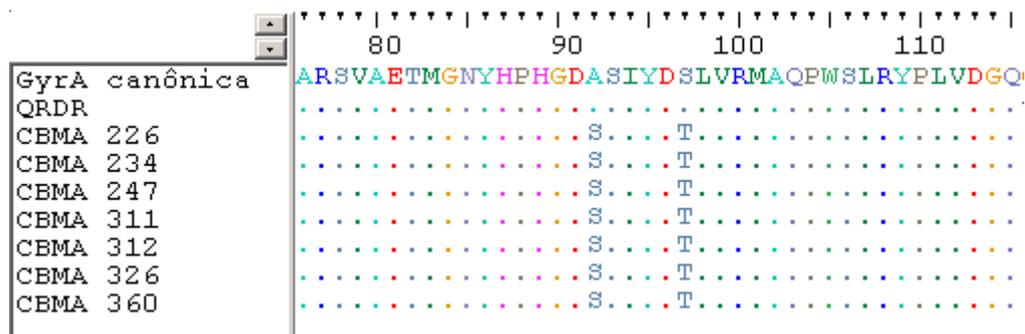


Figura 4.18. Alinhamento de aminoácidos da região QRDR de GyrA dos isolados CBMA.

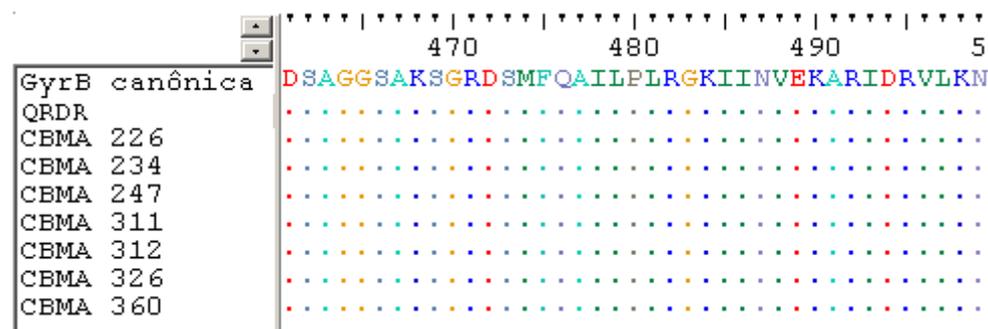


Figura 4.19. Alinhamento de aminoácidos da região QRDR de GyrB dos isolados CBMA.

Como todos os isolados apresentaram resistência ao ácido nalidíxico, e dentro dos genes identificados como parte do resistoma associados a esta classe de antibióticos, temos os genes *mfpA/B*. Verificamos a presença de cinco/quatro alelos de *mfpA/B* distribuídos entre os sete isolados (Figuras 4.20 e 4.21). Representantes dos distintos alelos foram avaliados no sistema heterólogo.

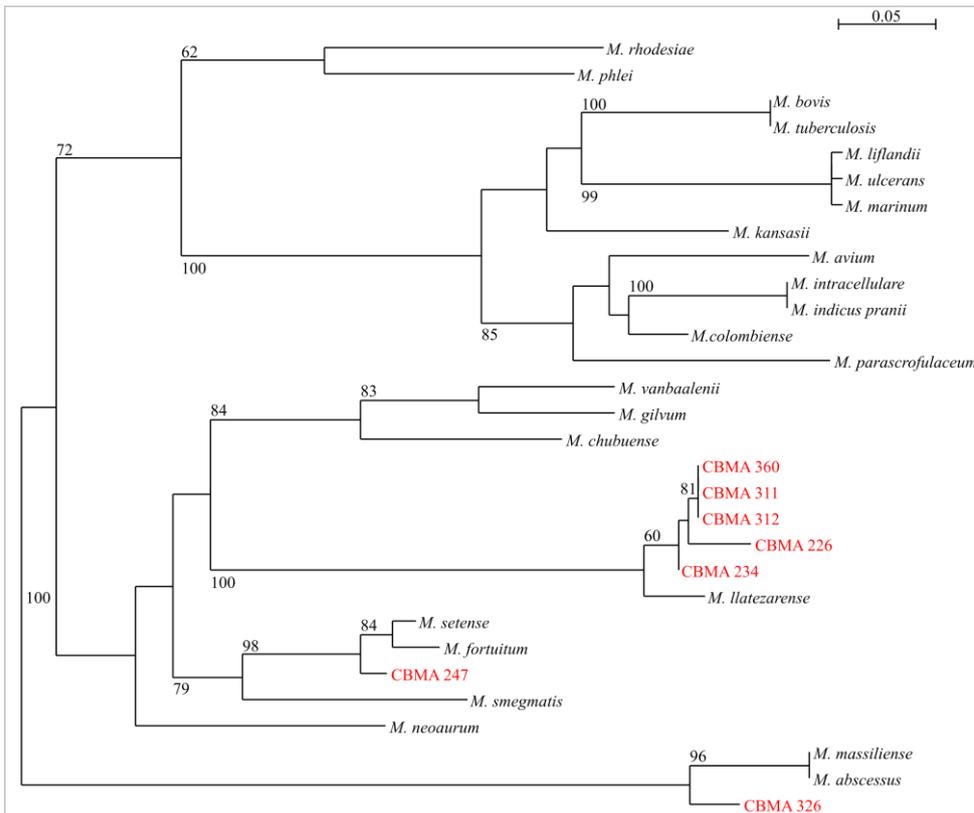


Figura 4.20. **Árvore de máxima verossimilhança da proteína MfpA de micobactérias.** Os isolados CBMA encontram-se destacados em vermelho. 100 replicatas de *bootstrap*.

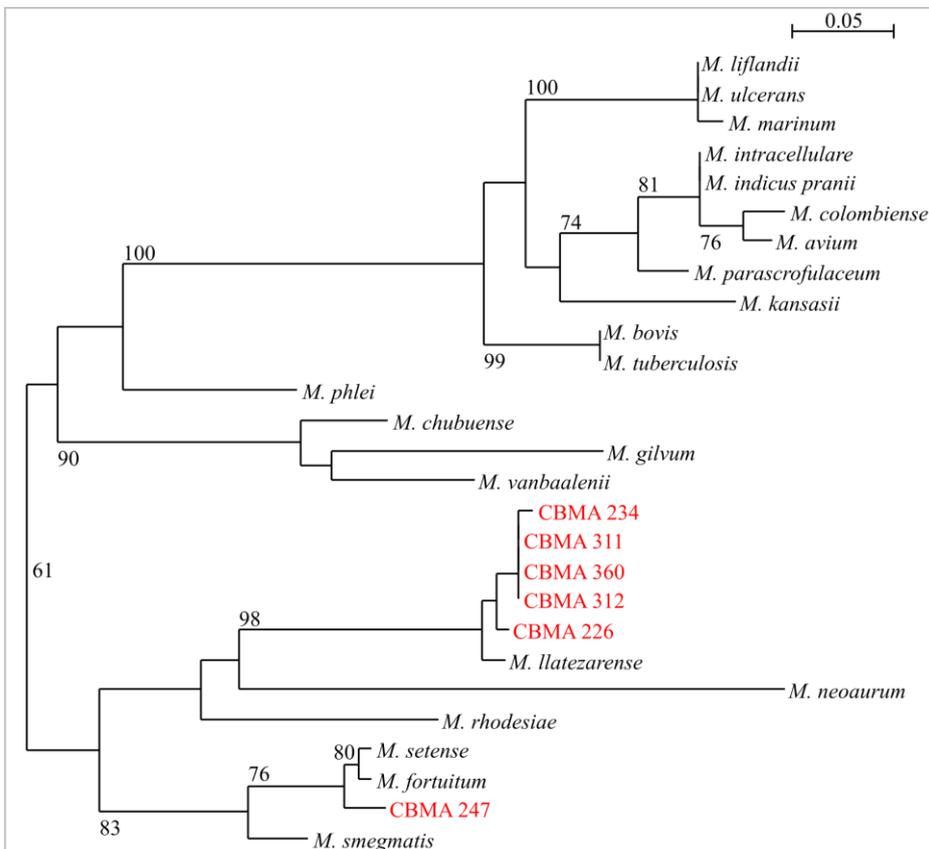


Figura 4.21. **Árvore de máxima verossimilhança da proteína MfpB de micobactérias.** Os isolados CBMA encontram-se destacados em vermelho. 100 replicatas de *bootstrap*.

Foi observada uma redução do diâmetro dos halos (diminuição da sensibilidade) de ácido nalidíxico, norfloxacin e ciprofloxacina no ensaio com as transformantes carregando os alelos *mfpA/B* 360 e *mfpA/B* 247 em relação aos halos obtidos nos ensaios com a *E. coli* selvagem (Tabela 4.14). Considerando que as micobactérias CBMA foram resistentes ao ácido nalidíxico (ausência de halo - resistência total), e que estes alelos de *mfpA/B* estão relacionados a uma diminuição da sensibilidade, podemos inferir que outro mecanismo de resistência a quinolonas está envolvido no fenótipo observado.

Tabela 4.14. Perfil de resistência a quinolonas das transformantes carregando os diferentes alelos *mfpA/B*.

Alelo	CIP	NAL	NOR
	(mm)		
MC1061 + <i>mfpA/B</i> 360	44	27	36
MC1061 + <i>mfpA/B</i> 247	39	27	39
MC1061	52	36	44

CIP, ciprofloxacina; NAL, ácido nalidíxico; NOR, norfloxacin.

4.5.4.3 Resistência a beta-lactâmicos

Várias beta-lactamases foram identificadas na predição *in silico*. Uma, em especial, na CBMA 326, apresentava alta identidade (80%) e cobertura (86%) com uma beta-lactamase de espectro estendido identificada e caracterizada em *M. abscessus* (*blaMab*). Como, de fato, a CBMA 326 apresenta o maior espectro de resistência aos beta-lactâmicos (Figura 4.15), fomos avaliar a funcionalidade deste gene (*bla326*) em sistema heterólogo.

Os resultados mostram que *bla326* é funcional e apresenta um espectro restrito de atividade para os beta-lactâmicos, já que confere resistência a cefalosporinas de segunda geração. De fato, foi observada perda total de sensibilidade das transformantes à cefalotina, que apresentaram uma MIC de 32 µg/ml para cefalotina em relação a MIC de 1,5 µg/ml observada na *E. coli* selvagem.

5 DISCUSSÃO

O bioma Mata Atlântica é um dos principais concentradores da biodiversidade no mundo. A microbiota do solo é o componente chave para a regulação e o funcionamento de um bioma. No único estudo, com abordagem genética, acerca da diversidade microbiana do solo deste bioma, no Rio de Janeiro, os principais filos de bactérias recuperados foram Acidobacteria, Proteobacteria e Verrucomicrobia. Além destes, um número limitado de isolados pertencentes aos Actinomycetes foi identificado, e dentre estes alguns pertenciam ao gênero *Mycobacterium* (Bruce et al. 2010). Considerando ambientes naturais, a diversidade de microrganismos é praticamente desconhecida. Particularmente em relação ao gênero *Mycobacterium*, os poucos estudos de prevalência e diversidade em ambientes naturais estão limitados a poucas regiões geográficas, a maioria delas no hemisfério norte (Khera 2012). Nestes estudos, poucas espécies já descritas, *Mycobacterium fortuitum*, *Mycobacterium senegalense*, *Mycobacterium chubuense*, *M. septicum*, entre outras, foram identificadas com base no gene 16S rRNA (Niva et al. 2006, Kopecky et al. 2011, Azadi et al. 2016).

Em nosso estudo, a partir da informação global do genoma *core*, determinamos a taxonomia e a diversidade genética de isolados do gênero *Mycobacterium*, provenientes do solo da Mata Atlântica. Curiosamente, nenhuma espécie já descrita foi identificada. Dentre os 14 isolados, todos correspondiam a novas espécies do gênero *Mycobacterium*, em um total de cinco novas espécies. Esta classificação foi possível através da taxonomia genômica, que permite acessar a diversidade de espécies, tanto cultiváveis, quanto não cultiváveis, inclusive a partir de dados metagenômicos (Thompson et al. 2013b). Até o presente, nenhum estudo de taxonomia genômica abordou o gênero *Mycobacterium* de modo geral, contudo, recentes estudos com isolados de micobactérias, tanto de crescimento rápido, quanto de crescimento lento, propuseram novas espécies neste gênero com base em análises genômicas (Ngeow et al. 2015, Choo et al. 2016). Atualmente, inclusive na clínica, para a identificação de micobactérias, é necessário informação genética, além das fenotípicas (Forbes 2016).

Quanto à diversidade das micobactérias CBMA, os 14 isolados foram agrupados em cinco espécies, duas delas tinham vários representantes. Uma dessas espécies possui relação com o grupo *M. septicum*, enquanto que a outra tem relação com o grupo *M. llatzerense*. *M. septicum* está presente em vários nichos,

incluindo ambiente e pacientes com doença pulmonar (Chin'ombe et al. 2016). Enquanto que *M. llatzerense* tem sido identificada, eventualmente, em amostras de solo e água, e recentemente também de casos clínicos de abscessos em humanos (Cárdenas et al. 2014, Dziejzinska et al. 2016). Uma das novas espécies, representada por um único isolado, CBMA 326, está relacionada filogeneticamente ao grupo da *M. abscessus*. Dentre as micobactérias de crescimento rápido, esta espécie é a mais prevalente em casos clínicos humanos (Lefebvre et al. 2016), sendo considerada um patógeno emergente devido também ao amplo conjunto de resistência aos antibióticos (Nessar et al. 2012).

A caracterização de plasmídeos e bacteriófagos de nosso estudo, ou seja, elementos do genoma acessório (Hacker & Kaper 2000), foi consequência do uso de programas específicos, como plasmidSPAdes e PHASTER, respectivamente. Observamos que anotadores, como o RAST, subestimaram a presença de plasmídeos nas CBMA, o que pode ser devido a falta de referências para plasmídeos de micobactérias em sua base de dados. Portanto, tanto para a montagem, quanto anotação, deve-se experimentar várias plataformas e/ou programas, no caso deste gênero de bactéria, em particular.

Existem milhares de micobacteriófagos descritos no gênero *Mycobacterium*, mas a maioria absoluta deles foi identificada em *M. smegmatis* (<http://phagesdb.org>). Portanto, os dois genomas completos de micobacteriófagos caracterizados neste estudo representariam os primeiros micobacteriófagos de uma espécie relacionada ao grupo da *M. llatzerense*. Além disso, quatro dos cinco plasmídeos identificados, estão nesta nova espécie, relacionada ao grupo da *M. llatzerense*. O outro plasmídeo (P4), foi identificado na CBMA226, que é uma nova espécie, também relacionada a este grupo. É interessante que até o momento, nenhum micobacteriófago e/ou plasmídeo foi identificado em espécies deste grupo. Portanto, nosso estudo foi o primeiro a identificar e caracterizar elementos do mobiloma relacionado ao grupo da *M. llatzerense*.

Considerando a diversidade de espécies de micobactéria, um número reduzido (29) de plasmídeos foi identificado até 2015 (Shintani et al. 2015). Plasmídeos de modo geral são caracterizados como lineares ou circulares; transferíveis ou mobilizáveis; e por grupo de incompatibilidade (Smillie et. al 2010). Em algumas das espécies de micobactéria, já foram identificados tanto plasmídeos circulares, quanto lineares (Ventura et al. 2007, Leão et al. 2013). Em relação à caracterização de plasmídeos transferíveis ou mobilizáveis, um dos marcadores é o

gene da relaxase, que pertence ao conjunto de genes de mobilização (MOB). Este marcador está presente em alguns dos plasmídeos já caracterizados em micobactérias (Ummels et al. 2014, Wang et al. 2015), mas não em todos.

Apesar de já terem sido identificados eventos de conjugação entre diferentes espécies de micobactéria, o conjunto de mecanismos/genes envolvido neste processo ainda não foi totalmente definido (Rabello et al. 2012, Ummels et al. 2014). Em um dos plasmídeos circulares (P2 - presente em CBMA 312 e 360), foi identificado um conjunto de genes relacionados à conjugação, semelhantes a genes de um plasmídeo de *M. marinum* (Ummels et al. 2014). A caracterização completa de P2 está no manuscrito submetido à revista Memórias do Instituto Oswaldo Cruz (anexo 3). Interessantemente, o plasmídeo P4, identificado no isolado CBMA 226, compartilha genes com P2, o que pode ser uma evidência de transferência lateral de genes mediada por plasmídeos entre estas duas novas espécies.

A emergência de resistência a partir da transferência de genes identificados em resistomas ambientais para clínica, assim como entre bactérias, é principalmente definida pelos determinantes de resistência associados com atividade direta de degradação ou modificação de antibióticos. Este novo fenótipo de resistência seria consequência de um único evento de transferência lateral de gene (Munita & Arias 2016). Assim, considerando os genes/mecanismos de resistência identificados nos genomas CBMA, foi possível analisar em detalhes, ou seja, associar genótipo ao fenótipo, dois mecanismos enzimáticos distintos, identificados no resistoma das CBMA: degradação de antibiótico (*bla326*) e modificação de antibiótico (*arr*).

Em uma linhagem de *M. abscessus* foi caracterizado um gene (*blaMab*) codificador de uma beta-lactamase de classe A que apresentava um espectro de atividade estendido (Soroka et al. 2013). Por sua vez, a beta-lactamase de classe A, *bla326*, identificada neste estudo, que tem identidade de 80% com *blaMab*, apresentou um espectro de atividade restrito. Curiosamente, a CBMA 326 apresenta resistência aos beta-lactâmicos amoxicilina, cefepime, ceftzidime e meropenem. Portanto, alguma outra beta-lactamase, dentre as demais identificadas no seu resistoma, e/ou outros mecanismos de resistência seriam os determinantes deste fenótipo.

Em um isolado do patógeno oportunista *M. smegmatis* foi identificada uma ADP-ribosiltransferase codificada pelo gene cromossomal *arr*. Esta ribosiltransferase atua modificando o antibiótico rifampicina, e o alelo presente neste isolado está associado a uma MIC de 256 µg/ml para rifampicina (Baysarowich et al. 2008). Nos

resistomas das CBMA 226, 234 e 326 foram identificados diferentes alelos do gene *arr*, e as mesmas apresentam distintos fenótipos de resistência para rifampicina: 1 µg/ml, 12 µg/ml e 32 µg/ml, respectivamente. Contudo, estes alelos não foram associados à perda de sensibilidade/resistência a rifampicina, logo estes fenótipos seriam consequência de outros mecanismos. De fato, mecanismos de efluxo de drogas, como por exemplo, *pstB*, Rv1258c e *efpA*, que estão associados a resistência a rifampicina em *M. tuberculosis* (Louw et al. 2009), foram identificados no resistoma destes genomas.

6 CONCLUSÕES

- Análises de sequências genômicas permite acessar a diversidade de linhagens e espécies, assim como realizar predições relacionadas ao resistoma e mobiloma;
- O nicho avaliado do bioma Mata Atlântica contém novas espécies de micobactéria;
- As métricas aplicadas foram consistentes para definir as espécies novas de micobactérias neste conjunto de genomas;
- Há uma diversidade de plasmídeos na espécie relacionada ao grupo *M. latzerense*, no qual ainda não havia identificação destes elementos.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Ahmed W, Staley C, Hamilton KA, Beale DJ, Sadowsky MJ, Toze S, et al. Amplicon-based taxonomic characterization of bacteria in urban and peri-urban roof-harvested rainwater stored in tanks. *Sci Total Environ.* 2017; 576:326-334. doi: 10.1016/j.scitotenv.2016.10.090.

Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* 2011;12:402. doi:10.1186/1471-2164-12-402.

Allen HK, Donato J, Wang HH, Cloud-Hansen KA, Davies J, Handelsman J. Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol.* 2010; Apr;8(4):251-9. doi: 10.1038/nrmicro2312.

Amaral GR, Dias GM, Wellington-Oguri M, Chimetto L, Campeão ME, Thompson FL, et al. Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. *Int J Syst Evol Microbiol.* 2014 Feb;64(Pt 2):357-65. doi: 10.1099/ijs.0.057927-0.

Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics.* 2016 Nov 15;32(22):3380-3387.

Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research.* 2016;44(Web Server issue):W16-W21. doi:10.1093/nar/gkw387.

Auch AF, Klenk H-P, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Standards in Genomic Sciences.* 2010;2(1):142-148. doi:10.4056/sigs.541628.

Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence

comparison. *Standards in Genomic Sciences*. 2010;2(1):117-134. doi:10.4056/sigs.531120.

Avalos E, Catanzaro D, Catanzaro A, et al. Frequency and Geographic Distribution of *gyrA* and *gyrB* Mutations Associated with Fluoroquinolone Resistance in Clinical Mycobacterium Tuberculosis Isolates: A Systematic Review. Kranzer K, ed. *PLoS ONE*. 2015;10(3):e0120470. doi:10.1371/journal.pone.0120470.

Azadi D, Shojaei H, Pourchangiz M, Dibaj R, Davarpanah M, Naser AD. Species diversity and molecular characterization of nontuberculous mycobacteria in hospital water system of a developing country, Iran. *Microb Pathog*. 2016 Nov;100:62-69. doi: 10.1016/j.micpath.2016.09.004.

Aziz RK, Bartels D, Best AA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*. 2008;9:75. doi:10.1186/1471-2164-9-75.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol*. 2012; 19(5): 455–477.

Baysarowich J, Koteva K, Hughes DW, et al. Rifamycin antibiotic resistance by ADP-ribosylation: Structure and diversity of Arr. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(12):4886-4891. doi:10.1073/pnas.0711939105.

Boolchandani M, Patel S, Dantas G. Functional Metagenomics to Study Antibiotic Resistance. *Methods Mol Biol*. 2017;1520:307-329.

Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ: Universal trees based on large combined protein sequence data sets. *Nature Genet* 2001, 28:281–285.

Bruce T, Martinez I, Maia Neto O, Vicente A, Kruger R, Thompson F. Bacterial community diversity in the Brazilian Atlantic forest soils. *Microb Ecol*. 2010;60:840–849

Cárdenas AM, Gomila M, Lalucat J, Edelstein PH. Abdominal Abscess Caused by Mycobacterium llatzerense. Carroll KC, ed. Journal of Clinical Microbiology. 2014;52(4):1287-1289. doi:10.1128/JCM.03525-13.

Chin'ombe N, Muzividzi B, Munemo E, Nziramasanga P. Molecular Identification of Nontuberculous Mycobacteria in Humans in Zimbabwe Using 16S Ribosequencing. The Open Microbiology Journal. 2016;10:113-123. doi:10.2174/1874285801610010113.

Choo SW, Dutta A, Wong GJ, Wee WY, Ang MY, Siow CC. Comparative Genomic Analysis Reveals a Possible Novel Non-Tuberculous Mycobacterium Species with High Pathogenic Potential. Floto AR, ed. PLoS ONE. 2016;11(4):e0150413. doi:10.1371/journal.pone.0150413.

Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol. 2014 Feb;64(Pt 2):316-24. doi: 10.1099/ijs.0.054171-0.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998 Jun 11;393(6685):537-44.

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001 Feb 22;409(6823):1007-11.

Colwell R.R. Polyphasic Taxonomy of Bacteria. In: Baltimore MD, editor. Symposium on Taxonomic Studies of Microorganisms by Instrumental Analysis of their Components and Metabolites International Conference on Culture Collections. Tokyo: University Park Press; 1968.

Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. Applied and Environmental Microbiology. 2013;79(24):7696-7701. doi:10.1128/AEM.02411-13.

D'Costa VM, McGrann KM, Hughes DW, Wright GD: Sampling the antibiotic resistome. *Science*. 2006, 311:374-377.

Duranti S, Lugli GA, Mancabelli L, Turrone F, Milani C, Mangifesta M. Prevalence of antibiotic resistance genes among human gut-derived bifidobacteria. *Appl Environ Microbiol*. 2016 Nov 18. pii: AEM.02894-16.

Dutilh BE, Backus L, Edwards RA, Wels M, Bayjanov JR, van Hijum SAFT. Explaining microbial phenotypes on a genomic scale: GWAS for microbes. *Briefings in Functional Genomics*. 2013;12(4):366-380. doi:10.1093/bfpgp/elt008.

Dziedzinska R, Makovcova J, Kaevska M, Slany M, Babak V, Moravkova M. Nontuberculous Mycobacteria on Ready-to-Eat, Raw and Frozen Fruits and Vegetables. *J Food Prot*. 2016 Aug;79(8):1452-6. doi: 10.4315/0362-028X.JFP-16-030.

Fernández-Cabezón L, Galán B, García JL. Engineering Mycobacterium smegmatis for testosterone production. *Microb Biotechnol*. 2016 Nov 17. doi: 10.1111/1751-7915.12433.

Forbes BA. Mycobacterial Taxonomy. *J Clin Microbiol*. 2016; Dec 7. pii: JCM.01287-16.

Galán B, Uhía I, García-Fernández E, Martínez I, Bahillo E, de la Fuente JL, et al. Mycobacterium smegmatis is a suitable cell factory for the production of steroidic synthons. *Microb Biotechnol*. 2016 Nov 2. doi: 10.1111/1751-7915.12429.

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, et al. Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol*. 2005; 3(9):733-9.

Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *The ISME Journal*. 2015;9(1):207-216. doi:10.1038/ismej.2014.106.

Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol.* 2015; 38(4):237-45. doi: 10.1016/j.syapm.2015.03.007.

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007; 57: 81–91.

Gouy M, Guindon S, Gascuel O. SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 2010 Feb;27(2):221-4. doi: 10.1093/molbev/msp259.

Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrobial Agents and Chemotherapy.* 2014;58(1):212-220. doi:10.1128/AAC.01310-13.

Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol.* 2000;54:641-79.

Hall-Stoodley L, Lappin-Scott H. Biofilm formation by the rapidly growing mycobacterial species *Mycobacterium fortuitum*. *FEMS Microbiol. Lett.* 1998; 168: 77–84.

Hernandez D, François P, Farinelli L, Østerås M, Schrenzel J. De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research.* 2008;18(5):802-809. doi:10.1101/gr.072033.107.

Honda JR, Hasan NA, Davidson RM, et al. Environmental Nontuberculous Mycobacteria in the Hawaiian Islands. Picardeau M, ed. *PLoS Neglected Tropical Diseases.* 2016;10(10):e0005068. doi:10.1371/journal.pntd.0005068.

Karlin S, Burge C. Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet.* 1995; 11(7):283-90.

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 2002;30(14):3059-3066.

Kelley DR, Schatz MC, Salzberg SL. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol*. 2010; 11(11): R116.

Khera T. The diversity and distribution of *Mycobacterium* species in varying ecological and climatic environments. School of Life Sciences [thesis]. Coventry (CV): University of Warwick; 2012.

Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 2005; 102:2567–2572.

Kopecky J, Kyselkova M, Omelka M, Cermak L, Novotna J, Grundmann G, et al. Environmental *Mycobacteria* Closely Related to the Pathogenic Species Evidenced in an Acidic Forest Wetland. *Soil Biology and Biochemistry*. 2001; 43, 697-700.

Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*. 2004;32(1):11-16. doi:10.1093/nar/gkh152.

Leão SC, Matsumoto CK, Carneiro A, et al. The Detection and Sequencing of a Broad-Host-Range Conjugative IncP-1 β Plasmid in an Epidemic Strain of *Mycobacterium abscessus* subsp. *bolletii*. Ahmed N, ed. *PLoS ONE*. 2013;8(4):e60746. doi:10.1371/journal.pone.0060746.

Lee H, Kim B-J, Kim B-R, Kook Y-H, Kim B-J. The Development of a Novel *Mycobacterium-Escherichia coli* Shuttle Vector System Using pMyong2, a Linear Plasmid from *Mycobacterium yongonense* DSM 45126T. Chatterji D, ed. *PLoS ONE*. 2015;10(3):e0122897. doi:10.1371/journal.pone.0122897.

Lefebvre AL, Dubée V, Cortes M, Dorchène D, Arthur M, Mainardi JL. Bactericidal and intracellular activity of β -lactams against *Mycobacterium abscessus*. *J Antimicrob Chemother*. 2016 Jun;71(6):1556-63. doi: 10.1093/jac/dkw022.

Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*. 2003;13(9):2178-2189. doi:10.1101/gr.1224503.

Lin S-H, Liao Y-C. CISA: Contig Integrator for Sequence Assembly of Bacterial Genomes. Watson M, ed. *PLoS ONE*. 2013;8(3):e60843. doi:10.1371/journal.pone.0060843.

Liu B, Pop M. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Research*. 2009;37(Database issue):D443-D447. doi:10.1093/nar/gkn656.

Löffler FE, Yan J, Ritalahti KM et al. *Dehalococcoides mccartyi* gen. nov., sp. nov., obligate organohalide-respiring anaerobic bacteria, relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidetes* classis nov., within the phylum *Chloroflexi*. *Int J Syst Evol Microbiol*. 2012. doi:10.1099/ijs.0.034926-0.

Louw GE, Warren RM, Gey van Pittius NC, McEvoy CRE, Van Helden PD, Victor TC. A Balancing Act: Efflux/Influx in Mycobacterial Drug Resistance . *Antimicrobial Agents and Chemotherapy*. 2009;53(8):3181-3189. doi:10.1128/AAC.01577-08.

Ma L, Deng F, Yang C, Guo C, Dang Z. Bioremediation of PAH-contaminated farmland: field experiment. *Environ Sci Pollut Res Int*. 2016 Nov 12.

Martínez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. *Nat Rev Microbiol*. 2015 Feb;13(2):116-23. doi: 10.1038/nrmicro3399.

Martínez JL, Coque TM, Lanza VF, de la Cruz F, Baquero F. Genomic and metagenomic technologies to explore the antibiotic resistance mobilome. *Ann N Y Acad Sci*. 2016 Nov 10. doi: 10.1111/nyas.13282.

Maurya AK, Nag VL, Kant S, Kushwaha RS, Dhole TN. Genotypic analysis of multidrug-resistant tuberculosis isolates from extra pulmonary tuberculosis cases in

tertiary care centers in Northern India. *Int J Mycobacteriol.* 2016 Dec;5 Suppl 1:S125-S126. doi: 10.1016/j.ijmyco.2016.09.027.

McArthur AG, Waglechner N, Nizam F, et al. The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy.* 2013;57(7):3348-3357. doi:10.1128/AAC.00419-13.

Mikalsen J, Olsen AB, Tengs T, Colquhoun DJ. *Francisella philomiragia* subsp. *noatunensis* subsp. nov., isolated from farmed Atlantic cod (*Gadus morhua* L.). *Int J Syst Evol Microbiol.* 2007; 57: 1960–1965.

Miller JR, Koren S, Sutton G. Assembly Algorithms for Next-Generation Sequencing Data. *Genomics.* 2010;95(6):315-327. doi:10.1016/j.ygeno.2010.03.001.

Moran RA, Anantham S, Holt KE, Hall RM. Prediction of antibiotic resistance from antibiotic resistance genes detected in antibiotic-resistant commensal *Escherichia coli* using PCR or WGS. *J Antimicrob Chemother.* 2016 Dec 30. pii: dkw511. doi: 10.1093/jac/dkw511.

Muangchinda C, Yamazoe A, Polrit D, Thoetkiattikul H, Mhuantong W, Champreda V, et al. Biodegradation of high concentrations of mixed polycyclic aromatic hydrocarbons by indigenous bacteria from a river sediment: a microcosm study and bacterial community analysis. *Environ Sci Pollut Res Int.* 2016 Dec 12.

Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. *Microbiology spectrum.* 2016; 4(2):10.1128/microbiolspec. VMBF-0016-2015. doi:10.1128/microbiolspec.VMBF-0016-2015.

Nessar R, Cambau E, Reyrat JM, Murray A, Gicquel B. *Mycobacterium abscessus*: a new antibiotic nightmare. *J Antimicrob Chemother.* 2012 Apr;67(4):810-8. doi: 10.1093/jac/dkr578.

Ngeow YF, Wong YL, Tan JL, et al. Identification of New Genomespecies in the *Mycobacterium terrae* Complex. Hoshino Y, ed. *PLoS ONE.* 2015;10(4):e0120789. doi:10.1371/journal.pone.0120789.

Niva M, Hernesmaa A, Haahtela K, Salkinoja-salonen M, Haukka K. Actinobacterial Communities of Boreal Forest Soil and Lake Water are Rich in Mycobacteria. *Boreal Environment Research*. 2006; 11, 45-53.

Noguera-Ortega E, Secanella-Fandos S, Eraña H, Gasión J, Rabanal RM, Luquin M, et al. Nonpathogenic *Mycobacterium brumae* Inhibits Bladder Cancer Growth In Vitro, Ex Vivo, and In Vivo. *European Urology Focus*. 2016; 2(1):67-76

Nyka W. Studies on the Effect of Starvation on Mycobacteria. *Infection and Immunity*. 1974;9(5):843-850.

Parvandar K, Mayahi M, Mosavari N, Pajoochi RA. Drug susceptibility testing of *Mycobacterium Avium* subsp. *Avium* isolates from naturally infected domestic pigeons to avian tuberculosis. *Int J Mycobacteriol*. 2016 Dec;5 Suppl 1:S217-S218.

Patel RK and Jain M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS One*. 2012; 7(2): e30619.

Paul Pottinger, L. Barth Reller, Kenneth J. Ryan. *Mycobacteria*. In: Kenneth J. Ryan, C. George Ray. *Sherris Medical Microbiology*. 6th edition. McGraw-Hill Medical; 2014. p. 439-440.

Pawlowski AC, Wang W, Koteva K, Barton HA, McArthur AG, Wright GD. A diverse intrinsic antibiotic resistome from a cave bacterium. *Nature Communications*. 2016;7:13803. doi:10.1038/ncomms13803.

Pérez-Losada M, Cabezas P, Castro-Nallar E, Crandall KA. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. *Infect Genet Evol*. 2013 Jun;16:38-53. doi: 10.1016/j.meegid.2013.01.009.

Pesciaroli M, Alvarez J, Boniotti MB, Cagiola M, Di Marco V, Marianelli C, et al. Tuberculosis in domestic animal species. *Res Vet Sci*. 2014 Oct;97 Suppl:S78-85. doi: 10.1016/j.rvsc.2014.05.015.

Petersen FC, Scheie AA. Genetic transformation in *Streptococcus mutans* requires a peptide secretion-like apparatus. *Oral Microbiol Immunol*. 2000 Oct;15(5):329-34.

Prakash O, Gihring TM, Dalton DD, Chin KJ, Green SJ, Akob DM, et al. *Geobacter daltonii* sp. nov., an Fe(III)- and uranium(VI)-reducing bacterium isolated from a shallow subsurface exposed to mixed heavy metal and hydrocarbon contamination. *Int J Syst Evol Microbiol*. 2010 Mar;60(Pt 3):546-53. doi: 10.1099/ijs.0.010843-0.

Primm TP, Lucero CA, Falkinham JO. Health Impacts of Environmental Mycobacteria. *Clinical Microbiology Reviews*. 2004;17(1):98-106. doi:10.1128/CMR.17.1.98-106.2004.

Rabello MC da S, Matsumoto CK, de Almeida LGP, et al. First Description of Natural and Experimental Conjugation between Mycobacteria Mediated by a Linear Plasmid. Johnson EA, ed. *PLoS ONE*. 2012;7(1):e29884. doi:10.1371/journal.pone.0029884.

Rastogi N, Frehel C, Rytter A, Ohayon H, Lesourd M, David HL. Multiple drug resistance in *Mycobacterium avium*: is the cell wall architecture responsible for the exclusion of antimicrobial agents? *Antimicrob Agents Chemother*. 1981; 20: 666–677.

Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, et al. (2009) Non Mycobacterial Virulence Genes in the Genome of the Emerging Pathogen *Mycobacterium abscessus*. *PLoS ONE* 4(6): e5660. doi:10.1371/journal.pone.0005660

Ritalahti KM, Justicia-Leon SD, Cusick KD et al. *Sphaerochaeta globosa* gen. nov., sp. nov. and *Sphaerochaeta pleomorpha* sp. nov., free-living, spherical spirochaetes. *Int J Syst Evol Microbiol*. 2012; 62: 210–216.

Rohwer F, Edwards R: The phage proteomic tree: a genome-based taxonomy for phage. *J Bacteriol* 2002, 184:4529–4535.

Rosselló-Mora R, Amann R. The species concept for prokaryotes. *FEMS Microbiol Rev.* 2001 Jan;25(1):39-67.

Saffo Z, Ognjan A. Mycobacterium smegmatis infection of a prosthetic total knee arthroplasty. *IDCases.* 2016; 5:80-82. doi:10.1016/j.idcr.2016.07.007.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014; 30(14):2068-9.

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. doi:10.3389/fmicb.2015.00242.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. *Genome Research.* 2009;19(6):1117-1123. doi:10.1101/gr.089532.108.

Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, de la Cruz F. Mobility of Plasmids. *Microbiol Mol Biol Rev.* 2010; 74(3): 434-52.

Snipen L, Ussery DW. Standard operating procedure for computing pangenome trees. *Standards in Genomic Sciences.* 2010;2(1):135-141. doi:10.4056/sigs.38923.

Soroka D, Dubée V, Soulier-Escrihuela O, Cuinet G, Hugonnet JE, Gutmann L, et al. Characterization of broad-spectrum Mycobacterium abscessus class A β -lactamase. *J Antimicrob Chemother.* 2014 Mar;69(3):691-6. doi: 10.1093/jac/dkt410.

Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; 33: 152–155.

Stahl DA, Urbance JW. The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *Journal of Bacteriology.* 1990;172(1):116-124.

Steingrube VA, Gibson JL, Brown BA, et al. PCR amplification and restriction endonuclease analysis of a 65-kilodalton heat shock protein gene sequence for taxonomic separation of rapidly growing mycobacteria. *Journal of Clinical Microbiology*. 1995;33(1):149-153.

Stinear TP, Seemann T, Harrison PF, et al. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Research*. 2008;18(5):729-741. doi:10.1101/gr.075069.107.

Tao J, Han J, Wu H, et al. *Mycobacterium* fluoroquinolone resistance protein B, a novel small GTPase, is involved in the regulation of DNA gyrase and drug resistance. *Nucleic Acids Research*. 2013;41(4):2370-2381. doi:10.1093/nar/gks1351.

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 of Clinical Microbiology*. 1993;31(2):175-178.

Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL. Microbial genomic taxonomy. *BMC Genomics*. 2013;14:913. doi:10.1186/1471-2164-14-913.

Thompson CC, Silva GG, Vieira NM, Edwards R, Vicente AC, Thompson FL. Genomic taxonomy of the genus *prochlorococcus*. *Microb Ecol*. 2013 Nov;66(4):752-62. doi: 10.1007/s00248-013-0270-8.

Thompson CC, Vicente ACP, Souza RC, et al. Genomic taxonomy of vibrios. *BMC Evolutionary Biology*. 2009;9:258. doi:10.1186/1471-2148-9-258.

Tindall BJ, Rosselló- Móra R, Busse HJ, Ludwig W, Kampfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; 60: 249–266.

Tritt A, Eisen JA, Facciotti MT, Darling AE. An Integrated Pipeline for de Novo Assembly of Microbial Genomes. Zhu D, ed. PLoS ONE. 2012;7(9):e42304. doi:10.1371/journal.pone.0042304.

Uchiya K, Takahashi H, Nakagawa T, Yagi T, Moriyama M, Inagaki T, et al. Characterization of a novel plasmid, pMAH135, from *Mycobacterium avium* subsp. *hominissuis*. PLoS ONE. 2015; 10(2): e0117797

Ummels R, Abdallah AM, Kuiper V, et al. Identification of a Novel Conjugative Plasmid in *Mycobacteria* That Requires Both Type IV and Type VII Secretion. mBio. 2014;5(5):e01744-14. doi:10.1128/mBio.01744-14.

Umrao J, Singh D, Zia A, Saxena S, Sarsaiya S, Singh S, et al. Prevalence and species spectrum of both pulmonary and extrapulmonary nontuberculous mycobacteria isolates at a tertiary care center. Int J Mycobacteriol. 2016; 5(3):288-293. doi: 10.1016/j.ijmyco.2016.06.008

Vandamme P., Pot B., Gillis M., de Vos P., Kersters K., Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol. Rev. 1996;60:407–438.

Vanlaere E, Baldwin A, Gevers D, Henry D, De Brandt E, LiPuma JJ, et al. Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. Int J Syst Evol Microbiol. 2009 Jan;59(Pt 1):102-11. doi: 10.1099/ijs.0.001123-0.

Varghese NJ, Mukherjee S, Ivanova N, et al. Microbial species delineation using whole genome sequences. Nucleic Acids Research. 2015;43(14):6761-6771. doi:10.1093/nar/gkv657.

Velu PP, Fernandes SE, Laurenson IF, Noble DD. Pulmonary *Mycobacterium marinum* infection: 'fish tank granuloma' of the lung. Scott Med J. 2016 Nov;61(4):203-206.

Ventura M, Canchaya C, Tauch A, et al. Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum . *Microbiology and Molecular Biology Reviews: MMBR*. 2007;71(3):495-548. doi:10.1128/MMBR.00005-07.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One*. 2014; 9(11): e112963.

Wallace JC, Port JA, Smith MN, Faustman EM. FARME DB: a functional antibiotic resistance element database. *Database (Oxford)*. 2017 Jan 10;2017. pii: baw165. doi: 10.1093/database/baw165.

Wang J, McIntosh F, Radomski N, Dewar K, Simeone R, Enninga J, et al. Insights on the emergence of *Mycobacterium tuberculosis* from the analysis of *Mycobacterium kansasii*. *Genome Biol Evol*. 2015 Feb 25;7(3):856-70. doi: 10.1093/gbe/evv035.

Wright. The 'effective number of codons' used in a gene. *Gene*. 1990, 1;87(1):23-9.

Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 2008;18(5):821-829. doi:10.1101/gr.074492.107.

Zhou ZF, Wang MX, Zuo XH, Yao YH. Comparative Investigation of Bacterial, Fungal, and Archaeal Community Structures in Soils in a Typical Oilfield in Jiangnan, China. *Arch Environ Contam Toxicol*. 2017; 72(1):65-77. doi: 10.1007/s00244-016-0333-1.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. *Bioinformatics*. 2013;29(21):2669-2677. doi:10.1093/bioinformatics/btt476.