

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

HELENA REGINA SALOMÉ D'ESPINDULA

ESTUDO DA RESISTÊNCIA ANTIMICROBIANA DE CEPAS DE *Mycobacterium leprae* ADVINDAS DE UMA POPULAÇÃO HIPERENDÊMICA ISOLADA DO NORTE DO BRASIL

CURITIBA

2018

HELENA REGINA SALOMÉ D'ESPINDULA

ESTUDO DA RESISTÊNCIA ANTIMICROBIANA DE CEPAS DE *Mycobacterium leprae* ADVINDAS DE UMA POPULAÇÃO HIPERENDÊMICA ISOLADA DO NORTE DO BRASIL

Tese apresentada ao Programa de Pós-Graduação em Ciências da Saúde, Escola de Medicina, Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de Doutora em Ciências da Saúde.

Orientador: Prof. Dr. Marcelo Távora Mira

Co-orientadora: Dr. Patricia Sammarco Rosa

CURITIBA

2018

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Luci Eduarda Wielganczuk – CRB 9/1118

D468e
2019

D’Espindula, Helena Regina Salomé
Estudo da resistência antimicrobiana de cepas de *Mycobacterium leprae* advindas de uma população hiperendêmica isolada do norte do Brasil / Helena Regina Salomé D’Espindula ; orientador: Marcelo Távora Mira ; co-orientadora: Patricia Sammarco Rosa. – 2019.
188 f. : il. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2019
Bibliografia: 112-123

1. Hanseníase. 2. Sequenciamento completo do genoma. 3. Resistência microbiana a medicamentos. I. Mira, Marcelo Távora. II. Rosa, Patricia Sammarco. III. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Ciências da Saúde. IV. Título.

CDD 20. ed. – 616.998



PUCPR

Pontifícia Universidade Católica do Paraná

Escola de Medicina

Programa de Pós-Graduação em Ciências da Saúde

ATA DA SESSÃO PÚBLICA DE EXAME DE TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE EM NÍVEL DE DOUTORADO DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ.

Aos vinte e um dias do mês de fevereiro de 2018 às 13hs e 30min., realizou-se a sessão pública de Defesa de Tese “ESTUDO DA RESISTENCIA ANTIMICROBIANA DE CEPAS DE *Mycobacterium leprae* ADVINDAS DE UMA POPULAÇÃO HIPERENDÊMICA ISOLADA DO NORTE DO BRASIL” apresentado por **Helena Regina Salomé D’Espindula** para obtenção do título de Doutora; Área de concentração: Medicina e áreas afins.

A banca examinadora foi composta pelos seguintes membros:

MEMBROS DA BANCA	ASSINATURA
Prof. Dr. Marcelo Távora Mira – Presidente (PUCPR)	
Profa. Dra. Patrícia Sammarco Rosa – (ILSL-SP)	
Prof. Dr. Flávio Alves Lara (FIOCRUZ-RJ)	
Prof. Dr. Marcelo Pilonetto (PUCPR)	
Prof. Dr. Javier Felipe Teodoro B. Senoret– (PUCPR)	

De acordo com as normas regimentais a Banca Examinadora deliberou sobre os conceitos a serem distribuídos e que foram os seguintes:

Prof. Dr. Marcelo Mira	Conceito <u>APROVADA</u>
Profa. Dra. Patrícia Sammarco Rosa	Conceito <u>APROVADA</u>
Prof. Dr. Flávio Alves Lara	<u>APROVADA</u>
Prof. Dr. Marcelo Pilonetto	Conceito <u>APROVADA</u>
Prof. Da. Javier Felipe Teodoro B. Senoret	Conceito <u>APROVADA</u>
	Parecer Final: <u>APROVADA</u>

Observações da Banca Examinadora:

Prof. Dr. Marcelo Távora Mira
Presidente da Banca Examinadora

Profa. Dra. Cristina Pellegrino Baena
Coordenadora do PPGCS-PUCPR

AGRADECIMENTOS

Aos sujeitos de pesquisa, que generosamente nos cederam um pouco do seu tempo, um pouco de sua história e um pequeno pedaço de si em prol da ciência;

Aos familiares e amigos pela paciência e apoio, mesmo nas horas mais difíceis;

Ao meu orientador pela contínua orientação e atenção dispensadas a este trabalho;

Aos colegas de bancada, pela convivência diária e pelas trocas de experiências e conhecimento;

Aos grandes e pequenos cientistas que desbravaram os caminhos que hoje trilho;

Aos colaboradores, cuja parceria expandiu as possibilidades desse projeto;

Às bancas avaliadoras, que se dispuseram a usar seu tempo e conhecimento em prol da minha evolução e deste trabalho;

Aos financiadores deste estudo (Ministério da Ciência, Tecnologia e Inovação / Conselho Nacional de Desenvolvimento Científico e Tecnológico / Secretaria de Ciência, Tecnologia e Insumos Estratégicos do Ministério da Saúde; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES; e Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná) que acreditaram nesta proposta e a viabilizaram.

Genius is one percent inspiration, ninety-nine percent perspiration

Thomas Alva Edison

Ando devagar porque já tive pressa
E levo esse sorriso porque já chorei demais
Hoje me sinto mais forte, mais feliz, quem sabe
Só levo a certeza de que muito pouco sei ou nada sei

Almir Sater e Renato Teixeira

RESUMO

A hanseníase é uma doença infecciosa causada pelo *Mycobacterium leprae* (*M. leprae*), um bacilo intracelular obrigatório de crescimento lento. O tratamento da doença é feito com a combinação de três antibióticos: dapsona, rifampicina e clofazemina (Poliquimioterapia - PQT) e terapias alternativas são escassas. Em iniciativa anterior iniciada em 2009 com participação de nosso grupo de pesquisa, foi constatado um índice alarmante de resistência à PQT na Vila Santo Antônio do Prata (ou Vila do Prata), uma ex-colônia de hansenianos, que ainda hoje é hiperendêmica para essa doença. Dessa forma, fez-se necessário descrever a exata extensão do problema de resistência primária e secundária do *M. leprae* aos componentes da PQT na Vila do Prata, bem como a natureza molecular dos isolados bacterianos. Essa nova iniciativa utilizou uma estratégia sistemática de recrutamento da população, seguida de confirmação laboratorial dos casos, cultivo *in vivo* do *M. leprae*, análise molecular dos genes clássicos de resistência, genotipagem de *SNPs* e variações em tandem dos isolados e, por fim, padronização do sequenciamento de genoma completo. Um total de 611 indivíduos foram examinados em atividade intensiva de atendimento dermatológico na Vila do Prata. Foram obtidas biópsias de 20 casos clinicamente identificados como hanseníase ativa. Constatou-se que, dentre 11 casos de recidiva, 5 (45,5%) tinham alguma resistência e, dentre os 9 casos novos, 2 (22,2%) tinham alguma resistência. Somando esses casos aos identificados no Estudo de 2009, obteve-se um total de 18 recidivas (das quais 13 tinham resistência a um ou mais antibióticos) e 19 novos casos (dos quais 6 tinham resistência a um ou mais antibióticos). A análise combinada dos dados permitiu a demonstração de transmissão intrafamiliar de casos de resistência secundária gerando casos de resistência primária. Finalmente, foi padronizado o protocolo de sequenciamento de genoma completo e foi definida uma proposta preliminar de fluxo da análise de dados do sequenciamento.

Palavras-chave: hanseníase, genoma, resistência, sequenciamento de segunda geração.

ABSTRACT

Leprosy is an infectious disease caused by *Mycobacterium leprae* (*M. leprae*), a mandatory slow-growing intracellular bacillus. The treatment of the disease is based on a combination of three antibiotics: dapsone, rifampicin and clofazimine (Multidrugtherapy - MDT) and alternative therapies are scarce. In an earlier initiative started in 2009 with the participation of our research group, an alarming index of resistance to MDT was found in Vila Santo Antônio do Prata, a former leprosy colony that is still hyperendemic to leprosy today. Thus, it was necessary to describe an exact extension of the primary and secondary resistance problem of *M. leprae* to the components of MDT in Vila do Prata, as well as the molecular nature of bacterial isolates. In this new initiative a systematic population recruitment strategy was used, followed by laboratory confirmation of the cases, *in vivo* *M. leprae* culture, molecular analysis of classical resistance genes, *SNP* typing and tandem variations and, finally, standardization of *M. leprae* genome sequencing. A total of 611 individuals were examined in intensive dermatological care in Vila do Prata. Biopsies were obtained from 20 identified cases of active leprosy, 11 relapses and 9 new cases; of these, five and two, respectively, had resistance to one or more antibiotics. Adding these data to the 2009-study, we obtained a total of 18 relapses (of which 13 had resistance to one or more antibiotics) and 19 new cases (of which 6 had resistance to one or more antibiotics) were observed. The combined analysis of the data led to the demonstration of intra-familial transmission of cases of secondary resistance in cases of primary resistance. Finally, the complete genome sequencing protocol was standardized and a preliminary proposal for flow of data analysis was defined.

Keywords: leprosy, genome, resistance, next generation sequencing.

LISTA DE ILUSTRAÇÕES

Figura 1 – Patogêneses da hanseníase.....	19
Figura 2 – Mecanismo de ação da DDS.....	22
Figura 3 – Mecanismo de ação da RFP.....	24
Figura 4 – Mecanismo de ação da OFX.....	26
Figura 5 – Representação esquemática das regiões determinantes de resistência a antibióticos (<i>drug resistance determining region - DRDR</i>) mais comumente relatadas no gene <i>folP</i> do <i>M. leprae</i>	28
Figura 6 – Representação esquemática das regiões determinantes de resistência a antibióticos (<i>drug resistance determining region - DRDR</i>) mais comumente relatadas no gene <i>rpoB</i> do <i>M. leprae</i>	28
Figura 7 – Representação esquemática das regiões determinantes de resistência a antibióticos (<i>drug resistance determining region - DRDR</i>) mais comumente relatadas no gene <i>gyrA</i> do <i>M. leprae</i>	28
Figura 8 - Prevalência estimada de hanseníase no mundo de 1966 a 1983...	35
Figura 9 - Prevalência e incidência de hanseníase no mundo de 1985 a 2000	36
Figura 10 - Prevalência e incidência de hanseníase no mundo de 2000 a 2016	37
Figura 11 – Variabilidade do marcador GAA 21 em estudos de <i>VNTRs</i> de <i>M. leprae</i>	43
Figura 12 – Variabilidade do marcador GTA 9 em estudos de <i>VNTRs</i> de <i>M. leprae</i>	44
Figura 13 – Variabilidade dos marcadores 6-3a e 21-3 em estudos de <i>VNTRs</i> de <i>M. leprae</i>	45
Figura 14 - Representação esquemática das etapas de fragmentação e ligação de adaptadores e <i>barcodes</i>	53
Figura 15 - Representação esquemática das etapas de <i>size-select</i> e amplificação da biblioteca	54
Figura 16 - Representação esquemática de preparo de <i>template</i>	55
Figura 17 - Representação esquemática do processo de sequenciamento em poço do chip.....	56
Figura 18 - Relação da escala <i>Phred</i> com erro e acurácia.....	57
Figura 19 - Representação esquemática do alinhamento das <i>reads</i> a uma sequência de referência	59
Figura 20 - Esquema do desenho de estudo do projeto global no qual este estudo está inserido.	63
Figura 21 - Localização da Vila Santo Antônio do Prata	74
Figura 22 – Representação esquemática das etapas de análise de dados	81
Figura 23 - Distribuição da idade dos 611 indivíduos examinados.....	83
Figura 24 - Carregamento dos chips contendo as amostras	85
Figura 25 - Sumário das <i>reads</i> utilizáveis no sequenciamento.....	86

Figura 26 - Qualidade das bases das amostras de acordo com a sua posição na <i>read</i>	87
Figura 27 – Distribuição de comprimento e qualidade de todas as <i>reads</i> da amostra P-84 (purificada com NaOH)	91
Figura 28 – Distribuição de comprimento e qualidade de todas as <i>reads</i> da amostra P-220 (purificada com NaOH)	92
Figura 29 – Distribuição de comprimento e qualidade de todas as <i>reads</i> da amostra ROS-MAD (purificada com NaOH e colagenase).....	93
Figura 30 – Distribuição de comprimento e qualidade de todas as <i>reads</i> da amostra LC (purificada com NaOH e colagenase)	94
Figura 31 – Distribuição de comprimento e qualidade de todas as <i>reads</i> da amostra P-84 (purificada com NaOH e colagenase)	95
Figura 32 – Cobertura ao longo do genoma no mapeamento da amostra ROS-MAD no <i>software</i> MIRA.....	99
Figura 33 – Cobertura ao longo do genoma no mapeamento da amostra LC no <i>software</i> MIRA.....	100
Figura 34 – Cobertura ao longo do genoma no mapeamento da amostra P-84 no <i>software</i> MIRA.....	101
Figura 35 – Comparação dos genomas obtidos na montagem <i>de novo</i> entre si e com o genoma de referência.....	102
Figura 36 – Comparação dos genomas obtidos no mapeamento (com o <i>software</i> MIRA) entre si e com o genoma de referência.....	103
Figura 37 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra ROS-MAD no <i>software</i> MIRA	105
Figura 38 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra ROS-MAD no <i>software</i> TMAP	106
Figura 39 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra LC no <i>software</i> MIRA.....	107
Figura 40 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra LC no <i>software</i> TMAP.....	108
Figura 41 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra P-84 no <i>software</i> MIRA.....	109
Figura 42 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra P-84 no <i>software</i> TMAP.....	110
Figura 43 – Representação esquemática das variantes encontradas nas amostras analisadas	111

LISTA DE TABELAS

Tabela 1 - Regime recomendado da Poliquimioterapia a partir de 1998.	26
Tabela 2 – Dados dos relatórios de monitoramento da resistência da OMS. ...	32
Tabela 3 – Marcadores do tipo <i>VNTR</i> já utilizados nos estudos de <i>M. leprae</i>	39
Tabela 4 – Marcadores utilizados em cada estudo de <i>VNTRs</i> de <i>M. leprae</i>	41
Tabela 5 – Tipagem por <i>SNP</i> de <i>M. leprae</i> em 4 tipos.	46
Tabela 6 – Tipagem por <i>SNP</i> de <i>M. leprae</i> em 16 subtipos.....	46
Tabela 7 - Comparação do genoma de algumas micobactérias.....	48
Tabela 8 - Escore <i>Phred</i> relacionado à probabilidade de erro.....	57
Tabela 9 – Reagentes utilizados para o sequenciamento	78
Tabela 10 – Programas utilizados na análise de dados.....	80
Tabela 11 – Distribuição categórica de doenças dermatológicas nos indivíduos com alguma doença dermatológica.....	84
Tabela 12 – Resultados do alinhamento com <i>M. musculus</i>	96
Tabela 13 – Resultados da montagem <i>de novo</i> no software <i>MIRA</i>	97
Tabela 14 – Resultados do mapeamento no software <i>MIRA</i>	98
Tabela 15 – Resultados comparativos do mapeamento com <i>M. leprae</i>	104

LISTA DE QUADROS

Quadro 1 – Classificação da hanseníase para fins terapêuticos de acordo com OMS	20
Quadro 2 – Escala Logarítmica de Ridley para baciloscopia de hanseníase. .	21
Quadro 3 – Tratamento recomendando pelo MS para doente com intolerância a um ou mais componentes da PQT.	29
Quadro 4 – Definições da OMS.....	31
Quadro 5 – Exemplo de linha de comando	58
Quadro 6 – <i>String</i> de busca.....	125
Quadro 7 - Comando de indexação da referência <i>M. musculus</i> no <i>software TMAP</i>	128
Quadro 8 - Comando de mapeamento das amostras utilizando a referência <i>M. musculus</i> no <i>software TMAP</i>	129
Quadro 9 - Comando de ordenação das amostras mapeadas no <i>software Samtools</i>	129
Quadro 10 - Comando de avaliação de qualidade de mapeamento no <i>software Qualimap</i>	130
Quadro 11 - Comando de separação das <i>reads</i> mapeadas no <i>software Samtools</i>	130
Quadro 12 - Comando de separação das <i>reads</i> mapeadas no <i>software Samtools</i>	131
Quadro 13 – Manifesto utilizado para a montagem <i>de novo</i> no <i>software MIRA</i>	131
Quadro 14 – Comando de montagem <i>de novo</i> no <i>software MIRA</i>	132
Quadro 15 – Comando de conversão do formato de arquivo alinhado no <i>software MIRA</i>	132
Quadro 16 – Comando de avaliação de qualidade do alinhamento no <i>software Qualimap</i>	132
Quadro 17 – Manifesto utilizado para o mapeamento com referência de <i>M. leprae</i> no <i>software MIRA</i>	133
Quadro 18 – Comando de mapeamento com referência de <i>M. leprae</i> no <i>software MIRA</i>	134
Quadro 19 – Comando de conversão do formato de arquivo mapeado no <i>software MIRA</i>	134
Quadro 20 – Comando de avaliação de qualidade do mapeamento no <i>software Qualimap</i>	134
Quadro 21 - Comando de indexação da referência <i>M. leprae</i> no <i>software TMAP</i>	135
Quadro 22 - Comando de mapeamento das amostras utilizando a referência <i>M. leprae</i> no <i>software TMAP</i>	136
Quadro 23 - Comando de ordenação das amostras mapeadas no <i>software Samtools</i>	136

Quadro 24 - Comando de avaliação de qualidade de mapeamento no <i>software Qualimap</i>	136
Quadro 25 - Comando de separação das <i>reads</i> mapeadas no <i>software Samtools</i>	137
Quadro 26 – Comando de ordenação das <i>reads</i> mapeadas no <i>software Picard</i>	137
Quadro 27 – Comando de marcação de duplicatas no <i>software Picard</i>	137
Quadro 28 – Comando de indexação no <i>software Picard</i>	137
Quadro 29 – Comando para obtenção de lista de variantes no <i>software TVC</i> 138	
Quadro 30 – Arquivo de parâmetros utilizado para obtenção de variantes no <i>software TVC</i>	139

LISTA DE ABREVIATURAS, SÍMBOLOS E SIGLAS

A260/230	<i>Ratio of the absorbance at 260 and 230nm</i> (Razão de absorvancia 260/230 nm)
A260/280	<i>Ratio of the absorbance at 260 and 280nm</i> (Razão de absorvancia 260/230 nm)
°C	Grau Celsius
µg	Micro-grama (= 10 ⁻⁶ grama)
µL	Micro-litro (= 10 ⁻⁶ litro)
Ala	Alanina
AQ	<i>Alignment Quality</i> (Qualidade do Alinhamento)
Arg	Arginina
Asn	Asparagina
Asp	Ácido Aspártico
Avg.	<i>Average</i> (média)
BLAST	<i>Basic Local Alignment Search Tool</i> (Ferramenta de Pesquisa de Alinhamento Local Básico)
CEP	Comitê de Ética em Pesquisa
CLF	Clofazimina
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
col.	Colaboradores
Conc.	Concentração
Cys	Cisteína
DDS	Dapsona
DECIT	Departamento de Ciência e Tecnologia
DHPS	Didropteroato sintetase
DNA	<i>Deoxyribonucleic acid</i> (ácido desoxirribonucleico)
EDTA	<i>Ethylenediamine tetraacetic acid</i> (ácido etilendiamino tetra-acético)
FastQC	<i>Quality control tool for high throughput sequence data</i>
FIOCRUZ	Fundação Oswaldo Cruz
folP	<i>Dihydropteroate synthase 1</i>
folP2	<i>Dihydropteroate synthase 2</i>
GI	<i>GenInfo Identifier</i>
Gln	Glutamina
Glu	Ácido Glutâmico
Gly	Glicina
gryA	<i>DNA gyrase subunit A</i> (subunidade A da DNA girase)
His	Histidina
IB	Índice baciloscópio
Ile	Isoleucina
ILSL	Instituto Lauro de Souza Lima
Ion PGM	<i>Ion Torrent Personal Genome Machine</i>
ISP	<i>Ion Sphere Particles</i>
Leu	Leucina
Lys	Lisina
<i>M. africanum</i>	<i>Mycobacterium africanum</i>
<i>M. bovis</i>	<i>Mycobacterium bovis</i>
<i>M. leprae</i>	<i>Mycobacterium leprae</i>
<i>M. lepromatosis</i>	<i>Mycobacterium lepromatosis</i> (agente etiológico da hanseníase lepromatosa difusa)
<i>M. musculus</i>	<i>Mus musculus</i> (camundongo)
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i> (agente etiológico da Tuberculose)
Mauve	<i>Multiple Alignment of Conserved Genomic Sequence with Rearrangements</i>

MB	Multibacilar
MCTI	Ministério da Ciência, Tecnologia e Inovação
MDT	<i>Multidrug therapy</i>
Met	Metionina
MIRA	Mimicking Intelligent Read Assembly
mL	Mili-litro (= 10 ⁻³ litro)
MS-SCTIE	Ministério da Saúde - Secretaria de Ciência, Tecnologia e Insumos Estratégicos
N	Normal
NaOH	Hidróxido de sódio
NCBI	<i>National Center for Biotechnology Information</i>
NGS	<i>Next-Generation Sequencing</i>
nm	Nano-metros (= 10 ⁻⁹ litro)
OFX	Ofloxacina
OMS	Organização Mundial da Saúde
PABA	Ácido para-aminobenzóico
pb	Pares de base
PB	Paucibacilar
PCR	<i>Polymerase Chain Reaction</i> (Reação em cadeia da polimerase)
PCR-RT	Real Time - <i>Polymerase Chain Reaction</i> (Reação em cadeia da polimerase em tempo real)
pg	Pico-grama (= 10 ⁻¹² grama)
Phe	Fenilalanina
PQT	Poliquimioterapia
Pro	Prolina
PUCPR	Pontifícia Universidade Católica do Paraná
RAST	Rapid Annotation using Subsystem Technology
ResFinder	<i>Resistance Finder</i>
RFLP	<i>Restriction Fragment Length Polymorphisms</i>
RFP	Rifampicina
RJ	Rio de Janeiro
RNA	<i>Ribonucleic acid</i> (ácido ribonucleico)
RPMI	<i>Roswell Park Memorial Institute medium</i>
rpoA	<i>RNA polymerase, alpha subunit</i> (subunidade alfa da RNA polimerase)
rpoB	<i>RNA polymerase, beta subunit</i> (subunidade beta da RNA polimerase)
rpoC	<i>DNA-directed RNA polymerase subunit beta</i> (subunidade beta da RNA polimerase DNA dirigida)
rpoT	<i>RNA polymerase sigma factor</i> (fator sigma da RNA polimerase)
Ser	Serina
SNP	<i>Single Nucleotide Polymorphism</i> (Polimorfismo de nucleotídeo único)
SP	São Paulo
TCLE	Termo de Consentimento Livre e Esclarecido
Thr	Treonina
TMAP	<i>Torrent Mapping Alignment Program</i>
Trp	Triptofano
TVC	<i>Torrent Variant Caller</i>
Tyr	Tirosina
UFPA	Universidade Federal do Pará
Val	Valina
VNTR	<i>Variable number tandem repeat</i> (Repetição em tandem de número variável)
WHO	<i>World Health Organization</i> (Organização Mundial da Saúde)

SUMARIO

1.	INTRODUÇÃO	18
1.1.	REVISÃO DE LITERATURA – PARTE 1 – HANSENÍASE	18
1.1.1.	Agente Etiológico	18
1.1.2.	Patogêneses	18
1.1.3.	Diagnóstico	20
1.1.4.	Tratamento e Resistência	21
1.1.4.1.	Tratamento com dapsona	21
1.1.4.2.	Tratamento com clofazimina	23
1.1.4.3.	Tratamento com rifampicina e resistências primárias	23
1.1.4.4.	Surgimento da Poliquimioterapia	25
1.1.4.5.	Tratamento com ofloxacina	26
1.1.4.6.	Multirresistência	27
1.1.4.7.	Tratamentos substitutivos	28
1.1.4.8.	Monitoramento da resistência	29
1.1.5.	Epidemiologia	34
1.1.6.	Variabilidade do <i>M. leprae</i>	38
1.1.6.1.	VNTRs	38
1.1.6.2.	Tipagem por <i>SNP</i>	45
1.1.6.3.	Estudos de genoma	48
1.2.	REVISÃO DE LITERATURA – PARTE 2 – SEQUENCIAMENTO DE GENOMA COMPLETO	51
1.2.1.	Sequenciamento de Segunda Geração	51
1.2.1.1.	Condições pré-sequenciamento	52
1.2.1.2.	Preparo de Biblioteca	52
1.2.1.3.	Preparo de <i>template</i>	55
1.2.1.4.	Sequenciamento	56
1.2.1.5.	Dados Brutos	57
1.2.1.6.	Análise de Dados	58
1.2.1.6.1.	Tipos de <i>software</i>	58
1.2.1.6.2.	Alinhamento	58

1.2.1.6.2.1.	Mapeamento	59
1.2.1.6.2.2.	Montagem <i>de novo</i>	60
2.	CONTEXTO E DESENHO DO ESTUDO	61
3.	JUSTIFICATIVA.....	64
4.	OBJETIVOS.....	65
4.1.	OBJETIVO GERAL.....	65
4.2.	OBJETIVOS ESPECÍFICOS	65
5.	MANUSCRITO	66
6.	MÉTODOS COMPLEMENTARES	74
6.1.	SUPORTE FINANCERO	74
6.2.	POPULAÇÃO DA VILA SANTO ANTONIO DO PRATA.....	74
6.2.1.	Recrutamento da amostra populacional.....	75
6.3.	CONSIDERAÇÕES ÉTICAS	76
6.4.	TESTE DE RESISTÊNCIA <i>in vivo</i>	76
6.5.	PADRONIZAÇÃO DO PROTOCOLO DE SEQUENCIAMENTO DE GENOMA COMPLETO DE <i>M. LEPRAE</i>	76
6.5.1.	Purificações do bacilo e extração de DNA.....	76
6.5.1.1.	Purificação com NaOH.....	77
6.5.1.2.	Purificação com colagenase-dispase e NaOH.....	77
6.5.2.	Etapas do sequenciamento.....	77
6.5.2.1.	Purificação com <i>beads</i>	78
6.5.2.2.	Preparo de biblioteca	79
6.5.2.3.	Preparo de <i>template</i>	79
6.5.2.4.	Sequenciamento	79
6.6.	PADRONIZAÇÃO DO PROTOCOLO DE ANÁLISE DOS DADOS DE SEQUENCIAMENTO DE GENOMA COMPLETO DE <i>M. LEPRAE</i>	80
6.7.	OUTRAS ATIVIDADES	82
7.	RESULTADOS E DISCUSSÃO COMPLEMENTARES.....	83
7.1.	RECRUTAMENTO DA AMOSTRA POPULACIONAL E COLETA DE MATERIAL BIOLÓGICO	83
7.2.	PADRONIZAÇÃO DE SEQUENCIAMENTO E ANÁLISE DE GENOMA COMPLETO	84
7.2.1.	Análise geral do sequenciamento	85

7.2.2.	Análise de qualidade do sequenciamento.....	87
7.2.3.	Análise do mapeamento com o <i>M. musculus</i>	96
7.2.4.	Análise comparativa da montagem <i>de novo</i> e mapeamento com referência no software <i>MIRA</i>	96
7.2.5.	Análise comparativa da qualidade dos mapeamentos com referência ...	104
7.2.6.	Análise de variantes.....	111
8.	CONCLUSÕES COMPLEMENTARES.....	112
	REFERÊNCIAS.....	113
	APENDICE I: REVISÃO SISTEMÁTICA DE ESTUDOS DE VNTRS EM <i>M. leprae</i> – PROTOCOLO	124
	APENDICE II: ANALISE DE DADOS DO SEQUENCIAMENTO DE GENOMA COMPLETO – PROTOCOLO E PRINCIPIO DO METODO	127
	ANEXO I: ARTIGO DE REVISÃO PUBLICADO NA <i>CLINICS IN DERMATOLOGY</i>	141
	ANEXO II: ARTIGO PUBLICADO NO <i>INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY</i>.....	151
	ANEXO III: CAPITULO DO LIVRO "BIOÉTICA E PSICOLOGIA: INTER-RELAÇÕES"	161
	ANEXO IV: MANUSCRITO "THE PRATA VILLAGE: AN ESTIMATE OF THE INCUBATION PERIOD AND THE PROPORTION OF NATURALLY SUSCEPTIBLE INDIVIDUALS TO LEPROSY IN A HUMAN POPULATION"	174

1. INTRODUÇÃO

1.1. REVISÃO DE LITERATURA – PARTE 1 – HANSENÍASE

A hanseníase é uma doença infecciosa, estigmatizante e debilitante, cujo mais antigo indício de existência é um esqueleto encontrado na Índia, datado de 2.000 a.C, com lesões características da doença (1).

1.1.1. Agente Etiológico

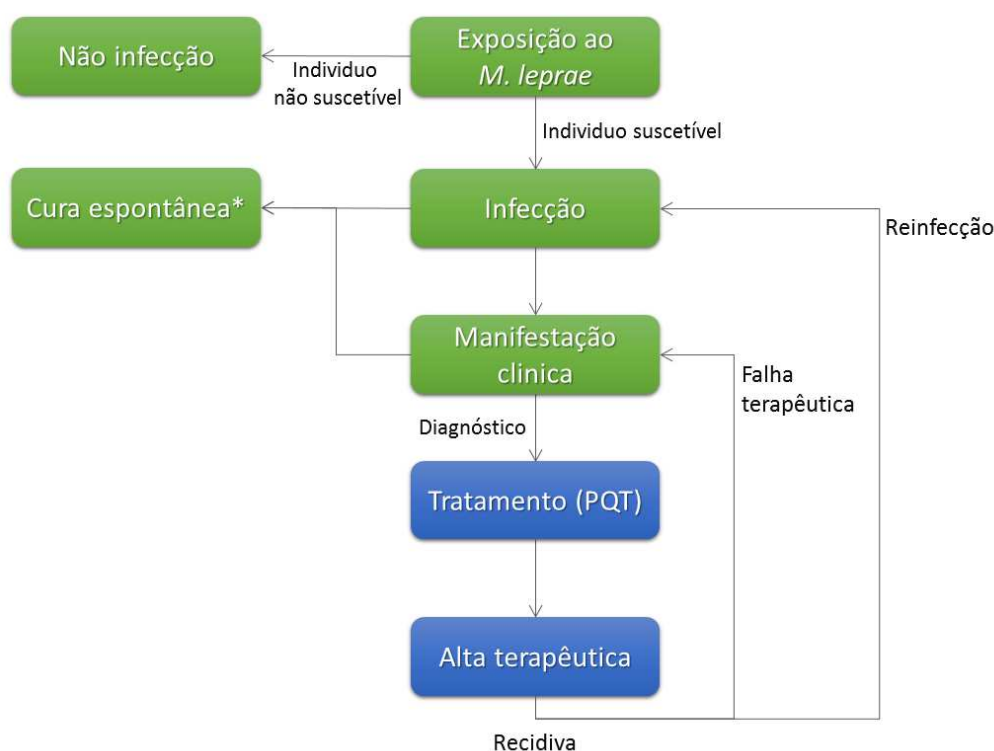
O agente etiológico da hanseníase é o *Mycobacterium leprae*, um bacilo que taxonomicamente pertence ao reino *Bacteria*, ao filo *Actinobacteria*, à classe *Actinobacteria*, à ordem *Corynebacteriales* e à família *Mycobacteriaceae*. O *M. leprae* mede de 1 a 8 µm, é álcool ácido-resistente (BAAR), gram positivo, sem motilidade, aeróbico, intracelular obrigatório e de crescimento lento. Ele possui uma parede celular de cerca de 20 nm de espessura, constituída de peptidoglicanos e cadeias polissacarídeas. Além disso, o bacilo apresenta uma cápsula caracterizada pela presença de componentes lipídicos e, até hoje, não é cultivável em meios artificiais (2, 3).

1.1.2. Patogêneses

A exposição ao *M. leprae* é necessária, mas não suficiente para o desenvolvimento da hanseníase. Caso o indivíduo exposto seja intrinsecamente suscetível ocorrerá infecção, acompanhada ou não de manifestações clínicas (**Figura 1**). Estes indivíduos são ditos suscetíveis à hanseníase *per se*, ou seja, à doença, independente da sua manifestação clínica, que se caracteriza por um amplo espectro de sinais e sintomas que variam conforme a resposta imunológica do hospedeiro (4).

Todo paciente diagnosticado com hanseníase tem direito a receber tratamento poliquimioterápico (PQT) gratuitamente. Após um curso de tratamento completo o paciente recebe alta terapêutica, porém, é possível a ocorrência de recidiva. Caso o indivíduo tenha sido curado e continue exposto à bactéria, pode ocorrer recidiva por reinfecção. Caso não tenha havido cura, o indivíduo pode sofrer reativação clínica, dita recidiva por falha terapêutica ou persistência. A falha terapêutica, por sua vez, pode derivar de um tratamento irregular, não adesão terapêutica ou resistência da bactéria à antibioticoterapia.

Figura 1 – Patogêneses da hanseníase.



Legenda: * A cura espontânea foi relatada em alguns casos esporádicos na época em que não havia um tratamento de efetividade terapêutica comprovada.

Fonte: a autora (2018), adaptado de (4).

1.1.3. Diagnóstico

O diagnóstico da hanseníase é essencialmente clínico e depende de exame cuidadoso da pele e nervos periféricos do paciente. Idealmente, toda a morfologia da pele deve ser observada, os nervos periféricos devem ser apalpados em busca de possíveis comprometimentos nervosos e pontos de anestesia (perda de sensibilidade) devem ser testados (especialmente em lesões cutâneas e áreas próximas a nervos possivelmente comprometidos). Um diagnóstico precoce e preciso possibilita o tratamento correto e diminui o desenvolvimento de sequelas (5). A partir de 1982 foi adotada uma classificação simplificada para fins de tratamento, proposta pela Organização Mundial da Saúde (OMS), que divide os pacientes em paucibacilares (PB) e multibacilares (MB), conforme critérios do **Quadro 1** (6).

Quadro 1 – Classificação da hanseníase para fins terapêuticos de acordo com OMS

Casos PB	Casos de hanseníase com 1 a 5 lesões de pele e ausência de bacilos detectáveis na baciloscopia
Casos MB	Casos de hanseníase com mais de 5 lesões de pele ou com qualquer envolvimento neural ou que teve presença de bacilos na baciloscopia (independentemente do número de lesões de pele)

Fonte: Traduzido e adaptado de (7)

A baciloscopia é um exame diagnóstico complementar especialmente recomendado em casos de: (i) diagnóstico diferencial com outras doenças dermatoneurológicas; (ii) dúvidas na classificação; (iii) casos suspeitos de recidiva. Usualmente, linfa é coletada das orelhas, cotovelo e/ou lesões cutâneas do paciente e o material é aplicado numa lâmina. Essa lâmina é, então, corada com a coloração de Coloração de Kinyoun (também conhecido como “Coloração de Ziehl-Neelsen a frio”) e observada ao microscópio sob o aumento de 100x na objetiva de imersão. O resultado é dado numa escala logarítmica na forma de um parâmetro denominado “índice baciloscópico” (IB) (**Quadro 2**) (8-10).

Quadro 2 – Escala Logarítmica de Ridley para baciloscopia de hanseníase.

Escala	Significado
0	Ausência de bacilos detectáveis em 100 campos examinados.
1+	Presença de 1 a 10 bacilos, em 100 campos examinados.
2+	Presença de 1 a 10 bacilos, em cada 10 campos examinados.
3+	Presença de 1 a 10 bacilos, em média, em cada campo examinado.
4+	Presença de 10 a 100 bacilos, em média, em cada campo examinado.
5+	Presença de 100 a 1.000 bacilos, em média, em cada campo examinado.
6+	Presença de mais de 1.000 bacilos, em média, em cada campo examinado.

Fonte: Adaptado de (8, 9)

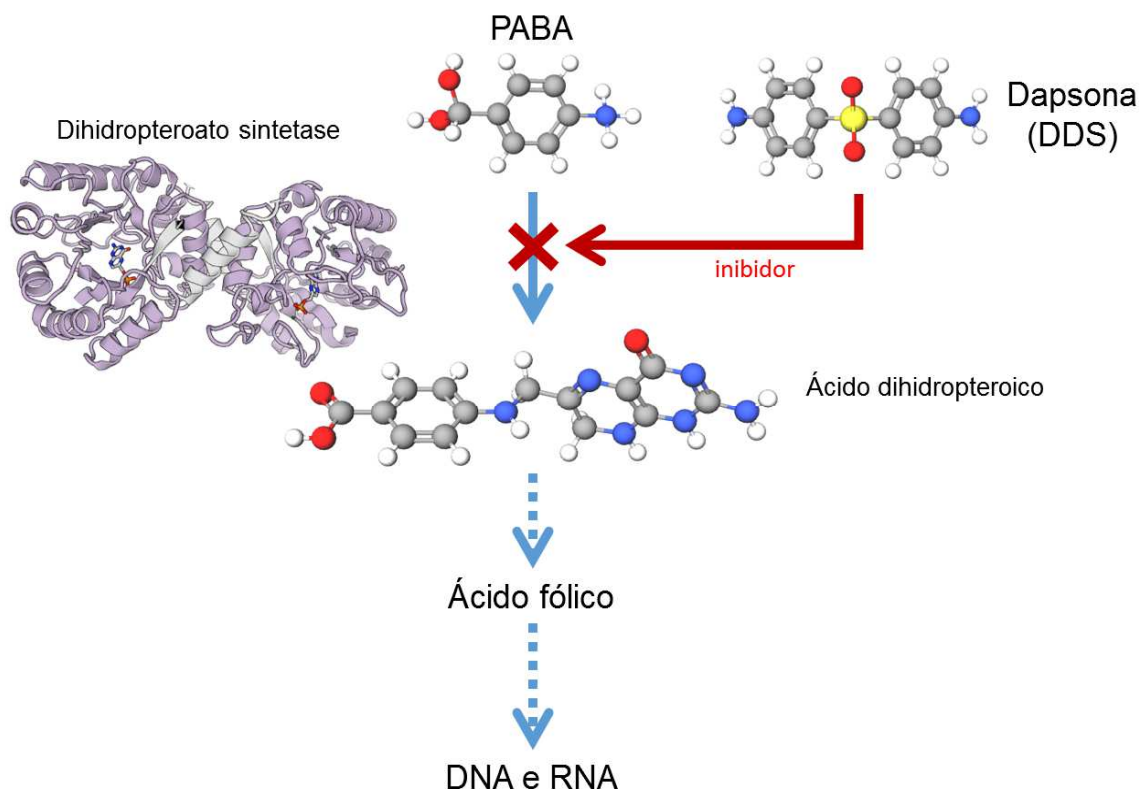
1.1.4. Tratamento e Resistência

Durante muitos anos, a hanseníase foi uma doença incurável e seu principal tratamento, de efetividade questionável, era baseado no uso do óleo de chaulmoogra (árvore da família das Flacourtiáceas) (11, 12).

1.1.4.1. Tratamento com dapsona

Apenas na década de 40 começaram a ser utilizados experimentalmente os derivados sulfônicos para fins de contenção da infecção pelo *M. leprae*, como o promim, a diazona e a dapsona (DDS, do inglês *diaminodiphenyl sulfone*) (13-16). Nos anos seguintes, a DDS passou a ser utilizada como terapia de escolha contra hanseníase, com base na sua efetividade, baixo custo e facilidade de administração (por via oral). A DDS é um agente bacteriostático que atua como análoga sintética do ácido para-aminobenzóico (PABA), necessário na síntese de ácido fólico e consequentemente na síntese de DNA e RNA pelas bactérias (**Figura 2**) (17-20).

Figura 2 – Mecanismo de ação da DDS



Legenda: A seta azul contínua representa a síntese do ácido dihidropteroico. A seta azul tracejada representa a série de reações que resultará na síntese do DNA. A seta e x vermelhos indicam a ação inibidora. Nas estruturas secundárias: as esferas cinzas representam átomos de carbono, as brancas de hidrogênio, as vermelhas de oxigênio, as azuis de nitrogênio e a amarela de enxofre.

Fonte: a autora (2018), utilizando figuras obtidas nos softwares MolView (21) e SWISS-MODEL (22).

Pouco tempo depois da instituição da monoterapia com DDS, foram observados que casos de hanseníase tratados com DDS por longos períodos por vezes deixavam de responder à droga e evoluíam para recidiva da doença (17, 23). Importante destacar que o *M. leprae* é incapaz de crescer em meios de cultura artificiais existentes, impossibilitando a confirmação da resistência *in vitro* através de técnicas clássicas da microbiologia, como, por exemplo, por antibiograma. Apenas em 1960 foi desenvolvida por Shepard uma técnica de inoculação de *M. leprae* em coxim plantar de camundongos (*in vivo*), a partir da qual tornou-se possível detectar resistência do bacilo (24). Utilizando essa técnica, em 1964, sete pacientes tratados ao longo de 13 a 15 anos com DDS tiveram as bactérias testadas para resistência e três delas se mostraram experimentalmente resistentes à DDS. Esses foram os primeiros casos comprovados de resistência à DDS na hanseníase (25).

Posteriormente, outros relatos de resistência à DDS começaram a surgir em várias partes do mundo. (17, 26, 27).

1.1.4.2. Tratamento com clofazimina

Em 1962, a clofazimina (CLF) foi utilizada pela primeira vez em três pacientes da Nigéria. Esses indivíduos já haviam sido tratados com DDS e a bactéria desenvolveu resistência a esse tratamento. Durante um ano de tratamento com CLF, foi relatada melhora clínica e redução do IB de todos os pacientes (28). Outro estudo do uso da CLF foi realizado em 1981, em 20 pacientes que tinham suspeita clínica de resistência ou tinham resistência comprovada à DDS. Todos os 20 pacientes relataram melhora clínica considerável.

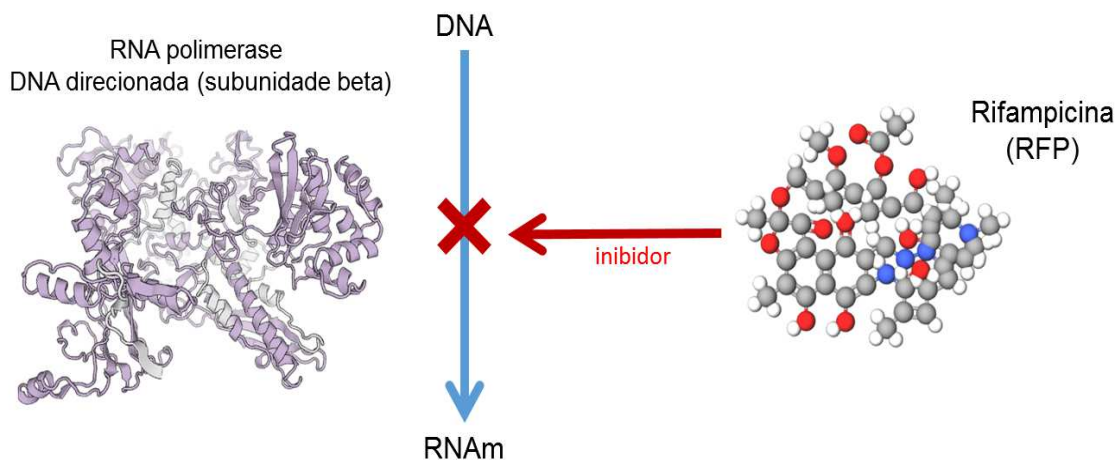
O mecanismo de ação da CLF ainda não foi completamente elucidado, sabe-se que ela se liga ao DNA bacteriano e possivelmente inibe sua função como molde, eventualmente levando a morte celular (29). Com isso o efeito da CLF é fracamente bactericida e, portanto, inadequado para que ela fosse utilizada como monoterapia de escolha para hanseníase. Além disso, essa droga causa, como efeito colateral, o desenvolvimento de uma pigmentação avermelhada na pele do paciente, especialmente visível na pele clara (30).

1.1.4.3. Tratamento com rifampicina e resistências primárias

Na década de 1970 foi realizado um estudo experimental e clínico que comprovou que a rifampicina (RFP) inibia a multiplicação de *M. leprae* em patas de camundongos. Além disso, nesse mesmo estudo, a RFP se mostrou mais efetiva que a DDS no tratamento da hanseníase (31).

A RFP é um derivado piperazínico, que possui um efeito altamente bactericida, atuando seletivamente sobre a enzima RNA polimerase da bactéria (codificado pelos genes *rpoA*, *rpoB*, *rpoC*, *rpoT*) (32), causando uma mudança de conformação que resulta na inativação da enzima e, conseqüentemente, no bloqueio da síntese de RNA mensageiro (**Figura 3**) (33).

Figura 3 – Mecanismo de ação da RFP



Legenda: A seta azul contínua representa a síntese do RNA mensageiro. A seta e x vermelhos indicam a ação inibidora. Nas estruturas secundárias: as esferas cinzas representam átomos de carbono, as brancas de hidrogênio, as vermelhas de oxigênio e as azuis de nitrogênio.

Fonte: a autora (2018), utilizando figuras obtidas nos softwares MolView (21) e SWISS-MODEL (22).

Mesmo após a experiência com o desenvolvimento de resistência devido à monoterapia sulfônica, a RFP também foi utilizada ocasionalmente em regime de monoterapia. Conseqüentemente, em 1976 foi inoculado em pata de camundongo a cepa de um paciente escandinavo com recidiva da doença após ser tratado por muitos anos com DDS e, posteriormente, com RFP, que se revelou o primeiro caso de *M. leprae* resistente a RFP (34).

Em 1977, Pearson e col. relataram que, de oito cepas inoculadas em patas de camundongo advindas de pacientes da Etiópia sem tratamento prévio, cinco delas apresentaram resistência à DDS, sendo esses os primeiros relatos de resistência primária à DDS em hanseníase (35). Mais tarde, outros trabalhos reportaram casos de *M. leprae* resistentes tanto à DDS quanto à RFP, obtidos principalmente de pacientes tratados em esquema de monoterapia. Por exemplo, um estudo acompanhou todos os pacientes tratados para hanseníase MB no período de 1975 a 1978 no Cebu (nas Filipinas), constatando que 42 apresentaram recidiva e foram suspeitos de terem desenvolvido resistência secundária à DDS. Utilizando a técnica de Shepard, 31 casos de resistência secundária à DDS foram confirmados. Conseqüentemente, foi feita uma busca por casos de resistência primária, inoculando em pata de camundongo amostras de 58 pacientes sem tratamento prévio. Duas (de 55 amostras que cresceram nos camundongos controle) também

criaram nos camundongos que receberam DDS, resultando numa prevalência de resistência primária à DDS estimada em 3,6% (36).

Estudo publicado em 1984 mostrou que 9/45 casos de recidiva eram causados por cepas resistentes à RFP *in vivo*. O tempo e constância de terapias prévias com RFP nesses nove pacientes foram variáveis (5 a 72 meses). Apenas uma dessas nove cepas era sensível à DDS, isolada do único paciente que não havia recebido monoterapia com este medicamento; todas as demais apresentaram também resistência à DDS (37).

1.1.4.4. Surgimento da Poliquimioterapia

Devido ao grande número de casos de resistência já relatados na década de 1980, concluiu-se que, para prevenir o aparecimento e disseminação de cepas resistentes, a solução seria utilizar um regime combinado de antibióticos. Na época, apenas quatro drogas estavam disponíveis e poderiam ser utilizadas nessa terapia combinada: RFP, DDS, CLF e etionamida. No caso de resistência prévia a alguma droga (em especial à DDS), o uso combinado com apenas mais um antibiótico equivaleria a uma monoterapia desse antibiótico, propiciando o desenvolvimento de resistência. Portanto, pelo menos três antibióticos deveriam ser utilizados. A DDS, por ser efetiva, barata e já amplamente utilizada na hanseníase, foi um dos medicamentos escolhidos para compor esse novo regime. A RFP, por ser bactericida, deveria também estar presente na combinação. Como a etionamida era mais hepatotóxica que a CLF, a CLF foi escolhida como terceiro componente. Em 1981 (6), a Organização Mundial da Saúde (OMS) passou a recomendar a poliquimioterapia (PQT) baseada nos supracitados antibióticos para hanseníase, com um regime de seis meses para pacientes PB e de pelo menos dois anos para pacientes MB. Ao longo dos anos, pequenas alterações foram propostas nesse regime: em 1988 (38), 1994 (39) e 1998 (40). Uma das principais mudanças realizadas foi a redução do tratamento para pacientes MB de 24 para 12 meses. Importante destacar que essa mudança não foi baseada em estudos clínicos e sim em conveniência operacional. O regime atualmente utilizado é detalhado na **Tabela 1** (12, 17).

Tabela 1 - Regime recomendado da Poliquimioterapia a partir de 1998.

	Pacientes MB (12 meses)	Pacientes PB (6 meses)
Mensalmente (Supervisionada)	RFP (600mg) + CLF (300mg) + DDS (100mg)	RFP (600 mg) + DDS (100mg)
Diária (Autoadministrada)	DDS (100mg) + CLF (50mg)	DDS (100mg)

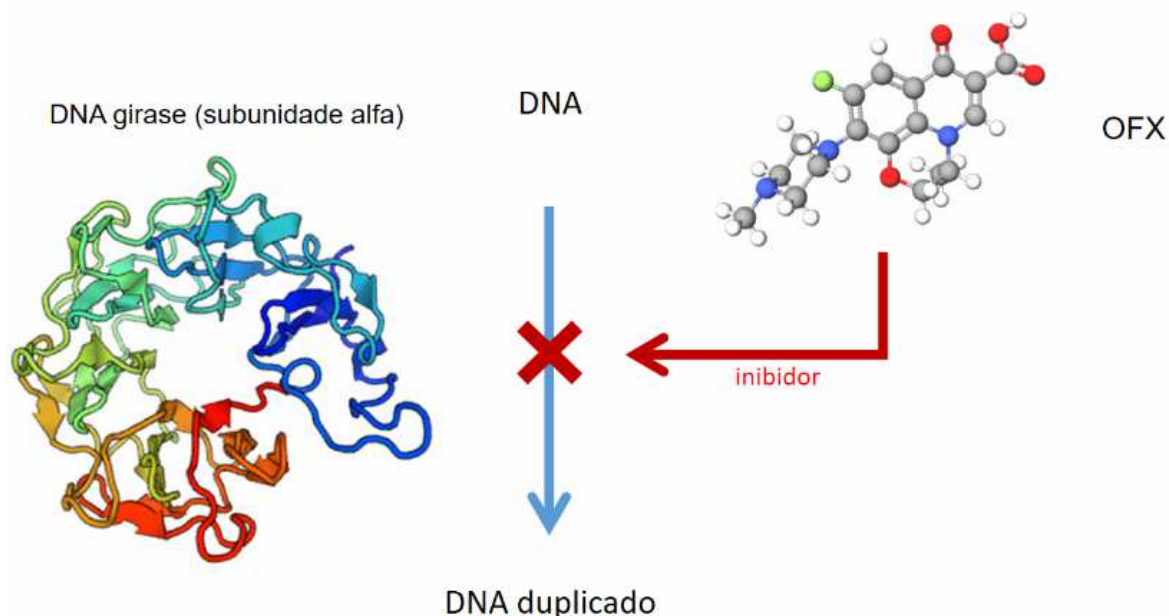
Abreviações: MB: multibacilar; PB: paucibacilar; DDS: dapsona; RFP: rifampicina; CLF: clofazimina mg: miligramas.

Fonte: Adaptado de (40)

1.1.4.5. Tratamento com ofloxacina

Em 1986 foram feitos testes do uso *in vivo* de ofloxacina (OFX) para tratamento da hanseníase, um antibiótico da família das quinolonas que se mostrou moderadamente bactericida contra o *M. leprae*. A OFX se liga à subunidade A da DNA girase (codificada pela *gyrA*), inibindo a replicação do DNA bacteriano (**Figura 4**). Desde então, a OFX passou a ser usada como droga de segunda linha, em caso de resistência a componentes da PQT (41, 42).

Figura 4 – Mecanismo de ação da OFX



Legenda:

A seta azul continua representa a duplicação do DNA. A seta e x vermelhos indicam a ação inibidora. Nas estruturas secundárias: as esferas cinzas representam átomos de carbono, as brancas de hidrogênio, as vermelhas de oxigênio, as azuis de nitrogênio e a verde de flúor.

Fonte: a autora (2018), utilizando figuras obtidas nos softwares MolView (21) e SWISS-MODEL (22).

1.1.4.6. Multirresistência

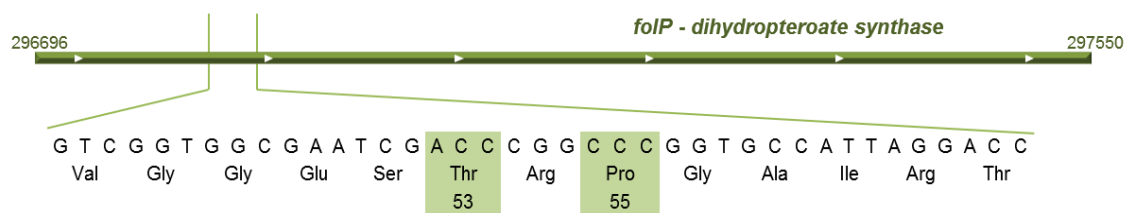
Em 1996, Shetty e col. estudaram três cepas de pacientes sem tratamento prévio para hanseníase, das quais duas apresentaram resistência *in vivo*; uma delas, apenas a uma droga (DDS) e a outra, a múltiplas drogas (DDS, RFP, CLF). Dessa forma, os autores descreveram, pela primeira vez, um caso de *M. leprae* multirresistente que era também um caso de resistência primária (43).

O segundo caso de *M. leprae* resistente a múltiplas drogas foi também o primeiro caso de resistência a fluoroquinolonas. A cepa foi isolada de um paciente do Mali previamente tratado com monoterapia dapsônica por 12 anos e terapia combinada de RFP e OFX por um mês. Essa cepa foi testada em pata de camundongo e apresentou resistência à DDS, RFP e OFX. Também foram investigadas as bases moleculares dessas resistências, tendo sido encontrada uma mutação no gene *rpoB* (TCG → TTG - Leu 425 por Ser) e uma no gene *gyrA* (GCA → ATA - Val 91 por Ala) (**Figura 5 a 7**) (44).

Em 1999, ao ser analisada a sequência de nucleotídeos dos genes *folP* e *folP2* (responsáveis pela codificação da enzima diidropteroato sintetase – DHPS – no *M. leprae*) em seis diferentes cepas de *M. leprae* resistentes à DDS, foram observadas duas mutações *missense* nos códons 53 e 55 do gene *folP* (**Figura 5**). Embora a mutação no códon 55 - ou em seus sítios homólogos - tenha sido relatada previamente em outras espécies bacterianas, a mutação no códon 53 estava sendo relatada pela primeira vez em bactérias (45). A enzima DHPS é responsável por processar tanto a DDS quanto o PABA (**Figura 5**), sendo, portanto, o principal *locus* de ocorrência de mutações associadas à resistência do bacilo à DDS. O mecanismo de resistência mais comum a essa droga, em geral, envolve a síntese de uma enzima que não reconhece – e, portanto, não processa – seletivamente a DDS (46-49).

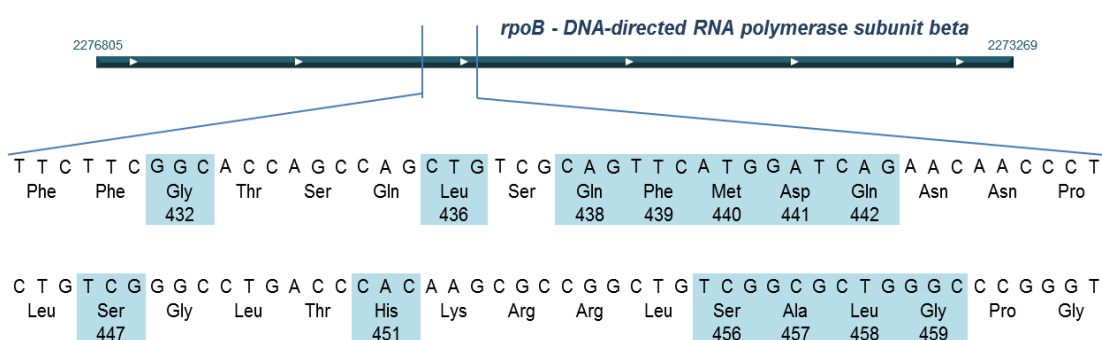
Pouco tempo depois, em 2000, Matsuoka e col. publicaram o terceiro caso de *M. leprae* multirresistente. Essa cepa era advinda de um paciente japonês tratado com monoterapia dapsônica por 20 anos, por DDS combinada com RFP por mais 11 anos, protocolo ao qual foi adicionada a OFX. Ainda assim, o paciente manteve um IB alto. O isolado foi testado para resistência em pata de camundongo, que revelou resistência à DDS, RFP, OFX e sparfloxacina, e sensibilidade à claritromicina e CLF (50).

Figura 5 – Representação esquemática das regiões determinantes de resistência a antibióticos (*drug resistance determining region - DRDR*) mais comumente relatadas no gene *folP* do *M. leprae*



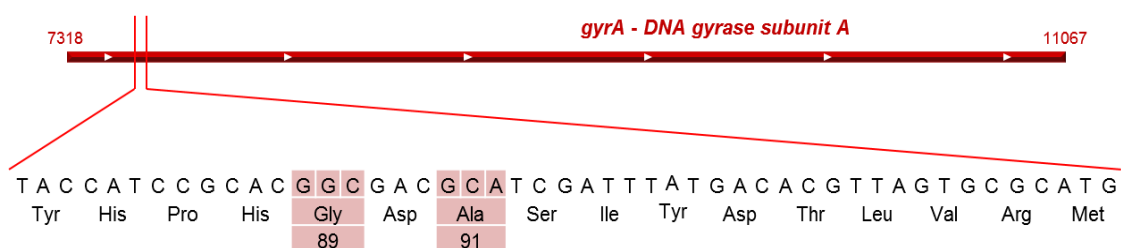
Fonte: a autora (2018) baseada na sequência do gene *folP* disponível em (51).

Figura 6 – Representação esquemática das regiões determinantes de resistência a antibióticos (*drug resistance determining region - DRDR*) mais comumente relatadas no gene *rpoB* do *M. leprae*



Fonte: a autora (2018) baseada na sequência do gene *rpoB* disponível em (52)

Figura 7 – Representação esquemática das regiões determinantes de resistência a antibióticos (*drug resistance determining region - DRDR*) mais comumente relatadas no gene *gyrA* do *M. leprae*



Fonte: a autora (2018) baseada na sequência do gene *gyrA* disponível em (53)

1.1.4.7. Tratamentos substitutivos

Atualmente no protocolo de hanseníase do Ministério da Saúde não existe um tratamento substitutivo específico para os casos de resistência. Nesses casos, os protocolos adotados são os mesmos de intolerância à medicação (**Quadro 3**) (10).

Quadro 3 – Tratamento recomendando pelo MS para doente com intolerância a um ou mais componentes da PQT.

Intolerância a	Forma clínica (duração)	Dose mensal (supervisionada)	Dose diária (auto administrada)	
DDS	PB (6 meses)	RFP (600mg) + CLF (300mg)	CLF (50mg)	
	MB (12 meses)	RFP (600mg) + CLF (300mg) + OFL (400mg)	CLF (50mg)	
		OU MNC (100mg)	OU MNC (100mg)	
RFP	PB (6 meses)	DDS (100mg) + OFX (400mg)	DDS (100mg) + OFX (400mg)	
		OU MNC (100mg)	OU MNC (100mg)	
	MB (12 meses)	DDS (100mg) + OFX (400mg) + CLF (300mg)	DDS (100mg) + OFX (400mg) + CLF (50mg)	
		OU MNC (100mg)	OU MNC (100mg)	
		CLF (300mg) + OFX (400mg)	CLF (50mg) + OFX (400mg)	
RFP e a DDS	PB (6 meses)	OU MNC (100mg)	OU MNC (100mg)	
		MB (primeiros 6 meses)	CLF (300mg) + OFX (400mg) + MNC (100mg)	CLF (50mg) + OFX (400mg) + MNC (100mg)
			CLF (300mg) + OFX (400mg)	CLF (50mg) + OFX (400mg)
	MB (18 meses subsequentes)	OU CLF (300mg) + MNC (100mg)	OU CLF (50mg) + MNC (100mg)	

Legenda: PB = paucibacilar; MB = multibacilar; DDS = dapsona; RFP = rifampicina; CLF = clofazemina; OFX = ofloxaxina; MNC = minociclina;

Fonte: Adaptado de (10)

1.1.4.8. Monitoramento da resistência

Em 2008, a OMS organizou um *workshop* para discutir os crescentes relatos de resistência publicados; dentre os pontos mais importantes, foi discutida a criação de um sistema de monitoramento da resistência em hanseníase (54). Dessa forma, foi publicado em 2009 o “*Guidelines for Global Surveillance of Drug Resistance in Leprosy*” que determina que os países endêmicos que participam do monitoramento devem coletar amostras de pacientes MB, com IB maior que “2+”, para avaliação molecular de resistência do bacilo à DDS, RFP e OFX nos casos de recidiva de hanseníase (55). Após essa proposta, a OMS publicou três relatórios do monitoramento de resistência (56-58), - descritos sucintamente na **Tabela 2** - quatro relatórios de reuniões anuais (2009-2012) sobre o monitoramento da resistência (59-62) e uma consulta em monitoramento de resistência (63), no qual constam dados referentes de 2010 a 2015. Nesse período, 5,2% (57/1086) eram casos de resistência secundária à RFP e 2,1% (16/776) eram casos de resistência primária à RFP (7).

Recentemente, um estudo foi realizado na Colômbia com DNA de *M. leprae* advindo de 243 indivíduos coletados como parte do programa de monitoramento de resistência da OMS de 2004 a 2013. Após o sequenciamento das DRDR, foram encontrados 3 isolados com mutações apenas no gene *rpoB* (Asp441Tyr, Ser456Leu, Ser458Met); 3 com mutações apenas no gene *folP* (Thr53Ala, Pro55Leu, Thr53Ala + Thr88Pro) e 1 isolado com mutações no *rpoB* (Ser456Met) e no *folP* (Pro55Leu). Também foram encontradas mutações nunca associadas à resistência no *rpoB* (Ala426Thre e Arg505Trp) no *folP* (Asp91His, Arg94Trp e Thr88Pro) e na *gyrA* (Ala107Leu) (64).

Outro estudo examinou biópsias de 25 pacientes da Guinéa-Conakry e identificou, molecularmente, 4 casos de resistências primárias (1 com a mutação Ser456Leu no *rpoB* e 3 com Thr53Arg no *folP*). Esse mesmo estudo sequenciou as três cepas (MI2-10, MI6-50 e MI6-55) com resistência à dapsona, identificando uma diferença de apenas 2 *SNPs* dentre elas, indicando que elas eram originalmente a mesma cepa (65).

Um estudo realizado na Índia, por sua vez verificou que dentre 239 recidivas e 11 casos novos 21,6% tinham resistência a um antibiótico e 6,8% tinham resistência a mais de um antibiótico (66).

Esses resultados levaram a OMS a rever o monitoramento de resistência à hanseníase, culminando com a publicação do “*A guide for surveillance of antimicrobial resistance in leprosy: 2017 update*” em 2017 (7). Esse documento se baseou nas definições apresentadas no **Quadro 4** e propõe que todos os casos de re-tratamento (exceto os de transferência) e parte dos casos novos (a quantidade pode variar conforme a disponibilidade de cada país) sejam testados para mutações associadas à resistência a DDS, RFP e OFX. É recomendado que sejam testados apenas aqueles casos que apresentem IB maior que 2+ para aumentar as chances de obter uma PCR positiva.

Por fim, um estudo recente analisou 1.932 cepas de *M. leprae* (obtidas entre 2009 e 2015 em centros de monitoramento de 19 países) e constatou que 8,0% (154) delas apresentaram algum grau de resistência. As resistências primárias e secundárias a RFP foram de 2,0% (16/789) e 5,2% (58/1143), respectivamente. Também foram constatados 20 casos (1,0 %) de resistência dupla a DDS e RFP (67).

Quadro 4 – Definições da OMS

Novo caso	Paciente diagnosticado com hanseníase que nunca foi tratado para a doença
Re-tratamento	Paciente diagnosticado com hanseníase que já recebeu tratamento para a doença no passado. Sendo sub-dividido em:
Após abandono	Paciente diagnosticado com hanseníase que abandonou o tratamento antes de completa-lo e retornou à unidade de saúde para completar o tratamento
Recidiva	Paciente que completou todo um tratamento de hanseníase no passado e que teve reaparecimento dos sinais e sintomas da doença que não foram considerados decorrentes de um estado reacional.
Transferência	Paciente que começou o tratamento num local e se reportou a um novo local para dar continuidade ao tratamento
Outros re-tratamentos	Casos de re-tratamento que não possam ser enquadrados em nenhuma das outras categorias

Fonte: Traduzido e adaptado de (7)

Tabela 2 – Dados dos relatórios de monitoramento da resistência da OMS.

<i>Pais</i>	<i>Período</i>	<i>Número de casos testados</i>	<i>Resistência a RFP</i>	<i>Resistência a DDS</i>	<i>Resistência a OFX (ou quinolonas)</i>	<i>Resistência simultânea a RFP e DDS</i>	<i>Taxa de resistência (resistência a cada 100 casos testados)</i>
Benin	2011-2015 ^a	70	1	3	0	0	5,7
Brasil	2009 ^b	135	2	4	0	NT	4,4
	2010-2015 ^a	244	3	2	2	25	13,1
Birmânia	2005-7 ^c	10	3	2	0	NT	50,0
	2009-10 ^{b,d}	44	4	1	0	NT	11,4
China	2009 ^b	10	NT	0	NT	NT	0,0
	2010 ^d	8	1	0	0	NT	12,5
	N.D. ^a	116	1	8	0	0	7,8
Colômbia	2006-10 ^{b,c}	36	7	7	0	NT	38,9
Filipinas	2009-2015 ^a	140	NI	4	NI	NI	2,9
Guinea	2013-2015 ^a	27	1	0	0	0	3,7
Índia	2009-10 ^{b,d}	58	5	0	2	NT	12,1
	2010-2015 ^a	661	40	24	15	NI	12,0
Indonésia	2003-2011 ^a	270	NI	5	NI	1	2,2
	2013-2015 ^a	70	1	3	NI	NI	5,7
Madagascar	2013-2015 ^a	212	1	1	0	0	0,9
Mali	2011 ^a	93	0	5	0	0	5,4
Moçambique	2013-2014 ^a	9	1	0	0	0	11,1
Myanmar	2010-2015 ^a	108	1	3	0	0	3,7

Nepal	2009-2015 ^a	178	10	2	0	2	7,9
Nigéria	2013-2015 ^a	40	1	1	0	0	5,0
Paquistão	2010 ^d	8	0	0	0	NT	0,0
Vietnam	2009 ^b	12	3	0	0	NT	25,0
	2010 ^d	6	3	0	0	NT	50,0
	2010-2015 ^a	290	0	20	0	NI	6,9
Iêmen	2010 ^d	3	0	0	0	NT	0,0

Legenda: As letras sobrescritas na coluna "Período" indicam o relatório fonte das informações, sendo: ^a: (63); ^b: (57); ^c:(58); ^d:(56);

N.D.: Ano de amostragem não declarado.

Fonte: a autora (2018) a partir de dados de (56-58, 63).

1.1.5. Epidemiologia

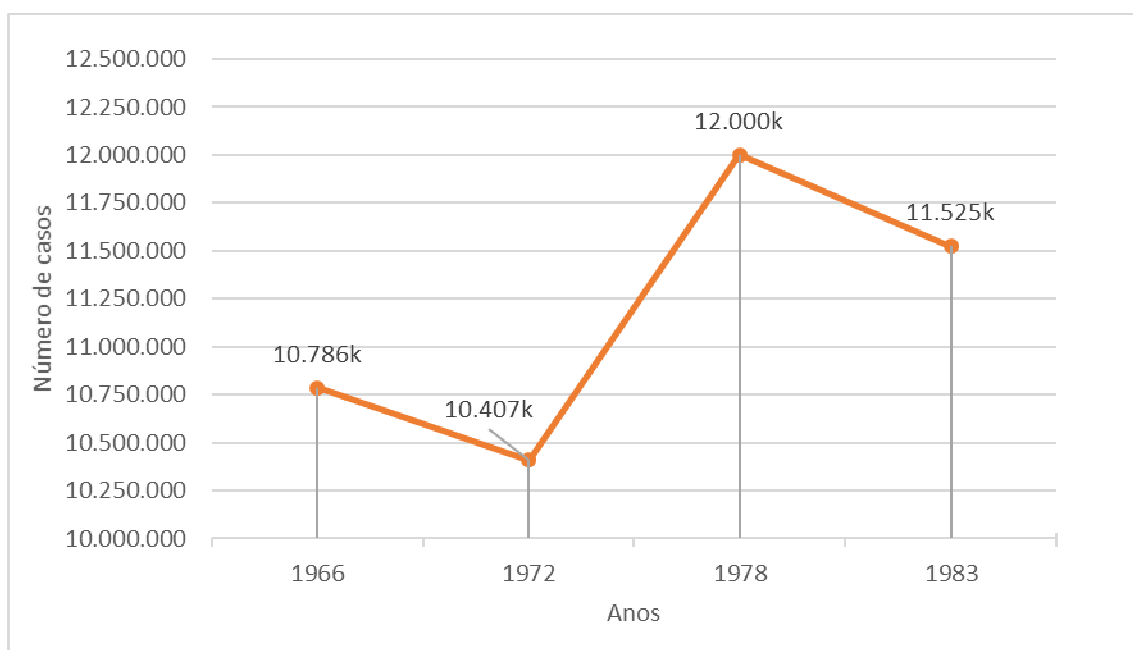
Em 1966 foi estimado que haviam 10.786.000 casos de hanseníase no mundo (68). Nas duas décadas seguintes, as estimativas ficaram entre 10 a 12 milhões de casos no mundo (**Figura 8**). A partir de 1985 essas estimativas passaram a se basear em dados mais precisos e a serem feitas anualmente, levando em consideração a prevalência e a incidência da doença (69). O paciente era considerado um “caso” para contabilização da prevalência enquanto não completava a PQT. Quando a PQT para pacientes MB reduziu de 2 para 1 ano (em 1998), a contagem de “casos” MB foi conseqüentemente reduzida pela metade (**Figura 9**) (17).

Em 1991, a OMS lançou uma campanha para que, utilizando a PQT, em 2000 a hanseníase fosse eliminada como problema de saúde pública, ou seja, houvesse menos de um caso a cada 10.000 pessoas (70). Essa campanha nunca teve a aquiescência do grupo de especialistas de hanseníase da OMS (*WHO leprosy expert committees*), pois alguns fatores claramente dificultavam essa eliminação, tais como i) o longo tempo de incubação (assintomático) da doença; ii) a falta de informação sobre possíveis fontes não humanas (ambientais e animais) de infecção; iii) a ausência de vacina ou outra medida efetiva de prevenção (17). Ainda assim, a partir 1995, a PQT passou a ser fornecida gratuitamente graças ao suporte da *The Nippon Foundation (TNF)* e *Sasakawa Memorial Health Foundation (SMHF)* do Japão (de 1995 a 1999) e posteriormente pela *Novartis* e *Novartis Foundation for Sustainable Development (NFSD)* (71). Nesse contexto, uma grande pressão foi feita para que os países alcançassem a meta de eliminação, levando, em alguns casos, à distorção das políticas públicas de saúde. Um dos exemplos mais característicos ocorreu na Índia, que era responsável pela maioria dos casos do mundo e que adotou medidas como: i) deixar de contabilizar os casos de hanseníase do tipo “lesão única” (cerca de um terço dos casos de hanseníase do país); ii) passar a ter a contabilidade dos casos “verificada” pelos gestores da campanha; iii) desencorajar a busca ativa de casos; iv) deixar de contabilizar como “caso” o paciente recém diagnosticado que havia recebido toda a medicação. Dessa forma, a Índia reduziu o número de novos casos de 559.938 em 2000 para 139.252 em 2006, uma queda de 75% em 6 anos, improvável para uma doença cujo período de incubação é de cerca

de 15 anos. Essa diferença é a principal responsável pelo decréscimo observado tanto na incidência quanto na prevalência global entre 2000 a 2006 (**Figura 10**) (17).

Em 2004 a OMS declarou que, de 131 países previamente endêmicos, apenas nove não atingiram a meta, a saber: Brasil, Congo, Madagascar, Nepal, Tanzânia, Angola, Libéria, Moçambique e República Centro-Africana (72). Porém, ao se observar a incidência mundial de 1985 a 2016 (**Figura 9 e 10**) nota-se que, com exceção de um pequeno aumento de 1997-2001 (atribuída ao incentivo na busca ativa de casos e implementação dos programas de eliminação) e uma redução de 2001 a 2005 (supracitado período de mudanças na contabilidade de casos na Índia) este parâmetro tem permanecido estável, indicando que a transmissão continua a ocorrer similarmente com o que ocorria 30 anos atrás (17, 73).

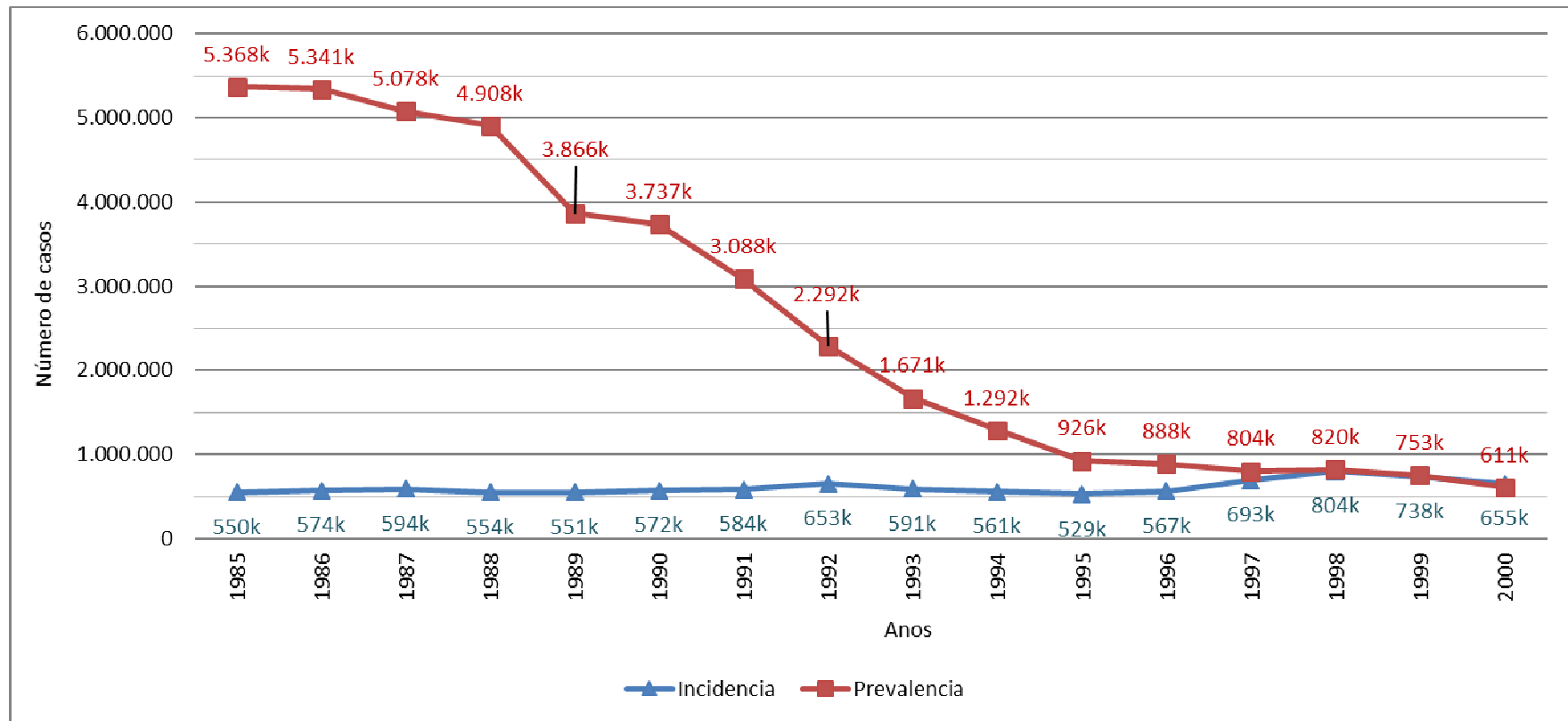
Figura 8 - Prevalência estimada de hanseníase no mundo de 1966 a 1983



Legenda: O símbolo “k” equivale à notação “vezes 10³”. Incidência é o número de novos casos reportados no ano. Prevalência é o número de casos registrados.

Fonte: a autora, baseado em dados de (68, 74-76)

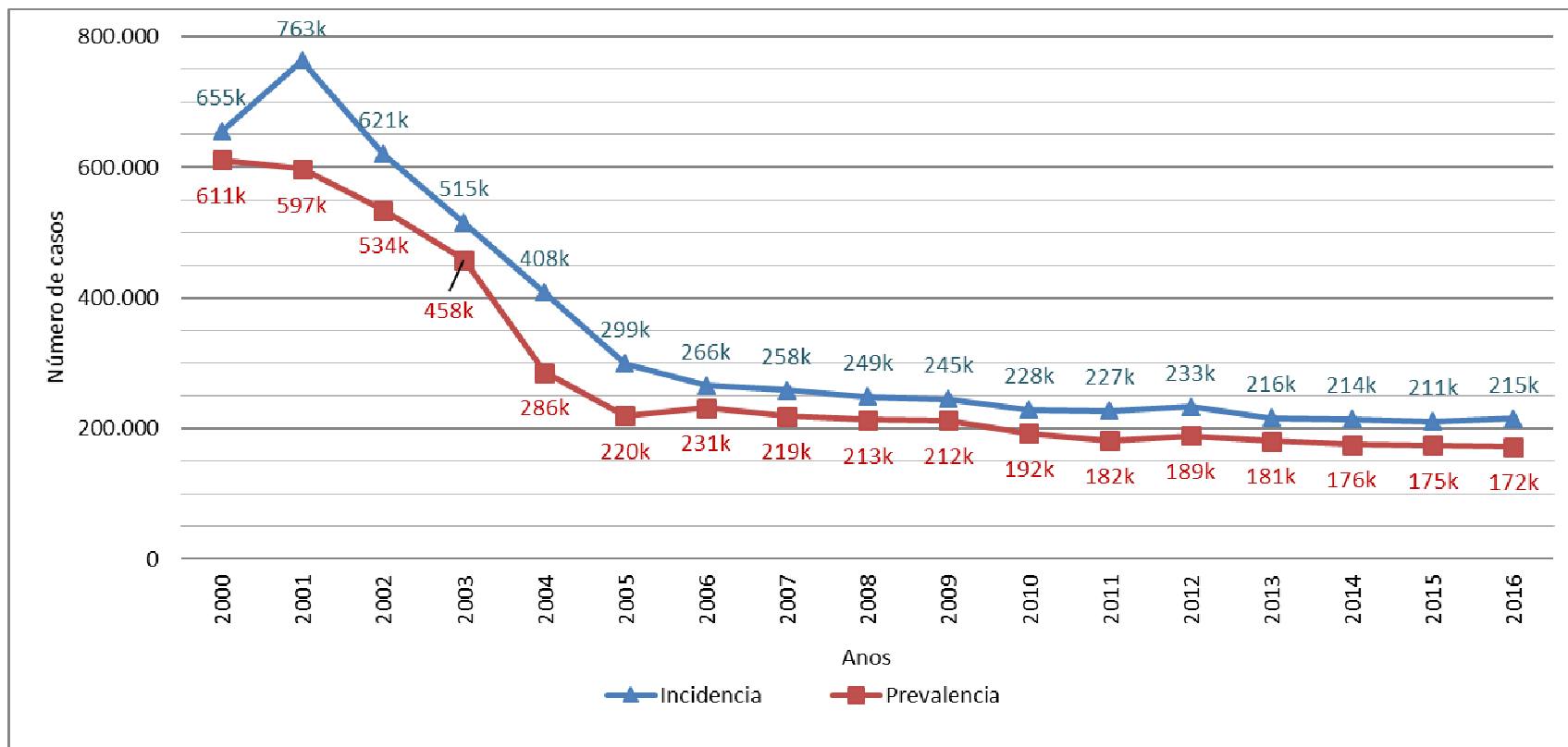
Figura 9 - Prevalência e incidência de hanseníase no mundo de 1985 a 2000



Legenda: Incidência é o número de novos casos reportados no ano. Prevalência é o número de casos registrados.

Fonte: A autora, baseado em dados de (77-89).

Figura 10 - Prevalência e incidência de hanseníase no mundo de 2000 a 2016



Legenda: O símbolo “k” equivale à notação “vezes 10³”. Incidência é o número de novos casos reportados no ano. Prevalência é o número de casos registrados.

Fonte: A autora, baseado em dados de (77-91)

1.1.6. Variabilidade do *M. leprae*

Idealmente para redução/eliminação de qualquer doença infecciosa é crítico conhecer os mecanismos de infecção e transmissão do patógeno. No caso da hanseníase, o longo tempo de incubação e a impossibilidade de cultivar o bacilo em meios artificiais dificultam consideravelmente o estudo desses mecanismos utilizando técnicas clássicas da microbiologia. Recentemente, graças ao desenvolvimento de ferramentas de análise molecular, diferentes abordagens foram propostas para delinear transmissão e variabilidade do *M. leprae* (92).

1.1.6.1. VNTRs

Dentre as primeiras técnicas moleculares utilizadas para avaliação da variabilidade do *M. leprae*, estão as comparações genômicas baseadas em *RFLP* (*restriction fragment-length polymorphism*) e sequenciamento de 16S-23S rDNA, que não conseguiam diferenciar dentre cepas de *M. leprae* de diferentes origens. Até que, em 2000, Masuoka (92) propôs o uso de uma região de repetição curta do gene *rpoT* (*RNA polymerase sigma factor* ou fator sigma da RNA polimerase) do *M. leprae* para diferenciação de cepas. Esse marcador foi testado em 51 isolados de 12 áreas geográficas distintas, tendo encontrado grupos (com 3 e 4 cópias) coerentes com a distribuição geográfica dos isolados. Posteriormente, esse marcador passou a ser conhecido como “6-3a”. Nos anos seguintes, diversos outros marcadores foram propostos. Uma revisão sistemática (**APENDICE I**) revelou que 47 marcadores de VNTRs (*Variable-Number Tandem-Repeat*) já foram testados para uso em estudos de variabilidade do *M. leprae*. No entanto, apenas 17 marcadores se estabeleceram como padrão de análise de VNTRs em *M. leprae* (cada um deles tendo sido testado em pelo menos 10 estudos e 500 cepas) (**Tabela 3 e 4**). Dentre esses 17 marcadores, o GAA 21 (**Figura 11**) e GTA 9 (**Figura 12**) são os que apresentaram maior variabilidade; os marcadores 6-3a e 21-3 os que apresentaram menor variabilidade (**Figura 13**).

Tabela 3 – Marcadores do tipo VNTR já utilizados nos estudos de *M. leprae*.

Marcador	Locus	Posição genoma (TN)		Sequência repetitiva	Cópias (TN)	Número de estudos ^e	Número de cepas ^f
GTA 9	interg ML2172 / ML2173	2583814	2583840	GTA	9	28	1603
GAA 21 ^a	interg ML2344 / ML2345	2785433	2785495	GAA	21	24	1274
6-7	ML1505	1816851	1816892	GCACCT	7	22	1427
AT 17	ML2183	2597735	2597768	AT	17	22	1301
21-3	ML0058	73074	73136	GATCAAGCCAGGAA TCAAGTT	3	22	1244
AC 9	interg nadC / ML1228	1452573	1452590	AC	9	20	1579
AC 8a	ML1285	1531185	1531200	AC	8	20	1421
27-5	ML0568	687053	687177	CCGGTGGTGCCGCC TGGTGGGTTTCCC	5	20	1355
12-5	ML1182	1381663	1381722	CCTCAACAACCT	5	20	1228
23-3	interg proV / proZ	2945487	2945555	ATAATACTGTAGTG AACGACATC	3	18	1286
GGT 5	interg ML2159 / ML2160	2567251	2567265	GGT	5	18	1187
AT 15	interg ML0798 / ML0799	948935	948964	AT	15	18	1135
AC 8b	interg ML1825 / cobU	2211035	2211050	AC	8	17	1145
TA 18	ML0830	984591	984626	TA	18	17	975
18-8	ML1334	1587621	1587764	CGGCCCGCCTG GCCAGTA	8	14	829
TA 10	interg cobM / ML1451	1744091	1744110	TA	10	12	878
6-3a	rpoT	1190343	1190360	GACATC	3	11	726
AT 14	interg clpC / pabB	308814	308841	AT	14	2	49
AGA 20 ^d	interg ML2344 / ML2345	2785435	2785494	AGA	20	2	35
C 20	interg ML0237 / metG	312039	312058	C	20	2	28
CG 6	ML2472	2947291	2947302	CG	6	2	27
ML1	?	?	?	?	?	1	53
25-2	ML2418	2893422	2893471	AGAATTTACCGGCGT TCA*ATAAGAA	2	1	43
G22	ML0163 e fadE13	229625	229646	G	22	1	27

TA13 ^c	interg clpC / pabB	308815	308840	TA	13	1	27
T8 A6	interg ML0258 / ML0259	337466	337479	T,A	8,6	1	27
G10a	ML0266	347280	347289	G	10	1	27
G10b	ML0350	442993	443002	G	10	1	27
T6 N7 T8	ML0415	514180	514200	T,N,T	6,7,8	1	27
G9	ML0823	976857	976865	G	9	1	27
G12	ML0946	1116443	1116454	G	12	1	27
10-4	ML0970	1139035	1139074	TTATTAATAA	4	1	27
AGT 5a	ML1073	1237528	1237542	AGT	5	1	27
AGT 5b	ML1118	1293503	1293517	AGT	5	1	27
G11	ML1126	1309544	1309554	G	11	1	27
A9	interg ML1196 / dinX	1414666	1414674	A	9	1	27
ACC 5	cobD	1980049	1980063	ACC	5	1	27
C16 G8	interg ML1647 / valV	1987156	1987179	C,G	16,8	1	27
CA6	interg argT / ML2107	2507097	2507108	CA	6	1	27
CACCG 3	interg ML2158 / ML2159	2562391	2562405	CACCG	3	1	27
ACT 5	ML2236	2656108	2656122	ACT	5	1	27
C9	phoR	2658192	2658200	C	9	1	27
TA9	interg mmpL2 / ML2376	2844971	2844988	TA	9	1	27
15-3	interg ML2454 / ML2455	2928131	2928175	CCTTCTTGGCC GGAG	3	1	27
AT10	interg ML2476 / ML2477	2951821	2951840	AT	10	1	27
6-3 ^b	ML1918?	?	?	?	3	1	27
7-3 ^b	ML0213?	?	?	?	3	1	27
TA8	interg ML2676 / ML2677	3221617	3221632	TA	8	1	0

Legenda: Marcadores apresentados em ordem decrescente de número de publicações que os utilizaram. ^{a)} marcador também conhecido como TCC 21. ^{b)} marcadores que não puderam ter sua localização confirmada na versão atual do genoma do *M. leprae* TN (NC_002677.1). ^{c)} marcador na mesma posição do marcador AT 14, mas que não pode ser incluso nesse, devido a diferenças no sistema de contagem de cópias. ^{d)} marcador na mesma posição do marcador GAA 21, mas que não pode ser incluso nesse, devido a diferenças no sistema de contagem de cópias. ^{e)} Número de estudos que genotiparam o marcador dentre os estudos revisados sistematicamente. ^{f)} Número de cepas testadas dos estudos revisados sistematicamente, cujos resultados foram obtidos e não foram dúbios. interg) marcador em região intergenica dentre os dois genes que seguem. TN) Valor referente a cepa *M. leprae* TN (NC_002677.1).?) Desconhecido.

Tabela 4 – Marcadores utilizados em cada estudo de *VNTRs* de *M. leprae*.

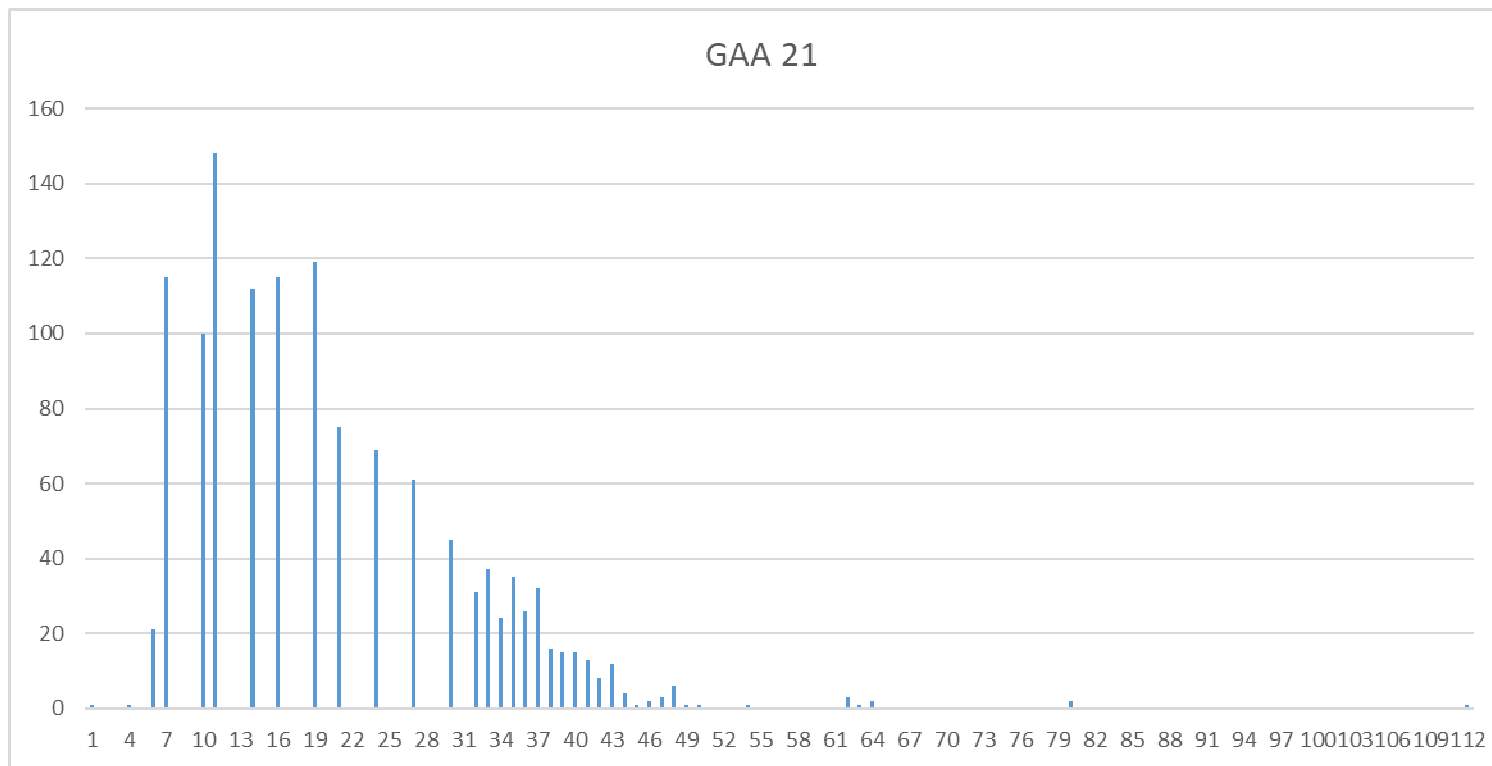
Estudo	Descrição	Amostra	Total de marcadores usados	Marcadores
Matsuoka, 2000	Por amostra	51	1	6-3a
Shin, 2000	Por amostra	34	1	GAA 21
Groathouse, 2004	Por amostra	4	9	GAA 21, AT 17, C 20, CG 6, 6-7, 12-5, 18-8, 21-3, 27-5
Truman, 2004	Por amostra	37	4	GAA 2, AT 17, TA 18, GTA 9
Young, 2004	Por amostra	77	3	GAA 21, GTA 9, AT 15
Monot, 2005	Por amostra	7	7	6-3a, GAA 21, AT 17, TA 18, GTA 9, AT 14, AT 15
Zhang, 2005	Por amostra	27	44	6-3a, AT 17, TA 18, GTA 9, C 20, CG 6, 6-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, TA 10, AC 8b, T8 A6, T6 7 T8, A9, G9, C9, G10a, G10b, G11, G12, C16 G8, G22, CA6, TA8, TA9, TA13, AT10, ACC 5, AGT 5a, AGT 5b, ACT 5, AGA 20, CACCG 3, 6-3b, 7-3, 10-4, 15-3
Taylor, 2006	Por amostra	6	3	GAA 21, GTA 9, 21-3
Weng, 2007	Só freq.	69	7	GAA 21, AT 17, TA 18, GTA 9, 6-7, AC 9, AC 8a
Monot, 2008	Só freq.	42	5	GAA 21, AT 17, GTA 9, AT 14, AT 15
Young, 2008	Só freq.	43	7	GAA 21, GTA 9, 6-7, 12-5, 21-3, 23-3, 25-2
Cardona-Castro, 2009	Por amostra	38	15	GAA 21, AT 17, TA 18, GTA 9, 6-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, AC 8b
Fontes, 2009	Por amostra	89	16	GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, TA 10, AC 8b
Gillis, 2009	Por amostra	4	13	GAA 21, AT 17, GTA 9, 6-7, 12-5, 21-3, 27-5, AC 9, AC 8a, 23-3, GGT 5, TA 10, AC 8b
Kimura, 2009	Por amostra	53	17	6-3a, GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, TA 10, AC 8b
Matsuoka, 2009	Só freq.	4	1	6-3a
Sakamuri, 2009-cont	Por amostra	85	17	6-3a, GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, TA 10, AC 8b
Sakamuri, 2009-pop	Por amostra	48	15	GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, AC 8b
Shinde, 2009	Por amostra	80	17	6-3a, GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, ML1, TA 10

Srisungnam, 2009	Por amostra	97	14	GAA 21, AT 17, TA 18, GTA 9, 6-7, 12-5, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, AC 8b
Xing, 2009	Por amostra	20	16	6-3a, GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, AC 8b
da Silva Rocha, 2011	Por amostra	16	4	GAA 21, AT 17, TA 18, GTA 9
Li, 2011	Por amostra	16	16	GAA 21, AT 17, TA 18, GTA 9, 6-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3, AT 15, GGT 5, TA 10, AC 8b
Taylor, 2011	Por amostra	8	3	GTA 9, 21-3
Truman, 2011	Por amostra	88	10	GTA 9, 6-7, 12-521-3, 27-5, AC 9, AC 8a, 23-3, GGT 5, AC 8b
Weng, 2011	Por amostra	130	7	AT 17, GTA 9, 6-7, AC 9, AC 8a, AT 15, TA 10
Fontes, 2012	Por amostra	78	16	GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3AT 15, GGT 5TA 10, AC 8b
Kuruwa, 2012	Por amostra	48	15	6-3a, GAA 21, AT 17GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8aAT 15, GGT 5TA 10, AC 8b
Cardona-Castro, 2013	Só freq.	0	2	12-5, 27-5
Weng, 2013	Por amostra	191	16	6-3a, GAA 21, AT 17, TA 18, GTA 96-718-8, 21-3, 27-5, AC 9, AC 8a, 23-3AT 15, GGT 5TA 10, AC 8b
Lavania, 2015	Só freq.	70	10	GTA 96-7, 12-521-3, 27-5, AC 9, AC 8a, 23-3GGT 5AC 8b
Fontes, 2017	Só freq.	159	17	6-3a, GAA 21, AT 17, TA 18, GTA 96-7, 12-5, 18-8, 21-3, 27-5, AC 9, AC 8a, 23-3AT 15, GGT 5TA 10, AC 8b
Bonnar, 2018	Por amostra	2	10	GTA 96-7, 12-521-3, 27-5, AC 9, AC 8a, 23-3GGT 5AC 8b
TOTAL		1719		

Legenda: Por amostra) resultado apresentado individualmente para cada amostra analisada; Por freq.) resultado apresentado numa tabela de frequência do número de cópias encontrado para cada marcador.

Fontes: Matsuoka, 2000 (92); Shin, 2000 (93); Groathouse, 2004 (94); Truman, 2004 (95); Zhang, 2005 (96); Taylor, 2006 (97); Weng, 2007 (98); Monot, 2008 (99); Young, 2008 (100); Cardona-Castro, 2009 (101); Fontes, 2009 (102); Kimura, 2009 (103); Sakamuri, 2009 (104); Shinde, 2009 (105); Srisungnam, 2009 (106); Xing, 2009 (107); da Silva Rocha, 2011 (108); Li, 2011 (109); Weng, 2011 (110); Weng, 2013 (111); Fontes, 2017 (112)

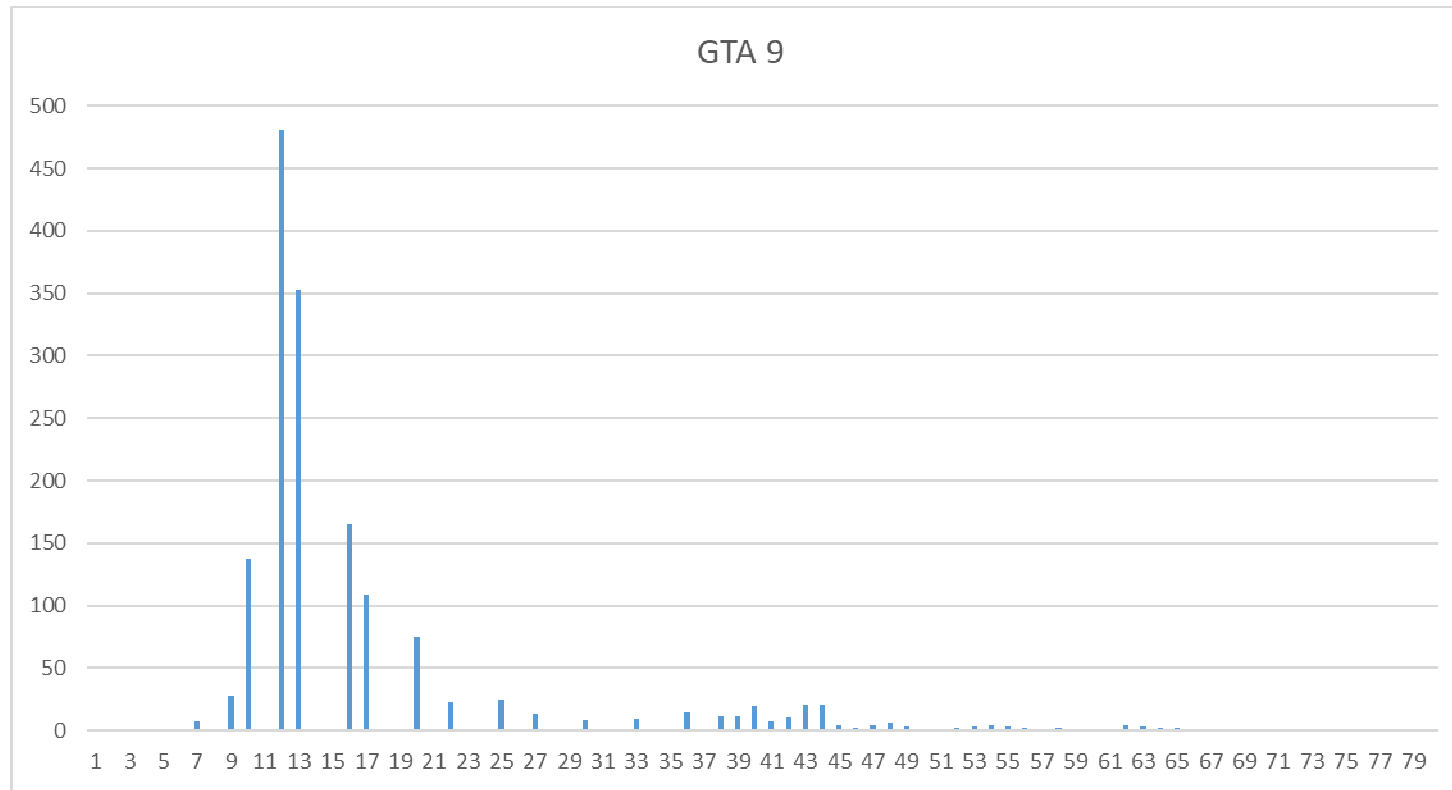
Figura 11 – Variabilidade do marcador GAA 21 em estudos de VNTRs de *M. leprae*



Legenda: O eixo x representa o número de repetições e o eixo y representa o número de amostras relatadas nos artigos.

Fonte: A autora (2018).

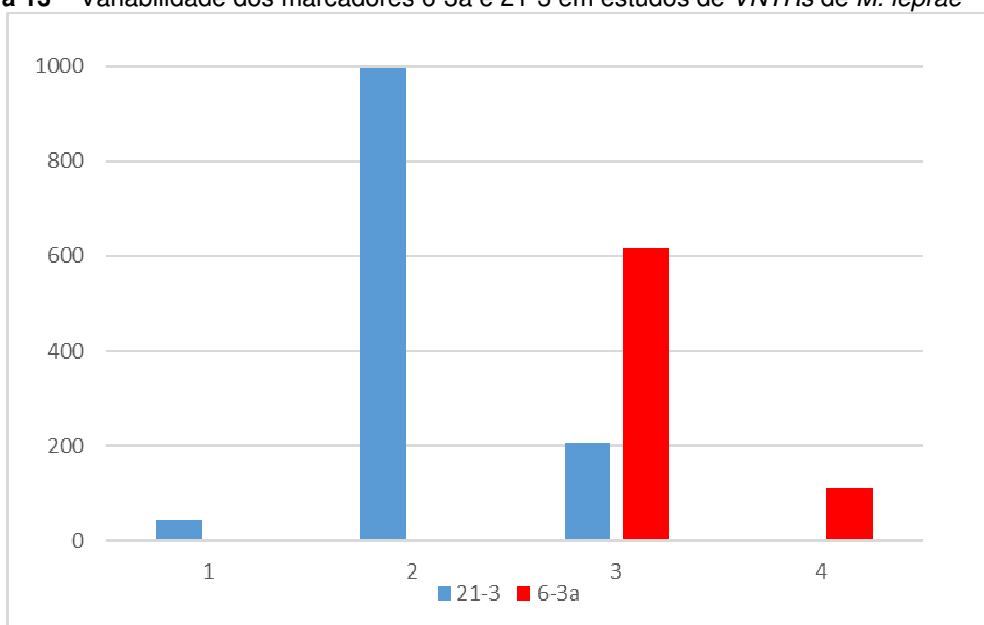
Figura 12 – Variabilidade do marcador GTA 9 em estudos de VNTRs de *M. leprae*



Legenda: O eixo x representa o número de repetições e o eixo y representa o número de amostras relatadas nos artigos.

Fonte: A autora (2018).

Figura 13 – Variabilidade dos marcadores 6-3a e 21-3 em estudos de VNTRs de *M. leprae*



Legenda: A cor azul representa a variabilidade do marcador 21-3 e a cor vermelha representa a variabilidade do marcador 6-3^a. O eixo x representa o número de repetições e o eixo y representa o número de amostras relatadas nos artigos.

Fonte: A autora (2018).

1.1.6.2. Tipagem por *SNP*

Em 2005 Monot e col. (113) avaliaram diversas cepas de *M. leprae* e concluíram que eles eram praticamente clonais. Neste estudo, 3 SNPs (*Single nucleotide polymorphism* ou Polimorfismo de nucleotídeo único) informativos foram localizados e testados em 175 cepas de 21 países de todos os cinco continentes. Os resultados indicaram que a combinação de marcadores se limitavam a apenas em 4 possibilidades, ou “*SNP types*”, numerados de 1 a 4 (**Tabela 5**). Foi ainda constatada uma correlação entre o *SNP type* e a origem geográfica das amostras, sendo o tipo 1 predominante na Ásia, Pacífico e no leste da África; o tipo 2 foi encontrado na Etiópia, Malawi, Nepal, Norte da Índia e Nova Caledónia; o tipo 3 estava relacionado à Europa, Norte da África e Américas; e o tipo 4 estava no oeste da África e no Caribe. Em 2009 (114), graças à disponibilidade de dados de genoma completo de outras cepas, foi possível identificar 78 SNPs, 2 inserções ou deleções (InDels) e 4 homopolímeros (HPT) informativos que possibilitaram que os quatro *SNP types* originais fossem subdivididos em 16 subtipos denominados de A a P (**Tabela 6**), e que também mostram grande correlação geográfica.

Tabela 5 – Tipagem por *SNP* de *M. leprae* em 4 tipos.

Marcador	Posição	Tipos			
		1	2	3	4
<i>SNP</i>	14.676	C			T
<i>SNP</i>	1.642.875	G	T		
<i>SNP</i>	2.935.685	A		C	

Fonte: Adaptado de (113)

Tabela 6 – Tipagem por *SNP* de *M. leprae* em 16 subtipos.

Marcador	Posição	Subtipos																
		1A	1B	1C	1D	2E	2F	2G	2H	3I	3J	3K	3L	3M	4N	4O	4P	
<i>SNP</i>	8.453	C																
<i>SNP</i>	298.572																	
<i>SNP</i>	347.385																	
<i>SNP</i>	383.599																	
<i>SNP</i>	481.476																	
<i>SNP</i>	485.138																	
<i>SNP</i>	494.674																	
<i>SNP</i>	508.481																	
<i>SNP</i>	694.090																	
<i>SNP</i>	890.453																	
<i>SNP</i>	904.824										G							
<i>SNP</i>	1.087.397																	
<i>SNP</i>	1.257.185																	
<i>SNP</i>	1.324.009																	
<i>SNP</i>	1.348.426																	
<i>SNP</i>	1.579.109																	
<i>SNP</i>	1.614.069																	
<i>SNP</i>	2.553.176																	
<i>SNP</i>	2.706.236																	
<i>SNP</i>	2.818.521																	
<i>SNP</i>	3.016.175																	
HPT	1.414.666																	
<i>SNP</i>	313.361	A																
<i>SNP</i>	840.584																	
<i>SNP</i>	1.143.423										G							
<i>SNP</i>	1.155.582																	
<i>SNP</i>	1.643.162																	
<i>SNP</i>	2.174.865																	
<i>SNP</i>	61.425		A								G							
<i>SNP</i>	832.152	T																
<i>SNP</i>	972.005											C						
<i>SNP</i>	1.339.813																	

SNP	1.605.956		
SNP	1.378		
SNP	2.011.747		
SNP	2.459.766		
SNP	3.102.778	A	C
SNP	1.104.232	C	G
SNP	2.751.783	A	G
SNP	73		
SNP	100.574		
SNP	330.125		
SNP	1.926.696	A	G
SNP	2.015.320		
SNP	2.462		
SNP	1.133.492		
SNP	1.295.192	T	G
SNP	2.235.048		
SNP	2.312.059	C	G
SNP	413.902		
SNP	642.131		
SNP	807.613	G	A
SNP	2.131.753		
SNP	3.090.722		
SNP	20.910		
SNP	527.964		
SNP	591.857		
SNP	733.811	G	A
SNP	2.815.502		
SNP	3.267.975		
SNP	1.079.902		
SNP	1.214.326		
SNP	1.313.204		
SNP	1.524.667		
SNP	1.575.600		
SNP	1.947.401		
SNP	2.096.406		
SNP	2.132.052		
SNP	2.239.468	G	A
SNP	22.719		
SNP	2.347.320		
SNP	2.414.869		
SNP	2.423.927		
SNP	365.373		
SNP	379.804		
SNP	467.512		
SNP	508.986		

SNP	537.330		
SNP	RNC1		
HPT	3.244.472		
HPT	3.041.556		
HPT	741.133		
Ins	978.589	-	T
Del	1.476.519	T	-

Legenda: Classificação baseada em *SNP* (polimorfismo de nucleotídeo único), *HPT* (homopolímero), *Ins* (inserções), *Del* (deleções). As bases mostradas representam apenas um alelo dentre as duas possibilidades de cada marcador.

Fonte: Adaptado de (114)

1.1.6.3. Estudos de genoma

Mais recentemente, o desenvolvimento de novas tecnologias permitiu que, além dos estudos de marcadores e genes específicos, outros aspectos moleculares do *M. leprae* fossem estudados. Em 2001, Stuart Cole e col. sequenciaram pela primeira vez o genoma completo do *M. leprae* (cepa TN) formado por 3.268.203pb num único DNA circular e sem plasmídeos, com um conteúdo médio de G+C de 57,8%. Esses valores são bem menores que os já descritos para o genoma da maioria das outras micobactérias (**Tabela 7**), por exemplo, o *M. tuberculosis*, que é composto de 4.411.532pb e 65,6% de conteúdo médio G+C (115).

Tabela 7 - Comparação do genoma de algumas micobactérias

Bactéria	<i>M. leprae</i>	<i>M. tuberculosis</i>	<i>M. bovis</i>	<i>M. africanum</i>	<i>M. lepromatosis</i>
Cepa	TN	H37Rv	1595	GM041182	Mx1-22A
Tamanho do genoma (Mb)	3,27	4,41	4,35	4,39	3,21
Conteúdo G+C (%)	57,8	65,6	65,6	65,6	57,9
Genes	2 770	4 008	4 220	4 243	2 856
Proteínas	1 605	3 906	4 026	4 045	1 477
Pseudogenes	1 115	30	143	147	1 331

Legenda: Mb) Mega bases

Fonte: (32, 116-119)

Além disso, a anotação das sequências de genoma e proteoma do *M. leprae* levou à surpreendente constatação que, de 2770 genes, apenas 1605 (57,94%) codificam proteínas, os demais se compõem de pseudogenes (quadros de leitura inativos com contrapartidas funcionais em outras micobactérias) ou sequências regulatórias (115, 120).

Assumindo que o genoma do *M. leprae* já tenha tido um tamanho similar ao da maioria das micobactérias ($\approx 4,4$ Mb), houve uma extensa redução e rearranjo durante a sua evolução (121). Devido a isso, muitas das vias comuns nas outras micobactérias foram reduzidas no *M. leprae*. No geral, as vias de degradação foram mais reduzidas que as vias de síntese. Por exemplo, foram encontradas mutações em genes envolvidos na regulação, desintoxicação, reparo de DNA, transporte e efluxo de metabólitos. Dessa forma, os genes responsáveis por transmissão, alojamento e sobrevivência no hospedeiro foram preservados, enquanto boa parte dos genes que podem ser compensados pelos parasitismos foi descartada. Esse processo é denominado "redução evolutiva" e tem sido observado em diversos patógenos humanos intracelulares obrigatórios, principalmente devido à inativação de genes cuja função não é mais necessária nesse nicho altamente especializado. Essa perda de diversas vias metabólicas também explica, em grande parte, a impossibilidade de cultivar o *M. leprae in vitro* (115, 120, 122, 123).

Mais recentemente, em 2018, foi realizado um amplo estudo de 154 genomas de *M. leprae* advindos de 25 países (124) que constatou que:

- (i) A diferença média entre os genomas analisados era de 114 *SNPs*;
- (ii) Alguns polimorfismos encontrados sofreram seleção positiva durante a evolução, o que pode ser consequência da pressão do sistema imune humano, da PQT ou de outras forças;
- (iii) A análise filogenética indicou que o ancestral do *M. leprae* se originou na Eurásia, provavelmente no Leste Asiático;
- (iv) Mutações no gene *nth* (que codifica a Endonuclease III) prejudicam o funcionamento sistema de reparo bacteriano levando a uma maior variabilidade das cepas. Dessa forma é possível que a ausência do *nth* favoreça o surgimento de mutações de resistência. O fato que todas as cepas que tinham mutação no *nth* também tinham resistência à antimicrobianos,

corroborar com essa possibilidade. Com isso esse gene passou a ser um candidato a marcador de risco de desenvolvimento de resistência;

- (v) Um alto número de casos de resistência a quinolonas em pacientes que estavam manifestando a hanseníase pela primeira vez (resistência primária);
- (vi) Foram relatadas mutações na *gyrB* (codificante da subunidade beta da DNA-girase) suspeitas de causar resistência a antimicrobianos no *M. leprae*;
- (vii) Foi observada uma frequência alta de mutações no *fadD9* e *ribD*, especialmente nas cepas com resistência a antibióticos. É possível que as mutações no gene *ribD* estejam relacionadas à resistência aos antibióticos derivados do Ácido 4-amino-salicílico. A função do gene *fadD9* ainda não é completamente elucidada de forma que seu papel na resistência antimicrobiana (especialmente a clofazemina) ainda deve ser investigada.

1.2. REVISÃO DE LITERATURA – PARTE 2 – SEQUENCIAMENTO DE GENOMA COMPLETO

1.2.1. Sequenciamento de Segunda Geração

Desde a década de 70, quando foi proposto o sequenciamento enzimático de DNA, houve uma evolução constante nas técnicas de sequenciamento e, nesse processo, diversas plataformas surgiram e várias delas se tornaram obsoletas. Essas plataformas são agrupadas em gerações, conforme suas características tecnológicas em geral e pela época em que foram desenvolvidas. A primeira geração de sequenciadores, por exemplo, é baseada na técnica de *Sanger*, que utiliza a eletroforese de *amplicons*, contendo di-deoxinucleotídeos para determinar uma única sequência (125, 126).

Somente na década de 2000 foi desenvolvida a segunda geração de sequenciadores (em inglês, *Next-Generation Sequencing – NGS*). Todas as plataformas dessa geração têm em comum a leitura paralela de diversos *amplicons* espacialmente separados e, conseqüentemente, um maior *throughput* de dados. Para todas elas, o processo de sequenciamento se compõe de três etapas principais: (i) o preparo de biblioteca; (ii) o preparo de *template*; e (iii) o sequenciamento propriamente dito. Cada plataforma, porém, utiliza diferentes métodos de detecção das bases, o que faz com que cada uma tenha suas peculiaridades, vantagens e desvantagens específicas (125, 126).

Neste trabalho, foram utilizadas ferramentas de sequenciamento de segunda geração por meio de semicondutor de íons (*Ion PGM*). Devido às especificidades dessa tecnologia, será apresentada aqui uma breve descrição de seus conceitos básicos e de seu funcionamento.

1.2.1.1. Condições pré-sequenciamento

É importante ressaltar que, em métodos de biologia molecular, a qualidade do material genético utilizado no processo é determinante na qualidade do resultado final. Nesse aspecto, o sequenciamento pode ser especialmente exigente. A qualidade do DNA usualmente é medida a partir da relação entre a absorvância de 260nm (faixa do comprimento de onda que os nucleotídeos, DNA, RNA etc. absorvem) e a absorvância de 280nm (faixa do comprimento de onda que as proteínas e outros contaminantes absorvem). O DNA ideal deve ter a relação 260/280 próxima a 1,8 para ser considerado livre de contaminantes. Devido às exigências do sequenciamento, também é utilizada uma segunda relação entre a absorvância 260nm e a absorvância de 230nm, que indica a contaminação da amostra com outros produtos geralmente advindos do processo de extração, como carboidratos, fenol, guanina, glicogênio, sais e EDTA. Quando puro, o DNA apresenta razão 260/230 próxima a 2,0 e, portanto, esse é o recomendado para fins de sequenciamento (125-127).

Também é importante ressaltar que o sequenciamento de DNA em plataformas de segunda geração não necessariamente utiliza *primers* para amplificação de sequências alvos. Isso facilita o sequenciamento de grandes extensões de DNA, como o genoma bacteriano completo; porém, sem a especificidade dos *primers*, todo material genético presente na amostra também será sequenciado. Sendo assim, o método é extremamente sensível à presença de DNA contaminante. No contexto de sequenciamento de genoma bacteriano, por exemplo, se o DNA for extraído a partir de material biológico do hospedeiro, o genoma deste poderá ser sequenciado juntamente com a sequência alvo.

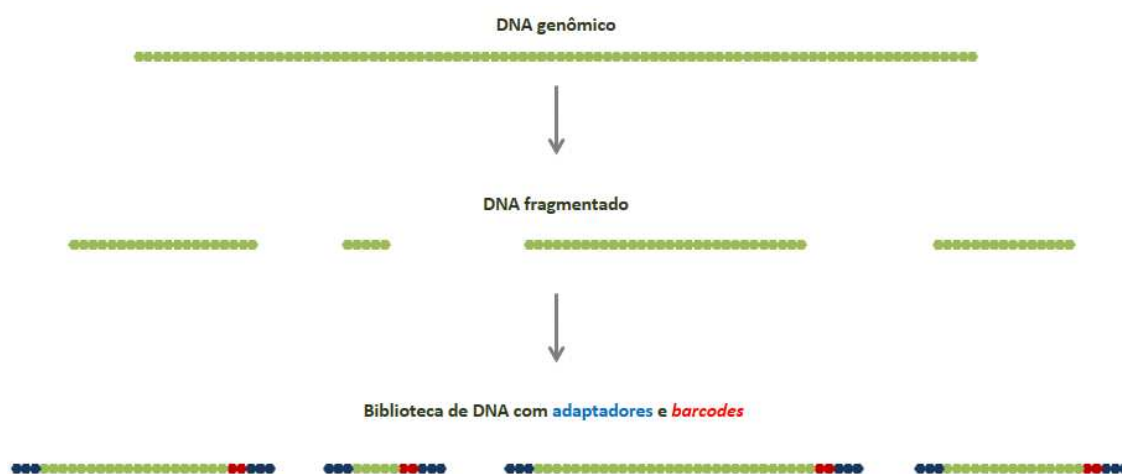
1.2.1.2. Preparo de Biblioteca

O preparo de biblioteca é a etapa de preparação do material genético que será sequenciado, sendo a etapa menos automatizada do processo. O objetivo desta etapa é obter fragmentos pequenos de nucleotídeos, de tamanho uniforme, ligados a adaptadores. Para isso, o protocolo varia conforme o material genético utilizado e a química de sequenciamento escolhida. Para sequenciamento de DNA

genômico bacteriano por tecnologia *Ion Torrent*, o preparo de biblioteca se compõe de quatro principais sub etapas, detalhadas a seguir:

- **Fragmentação:** O DNA genômico íntegro é muito extenso para ser sequenciado e deve passar por fragmentação enzimática para adequá-lo ao sequenciamento. A enzima utilizada nessa etapa é sensível à presença de EDTA, razão pela qual a presença deste na amostra pode comprometer o processo. Ao final da fragmentação, é realizada uma purificação com *beads* (**Figura 9**).
- **Ligação de Adaptadores e *Barcodes*:** Nessa etapa, os fragmentos são ligados a duas sequências curtas de DNA, conhecidas como adaptadores e *barcodes*. Os adaptadores uniformizam as extremidades de todos os fragmentos, o que possibilita a sua amplificação e preparo de *template* com um iniciador universal. Os *barcodes* são sequências que servem para identificar as amostras de origem, no caso de um sequenciamento contendo mais de uma amostra simultaneamente. Ao final do processo de ligação, é realizada uma purificação com *beads* (**Figura 14**).

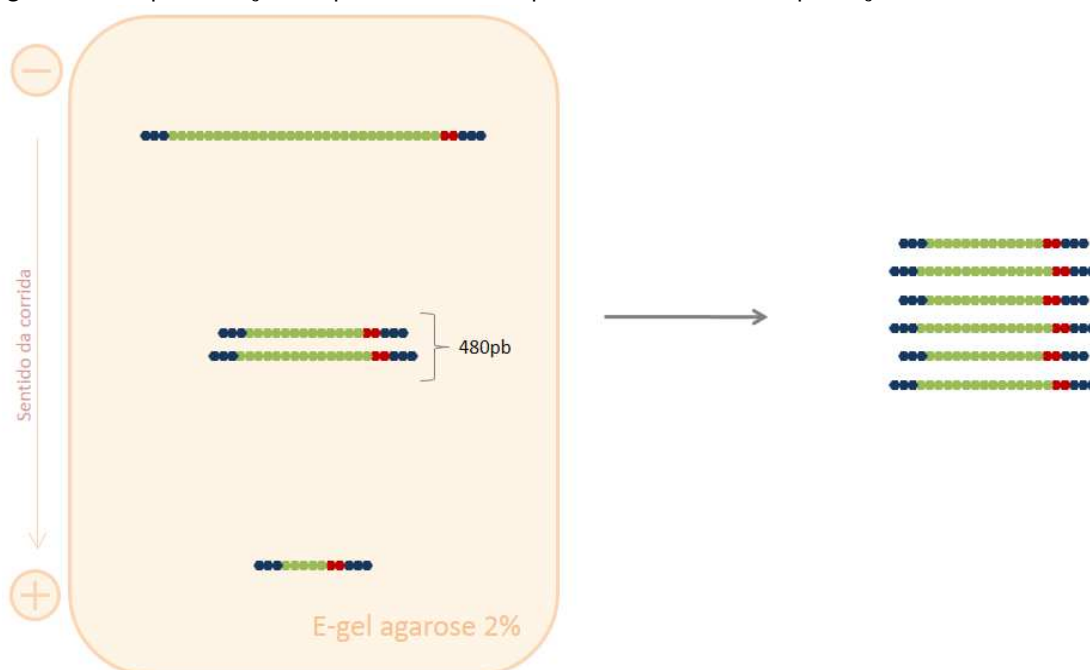
Figura 14 - Representação esquemática das etapas de fragmentação e ligação de adaptadores e *barcodes*



Fonte: a autora (2018).

- **Size-select:** Nessa etapa, são selecionados os fragmentos com o tamanho ideal (no caso deste estudo, próximo a 480pb). Esse processo é realizado em um gel comercial de agarose a 2% de alta sensibilidade que facilita a recuperação das amostras. A visualização desse gel é um ponto de checagem do preparo de biblioteca, sendo necessário secundariamente para definir a amplificação ou não da biblioteca, e, em alguns casos, o reinício do preparo da biblioteca, devido à perda de amostra em etapas prévias (**Figura 15**).

Figura 15 - Representação esquemática das etapas de *size-select* e amplificação da biblioteca



Fonte: a autora (2018).

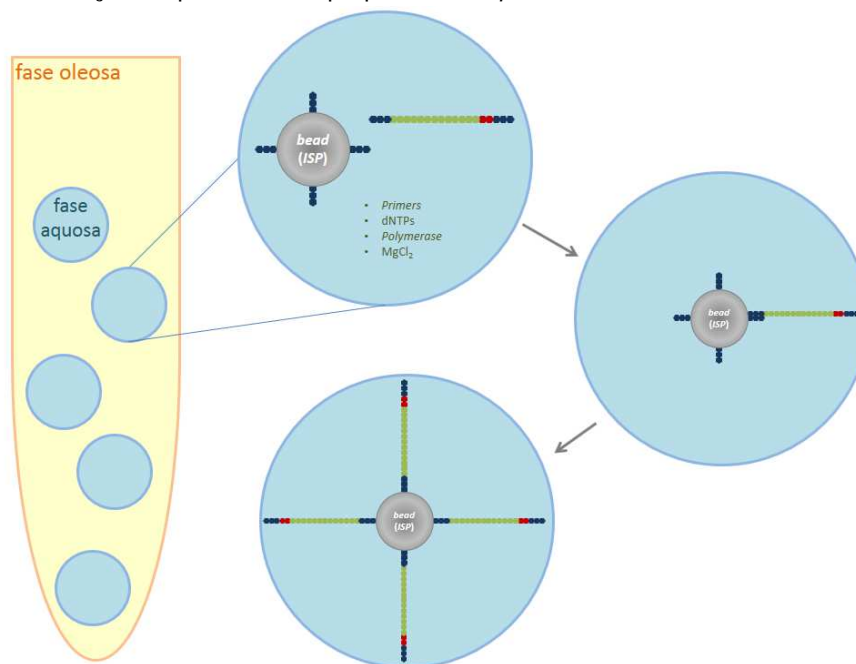
- **Amplificação:** Quando a presença de amostra é constatada no gel, a mesma deverá ser amplificada por Reação em cadeia da polimerase (*Polymerase Chain Reaction* – PCR), utilizando um iniciador universal que se liga aos adaptadores das amostras. Essa etapa de PCR sabidamente pode inserir erros nas sequências e, portanto, é realizada com uma enzima de alta fidelidade. Após a amplificação, a biblioteca é purificada com *beads* e quantificada (**Figura 15**).

1.2.1.3. Preparo de *template*

O objetivo desta etapa é produzir bibliotecas de DNA clonais, isoladas e ligadas a uma superfície. Nas plataformas com tecnologia *Ion Torrent*, o preparo de *template* é automatizado e as bibliotecas são ligadas a *beads* de sequenciamento denominadas *Ion Sphere Particles (ISP)*. Para isso, o preparo de *template* se subdivide em três etapas:

- **PCR em emulsão:** Nesse ponto, a(s) biblioteca(s) preparada(s) é(são) submetida(s) a uma PCR em emulsão, na qual micelas aquosas servem como mini-reatores em um meio oleoso. Idealmente, cada micela deverá conter uma *bead* de sequenciamento (a qual tem *primers* complementares aos adaptadores ligados à sua superfície), um único fragmento da biblioteca e os demais componentes necessários à PCR. Ao final do processo, a superfície de cada *ISP* estará coberta de cópias de um único fragmento da biblioteca (125) (**Figura 16**).

Figura 16 - Representação esquemática de preparo de *template*.



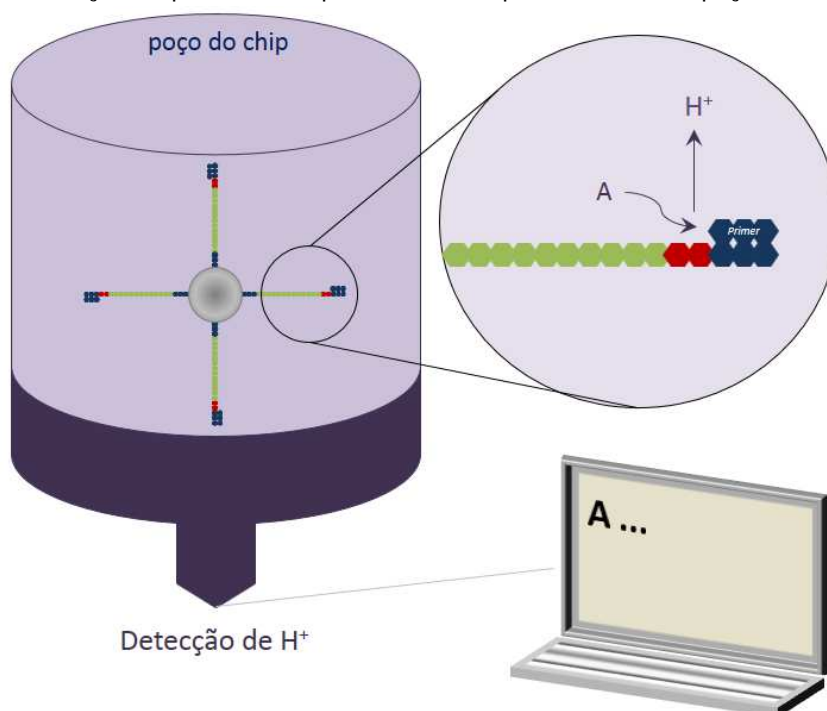
Fonte: a autora (2018).

- **Enriquecimento:** Nessa etapa, as *ISP* negativas (nas quais não houve amplificação) são descartadas e as *ISP* positivas são, portanto, concentradas.

1.2.1.4. Sequenciamento

O sequenciamento se compõe de duas etapas: o preenchimento do chip com a solução contendo as *ISP* positivas e o sequenciamento, propriamente dito. Um chip de sequenciamento é um sensor eletrônico num circuito integrado, capaz de detectar os íons de hidrogênio liberados durante a síntese de DNA pela DNA polimerase. Esse chip tem 1,2 milhão de poços, nos quais são inseridas as *ISP* para o sequenciamento paralelo de diversos fragmentos. O sequenciamento consiste, portanto, no fluxo alternado de nucleotídeos no chip (chamado *flow order*), que, quando incorporados à síntese, liberam os íons H^+ , que são detectados por alteração de pH (chamado de *flow signal*) e computados (**Figura 17**). (128).

Figura 17 - Representação esquemática do processo de sequenciamento em poço do chip



Fonte: a autora (2018).

1.2.1.5. Dados Brutos

A partir da leitura de bases de um fragmento, numa *bead*, num poço do chip, é gerada uma sequência curta, conhecida como *read*. A cada base nesta *read* é atribuído um escore de qualidade baseado na escala *Phred* (*Phred* Quality score, Q): uma escala logarítmica baseada em probabilidade de erro (129, 130) (**Tabela 8**). Ainda, o *output* da plataforma *Ion Torrent* inclui os dados de *flow order* e *flow signal*.

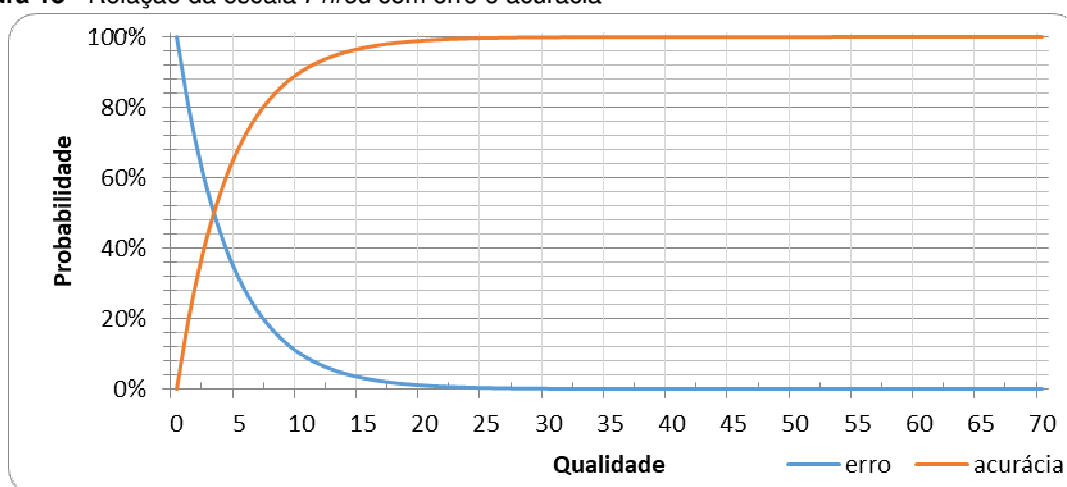
Tabela 8 - Escore *Phred* relacionado à probabilidade de erro.

Escore de qualidade <i>Phred</i> (Q)	Probabilidade de determinação errônea da base	Acurácia
10	1 em 10	90%
20	1 em 100	99%
30	1 em 1000	99,90%
40	1 em 10.000	99,99%
50	1 em 100.000	99,999%
60	1 em 1.000.000	99,9999%

Fonte: a autora, adaptado de (131)

Quanto maior o escore, menor a probabilidade de erro e maior a acurácia da base sequenciada. Como pode ser observado na **Figura 18**, até o valor de Q20 há uma diferença considerável quanto à qualidade da base, mas acima desse valor essa diferença fica pequena. Por isso, o valor de Q20 é frequentemente usado como *cut-off*.

Figura 18 - Relação da escala *Phred* com erro e acurácia



Legenda: No eixo X está a escala *Phred* e no eixo Y probabilidade em porcentagem. A linha azul representa a probabilidade de erro e a vermelha representa a acurácia (1-erro).

Fonte: a autora (2018).

1.2.1.6. Análise de Dados

1.2.1.6.1. Tipos de *software*

Existe hoje uma grande variedade de programas disponíveis para realizar análises de dados advindos de sequenciamento. Eles podem ser classificados de maneira geral de diversas formas:

- Por critérios financeiros (pagos ou gratuitos);
- Adequados ao uso em diferentes sistemas operacionais (Windows, Linux, Mac etc);
- Por modo de uso (com interface e/ou linha de comando);

Neste estudo, foram utilizados programas gratuitos voltados para uso na plataforma Linux. Tendo em vista que os dados de sequenciamento tem peculiaridades relativas a tecnologia utilizada, foi dada preferência a programas que levassem em consideração as características da plataforma *Ion Torrent*. Nesse contexto, a maioria dos programas utilizados funcionava por linha de comando. Usualmente, a linha de comando se inicia com o programa a ser executado, seguindo-se - caso necessário - a sub-função a ser utilizada e as opções (ou parâmetros) do programa. Geralmente, as opções são identificadas pelo programa por "-" seguido de uma letra ou "--" seguido de uma palavra. Usualmente os programas têm minimamente um arquivo de entrada (*input*) e um arquivo de saída (*output*) (**Quadro 5**).

Quadro 5 – Exemplo de linha de comando

```
programa subfuncao --opcao1 valor1 --opcao2 valor2 arquivo.entrada arquivo.saida
```

1.2.1.6.2. Alinhamento

Na etapa de alinhamento as *reads* são alinhadas e comparadas a uma sequência referência. O número de vezes que uma base é computada em diversas *reads* é conhecido como cobertura da base. A média de cobertura de todas as bases sequenciadas e a uniformidade dessa cobertura são alguns dos critérios de

qualidade um sequenciamento. O *Alignment Quality (AQ)*, que em alguns casos também é baseado na escala *Phred*, também é uma medida de qualidade de alinhamento.

Para genomas pequenos e/ou de procariotos as regiões de pouca ou nenhuma cobertura são conhecidas como *gap*. O trecho de bases alinhadas entre um *gap* e outro é denominado *contig* (**Figura 19**).

Figura 19 - Representação esquemática do alinhamento das *reads* a uma sequência de referência



Fonte: A autora (2018).

Conseqüentemente um genoma completamente sequenciado, sem *gaps*, seria composto de um único *contig*. Para melhor mensurar os *contigs* são utilizadas as medidas tais como “N50” e “comprimento N50”. O “N50” representa o menor número de *contigs* necessários para que 50% do genoma seja representando enquanto “comprimento N50” representa o comprimento em pares de base gerado com 50% dos *contigs*. De forma similar existem outras medidas tais com N90 (para 90%) e N95 (para 95%).

1.2.1.6.2.1. Mapeamento

Mapeamento o processo de alinhamento de dados que se utiliza de uma sequência referência para a organização das *reads*. Um bom mapeamento tem grande dependência da qualidade e coerência da referência utilizada. O uso referência inadequada poderá induzir a amostra a um alinhamento errôneo. Além disso no mapeamento é difícil a localização de elementos móveis, translocações, etc.

1.2.1.6.2.2. Montagem *de novo*

A montagem *de novo* é um alinhamento sem referência, baseado apenas nas similaridades entre as *reads*. É o método utilizado principalmente quando não há um genoma de referência adequando. Um fator limitante desse método é o comprimento das *reads*. Quando muito curtas dificultarão a montagem que gerará um grande número de *gaps* e, conseqüentemente, de *contigs*. Esse método porém possibilita a descoberta de elementos móveis, translocações etc.

2. CONTEXTO E DESENHO DO ESTUDO

O Lazaropolis do Prata foi a primeira colônia agrícola de leprosos fundada no Brasil na década de 1920 (132) quando foi adotada a política de internação compulsória no país, considerada a melhor profilaxia para uma doença, até então, sem cura. Essa política seguiu até 1962 (133) quando a disponibilidade de alguns antibióticos eficazes no tratamento da doença passou a permitir a reintegração dos pacientes à sociedade.

Mais recentemente, a ex-colônia passou a ser denominada de Vila Santo Antônio do Prata e nela habitam os antigos internos e seus descendentes. Devido, em grande parte, ao estigma da doença, o fluxo migratório é praticamente nulo e a população permanece relativamente isolada. No período 2006 a 2007 toda a população da Vila foi recrutada (2005 indivíduos) pelo nosso grupo de pesquisa para um estudo epidemiológico (**ANEXO IV**). Neste estudo, foi constatado que 257 pessoas tinham ou tiveram hanseníase, o que representava uma prevalência de 12,82%. Essa taxa de prevalência foi maior que a taxa de prevalência do Brasil todo, de 0,0321% (no ano de 2007) e muito maior que a recomendação da OMS de menos de 0,01% (menos de 1 casos a cada 10.000 habitantes). Portanto a Vila era hiperendêmica para a hanseníase, o que, se considerado juntamente com o isolamento e a homogeneidade de fatores ambientais faz dessa localidade um ambiente propício para o estudo de vários aspectos da doença, um “laboratório natural”.

Assim sendo, a partir 2009 nosso grupo de pesquisa participou de um projeto colaborativo multicêntrico de larga escala (doravante denominado “Estudo-2009”) cujo objetivo foi investigar suspeitas de casos de resistência dentre os casos de recidiva de hanseníase. Constatou-se que 9 dentre 12 casos de recidiva (75%) e 4 dentre 10 novos casos (40%) tinham resistência a um ou mais antibióticos (ver item “5. Manuscrito” para maiores detalhes). Esses índices estavam muito acima das estimativas globais da OMS de 5,2% e 2,1% de resistência secundária e primária, respectivamente. Tendo em vista esses alarmantes resultados, decidiu-se submeter um projeto de estudo de *follow-up* (doravante denominado “Estudo-2013”). Essa nova proposta tinha como objetivo avaliar a extensão dos casos de hanseníase causados por *M. leprae* resistente aos componentes da PQT na Vila do Prata, bem

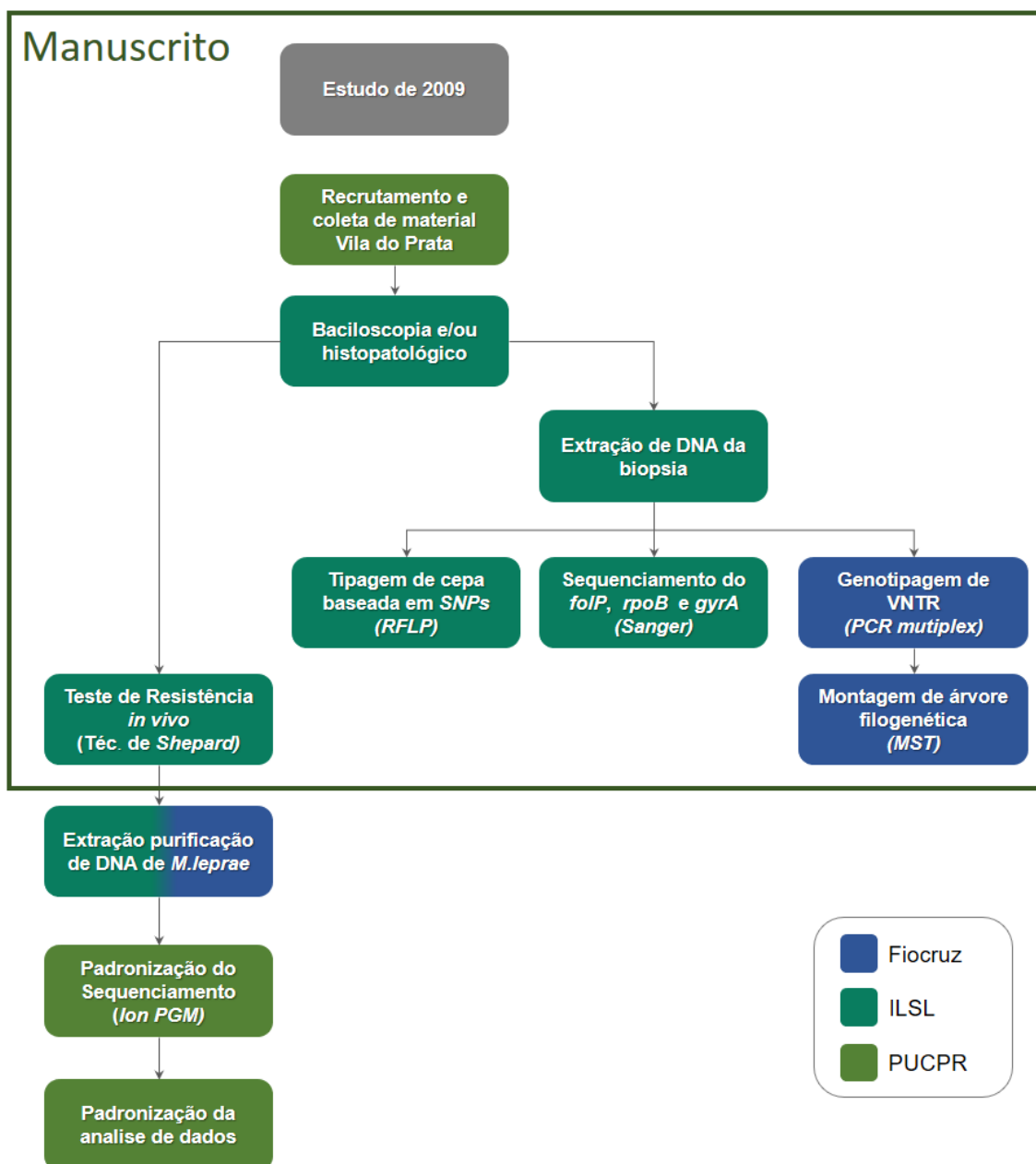
como ampliar o entendimento da dinâmica de transmissão destes isolados. É no contexto do Estudo-2013 que está inserida esta tese.

Brevemente, no âmbito do “Estudo-2013”, foram recrutados indivíduos com hanseníase da Vila do Prata, identificados através de mutirão dermatológico e busca ativa de casos em toda a Vila. Os indivíduos que aceitaram participar do estudo tiveram amostras de biópsias de lesões hansênicas coletadas. *M. leprae* isolado das biópsias foi utilizado para a determinação do perfil de resistência, por meio da técnica de inoculação em coxim plantar (técnica de Shepard) em camundongos BALB/c e nude (24). O DNA total obtido das amostras de biópsia foi submetido a (i) análise molecular de resistência através do sequenciamento (*Sanger*) de regiões específicas dos genes *folP*, *rpoB* e *gyrA*; (ii) genotipagem de *SNPs* *M. leprae*-específicos por *RFLP* (*Restriction Fragment Length Polymorphisms*) para caracterização molecular (113); (iii) análise de epidemiologia molecular utilizando repetições em *tandem* de número variável - *Variable number tandem repeat* - *VNTR* (utilizando técnica de reação em cadeia da polimerase múltipla ou *PCR multiplex*). O DNA obtido das amostras cultivadas *in vivo* foi submetido à análise comparativa de sequências de genomas completos dos *M. leprae* (sequenciamento no *Ion PGM™*).

A **Figura 20** apresenta o desenho de estudo do “Estudo-2013” de forma esquemática. Importante destacar que a coordenação deste foi baseada na Pontifícia Universidade Católica do Paraná (PUCPR) e foi realizado em parceria com as seguintes instituições:

- Instituto Lauro de Souza Lima (ILSL) – Bauru (SP);
- Universidade Federal do Pará (UFPA) - Belém (PA);
- Fundação Oswaldo Cruz (FIOCRUZ) - Rio de Janeiro (RJ).

Figura 20 - Esquema do desenho de estudo do projeto global no qual este estudo está inserido.



Legenda: As cores das etapas correspondem ao local de sua realização. O recrutamento populacional teve participação de todas as entidades envolvidas no projeto global, mas está em verde pois a sua organização e coordenação foi feita pela PUCPR.

O quadrado “Manuscript” destaca as etapas descritas no manuscrito em preparação (ver item 5. Manuscrito).

RFLP = *Restriction Fragment Length Polymorphisms* (Polimorfismo no Comprimento do Fragmento de Restrição); *Ion PGM™* = *Ion Torrent Personal Genome Machine System*; *VNTR* = *Variable number tandem repeat* (Repetição em tandem de número variável); *PCR* = *Polymerase Chain Reaction* (Reação em cadeia da polimerase); *MST* – *Minimum Spanning Tree*

Fonte: A autora (2018).

3. JUSTIFICATIVA

Há quatro fatores principais que motivaram o desenvolvimento deste estudo:

- I. A existência já comprovada de índices alarmantes de resistência à PQT entre os isolados da Vila Santo Antônio do Prata. Sob o ponto de vista da saúde pública e para fins epidemiológicos, fez-se necessário estender essa análise, através de busca ativa e sistemática de novos casos de hanseníase causada por isolados resistentes de *M. leprae* em toda a Vila Santo Antônio do Prata;
- II. A demonstração da dinâmica de surgimento e transmissão de cepas resistentes de *M. leprae* em uma população modelo será um avanço no entendimento dos mecanismos de seleção e emergência de resistência do *M. leprae* ao tratamento;
- III. A oportunidade de se aliar pesquisa básica com ações educacionais e de promoção de saúde, com impacto direto sobre o controle da endemia e a qualidade de vida da população alvo do estudo;
- IV. Os resultados encontrados poderão levar à revisão dos protocolos de tratamento em outros sítios do Brasil.

4. OBJETIVOS

4.1. OBJETIVO GERAL

Descrever a exata extensão do problema de resistência primária e secundária do *M. leprae* aos componentes da poliquimioterapia em uma população hiperendêmica isolada do norte do Brasil, bem como a natureza molecular dos isolados bacterianos obtidos, aplicando uma estratégia sistemática de recrutamento da população em estudo.

4.2. OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste estudo são:

- I. Produzir um levantamento de base populacional dos casos de hanseníase na Vila do Prata através de busca ativa e atividade intensiva de atendimento dermatológico ambulatorial;
- II. Obter o perfil de resistência (*in vivo* e molecular) das cepas obtidas;
- III. Obter o perfil de variabilidade molecular (*VNTRs* e tipagem por *SNP*) das cepas obtidas;
- IV. Cruzar os dados dos pacientes com os perfis de resistência e de variabilidade molecular de forma a observar a transmissão das cepas;
- V. Padronizar o protocolo de sequenciamento de genoma e análise primária de dados dos genomas das cepas *M. leprae* em plataforma *Ion PGM™ (Personal Genome Machine)*.

Emergence and Transmission of Drug-/Multidrug-resistant *Mycobacterium leprae* in a Former Leprosy Colony in the Brazilian Amazon

Patricia S. Rosa,^{1,a} Helena R. S. D'Espindula,^{2,a} Ana C. L. Melo,³ Amanda N. B. Fontes,⁴ Amanda J. Finardi,¹ Andréa F. F. Belone,¹ Beatriz G. C. Sartori,¹ Carla A. A. Pires,⁵ Cleverton T. Soares,¹ Flávio B. Marques,¹ Francisco J. D. Branco,³ Ida M. F. D. Baptista,¹ Lázara M. Trino,¹ Luciana R. V. Fachin,¹ Marília B. Xavier,^{5,b} Marcos C. Floriano,⁷ Somei Ura,¹ Suzana M. Diório,¹ Wladimir F. B. Delanina,¹ Milton O. Moraes,⁴ Marcos C. L. Virmond,¹ Philip N. Suflys,⁴ and Marcelo T. Mira²

¹Division of Research and Education, Instituto Lauro de Souza Lima, Bauru, São Paulo, ²Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, ³Centro de Referência Nacional em Dermatologia Sanitária Dona Libânia, Fortaleza, Ceará, ⁴Laboratory of Molecular Biology Applied in Mycobacteria, Fundação Oswaldo Cruz, Rio de Janeiro, ⁵Core of Tropical Diseases, Universidade Federal do Pará, Belém, ⁶Center for Biological and Health Sciences, Universidade do Estado do Pará, Belém, and ⁷Department of Dermatology, Universidade Federal de São Paulo, Brazil

Background. Leprosy has been treated with multidrug therapy, which has been distributed for free across the globe and regarded as highly efficient. However, the impossibility of growing *Mycobacterium leprae* in axenic media has historically impaired assessments of *M. leprae* resistance, a parameter only recently detectable through molecular methods.

Methods. A systematic, population-based search for *M. leprae* resistance in suspected leprosy relapse cases and contacts was performed in Prata Village, an isolated, hyperendemic, former leprosy colony located in the Brazilian Amazon. Results led to an extended active search involving the entire Prata population. Confirmed leprosy cases were investigated for bacterial resistance using a combination of in vivo testing and direct sequencing of resistance genes *folP1*, *rpoB*, and *gyrA*. A molecular epidemiology analysis was performed using data from 17 variable number tandem repeats (VNTR).

Results. *Mycobacterium leprae* was obtained from biopsies of 37 leprosy cases (18 relapses and 19 new cases): 16 (43.24%) displayed drug-resistance variants. Multidrug resistance to rifampicin and dapsone was observed in 8 relapses and 4 new cases. Single resistance to rifampicin was detected in 1 new case. Resistance to dapsone was present in 2 relapses and 1 new case. Combined molecular resistance and VNTR data revealed evidence of intra-familial primary transmission of resistant *M. leprae*.

Conclusions. A comprehensive, population-based systematic approach to investigate *M. leprae* resistance in a unique population revealed an alarming scenario of the emergence and transmission of resistant strains. These findings may be used for the development of new strategies for surveillance of drug resistance in other populations.

Keywords. leprosy; *M. leprae*; multidrug resistance; primary resistance; transmission.

Over the past 20 years, the global number of new cases of leprosy has remained stable, irrespective of available effective treatment, suggesting that better prophylactic and diagnostic tools are necessary to improve disease control and achieve reduced incidence rates; in this scenario, maintaining high therapeutic efficacy is of critical importance [1].

The first cases of *Mycobacterium leprae* secondary and primary drug resistance (DR) to dapsone (DDS) were reported in 1964 [2] and 1977 [3], respectively; the first case of resistance to rifampicin (RIF) was described in 1976 [4]. In 1981, the World Health Organization (WHO) recommend multidrug therapy (MDT) against leprosy, composed by DDS, RIF, and clofazimine [5]; in 1996, the first case of primary multidrug resistance (MDR) was reported [6], followed by the description of the first case of MDR to DDS, RIF, and ofloxacin [7]. Clofazimine resistance has hardly ever been detected in *M. leprae*, possibly due to the absence of a homolog for the Rv0678 efflux pump found in *M. tuberculosis* [6, 8, 9]

The inability of *M. leprae* to grow in artificial media has been a major limitation for the detection of DR/MDR in leprosy: the classic Shepard's method, based on bacterial growth in the mouse footpad, is a labor-intensive and time-consuming procedure [10]. In the 1990s, molecular tools became available for sequencing drug resistance-determining regions (DRDR) of *M. leprae* genes *folP1*, *rpoB*, and *gyrA*, associated with resistance

Received 2 April 2019; editorial decision 13 June 2019; accepted 28 June 2019; published online July 1, 2019.

^aP. S. R. and H. R. S. D. contributed equally to this manuscript.

Correspondence: M. T. Mira, Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição, 1155, Prado Velho, CEP 80215-901, Curitiba, Paraná, Brazil (m.mira@pucpr.br).

Clinical Infectious Diseases® 2019;XX(X):1-8

© The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com DOI: 10.1093/cid/ciz570

to DDS, RIF, and ofloxacin, respectively [11]. Consequently, there has been an increase in the number of sporadic case reports of DR and MDR isolates of *M. leprae* [11–14]. In 2009, the WHO launched a drug resistance surveillance program restricted to references and sentinel centers, which focused on pretreated individuals rediagnosed with leprosy [15].

To date, reports of *M. leprae* resistance rates range from 2.05% (of 243 isolates in Colombia) [16] to 16% (of 24 isolates in Guinea-Conakry) [17]. An Indian study of 239 relapses and 11 new cases found 21.6% of cases to be DR and 6.8% to be MDR [18]. Finally, a recent, large study of 1932 *M. leprae* strains obtained between 2009 and 2015 in sentinel centers of 19 countries determined that 8.0% (154) of the isolates presented some degree of resistance. Primary and secondary resistance to RIF were 2.0% (16/789) and 5.2% (58/1143), respectively; 20 cases (1.0%) were resistant to both DDS and RIF [19, 20].

Here, we present a population-based study on *M. leprae* resistance to MDT, performed in a hyperendemic population of a former leprosy colony located in the Brazilian Amazon. We propose that DR/MDR contributes to the maintenance of the endemicity in the village through reactivation/relapse and primary drug-resistant disease.

SUBJECTS AND METHODS

Ethics Statement

The ethics committees for human research of all involved institutions approved all methods and procedures used in this study (protocol number PUCPR 274.776) and written informed consent was obtained from all participants. The animal experiments were conducted in accordance with the guidelines of the Brazilian Committee for Care and Use of Laboratory Animals (protocol number ILSL 07/1).

Target Population

This study involved the entire population of the Vila do Santo Antonio do Prata (Prata village), located in the Amazonian state of Pará, Brazil. In the 1920s, the village became a venue for

compulsory isolation of leprosy patients who were diagnosed across the northern and northeastern states of Brazil. Isolation was mandatory until 1962; however, the strong social stigma associated with leprosy has been very present and, to date, emigration of affected individuals from the village and immigration of nonaffected individuals to the village has been limited. As a result, the Prata Village remains mostly socially and geographically isolated and, despite efforts towards disease control, leprosy is still highly prevalent and homogeneously distributed across the entire community [21].

At the time of enrollment, the village had an estimated population of around 3000 individuals, mainly composed of descendants of founding, leprosy-affected individuals. A previous, population-based genetic epidemiology study of the Prata village revealed a cumulative prevalence of leprosy of 12.8%, among the highest ever reported [21]. The same study revealed the presence of a strong genetic component controlling susceptibility to leprosy in the community, a finding compatible with the hypothesis that, due to the history of the village, genetic risk factors for leprosy are enriched within this population [21]. All inhabitants share the same limited environment: the village has only 1 active church, 2 elementary schools, 1 social club, and 1 large central square, used by the population for leisure activities such as soccer games and outdoor celebrations. In this context, it is reasonable to assume that all individuals are constantly, equally, and heavily exposed to *M. leprae*. Combined, these peculiar characteristics present the Prata village as a unique model for epidemiological, population-based leprosy studies.

Recruitment Strategy and Procedures

Recruitment followed a 2-stage strategy. In the first stage, in 2009, following WHO recommendations, we investigated *M. leprae* resistance, focusing on suspected cases of relapse [15]—as defined by individuals with new skin lesions and/or enlarged tender nerves who had been released from treatment

Table 1. Epidemiological Description of the Enrolled Patients

	2009		2013	
	Relapse	Contacts	Individuals	
Number of suspected cases examined	117	85	611	
...	Relapse	New case	Relapse	New case
Leprosy cases	12	10	11	9
Sex				
Female	5	7	3	5
Male	7	3	8	4
Clinical form at diagnosis, WHO				
MB	11	2	10	1
PB	1	8	1	8
Age range at diagnosis, years	10 to 43	8 to 79	10 to 49	10 to 69
Mean age at first diagnosis, years (SD)	24.5 (11.7)	32.2 (22.7)	26.3 (12.3)	34 (17.6)

Abbreviations: MB, multibacillary; PB, paucibacillary; SD, standard deviation; WHO, World Health Organization.

at least 5 years before—and household contacts. The resulting high number of cases of *M. leprae* with MDR led to the second stage, in 2013, of active case-finding, expanding the survey to the entire population (Table 1). Results of the DR/MDR analysis for all samples/patients were immediately released to the local health officials in order to inform therapeutic decisions. In 2017, individuals identified as cases or relapses in the first and/or second stage of recruitment were reached for follow-up. Detailed descriptions of the recruitment strategy and procedures are available in the [Supplementary Methods](#).

***Mycobacterium leprae* Resistance and Molecular Epidemiology Analyses**

Skin lesion biopsies were obtained and processed for diagnosis and estimation of the number of acid-fast bacilli. *Mycobacterium leprae* suspensions were inoculated in BALB/c mouse footpads for in vivo DDS and RIF resistance testing. *Mycobacterium leprae* DNA was obtained for molecular testing. Specific DRDR of genes *folP1*, *rpoB*, and *gyrA* were genotyped by Sanger sequencing; 17 short tandem repeats were selected and genotyped by fluorescence-based capillary electrophoresis. The individual discriminatory power of each short tandem repeat marker was calculated using the Hunter-Gaston discriminatory index (HGDI). An unweighted pair group method with arithmetic mean–based similarity matrix was created for the generation of a dendrogram and minimum spanning trees [22]. Detailed descriptions of these experimental procedures are available in the [Supplementary Methods](#).

RESULTS

Clinical and Epidemiological Data

In the first stage, 117 suspected cases of leprosy relapse and 85 contacts were evaluated; 12 cases of relapse were confirmed and 10 household contacts with active leprosy were identified. A total of 611 individuals were evaluated during the second stage of recruitment (55 of these were evaluated both in 2009 and 2013), and 20 cases of active leprosy were detected: 11 relapses and 9 new cases. Of the relapse cases, 5 had already been diagnosed with a leprosy relapse in 2009. Table 1 summarizes the epidemiological data of all diagnosed patients and Table 2 details the data of all diagnosed patients; information on previous treatments was available for all patients (Supplementary Table 1).

In 2017, the research team located the clinical records of 27 of the 37 leprosy patients diagnosed during the study. According to these records, 23 had received regular MDT; of these, 11 had been characterized as drug sensitive or inconclusive; 9 were resistant to DDS and RIF; 2 were resistant to DDS; and 1 was resistant to RIF. Also, of these 23 patients, 17 had completed the therapeutic course; 2 had died; 1 was still under treatment; and 3 did not have treatment information available. Our team of clinicians examined 5 out of these 23 patients after treatment:

1 sensitive and 4 resistant. All 5 presented full remission of the lesions. Out of the 27, 4 follow-up patients had received RIF, ofloxacin, and minocycline therapy: 1 was sensitive to all antibiotics tested and dropped out of treatment, and 3 were MDR, had completed the therapeutic course, and were clinically inactive after completion of treatment.

In Vivo Resistance

In 2009, skin biopsies were obtained from the 12 relapse cases and used for mouse footpad inoculation. There were 6 samples that yielded sufficient bacilli: 2 (PA-043 and PA-084) were resistant to both DDS and RIF; 1 (PA-155) was resistant to DDS; and 2 (PA-006 and PA-012) were susceptible to both DDS and RIF. The sixth sample, PA-192, did not show any growth (Supplementary Table 2). Skin biopsies for mouse footpad inoculation were also obtained from 8 relapses and 6 new cases detected in 2013. Of these, 3 presented sufficient bacilli for inoculation: 2 (PA-012 and PA-208) did not grow in any mouse group and 1 (PA-034) showed susceptibility to both DDS and RIF (Supplementary Table 2).

Genotyping of *Mycobacterium leprae* Drug Resistance–Associated Genes

A detailed description of the variants detected is available in Table 2. In summary, MDR *M. leprae* was detected in 40.9% (9/22) of the leprosy cases detected in 2009: 7 relapses and 2 new cases. In addition, a single *rpoB* mutation was detected in 1 new case, and a single *folP1* mutation was detected in 2 relapses and 1 new case. No *gyrA* gene variants were detected in any of the patients (Table 2). Out of the additional 15 leprosy cases identified only in 2013, 1 relapse case and 2 new cases presented MDR. Mutations outside of DRDR regions were detected: CCG100TGG in *folP1* (PA-248), CTA470CAA in *rpoB* (PA-034), and TTA97TTT in *gyrA* (PA-214 and PA-243; Table 2). Of note, 100% concordance on sensitivity/resistance statuses was observed across the molecular and in vivo analyses.

Familial Occurrences for *Mycobacterium leprae* Resistance Genes

Our strategy to include household contacts as primary targets of the surveys allowed the detection of 5 multiplex pedigrees displaying *M. leprae*–resistant isolates (Figure 1):

1. In the most remarkable case of familial clustering of *M. leprae* resistance, patient PA-043 was diagnosed with leprosy relapse in 2009. A contact examination revealed that his wife (PA-073) and his father-in-law (PA-074) were also affected. In 2013, the brother-in-law of PA-074 (PA-221) was diagnosed with leprosy relapse; an examination of his contacts led to the diagnosis of his daughter (PA-220) as a new leprosy case. Cases PA-043, PA-073, PA-074, and PA-220 were MDR, with PA-043 also showing MDR in vivo. Patient PA-221 did not yield a polymerase chain reaction product for DRDR analysis (Figure 1A).

Table 2. Clinical Information and Resistance Status of the Isolates

Family	ID	Recruitment ^a	Disease Status	Clinical Form at Diagnosis, WHO	Resistance ^b			Resistance Genotype		
					Dapsone	Rifampin	Ofloxacin	<i>folP1</i>	<i>rpoB</i>	<i>gyrA</i>
A	PA-043	2009 and 2013	Relapse	MB	Resistant ^c	Resistant ^c	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)/GAT44TTAT (Asp-Tyr) ^d	WT
A	PA-073	2009	New case	PB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
A	PA-074	2009	New case	PB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
A	PA-220	2013	New case	PB	Resistant	Resistant	Unknown	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	PCR neg
A	PA-221	2013	Relapse	MB	Susceptible	Susceptible	Susceptible	WT	WT	WT
B	PA-084	2009 and 2013	Relapse	MB	Resistant ^c	Resistant ^c	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
B	PA-085	2009 and 2013	Relapse	PB	Resistant	Resistant	Unknown	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	PCR neg
C	PA-058	2009	New case	PB	Susceptible	Susceptible	Susceptible	WT	WT	WT
C	PA-103	2009	New case	PB	Resistant	Unknown	Unknown	CCC55CGC (Pro-Arg)	PCR neg	PCR neg
D	PA-118	2009	New case	PB	Susceptible	Susceptible	Susceptible	WT	WT	WT
D	PA-119	2009	New case	PB	Unknown	Resistant	Susceptible	PCR neg	TCG456ATG (Ser-Met)	WT
E	PA-192	2009 and 2013	Relapse	MB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
E	PA-199	2009	New case	MB	Unknown	Unknown	Unknown	PCR neg	PCR neg	PCR neg
...	PA-006	2009	Relapse	MB	Susceptible	Susceptible	Susceptible	WT	WT	WT
...	PA-012	2009 and 2013	Relapse	MB	Susceptible ^c	Susceptible ^c	Susceptible	PCR neg	PCR neg	WT
...	PA-016	2009	Relapse	MB	Resistant	Resistant	Susceptible	ACC53GCC (Thr-Ala)	TCG456TTG (Ser-Leu)	WT
...	PA-034	2013	Relapse	MB	Susceptible	Susceptible ^c	Susceptible	WT	CTA470CAA (Leu-Gln)	WT
...	PA-040	2009	Relapse	MB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
...	PA-067	2009	Relapse	MB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
...	PA-070	2009	New case	PB	Unknown	Susceptible	Susceptible	PCR neg	WT	WT
...	PA-131	2009	Relapse	MB	Susceptible	Susceptible	Susceptible	WT	WT	WT
...	PA-146	2009	Relapse	MB	Resistant	Susceptible	Susceptible	CCC55CGC (Pro-Arg)	WT	WT
...	PA-155	2009	Relapse	MB	Resistant ^c	Susceptible	Susceptible	CCC55CTC (Pro-Leu)	WT	WT
...	PA-164	2009	New case	PB	Unknown	Unknown	Unknown	PCR neg	PCR neg	PCR neg
...	PA-167	2009	New case	MB	Unknown	Unknown	Susceptible	PCR neg	PCR neg	WT
...	PA-208	2013	Relapse	MB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
...	PA-213	2013	New case	PB	Unknown	Unknown	Unknown	PCR neg	PCR neg	PCR neg
...	PA-214	2013	Relapse	MB	Susceptible	Susceptible	Unknown	WT	WT	TTA97TTT (Leu-Phe)
...	PA-215	2013	New case	PB	Unknown	Unknown	Unknown	PCR neg	PCR neg	PCR neg
...	PA-232	2013	New case	PB	Unknown	Unknown	Unknown	PCR neg	PCR neg	PCR neg
...	PA-235	2013	New case	PB	Resistant	Resistant	Susceptible	CCC55CGC (Pro-Arg)	TCG456ATG (Ser-Met)	WT
...	PA-240	2013	Relapse	MB	Unknown	Unknown	Unknown	PCR neg	PCR neg	PCR neg
...	PA-243	2013	New case	PB	Susceptible	Unknown	Unknown	WT	PCR neg	TTA97TTT (Leu-Phe)
...	PA-245	2013	New case	PB	Susceptible	Unknown	Susceptible	WT	PCR neg	WT
...	PA-246	2013	New case	MB	Susceptible	Unknown	Susceptible	WT	PCR neg	WT

Table 2. Continued

Family	ID	Recruitment ^a	Disease Status	Clinical Form at Diagnosis, WHO	Resistance ^b			Resistance Genotype			
					Dapsone	Rifampin	Ofloxacin	folP1	rpoB	gyrA	
...	PA-248	2013	Relapse	MB	Unknown	Susceptible	Susceptible	Susceptible	CGG100TGG (Arg-Trp)	WT	WT
...	PA-254	2013	New case	PB	Susceptible	Susceptible	Susceptible	Susceptible	WT	WT	WT

Abbreviations: ID, identification number; MB, multibacillary; PB, paucibacillary; PCR neg, not amplified polymerase chain reaction; WHO, World Health Organization classification; WT, wild type.
^aFirst step in 2009 and second step in 2013.
^bResistance inferred from mutations in drug resistance-determining regions that have been proved to cause resistance.
^cResistance confirmed on mice test (in vivo).
^dVariant GAT441TAT (Asp-Tyr) has not been detected in any other members of the family.

2. Patient PA-084 and his wife, PA-085, were both diagnosed with relapses in 2009 and 2013, both with MDR strains (Figure 1B).
3. In 2009, 2 brothers, PA-103 (CCC55CGC in *folP1*) and PA-058 (wild-type), were diagnosed as new leprosy cases (Figure 1C).
4. In 2009, 2 sisters, PA-118 (wild-type) and PA-119 (TCG456ATG mutation in *rpoB*), were diagnosed as new leprosy cases (Figure 1D).
5. An MDR patient, PA-192, was identified as a relapse in both 2009 and 2013. Upon a contact examination in 2009, his mother-in-law (PA-199) was diagnosed as a new leprosy case; unfortunately, none of the DRDR genes of PA-199 yielded polymerase chain reaction product for DR/MDR investigation (Figure 1E).

Multiple-Locus Variable Number of Tandem Repeat Analysis Typing

A total of 42 samples, derived from 37 different patients (5 patients presented leprosy both in 2009 and 2013), were genotyped for the variable number tandem repeat markers. All 17 markers were successfully genotyped in 29 samples; the remaining samples had 9 to 16 genotyped markers (Suppl. Table 3). Of the evaluated markers, 5 were nonpolymorphic (HGDI = 0): AC8b, GGT5, 6-3A, 21-3, and 23-3. The largest allelic variability was detected for markers TA18, AT15, and AT17, with 14, 10, and 9 alleles, respectively, and HGDI scores between 0.77–0.89. The remaining markers presented HGDI scores ranging between 0.10–0.70 and have been ordered by decreasing variability as follows: GTA9 (6 alleles), AC8a and GGA21 (both 5 alleles), TA10 (4 alleles), AC9 and 6-7 (both with 3 alleles), 27-5 and 12-5 (both with 2 alleles), and 18-8 (2 alleles; Supplementary Table 3).

An analysis revealed the existence of 5 clusters (Figure 2): 1 included all 6 samples of 5 patients belonging to pedigree A; 1 corresponded to 4 samples from the 2 patients of pedigree B; 1 included the 2 samples of pedigree D; 1 was composed of both samples obtained from patient PA-192 (2009 and 2013); and 1 included patients PA-213 and PA-254. Of note, for all 5 cases that had samples collected in 2009 and 2013, only 1 (PA-012) did not have the 2 samples included in the same cluster, due to a difference in the AT15 marker.

DISCUSSION

Today, antimicrobial resistance is arguably the main challenge in fighting human infection. Yet, little is known about antimicrobial resistance in leprosy and most of the information on leprosy resistance to treatment comes from sporadic reports of resistant cases, reinforced by a few investigations of larger samples [16, 18, 23] and, more recently, a large study of global reach [19, 20]. These studies show a proportion of resistant *M. leprae* significantly higher than those reported before by the WHO [19, 24], indicating that the real extent of the problem

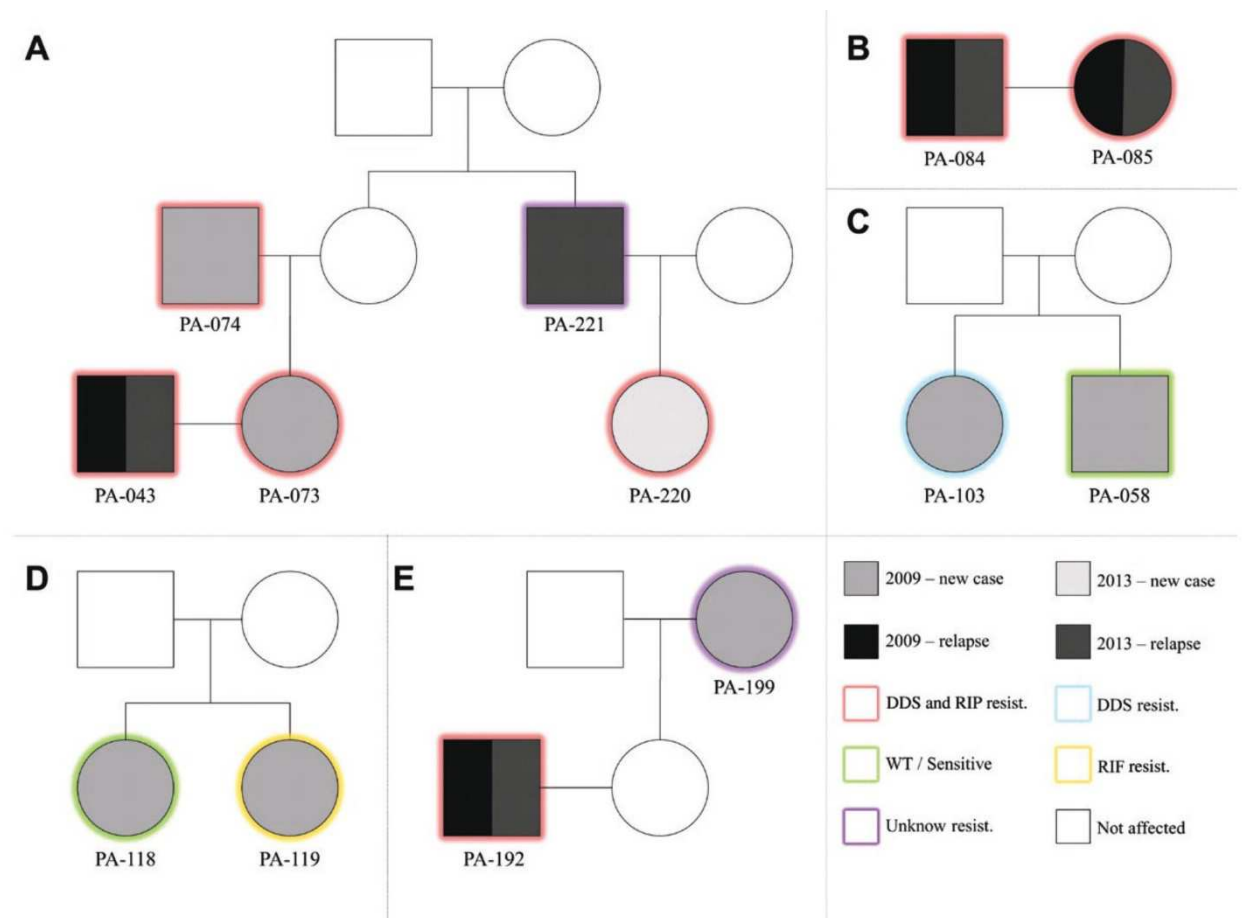


Figure 1. Male and female individuals are represented by squares and circles, respectively. Unfilled shapes are unexamined individuals. Labels under the squares/circles describe patients' IDs. Different fills indicate the year of diagnosis and whether it is a case or a relapse. The border color indicates the molecular resistance profile. Abbreviations: DDS, dapson; ID, identification number; RIF, rifampicin; WT, wild-type.

of *M. leprae* resistance to antibiotics is still unknown. This becomes even more critical in the scenario of a high hidden prevalence of leprosy, as recently suggested [25].

Here, we present the results of the first population-based, systematic report of *M. leprae* resistance in an isolated, hyperendemic population with decades of leprosy history. Our combined strategy of relapse monitoring and active searches for new cases has resulted in the identification of 37 leprosy cases, 16 (43.2%) of them showing some degree of resistance and 12 (32.4%) of which were resistant to both DDS and RIF. These high proportions of resistant *M. leprae* are likely due to the peculiar nature of the studied population, which has been heavily exposed to leprosy for almost 100 years: a combination of increased natural susceptibility and environmental factors likely favors the emergence, persistence, and propagation of resistant *M. leprae* through undiagnosed cases and asymptomatic carriers. Also, some relapse patients were first treated with DDS monotherapy in the 1960s and some have records of up to 4

rediagnoses, due to relapses or reinfections. Thus, the high level of resistance observed in the Prata Village is likely exceptional and not necessarily extensive to other populations. Still, the Prata Village, if seen as a natural experiment and interpreted accordingly, alerts us to the existence of a possible, elusive scenario of emergence of *M. leprae*-resistant strains that, due to disease characteristics such as an extremely long incubation period, may manifest as a public health problem only years ahead.

The combined molecular analysis of resistance and phylogenetic markers between isolates obtained from members of the same Prata Village families strongly suggests transmission along the pedigrees, in particular, of resistant strains characterizing primary resistance. This is particularly evident for pedigree A, which displayed at least 3 cases of primary, double-resistant leprosy (Figure 1) and phylogenetic data positioning all *M. leprae* isolates into a single cluster (Figure 2). Complete concordance of both resistance and phylogenetic profiles was also observed for pedigree B. For pedigree D, phylogenetic data

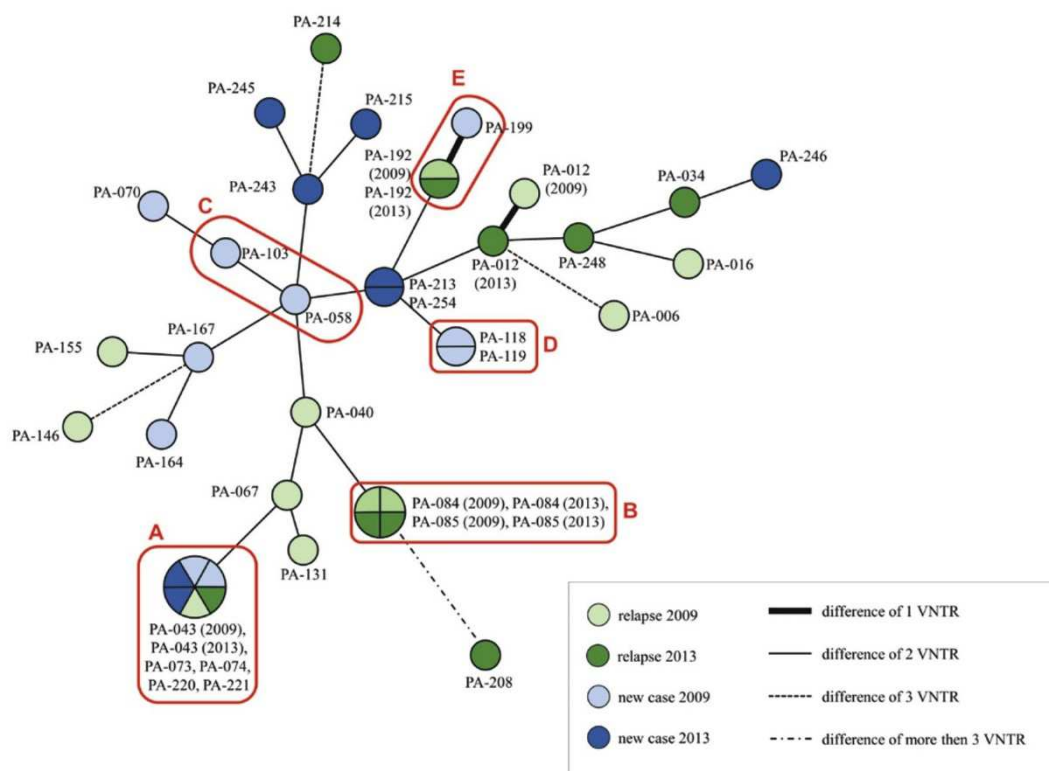


Figure 2. The lengths of the branches of the tree represent the distances between the genetic standards. Red boxes define individuals from the same pedigree, as labeled by the letters A (pedigree A of Figure 1) to E (pedigree E of Figure 1). Abbreviation: VNTR, variable number tandem repeats.

placed both isolates into the same cluster; however, the resistance profiles are distinct, likely due to the emergence of RIF resistance in 1 of the sisters. For pedigree C, distinct phylogenetic and resistance profiles suggest different isolates caused leprosy independently across siblings. For pedigree E, phylogenetic data differed only by a single AT17 variant; unfortunately, it was not possible to obtain resistance information for the primary case, which would have been particularly valuable since the index case is a multiple-relapse, double-resistant patient. Since marker AT17 is highly unstable and polymorphic, it is possible that the discordance is due to natural molecular changes observed within the same isolate. A similar explanation may apply to the single divergence (marker AT15, patient PA-012) among the 5 patients for whom samples were obtained both in 2009 and 2013. This great similarity between strains of the same patients, collected at different stages, indicates that the disease in these cases was caused by the same strain at both moments. Of note, a complementary phylogenetic analysis indicated the presence of *M. leprae* Single Nucleotide Polymorphism types 3 and 4 in the Prata Village—with a 76.2% predominance of the later—that is compatible with the mixed ethnic background of the Prata Village [21] (data not shown).

3 of the alleles in the DRDR region of the genes detected in the study haven't been previously associated with *M. leprae* resistance: (1) CTA470CAA (Leu-Gln) in *rpoB*, detected in 1 relapse patient; (2) CGG100TGG (Arg-Trp) in *folP1*, also detected in 1 relapse patient; and (3) TTA97TTT (Leu-Phe) in *gyrA*, detected in 1 relapse and 1 new case. All isolates harboring these mutations were inoculated in the mouse footpad; however, only *rpoB* CTA470CAA (Leu-Gln) yielded positive growth, showing sensitivity to DDS and RIF in vivo. Thus, both *folP1* and *gyrA* mutations remain as natural candidates for further investigation of a potential role in leprosy resistance.

The long-term follow-up indicates that a minor proportion of the patients resistant do DDS, RIF, or both had received regular MDT, perhaps following outdated guidelines, combined with complex administrative and bureaucratic procedures, in order to guarantee treatment. The impact of such deleterious effects upon the emergence and maintenance of *M. leprae* resistance is unknown.

We are aware of the limitations of our study. First, despite important technological advances, working with *M. leprae* still poses challenges, such as obtaining enough quantities from biological samples to successfully achieve bacterial growth in mice—particularly difficult for paucibacillary patients—in

order to allow in vivo testing of resistance and access to enough DNA to perform molecular analyses. To address this limitation, molecular data have been re-checked by a reference laboratory with 100% concordance. Second, a recent genomic analysis suggests the existence of novel targets involved in antimicrobial resistance in leprosy that haven't yet been tested in the Prata population [26, 27]; including these novel markers through a follow-up study involving whole-genome sequencing would produce a more comprehensive description of the resistance molecular profile and a higher resolution pattern of transmission, possibly increasing resistance rates even further.

In summary, the results of our systematic study revealed the underdetection of primary and secondary resistance and dissemination of *M. leprae*-resistant strains in the context of a very unique population. It is unlikely that the scenario observed in the Prata Village also applies to less atypical contexts; however, we expect that our data can raise awareness for a potential, alarming scenario that demands immediate action from leprosy control authorities in this particular region of Brazil and calls for better surveillance in other regions of the globe.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. The authors thank the patients for agreeing to participate in the study, as well as Masanori Matsuoka and Masanori Kai from the National Institute of Health in Japan for quality control of sequenced samples.

Financial support. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant numbers MCT/CNPq/MS-SCTIE-DECIT 35/2005, MCT/CNPq/CT-Saúde/MS/SCTIE/DECIT Nº 034/2008, and MCTI/CNPq/MS-SCTIE – Decit Nº 40/2012).

Potential conflict of interests. M. T. M. was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant number CNPq-PQ 2). H. R. S. D. was supported by Fundação Araucária (grant number FA 247/2015). P. N. S. was supported by the Damien Foundation. All other authors report no potential conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Smith CS, Aerts A, Saunderson P, Kawuma J, Kita E, Virmond M. Multidrug therapy for leprosy: a game changer on the path to elimination. *Lancet Infect Dis* **2017**; 17:e293–7.
2. Pettit JH, Rees RJ. Sulphone resistance in leprosy: an experimental and clinical study. *Lancet* **1964**; 2:673–4.

3. Pearson JM, Haile GS, Rees RJ. Primary dapsone-resistant leprosy. *Lepr Rev* **1977**; 48:129–32.
4. Jacobson RR, Hastings RC. Rifampin-resistant leprosy. *Lancet* **1976**; 2:1304–5.
5. World Health Organization. Chemotherapy of leprosy for control programmes. *World Health Organ Tech Rep Ser* **1982**; 675:1–33.
6. Shetty VP, Uplekar MW, Antia NH. Primary resistance to single and multiple drugs in leprosy—a mouse footpad study. *Lepr Rev* **1996**; 67:280–6.
7. Cambau E, Perani E, Guillemin I, Jamet P, Ji B. Multidrug-resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. *Lancet* **1997**; 349:103–4.
8. Machado D, Lecorche E, Mougari F, Cambau E, Viveiros M. Insights on *Mycobacterium leprae* efflux pumps and their implications in drug resistance and virulence. *Front Microbiol* **2018**; 9:1–10.
9. Yew WW, Liang D, Chan DP, Shi W, Zhang Y. Molecular mechanisms of clofazimine resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* **2017**; 72:2943–4.
10. Shepard CC. The experimental disease that follows the injection of human leprosy bacilli into foot-pads of mice. *J Exp Med* **1960**; 112:445–54.
11. Contreras Mejia Mdel C, Porto Dos Santos M, Villarouco da Silva GA, et al. Identification of primary drug resistance to rifampin in *Mycobacterium leprae* strains from leprosy patients in Amazonas State, Brazil. *J Clin Microbiol* **2014**; 52:4359–60.
12. Matsuoka M. Drug resistance in leprosy. *Jpn J Infect Dis* **2010**; 63:1–7.
13. Williams DL, Gillis TP. Drug-resistant leprosy: monitoring and current status. *Lepr Rev* **2012**; 83:269–81.
14. da Silva Rocha A, Cunha Md, Diniz LM, et al. Drug and multidrug resistance among *Mycobacterium leprae* isolates from Brazilian relapsed leprosy patients. *J Clin Microbiol* **2012**; 50:1912–7.
15. World Health Organization. Guidelines for global surveillance of drug resistance in leprosy. New Delhi, India: WHO Regional Office for South-East Asia, **2009**.
16. Beltrán-Alzate C, López Diaz F, Romero-Montoya M, et al. Leprosy drug resistance surveillance in Colombia: the experience of a sentinel country. *PLOS Negl Trop Dis* **2016**; 10:e0005041.
17. Avanzi C, Busso P, Benjak A, et al. Transmission of drug-resistant leprosy in Guinea-Conakry detected using molecular epidemiological approaches. *Clin Infect Dis* **2016**; 63:1482–4.
18. Lavania M, Singh I, Turankar RP, et al. Molecular detection of multidrug-resistant *Mycobacterium leprae* from Indian leprosy patients. *J Glob Antimicrob Resist* **2018**; 12:214–9.
19. Cambau E, Saunderson P, Matsuoka M, et al. Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009–2015. *Clin Microbiol Infect* **2018**; 24:1305–10.
20. Cambau E, Saunderson P, Matsuoka M, et al; World Health Organization Surveillance Network of Antimicrobial Resistance in Leprosy. Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009–15. *Clin Microbiol Infect* **2018**; 24:1305–10.
21. Lázaro FP, Werneck RI, Mackert CC, et al. A major gene controls leprosy susceptibility in a hyperendemic isolated population from north of Brazil. *J Infect Dis* **2010**; 201:1598–605.
22. Fontes ANB, Lima LNGC, Mota RMS, et al. Genotyping of *Mycobacterium leprae* for better understanding of leprosy transmission in Fortaleza, Northeastern Brazil. *PLOS Negl Trop Dis* **2017**; 11:e0006117.
23. Avanzi C, Del-Pozo J, Benjak A, et al. Red squirrels in the British Isles are infected with leprosy bacilli. *Science* **2016**; 354:744–7.
24. World Health Organization. A guide for surveillance of antimicrobial resistance in leprosy: 2017 update. New Delhi, India: WHO Regional Office for South-East Asia, **2017**.
25. Salgado CG, Barreto JG, da Silva MB, et al. Are leprosy case numbers reliable? *Lancet Infect Dis* **2018**; 18:135–7.
26. Singh P, Benjak A, Carat S, et al. Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3. *Clin Microbiol Infect* **2014**; 20:O619–22.
27. Benjak A, Avanzi C, Singh P, et al. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. *Nat Commun* **2018**; 9:1–11.

6. MÉTODOS COMPLEMENTARES

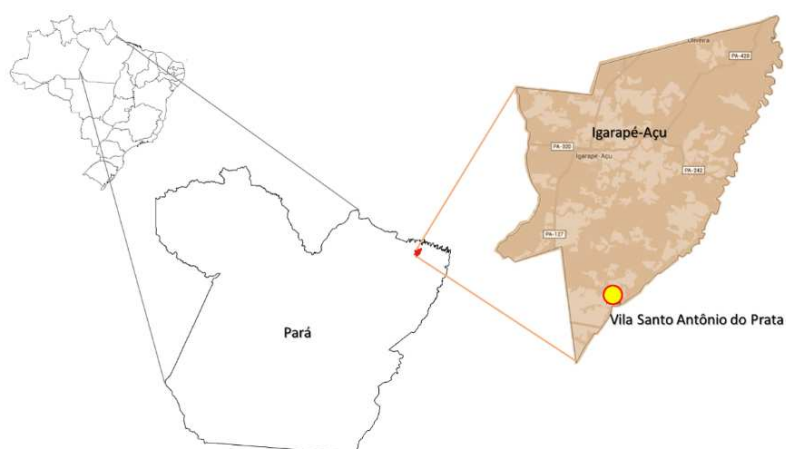
6.1. SUPORTE FINANCERO

O projeto de origem foi submetido e aprovado para financiamento no edital DECIT/CNPq, chamada MCTI/CNPq/MS-SCTIE - Decit N° 40/2012 - Pesquisa em Doenças Negligenciadas, em um total de R\$ 236.021,20 (processo 403889/2012-0). Esse edital já foi finalizado e os relatórios finais técnico e financeiro já foram aprovados.

6.2. POPULAÇÃO DA VILA SANTO ANTONIO DO PRATA

A população alvo deste estudo reside na Vila Santo Antônio do Prata (antiga “Colônia do Santo Antônio do Prata” ou “Lazarópolis do Prata”), localizada no município de Igarapé-açu, cerca de 110 km a leste de Belém, capital do estado do Pará, norte do Brasil (ver **Figura 21**). A Vila do Prata foi fundada por frades franciscanos em 1898 e, mais tarde, passou a ser sítio de isolamento compulsório (no Brasil de 1924 a 1962) de pacientes hansenianos do estado do Pará e Maranhão. Desde sua fundação até hoje, o fluxo migratório espontâneo tem sido praticamente ausente, provavelmente devido ao estigma que ainda acompanha a doença (134).

Figura 21 - Localização da Vila Santo Antônio do Prata



Legenda: Mapa do Brasil, com destaque para o estado do Pará e para o município de Igarapé-Açu (em vermelho) onde está sinalizado (em amarelo) a localização da Vila Santo Antônio do Prata.

Localização aproximada: 1°18'14.1"S 47°35'57.3"W

Fonte: a autora (2018).

6.2.1. Recrutamento da amostra populacional

Para o recrutamento da amostra populacional deste estudo, foi aplicada estratégia de atendimento dermatológico de demanda espontânea, aliada à busca ativa de casos de hanseníase, dirigida por agentes comunitários de saúde locais, em atividade intensiva de uma semana. A equipe foi composta por cinco médicos dermatologistas, três profissionais de saúde (de apoio, para coleta de material) e seis agentes comunitários de saúde da Vila do Prata. Essa equipe foi distribuída em três grupos, encarregados de atender a demanda por consulta dermatológica; além disso, quando indicado pelos agentes comunitários de saúde, foram feitas visitas em domicílios. Foi dada especial atenção aos possíveis casos entre as pessoas que conviveram com doentes (comunicantes), bem como àqueles já tratados com PQT, mas suspeitos de reativação da doença. A presença de moradores conhecidos na equipe (os agentes comunitários de saúde) também teve como objetivo reduzir a eventual resistência da população.

Todos os pacientes atendidos no mutirão foram investigados para doenças dermatológicas em geral e hanseníase em particular, para as quais receberam tratamentos adequados, quando necessário. Quando houve suspeita clínica de hanseníase, o indivíduo foi submetido a exame de baciloscopia e exame histopatológico de biópsia para a confirmação da doença. As amostras biológicas que foram confirmadas como hanseníase ativa tiveram as biópsias encaminhadas para as demais análises.

Para cada paciente, comunicantes e indivíduos da população em geral com suspeita clínica de hanseníase ativa, foram registrados dados clínicos-epidemiológicos pelo médico, durante a avaliação dermatológica. Dados do prontuário também foram inclusos, se relevantes. A classificação clínica de casos foi realizada de acordo com os critérios Ridley & Jopling (135).

6.3. CONSIDERAÇÕES ÉTICAS

As amostras biológicas (biópsias) foram coletadas dos pacientes, mas nenhum experimento foi realizado utilizando-se marcadores genéticos de DNA humano; toda a investigação se concentrou nas variações de genoma do *M. leprae*. Ainda assim, todos os sujeitos recrutados (ou seus representantes legais), ao concordarem em participar do estudo e ceder uma amostra de biópsia, assinaram o Termo de Consentimento Livre e Esclarecido, devidamente aprovado pelo Comitê de Ética em Pesquisa (CEP) da PUCPR (Parecer 274.776). Os experimentos foram conduzidos segundo as normas brasileiras de experimentação animal, assim como a aprovação do comitê para uso dos mesmos (ILSL – protocolo 07/1).

6.4. TESTE DE RESISTÊNCIA *in vivo*

As amostras biológicas (biópsias) coletadas dos pacientes da Vila do Prata foram encaminhadas para extração de DNA (para análises de biologia molecular) e, em paralelo, ao cultivo *in vivo*, realizado em instituição parceira (ILSL). O DNA extraído diretamente da amostra de biópsia tende a conter grandes quantidades de DNA do hospedeiro humano e pouca quantidade de DNA da bactéria, o que é inconveniente para o sequenciamento. Assim, optou-se por utilizar a amostra com crescimento de *M. leprae in vivo*, processo que requer de 8 a 12 meses de cultivo (em camundongo nude). O cultivo e extração do bacilo da pata foram realizados conforme protocolo previamente descrito (136).

6.5. PADRONIZAÇÃO DO PROTOCOLO DE SEQUENCIAMENTO DE GENOMA COMPLETO DE *M. LEPRAE*

6.5.1. Purificações do bacilo e extração de DNA

Para o sequenciamento do DNA genômico bacteriano por sequenciamento de segunda geração, é necessário minimizar a contaminação da amostra com o DNA do camundongo. Para isso, duas diferentes técnicas de purificação do bacilo extraído do coxim plantar foram testadas. Posteriormente, amostras purificadas por

estes métodos foram submetidas a um sequenciamento para avaliação e validação dos resultados.

6.5.1.1. Purificação com NaOH

Essa purificação foi realizada por nossos parceiros no ILSL. O tecido advindo da pata foi triturado e centrifugado (136). Os precipitados foram então suspensos em NaOH 0,1N estéril, centrifugado e acrescido em meio RPMI 1640 estéril. Por fim, o material foi homogeneizado 10 vezes por passagem em seringa de insulina de 1mL para dissolver os grumos dos bacilos. A extração de DNA foi realizada conforme protocolo otimizado de extração do DNA de micobactérias em pequena escala, proposto por Santos e col. (137).

6.5.1.2. Purificação com colagenase-dispase e NaOH

Essa purificação foi realizada por nossos parceiros na FIOCRUZ. O tecido advindo da pata do camundongo foi picado em pedaços pequenos com tesoura pequena e pinça estéreis. Em seguida, foi adicionado meio RPMI 1640, solução de colagenase-dispase 2x e ampicilina para concentração final de 50µg/mL. A placa foi homogeneizada por agitação suave e incubada por 2 horas a 33°C. Após esse período, a solução foi homogeneizada por pipetagem, centrifugada e lavada duas vezes com água estéril livre de pirogênio, sendo re-centrifugada a cada lavagem. Após esse processo o material foi purificado com NaOH conforme item anterior (“6.5.1.1 Purificação com NaOH”) (138).

6.5.2. Etapas do sequenciamento

As etapas de sequenciamento para validação do processamento de amostras foram realizadas de acordo com protocolos e recomendações do fabricante (139-141). Os reagentes utilizados em cada etapa estão descritos na **Tabela 9**.

Tabela 9 – Reagentes utilizados para o sequenciamento

Etapa	Sub-etapa	Reagentes (Cat nº)
Preparo de Biblioteca	Fragmentação	Ion Xpress™ Plus Fragment Library Kit (4471269)
	Purificação da fragmentação	Agencourt® AMPure® XP Kit Beckman Coulter (A63881)
	Ligação de adaptadores e <i>barcode</i>	Ion Xpress™ <i>Barcode</i> Adapters 1-16 Kit" (4471250)
	Purificação da ligação	Agencourt® AMPure® XP Kit Beckman Coulter (A63881)
	Seleção de tamanho	E-Gel® SizeSelect™ Agarose Gels, 2%" (G6610-02)
	Amplificação	Ion Xpress™ Plus Fragment Library Kit (4471269)
	Purificação com <i>bead</i>	Agencourt® AMPure® XP Kit Beckman Coulter (A63881)
	Quantificação	Ion Library Quantitation Kit" (4468802) ou High Sensitivity DNA Analysis Kits - Agilent Technologies (5067-4626)
Preparo de <i>template</i>	PCR em emulsão	Ion PGM™ <i>Template</i> OT2 400 Kit" (4479878)
	Enriquecimento das <i>beads</i>	Ion PGM™ <i>Template</i> OT2 400 Kit" (4479878)
		Ion PGM™ <i>Enrichment Beads</i> " (4478525)
Sequenciamento	Sequenciamento	Ion PGM™ Sequencing 400 Kit" (4482002)
		Ion 318™ Chip v2 (4484355)

Legenda: Cat nº) Número de referência do produto no catálogo da indústria.

Fonte: a autora (2018).

6.5.2.1. Purificação com *beads*

Durante a padronização do sequenciamento constatou-se a necessidade de adição de processo de purificação com *beads* adaptado de outro protocolo para *Ion Torrent*, voltado para purificação para amostras de DNA de qualidade reduzida. O diferencial deste é a não eluição do material genético das *beads* ao fim da purificação, o que consequentemente faz com que a fragmentação seja feita com o material ainda ligado às *beads* (142).

6.5.2.2. Preparo de biblioteca

Amostras de DNA genômico foram fragmentadas enzimaticamente e purificadas com *beads* paramagnéticas. Os fragmentos gerados foram ligados aos adaptadores e *barcodes* e novamente purificados, sendo doravante consideradas como “bibliotecas não amplificadas”. Essas bibliotecas foram aplicadas em gel de agarose a 2% para seleção dos fragmentos com tamanho de cerca de 480pb, ideais para a fase de sequenciamento. As bibliotecas foram amplificadas por PCR, novamente purificadas e quantificadas, utilizando PCR em tempo real ou, alternativamente, utilizando o *2100 Bioanalyzer Instruments*. As concentrações obtidas foram utilizadas para o preparo das soluções de trabalho das bibliotecas, com concentrações padronizadas pelo protocolo (139).

6.5.2.3. Preparo de *template*

Nessa etapa, foi feito um *pool* das soluções de trabalho das bibliotecas, a ser submetido a uma PCR em emulsão no *Ion OneTouch™ 2 Instrument*, para amplificação clonal dos fragmentos da biblioteca na superfície das *beads*. Em seguida, as *beads* foram enriquecidas, utilizando o equipamento *Ion OneTouch™ ES*.

6.5.2.4. Sequenciamento

Primeiramente, foi feito o planejamento virtual da corrida de sequenciamento no *Torrent Browser* e a limpeza e inicialização do sequenciador *Ion PGM™*. As esferas carregadas com as bibliotecas amplificadas foram incubadas para acoplamento dos *primers* de sequenciamento. Em seguida, o *chip* foi testado no sequenciador (*Chip Check*), a amostra foi injetada no *chip* e este foi colocado no *Ion PGM™*, tendo sido, então, iniciada a corrida de sequenciamento.

6.6. PADRONIZAÇÃO DO PROTOCOLO DE ANÁLISE DOS DADOS DE SEQUENCIAMENTO DE GENOMA COMPLETO DE *M. LEPRAE*

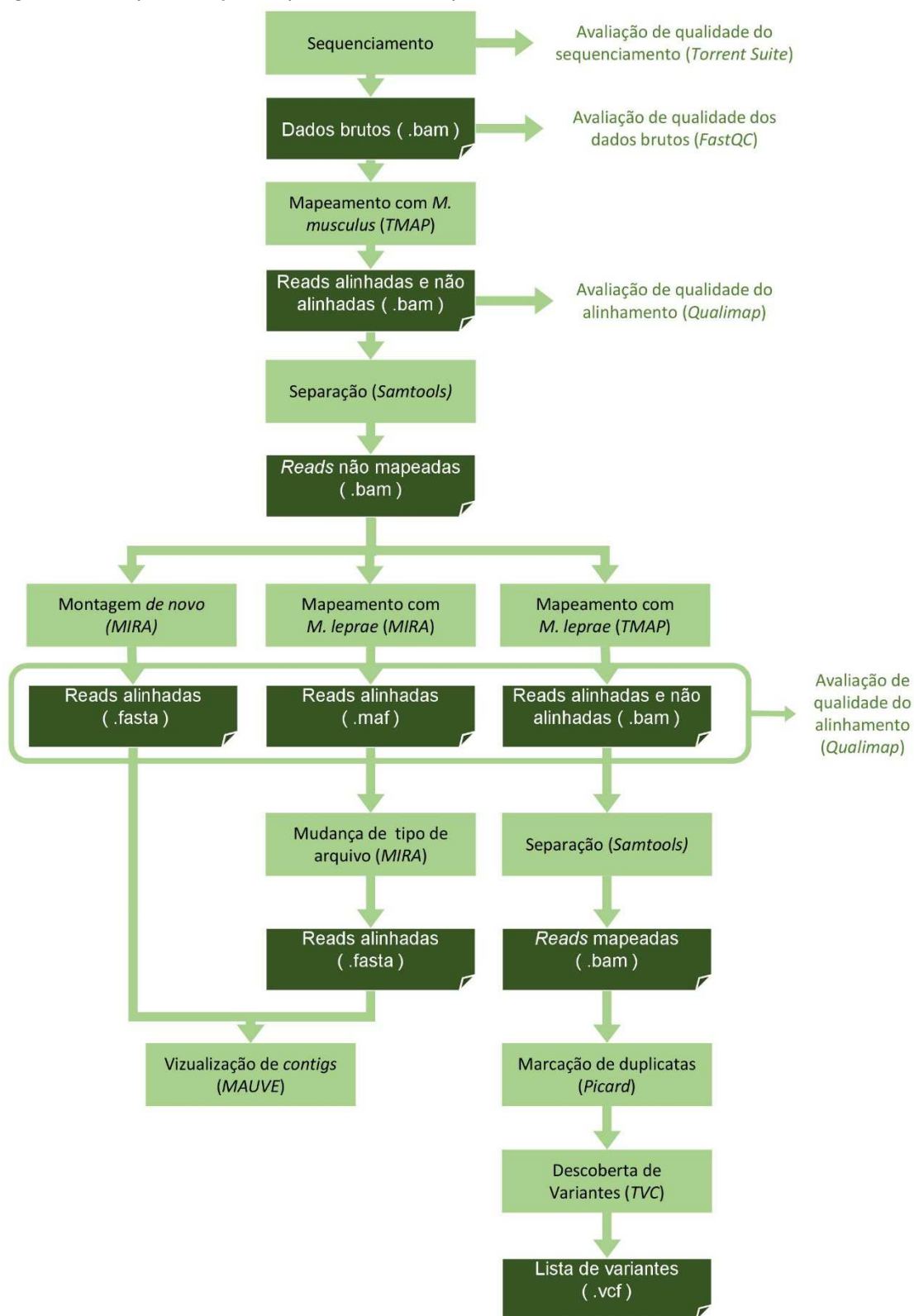
Os programas utilizados na padronização de análises de dados estão sucintamente descritos na **Tabela 10**. Uma apresentação esquemática das análises realizadas pode ser observada na **Figura 22** e a descrição mais pormenorizada dos parâmetros utilizados pode ser consultada no **APÊNDICE II**.

Tabela 10 – Programas utilizados na análise de dados

Programa	Versões	Objetivo	Fontes
Samtools	1.5 e 1.6	Utilitários para manipulação de alinhamentos no formato SAM incluindo ordenação, fusão, indexação etc	(143)
TMAP	5.4.11	<i>Software</i> de alinhamento (apenas com referência) otimizado para sequencias de <i>Ion Torrent</i> *	(144)
TVC	5.4-11	Identificação de variantes (<i>SNP</i> , <i>InDels</i> etc) otimizado para dados de <i>Ion Torrent</i> *	(145)
Torrent Suite	5.0.2	<i>Software</i> nativo da plataforma <i>Ion Torrent</i> , responsável pelo processamento de dados do instrumento (<i>Ion PGM</i>) e pela análise de qualidade do sequenciamento	(145)
MIRA	4.0	Faz alinhamento/ mapeamento em sequências contiguas (<i>contigs</i>) e conversão dos arquivos de alinhamento/mapeamento	(146)
Mauve	2015-02-13	Sistema de alinhamento e visualização de múltiplos genomas	(147)
FastCQ	0.10.1	Ferramenta de controle de qualidade de dados brutos (<i>raw data</i>) de sequenciamento	(148)
Qualimap	2.2.1	Gera relatório de qualidade de dados alinhados/mapeados	(149)
Prinseq	0.20.4	Gera relatório de qualidade de dados brutos e realiza filtragem desses dados.	(150)
Picard	2.9.0-1	Marcação de <i>reads</i> duplicadas	(151)

Legenda: TMAP) *Torrent Mapping Alignment Program*. TVC) *Torrent Variant Caller*. MIRA) *Mimicking Intelligent Read Assembly*. *)A otimização de alguns softwares para *Ion Torrent* em geral leva em consideração o flow order e flow signal, atributos específicos dessa plataforma.

Figura 22 – Representação esquemática das etapas de análise de dados



Fonte: a autora (2018)

6.7. OUTRAS ATIVIDADES

Além das atividades relacionadas diretamente a esse estudo, a aluna atuou como colaboradora de outros projetos que resultaram em:

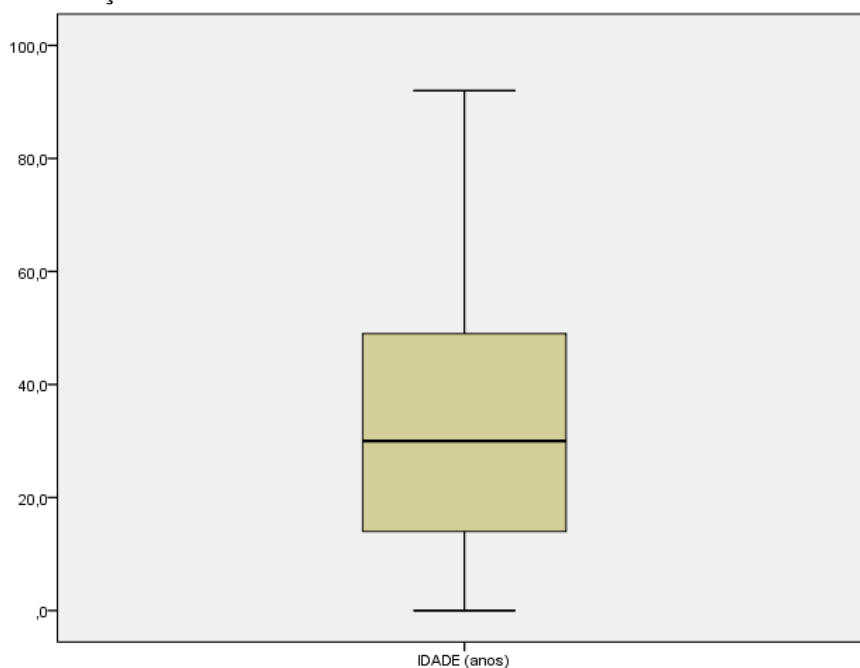
- Um artigo de revisão publicado na *Clinics in Dermatology* (**Anexo I**);
- Um artigo publicado no *International Journal of Systematic and Evolutionary Microbiology* (**Anexo II**);
- Um capítulo de livro: “Bioética, Biotecnologia e Biopoder” (**Anexo III**);
- Manuscrito em fase final de preparação (**Anexo IV**) a ser submetido em breve;
- Um *software*, denominado *pseudosib* em fase final de desenvolvimento (disponível sob solicitação no site: <https://gitlab.com/helena.falena/pseudosib>).

7. RESULTADOS E DISCUSSÃO COMPLEMENTARES

7.1. RECRUTAMENTO DA AMOSTRA POPULACIONAL E COLETA DE MATERIAL BIOLÓGICO

Todos os pacientes atendidos no mutirão de atendimento médico de 2013 foram investigados para doenças dermatológicas em geral e hanseníase em particular. Esta estratégia de recrutamento resultou em 611 consultas dermatológicas, sendo 365 (59,7%) indivíduos do sexo feminino, 225 (36,8%) do sexo masculino e 21 (3,4%) não declarados. A idade dos indivíduos examinados variou de menos de 1 ano completo a 92 anos de idade, sendo a média de 32,8 anos (desvio padrão de 21,3 anos), conforme distribuição representada na **Figura 23**.

Figura 23 - Distribuição da idade dos 611 indivíduos examinados.



Legenda:

Gráfico *box-plot* da distribuição de idade dos indivíduos examinados.

A linha preta central representa a mediana, os limites superior e inferior do quadrado marrom representam o intervalo do primeiro ao terceiro quartil. Externamente ao quadrado, as linhas horizontais inferior e superior representam 10% e 90% dos dados, respectivamente.

Fonte: A autora, no *software* SPSS (2017).

Dentre os examinados, 98 pessoas possuíam características compatíveis com hanseníase no exame clínico, incluindo pacientes já tratados, com lesões residuais e estados reacionais. Destes, 52 (8,51%) foram considerados casos suspeitos de hanseníase ativa e tiveram amostras de biópsia de pele coletadas. Os indivíduos que tiveram outras doenças dermatológicas diagnosticadas (**Tabela 11**) foram encaminhados para tratamento adequado.

Tabela 11 – Distribuição categórica de doenças dermatológicas nos indivíduos com alguma doença dermatológica.

Tipo de doença	Frequência^a	Porcentagem (%)^a
Hanseníase (ativa ou não, lesões residuais e estados reacionais)	98	33,11
Infectocontagiosas^b	66	22,3
Discromias	50	16,89
Inflamatórias	39	13,18
Neoplásicas	31	10,47
Outros	12	4,05
TOTAL	296	100

Legenda:

a) As frequências e porcentagens podem exceder a 100% devido a indivíduos que tinham mais de uma doença em simultâneo.

b) Exceto a própria Hanseníase.

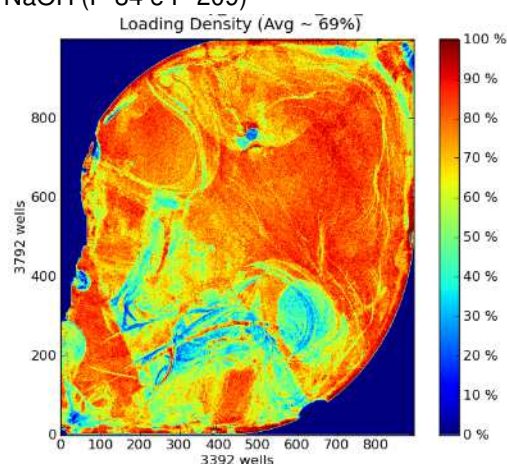
7.2. PADRONIZAÇÃO DE SEQUENCIAMENTO E ANÁLISE DE GENOMA COMPLETO

O DNA utilizado no sequenciamento de genoma completo foi obtido a partir das cepas de *M. leprae* cultivadas em pata de camundongo nude. Dessa forma, dentre as amostras obtidas no Prata, apenas as cepas PA-084 e PA-220 apresentaram crescimento suficiente *in vivo* e tiveram seu DNA sequenciado. Outras duas cepas advindas de pata de camundongo, a cepa ROS-MAD e a cepa LC, foram cedidas pelo Instituto Lauro de Souza Lima.

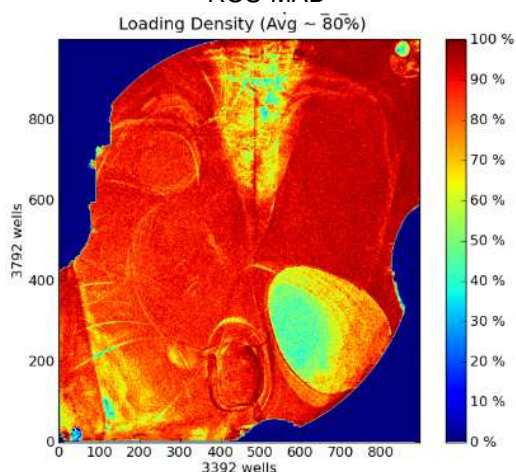
7.2.1. Análise geral do sequenciamento

O carregamento dos chips de sequenciamento variou de 55 a 80% (**Figura 24**): O enriquecimento das *beads* ficou entre 97 a 100%, das quais 50 a 66% geraram *reads* monoclonais. De 7 a 30% dessas *reads* foram descartadas por baixa qualidade (**Figura 25**).

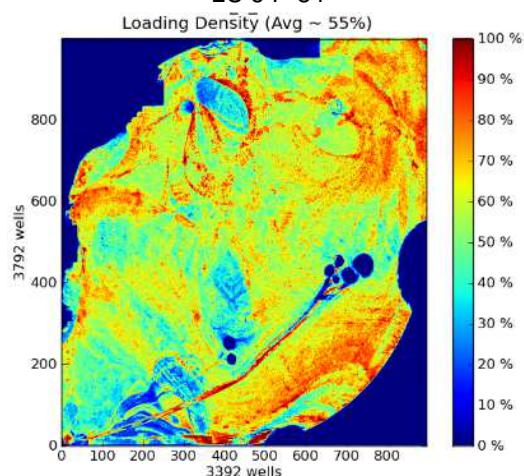
Figura 24 - Carregamento dos chips contendo as amostras
A) purificadas apenas com NaOH (P-84 e P-209)



B) purificadas com NaOH e colagenase –
ROS-MAD



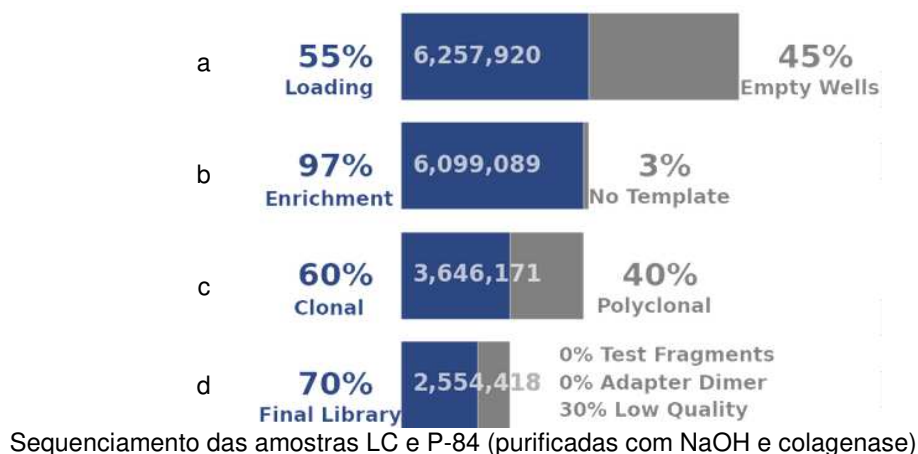
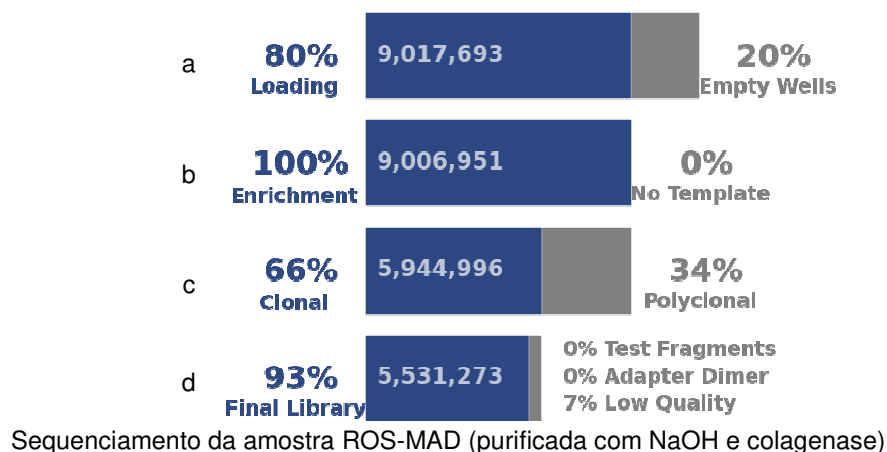
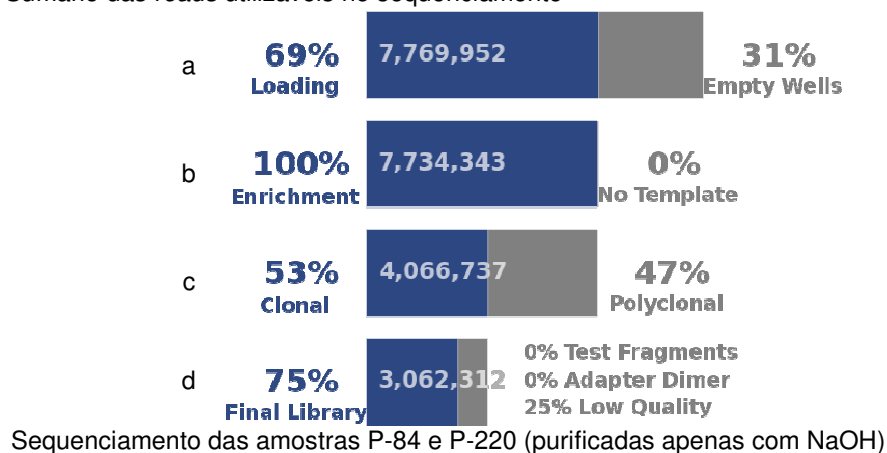
C) purificadas com NaOH e colagenase –
LC e P-84



Representação do chip de sequenciamento com o carregamento das *beads* nos poços sendo representado pelas cores. Quanto mais próximo ao vermelho (cores mais quentes), maior a porcentagem de carregamento de *beads* na área e, quanto mais próximo ao azul (cores mais frias), menor a porcentagem de carregamento.

Fonte: *Torrent Suite 5.0.2*

Figura 25 - Sumário das *reads* utilizáveis no sequenciamento



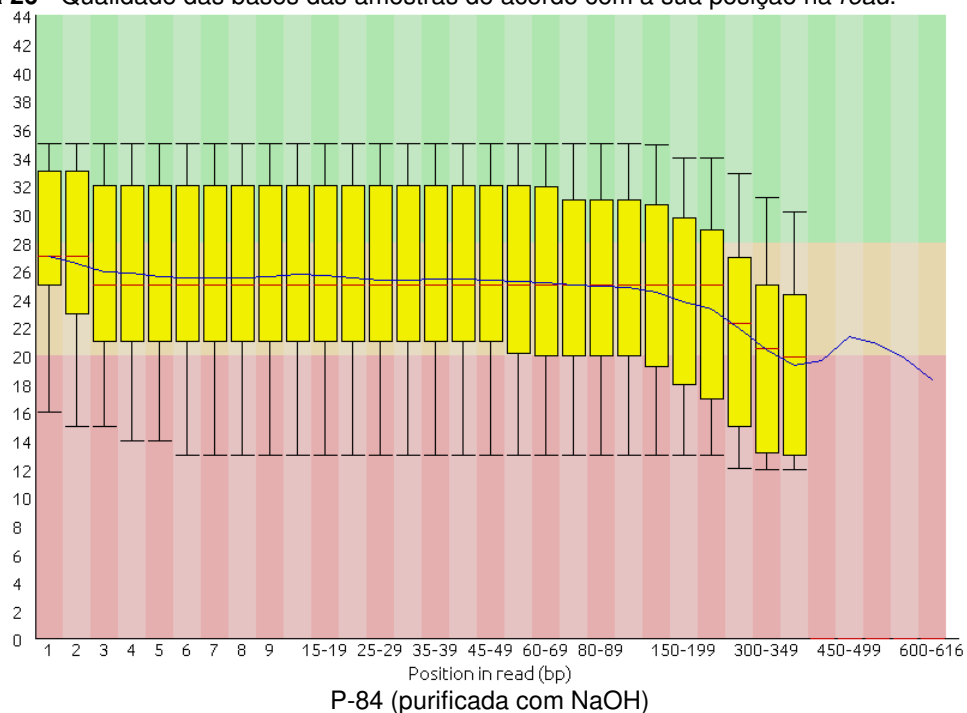
De cima para baixo, cada linha representa as filtragens subsequentes dos dados. Em cinza estão representados os dados descartados e em azul os mantidos. a) filtragem dos poços não carregados com *beads*; b) filtragem das *beads* não carregadas; c) filtragem das *beads* policlonais e; d) filtragem das bibliotecas de baixa qualidade.

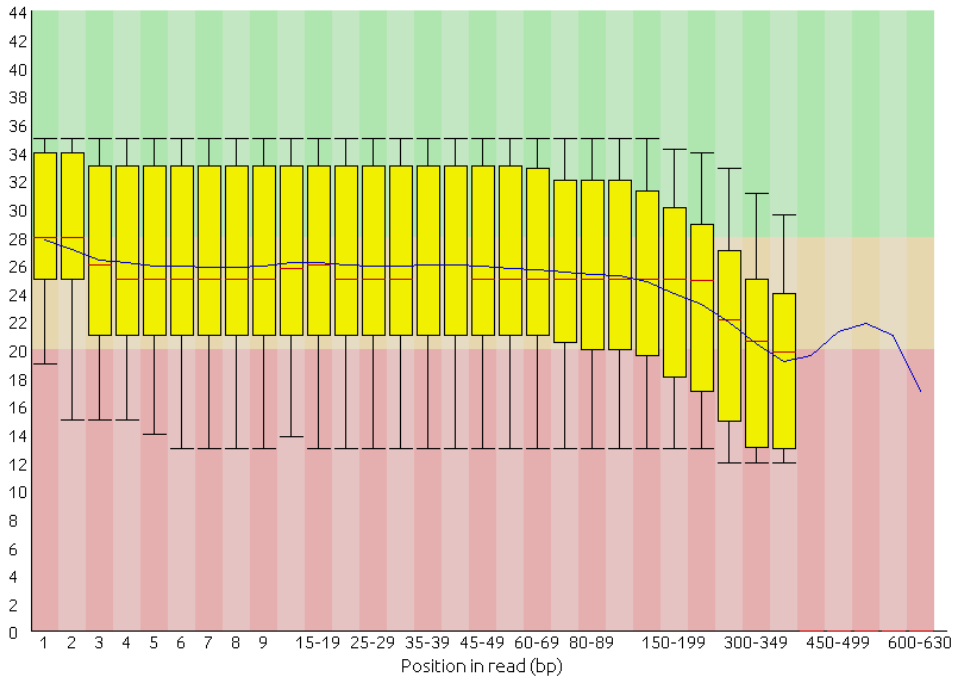
Fonte: *Torrent Suite 5.0.2*

7.2.2. Análise de qualidade do sequenciamento

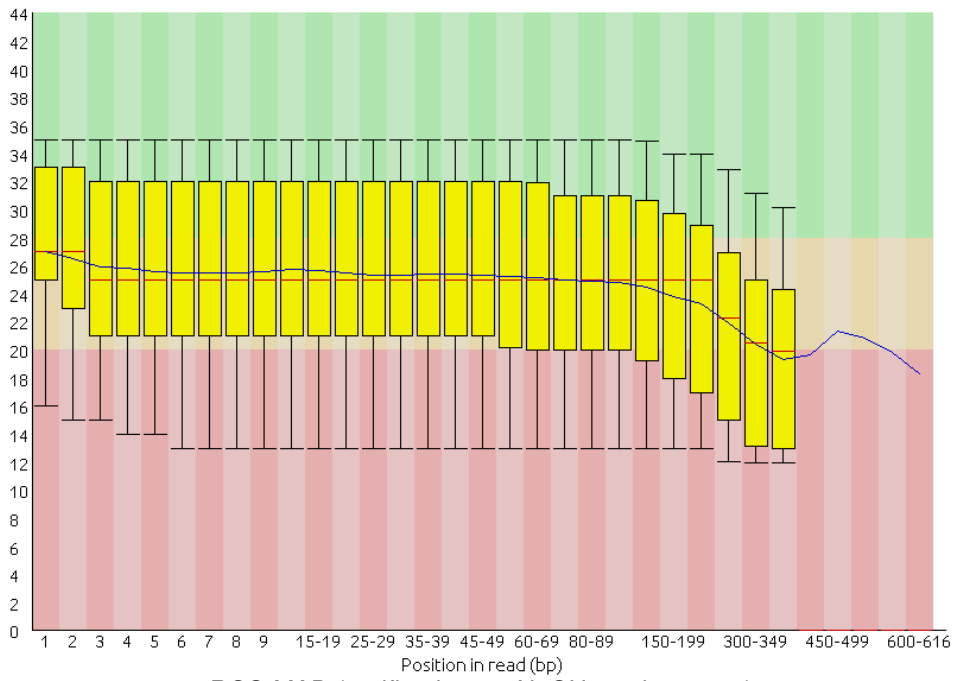
Usando o *plugin FastQC*, foi analisada a qualidade da base de acordo com a sua posição na *read* e a média e mediana de qualidade por base da *read* variou de Q20 a Q30 em todas as amostras. Também foi possível notar que as bases localizadas no final da *read* tenderam a apresentar uma qualidade inferior às demais (**Figura 26**). A diminuição da qualidade de detecção de bases no decorrer do sequenciamento é um fenômeno esperado nas plataformas de sequenciamento de segunda geração (152).

Figura 26 - Qualidade das bases das amostras de acordo com a sua posição na *read*.

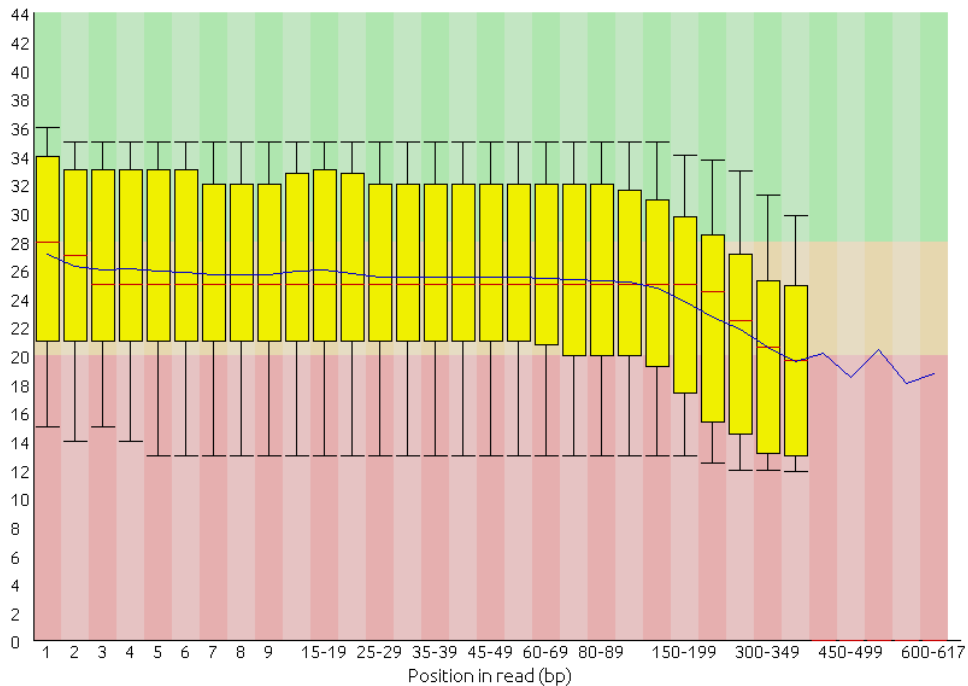




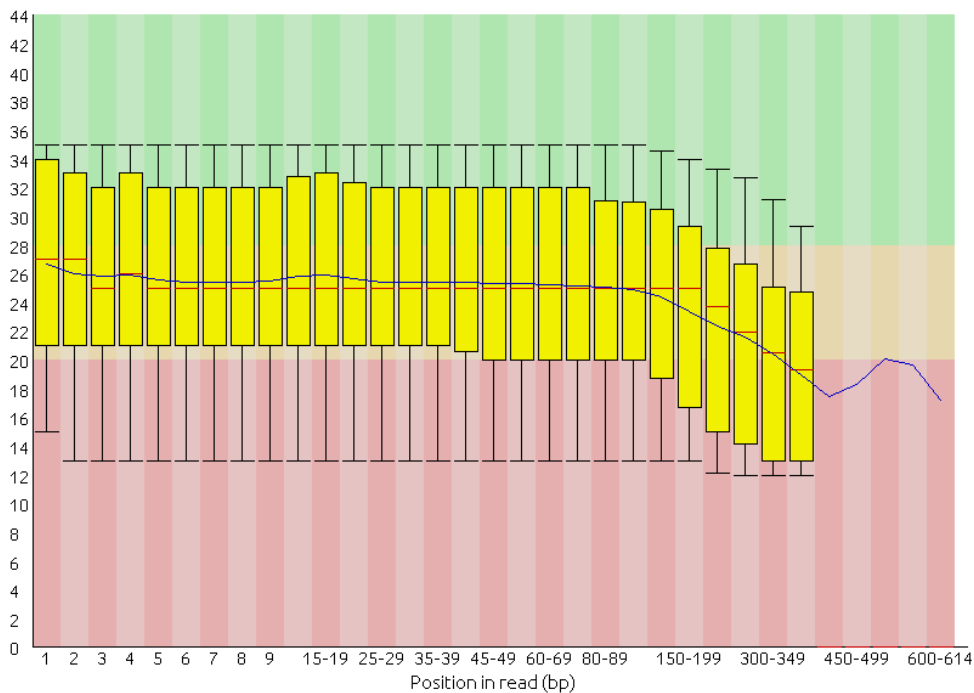
P-220 (purificada com NaOH)



ROS-MAD (purificada com NaOH e collagenase)



LC (purificada com NaOH e collagenase)



P-84 (purificada com NaOH e collagenase)

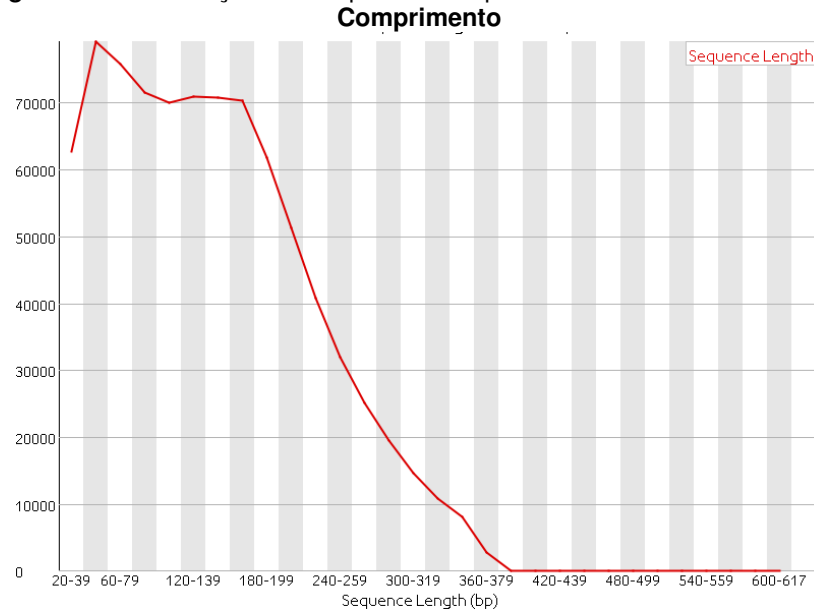
O eixo X mostra a posição em pb na *read*. O eixo Y mostra o escore de qualidade (Q) na escala *Phred*. Quanto maior o escore, melhor a qualidade da base. O fundo do gráfico divide as qualidades em três grupos: a qualidade ótima (verde), a qualidade média (laranja) e a qualidade baixa (vermelha).

A caixa amarela representa o inter-quartil (25-75%) dos índices de qualidade de todas das bases por posição na *read*. Os traços pretos marcam os limites de 10% a 90% desses índices. As linhas vermelhas representam a mediana e a linha azul representa a média.

Fonte: *FastQC*

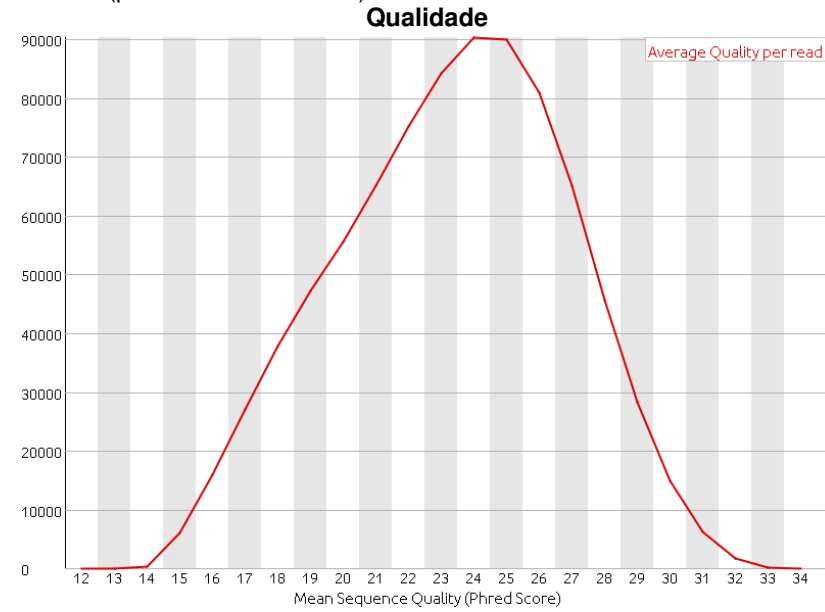
Utilizando a mesma ferramenta, foi avaliada a distribuição das *reads* conforme seu comprimento e a qualidade das *reads* (**Figura 27 a 31**). A amostra P-84 (purificada apenas com NaOH) obteve a pior distribuição com a moda de comprimento sendo menor que 100pb. A amostra ROS-MAD (purificada com NaOH e colagenase), por sua vez, apresentou a melhor distribuição, com a moda sendo cerca de 300pb. As demais amostras tiveram uma moda próxima de 150pb. Todos os comprimentos foram menores que o esperado, considerando o uso de um kit com uma meta de 400pb. Isso, porém, não prejudicou a qualidade dessas *reads* e, portanto, pode ser dada continuidade à análise.

Figura 27 – Distribuição de comprimento e qualidade de todas as *reads* da amostra P-84 (purificada com NaOH)



No eixo X está o comprimento das *reads* em pares de base. No eixo Y está a contagem de *reads*.

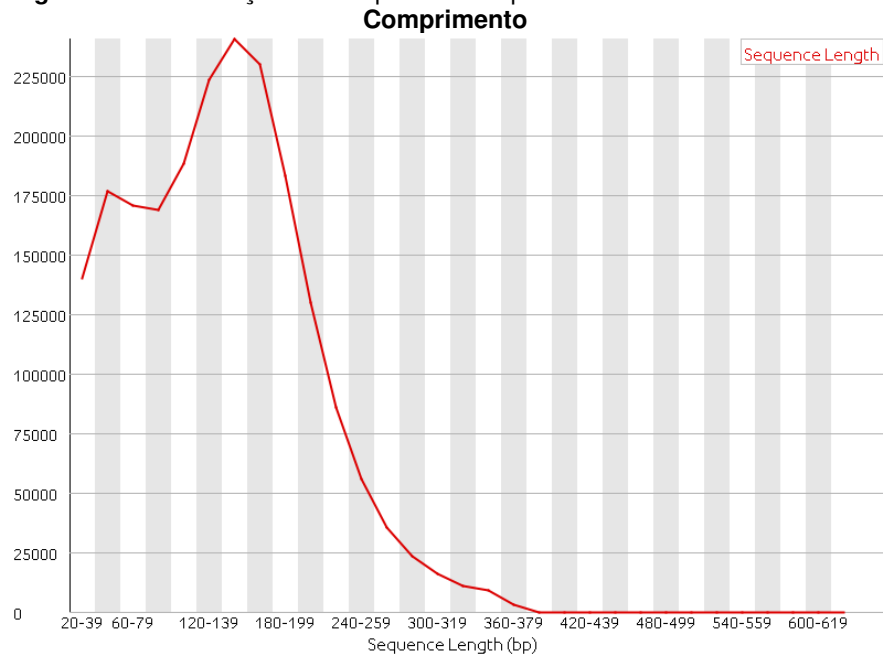
Fonte: *FastQC*



O eixo X é a média de qualidade da *read* na escala *Phred* e o eixo Y é a contagem de *reads*.

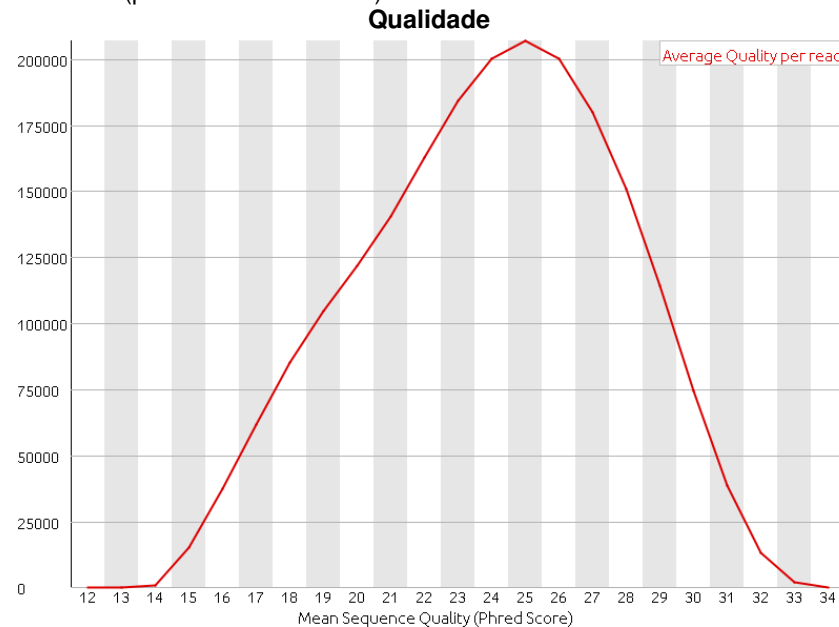
Fonte: *plugging FastQC*

Figura 28 – Distribuição de comprimento e qualidade de todas as *reads* da amostra P-220 (purificada com NaOH)



No eixo X está o comprimento das *reads* em pares de base. No eixo Y está a contagem de *reads*.

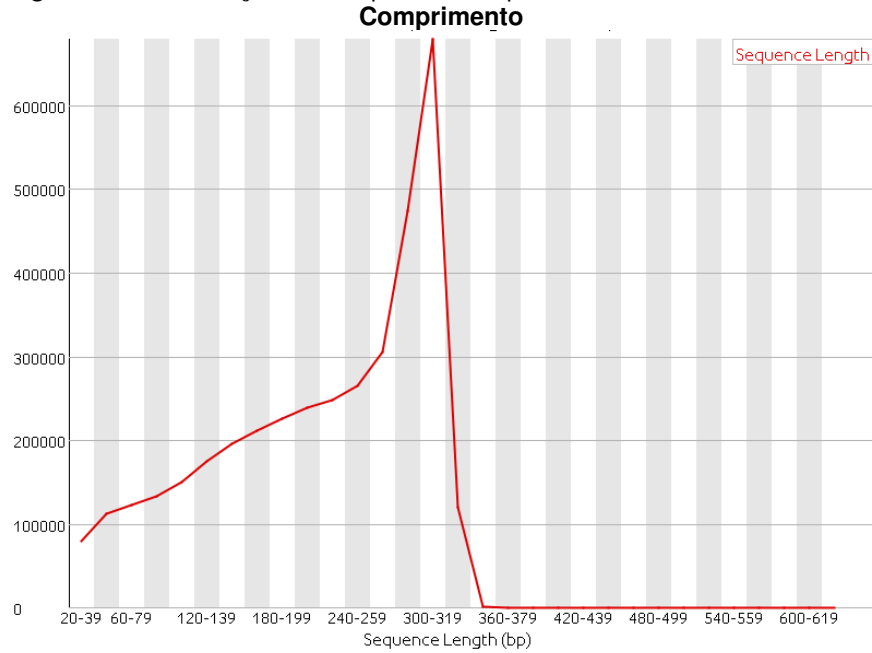
Fonte: *FastQC*



O eixo X é a média de qualidade da *read* na escala *Phred* e o eixo Y é a contagem de *reads*.

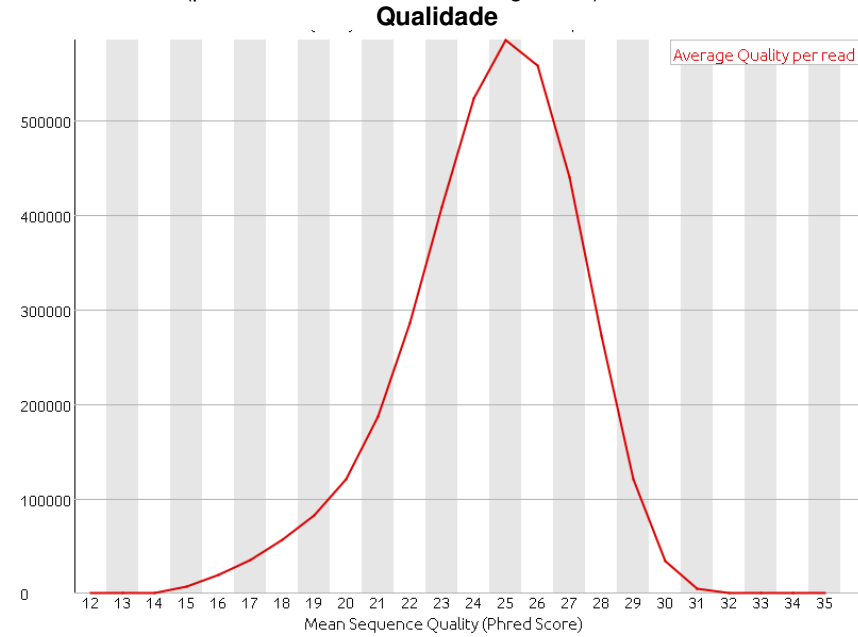
Fonte: *plugging FastQC*

Figura 29 – Distribuição de comprimento e qualidade de todas as *reads* da amostra ROS-MAD (purificada com NaOH e colagenase)



No eixo X está o comprimento das *reads* em pares de base. No eixo Y está a contagem de *reads*.

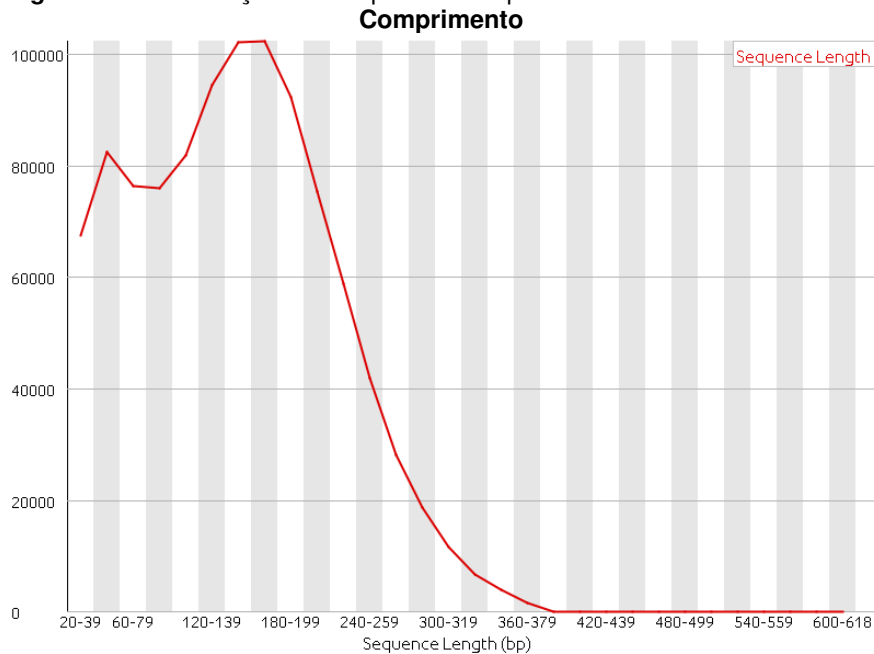
Fonte: *FastQC*



O eixo X é a média de qualidade da *read* na escala *Phred* e o eixo Y é a contagem de *reads*.

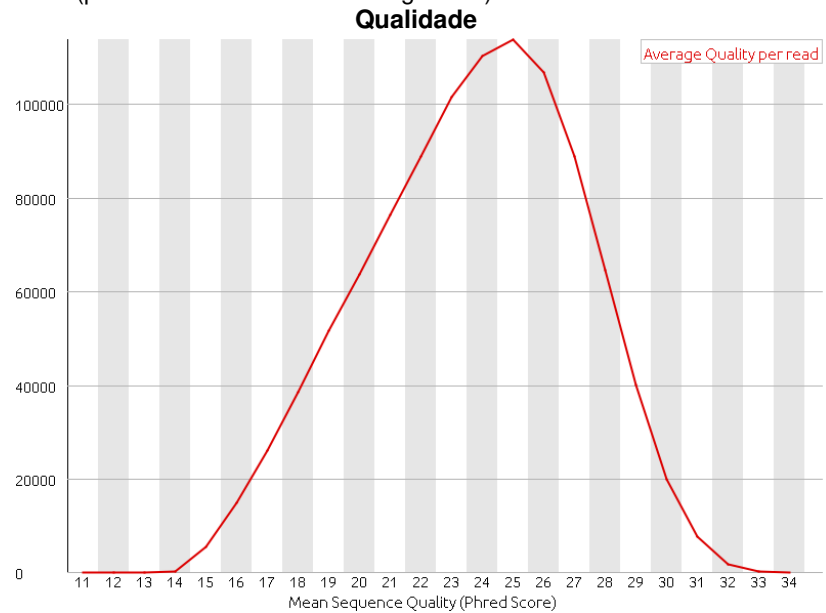
Fonte: *plugging FastQC*

Figura 30 – Distribuição de comprimento e qualidade de todas as *reads* da amostra LC (purificada com NaOH e colagenase)



No eixo X está o comprimento das *reads* em pares de base. No eixo Y está a contagem de *reads*.

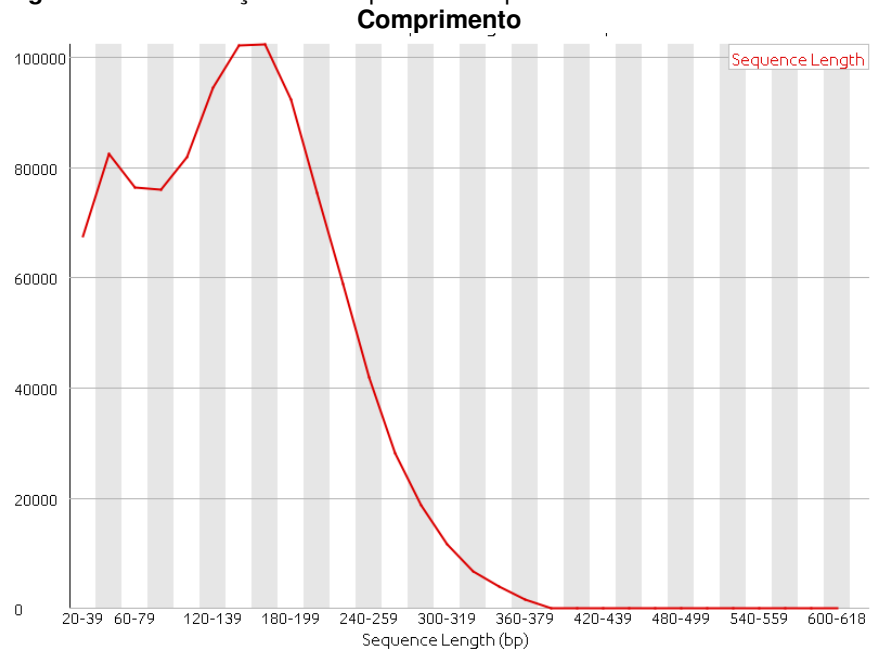
Fonte: *FastQC*



O eixo X é a média de qualidade da *read* na escala *Phred* e o eixo Y é a contagem de *reads*.

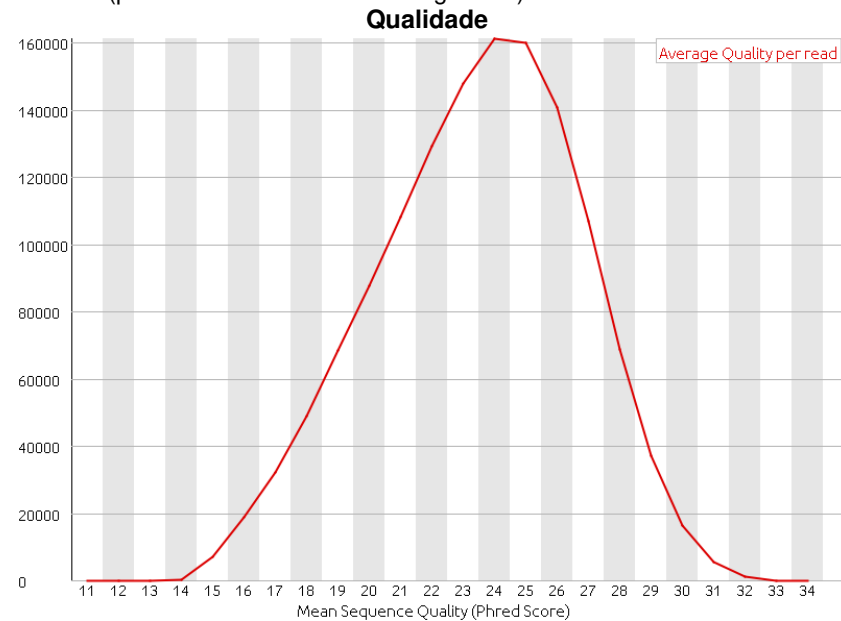
Fonte: *plugging FastQC*

Figura 31 – Distribuição de comprimento e qualidade de todas as *reads* da amostra P-84 (purificada com NaOH e colagenase)



No eixo X está o comprimento das *reads* em pares de base. No eixo Y está a contagem de *reads*.

Fonte: *FastQC*



O eixo X é a média de qualidade da *read* na escala *Phred* e o eixo Y é a contagem de *reads*.

Fonte: *plugging FastQC*

7.2.3. Análise do mapeamento com o *M. musculus*

Para as amostras purificadas apenas com NaOH (**Tabela 12**), mais de 95% das *reads* alinharam ao genoma do *M. musculus*. Isso demonstrou que esse método de purificação foi ineficiente e impossibilitou a análise das cepas de *M. leprae*. As amostras purificadas com NaOH e colagenase (**Tabela 12**), por sua vez, alinharam de 15 a 45% com o camundongo. Dessa forma, pode ser dado prosseguimento a análise às amostras purificadas com NaOH e colagenase. Dessa forma padronizamos uma metodologia de purificação do *M. leprae* adequada para o uso em sequenciamento em futuros trabalhos.

As análises demonstram que o método de purificação com NaOH e colagenase foi eficiente e pode ser utilizado em outros estudos que se proponham a sequenciar cepas *M. leprae* de forma a ter um melhor desempenho do processo, especialmente tendo em vista a alta complexidade e custo do cultivo *in vivo* desta bactéria. É possível que este método seja útil também para purificação de outras bactérias, especialmente mycobacterias, mas essa possibilidade ainda deve ser testada.

Tabela 12 – Resultados do alinhamento com *M. musculus*

Extração	Amostra	Total reads	Reads alinhadas (%)
NaOH	P-84	839.089	806.434 (96,11)
NaOH	P-220	2.097.474	2.032.776 (96,92)
colagenase-dispase e NaOH	ROS-MAD	3.745.807	1.696.931 (45,3)
colagenase-dispase e NaOH	LC	1.023.164	158.090 (15,45)
colagenase-dispase e NaOH	P-84	1.349.839	234.867 (17,4)

Legenda: Total de *reads*: Total de *reads* obtidas no sequenciamento para cada amostra;

7.2.4. Análise comparativa da montagem *de novo* e mapeamento com referência no software *MIRA*

Quando comparamos a montagem *de novo* (**Tabela 13**) com o mapeamento com o genoma de referência (**Tabela 14**) para as três amostras foi possível notar que a montagem *de novo* gerou de 200 a 300 *contigs* enquanto o mapeamento gerou um único *contig* com uma cobertura uniforme (**Figuras 32 e 33**). Essa diferença pode ser visualizada nas **Figuras 34 e 35**.

Tabela 13 – Resultados da montagem *de novo* no *software MIRA*

Amostra	ROS-MAD				LC				P-84			
Cobertura media (vezes)	120,88				33,48				43,85			
	Maiores contigs		Todos os contigs		Maiores contigs		Todos os contigs		Maiores contigs		Todos os contigs	
Número de contigs	211		7.263		295		1.298		320		2438	
Total consensus*	3.301.834	101,0%	6.526.908	199,7%	3.275.894	100,2%	3.639.259	111,4%	3.300.281	101,0%	4.124.805	126,2%
Maior contig*	83.927	2,6%	83.927	2,6%	111.222	3,4%	111.222	3,4%	101.949	3,1%	101.949	3,1%
Tamanho contig N50*	30.429	0,9%	1.888	0,1%	22.011	0,7%	19.356	0,6%	20.686	0,6%	15.209	0,5%
Tamanho contig N90*	8.834	0,3%	372	0,0%	6.155	0,2%	639	0,0%	5.602	0,2%	417	0,0%
Tamanho contig N95*	5.759	0,2%	335	0,0%	3.319	0,1%	388	0,0%	3.278	0,1%	339	0,0%
Media da qualidade do consenso**	78		46		75		46		76		44	
Tempo de alinhamento***	6h36m28s				51m22s				1h25m43ds			

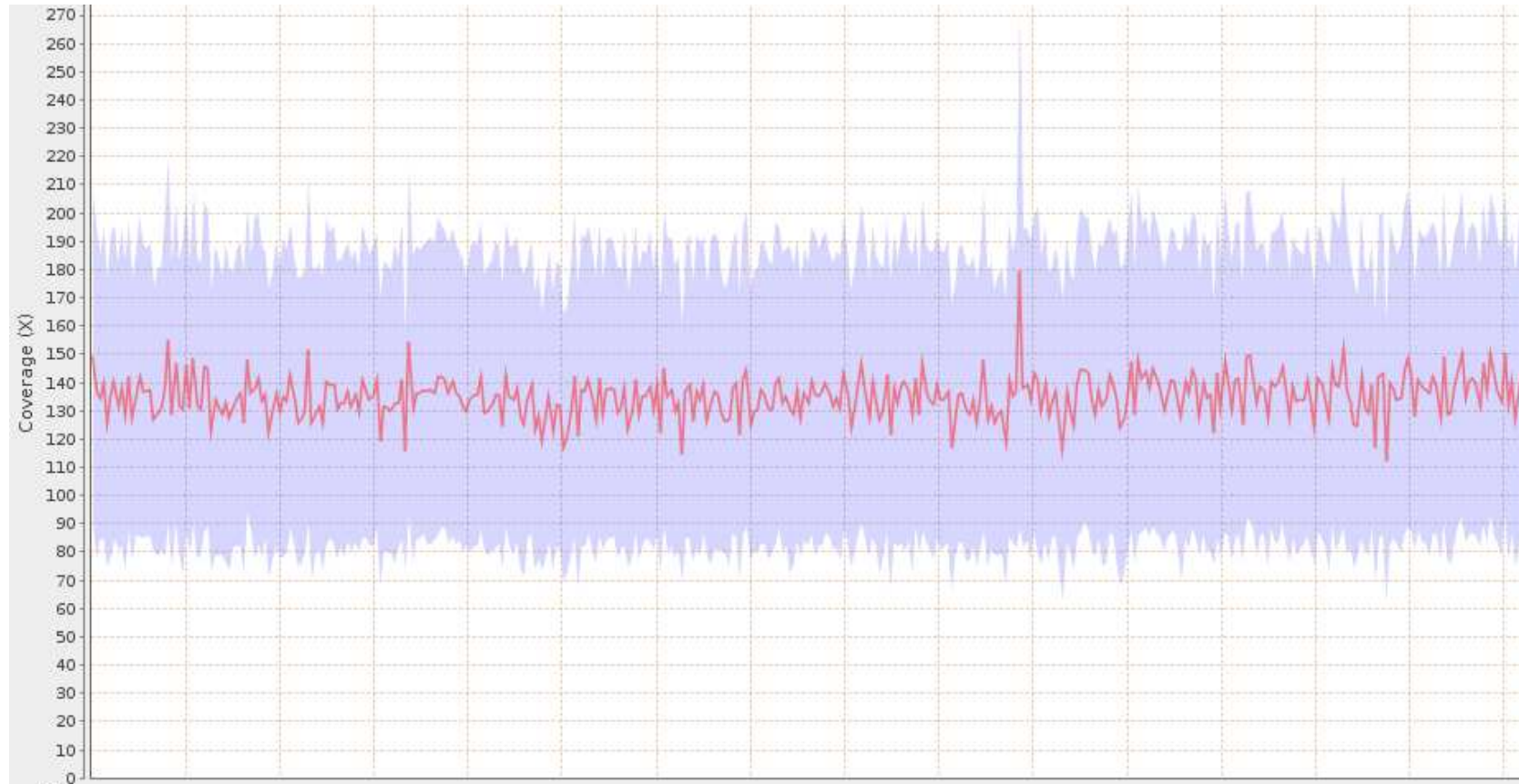
Legenda: *Porcentagem baseada no tamanho do genoma de referência (3.268.203 pb). **Qualidade calculada com método próprio do *software MIRA*. *** Tempo de alinhamento considerando o comando apresentado no item 5.5.3.2 e o uso da função “time” em linha de comando. Por se tratar de uma montagem de novo o tamanho do genoma obtido pode ser maior do que o da referência e, portanto, a porcentagem obtida pode exceder 100%

Tabela 14 – Resultados do mapeamento no *software MIRA*

Amostra	ROS-MAD				LC				P-84			
	1.902.646				747.455				968.602			
Cobertura media (vezes)	137,03				38,09				50,08			
	Maiores contigs		Todos os contigs		Maiores contigs		Todos os contigs		Maiores contigs		Todos os contigs	
Número de contigs	1		1		1		1		1		2438	
Total consensus*	3.268.580	100,01%	3.268.580	100,01%	3.267.974	99,99%	3.267.974	99,99%	3.267.936	99,99%	3.267.936	99,99%
Maior contig*	3.268.580	100,01%	3.268.580	100,01%	3.267.974	99,99%	3.267.974	99,99%	3.267.936	99,99%	3.267.936	99,99%
Tamanho contig N50*	3.268.580	100,01%	3.268.580	100,01%	3.267.974	99,99%	3.267.974	99,99%	3.267.936	99,99%	3.267.936	99,99%
Tamanho contig N90*	3.268.580	100,01%	3.268.580	100,01%	3.267.974	99,99%	3.267.974	99,99%	3.267.936	99,99%	3.267.936	99,99%
Tamanho contig N95*	3.268.580	100,01%	3.268.580	100,01%	3.267.974	99,99%	3.267.974	99,99%	3.267.936	99,99%	3.267.936	99,99%
Media da qualidade do consenso**	74		74		75		75		74		74	
Tempo de alinhamento***	6h15m50s				1h27m48s				2h33m51s			

Legenda: *Porcentagem baseada no tamanho do genoma de referência (3.268.203 pb). **Qualidade calculada com método próprio do *software MIRA*. *** Tempo de alinhamento considerando o comando apresentado no item 5.5.3.2 e o uso da função “time” em linha de comando.

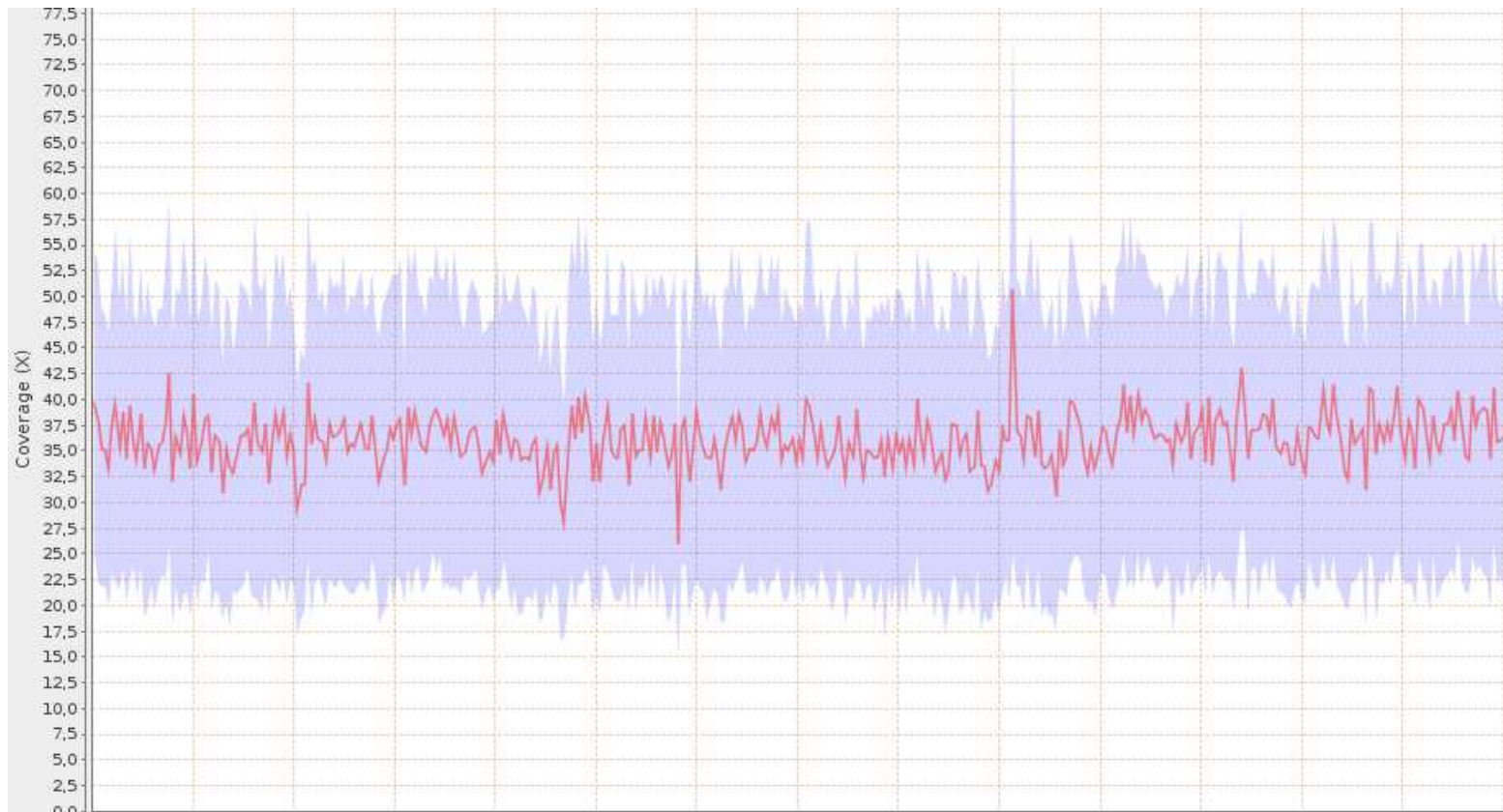
Figura 32 – Cobertura ao longo do genoma no mapeamento da amostra ROS-MAD no *software* MIRA



Legenda: O eixo x representa todo o genoma de referência; A linha vermelha representa a cobertura média obtida em cada base e a área roxa representa o desvio padrão da cobertura para cada base.

Fonte: a autora utilizando *software* Qualimap (2018)

Figura 33 – Cobertura ao longo do genoma no mapeamento da amostra LC no *software* MIRA



Legenda: O eixo x representa todo o genoma de referência; A linha vermelha representa a cobertura média obtida em cada base e a área roxa representa o desvio padrão da cobertura para cada base.

Fonte: a autora utilizando *software* Qualimap (2018)

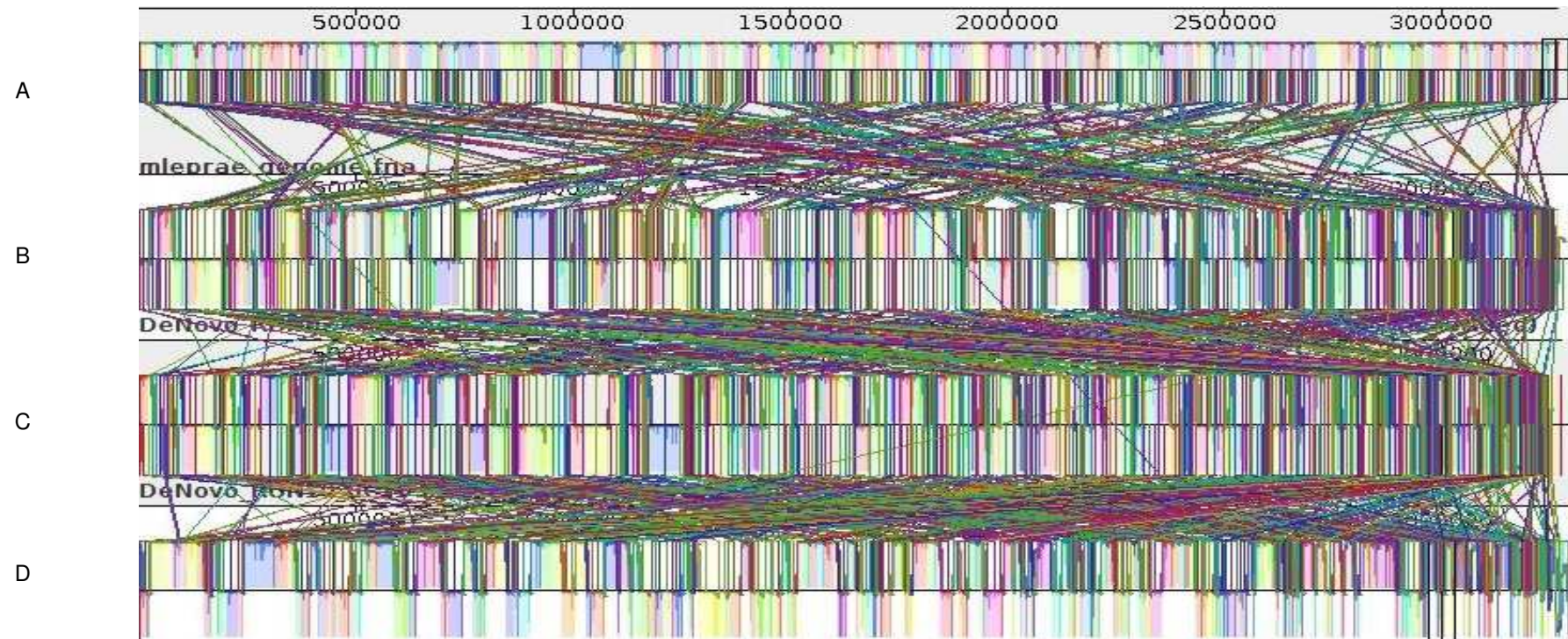
Figura 34 – Cobertura ao longo do genoma no mapeamento da amostra P-84 no *software* MIRA



Legenda: O eixo x representa todo o genoma de referência; A linha vermelha representa a cobertura média obtida em cada base e a área roxa representa o desvio padrão da cobertura para cada base.

Fonte: a autora utilizando *software* Qualimap (2018)

Figura 35 – Comparação dos genomas obtidos na montagem *de novo* entre si e com o genoma de referência

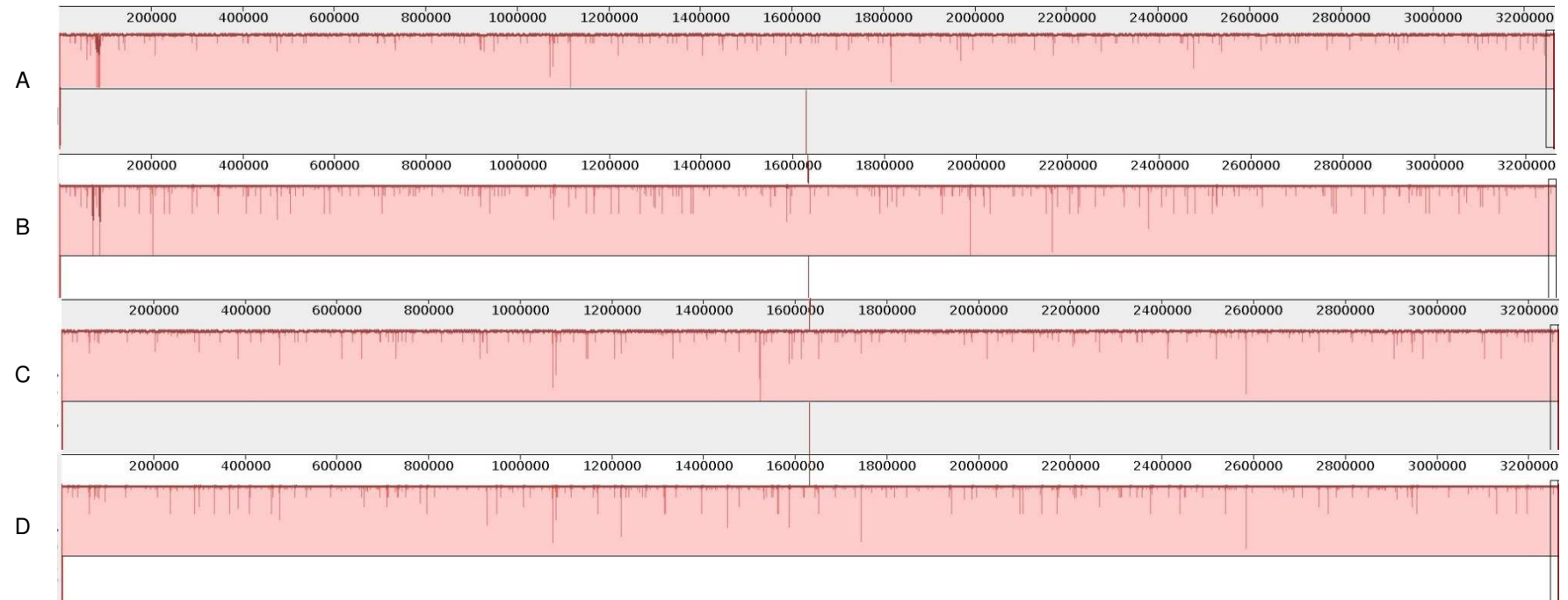


Legenda: Cada bloco de sequencias homologas é representado por uma cor, as linhas conectam dois blocos homólogos entre genomas.

A) Genoma de referência do *M. leprae* (cepa TN); B) Amostra ROS-MAD; C) Amostra LC; D) Amostra P-84;

Fonte: a autora utilizando *software Mauve* (2018)

Figura 36 – Comparação dos genomas obtidos no mapeamento (com o *software MIRA*) entre si e com o genoma de referência



Legenda: Cada bloco de sequencias homologas é representado por uma cor, as linhas conectam dois blocos homólogos entre genomas.

A) Genoma de referência do *M. leprae* (cepa TN); B) Amostra ROS-MAD; C) Amostra LC; D) Amostra P-84;

Fonte: a autora utilizando *software Mauve* (2018)

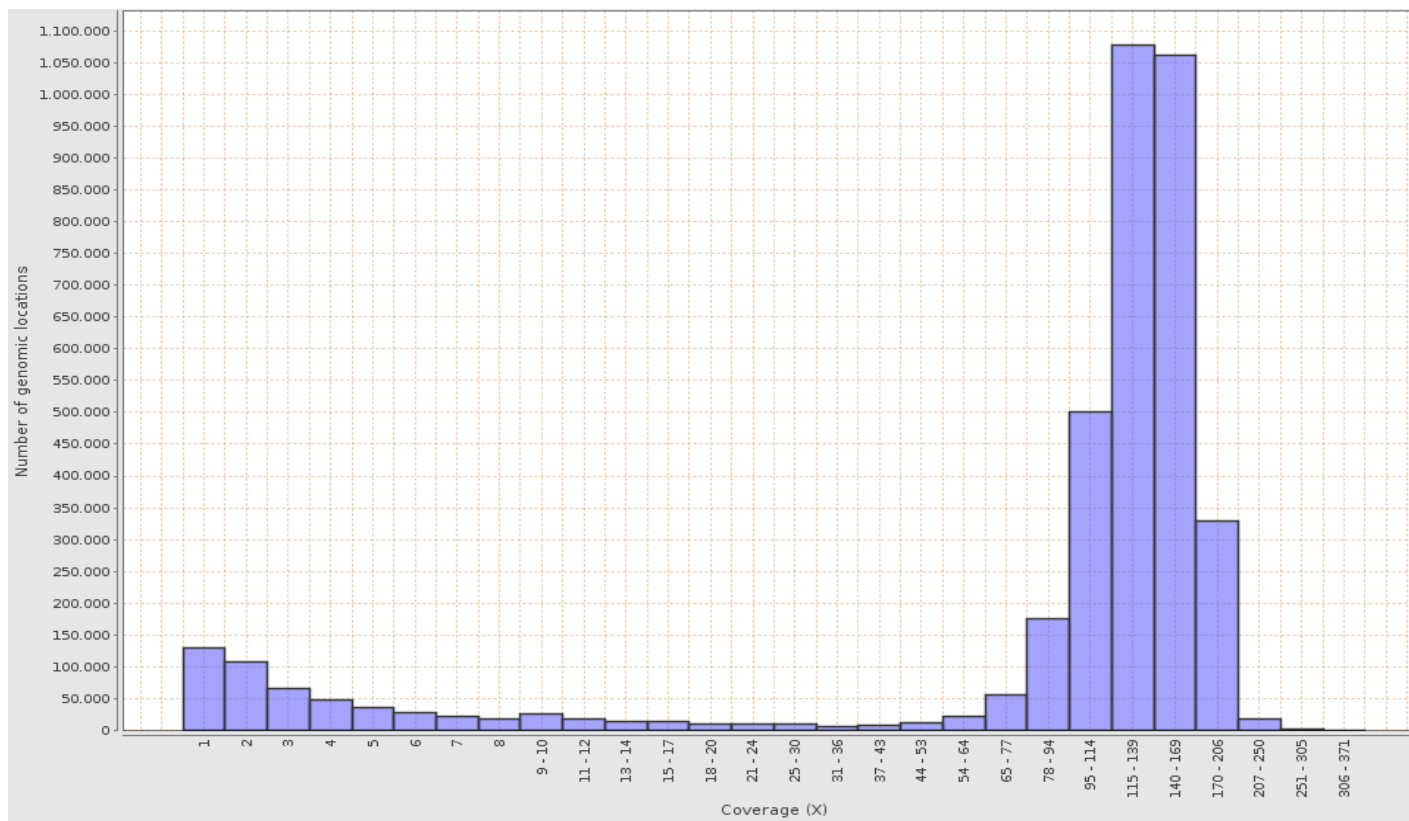
7.2.5. Análise comparativa da qualidade dos mapeamentos com referência

Comparando os mapeamentos realizados nos *softwares MIRA* e *TMAP*, é possível observar que para todas as amostras o *software MIRA* teve um maior número de *reads* mapeadas do que o *software TMAP*. Porém a média de cobertura obtida foi muito parecida e no geral o desvio padrão da cobertura foi ligeiramente menor para o *software TMAP* (**Tabela 15**). A qualidade do mapeamento não pode ser comparada devido aos diferentes algoritmos utilizados para avaliação de qualidade nos dois programas, no entanto foi possível comparar a distribuição da cobertura obtida entre os mapeamentos para cada amostra (**Figuras 37 a 42**) que se mostraram bastante semelhante, tendo o mapeamento com *MIRA* apresentado maior número de regiões de baixa cobertura, o que condiz com o maior desvio padrão da cobertura. Dessa forma o mapeamento com *TMAP* foi considerado melhor para dar continuidade as análises.

Tabela 15 – Resultados comparativos do mapeamento com *M. leprae*

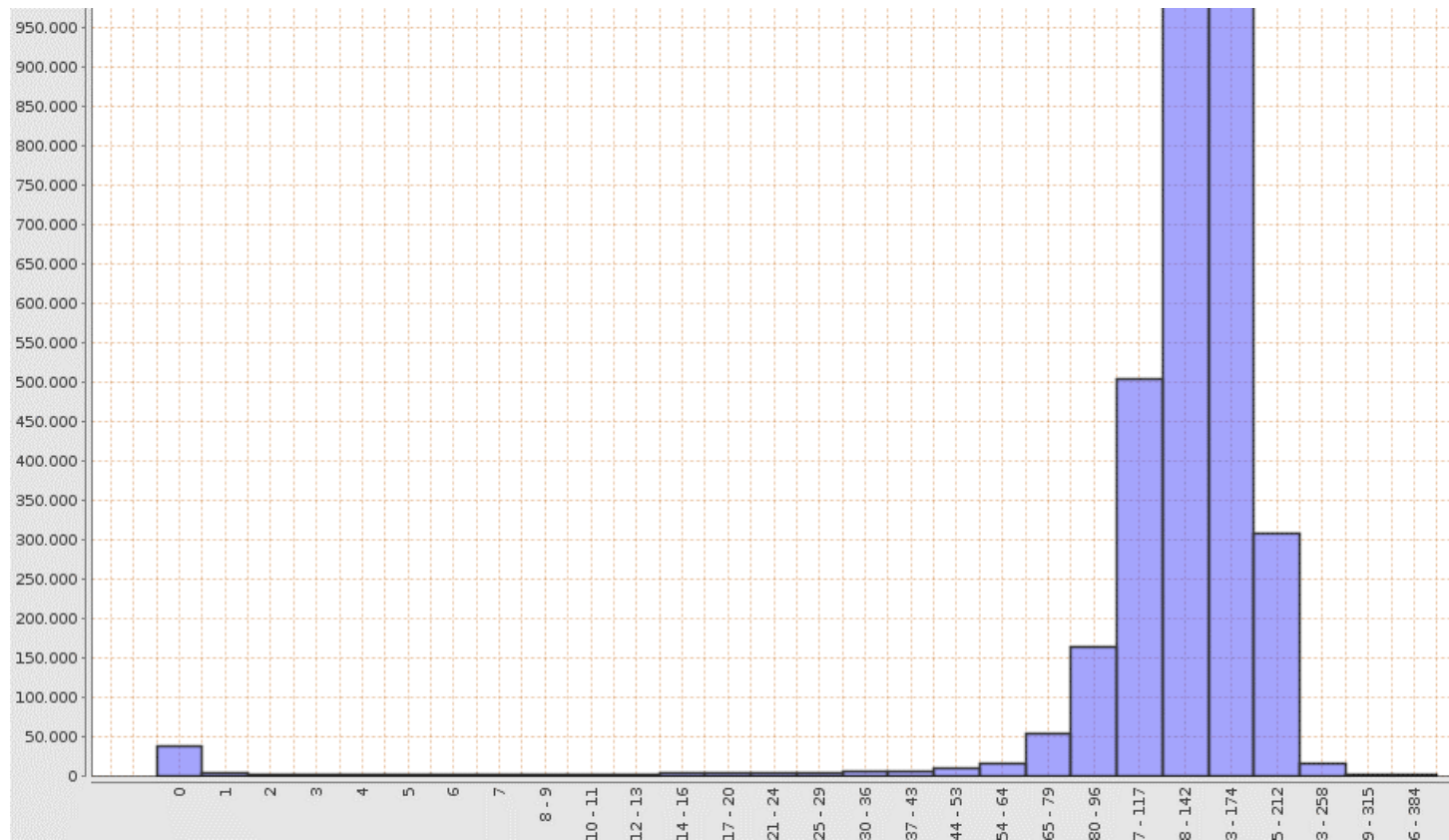
Amostra	Mapeamento <i>MIRA</i>			Mapeamento <i>TMAP</i>		
	ROS-MAD	LC	P-84	ROS-MAD	LC	P-84
Reads mapeadas	2.724.801	1.266.899	1.596.841	1.976.587	833.090	1.069.814
Media de cobertura	135,033	36,0892	48,0828	135,9446	36,589	48,6438
Desvio padrão da cobertura	56,2235	13,795	20,1294	34,9376	11,8177	16,6731

Figura 37 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra ROS-MAD no *software* MIRA



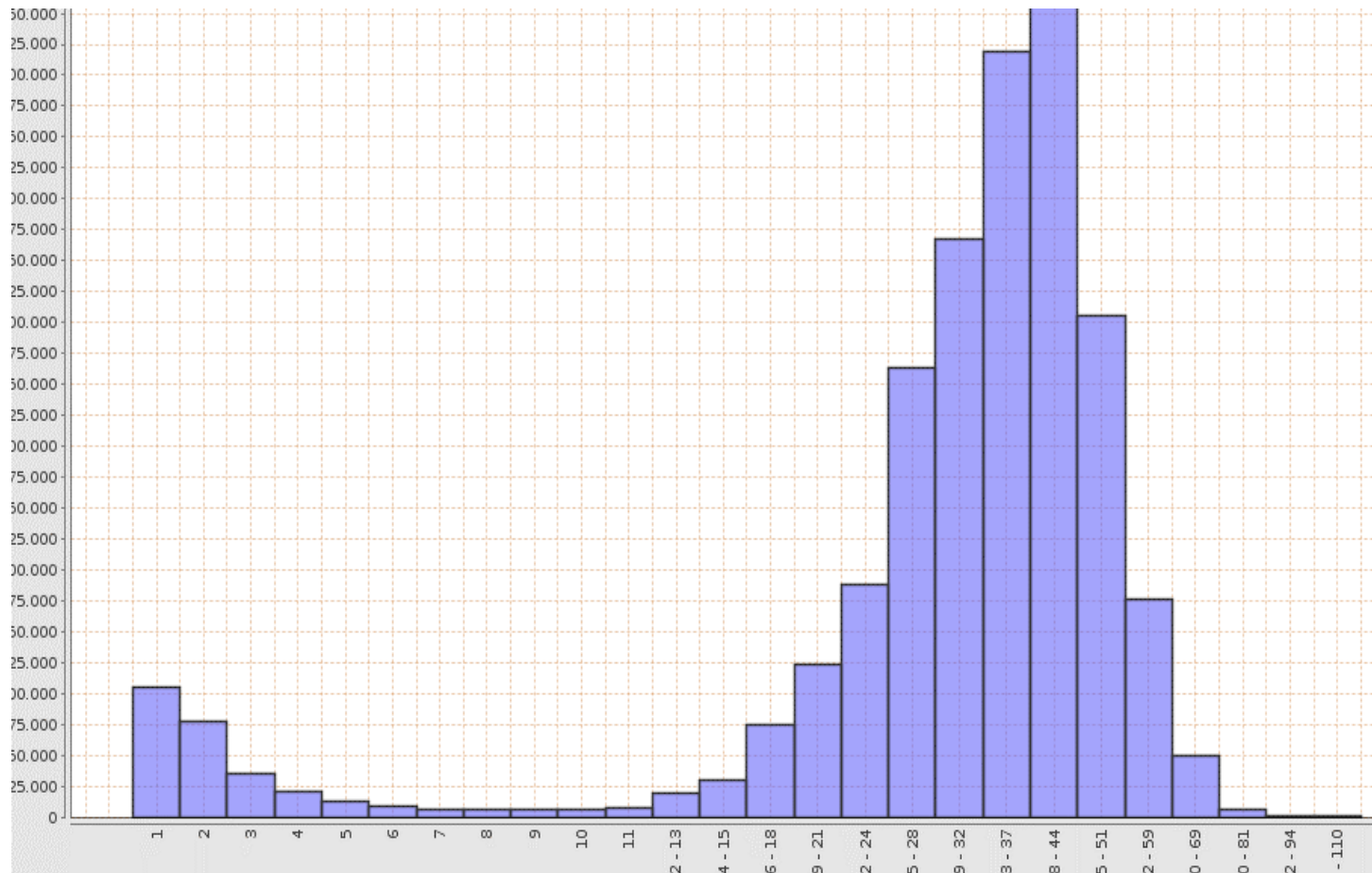
Fonte: a autora utilizando *software* Qualimap (2018)

Figura 38 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra ROS-MAD no *software* TMAP



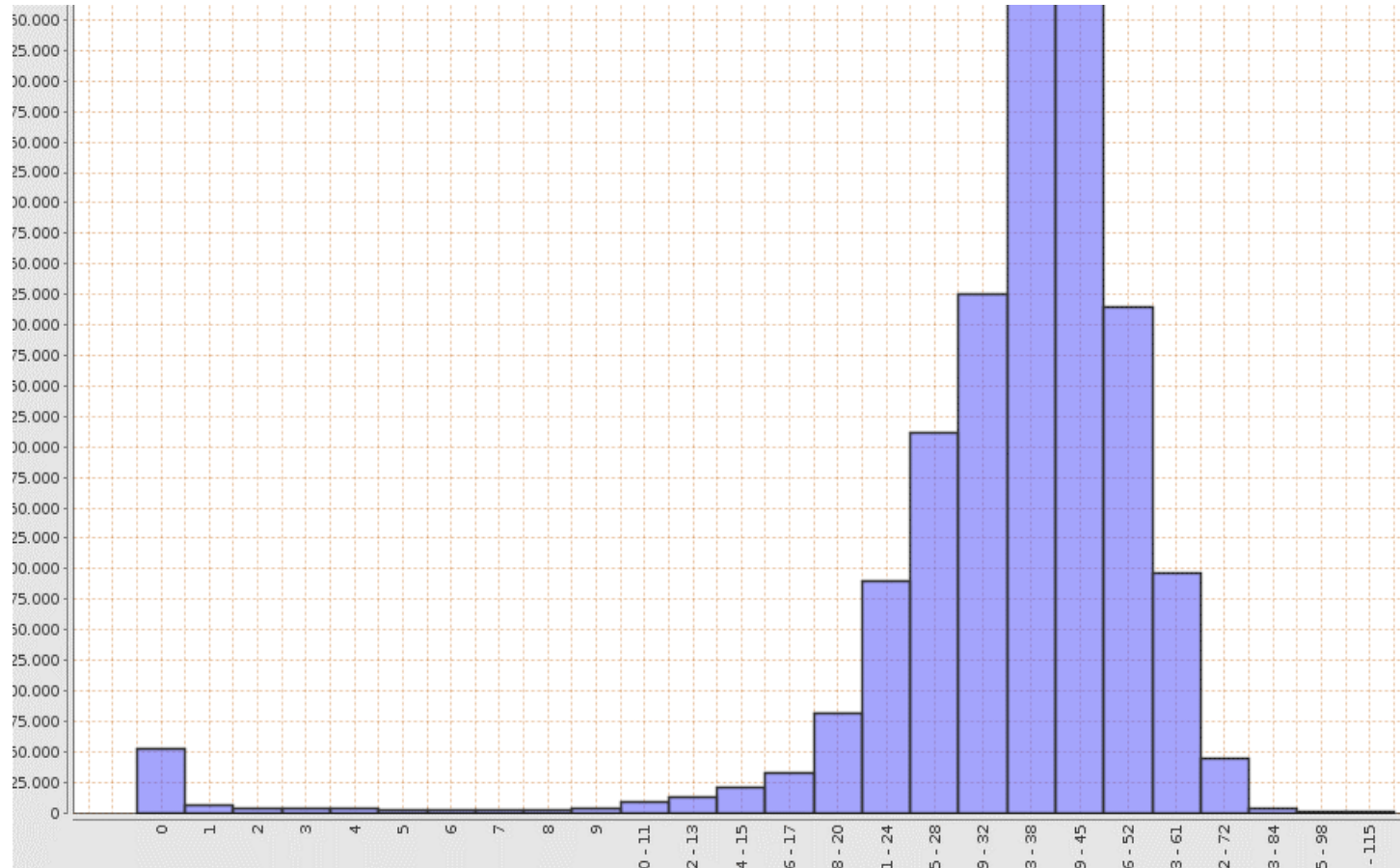
Fonte: a autora utilizando *software* Qualimap (2018)

Figura 39 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra LC no *software* MIRA



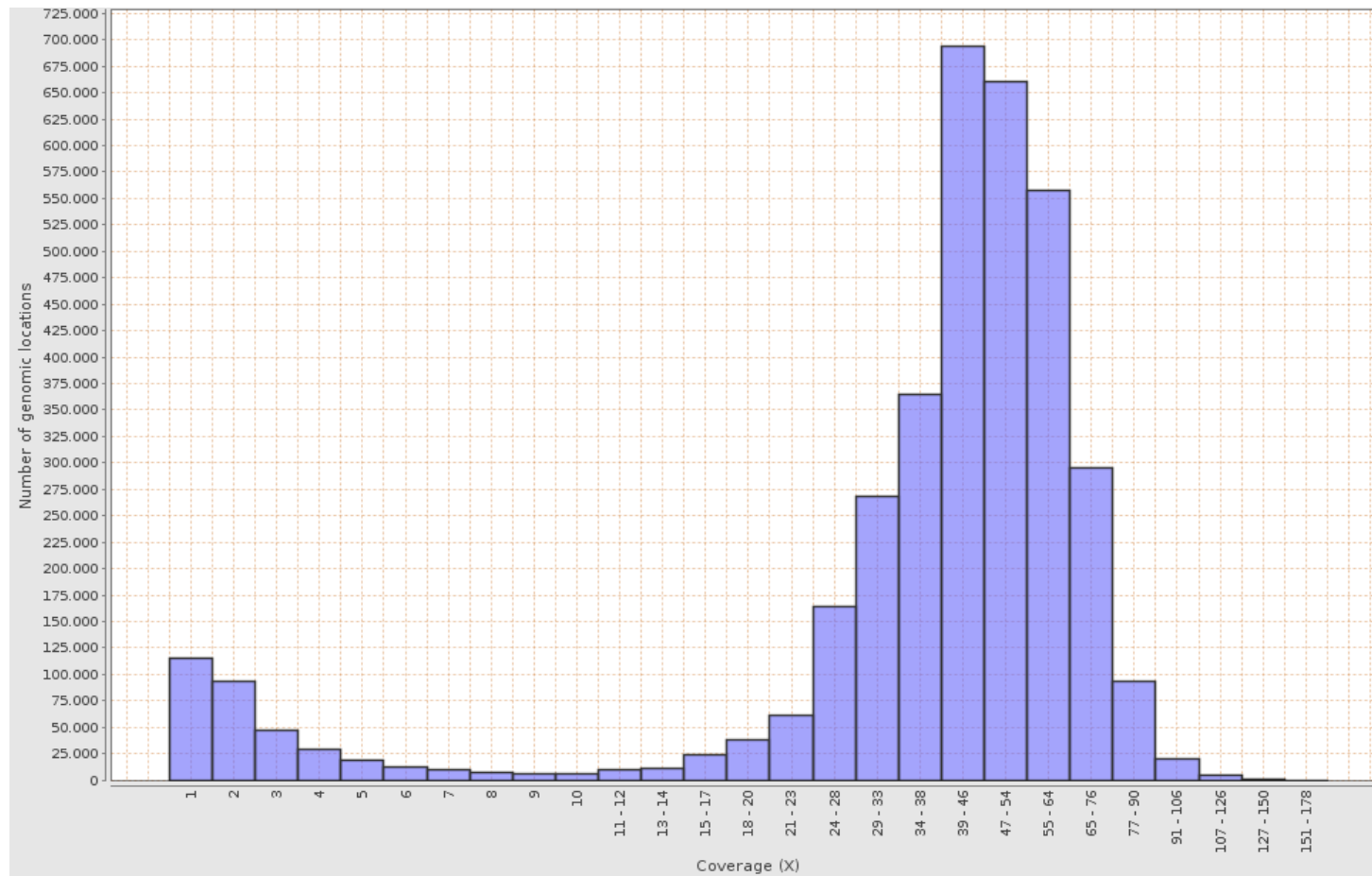
Fonte: a autora utilizando *software* Qualimap (2018)

Figura 40 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra LC no *software TMAP*



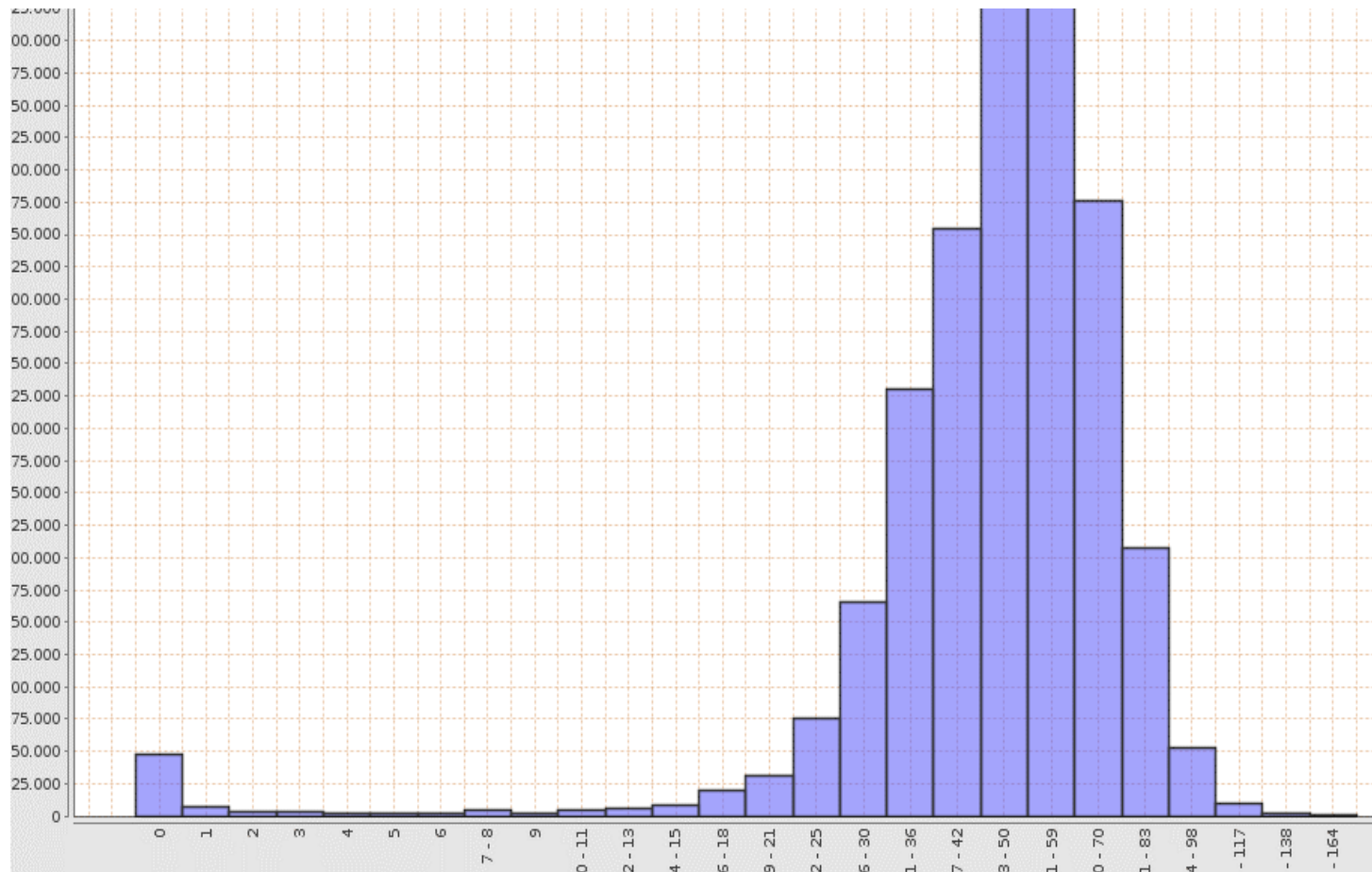
Fonte: a autora utilizando *software Qualimap* (2018)

Figura 41 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra P-84 no *software* MIRA



Fonte: a autora utilizando *software Qualimap* (2018)

Figura 42 – Histograma do número de regiões genômicas por cobertura no mapeamento da amostra P-84 no *software* TMAP

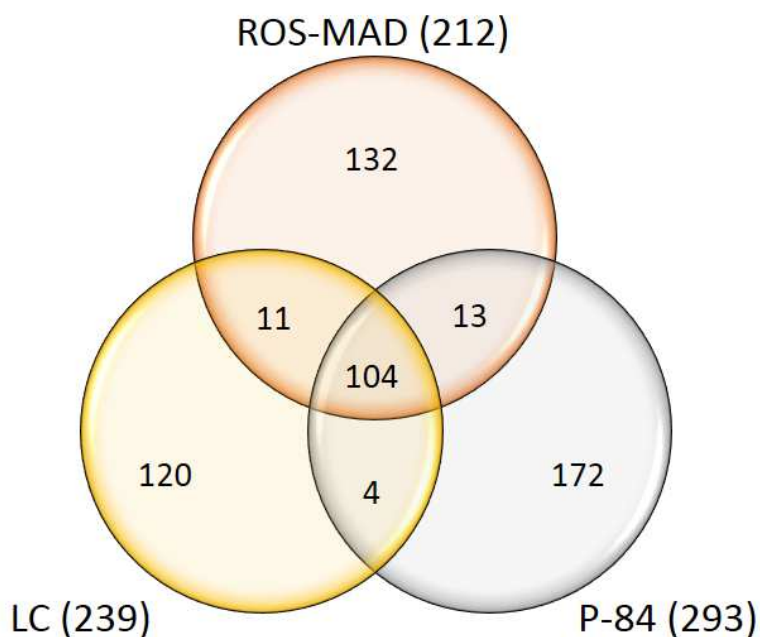


Fonte: a autora utilizando *software* Qualimap (2018)

7.2.6. Análise de variantes

A chamada de variantes foi feita com o *software TVC* a partir do mapeamento com o *TMAP*. Foram encontradas 212 variantes para a amostra ROS-MAD, 239 para a amostra LC e 293 para a amostra P-84. O número de variantes em comum entre as amostras está descrito de forma esquemática na **Figura 43**.

Figura 43 – Representação esquemática das variantes encontradas nas amostras analisadas



Fonte: a autora.

A detecção de variantes nos genes de resistência (*folP*, *rpoB* e *gyrA*) foi coerente com as mutações identificadas no sequenciamento por *Sanger*, o que as valida. Padronizadas as etapas de análise primária, será possível realizar análises de alta complexidade desses e de futuros genomas, tais como filtragem fina das variantes, previsão computacional de impacto funcional, comparação da conservação dos nucleotídeos entre cepas e espécies disponíveis em bases de dados, anotação genômica, etc.

8. CONCLUSÕES COMPLEMENTARES

O protocolo de sequenciamento foi padronizado e foi constatada que a purificação com NaOH e colagenase foi o método de purificação mais adequado. Constatamos que o mapeamento com referência foi mais adequado que a montagem *de novo*. E concluímos que o mapeamento no *software MIRA* permite uma melhor visualização dos *contigs*, mas o mapeamento com o *software TMAP* foi melhor para a obtenção de variantes.

REFERÊNCIAS

1. Robbins G, Tripathy VM, Misra VN, Mohanty RK, Shinde VS, Gray KM, et al. Ancient skeletal evidence for leprosy in India (2000 B.C.). *PloS one*. 2009;4(5):e5669.
2. Clapasson A, Canata S. Microbiology. In: Nunzi E, Massone C, editors. *Leprosy*. Italia: Springer-Verlag; 2012. p. 15-8.
3. National Center for Biotechnology Information. *Mycobacterium leprae*: Causative agent of human leprosy 2018 [24/01/2018]. Available from: <https://www.ncbi.nlm.nih.gov/genome/903>.
4. Sauer ME, Salomao H, Ramos GB, D'Espindula HR, Rodrigues RS, Macedo WC, et al. Genetics of leprosy: expected and unexpected developments and perspectives. *Clinics in dermatology*. 2015;33(1):99-107.
5. Nunzi E, Massone C, Noto S. Diagnostic Work-up. In: Nunzi E, Massone C, editors. *Leprosy*. Italy: Springer-Verlag; 2012.
6. World Health Organization. Chemotherapy of leprosy for control programmes. World Health Organization technical report series. 1982;675:1-33.
7. World Health Organization. A guide for surveillance of antimicrobial resistance in leprosy: 2017 update. 2017.
8. Clapasson A, Canata S. Laboratory Investigations. In: Nunzi E, Massone C, editors. *Leprosy*. Italia: Springer-Verlag; 2012. p. 55-62.
9. Talhari S, Goncalves Hd, Pontes MAA, Silveira MIS, Orsi AT. Diagnóstico. In: Talhari S, Penna GO, Goncalves Hd, Oliveira MLW, editors. *Hanseníase*. Manaus: DiLivros; 2015.
10. Brasil MS. Diretrizes para vigilância, atenção e eliminação da Hanseníase como problema de saúde pública : manual técnico-operacional. In: *Doenças MdSSdVeSDdVd, Transmissíveis*, editors. Brasília2016.
11. Santos FSDd, Souza LPAd, Siani AC. [Chaulmoogra oil as scientific knowledge: the construction of a treatment for leprosy]. *História, Ciências, Saúde-Manguinhos*. 2008;15:29-46.
12. Noordeen SK. History of chemotherapy of leprosy. *Clinics in dermatology*. 2016;34(1):32-6.
13. Faget GH, Johansen FA, Ross H. Public Health Reports, December 11, 1942: Sulfanilamide in the treatment of leprosy. *Public health reports*. 1942;90(6):486-9.
14. Faget GH, Pogge RC, Johansen FA, Dinan JF, Prejean BM, Eccles CG. The prominent treatment of leprosy: a progress report. *Public health reports*. 1943;58(48):1729-41.
15. Faget GH, Johansen FA, Ross H. Public Health Reports, December 11, 1942: Sulfanilamide in the treatment of leprosy. *Public health reports*. 1975;90(6):486-9.

16. Lowe J. Treatment of leprosy with diamino-diphenyl sulphone by mouth. *Lancet*. 1950;1(6596):145-50.
17. Gelber RH, Grosset J. The chemotherapy of leprosy: an interpretive history. *Leprosy review*. 2012;83(3):221-40.
18. World Health Organization. EXPERT committee on leprosy; first report. World Health Organization technical report series. 1953;71:1-28.
19. World Health Organization. WHO expert committee on leprosy: Second Report. World Health Organization technical report series. 1959;189.
20. Seydel JK, Richter M, Wempe E. Mechanism of action of the folate blocker diaminodiphenylsulfone (dapson, DDS) studied in *E. coli* cell-free enzyme extracts in comparison to sulfonamides (SA). *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1980;48(1):18-29.
21. Bergwerf H. MolView: an attempt to get the cloud into chemistry classrooms. DivCHED CCCE: Committee on Computers in Chemical Education. 2015.
22. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*. 2014;42(Web Server issue):W252-8.
23. Erickson PT. Relapse following apparent arrest of leprosy by sulfone therapy. *Public health reports*. 1950;65(36):1147-57.
24. Shepard CC. The Experimental Disease That Follows the Injection of Human Leprosy Bacilli into Foot-Pads of Mice. *The Journal of experimental medicine*. 1960;112(3):445-54.
25. Pettit JH, Rees RJ. Sulphone Resistance in Leprosy. An Experimental and Clinical Study. *Lancet*. 1964;2(7361):673-4.
26. Pearson JM, Rees RJ, Waters MF. Sulphone resistance in leprosy. A review of one hundred proven clinical cases. *Lancet*. 1975;2(7924):69-72.
27. Gelber RH, Rea TH, Murray LP, Siu P, Tsang M, Byrd SR. Primary dapson-resistant Hansen's disease in California. Experience with over 100 *Mycobacterium leprae* isolates. *Archives of dermatology*. 1990;126(12):1584-6.
28. Pettit JH, Rees RJ. Studies on sulfone resistance in leprosy. 2. Treatment with a riminophenazine derivative (B.663). *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1966;34(4):391-7.
29. Maddison JE, Watson ADJ, Elliott J. Chapter 8 - Antibacterial drugs. *Small Animal Clinical Pharmacology (Second Edition)*. Edinburgh: W.B. Saunders; 2008. p. 148-85.
30. Browne SG, Harman DJ, Waudby H, McDougall AC. Clofazimine (Lamprene, B663) in the treatment of lepromatous leprosy in the United Kingdom. A 12 year review of 31 cases, 1966-1978. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1981;49(2):167-76.
31. Rees RJ, Pearson JM, Waters MF. Experimental and clinical studies on rifampicin in treatment of leprosy. *British medical journal*. 1970;1(5688):89-92.

32. National Center for Biotechnology Information. Genome Assembly and Annotation report: *Mycobacterium leprae* TN - NC_002677.1: NCBI; 2017 [updated 26/08/2016; cited 2017 18/09/2017]. Available from: http://www.ncbi.nlm.nih.gov/genome/903?genome_assembly_id=169891.
33. Maggi N, Pasqualucci CR, Ballotta R, Sensi P. Rifampicin: A New Orally Active Rifamycin. *Chemotherapy*. 1966;11(5):285-92.
34. Jacobson RR, Hastings RC. Rifampin-resistant leprosy. *Lancet*. 1976;2(7998):1304-5.
35. Pearson JM, Haile GS, Rees RJ. Primary dapsone-resistant leprosy. *Leprosy review*. 1977;48(2):129-32.
36. Guinto RS, Cellona RV, Fajardo TT, de la Cruz EC. Primary dapsone-resistant leprosy in Cebu, Philippines. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1981;49(4):427-30.
37. Guelpa-Lauras CC, Grosset JH, Constant-Desportes M, Brucker G. Nine cases of rifampin-resistant leprosy. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1984;52(1):101-2.
38. WHO Expert Committee on Leprosy. Sixth Report—TRS 768. World Health Organization technical report series. 1988;768:1-51.
39. World Health Organization Study Group. Chemotherapy of leprosy - TRS 847. World Health Organization technical report series. 1994;847:1-24.
40. WHO Expert Committee on Leprosy. Seventh Report—TRS 874. World Health Organization technical report series. 1998;874:1-43.
41. Matsuoka M. Drug resistance in leprosy. *Japanese journal of infectious diseases*. 2010;63(1):1-7.
42. Saito H, Tomioka H, Nagashima K. In vitro and in vivo activities of ofloxacin against *Mycobacterium leprae* infection induced in mice. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1986;54(4):560-2.
43. Shetty VP, Uplekar MW, Antia NH. Primary resistance to single and multiple drugs in leprosy--a mouse footpad study. *Leprosy review*. 1996;67(4):280-6.
44. Cambau E, Perani E, Guillemin I, Jamet P, Ji B. Multidrug-resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. *Lancet*. 1997;349(9045):103-4.
45. Kai M, Matsuoka M, Nakata N, Maeda S, Gidoh M, Maeda Y, et al. Diaminodiphenylsulfone resistance of *Mycobacterium leprae* due to mutations in the dihydropteroate synthase gene. *FEMS microbiology letters*. 1999;177(2):231-5.
46. Dallas WS, Gowen JE, Ray PH, Cox MJ, Dev IK. Cloning, sequencing, and enhanced expression of the dihydropteroate synthase gene of *Escherichia coli* MC4100. *Journal of bacteriology*. 1992;174(18):5961-70.

47. Fermer C, Kristiansen BE, Skold O, Swedberg G. Sulfonamide resistance in *Neisseria meningitidis* as defined by site-directed mutagenesis could have its origin in other species. *Journal of bacteriology*. 1995;177(16):4669-75.
48. Lopez P, Espinosa M, Greenberg B, Lacks SA. Sulfonamide resistance in *Streptococcus pneumoniae*: DNA sequence of the gene encoding dihydropteroate synthase and characterization of the enzyme. *Journal of bacteriology*. 1987;169(9):4320-6.
49. Chaitanya VS, Das M, Bhat P, Ebenezer M. Computational Modelling of Dapsone Interaction with Dihydropteroate Synthase in *Mycobacterium leprae*; Insights into Molecular Basis of Dapsone Resistance in Leprosy. *Journal of cellular biochemistry*. 2015.
50. Matsuoka M, Kashiwabara Y, Namisato M. A *Mycobacterium leprae* isolate resistant to dapsone, rifampin, ofloxacin and sparfloxacin. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 2000;68(4):452-5.
51. National Center for Biotechnology Information. folP - dihydropteroate synthase [*Mycobacterium leprae* TN] - Gene ID: 908646: NCBI; 2017 [updated 24/09/2016; cited 2017 20/12/2017]. Available from: <https://www.ncbi.nlm.nih.gov/gene/908646>.
52. National Center for Biotechnology Information. rpoB - DNA-directed RNA polymerase subunit beta [*Mycobacterium leprae* TN] - Gene ID: 910599: NCBI; 2017 [updated 27/08/2016; cited 2017 20/12/2017]. Available from: <https://www.ncbi.nlm.nih.gov/gene/910599>.
53. National Center for Biotechnology Information. gyrA - DNA gyrase subunit A [*Mycobacterium leprae* TN] - Gene ID: 908154: NCBI; 2017 [updated 08/07/2017; cited 2017 20/12/2017]. Available from: <https://www.ncbi.nlm.nih.gov/gene/908154>.
54. World Health Organization. Report of the Workshop on Sentinel Surveillance for Drug Resistance in Leprosy: Hanoi, Viet Nam, 20-22 October 2008. 2009 SEA-GLP-2009.1.
55. World Health Organization. Guidelines for Global Surveillance of Drug Resistance in Leprosy. 2009.
56. World Health Organization. Surveillance of drug resistance in leprosy: 2010. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2011;86(23):237.
57. World Health Organization. Surveillance of drug resistance in leprosy: 2009. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2010;85(29):281.
58. World Health Organization. Drug resistance in leprosy: reports from selected endemic countries. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2009;84(26):264-7.
59. World Health Organization. Sentinel Surveillance for Drug Resistance in Leprosy: Report of the Workshop; Paris, France, 26-27 October 2009. 2010 SEA-GLP-2010.2.

60. World Health Organization. Meeting on Sentinel Surveillance for Drug Resistance in Leprosy: A Report; 9-10 October 2010, Tokyo, Japan. 2011 SEA-GLP-2011.1.
61. World Health Organization. Report of Meeting on Sentinel Surveillance for Drug Resistance in Leprosy: 22-23 August 2011, Hyderabad, India. 2012 SEA-GLP-2012.4.
62. World Health Organization. Sentinel Surveillance for Drug Resistance in Leprosy: Report of the WHO Global Leprosy Programme Meeting; Cotonou, Benin, 12-13 November 2012. 2013 SEA-GLP-2013.1.
63. World Health Organization. Antimicrobial resistance in Leprosy. Report of a global consultation. 2016 27-28 out. 2016. Report No.: SEA-GLP-2016.5.
64. Beltran-Alzate C, Lopez Diaz F, Romero-Montoya M, Sakamuri R, Li W, Kimura M, et al. Leprosy Drug Resistance Surveillance in Colombia: The Experience of a Sentinel Country. PLoS neglected tropical diseases. 2016;10(10):e0005041.
65. Avanzi C, Busso P, Benjak A, Loiseau C, Fomba A, Doumbia G, et al. Transmission of Drug-Resistant Leprosy in Guinea-Conakry Detected Using Molecular Epidemiological Approaches. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2016;63(11):1482-4.
66. Lavania M, Singh I, Turankar RP, Ahuja M, Pathak V, Sengupta U, et al. Molecular detection of multi drug resistant Mycobacterium leprae from Indian leprosy patients. Journal of global antimicrobial resistance. 2017.
67. Cambau E, Saunderson P, Matsuoka M, Cole ST, Kai M, Suffys P, et al. Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009-2015. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2018.
68. Bechelli LM, Martinez Dominguez V. The leprosy problem in the world. Bulletin of the World Health Organization. 1966;34(6):811-26.
69. Noordeen SK, Lopez Bravo L, Sundaresan TK. Estimated number of leprosy cases in the world. Leprosy review. 1992;63(3):282-7.
70. World Health Assembly. Forty-fourth World Health Assembly, Geneva, 6-16 May 1991: resolutions and decisions, annexes. Geneva: World Health Organization; 1991.
71. World Health Organization. WHO donated MDT 2018 [31/01/2018]. Available from: <http://www.who.int/lep/mdt/donation/en/>.
72. World Health Organization. Global Strategy for Further Reducing the Leprosy Burden and Sustaining Leprosy Control Activities (Plan period: 2006-2010). 2005.
73. World Health Organization. WEEKLY EPIDEMIOLOGICAL RECORD: Leprosy elimination campaigns – reaching every patient in every village. 1997 11 July 1997. Report No.
74. Bechelli LM, Martinez Dominguez V. Further information on the leprosy problem in the world. Bulletin of the World Health Organization. 1972;46:523-36.

75. WHO expert committee on leprosy. World Health Organization technical report series. 1977(607):7-48.
76. Dharmendra. Epidemiology of leprosy in relation to control (WHO Technical Report Series no. 716 of 1985). Indian journal of leprosy. 1986;58(1):1-16.
77. World Health Organization. Global leprosy situation, 2004. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2005;80(13):118-24.
78. World Health Organization. Global leprosy situation, 2009. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2009;84(33):333-40.
79. World Health Organization. Global leprosy situation, 2010. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2010;85(35):337-48.
80. World Health Organization. Leprosy update, 2011. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2011;86(36):389-99.
81. World Health Organization. Global leprosy situation, 2012. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2012;87(34):317-28.
82. World Health Organization. Global leprosy: update on the 2012 situation. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2013;88(35):365-79.
83. World Health Organization. Global leprosy update, 2013; reducing disease burden. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2014;89(36):389-400.
84. World Health Organization. Global leprosy situation, 2005. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2005;80(34):289-95.
85. World Health Organization. Global leprosy situation, 2006. Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2006;81(32):309-16.
86. World Health Organization. Global leprosy situation, 2007 (additional information). Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations. 2007;82(44):388.

87. World Health Organization. Global leprosy situation, 2008 (additional information). *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2008;83(50):459.
88. Daumerie D. Implementation of MDT. In: Sansarricq H, editor. *Multidrug therapy against leprosy: Development and implementation over the past 25 years*. Geneva: WHO Document Production Services; 2004. p. 45-57.
89. World Health Organization. Global leprosy update, 2014: need for early case detection. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2015;90(36):461-76.
90. World Health Organization. Global leprosy update, 2016: accelerating reduction of disease burden. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2017;92(35):501-19.
91. World Health Organization. Global leprosy update, 2015: time for action, accountability and inclusion. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2016;91(35):405-20.
92. Matsuoka M, Maeda S, Kai M, Nakata N, Chae GT, Gillis TP, et al. *Mycobacterium leprae* typing by genomic diversity and global distribution of genotypes. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 2000;68(2):121-8.
93. Shin YC, Lee H, Lee H, Walsh GP, Kim JD, Cho SN. Variable numbers of TTC repeats in *Mycobacterium leprae* DNA from leprosy patients and use in strain differentiation. *Journal of clinical microbiology*. 2000;38(12):4535-8.
94. Groathouse NA, Rivoire B, Kim H, Lee H, Cho SN, Brennan PJ, et al. Multiple polymorphic loci for molecular typing of strains of *Mycobacterium leprae*. *Journal of clinical microbiology*. 2004;42(4):1666-72.
95. Truman R, Fontes AB, De Miranda AB, Suffys P, Gillis T. Genotypic variation and stability of four variable-number tandem repeats and their suitability for discriminating strains of *Mycobacterium leprae*. *Journal of clinical microbiology*. 2004;42(6):2558-65.
96. Zhang L, Budiawan T, Matsuoka M. Diversity of potential short tandem repeats in *Mycobacterium leprae* and application for molecular typing. *Journal of clinical microbiology*. 2005;43(10):5221-9.
97. Taylor GM, Watson CL, Bouwman AS, Lockwood DNJ, Mays SA. Variable nucleotide tandem repeat (VNTR) typing of two palaeopathological cases of lepromatous leprosy from Mediaeval England. *Journal of Archaeological Science*. 2006;33(11):1569-79.
98. Weng X, Wang Z, Liu J, Kimura M, Black WCt, Brennan PJ, et al. Identification and distribution of *Mycobacterium leprae* genotypes in a region of high leprosy prevalence in China: a 3-year molecular epidemiological study. *Journal of clinical microbiology*. 2007;45(6):1728-34.

99. Monot M, Honore N, Baliere C, Ji B, Sow S, Brennan PJ, et al. Are variable-number tandem repeats appropriate for genotyping *Mycobacterium leprae*? *Journal of clinical microbiology*. 2008;46(7):2291-7.
100. Young SK, Ponnighaus JM, Jain S, Lucas S, Suneetha S, Lockwood DN, et al. Use of short tandem repeat sequences to study *Mycobacterium leprae* in leprosy patients in Malawi and India. *PLoS neglected tropical diseases*. 2008;2(4):e214.
101. Cardona-Castro N, Beltran-Alzate JC, Romero-Montoya IM, Melendez E, Torres F, Sakamuri RM, et al. Identification and comparison of *Mycobacterium leprae* genotypes in two geographical regions of Colombia. *Leprosy review*. 2009;80(3):316-21.
102. Fontes AN, Sakamuri RM, Baptista IM, Ura S, Moraes MO, Martinez AN, et al. Genetic diversity of *Mycobacterium leprae* isolates from Brazilian leprosy patients. *Leprosy review*. 2009;80(3):302-15.
103. Kimura M, Sakamuri RM, Groathouse NA, Rivoire BL, Gingrich D, Krueger-Koplin S, et al. Rapid variable-number tandem-repeat genotyping for *Mycobacterium leprae* clinical specimens. *Journal of clinical microbiology*. 2009;47(6):1757-66.
104. Sakamuri RM, Harrison J, Gelber R, Saunderson P, Brennan PJ, Balagon M, et al. A continuation: study and characterisation of *Mycobacterium leprae* short tandem repeat genotypes and transmission of leprosy in Cebu, Philippines. *Leprosy review*. 2009;80(3):272-9.
105. Shinde V, Newton H, Sakamuri RM, Reddy V, Jain S, Joseph A, et al. VNTR typing of *Mycobacterium leprae* in South Indian leprosy patients. *Leprosy review*. 2009;80(3):290-301.
106. Srisungnam S, Rudeeaneksing J, Lukebua A, Wattanapokayakit S, Pasadorn S, Mahotarn K, et al. Molecular epidemiology of leprosy based on VNTR typing in Thailand. *Leprosy review*. 2009;80(3):280-9.
107. Xing Y, Liu J, Sakamuri RM, Wang Z, Wen Y, Vissa V, et al. VNTR typing studies of *Mycobacterium leprae* in China: assessment of methods and stability of markers during treatment. *Leprosy review*. 2009;80(3):261-71.
108. da Silva Rocha A, Cunha Dos Santos AA, Pignataro P, Nery JA, de Miranda AB, Soares DF, et al. Genotyping of *Mycobacterium leprae* from Brazilian leprosy patients suggests the occurrence of reinfection or of bacterial population shift during disease relapse. *Journal of medical microbiology*. 2011;60(Pt 10):1441-6.
109. Li W, Sakamuri RM, Lyons DE, Orcullo FM, Shinde V, Dela Pena EL, et al. Transmission of dapson-resistant leprosy detected by molecular epidemiological approaches. *Antimicrobial agents and chemotherapy*. 2011;55(11):5384-7.
110. Weng X, Vander Heiden J, Xing Y, Liu J, Vissa V. Transmission of leprosy in Qiubei County, Yunnan, China: insights from an 8-year molecular epidemiology investigation. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2011;11(2):363-74.
111. Weng X, Xing Y, Liu J, Wang Y, Ning Y, Li M, et al. Molecular, ethno-spatial epidemiology of leprosy in China: novel insights for tracing leprosy in endemic and non endemic provinces. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2013;14:361-8.

112. Fontes ANB, Lima L, Mota RMS, Almeida RLF, Pontes MA, Goncalves HS, et al. Genotyping of *Mycobacterium leprae* for better understanding of leprosy transmission in Fortaleza, Northeastern Brazil. *PLoS neglected tropical diseases*. 2017;11(12):e0006117.
113. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, Lacroix C, et al. On the origin of leprosy. *Science*. 2005;308(5724):1040-2.
114. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nature genetics*. 2009;41(12):1282-9.
115. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. *Nature*. 2001;409(6823):1007-11.
116. National Center for Biotechnology Information. Genome Assembly and Annotation report: *Mycobacterium tuberculosis* H37Rv - NC_000962.3: NCBI; 2017 [updated 02/08/2016; cited 2015 18/09/2017]. Available from: http://www.ncbi.nlm.nih.gov/genome/166?genome_assembly_id=159857.
117. National Center for Biotechnology Information. Genome Assembly and Annotation report: *Mycobacterium bovis* strain 1595 - NZ_CP012095.1: NCBI; 2017 [updated 10/04/2017; cited 2015 18/09/2017]. Available from: https://www.ncbi.nlm.nih.gov/genome/161?genome_assembly_id=236414.
118. National Center for Biotechnology Information. Genome Assembly and Annotation report: *Mycobacterium lepromatosis* strain Mx1-22A - NZ_JRPY00000000.1: NCBI; 2017 [updated 10/04/2017; cited 2017 18/09/2017]. Available from: https://www.ncbi.nlm.nih.gov/genome/36766?genome_assembly_id=229003.
119. National Center for Biotechnology Information. Genome Assembly and Annotation report: *Mycobacterium africanum* GM041182 - NC_015758.1: NCBI; 2017 [updated 19/05/2017; cited 2017 18/09/2017]. Available from: https://www.ncbi.nlm.nih.gov/genome/10238?genome_assembly_id=173297.
120. Eiglmeier K, Parkhill J, Honore N, Garnier T, Tekaia F, Telenti A, et al. The decaying genome of *Mycobacterium leprae*. *Leprosy review*. 2001;72(4):387-98.
121. Gomez-Valero L, Rocha EP, Latorre A, Silva FJ. Reconstructing the ancestor of *Mycobacterium leprae*: the dynamics of gene loss and genome reduction. *Genome research*. 2007;17(8):1178-85.
122. Andersson JO, Andersson SG. Insights into the evolutionary process of genome degradation. *Current opinion in genetics & development*. 1999;9(6):664-71.
123. Vissa VD, Brennan PJ. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome biology*. 2001;2(8):REVIEWS1023.
124. Benjak A, Avanzi C, Singh P, Loiseau C, Girma S, Busso P, et al. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. *Nature communications*. 2018;9(1):352.
125. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. *Molecular genetics and metabolism*. 2013;110(1-2):3-24.

126. Metzker ML. Sequencing technologies - the next generation. *Nature reviews Genetics*. 2010;11(1):31-46.
127. Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *BioTechniques*. 1997;22(3):474-6, 8-81.
128. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011;475(7356):348-52.
129. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research*. 1998;8(3):186-94.
130. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research*. 1998;8(3):175-85.
131. Phred - Quality Base Calling [cited 2015 25/09/2015]. Available from: <http://www.phrap.com/phred/>.
132. Decreto nº 16.300, de 31 de dezembro de 1923, (1932).
133. Decreto nº 968, de 7 de maio de 1962, (1962).
134. Lazaro FP, Werneck RI, Mackert CC, Cobat A, Prevedello FC, Pimentel RP, et al. A major gene controls leprosy susceptibility in a hyperendemic isolated population from north of Brazil. *The Journal of infectious diseases*. 2010;201(10):1598-605.
135. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1966;34(3):255-73.
136. Trombone AP, Pedrini SC, Diorio SM, Belone Ade F, Fachin LR, do Nascimento DC, et al. Optimized protocols for *Mycobacterium leprae* strain management: frozen stock preservation and maintenance in athymic nude mice. *Journal of visualized experiments : JoVE*. 2014(85).
137. Santos AR, de Miranda AB, Lima LM, N. Suffys P, Degraeve WM. Method for high yield preparation in large and small scale of nucleic acids from mycobacteria. *Journal of Microbiological Methods*. 1992;15(2):83-94.
138. Neumann Ada S, Dias Fde A, Ferreira Jda S, Fontes AN, Rosa PS, Macedo RE, et al. Experimental Infection of *Rhodnius prolixus* (Hemiptera, Triatominae) with *Mycobacterium leprae* Indicates Potential for Leprosy Transmission. *PloS one*. 2016;11(5):e0156037.
139. Life technologies. Ion PGM™ Sequencing 400 Kit - MAN0007242 - 2.0: Thermo Fisher Scientific Inc.; 2013. 70 p.
140. Life technologies. Ion Xpress™ Plus gDNA Fragment Library Preparation - MAN0009847 - C.0: Thermo Fisher Scientific Inc.; 2014. 84 p.
141. Life technologies. Ion PGM™ Template OT2 400 Kit - MAN0007218 - A.0: Thermo Fisher Scientific Inc.; 2014. 122 p.

142. Life technologies. Ion TargetSeq™ Exome Enrichment for the Ion Proton™ System - MAN0006730 - 5.0: Thermo Fisher Scientific Inc.; 2013. 58 p.
143. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
144. Homer N. TMAP: the torrent mapping program. 2010.
145. Torrent Suite 2017 [updated 15/09/2017/22/11/2017]. Available from: <https://github.com/iontorrent/TS>.
146. Chevreux B, Wetter T, Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*. 1999;99:45-56.
147. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*. 2010;5(6):e11147.
148. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data 2010 [cited 2015 23/09/2015]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
149. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32(2):292-4.
150. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863-4.
151. Picard 2017. Available from: <http://broadinstitute.github.io/picard>.
152. Bioinformatics Babraham. Per Base Sequence Quality 2015 [cited 2015 24/09/2015]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/2%20Per%20Base%20Sequence%20Quality.html>.
153. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-95.
154. Thermo Fisher Scientific. [iontorrent/TS/plugin/variantCaller/pluginMedia/parameter_sets/4_2/ampliseq_germline_highstringency_pgm_parameters.json](https://github.com/iontorrent/TS/blob/master/plugin/variantCaller/pluginMedia/parameter_sets/4_2/ampliseq_germline_highstringency_pgm_parameters.json) 2015. Available from: https://github.com/iontorrent/TS/blob/master/plugin/variantCaller/pluginMedia/parameter_sets/4_2/ampliseq_germline_highstringency_pgm_parameters.json.

**APENDICE I: REVISÃO SISTEMÁTICA DE ESTUDOS DE VNTRs EM *M. leprae* –
PROTOCOLO**

1. DESENVOLVIMENTO

a. Busca

Para essa revisão sistemática foram buscados artigos na base de dados *PubMed*, (<https://www.ncbi.nlm.nih.gov/pubmed/>) em janeiro de 2018. Para a busca foi utilizado o *query* abaixo:

Quadro 6 – String de busca

("minisatellite repeats"[MeSH Terms] OR ("minisatellite"[All Fields] AND "repeats"[All Fields]) OR "minisatellite repeats"[All Fields] OR "vntr"[All Fields] OR "STR"[All Fields]) AND leprae[All Fields]

Para essa busca foram encontrados 53 resultados.

b. Filtragem

Seguiu-se então uma etapa de filtragem desses artigos baseada na leitura de seus títulos e resumos. Os critérios de exclusão adotados foram:

- Artigos de temas diferente do foco dessa revisão;
- Artigos em idioma que não português, inglês, francês, italiano ou espanhol;
- Artigos em duplicata;

Após essa filtragem foram excluídos 9 artigos por inadequação do tema e 4 artigos em outro idioma (chinês). Os demais artigos foram lidos na íntegra. Dentre eles 12 eram artigos de revisão e foram utilizados apenas para obtenção de novas referências de interesse, quando cabível. Os demais 24 artigos relatavam experimentos de genotipagem de *VNTRs* em *M. leprae* e tiveram os dados tabelados. Dentre as referências dos artigos lidos foram encontrados outros estudos de *VNTR* em *M. leprae* que não constavam não busca em base de dados, totalizando 33 artigos cujos resultados foram tabelados.

a. Organização dos dados

Primeiramente foram listados os marcadores utilizados em cada estudo. Importante ressaltar que considerando alterações nas bases de dados moleculares ao longo dos anos todas as posições de marcadores foram conferidas e, se necessário, atualizadas para corresponder a versão NC_002677.1 da cepa TN de *M. leprae* (sequência de referência). Foi constatado que os artigos utilizados apresentavam seus resultados de duas formas diferentes:

- Por amostra: resultado apresentado individualmente para cada amostra analisada;
- Por frequência: resultado apresentado numa tabela de frequência do número de cópias encontrado para cada marcador.

Ambas as formas foram tabuladas e tabelas e gráficos dos dados foram gerados utilizando o *software* Excel. Em seguida foi gerada:

- Uma tabela contendo todos os marcadores estudados, suas posições em relação a cepa de referência, o número de cópias na cepa de referência, o número de artigos que utilizou esse marcador e o número de cepas (de resultados conhecidos e não dúbios) já genotipadas para cada marcador (**Tese – Tabela 3**);
- Uma tabela contendo quais os marcadores foram genotipados em cada artigo (**Tese - Tabela 4**);
- Uma única tabela com todas as frequências de todos os marcadores de todos os artigos (Não apresentada devido ao seu tamanho);
- Gráficos da distribuição de frequência dos marcadores mais e menos variáveis (dentre aqueles utilizados em mais de 10 artigos) (**Tese – Figuras 11 a 13**);

•

**APENDICE II: ANALISE DE DADOS DO SEQUENCIAMENTO DE GENOMA
COMPLETO – PROTOCOLO E PRINCIPIO DO METODO**

1. ANÁLISE PRIMÁRIA DA QUALIDADE

Para análise geral da qualidade dos experimentos de sequenciamento foi utilizado o *software* nativo *Torrent Suite*. Foi analisada a cobertura do chip, o carregamento das *beads*, a taxa de policlonalidade e o número de *reads* obtidas. Em seguida, foi utilizado o *software* *FastQC* para avaliação do comprimento das *reads*, da qualidade das bases e da qualidade das *reads*.

2. MAPEAMENTO COM O GENOMA DO *M. musculus*

2.1. Obtenção e indexação da referência (*M. musculus*)

Antes do alinhamento foi necessário fazer o download de uma sequência de referência de bases de dados. Foi utilizada o *Mus musculus* GRCm38.p5 (*Genome Reference Consortium Mouse Reference 38*) cujo código é GCA_000001635.7. Após o download a referência foi indexada com o mesmo *software* de mapeamento *TMAP* (**Quadro 7**). A única opção utilizada foi a “-v” para ativar o modo verbose, que permite um acompanhamento mais detalhado do processo.

Quadro 7 - Comando de indexação da referência *M. musculus* no *software* *TMAP*

```
tmap index -v -f mmusculus_genome_GRCm38_p5.fna
```

2.2. Mapeamento (*M. musculus*)

Para o mapeamento foi utilizado o *software* *TMAP* que se baseia no método de Burrows–Wheeler (*Burrows–Wheeler transform - BWT*) (144, 153). É importante destacar que os dados brutos gerados são no formato BAM não alinhado. Foi utilizada a sub-função “map4” (ou *mapping procedure #4*) que é o mapeamento com referência adequado para *reads* longas (≥ 100 bp)

As opções utilizadas foram (**Quadro 8**):

- “--verbose” para ativar o modo verbose;
- “--rand-read-name” para ativar a randomização baseada no nome da *read*;
- “--aln-output-mode 1” retorna o mapeamento com o melhor escore. Se mais de um mapeamento obtiver o mesmo escore, um mapeamento randômico com esse escore é retornado;
- “--sam-flowspace-tags” para manter os dados de *flow order* e *flow signal*;
- “--output-type 0” determina que o arquivo de saída será no formato SAM;

Quadro 8 - Comando de mapeamento das amostras utilizando a referência *M. musculus* no software *TMAP*

```
tmap map4 \  
  --aln-output-mode 1 \  
  --verbose \  
  --rand-read-name \  
  --sam-flowspace-tags \  
  --output-type 0 \  
  -f mmusculus_genome_GRCm38_p5.fna \  
  -r $RAW-LIB_X.bam \  
  -i bam \  
  -s $LIB_X_TMAP-Mm-a1.sam
```

2.3. Avaliação de qualidade do mapeamento *M. musculus*

Para avaliação da qualidade do mapeamento primeiramente foi necessário ordenar o arquivo obtido no mapeamento de acordo com as coordenadas genômicas das *reads*, para isso foi utilizado o software *Samtools* (**Quadro 9**). A única opção utilizada foi a “--threads” que possibilita o uso de diversos processadores em simultâneo de forma a realizar as análises num tempo otimizado.

Quadro 9 - Comando de ordenação das amostras mapeadas no software *Samtools*

```
samtools sort \  
  -o $LIB_X_TMAP-Mm-a1-sorted.bam \  
  --threads 11 \  
  $LIB_X_TMAP-Mm-a1.sam
```

Para a análise de qualidade foi utilizado o *software Qualimap* tendo como entrada arquivo BAM previamente mapeado e ordenado (**Quadro 10**). A saída foi feita em dois formatos, HTML no diretório \$DIR e em PDF. A opção “-nt” se refere ao número de processadores utilizados e a opção “--java-mem-size” para adequar o uso de memória para o processamento.

Quadro 10 - Comando de avaliação de qualidade de mapeamento no *software Qualimap*

```
qualimap bamqc \  
-bam LIB_X_TMAP-Mm-a1-sorted.bam \  
-nt 10 \  
-outdir $DIR \  
-outfile report-Mm-$LIB_X.pdf \  
--java-mem-size=4G
```

2.4. Geração de arquivo com as *reads* não mapeadas (*M. musculus*)

As amostras, às quais seria dado prosseguimento na análise, foram selecionadas com base na avaliação de qualidade, e, mais especificamente, da porcentagem de *reads* alinhadas. Para as amostras selecionadas foram filtradas apenas as *reads* não mapeadas com o genoma do camundongo utilizando o *software Samtools* na sua sub-função “view” (**Quadro 11**). A opção “-b” indica que a saída deve ser no formato BAM e a opção “-f 4” indica que devem ser retiradas as *reads* marcadas com a *flag* “4”, que são as *reads* que foram alinhadas.

Quadro 11 - Comando de separação das *reads* mapeadas no *software Samtools*

```
samtools view \  
-b \  
-f 4 \  
$LIB_X_TMAP-Mm-a1.sam \  
> $LIB_X_TMAP-Mm-a1-unmapped.bam
```

3. MONTAGEM *de novo* (UTILIZANDO O *MIRA*)

3.1. Mudança de formato de *input*

Para a montagem *de novo* o formato de entrada de dados deve ser FASTQ, portanto antes do mapeamento foi necessário mudar o formato do arquivo de BAM para FASTQ utilizando o *software Samtools* (**Quadro 12**).

Quadro 12 - Comando de separação das *reads* mapeadas no *software Samtools*

```
samtools fastq \  
--threads 10 \  
$LIB_TMAP-Mm-a1-unmapped.bam \  
> $LIB_TMAP-Mm-a1-unmapped-FASTQ.fastq
```

3.2. Montagem *de novo* (utilizando o *MIRA*)

Para o alinhamento com o *software MIRA* é necessário criar um arquivo com as especificações desejadas, denominado manifesto (**Quadro 13**). Nesse arquivo consta i) o nome do projeto (opção “project”), ii) o tipo de montagem que deve ser feito (opção “job”) que no caso é o montagem *de novo* e acurado de um genoma, iii) o nome do grupo de reads (opção “readgroup”), iv) o arquivo a ser alinhado (opção “data”), v) o tipo de tecnologia de sequenciamento (opção “technology”) e, por fim vi) os parâmetros avançados (opção “parameters”). O único parâmetro avançado utilizado foi o “-NW:cac=warn” que evita que o programa interrompa o mapeamento caso haja mais do que 80x de cobertura prevista.

Quadro 13 – Manifesto utilizado para a montagem *de novo* no *software MIRA*

```
# Arquivo manifest-DeNovo-$LIB.conf  
  
project = DeNovo_$LIB_07-11-2017  
job = genome,denovo,accurate  
readgroup = $LIB  
data = $LIB_TMAP-Mm-a1-unmapped-FASTQ.fastq  
technology = iontor  
parameters = -NW:cac=warn
```


Isso simplifica a linha de comando de montagem *de novo* (**Quadro 14**) que recebe como entrada apenas o manifesto e gera um arquivo de registro do processo (ou *log*) e uma pasta contendo diversos arquivos com os resultados de mapeamento (nos formatos MAF, FASTA etc) e informações do processo. A única opção utilizada na linha de comando foi a “-t” que se refere ao número de processadores utilizados.

Quadro 14 – Comando de montagem *de novo* no *software MIRA*

```
mira -t 10 manifest-DeNovo- $\$LIB$ .conf >&log_assembly-DeNovo- $\$LIB$ .txt
```

3.3. Avaliação de qualidade do mapeamento *M. leprae*

Para avaliação da qualidade do alinhamento no *software Qualimap* foi primeiramente necessário converter o formato do arquivo de alinhamento que é dado pelo *MIRA* no formato MAF para o formato SAM. Para essa conversão foi utilizado o próprio *software MIRA*, na sua subfunção “miraconvert” (**Quadro 15**).

Quadro 15 – Comando de conversão do formato de arquivo alinhado no *software MIRA*

```
miraconvert \  
  DeNovo_ $\$LIB$ _assembly/DeNovo_ $\$LIB$ _d_results/DeNovo_ $\$LIB$ _out.maf \  
  DeNovo_ $\$LIB$ _out.sam \  
  > log_conv-DeNovo_ $\$LIB$ .txt
```

Na análise pelo *software Qualimap* a saída foi feita, novamente, nos formatos HTML e PDF (**Quadro 16**). A opção “-nt” se refere ao número de processadores utilizados e a opção “--java-mem-size” para adequar o uso de memória para o processamento.

Quadro 16 – Comando de avaliação de qualidade do alinhamento no *software Qualimap*

```
qualimap bamqc -bam \  
  /DeNovo_ $\$LIB$ _out.sam \  
  -outdir Qualimap/ \  
  -outfile Qualimap/Qualimap-DeNovo_ $\$LIB$ .pdf \  
  -outformat PDF:HTML
```

4. MAPEAMENTO COM O GENOMA DE REFERÊNCIA DO *M. leprae* (UTILIZANDO O *MIRA*)

4.1. Download da referência (*M. leprae*)

Para o mapeamento com referência no *software MIRA* é recomendado que a referência esteja nos formatos GFF ou GenBank. Aqui utilizamos o formato do GenBank GB obtido no NCBI para o *M. leprae* cepa TN, código NC_002677.1. Destaca-se que o arquivo utilizado foi o “GenBank(full)” sem mostrar *GI (GenInfo Identifier)*.

4.2. Mapeamento (*M. leprae*)

Para o mapeamento com o *software MIRA* foi elaborado um novo manifesto (**Quadro 17**). Nesse arquivo consta o nome do projeto (opção “project”), o tipo de alinhamento que deve ser feito (opção “job”) que no caso é o mapeamento com referência e acurado de um genoma, o arquivo da referência (opção “data” após o “is_reference”), o nome do grupo de *reads* (opção “readgroup”), o arquivo a ser mapeado (opção “data”) e o tipo de tecnologia de sequenciamento (opção “technology”). Nesse mapeamento não houve necessidade de utilizar qualquer parâmetro avançado. Destaca-se que o arquivo a ser mapeado utilizado foi o FASTQ obtido no item 5.5.3.1.

Quadro 17 – Manifesto utilizado para o mapeamento com referência de *M. leprae* no *software MIRA*

```
# Arquivo manifest-Map-$LIB.conf  
  
project = Mapeamento_$LIB  
job = genome,mapping,accurate  
readgroup  
is_reference  
data = mleprae_genome.gb  
readgroup = $LIB  
data = $LIB_TMAP-Mm-a1-unmapped-FASTQ.fastq  
technology = iontor
```

A linha de comando do *software MIRA* (**Quadro 18**) recebeu como entrada apenas o manifesto e gerou um arquivo de registro do processo (ou *log*). A única opção utilizada na linha de comando foi a “-t” que se refere ao número de processadores utilizados.

Quadro 18 – Comando de mapeamento com referência de *M. leprae* no *software MIRA*

```
mira -t 10 manifest-Map-$LIB.conf >&log_assembly-Map-$LIB.txt
```

4.3. Avaliação de qualidade do mapeamento *M. leprae*

Para avaliação da qualidade do alinhamento no *software Qualimap* foi primeiramente necessário converter o formato do arquivo de alinhamento que é dado pelo *MIRA* no formato MAF para o formato SAM. Para essa conversão foi utilizado o próprio *software MIRA*, na sua subfunção “miraconvert” (**Quadro 19**).

Quadro 19 – Comando de conversão do formato de arquivo mapeado no *software MIRA*

```
miraconvert \  
  Mapeamento_$LIB_out.maf \  
  Mapeamento_$LIB_out.sam \  
  > log_conv-Mapeamento_$LIB.txt
```

Na análise pelo *software Qualimap* a saída foi feita, novamente, nos formatos HTML e PDF (**Quadro 20**). A opção “-nt” se refere ao número de processadores utilizados e a opção “-java-mem-size” para adequar o uso de memória para o processamento.

Quadro 20 – Comando de avaliação de qualidade do mapeamento no *software Qualimap*

```
qualimap bamqc \  
  -bam Mapeamento_$LIB_assembly/SAM/Mapeamento_$LIB_out.sam \  
  -outdir Qualimap/ \  
  -outfile Mapeamento_$LIB.pdf \  
  -outformat PDF:HTML
```

5. MAUVE

Para comparação dos contigs obtidos no mapeamento com a sequência de referência foi utilizado a opção “Align with progressiveMauve” do *software* MAUVE. Para tanto foi utilizada a referência em FASTA e os arquivos de cada amostra com todos os contigs “padded” também no formato FASTA.

6. MAPEAMENTO COM O GENOMA DO *M. leprae* (UTILIZANDO TMAP)

6.1. Download e indexação da referência (*M. leprae*)

Antes do alinhamento foi necessário fazer o download de uma sequência de referência de bases de dados. Foi utilizada o *M. leprae* cepa TN, código NC_002677.1, a mesma que foi utilizada no alinhamento com o *software* MIRA. Após o download da referência no formato FASTA, a mesma foi indexada com o *software* TMAP (**Quadro 21**). A única opção utilizada foi a “-v” para ativar o modo verbose, que permite um acompanhamento mais detalhado do processo.

Quadro 21 - Comando de indexação da referência *M. leprae* no *software* TMAP

```
tmap index -v -f mleprae_genome.fna
```

6.2. Mapeamento (*M. leprae*)

Para o mapeamento com referência do *M. leprae* foi utilizada novamente a sub-função “map4” (ou *mapping procedure #4*) que é o mapeamento com referência adequado para *reads* longas (≥ 100 bp). As opções utilizadas foram (**Quadro 22**):

- “--verbose” para ativar o modo verbose;
- “--rand-read-name” para ativar a randomização baseada no nome da *read*;
- “--aln-output-mode 0” retorna o mapeamento com o melhor escore apenas se todos ou outros mapeamentos tiverem um escore pior, caso contrário retorna não mapeado;
- “--sam-flowspace-tags” para manter os dados de *flow order* e *flow signal*;
- “--output-type 0” determina que o arquivo de saída será no formato SAM;

Quadro 22 - Comando de mapeamento das amostras utilizando a referência *M. leprae* no software *TMAP*

```
tmap map4 \  
  --aln-output-mode 0 \  
  --verbose \  
  --rand-read-name \  
  --sam-flowspace-tags \  
  --output-type 0 \  
  -f ~/Documents/Helena/data/genome/mleprae/mleprae_genome.fna \  
  -r $LIB_X_TMAP-Mm-a1-unmapped.bam \  
  -i bam \  
  -s $LIB_X_TMAP-Mm-a1-M1-a0.sam
```

6.3. Avaliação de qualidade do mapeamento (*M. leprae*)

Para avaliação da qualidade do mapeamento pelo *software* primeiramente foi necessário ordenar por coordenada genômica o arquivo obtido no mapeamento para isso foi utilizado o *software Samtools* (**Quadro 23**). A única opção utilizada foi a “*--threads*” que possibilita o uso de diversos processadores em simultâneo de forma otimizar o tempo de análises.

Quadro 23 - Comando de ordenação das amostras mapeadas no *software Samtools*

```
samtools sort \  
  -o $LIB_X_TMAP-Mm-a1-M1-a0-sorted.bam \  
  --threads 11 \  
  $LIB_X_TMAP-Mm-a1-M1-a0.sam
```

Para a análise de qualidade foi utilizado o *software Qualimap* tendo como entrada arquivo BAM previamente mapeado e ordenado (**Quadro 24**). A saída foi feita em dois formatos, HTML no diretório \$DIR e em PDF. A opção “*-nt*” se refere ao número de processadores utilizados e a opção “*--java-mem-size*” para adequar o uso de memória para o processamento.

Quadro 24 - Comando de avaliação de qualidade de mapeamento no *software Qualimap*

```
qualimap bamqc \  
  -bam $LIB_X_TMAP-Mm-a1-M1-a0-sorted.bam \  
  -nt 10 \  
  -outdir $DIR \  
  -outfile report-M1-$LIB_X.pdf \  
  --java-mem-size=4G
```

6.4. Gerando arquivo somente com as *reads* mapeadas com *M. leprae*

Após a avaliação de qualidade, foram filtradas apenas as *reads* mapeadas com o genoma do *M. leprae* utilizando o *software Samtools* na sua sub-função “view” (**Quadro 25**). A opção “-b” indica que a saída deve ser no formato BAM e a opção “-F 4” indica que devem ser obtidas apenas as *reads* marcadas com a *flag* “4” no arquivo de entrada, ou seja apenas as *reads* que foram alinhadas.

Quadro 25 - Comando de separação das *reads* mapeadas no *software Samtools*

```
samtools view \  
-b -F 4 \  
$LIB_X_TMAP-Mm-a1-M1-a0.sam  
> $LIB_X_TMAP-Mm-a1-M1-a0-mapped.sam
```

7. BUSCA DE VARIANTES (TVC)

7.1. Ordenação, marcação de duplicatas e indexação

Antes de buscar as variantes os arquivos precisavam ser ordenados (**Quadro 26**), as duplicatas marcadas (**Quadro 27**) e por fim indexados (**Quadro 28**). Todas essas etapas foram realizadas utilizando o *software Picard*

Quadro 26 – Comando de ordenação das *reads* mapeadas no *software Picard*

```
java -jar $PATH_PICARD SortSam \  
INPUT=$LIB_X_TMAP-Mm-a1-M1-a0-mapped.sam \  
OUTPUT=$LIB_X_TMAP-Mm-a1-M1-a0 -picard-sorted.bam \  
SORT_ORDER=coordinate
```

Quadro 27 – Comando de marcação de duplicatas no *software Picard*

```
java -jar $PATH_PICARD MarkDuplicates \  
INPUT=$LIB_X_TMAP-Mm-a1-M1-a0 -picard-sorted.bam \  
OUTPUT=$LIB_X_TMAP-Mm-a1-M1-a0-dedup_reads.bam \  
METRICS_FILE=$LIB_X_metrics.txt
```

Quadro 28 – Comando de indexação no *software Picard*

```
java -jar $PATH_PICARD BuildBamIndex \  
INPUT=$LIB_X_TMAP-Mm-a1-M1-a0-dedup_reads.bam
```

7.2. Descoberta de variantes

Para a descoberta de variantes foi utilizado o *software TVC* que utiliza os dados de *flow order* (ordem de entrada de nucleotídeos no sequenciador durante a corrida) e *flow signal* (intensidade da alteração de pH em cada poço do chip a cada introdução de nucleotídeo no sequenciador) como um dos parâmetros de qualidade para descoberta de variantes (**Quadro 29**). Foram utilizadas as opções “-n” para determinar o número de processadores utilizados, a opção “-r” para determinar o arquivo de referência, opção “-b” para determinar o arquivo BAM da amostra (*input*), opção “—force-sample-name” para determinar o nome da amostra, opção “-O” para determinar o diretório de *output*, opção “--parameters-file” para indicar o arquivo de parâmetros, opção “--output-vcf” para determinar o arquivo de saída de variantes do tipo *SNP* e a opção “--assembly-vcf” para determinar o arquivo de saída das variantes do tipo *InDel*.

Quadro 29 – Comando para obtenção de lista de variantes no *software TVC*

```
tvc \  
-n 11 \  
-r mleprae_genome.fna \  
-b =$LIB_X_TMAP-Mm-a1-M1-a0-dedup_reads.bam \  
--force-sample-name $LIB_X \  
-O $DIR \  
--parameters-file $PARAMETERS.json \  
--output-vcf $LIB_X-small_variants.vcf \  
--assembly-vcf $LIB_X-indel_assembly.vcf
```

O arquivo de parâmetros utilizado foi o "Generic - PGM - Germ Line - High Stringency" proposto na versão 4.2 do *Torrent Suit* (154) (**Quadro 30**). O único parâmetro modificado foi o "snp_min_allele_freq" que determina a frequência mínima para que uma variante seja considerada um *SNP*.

Quadro 30 – Arquivo de parâmetros utilizado para obtenção de variantes no *software TVC*

```
{
  "torrent_variant_caller": {
    "snp_min_allele_freq": "0.7999999999999999",
    "snp_strand_bias": "0.9499999999999996",
    "hotspot_min_coverage": 20,
    "hotspot_min_cov_each_strand": 3,
    "snp_strand_bias_pval": 1,
    "hotspot_min_allele_freq": "0.1499999999999999",
    "snp_min_variant_escore": 10,
    "mnp_min_variant_escore": 10,
    "hotspot_strand_bias": "0.9499999999999996",
    "hp_max_length": 8,
    "filter_insertion_predictions": "0.20000000000000001",
    "indel_min_variant_escore": 10,
    "indel_min_coverage": 20,
    "heavy_tailed": 3,
    "outlier_probability": "0.01",
    "indel_strand_bias_pval": 1,
    "data_quality_stringency": "8.5",
    "snp_min_cov_each_strand": 3,
    "mnp_strand_bias": "0.9499999999999996",
    "mnp_strand_bias_pval": 1,
    "hotspot_strand_bias_pval": 1,
    "hotspot_min_variant_escore": 10,
    "indel_strand_bias": "0.8499999999999998",
    "downsample_to_coverage": 400,
    "filter_unusual_predictions": "0.25",
    "indel_min_allele_freq": "0.1499999999999999",
    "mnp_min_allele_freq": "0.1499999999999999",
    "do_snp_realignment": 0,
    "mnp_min_cov_each_strand": 3,
    "mnp_min_coverage": 20,
    "prediction_precision": 1,
    "indel_min_cov_each_strand": 3,
    "filter_deletion_predictions": "0.20000000000000001",
    "suppress_recalibration": 0,
    "snp_min_coverage": 20
  },
  "meta": {
    "repository_id": "$Rev$",
    "ts_version": "4.2",
    "name": "Generic - PGM - Germ Line - High Stringency",
    "trimreads": true,
    "tooltip": "Optimized for allele frequencies >10% and high specificity",
    "tvcargs": "tvc",
    "based_on": "ampliseq_germline_highstringency_pgm_parameters.json",
    "built_in": true,
    "configuration": "germline_high_stringency",

    "compatibility": {
      "chip": [
        "pgm"
      ],
      "frequency": [
        "germline"
      ],
      "stringency": [
        "high"
      ],
      "library": [
        "wholegenome",
        "ampliseq",
      ]
    }
  }
}
```



```

        "targetseq"
    ]
}
},
"long_indel_assembler": {
    "short_suffix_match": 5,
    "min_indel_size": 4,
    "min_var_count": 5,
    "min_var_freq": "0.14999999999999999",
    "output_mnv": 0,
    "max_hp_length": 8,
    "relative_strand_bias": "0.80000000000000004",
    "kmer_len": 19
},
"freebayes": {
    "gen_min_coverage": 6,
    "allow_mnps": 0,
    "allow_complex": 0,
    "min_mapping_qv": 8,
    "read_snp_limit": 10,
    "allow_indels": 1,
    "read_max_mismatch_fraction": 1,
    "gen_min_alt_allele_freq": "0.14999999999999999",
    "allow_snps": 1,
    "gen_min_indel_alt_allele_freq": "0.14999999999999999"
}
}

```

Legenda: Destacado em negrito está o parametro modificado. Originalmente valor atribuido era "0.14999999999999999".

ANEXO I: ARTIGO DE REVISÃO PUBLICADO NA *Clinics in Dermatology*



Genetics of leprosy: Expected and unexpected developments and perspectives

Monica E.D. Sauer, BPharm, PhD candidate^{a,1}, Heloisa Salomão, BPharm, MS^{a,1},
Geovana B. Ramos, BPharm, PhD candidate^a,
Helena R.S. D'Espindula, BPharm, MS candidate^a,
Rafael S.A. Rodrigues, BPharm, PhD candidate^a,
Wilian C. Macedo, BPharm, MS^a, Renata H.M. Sindeaux, PhD^b, Marcelo T. Mira, PhD^{a,b}.

^aGroup for Advanced Molecular Investigation, Graduate Program in Health Sciences, School of Medicine, Pontifical Catholic University of Paraná, Curitiba, Paraná, Brazil

^bSchool of Health and Biological Sciences, Pontifical Catholic University of Paraná, Curitiba, Paraná, Brazil

Abstract A solid body of evidence produced over decades of intense research supports the hypothesis that leprosy phenotypes are largely dependent on the genetic characteristics of the host. The early evidence of a major gene effect controlling susceptibility to leprosy came from studies of familial aggregation, twins, and Complex Segregation Analysis. Later, linkage and association analysis, first applied to the investigation of candidate genes and chromosomal regions and more recently, to genome-wide scans, have revealed several leukocyte antigen complex and nonleukocyte antigen complex gene variants as risk factors for leprosy phenotypes such as disease per se, its clinical forms and leprosy reactions. In addition, powerful, hypothesis-free strategies such as Genome-Wide Association Studies have led to an exciting, unexpected development: Leprosy susceptibility genes seem to be shared with Crohns and Parkinson's diseases. Today, a major challenge is to find the exact variants causing the biological effect underlying the genetic associations. New technologies, such as Next Generation Sequencing that allows, for the first time, the cost and time-effective sequencing of a complete human genome, hold the promise to reveal such variants. Strategies can be developed to study the functional effect of these variants in the context of infection, hopefully leading to the development of new targets for leprosy treatment and prevention.

© 2015 Elsevier Inc. All rights reserved.

Genetics of host susceptibility to infectious diseases

The burden of infectious diseases has been massive throughout history of mankind. Infections have been

responsible for a strong selective pressure; yet, some of them are, still today, major public health problems. Recent advances, such as the development of vaccines and antibiotics, combined with a general increase of the education and socio-economical level of human populations led to an increase of life expectancy, but not to eradication of infectious diseases.¹ To understand this scenario, it is necessary to consider a very complex interplay between environmental (microbial and nonmicrobial) and human

Corresponding author. Tel.: +55 (41) 3271-2030; fax: +55 (41) 3271-1657.
E-mail address: m.mira@pucpr.br (M.T. Mira).

¹ These authors share first authorship.

(genetic and nongenetic) factors that determines immunity to infection or its clinical outcome.¹

A classic feature of human infections is that only a proportion of exposed individuals develop clinical disease.² Accumulating evidence suggests that host genetic factors play a particularly important role in controlling susceptibility to these diseases.³ Some of the most compelling evidence that human genetics does indeed determine the occurrence of infection comes from primary immunodeficiencies (PIDs). The PIDs are typically monogenic (Mendelian) disorders that impair host defense mechanisms and result in predisposition to multiple infectious diseases. The PIDs are responsible for more than 200 known clinical syndromes, at least 100 of which presenting a well-defined molecular genetic basis. Examples of PIDs include mutations in genes encoding proteins of the IL-12/23-IFN- γ pathway associated with the Mendelian Susceptibility to Mycobacterial Diseases syndrome; complement defects associated with *Neisseria sp* invasive disease; X-linked lymphoproliferative disease associated with Epstein-Barr virus infection; mutations in genes *EVER1* or *EVER2* associated with Epidermodysplasia *verruciformis*; and apolipoprotein L-1 deficiency, associated with trypanosomiasis.^{1,4-6}

The profound influence of the genetic make-up of the host over resistance to infection has been investigated in several models. Studies in mice, based on reverse (gene-targeted knock-out and knock-in mutations) and forward genetics (natural mutation and random mutagenesis), have provided important insights into the mechanisms controlling infection and immunity in human, natural conditions.⁷ In human genetics, epidemiologic studies of adopted individuals showed that predisposition to infection were largely inherited, paradoxically, even more than diseases associated with less obvious environmental risk factors, such as cancer.^{1,8} Studies comparing concordance rates between monozygotic and dizygotic twins have provided powerful evidence for the existence of a host genetic background controlling susceptibility to different infectious diseases.¹ Finally, several genes have been associated with diseases such as AIDS (*HLA*, *MICA*, *PSORS1 C3*, *KIR*, and *CCR5*); hepatitis B (genes *HLA*); tuberculosis (*MBL*, *VDR*, *NRAMP1*, genes *HLA*); malaria (*HBB*, *SCO1*, *DDC*); and meningococcal disease (*CFH*).^{5,9,10}

In this context, leprosy presents as a very good model for the study of genetic predisposition to infection: *Mycobacterium leprae*, the etiologic agent, is known for its limited diversity between strains of different locations^{11,12}; this near clonal characteristic, together with the observation of a wide range of leprosy clinical phenotypes, strongly suggests that most of the disease variability, including susceptibility to leprosy per se, is dependent upon the genetic background of the host.¹³

Genetics of leprosy

Today, it is widely accepted that the exposure to *M leprae* is necessary but not sufficient to trigger the outcome of the disease, and different sets of genes modify host susceptibility

to leprosy in three different stages, namely: (i) the control of infection per se, that is, the disease regardless of its clinical form manifestation; (ii) after the infection is established, the definition of different clinical forms of the disease; and (iii) the risk of developing leprosy reactions (Figure 1).

Observational studies indicate the presence of a familial component to susceptibility to the leprosy,¹⁴ as well as increased concordance rates of disease per se and its clinical forms among monozygotic compared with dizygotic twins.^{15,16} In addition, several Complex Segregation Analyses have been performed for leprosy in different populations,¹⁷⁻¹⁹ aiming to identify the best-fit model of inheritance of the phenotype, given a collection of pedigrees. All of these studies supported a polygenic model of inheritance that includes a major gene effect.

Taken together, these results indicate the existence of a strong genetic component controlling susceptibility to leprosy; however, these observational designs do not provide any information about the exact nature of the genetic factors involved, that is, the identity and number of genes, as well as the genetic variants of these genes, causative of the leprosy phenotypes. For that, molecular studies are necessary, and a vast number of these studies have been conducted over the past decades. As a result, several candidate chromosomal regions and genes have been described, such as the MHC/HLA-linked genes of class I and II, *TNFA*, *LTA*, *MICA*, *MICB*, as well as non-HLA genes, such as *CCDC122*, *IFNG*, *IL10*, *IL12 B*, *IL23 R*, *KIR*, *LACC1* (formerly *C13 orf31*), *LTA4 H*, *LRRK2*, *MRC1*, *NOD2*, *PARK2/PACRG*, *RIPK2*, *SLC11A1* (formerly *NRAMP1*), *TAP*, *TLR*, *TNFSF15*, and *VDR*. Among these, a few candidates have been consistently replicated by independent studies and/or successfully investigated by functional studies. A brief summary of selected genes is presented in Table 1 and expanded next.

Major histocompatibility complex genes

The major histocompatibility complex (MHC), in humans known as the leukocyte antigen complex (HLA), is a cluster of highly polymorphic genes contained within a 3.6 megabase (Mb) interval located on chromosome 6p21. Most of these genes encode for proteins that are essential players in the processing and binding of antigenic peptides during the immune response. The HLA region is organized in 3 classes: HLA class I contains subclasses HLA-A, B and -C, which present antigenic peptides of intra-cellular origin to CD8+ T cells; HLA class II includes subclasses HLA-DR, DQ, DM, and -DP, that primarily bind peptides of extra-cellular origin and present them to CD4+T cells, and HLA class III contains genes coding for cytokines, such as tumor necrosis factor alpha (TNFA) and lymphotoxin alpha (LTA), for enzymes involved in steroid synthesis, for heat-shock proteins and for other intermediates of the immune response mechanisms.²⁰

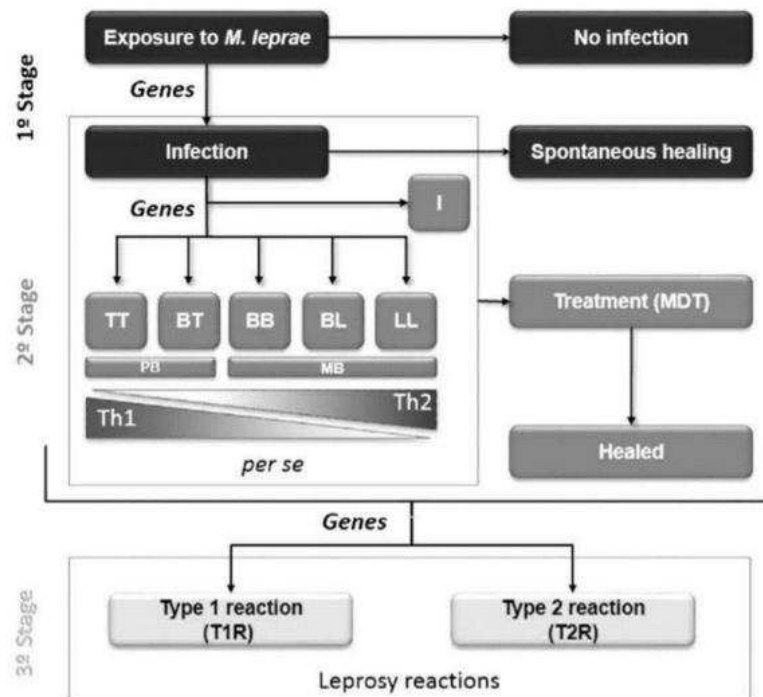


Fig. 1 A schematic stage model of genetic susceptibility to leprosy. The exposure to the mycobacteria does not always result in infection. In an initial stage, a first group of genes confers susceptibility to infection (leprosy per se). Among the individuals who develop the disease, a second group of genes determines the type of host immune response and subsequent leprosy subtype. Finally, a third group of genes confers the risk of developing leprosy reactions. TT, Tuberculoid-tuberculoid; BT, Borderline-tuberculoid; BB, Borderline-borderline; BL, Borderline-lepromatous; LL, Lepromatous-lepromatous; I, Indeterminate; PB, Paucibacillary; MB, Multibacillary; Th1, T-helper 1; Th2, T-helper 2; MDT, Multidrug therapy.

The crucial role of the HLA region in immune response regulation makes it the most exhaustively studied genomic candidate region in infection.²¹ It seems reasonable to assume a particularly important role for HLA genes in leprosy, given that clinical manifestation of the disease depends on the type of immune response shown by the host. The exchanges between Th1 and Th2 types of immune response may be partially controlled by a mechanism of antigen presentation involving HLA molecules.²² A large number of linkage and association studies reported the involvement of HLA alleles and haplotypes as important genetic factors controlling susceptibility to leprosy. In particular, HLA-DR alleles have been consistently associated with leprosy²³; Several studies reported an association of the HLA-DR2 alleles HLA-DRB1*04, DRB1*10, DRB1*12, DRB1*15, and DRB1*16 with susceptibility or resistance to leprosy in populations around the world.^{24–30} In addition, HLA-DR3 alleles were also found to be associated with leprosy susceptibility in two different populations.^{31,32}

HLA class I has been also intensively studied in leprosy, and HLA-A*2, A*11, B*40, and C*7 are some examples of alleles detected more often among leprosy cases compared with unaffected controls.³³ A recent study performed a high-resolution association scan of a 1.9 Mb region in the HLA complex in a Vietnamese population, followed by step-wise replication in an

independent sample from Vietnam and a sample from North India. The result was the identification of eight intergenic HLA class I region single nucleotide polymorphisms (SNPs) as novel genetic risk factors for leprosy per se, particularly implicating the HLA-C in leprosy susceptibility.³⁴

There is cumulative evidence that class III genes *TNFA* and *LTA* are involved in the immune response against leprosy.³⁵ *TNFA* encodes TNF- α , a proinflammatory and immunostimulatory cytokine that belongs to the TNF superfamily. This cytokine is involved in the regulation of a wide spectrum of biological processes, including the modulation of innate and adaptive immune responses. It is mainly secreted by macrophages, and functions as an important modulator of antigen presentation, through class II molecules, and cytokine production necessary for effective leukocyte response. In leprosy, a large body of functional experimental data indicates that TNF- α plays a central role by mediating the protective response to *M. leprae* invasion. Genetic studies have consistently indicated that *TNFA* variants can influence leprosy phenotypes.^{36,37} Recently, a large association study involving 4 population samples and more than 2500 individuals, followed by a meta-analysis, confirmed association between a promoter variant of *TNFA* and leprosy, interestingly, the effect seems to be restricted to the Brazilian samples.³⁷

Table 1 Selected leprosy associated genes replicated in at least two distinct population samples

Official symbol	Official full name	Function
<i>HLA-DRB1</i>	<i>Major histocompatibility complex, class II, DRB1</i>	Heterodimer anchored in the membrane that present antigenic peptides of extra-cellular origin to CD8+ T cells
<i>TNFA</i>	<i>Tumor necrosis factor alpha</i>	Proinflammatory cytokine
<i>LTA</i>	<i>Lymphotoxin alpha</i>	Proinflammatory cytokine
<i>IL10</i>	<i>Interleukin 10</i>	Anti-inflammatory cytokine
<i>PARK2</i>	<i>Parkinson protein 2, E3 ubiquitin protein ligase</i>	E3 ubiquitin ligase
<i>NOD2</i>	<i>Nucleotide-binding oligomerization domain containing 2</i>	Cytoplasmic receptor that plays a role in the immune response to intracellular bacterial lipopolysaccharides
<i>RIPK2</i>	<i>Receptor-interacting serine-threonine kinase 2</i>	Member of the receptor-interacting protein (RIP) family of serine/threonine protein kinases
<i>LACC1</i>	<i>Laccase (multicopper oxidoreductase) domain containing 1</i>	Unknown
<i>CCDC122</i>	<i>Coiled-coil domain containing 122</i>	Unknown

LTA is also a member of the TNF superfamily, but compared with TNF- α , much less is known about its function.³⁵ This cytokine is produced by lymphocytes and forms heterotrimers with lymphotoxin-beta, which anchor lymphotoxin-alpha to the cell surface. LTA signaling has proven important in controlling infection by intracellular pathogens, including *Mycobacterium tuberculosis* and *M leprae* among others.^{38,39} Interestingly, it seems that, in leprosy, the *LTA* genetic effect is more pronounced in patients with early onset of disease.⁴⁰ A recent study using knockout mice showed that the combination of TNF and LTA are necessary to the formation and maintenance of granulomas in response to *M leprae*: LTA seems to regulate granuloma formation whereas TNF- α is responsible for its integrity.^{35,39}

Finally, variants of additional HLA-linked genes, such MICA (MHC class I polypeptide-related sequence A), MICB (MHC class I polypeptide-related sequence B), and TAP (Transporter 1, ATP-binding cassette, sub-family B [MDR/TAP]), have also been described in association with leprosy phenotypes in different populations.²²

Of note, the interpretation of genetic association studies involving the HLA complex requires caution, due to the close proximity and very high gene density typical of this locus: Once positive association is detected between a leprosy phenotype and an MHC/HLA marker, the challenge is to determine if the causative gene is the one being investigated or another in close proximity to the marker used, a phenomenon called linkage disequilibrium.²²

Non-HLA genes

Interleukin 10

The *Interleukin 10* (*IL10*) gene is located at the 1q31-q32 chromosomal region and encodes for the anti-inflammatory cytokine IL-10, secreted by cells of the monocyte/macrophage

lineage and T-cell subsets such as Type 1 Tr, regulatory T, and T-helper 17.^{35,41} The IL-10 exerts its anti-inflammatory actions by blocking the production of proinflammatory cytokines by macrophages and their ability to serve as antigen-presenting or costimulatory cells.⁴¹ More specifically, this cytokine inhibits the production of IL-1, IL-6, and TNF- α in LPS- and IFN- γ -activated macrophages.^{36,41-43}

High levels of IL-10 are observed in multibacillary/lepomatous leprosy patients compared with paucibacillary/tuberculoid patients and a low TNF- α /IL-10 ratio is correlated to disease progression.^{35,42} Genetic epidemiology studies have been consistently reporting association between leprosy and SNPs located at the *IL10* gene.⁴⁴⁻⁴⁹ The exact reason why *IL10* polymorphisms are associated with leprosy is yet to be cleared. As this cytokine suppress the production of inflammatory mediator and boosts the development of Th2 immunity,⁵⁰ it is plausible that these polymorphisms somehow change *IL10* expression, directing the patient towards one of the poles of the leprosy clinical spectrum.

PARK2/PACRG (Parkinson protein 2, E3 ubiquitin protein ligase/Parkin co-regulated gene)

A genome-wide linkage analysis conducted in a Vietnamese population mapped a leprosy susceptibility locus to chromosome 6q25-27, an effect distributed along the entire leprosy clinical spectrum.⁵¹ In a follow-up study, the same group performed a systematic association scan of the candidate region and found 17 SNPs associated with leprosy susceptibility,⁵² 15 of them located in and around the promoter region shared by two genes: *PARK2* and *PACRG*. These results were validated in a separate set of unrelated individuals from Brazil. Later, an independent case control study found significant association between leprosy and *PARK2/PACRG* SNPs in an Indian population sample; however, the signal did not resist correction for multiple testing.⁵³ A study conducted in a geographically isolated Croatian community with a well-documented history

of leprosy showed that the protective alleles of two *PARK2* SNPs associated with the disease in Vietnam and Brazil were enriched in this population,⁵⁴ suggesting positive selection. More recently, a study involving a Vietnamese and an Indian population sample confirmed the *PARK2/PACRG* effect and revealed that age at diagnosis and differences in linkage disequilibrium patterns across different ethnicities are important for the correct interpretation of these association results.⁵⁵ Curiously, association has been also reported between the *PARK2/PACRG* leprosy polymorphisms and typhoid and paratyphoid fever in an Indonesian population.⁵⁶ The finding raised the hypothesis that the *PARK2/PACRG* genetic effect would not be specific to infection with *M leprae*, but related to host responses against intracellular parasites.

The *PARK2* gene encodes Parkin, an E3 ubiquitin ligase involved with the ubiquitin-proteasome complex that mediates the targeting of protein substrates for proteasomal degradation.¹³ Replicated association between leprosy and *PARK2/PACRG* variants revealed a new ubiquitin-dependent pathway of immunity to infection with *M leprae*, an idea supported by a functional study which demonstrated that proteasome function is important for *M leprae*-induced apoptosis.⁵⁷

The Genome Wide Association Studies genes

Genome Wide Association Studies (GWAS) are a powerful study design based on extensive coverage of the entire genome with hundreds of thousands of markers, genotyped in one single experiment, that capture the vast majority of common variants in the genome sequence.^{58,59} The genotyping data are then used in association testing that, if performed in samples large enough to achieve an adequate statistical power, allows for the identification of very small genetic effects, without the need of a previous hypothesis. The first GWAS on leprosy included 491,883 SNPs genotyped in 706 cases and 1225 controls from Eastern China. A total of 93 SNPs showed association with leprosy at the GWAS significance level; these SNPs were then tested in three independent replication sets totaling 3254 patients and 5955 controls from Eastern and Southern China. As a result, 15 SNPs distributed in six genes, *CCDC122*, *LACCI*, *NOD2*, *TNFSF15*, *RIPK2*, and the *HLA-DR DQ*, were consistently associated with leprosy across all samples. In addition, a trend toward association was detected between leprosy and one SNP of *LRRK2*.⁶⁰ In 2011, the same group published an expanded GWAS by combining their first data set with additional control subjects two additional genes were identified associated with leprosy: *IL23R* and *RAB32*.⁶¹ Later, another study identified a relative increase in *IL23R* gene copy number significantly associated with paucibacillary leprosy.⁶²

Given the nature of the GWAS, association studies involving a tremendous number of tests performed on a single experiment, therefore under strong inflation of type I error (false positive), these results, although exciting, must be validated by replication and/or by functional independent studies.⁶³

Association between leprosy and *HLA-DR-DQ* was replicated in an Indian population⁶⁴ and the *LACCI* and *CCDC122* signals were replicated in an Indian and African population.⁶⁵ More recently, a family-based replication study conducted in 474 Vietnamese leprosy families re-tested all 16 SNPs associated with leprosy in the Chinese original GWAS; six of them, located at *CCDC12*, *LACCI*, *NOD2*, *RIPK2*, and the *HLA-DR-DQ* genes were replicated.⁶⁶ Association between *NOD2* and leprosy has also been replicated in Nepal.⁶⁷

Several of the proteins encoded by these genes are involved in microbial sensing and in the early immune and inflammatory responses.⁶³ *NOD2* is located on chromosome 16q12 and encodes an intracellular receptor that recognizes a muramyl dipeptide component of the bacterial wall. After the interaction with activated *NOD2* molecules, *RIPK2* undergoes poly-ubiquitination mediated by an E3 ubiquitin ligase and promotes the activation of the TGF- β -activated kinase 1 (TAK1) complex. The activated TAK1 complex, again via poly-ubiquitination of a mediator, leads to degradation of transcriptional regulator nuclear factor κ B (NF- κ B) repressor I κ B, releasing NF- κ B to promote the transcription of pro-inflammatory genes, one of them, *TNFSF15*.^{63,68} A functional study reinforced the importance of the *NOD2* cascade in leprosy, by demonstrating that after *NOD2* from monocytes interact with *M leprae*'s muramyl dipeptide, a distinct interleukin-32-dependent induction of innate immune responses takes place, leading to the differentiation of monocytes into dendritic cells.⁷¹ These antigen-presenting cells are competent to define the adaptive immune response in leprosy.⁷²⁻⁷⁴

Genetics of leprosy reactions

LRs are sudden and intense inflammatory processes that affect individuals at all stages of the disease, from diagnosis, during treatment and even in the post-cure. The pathophysiological mechanisms involved are still widely unknown. Leprosy reactions are classified as type 1 (T1R or reversal reaction), which commonly affects patients at the tuberculoid side of the clinical spectrum; or type 2 (T2R, or *Erythema Nodosum Leprosum*), which affects mainly patients from the lepromatous pole of the disease.⁷⁵⁻⁷⁹ Only recently, human genetic epidemiology tools have been applied to the investigation of the molecular mechanisms controlling susceptibility to this extreme leprosy phenotype, as recently reviewed by Fava and cols.^{77,80}

The first evidence of association between LR and genetic polymorphisms came from studies involving Toll-like receptor (TLR) genes. An investigation involving a Nepalese population sample revealed polymorphisms on *TLR1* and *TLR2* associated with higher risk for T1R.^{81,82} A functional polymorphism of *TLR1*, which causes a substitution of asparagine to serine (N248S), was found associated with susceptibility to leprosy reactions in a Bangladeshi

population sample.⁸³ Interestingly, a recent study detected association between the same N248S polymorphism and susceptibility to leprosy in a Brazilian population sample.⁸⁴ The TLRs are transmembrane proteins that play a critical role in the inflammatory response to microbial pathogens.^{64,85} *TLR1* is located on chromosome 9q33.1 and its protein forms a heterodimer with TLR2 or TLR6 and mediates the recognition of several mycobacterial motifs: The heterodimer TLR1/TLR2 is involved in the recognition of *M leprae*,⁸⁶ and TLR1/TLR6 seems to be related to *M leprae* persistence in Schwann cells.⁸⁷

A prospective study of a Brazilian population sample resulted in strong evidence implicating variants of the *IL6* gene with susceptibility to T2R. Upon diagnosis, leprosy patients were monitored for at least 1 year for the occurrence of LR. Patients who developed T1R or T2R within the follow-up period were included in the group of cases, and leprosy patients who did not develop reactions were used to compose the control group. Cases of T1R and T2R were matched with controls by clinical form of leprosy and compared for the allele frequencies of markers physically covering the entire *IL6* gene. No association was observed between the *IL6* markers and T1R. Two independent signals of association with T2R were detected; one of them was captured by SNP rs1800795, a variant with known impact over *IL6* expression. These results support an important role of *IL6* in the development of T2R.⁷⁷

Additional studies on genetics of LR have implicated variants of *NOD2* and *VDR* as risk factors for the occurrence of T2R and T1R, respectively.^{67,88} Finally, a study in a Brazilian sample demonstrated an association between a SNP of *SCL11A1* with leprosy reactions.⁸⁹ These findings are yet to be replicated.

Leprosy, Crohn's, and Parkinson's diseases: a common genetic background?

In the past recent years, interesting findings concerning the genetic control of complex diseases have been observed: Some disorders, apparently unrelated, share genetic risk factors. In this context, the identification of leprosy susceptibility genes revealed an unexpected overlap with inflammatory bowel conditions and Parkinson's disease.⁹

Inflammatory bowel disease (IBD) is characterized by a chronic, relapsing intestinal inflammation. The two major types of IBD are Crohn's disease (CD) and ulcerative colitis (UC). The identification of *Mycobacterium avium* subspecies *paratuberculosis* RNA in mucosal samples of CD and UC patients suggested that the development of the disease, at least in some individuals, might be triggered by mycobacterial infection.^{90,91} This hypothesis has been supported by recent genetic studies revealing CD susceptibility genes that encode proteins responsible for recognition of bacterial structures and/or are present in immunologic pathways.^{9,92}

Some of these CD susceptibility loci are shared with leprosy,⁶³ as clearly exposed by the Chinese leprosy GWAS: Five of the newly described leprosy susceptibility genes have been previously associated with CD: *TNFSF15*, *NOD2*, *LACCI1*, *LRRK2*, and *IL23R*.^{60,61} Motivated by this unexpected finding, the same group later performed a systematic, comprehensive association study testing all previously described IBD loci as leprosy susceptibility candidates in a large Chinese leprosy sample, again, two IBD loci were associated with leprosy: *IL18RAP/IL18RI*, and *IL12B*.⁹³ These outstanding results corroborate the hypothesis that IBD, mainly CD, and leprosy share genetic risk factors and sum additional evidence supporting the role of an infectious agent participating in the initial events leading to CD manifestation.

One additional intriguing finding of the Chinese study is that the *in silico* analysis that places *NOD2*, *RIPK2*, and *TNFSF15* in the same pathway also included *PARK2* and *LRRK2*, genes encoding proteins that directly interact.⁶⁰ Strikingly, *PARK2* and *LRRK2* are well known Parkinson's disease (PD) susceptibility genes. Of particular interest, *LRRK2*, shown to be expressed in macrophage and monocytes,⁹⁴ harbors variants that have been associated with CD⁹⁵ and Parkinson's disease, as well as leprosy.

Based on this results, one can speculate that because *NOD2/RIPK2* initiates a signaling process that involves an ubiquitination process through TRAF6 (TNF receptor-associated factor 6), an E3 ubiquitin ligase, it is possible that parkin, also an E3, plays a role in this process; in addition, *LRRK2* is thought to regulate the ligase activity of *PARK2*⁹⁶; therefore, it may also take part in the signaling control. These observations suggest the existence of a partially shared genetic control of susceptibility to an infectious disease, an inflammatory disease and a neurodegenerative disorder. The complete elucidation of the cross talk between those susceptibility genes is a difficult, yet tremendously exciting task.

Future perspectives

Classic genetic studies on susceptibility to leprosy have been focusing on the identification of common variants that could explain predisposition to disease and, as a result, several common variants were described to be associated with leprosy phenotypes. The assumption made is that a set of these variants in one or several genes of a biochemical pathway would act together to contribute to a clinical outcome. These findings cannot explain the totality of the large genetic effect observed in descriptive genetic epidemiology studies. Interesting, this scenario remains true for a number of complex traits.⁹⁷

With the popularization of the GWAS, it has become increasingly clear that a large part of the genetic effect controlling disease susceptibility was missing⁹⁸: With rare

exceptions, more than 90-95% of the heritable component of a disease has been left unexplained after extensive GWAS on several complex diseases, giving rise to the term missing heritability.^{98,99} The idea behind these missing effect is that common genetic variability is unlikely to explain the entire genetic predisposition to disease.^{98,100,101}

As a result of intense debate, a new scenario of not one, but two major hypothesis has risen, both aiming to offer a better understanding on how different classes of genetic variations can account for a specific outcome: The Common Disease-Common Variant (CDCV) and the Common Disease Multiple Rare Variant (CDMRV) hypothesis. The first one argues that common variants with small effect are responsible for the genetic susceptibility to common diseases; the CDMRV hypothesis defends the idea that multiple rare alleles of large effect, explains the genetic susceptibility to common diseases.^{100,101}

In fact, the idea that rare variants are behind of human susceptibility to common diseases is not new. A remarkable example of the impact of a rare variant over disease phenotypes has been provided by Altare et al. in 1998¹⁰²: By applying molecular biology tools a genetic analysis on the study of a single young girl presenting generalized, atypical *M. bovis*-BCG infection. The authors found a homozygous, 4.4 kb-long deletion in the gene *IL12B* that impaired IL12-dependent, IFN- mediated response against a non-virulent mycobacteria. The finding, that clearly implicates the IL-12/IFN- axis as critical for the control of the immune response against mycobacterial infection, represents strong support to the idea that a rare structural, homozygous variant can underlie the mechanism controlling host susceptibility to infection. This study and others provided rising evidence that rare variations are important pieces of the puzzle of human phenotypic variation.¹⁰³ Identifying these rare variants without previous indication of their possible location used to be a daunting task, given (i) the need to study rare, extreme cases of disease; and (ii) the limited throughput and the high costs of classic, Sanger-based methods for genome sequencing.

The CDMRV hypothesis gained momentum with the very recent advent of next generation genetic analysis platforms capable of sequencing massive segments of the human genome, whole exomes or even genomes, in short time frames and for a reasonable cost. By reducing the time and cost limitations to a minimum challenge, these platforms of next-generation sequencing (NGS) technologies, also known as massively parallel DNA sequencing, are ideal tools to be used on the search of such rare variations.^{100,104}

Conclusions

The NGS technology has the potential to revolutionize our understanding on how genes or genomic regions are involved in the pathogenesis of human diseases.¹⁰⁵ The use of NGS can be directed to the identification of causative

disease mutation by resequencing the whole genome (or exome) of a small number of affected individuals, typically displaying extreme phenotypes of the disease. The approach has been successfully applied to determine the genetic basis of rare disorders, much of them Mendelian, through the study of a small number of affected individuals. In this scenario, an interesting question would be whether the same strategy could be applied to the identification of rare mutations possibly contributing to the risk of occurrence of a complex disease, such as common infections. In this much more complex context, leprosy has been considered as an excellent model to the study of genetic susceptibility to common infectious diseases.¹⁰⁶ It is reasonable to believe that innovative approaches based on NGS technology could help to unravel much of the "missing heritability" observed in leprosy and other infectious diseases. Also, classic experimental design such as linkage analysis can be coupled to these approaches to increase its power.

References

1. Casanova JL, Abel L. Inborn errors of immunity to infection: The rule rather than the exception. *J Exp Med*. 2005;202:197-201.
2. Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. *Nat Rev Genet*. 2012;13:175-188.
3. Alcais A, Abel L, Casanova JL. *Hum Genet* of infectious diseases: Between proof of principle and paradigm. *J Clin Invest*. 2009;119:2506-2514.
4. Picard C, Casanova JL, Abel L. Mendelian traits that confer predisposition or resistance to specific infections in humans. *Curr Opin Immunol*. 2006;18:383-390.
5. Abel L, Casanova JL. *Hum Genet* of infectious diseases: Fundamental insights from clinical studies. *Semin Immunol*. 2006;18:327-329.
6. Quintana-Murci L, Alcais A, Abel L, et al. Immunology in nature: Clinical, epidemiologic and evolutionary genetics of infectious diseases. *Nat Immunol*. 2007;8:1165-1171.
7. Casanova JL, Abel L. The human model: A genetic dissection of immunity to infection in natural conditions. *Nat Rev Immunol*. 2004;4:55-66.
8. Sorensen TI, Nielsen GG, Andersen PK, et al. Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med*. 1988;318:727-732.
9. Orlova M, Di Pietrantonio T, Schurr E. Genetics of infectious diseases: Hidden etiologies and common pathways. *Clin Chem Lab Med*. 2011;49:1427-1437.
10. Bellamy R. Genetic susceptibility to tuberculosis. *Clin Chest Med*. 2005;26:233-246. [vi].
11. Monot M, Honore N, Garnier T, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet*. 2009;41:1282-1289.
12. Truman RW, Singh P, Sharma R, et al. Probable zoonotic leprosy in the southern United States. *N Engl J Med*. 2011;364:1626-1633.
13. Schurr E, Alcais A, de Leseleuc L, et al. Genetic predisposition to leprosy: A major gene reveals novel pathways of immunity to *Mycobacterium leprae*. *Semin Immunol*. 2006;18:404-410.
14. Shields ED, Russell DA, Pericak-Vance MA. Genetic epidemiology of the susceptibility to leprosy. *J Clin Invest*. 1987;79:1139-1143.
15. Mohamed AP, Ramanujam K. Leprosy in twins. *Int J Lepr*. 1966;34:405-407.
16. Chakravarti MR, Vogel F. A twin study on leprosy. Georg Thieme Verlag. 1973:1-123.
17. Abel L, Vu DL, Oberti J, et al. Complex segregation analysis of leprosy in southern Vietnam. *Genet Epidemiol*. 1995;12:63-82.
18. Abel L, Demenais F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. *Am J Hum Genet*. 1988;42:256-266.

19. Lazaro FP, Werneck RI, Mackert CC, et al. A major gene controls leprosy susceptibility in a hyperendemic isolated population from north of Brazil. *J Infect Dis.* 2010;201:1598-1605.
20. Prado Montes de Oca E. Human Polymorphisms as Clinical Predictors in Leprosy. *J Trop Med.* 2011, Article ID 923943, 14 pages (1-14) doi: 10.1155/2011/923943.
21. Trowsdale J. The MHC, disease and selection. *Immunol Lett.* 2011;137:1-8.
22. Mira MT. Genetic host resistance and susceptibility to leprosy. *Microbes Infect.* 2006;8:1124-1131.
23. Fava VM, Mira MT. Genetics of leprosy. In: Nunzi E, Massone C, eds. *Leprosy: A Practical Guide.* Springer; 2012. p. 19-26.
24. Schauf V, Ryan S, Scollard D, et al. Leprosy associated with HLA-DR2 and DQw1 in the population of northern Thailand. *Tissue Antigens.* 1985;26:243-247.
25. Soebono H, Giphart MJ, Schreuder GM, et al. Associations between HLA-DRB1 alleles and leprosy in an Indonesian population. *Int J Lepr Other Mycobact Dis.* 1997;65:190-196.
26. Tosh K, Ravikumar M, Bell JT, et al. Variation in MICA and MICB genes and enhanced susceptibility to paucibacillary leprosy in South India. *Hum Mol Genet.* 2006;15:2880-2887.
27. Motta PM, Cech N, Fontan C, et al. Role of HLA-DR and HLA-DQ alleles in multibacillary leprosy and paucibacillary leprosy in the province of Chaco (Argentina). *Enferm Infecc Microbiol Clin.* 2007;25:627-631.
28. Vanderborgh PR, Pacheco AG, Moraes ME, et al. HLA-DRB1*04 and DRB1*10 are associated with resistance and susceptibility, respectively, in Brazilian and Vietnamese leprosy patients. *Genes Immun.* 2007;8:320-324.
29. da Silva SA, Mazini PS, Reis PG, et al. HLA-DR and HLA-DQ alleles in patients from the south of Brazil: Markers for leprosy susceptibility and resistance. *BMC Infect Dis.* 2009;9:134.
30. Zhang F, Liu H, Chen S, et al. Evidence for an association of HLA-DRB1*15 and DRB1*09 with leprosy and the impact of DRB1*09 on disease onset in a Chinese Han population. *BMC Med Genet.* 2009;10:133.
31. van Eden W, de Vries RR, D'Amato J, et al. HLA-DR-associated genetic control of the type of leprosy in a population from Surinam. *Hum Immunol.* 1982;4:343-350.
32. Gorodezky C, Flores J, Arevalo N, et al. Tubercloid leprosy in Mexicans is associated with HLA-DR3. *Lepr Rev.* 1987;58:401-406.
33. Shankarkumar U. HLA associations in leprosy patients from Mumbai, India. *Lepr Rev.* 2004;75:79-85.
34. Alter A, Huong NT, Singh M, et al. Human leukocyte antigen class I region single-nucleotide polymorphisms are associated with leprosy susceptibility in Vietnam and India. *J Infect Dis.* 2011;203:1274-1281.
35. Cardoso CC, Pereira AC, de Sales Marques C, et al. Leprosy susceptibility: Genetic variations regulate innate and adaptive immunity, and disease outcome. *Future Microbiol.* 2011;6:533-549.
36. Misch EA, Berrington WR, Vary Jr JC, et al. Leprosy and the human genome. *Microbiol Mol Biol Rev.* 2010;74:589-620.
37. Cardoso CC, Pereira AC, Brito-de-Souza VN, et al. TNF 308 G>A single nucleotide polymorphism is associated with leprosy among Brazilians: A genetic epidemiology assessment, meta-analysis, and functional study. *J Infect Dis.* 2011;204:1256-1263.
38. Roach DR, Briscoe H, Saunders B, et al. Secreted lymphotoxin-alpha is essential for the control of an intracellular bacterial infection. *J Exp Med.* 2001;193:239-246.
39. Hagge DA, Saunders BM, Ebenezer GJ, et al. Lymphotoxin-alpha and TNF have essential but independent roles in the evolution of the granulomatous response in experimental leprosy. *Am J Pathol.* 2009;174:1379-1389.
40. Alcais A, Alter A, Antoni G, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat Genet.* 2007;39:517-522.
41. Hsieh CS, Heimberger AB, Gold JS, et al. Differential regulation of T helper phenotype development by interleukins 4 and 10 in an alpha beta T-cell-receptor transgenic system. *Proc Natl Acad Sci U S A.* 1992;89:6065-6069.
42. Ishida H, Hastings R, Thompson-Snipes L, et al. Modified immunologic status of anti-IL-10 treated mice. *Cell Immunol.* 1993;148:371-384.
43. Kuhn R, Lohler J, Rennick D, et al. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell.* 1993;75:263-274.
44. Santos AR, Suffys PN, Vanderborgh PR, et al. Role of tumor necrosis factor-alpha and interleukin-10 promoter gene polymorphisms in leprosy. *J Infect Dis.* 2002;186:1687-1691.
45. Moraes MO, Pacheco AG, Schonkeren JJ, et al. Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy. *Genes Immun.* 2004;5:592-595.
46. Fitness J, Floyd S, Wardorff DK, et al. Large-scale candidate gene study of leprosy susceptibility in the Karonga district of northern Malawi. *Am J Trop Med Hyg.* 2004;71:330-340.
47. Malhotra D, Darvishi K, Sood S, et al. IL-10 promoter single nucleotide polymorphisms are significantly associated with resistance to leprosy. *Hum Genet.* 2005;118:295-300.
48. Pereira AC, Brito-de-Souza VN, Cardoso CC, et al. Genetic, epidemiologic and biological analysis of interleukin-10 promoter single-nucleotide polymorphisms suggests a definitive role for 819 C/T in leprosy susceptibility. *Genes Immun.* 2009;10:174-180.
49. Franceschi DS, Mazini PS, Rudnick CC, et al. Influence of TNF and IL10 gene polymorphisms in the immunopathogenesis of leprosy in the south of Brazil. *Int J Infect Dis.* 2009;13:493-498.
50. Kang TJ, Yeum CE, Kim BC, et al. Differential production of interleukin-10 and interleukin-12 in mononuclear cells from leprosy patients with a Toll-like receptor 2 mutation. *Immunology.* 2004;112:674-680.
51. Mira MT, Alcais A, Van Thuc N, et al. Chromosome 6 q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet.* 2003;33:412-415.
52. Mira MT, Alcais A, Nguyen VT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature.* 2004;427:636-640.
53. Malhotra D, Darvishi K, Lohra M, et al. Association study of major risk single nucleotide polymorphisms in the common regulatory region of PARK2 and PACRG genes with leprosy in an Indian population. *EJHG.* 2006;14:438-442.
54. Bakija-Konsuo A, Mulic R, Boraska V, et al. Leprosy epidemics during history increased protective allele frequency of PARK2/PACRG genes in the population of the Mljet Island, Croatia. *Eur J Med Genet.* 2011;54:e548-e552.
55. Alter A, Fava VM, Huong NT, et al. Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Hum Genet.* 2013;132:107-116.
56. Ali S, Vollaard AM, Widjaja S, et al. PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever. *Clin Exp Immunol.* 2006;144:425-431.
57. Fulco TO, Lopes UG, Samo EN, et al. The proteasome function is required for Mycobacterium leprae-induced apoptosis and cytokine secretion. *Immunol Lett.* 2007;110:82-85.
58. Manry J, Quintana-Murci L. A genome-wide perspective of human diversity and its implications in infectious disease. *Cold Spring Harb Perspect Med.* 2013;3:a012450.
59. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. *Genomics Inform.* 2012;10:117-122.
60. Zhang FR, Huang W, Chen SM, et al. Genomewide association study of leprosy. *N Engl J Med.* 2009;361:2609-2618.
61. Zhang F, Liu H, Chen S, et al. Identification of two new loci at IL23 R and RAB32 that influence susceptibility to leprosy. *Nat Genet.* 2011;43:1247-1251.
62. Ali S, Srivastava AK, Chopra R, et al. IL12 B SNPs and copy number variation in IL23 R gene associated with susceptibility to leprosy. *J Med Genet.* 2013;50:34-42.
63. Schurr E, Gros P. A common genetic fingerprint in leprosy and Crohn's disease? *N Engl J Med.* 2009;361:2666-2668.

64. Wong SH, Gochhait S, Malhotra D, et al. Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* 2010;6:e1000979.
65. Wong SH, Hill AV, Vannberg FO, et al. Genomewide association study of leprosy. *N Engl J Med.* 2010;362:1446-1447. [author reply 1447-1448].
66. Grant AV, Alter A, Huong NT, et al. Crohn's disease susceptibility genes are associated with leprosy in the Vietnamese population. *J Infect Dis.* 2012;206:1763-1767.
67. Berrington WR, Macdonald M, Khadge S, et al. Common polymorphisms in the NOD2 gene region are associated with leprosy and its reactive states. *J Infect Dis.* 2010;201:1422-1435.
68. Hitotsumatsu O, Ahmad RC, Tavares R, et al. The ubiquitin-editing enzyme A20 restricts nucleotide-binding oligomerization domain containing 2-triggered signals. *Immunity.* 2008;28:381-390.
69. Kanneganti TD, Lamkanfi M, Nunez G. Intracellular NOD-like receptors in host defense and disease. *Immunity.* 2007;27:549-559.
70. Hasegawa M, Fujimoto Y, Lucas PC, et al. A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF-kappaB activation. *EMBO J.* 2008;27:373-383.
71. Schenk M, Krutzik SR, Sieling PA, et al. NOD2 triggers an interleukin-32-dependent human dendritic cell program in leprosy. *Nat Med.* 2012;18:555-563.
72. Krutzik SR, Tan B, Li H, et al. TLR activation triggers the rapid differentiation of monocytes into macrophages and dendritic cells. *Nat Med.* 2005;11:653-660.
73. Sieling PA, Chatterjee D, Porcelli SA, et al. CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science.* 1995;269:227-230.
74. Sieling PA, Jullien D, Dahlem M, et al. CD1 expression by dendritic cells in human leprosy lesions: Correlation with effective host immunity. *J Immunol.* 1999;162:1851-1858.
75. Alter A, Grant A, Abel L, et al. Leprosy as a genetic disease. *Mamm Genome.* 2011;22:19-31.
76. Sampaio LH, Stefani MM, Oliveira RM, et al. Immunologically reactive M. leprae antigens with relevance to diagnosis and vaccine development. *BMC Infect Dis.* 2011;11:26.
77. Sousa AL, Fava VM, Sampaio LH, et al. Genetic and immunologic evidence implicates interleukin 6 as a susceptibility gene for leprosy type 2 reaction. *J Infect Dis.* 2012;205:1417-1424.
78. Kahawita IP, Walker SL, Lockwood DNJ. Leprosy type 1 reactions and erythema nodosum leprosum. *An Bras Dermatol.* 2008;83:75-82.
79. Britton WJ, Lockwood DN. Leprosy. *Lancet.* 2004;363:1209-1219.
80. Fava V, Orlova M, Cobat A, et al. Genetics of leprosy reactions: An overview. *Mem Inst Oswaldo Cruz.* 2012;107(Suppl 1):132-142.
81. Misch EA, Hawn TR. Toll-like receptor polymorphisms and susceptibility to human disease. *Clin Sci (Lond).* 2008;114:347-360.
82. Bochud PY, Sinsimer D, Aderem A, et al. Polymorphisms in Toll-like receptor 4 (TLR4) are associated with protection against leprosy. *Eur J Clin Microbiol Infect Dis.* 2009;28:1055-1065.
83. Schuring RP, Hamann L, Faber WR, et al. Polymorphism N248 S in the human Toll-like receptor 1 gene is related to leprosy and leprosy reactions. *J Infect Dis.* 2009;199:1816-1819.
84. de Sales Marques C, Brito-de-Souza VN, Albuquerque Guerreiro LT, et al. Toll-like receptor 1 (TLR1) N248 S single nucleotide polymorphism is associated with leprosy risk and regulates immune activation during mycobacterial infection. *J Infect Dis.* 2013;208:120-129.
85. Bochud PY, Hawn TR, Siddiqui MR, et al. Toll-like receptor 2 (TLR2) polymorphisms are associated with reversal reaction in leprosy. *J Infect Dis.* 2008;197:253-261.
86. Krutzik SR, Ochoa MT, Sieling PA, et al. Activation and regulation of Toll-like receptors 2 and 1 in human leprosy. *Nat Med.* 2003;9:525-532.
87. Mattos KA, Oliveira VG, D'Avila H, et al. TLR6-driven lipid droplets in Mycobacterium leprae-infected Schwann cells: Immunoinflammatory platforms associated with bacterial persistence. *J Immunol.* 2011;187:2548-2558.
88. Sapkota BR, Macdonald M, Berrington WR, et al. Association of TNF, MBL, and VDR polymorphisms with leprosy phenotypes. *Hum Immunol.* 2010;71:992-998.
89. Teixeira MA, Silva NL, Ramos AL, et al. NRAMP1 gene polymorphisms in individuals with leprosy reactions attended at two reference centers in Recife, northeastern Brazil. *Rev Soc Bras Med Trop.* 2010;43:281-286.
90. Greenstein RJ. Is Crohn's disease caused by a mycobacterium? Comparisons with leprosy, tuberculosis, and Johne's disease. *Lancet Infect Dis.* 2003;3:507-514.
91. Jeyanthan M, Boutros-Tadros O, Radhi J, et al. Visualization of Mycobacterium avium in Crohn's tissue by oil-immersion microscopy. *Microbes Infect.* 2007;9:1567-1573.
92. Behr MA, Schurr E. Mycobacteria in Crohn's disease: A persistent hypothesis. *Inflamm Bowel Dis.* 2006;12:1000-1004.
93. Liu H, Irwanto A, Tian H, et al. Identification of IL18 RAP/IL18 R1 and IL12 B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. *Am J Hum Genet.* 2012;91:935-941.
94. Thevenet J, Pescini Gobert R, Hooft van Huijsdijnen R, et al. Regulation of LRRK2 expression points to a functional role in human monocyte maturation. *PLoS One.* 2011;6:e21519.
95. Umeno J, Asano K, Matsushita T, et al. Meta-analysis of published studies identified eight additional common susceptibility loci for Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis.* 2011;17:2407-2415.
96. Smith WW, Pei Z, Jiang H, et al. Leucine-rich repeat kinase 2 (LRRK2) interacts with parkin, and mutant LRRK2 induces neuronal degeneration. *Proc Natl Acad Sci U S A.* 2005;102:18676-18681.
97. Frazer KA, Murray SS, Schork NJ, et al. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009;10:241-251.
98. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-753.
99. Maher B. Personal genomes: The case of the missing heritability. *Nature.* 2008;456:18-21.
100. Zhang J, Chiodini R, Badr A, et al. The impact of next-generation sequencing on genomics. *J Genet Genomics.* 2011;38:95-109.
101. Schork NJ, Murray SS, Frazer KA, et al. Common versus rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* 2009;19:212-219.
102. Altare F, Lammis D, Revy P, et al. Inherited interleukin 12 deficiency in a child with bacille Calmette-Guerin and Salmonella enteritidis disseminated infection. *J Clin Invest.* 1998;102:2035-2040.
103. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* 2001;69:124-137.
104. Metzker ML. Sequencing technologies-the next generation. *Nat Rev Genet.* 2010;11:31-46.
105. Novelli G, Predazzi IM, Mango R, et al. Role of genomics in cardiovascular medicine. *World J Cardiol.* 2010;2:428-436.
106. Alter A, Alcais A, Abel L, et al. Leprosy as a genetic model for susceptibility to common infectious diseases. *Hum Genet.* 2008;123:227-235.

**ANEXO II: ARTIGO PUBLICADO NO *INTERNATIONAL JOURNAL OF
SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY***

Emended description of the genus *Phytobacter*, its type species *Phytobacter diazotrophicus* (Zhang 2008) and description of *Phytobacter ursingii* sp. nov.

Marcelo Pillonetto,^{1,2} Lavinia N. Arend,² Helisson Faoro,³ Helena R. S. D'Espindula,¹ Jochen Blom,⁴ Theo H. M. Smits,⁵ Marcelo T. Mira¹† and Fabio Rezzonico^{5,*}†

Abstract

The species *Phytobacter diazotrophicus* and the associated genus *Phytobacter* were originally described by Zhang *et al.* [*Arch Microbiol* **189** (2008), 431–439] on the basis of few endophytic nitrogen-fixing bacteria isolated from wild rice (*Oryza rufipogon*) in China. In this study, we demonstrate that a number of clinical isolates that were either described in the literature, preserved in culture collections, or obtained during a 2013 multi-state sepsis outbreak in Brazil also belong to the same genus. 16S rRNA gene sequencing, multilocus sequence analysis based on *gyrB*, *rpoB*, *atpD* and *infB* genes, as well as digital DNA–DNA hybridization support the existence of a second species within the genus *Phytobacter*. All isolates from the recent Brazilian outbreak, along with some older American clinical strains, were found to belong to the already described species *Phytobacter diazotrophicus*, whereas three clinical strains retrieved in the USA over a time span of almost four decades, could be assigned to a new *Phytobacter* species. Implementation of an extended set of biochemical tests showed that the two *Phytobacter* species could phenotypically be discriminated from each other by the ability to utilize L-sorbose and D-serine. This feature was limited to the strains of the novel species described herein, for which the name *Phytobacter ursingii* sp. nov. is proposed, with ATCC 27989^T (=CNCTC 5729^T) as the designated type strain. An emended description of the species *Phytobacter diazotrophicus* and of the genus *Phytobacter* is also provided.

Species identification based on biochemical reactions within the former *Erwinia herbicola*–*Enterobacter agglomerans* complex (EEC) has often yielded inconclusive results [1–5]. In 1984, Brenner *et al.* defined 13 biotypes within the EEC based on DNA–DNA hybridization [6]. In the past, many clinical isolates belonging to the EEC have been repeatedly misidentified as *Pantoea* (*Enterobacter*) *agglomerans* or *Pantoea* spp. by commercial automated systems, resulting in several case and outbreak reports involving EEC bacteria erroneously attributed to *Pantoea* species [7–12]. With the increasing availability of 16S rRNA gene sequences [13–15], *gyrB* sequences [16, 17], and the use of multi-locus sequence analysis (MLSA) [18, 19], and more recently of whole-genome sequencing (WGS) [20, 21], the taxonomy of the *Enterobacteriaceae* has undergone major changes. Several of Brenner's original biotypes can be now assigned to species such as *Leclercia adecarboxylata* [22], *Pantoea agglomerans*

and *Pantoea dispersa* [23], *Pantoea ananatis* [24], *Erwinia aphidicola* [25], *Pantoea septica*, *Pantoea eucrina* and *Pantoea brenneri* [26], *Kosakonia cowanii* [27], *Enterobacter* sp. [28], and *Buttiauxella* spp. However, Brenner's Biotype XII has neither been extensively characterized nor assigned to a specific genus or species yet.

During our investigation of a multi-state sepsis outbreak occurring in Brazil caused by contaminated total parenteral nutrition (M. Pillonetto, L. Arend, S. M. T. Gomes, M. A. A. Oliveira, L. N. Timm, A. F. Martins, A. L. Barth, A. Mazzetti, L. Hersemann, T. H. M. Smits, M. T. Mira and F. Rezzonico, unpublished results) preliminary genetic analysis suggested that the enterobacterial species involved bore similarities both with Brenner's Biotype XII isolates from outbreaks occurring in the 1970s in the USA [6, 29–31] and with *Phytobacter diazotrophicus* [32], an endophytic

Author affiliations: ¹Core for Advanced Molecular Investigation, Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba-PR, Brazil; ²Molecular Bacteriology Division, Central Public Health Laboratory, State of Paraná, São José dos Pinhais-PR, Brazil; ³Carlos Chagas Institute, FIOCRUZ, Curitiba-PR, Brazil; ⁴Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany; ⁵Environmental Genomics and Systems Biology Research Group, Institute of Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland.

***Correspondence:** Fabio Rezzonico, fabio.rezzonico@zhaw.ch

Keywords: sepsis; outbreak; total parenteral nutrition; *Erwinia herbicola*; *Pantoea agglomerans*; misidentification.

Abbreviations: ANI, average nucleotide identity; DDH, DNA–DNA hybridization; MLSA, multi-locus sequence analysis; WGS, whole-genome sequencing.

†These authors contributed equally to this work.

Four supplementary tables and one supplementary figure are available with the online Supplementary Material.

nitrogen-fixing enterobacterial species isolated from wild rice in China. We thus decided to investigate further and 5110RM was selected as representative isolate of the Brazilian outbreak to undergo detailed analysis, along with culture collection strains previously assigned to Brenner's Biotype XII [3, 33] and type strain *Phytobacter diazotrophicus* DSM 17806^T. Two other strains, Bisph2 [34] and CAV1151 [35], were not directly available, but were included in the study on the basis of their complete genome sequences (Table 1).

Amplification and sequencing of the nearly complete 16S rRNA gene was performed as previously described [36]. Sequences of 1530 bp were obtained and used to generate a maximum likelihood tree together with entries of related species retrieved from the GenBank database by means of a BLASTN search. Whenever it was possible, only type strains were selected for the analysis. Isolate 5110RM was found to cluster well within the group of organisms previously assigned to Brenner's Biotype XII [33], which includes its archetypal strain LMG 5337 [6] and clinical strains ATCC 27981 and ATCC 27990 [3, 29–31]. Glyphosate-degrading soil bacterium [*Enterobacter*] sp. Bisph2 [34] and *Phytobacter diazotrophicus* type strain DSM 17806^T were found to belong to this group as well. A nearby, but separate, clade was formed by clinical strains ATCC 27982, ATCC 27989^T [3, 29–31] and [*Kluyvera intermedia*] CAV1151 [35], possibly indicating a second separate species within the genus *Phytobacter* (Fig. 1). All the above strains, together with a multi-drug resistant isolate unrelated to the initial outbreak

(10289RM) obtained in November 2015 during a routine survey for antibiotic resistances in a hospital in Curitiba (Brazil), were included in the ensuing analyses (Table 1).

MLSA was performed according to the protocol targeting the four standard housekeeping genes proposed by Brady *et al.* [18]. Using the MEGA7 software [37], concatenated *atpD*, *gyrB*, *infB* and *rpoB* sequences (2548 bp) were aligned and compared with the corresponding gene sequences of reference strains of species belonging to nearby taxa, including the genera *Kosakonia*, *Enterobacter*, *Citrobacter* and *Kluyvera* (Table S1, available in the online Supplementary Material). A maximum likelihood tree based on concatenated gene sequences was obtained (Fig. 2), which further supported the distinctiveness of the two clades that were already recognizable in the 16S rRNA tree (Fig. 1). Both branch depth and gene identities strongly suggested that they represent two separate species within the genus *Phytobacter*. The distinction between the two *Phytobacter* species is clearly supported also when using the amino acid sequences to perform MLSA (Fig. S1). Identity between *Phytobacter diazotrophicus* DSM 17806^T and strain ATCC 27989^T, which was selected as type strain for the second candidate *Phytobacter* species on the basis that the latter was already available both in our lab and in two separate culture collections, was 97.2% for *atpD*, 93.1% for *gyrB*, 96.7% for *infB*, and 95.9% for *rpoB*. Evolutionary divergence over sequence pairs for the MLSA gene set was estimated to be 4.2% between the clade of *Phytobacter*

Table 1. List of *Phytobacter* strains used in this study

Name	Source	Origin	Isolation date	Other designations	References	Previous identification*
<i>Phytobacter diazotrophicus</i>						
5110RM	Total parenteral nutrition	Brazil, Curitiba	5 Dec 2013	–	This work	–
10289RM	Hospital (surveillance swab)	Brazil, Curitiba	10 Nov 2015	–	This work	–
ATCC 27981	Sputum	USA, South Carolina	1974	CDC 865–74	[33]	<i>Pantoea agglomerans</i> †
ATCC 27990	Human, clinical	USA	1972	CDC 1134–72	[33]	<i>Pantoea agglomerans</i> †
Bisph2	Sandy soil	Algeria, Biskra	Unknown	–	[34]	<i>Enterobacter</i> sp.‡
DSM 17806 ^T	Wild rice (<i>Oryza rufipogon</i>)	PR China, Lingshui	May 2004	LS 8	[32]	<i>Phytobacter diazotrophicus</i>
LMG 5337	IV fluid	USA, Virginia	1971	CDC 219–71, ATCC 29915	[33]	Unidentified sp.§
<i>Phytobacter ursingii</i> sp. nov.						
ATCC 27982	IV fluid	USA	1972	CDC 2214–72	[3]	<i>Pantoea agglomerans</i> †
ATCC 27989 ^T	Sputum	USA, South Carolina	1974	CDC 1300–74	[33]	<i>Pantoea agglomerans</i> †
CAV1151	Perirectal	USA, Virginia	Sep 2009	–	[35]	<i>Kluyvera intermedia</i>

IV, Intravenous.

*Strain not available in our lab; results based only on genome sequence retrieved from GenBank.

†Still currently listed as *Pantoea agglomerans* in the ATCC catalogue despite being excluded from this species in the cited reference.

‡Genome sequence deposited as *Enterobacter* sp., classified as *Kosakonia* sp. by [42].

§As listed in the BCCM/LMG catalogue, excluded from *Pantoea agglomerans* in the cited reference, still listed as the latter species in the ATCC catalogue.

||Genome sequence deposited as *Kluyvera intermedia*, classified as *Kosakonia* sp. by [42].

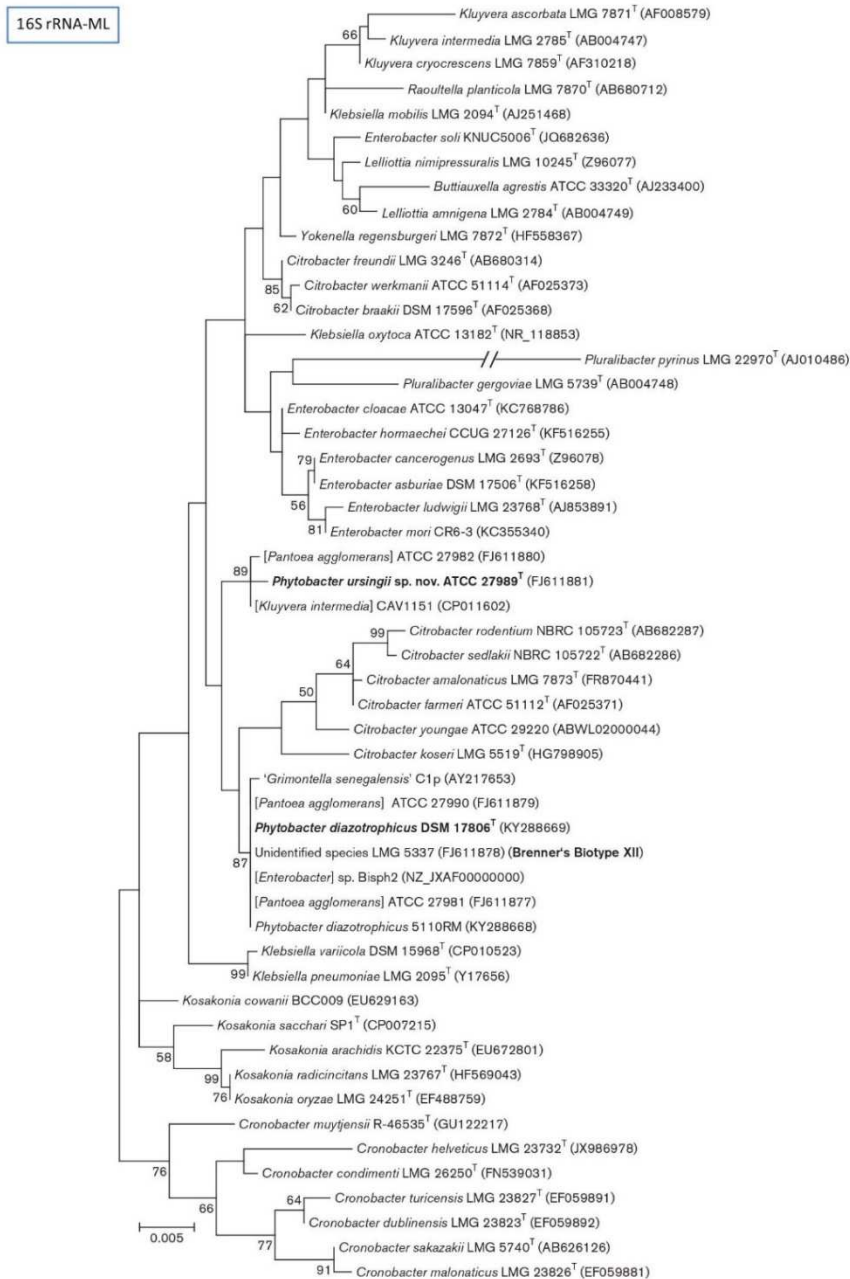


Fig. 1. Maximum likelihood tree based on nearly complete 16S rRNA gene sequences showing the phylogenetic relationship between species of the genus *Phytobacter* and related taxa within the family *Enterobacteriaceae*. The genus *Cronobacter* was used to root the tree. Numbers at branching points are bootstrap percentage values (>50%) based on 500 replications. Bar, 0.5% nucleotide substitutions.

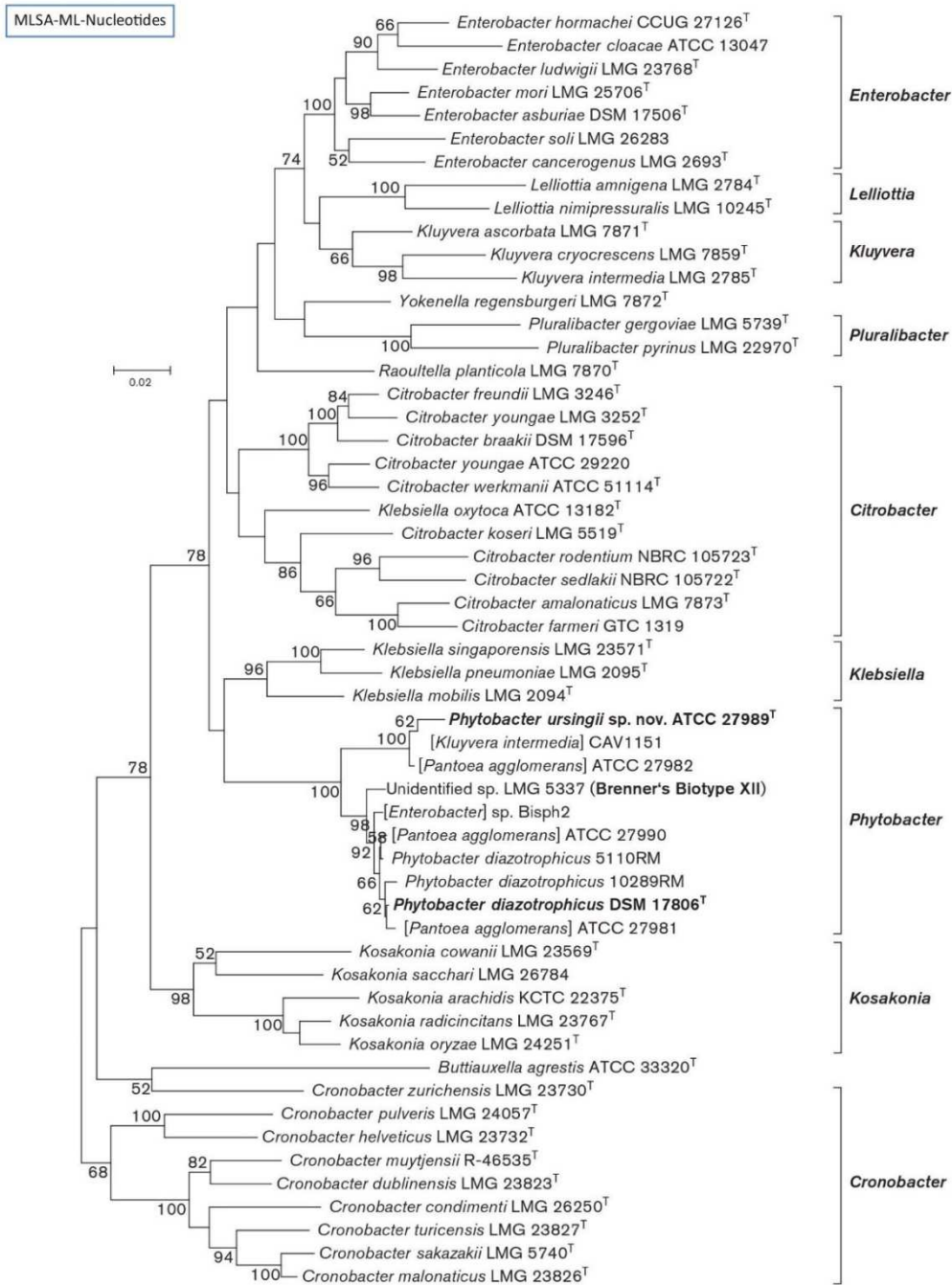


Fig. 2. Maximum likelihood tree based on concatenated partial *atpD*, *gyrB*, *infB* and *rpoB* gene sequences (2548 bp), showing the phylogenetic relationship between species of the genus *Phytobacter* and related taxa within the family *Enterobacteriaceae*. The genus *Cronobacter* was used to root the tree. Numbers at branching points are bootstrap percentage values (>50%) based on 500 replications. Bar, 2% nucleotide substitutions.

diazotrophicus DSM 17806^T ($n=9$) and that of strain ATCC 27989^T ($n=3$), whereas the mean distance to other genera was in all cases at least 9.0% (Table S2). Within the species, estimates of evolutionary divergences were 0.9 and 1.1% for *Phytobacter diazotrophicus* and the clade of strain ATCC 27989^T, respectively (Table S3).

Available *Phytobacter* strains (Table 1) were submitted to manual phenotypic testing using API 20E, API 50 CHE (bioMérieux) and Biolog GN2 as well as to automated testing using VITEK 2. The biochemical profiles obtained largely matched those obtained for the original Brenner's Biotype XII ($n=14$) [6]: all strains produced acid and gas from glucose, were D-adonitol, inositol and triple amino acid (lysine, ornithine and arginine) negative, and citrate, Voges-Proskauer test, lactose, sucrose, salicin, D-sorbitol, L-arabinose, L-rhamnose, malonate, acetate, D-xylose, trehalose, cellobiose and aesculin positive. Nitrate was reduced to nitrite and no yellow pigment was produced. Deviating from the original description, all the strains analysed in this work were motile and produced indole, which was true for only one of the fourteen strains analysed by Brenner. The latter was most probably CDC 219-71 (=LMG 5337), which is the only isolate present in both data sets. If correct, this discrepancy suggests that the original Biotype XII represents a broader taxonomical unit with respect to the two species described herein. Overall comparison of the manual and automated biochemical profiles revealed that the two *Phytobacter* species could be differentiated from each other on the basis of L-sorbose and D-serine, with both substrates being used exclusively by the strains clustering with ATCC 27989^T. Positive indole and nitrite production as well as negative N-acetyl-glucosaminidase, L-proline arylamidase, phosphatase and glycine arylamidase tests allow discrimination of *Phytobacter* from the nearby genus *Kosakonia* (Table 2). The results of additional tests are presented in the footnote of Table 2.

Draft genomes of *Phytobacter diazotrophicus* DSM 17806^T, of outbreak isolate 5110RM and of strain ATCC 27989^T were obtained using the Illumina MiSeq platform. Genomic DNA was extracted from pure bacterial cultures using the NucleoSpin Tissue kit (Macherey-Nagel) and sheared to a mean size of 700 bp using a Covaris S220 focused-ultrasonicator at the Functional Genomics Centre in Zürich, Switzerland. Library preparation was performed with the TruSeq Nano DNA Library Prep kit for NeoPrep (Illumina), selecting the 550 bp insert size workflow. Paired-end reads (301 bp) were generated on the Illumina MiSeq sequencer. The genomes were *de novo* assembled using the SeqMan NGen software included in the DNASTAR Lasergene genomics package version 12. Genome sequences of [*Enterobacter*] sp. Bisph2 (NZ_JXAF000000000) and [*Kluyvera intermedia*] CAV1151 (NZ_CP011602) were used as templates to refine the assemblies.

The draft genome sequences were annotated using GenDB [38] and included in an EDGAR 2.0 project [39] for comparative genomics with related genome sequences. A core genome tree was generated, which included available genomes of closely related organisms based on the MLSA phylogeny (not all relatives are represented by a genome in GenBank) (Fig. 3). The resulting tree shows a similar topology to the one obtained with the MLSA approach (Fig. 2), confirming the existence of two separate species within the genus *Phytobacter*, with outbreak strain 5110RM and [*Enterobacter*] sp. Bisph2 being closely related to *Phytobacter diazotrophicus* DSM 17806^T while [*Kluyvera intermedia*] CAV1151 clusters separately with strain ATCC 27989^T (Fig. 3). WGS data were used to perform digital DNA-DNA hybridization (DDH) using Genome-to-Genome Distance Calculator as described by Henz *et al.* [40] (version 2.1; <http://ggdc.dsmz.de>) and average nucleotide identity (ANI) as described by Goris *et al.* [41] using EDGAR 2.0 [39] and/or ANI Calculator (<http://enve-omics.ce.gatech.edu/ani/>).

Table 2. Differential phenotypic characteristics between *Phytobacter* spp. and the nearby genus *Kosakonia* [27]

The complete results for the biochemical panels tested below are listed in Table S4.

Test substrate	<i>Phytobacter diazotrophicus</i> ^a DSM 17806 ^T	<i>Phytobacter ursingii</i> sp. nov. ^b ATCC 27989 ^T	<i>Kosakonia</i>
API 20E			
Indole production	Positive	Positive	Negative
API 50CHE			
L-Sorbose	Negative	Positive	NR ^c
BIOLOG GN2			
D-Serine	Negative	Positive	Variable (30–49%)
VITEK 2			
β-N-acetyl-glucosaminidase	Negative	Negative	Variable (10–29%)
L-Proline arylamidase	Negative	Negative	Variable (10–29%)
Phosphatase	Negative	Negative	Non negative ^d
Assimilation of glycine arylamidase	Negative	Negative	Non negative ^d

a, $n=29$ (25 outbreak strains+DSM 17806^T+ATCC 27990+ATCC 27981+LMG 5337); b, $n=2$ (ATCC 27982+ATCC 27989); c, NR, not rated; d, this substrate is not reported by Brady *et al.* [18], but strain *Kosakonia* sp. ATCC 27988 yielded a positive reaction in this work.

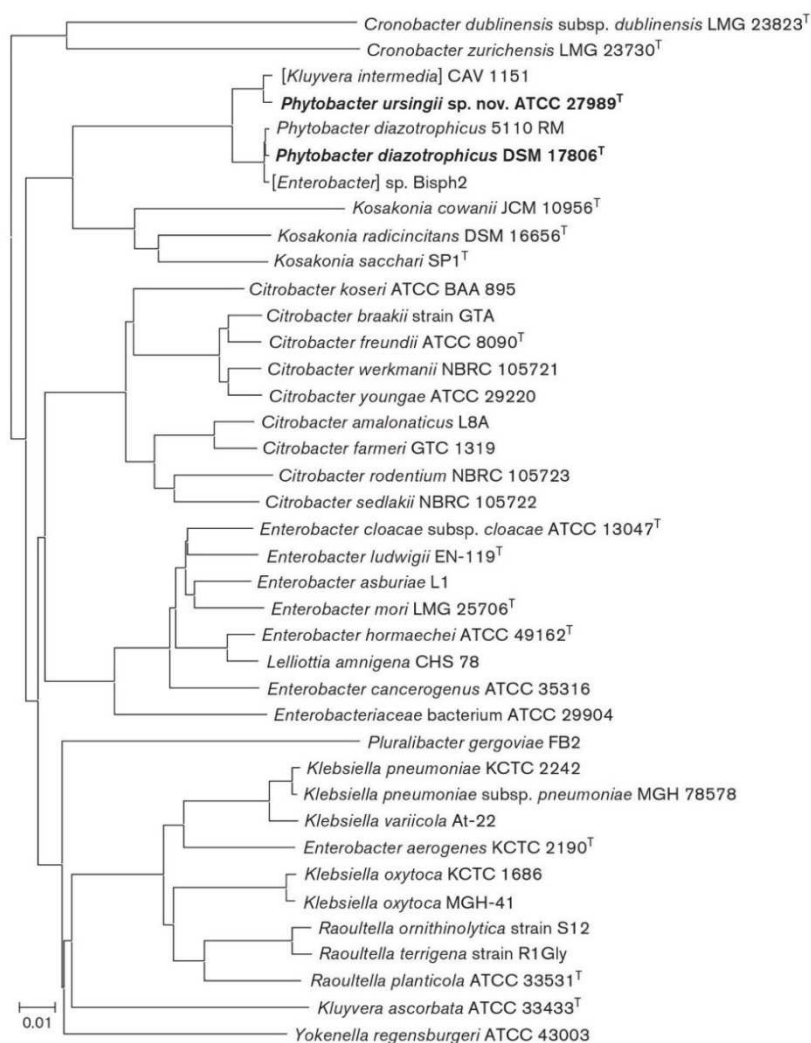


Fig. 3. Phylogenomic tree based on the core genomes of strains DSM 17806^T, 5110RM and ATCC 27989^T and strains of related species of the family Enterobacteriaceae as determined by EDGAR 2.0 (39). The genomes of the genus *Cronobacter* were used as outgroup. Owing to the large size of the alignments, bootstrap values could not be calculated for the tree. Bar, 1% nucleotide substitutions.

Minimal cut-off points of 70% DDH (Table 3) and 95% ANI (Table 4) were considered to represent species delineation. These values were exceeded within the clade that included strains *Phytobacter diazotrophicus* DSM 17806^T, isolate 5110RM and [*Enterobacter*] sp. Bisph2 as well as within the one composed by strain ATCC 27989^T and [*Kluyvera intermedia*] CAV1151. On the other hand, these thresholds were not attained when the comparison was performed between isolates belonging to the two different clades, although DDH and ANI values were larger than 45

and 90%, respectively, confirming that the two groups represent distinct species within the genus *Phytobacter*. In turn, the latter appears to be well distinct from nearby genera such as *Kosakonia*, *Citrobacter* or *Enterobacter*.

To conclude, the genotypic and phenotypic data presented in this study demonstrate that 5110RM, the archetypal isolate of the multi-state sepsis outbreak that occurred in 2013 in Brazil, belongs to *Phytobacter diazotrophicus*, a previously described species of which the type strain DSM 17806^T was originally isolated from wild-rice in China [32].

Table 3. Digital DDH of *Phytobacter* and *Kosakonia* species using Genome-to-Genome Distance Calculator (<http://ggdc.dsmz.de/>)

DDH values of over 70 % are shaded.

Strain	DSM 17806 ^T	5110RM	Bisph2	ATCC 27989 ^T	CAV1151	DSM 16656 ^T	SP1 ^T	JCM 10956 ^T
<i>Phytobacter diazotrophicus</i> DSM 17806 ^T	100	92.9	89.7	45.4	45.4	23.1	22.9	22.4
<i>Phytobacter diazotrophicus</i> 5110RM	92.9	100	90.5	45.4	46.6	23.4	22.9	22.6
[<i>Enterobacter</i>] sp. Bisph2	89.7	90.5	100	45.4	45.6	23.2	22.9	22.5
<i>Phytobacter ursingii</i> sp. nov. ATCC 27989 ^T	45.4	45.4	45.4	100	81.8	22.9	22.9	22.5
[<i>Kluyvera intermedia</i>] CAV1151	45.4	46.6	45.6	81.8	100	23.4	22.9	22.4
<i>Kosakonia radincitans</i> DSM 16656 ^T	23.1	23.4	23.2	22.9	23.4	100	27.6	25.4
<i>Kosakonia sacchari</i> SP1 ^T	22.9	22.9	22.9	22.9	22.9	27.6	100	25.9
<i>Kosakonia cowanii</i> JCM 10956 ^T	22.4	22.6	22.5	22.5	22.4	25.4	25.9	100

The same species includes further clinical strains isolated in the 1970s in the USA (i.e. ATCC 27981, ATCC 27990 and LMG 5337) [3, 6, 29–31], a 2015 multi-drug resistant isolate from a hospital environment in Curitiba (Brazil) unrelated to the 2013 multi-state sepsis outbreak, as well as Bisph2, a glyphosate-degrading bacterium isolated from a sandy soil in Algeria [34]. It is noteworthy that a nosocomial outbreak associated with contaminated anticoagulant citrate dextrose solution that occurred in 2010 in São Paulo (Brazil) can also be attributed to *Phytobacter diazotrophicus* on the basis of the *gyrB* gene sequence of the isolated bacteria, which was 100 % identical to ATCC 27990 [9], even if in the original report the organism was misidentified as *Pantoea agglomerans* [5]. Additionally, clinical strains CAV1151 [35], ATCC 27982 and ATCC 27989^T [3, 6, 29–31] were found to represent a novel species within the genus *Phytobacter*, for which the name *Phytobacter ursingii* sp. nov. is proposed.

EMENDED DESCRIPTION OF GENUS *PHYTOBACTER*

The description is based on Zhang *et al.* [32] and this study. Bacteria belonging to this genus are chemoorganotrophic and able to fix nitrogen. Cells are straight rods 0.6 to 0.7 by 0.9 to 1.2 µm in size, Gram-negative, facultative anaerobic and motile (amphitrichous flagella), non-encapsulated and non spore forming. Growth occurs at 5–40 °C, with an optimum between 28 and 37 °C. Optimum pH for growth is 7.0 (range is 4.0–10.0). Colonies growing on MacConkey agar

are circular, convex and smooth with non-entire margins and usually an elevated centre. They are not yellow-pigmented on nutrient agar up to 7 days of incubation and their diameter reaches 4 mm within 3 days at 30 °C. Bacteria are lactose fermenting, cytochrome oxidase negative and catalase positive. Glucose is fermented with the production of gas. Lysine and ornithine are not decarboxylated and arginine is not hydrolysed. The cells produce indole. Voges-Proskauer test, citrate (Simmon's) and aesculin are positive. Nitrate is reduced to nitrite. Cells do not produce hydrogen sulfide [Sulfide Indole Motility (SIM) agar] and urease. Acid is produced at 35±1 °C within 2 days from the following carbon sources: L-rhamnose, glycerol, L-arabinose, D-xylitol, D-mannitol, lactose, sucrose, trehalose, cellobiose, sorbitol, arabinose, dulcitol, salicin. Cells do not produce acid from inositol and melibiose. In contrast with what was reported by Zhang [32], all tested isolates including DSM 17806^T were tryptophan deaminase and lysine decarboxylase negative in the API20E test. *Phytobacter* can be differentiated from related enterobacterial genera based both on physiological and biochemical tests and on DNA hybridization. The type species is *Phytobacter diazotrophicus* [32].

EMENDED DESCRIPTION OF *PHYTOBACTER* *DAZOTROPHICUS*

The description of this taxon is the same as the genus *Phytobacter* cited above and is based on Zhang *et al.* [32] and this study. The type strain is *Phytobacter diazotrophicus* (=DSM

Table 4. Mean ANI values for all available genomes of *Phytobacter* and *Kosakonia* spp. as determined in EDGAR 2.0

ANI values over 95 % are shaded.

Strain	DSM 17806 ^T	5110RM	Bisph2	ATCC 27989 ^T	CAV1151	DSM 16656 ^T	SP1 ^T	JCM 10956 ^T
<i>Phytobacter diazotrophicus</i> DSM 17806 ^T	100	99.21	98.93	92.07	92.09	83.36	83.30	82.99
<i>Phytobacter diazotrophicus</i> 5110RM	99.21	100	98.99	92.04	92.36	83.53	83.18	82.93
[<i>Enterobacter</i>] sp. Bisph2	98.93	98.99	100	92.10	92.13	83.45	83.30	83.00
<i>Phytobacter ursingii</i> sp. nov. ATCC 27989 ^T	92.07	92.08	92.09	100	98.09	83.30	83.31	83.01
[<i>Kluyvera intermedia</i>] CAV1151	92.11	92.36	92.13	98.09	100	83.89	83.35	82.94
<i>Kosakonia radincitans</i> DSM 16656 ^T	83.36	83.57	83.41	83.23	83.79	100	85.75	84.62
<i>Kosakonia sacchari</i> SP1 ^T	83.42	83.40	83.33	83.26	83.36	85...83	100	84.99
<i>Kosakonia cowanii</i> JCM 10956 ^T	82.94	82.93	82.94	82.83	82.89	84.57	84.90	100

17806^T=LMG 23328^T) isolated from wild-rice *Oryza rufipogon* in China, but several clinical isolates are also reported. Based on spectrophotometry the DNA G+C content of type strain DSM 17806^T is 58.6±0.5 mol% [32]. Assembly of WGS data of DSM 17806^T yields a draft genome size of 5.58 Mb, with a DNA G+C content of 53.05 mol%.

DESCRIPTION OF *PHYTOBACTER URSINGII* SP. NOV.

Phytobacter ursingii (ur.sing'i.i. N.L. gen. masc. n. ursingii, named in honour of the Swedish bacteriologist and taxonomist Jan Ursing). The description of this taxon is the same as that of the genus *Phytobacter* cited above. The species can be differentiated from *Phytobacter diazotrophicus* on the basis of its ability to metabolize D-serine and L-sorbose. The type strain is *Phytobacter ursingii* ATCC 27989^T (=CNCTC 5729^T), isolated in 1974 from human sputum in South Carolina, USA. Other clinical sources have also been reported. Assembly of WGS data of ATCC 27989^T yielded a draft genome size of 5.75 Mb, with a DNA G+C content of 52.92 mol%.

Funding information

This study was supported by the Central Public Health Laboratory of Paraná (LACEN/PR), the Core for Advanced Molecular Investigation at Pontifícia Universidade Católica do Paraná (PUCPR), the FLASH Research Programme of the École polytechnique fédérale de Lausanne (EPFL) Leading House, within the 2013–16 Brazilian-Swiss Joint Research Programme (BSJRP) mandated by the State Secretariat for Education, Research and Innovation (SERI) and the Life Science and Facility Management Department at Zurich University of Applied Sciences (ZHAW).

Acknowledgements

We are thankful to Monica Dallman at the Core for Advanced Molecular Investigation at PUCPR and Joël Pothier at the Institute of Natural Resource Sciences at ZHAW Wädenswil for their support with NGS. We should further like to thank Karin O. Kalluff at LACEN/PR for technical assistance.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Kwon SW, Go SJ, Kang HW, Ryu JC, Jo JK. Phylogenetic analysis of *Erwinia* species based on 16S rRNA gene sequences. *Int J Syst Bacteriol* 1997;47:1061–1067.
- Cruz AT, Cazacu AC, Allen CH. *Pantoea agglomerans*, a plant pathogen causing human disease. *J Clin Microbiol* 2007;45:1989–1992.
- Rezzonico F, Smits THM, Montesinos E, Frey JE, Duffy B. Genotypic comparison of *Pantoea agglomerans* plant and clinical strains. *BMC Microbiol* 2009;9:204.
- Rezzonico F, Stockwell VO, Tonolla M, Duffy B, Smits TH. *Pantoea* clinical isolates cannot be accurately assigned to species based on metabolic profiling. *Transpl Infect Dis* 2012;14:220–221.
- Rezzonico F, Smits THM, Duffy B. Misidentification slanders *Pantoea agglomerans* as a serial killer. *J Hosp Infect* 2012;81:137–139.
- Brenner DJ, Fanning GR, Leete Knutson JK, Steigerwalt AG, Krichevsky MI. Attempts to classify Herbicola group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hybridization and phenotypic tests. *Int J Syst Bacteriol* 1984;34:45–55.
- Bergman KA, Arends JP, Schölvinck EH. *Pantoea agglomerans* septicemia in three newborn infants. *Pediatr Infect Dis J* 2007;26:453–454.
- Bicudo EL, Macedo VO, Carrara MA, Castro FF, Rage RI. Nosocomial outbreak of *Pantoea agglomerans* in a pediatric urgent care center. *Braz J Infect Dis* 2007;11:281–284.
- Boszczowski I, Nóbrega de Almeida Júnior J, Peixoto de Miranda EJ, Pinheiro Freire M, Guimarães T et al. Nosocomial outbreak of *Pantoea agglomerans* bacteraemia associated with contaminated anticoagulant citrate dextrose solution: new name, old bug? *J Hosp Infect* 2012;80:255–258.
- Gonçalves CR, Vaz TM, Leite D, Pisani B, Simoes M et al. Molecular epidemiology of a nosocomial outbreak due to *Enterobacter cloacae* and *Enterobacter agglomerans* in Campinas, São Paulo, Brazil. *Rev Inst Med Trop Sao Paulo* 2000;42:1–7.
- Habsah H, Zeehaida M, van Rostenberghe H, Noraida R, Wan Pauzi WI et al. An outbreak of *Pantoea* spp. in a neonatal intensive care unit secondary to contaminated parenteral nutrition. *J Hosp Infect* 2005;61:213–218.
- van Rostenberghe H, Noraida R, Wan Pauzi WI, Habsah H, Zeehaida M et al. The clinical picture of neonatal infection with *Pantoea* species. *Jpn J Infect Dis* 2006;59:120–121.
- Drancourt M, Bollet C, Cartioz A, Martelin R, Gayral J-P et al. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 2000;38:3623–3630.
- Hauben L, Moore ER, Vauterin L, Steenackers M, Mergaert J et al. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst Appl Microbiol* 1998;21:384–397.
- Pavan ME, Franco RJ, Rodriguez JM, Gadaleta P, Abbott SL et al. Phylogenetic relationships of the genus *Kluyvera*: transfer of *Enterobacter intermedius* Izard et al. 1980 to the genus *Kluyvera* as *Kluyvera intermedia* comb. nov. and reclassification of *Kluyvera cochleae* as a later synonym of *K. intermedia*. *Int J Syst Evol Microbiol* 2005;55:437–442.
- Dauga C. Evolution of the *gyrB* gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. *Int J Syst Evol Microbiol* 2002;52:531–547.
- Delmas J, Breyse F, Devulder G, Flandrois J-P, Chomar M. Rapid identification of *Enterobacteriaceae* by sequencing DNA gyrase subunit B encoding gene. *Diagn Microbiol Infect Dis* 2006;55:263–268.
- Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J et al. Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Syst Appl Microbiol* 2008;31:447–460.
- Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* 2015;38:237–245.
- Rosselló-Móra R, Amann R. Past and future species definitions for *Bacteria* and *Archaea*. *Syst Appl Microbiol* 2015;38:209–216.
- Dunne WM, Westblade LF, Ford B. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *Eur J Clin Microbiol Infect Dis* 2012;31:1719–1726.
- Tamura K, Sakazaki R, Kosako Y, Yoshizaki E. *Leclercia adecarboxylata* gen. nov., comb. nov., formerly known as *Escherichia adecarboxylata*. *Curr Microbiol* 1986;13:179–184.
- Gavini F, Mergaert J, Beji A, Mielcarek C, Izard D et al. Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and Description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* 1989;39:337–345.
- Mergaert J, Verdonck L, Kersters K. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. *Int J Syst Bacteriol* 1993;43:162–173.

25. Harada H, Oyaizu H, Kosako Y, Ishikawa H. *Erwinia aphidicola*, a new species isolated from pea aphid, *Acyrtosiphon pisum*. *J Gen Appl Microbiol* 1997;43:349–354.
26. Brady CL, Cleenwerck I, Venter SN, Engelbeen K, de Vos P et al. Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrinea* sp. nov., *Pantoea breneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. *Int J Syst Evol Microbiol* 2010;60:2430–2440.
27. Brady C, Cleenwerck I, Venter S, Coutinho T, de Vos P. Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii*, *E. radincintans*, *E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as *Kosakonia cowanii* comb. nov., *Kosakonia radincintans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis*, *E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. *Syst Appl Microbiol* 2013;36:309–319.
28. Minogue TD, Daligault HE, Davenport KW, Bishop-Lilly KA, Bruce DC et al. Draft genome assemblies of *Enterobacter aerogenes* CDC 6003-71, *Enterobacter cloacae* CDC 442-68, and *Pantoea agglomerans* UA 0804-01. *Genome Announc* 2014;2:e01073-14.
29. Maki DG, Rhame FS, Mackel DC, Bennett JV. Nationwide epidemic of septicemia caused by contaminated intravenous products. *Am J Med* 1976;60:471–485.
30. Ewing WH, Fife MA. *Enterobacter agglomerans* (Beijerinck) comb. nov. (the Herbicola-Lathyri bacteria). *Int J Syst Bacteriol* 1972;22:4–11.
31. Felts SK, Schaffner W, Melly MA, Koenig MG. Epidemiologic, clinical, and laboratory investigation of an outbreak in one hospital. *Ann Intern Med* 1972;77:881–890.
32. Zhang GX, Peng GX, Wang ET, Yan H, Yuan QH et al. Diverse endophytic nitrogen-fixing bacteria isolated from wild rice *Oryza rufipogon* and description of *Phytobacter diazotrophicus* gen. nov. sp. nov. *Arch Microbiol* 2008;189:431–439.
33. Rezzonico F, Vogel G, Duffy B, Tonolla M. Application of whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification and clustering analysis of *Pantoea* species. *Appl Environ Microbiol* 2010;76:4497–4509.
34. Benstama O, Boulahrouf A. High-quality draft genome sequence of *Enterobacter* sp. Bisph2, a glyphosate-degrading bacterium isolated from a sandy soil of Biskra, Algeria. *Genom Data* 2016;8:61–66.
35. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A et al. Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene *bla_{KPC}*. *Antimicrob Agents Chemother* 2016;60:3767–3778.
36. Rezzonico F, Smits TH, Born Y, Blom J, Frey JE et al. *Erwinia gerundensis* sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *Int J Syst Evol Microbiol* 2016;66:1583–1592.
37. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
38. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T et al. GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* 2003;31:2187–2195.
39. Blom J, Kreis J, Spänig S, Juhre T, Bertelli C et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 2016;44:W22–W28.
40. Henz SR, Huson DH, Auch AF, Nieselt-Struwe K, Schuster SC. Whole-genome prokaryotic phylogeny. *Bioinformatics* 2005;21:2329–2335.
41. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P et al. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81–91.
42. de Maayer P, Cowan DA. Flashy flagella: flagellin modification is relatively common and highly versatile among the *Enterobacteriaceae*. *BMC Genomics* 2016;17:377.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.

**ANEXO III: CAPITULO DO LIVRO "BIOÉTICA E PSICOLOGIA: INTER-
RELAÇÕES"**

**BIOÉTICA, BIOTECNOLOGIA
E BIOPODER:
sutis (bio)relações**

*Thereza Salomé D' Espindula
Helena Regina Salomé D' Espindula
Daiane Priscila Simão-Silva*

O biopoder, manifestação de poder estudada por Foucault e definido como um poder da vida e sobre a vida, está presente em contextos que têm relação com saúde/doença, na medida em que controla a saúde dos corpos, onde doença, tratamento e cura são tidos como problemas ao mesmo tempo políticos e econômicos. Submetidos ao biopoder, indivíduo e população são alvos de saber e objetos da medicina, que passa a atuar nas necessidades mais concretas dos seres humanos. O desenvolvimento gigantesco da biotecnologia no último século trouxe maior facilidade em identificar doenças do corpo – aquelas devidamente localizadas em órgãos e capazes de serem detectadas por exames específicos – mas ainda não eximiu o ser humano de uma série de sofrimentos.

De modo diverso, a ciência da psicologia preocupa-se com o indivíduo em suas questões subjetivas, em sua necessidade de auxílio para escuta e acolhimento. A saúde, conforme preconizada pela Organização Mundial da Saúde (OMS) é um “estado de completo bem-estar físico, mental e social e não somente ausência de afeções e enfermidades” (WHO [1946] 2005, p. 1). Ainda que haja controvérsias sobre a possibilidade de alguém ser tão completamente saudável a ponto de fazer jus a esse conceito, não há como negar que uma assistência à saúde cindida em cuidados com o corpo e cuidados com o indivíduo em si mesmo não trará vantagens ou as trará de forma bastante limitada.

Segundo Bezerra Jr. (2015, p. 44), a biopolítica e o biopoder constituem um pano de fundo numa paisagem mundial contemporânea permeada por inúmeros elementos. É neste cenário que se configura o desafio de compreender a “complexidade e as ambiguidades e contradições desse processo” como ponto de partida para pensar e agir de maneira crítica. Desta forma, o presente capítulo objetiva analisar criticamente o biopoder, com suas biotecnologias, em relação à Psicologia e à Bioética, no que diz respeito às condições de desigualdade que possa gerar. Neste contexto, pretende-se apresentar como a biotecnologia instrumentaliza essa forma de poder, assim como

discutir as relações estabelecidas entre a Psicologia, a Bioética e o *biopoder*. O estudo consistiu numa revisão de literatura e análise crítica pautada no conceito de biopoder de Foucault. O texto inicia com uma breve introdução ao contexto histórico de como a saúde e a doença vêm sendo compreendidas socialmente até chegar ao ponto em que o conhecimento sobre a saúde é utilizado como uma forma de poder sobre a vida, o *biopoder*.

1. Contexto histórico saúde/doença

A doença, para os povos primitivos, não se estabelecia numa relação de causa e efeito; mesmo ao longo da Idade Média, ela ainda era vista como um sinônimo de ritual de purificação, cuja cura só era atingida caso houvesse o merecimento. As características atribuídas tanto à saúde quanto às doenças ainda não eram suficientes para explicar-lhes a origem e a distribuição.

O conceito de corpo social, desenvolvido a partir do fim do século XVII, amadurece também o processo de formação de algo que viria a ser a saúde pública. (SCLAR, 1987). Assim, surge também a concepção da relação política entre saúde e sociedade, entre corpo individual e corpo social, com ações de saúde pública ocorrendo tanto sobre as pessoas quanto sobre o meio ambiente. A medicina passa a explorar os cadáveres, buscando a localização mais precisa das doenças: “A partir do cadáver, paradoxalmente se percebe a doença viver” (FOUCAULT, 2003, p. 169).

O século XIX traz consigo uma transformação relativa à organização e gestão do poder, reorganizando-o em suas relações. Com o advento da sociedade industrial, o desenvolvimento social já não poderia atuar de forma autônoma; problemas oriundos de um rápido crescimento da urbanização, a explosão demográfica e um crescente desenvolvimento científico e tecnológico em praticamente todas as áreas do conhecimento, apontam para a necessidade de uma tecnologia de intervenção social. Isto é, um poder regulador da vida, sendo seu objeto específico de interesse o homem enquanto ser vivo e espécie biológica, o qual pudesse estar, ao mesmo tempo, submetido à produção e preparado para atividades que envolvessem o consumo.

Relações entre ambiente social e sobrevivência principiam. O conceito de corpo social, supostamente baseado na divisão do trabalho, na realidade explica um sistema de poder, tendo em vista que a sociedade vem adquirindo rapidamente um maior grau de

complexidade e de organização, a qual clamará por registros, dados e cálculos estatísticos, capazes de serem contabilizados, tabulados e mesmo comparados a fim de criar indicadores. O raciocínio estatístico ampliou-se, levando as observações a serem colhidas nos locais de ocorrência das doenças e/ou epidemias. Dessa forma, os indicadores poderiam servir de base para estratégias de intervenção sobre populações (D'ESPÍNDULA, 2011).

Como uma nova forma de poder regulador da vida e sobre a vida, indicadores de natalidade, mortalidade, controle de patologias e outras formas de controle serão pensadas pelo viés econômico, assim como as intervenções sobre a criança, o adolescente, o adulto e o idoso, o trabalho, a urbanização e a higiene. Todas elas tomadas como estratégias de poder *da* vida e *sobre* a vida, que possibilitam um controle das populações. Os números e quantidades aceitos como de referência serão norteadores para estratégias econômicas e de saúde, fazendo a cisão entre aquilo que será considerado dentro de um padrão normal ou não. O conhecimento na área de saúde vem avançando e, desde então, aumentando cada vez mais a possibilidade de atuação de uma nova espécie de poder, ao mesmo tempo um poder *da* vida e *sobre* ela: o biopoder (D'ESPÍNDULA, 2011).

2. O Biopoder

As estratégias biopolíticas apresentam uma estreita ligação entre os quadros epidemiológicos das diferentes partes do mundo, e já se faz mister a exigência de uma luta de grandes proporções contra as doenças. Para Foucault, a biopolítica deve exercer uma relação positiva com a vida, para que seja possível aumentar tanto a força quanto o equilíbrio das populações, eliminando ameaças à vida (D'ESPÍNDULA, 2011). O corpo sujeito ao biopoder está também sujeito a ser entendido como algo relativo, como corpo capaz de ser moldado às exigências deste.

Foucault atribuía uma atuação capilar ao biopoder, isto é, como um poder que não está nas mãos de um ou de outro, mas disperso por entre todos e incorporado à trajetória das nações. Desse modo, a área de saúde, como área que toma para si o cuidado com a vida humana, articula o conhecimento científico e as estratégias de poder correspondentes aos corpos dos indivíduos e aos processos biológicos. As práticas sociais sobre corpos e populações são o exercício da biopolítica em sua plenitude. Com alvos do saber, particularmente do saber médico, a ação se dá voltada para o auxílio do ser humano na

manutenção da saúde, com vistas a uma produtividade que não necessite ser interrompida e a um consumo sem trégua, ambos baseados nas mais modernas conquistas científicas.

A biopolítica, segundo Rose (2014), pode ser compreendida como uma disputa pela vitalidade:

penso em todas aquelas disputas sobre a administração da vida, a modelagem da vida, a organização da reprodução, o acesso ao aborto, à testagem genética, se os indivíduos deveriam ou não ser geneticamente testados, ou se deveriam ter acesso à assistência médica personalizada, e assim por diante. Desse modo, tenho argumentado que, cada vez mais, a política – a biopolítica – se tornou uma política em torno de questões vitais (2014, p. 10).

As questões vitais trazem novas formas de compreender e responder aos anseios humanos, a partir da linguagem da biologia ou da biomedicina. Nesse aspecto, o corpo humano é compreendido como um tipo de máquina vital: “elas acabaram se tornando uma característica da biopolítica contemporânea” (ROSE, 2014, p. 10).

Neste contexto, os saberes oriundos da área de saúde, em particular o saber médico, são cada vez mais sinônimos de uma verdade irrefutável; assentam-se sobre um desenvolvimento científico e biotecnológico jamais visto, o qual faz crer aos indivíduos vulnerabilizados por enfermidades, tanto quanto aos profissionais de saúde, que não podem prescindir deles.

3. Desenvolvimento Biotecnológico

Os saberes da medicina ganham cada vez mais espaço e confiança, na medida em que se tornam mais minuciosos os diagnósticos e mais complexa a terapêutica. E o mesmo contexto que propiciou a evolução da medicina e os avanços da biotecnologia contribuiu sobremaneira para uma progressiva medicalização do Estado, das cidades e da população. Estruturada sob um discurso dicotômico de saúde-doença e/ou normal-patológico, a área de saúde se retroalimenta, fazendo com que o indivíduo só venha a se sentir melhor após uma intervenção.

A vida, nesta esfera, vem ganhando um papel de destaque, como objeto de preocupação a partir de decisões relativas a corpos doentes. Constata-se um aumento no uso de medicamentos, ao que parece, por absoluta intolerância à frustração; ao enfrentar situações que outrora eram resolvidas deixando que as emoções viessem à tona, atualmente faz-se necessário evitar o estado depressivo, ansioso e se desvencilhar da dor, com a ajuda dos medicamentos que “normalizam” comportamentos e eliminam sintomas

de sofrimento, negligenciando a busca de significado (D'ESPÍNDULA, 2013). Neste sentido, o homem nada mais é do que um ser ontologicamente enfermo, para o qual o desenvolvimento biotecnológico se mostra capaz de atender aos anseios de cura, e também aos de consumo, fazendo surgir novas formas de risco e de controle assim como de desigualdade.

A desigualdade se apresenta também no que concerne ao desfrute de recursos da medicina de ponta, que dispõe de terapias de alto custo, como, por exemplo, de medicamentos para fins oncológicos, ou terapias com uso de anticorpos monoclonais, células tronco, ou ainda testes de diagnóstico genético, ou mesmo em procedimentos estéticos como realização de cirurgias plásticas, uso da toxina botulínica ou silicone. Frequentemente até mesmo a justiça é chamada a intervir na liberação de fármacos ou recursos médicos, estejam eles disponíveis ou não no país, denotando uma busca por equidade ao que se apresenta como novas formas de geração de desigualdade de acesso e distribuição.

Dores emocionais e físicas se entrelaçam e necessitam de maior acolhimento, a fim de reverter quadros de ansiedade, depressão, doenças físicas e outros, os quais vistos de uma forma puramente biológica, só contribuem para beneficiar a indústria farmacêutica, muito pouco oferecendo ao portador do corpo doente. O ser humano carece de sentido de vida e busca solução a curto prazo, pois, na celeridade imposta ao cotidiano e à vida, de pouco ou nenhum tempo dispõe para reflexão. É neste contexto que, em seu papel terapêutico, o psicólogo deverá ter o cuidado de promover um bem-estar baseado no autoconhecimento, mostrando o quanto esse processo é ou não natural, percebendo e fazendo perceber ao atendido as nuances e sutilezas impostas pelo biopoder, sua “rotulação” e medicalização/terapeutização.

O biopoder na atualidade, por meio de suas biotecnologias, passa a apresentar o aperfeiçoamento como uma forma de curar a humanidade de si mesma. E curar não apenas no sentido de restabelecer a saúde perdida, mas como forma de superação de limites impostos pela própria natureza humana. O aperfeiçoamento biotecnológico da “natureza humana” é defendido por alguns autores como uma obrigação moral de promoção de qualidade de vida e da saúde, ou ainda como uma forma de “resistir ao desaparecimento. Por isso, é justo intervir, não devemos ter medo,” afirma o bioeticista Maurizio Mori (2003, *online*).

A biotecnologia fornece uma gama de oportunidades de cura e aprimoramento das capacidades humanas, por meios terapêuticos que passaram a ser utilizados para fins não

terapêuticos. Como, por exemplo, a ingestão de anticolinesterásico (medicamento utilizado para atraso na progressão da doença de Alzheimer) para melhoria da memória em jovens, o uso de hormônios visando o maior crescimento de crianças de estatura mediana e o uso de inibidores de apetite e suplementos alimentares para um corpo mais belo. Outro exemplo é a aplicação da terapia gênica em esportistas, resultando em *doping* genético. Alguns autores defendem como apropriado “usar meios médicos para prevenir o sofrimento e melhorar o bem-estar, mesmo que a fonte do problema não seja necessariamente uma doença” (PURDY, 2001, p. 258). Deste modo, é identificado o conflito ético que os sociólogos têm chamado de “medicalização” para se referir ao processo pelo qual, problemas “não médicos” são compreendidos e tratados como problemas “médicos” (CONRAD, 2007).

Os avanços biomédicos passam a ser uma necessidade em todas as esferas, até mesmo na geração da vida. As técnicas de reprodução assistida (RA), propostas como uma importante resolução de problemas procriativos, cada vez mais estão sendo ofertadas no contexto de possibilidades de testagem e seleção embrionária. Segundo Knoppers e Isasi (2004) o uso de testes genéticos pré-implantacionais (PGD) para fins não-médicos estão aumentando, principalmente para seleção do sexo de embriões exclusivamente por razões sociais ou culturais.

Num contexto cultural de valorização do presente, sem pensar no passado ou futuro, com foco na realização materialista e imediatista, na cura ou aperfeiçoamento de si ou de sua prole, podemos ver o paradoxo de busca de identidade e autenticidade permeada por fantasias de melhorar o desempenho individual, buscando assim a transformação de características individuais. É o que Bauman descreve como uma busca frenética por identidade e felicidade, de forma individualizada e consumidora:

uma vez que os bens capazes de tornar a vida mais feliz começam a se afastar dos domínios não-monetários para o mercado de mercadorias, não há como os deter; o movimento tende a desenvolver um impulso próprio e se torna autopropulsor e autoacelerador, reduzindo ainda mais o suprimento de bens que, pela sua natureza, só podem ser produzidos pessoalmente e só podem florescer em ambientes de relações humanas intensas e íntimas (2009, p. 16).

Desconsidera-se, nesse imediatismo, que a saúde pode envolver inúmeros fatores, ações e serviços, exigindo uma avaliação correta da condição do usuário de saúde. Isso inclui um sem número de estratégias – como o acesso à informação e a tratamentos adequados – concentradas na vida e direcionadas à população em conjunto.

4. Sutis (bio)relações

O campo da psicologia não difere das demais áreas da saúde em relação ao biopoder; o psicólogo, em sua relação com o atendido, sujeita ambos às relações de poder vigentes. Embora ao psicólogo seja vedada a prescrição de medicamentos ou exames, o que por si só reduz as formas coercitivas do biopoder, há ainda os testes de seu uso exclusivo, os inventários e as escalas, capazes de enquadrar o ser humano numa dada tabela, fornecendo indicadores de sua personalidade, suas aptidões, seu perfil familiar, dentre outros. Mesmo psicólogos que não trabalham utilizando esse tipo de material, podem estar presentes em escolas, empresas, hospitais, setores como os de Recursos Humanos e outros locais onde, de alguma forma, o atendido procura adquirir um comportamento mais apropriado, mais dentro do esperado pela sociedade, que faça cessar suas angústias e sofrimentos. Percebe-se assim o psicólogo, tanto quanto seu atendido, inseridos em contextos institucionais que vão exigir de ambos certos posicionamentos consoantes com essas instituições.

Prado Filho e Trisotto lembram que “porque a Psicologia, ao mesmo tempo em que se dedica ao homem, acaba tomando-o por objeto do seu olhar e das suas práticas, exercendo quotidianamente poderes efetivos sobre ele” (2006, p. 46). Com a exigência de se portar como ser social e sociável, o homem de hoje deixa de lado a preocupação com a salvação da alma – alvo de sua inquietude de outros tempos – e passa a uma preocupação com o corpo de forma nunca dantes imaginada; o corpo passa a pertencer ao indivíduo como foco do olhar e dos saberes emergentes. Nele se implantam as patologias, tanto quanto as normalidades, numa valorização condizente com o padrão de desenvolvimento biotecnológico atual. Inserido num torvelinho de cuidados e recompensas ditados pelo mercado, o corpo passa a ser alvo de sacrifícios, submetido a regras infundáveis de bem-estar e qualidade de vida, regras estas muitas vezes ditadas mais pela mídia do que propriamente por necessidades individuais.

A fim de ser aceito e valorizado, o homem do século XXI submete seu corpo a disciplinas constantes em busca do corpo perfeito, dispensando pouca ou nenhuma atenção para questões emocionais vinculadas a essas exigências cotidianas. O homem, numa eterna e inalcançável busca por um corpo sem falhas, relega a segundo plano pontos cruciais do ciclo vital que também estão ligados ao corpo, como a dor, o sofrimento, a felicidade e a morte. Age como se fosse possível manter-se afastado de tais condições,

como se elas não fizessem parte do viver humano. É bem verdade que a dor já pode, em muitos casos, ser minimizada; porém, para o sofrimento, o processo não se dá da mesma forma.

Quando o corpo se esforça por corresponder às expectativas de seus pares e tenta alcançar os modelos impostos, numa incessante busca pela saúde e perfeição, ele está, ao mesmo tempo, inserido em uma realidade biopolítica, visto que sofre sacrifícios tão intensos quanto num passado distante, porém de formas muito mais sutis, consumindo alimentos, exercícios, exames e medicamentos em busca utópica de saúde perfeita e satisfação total consigo mesmo. Visto que tais atitudes tendem facilmente a impedir o aflorar natural das emoções, desencadeiam-se amiúde fenômenos como depressão, queda na autoestima, distorções na autoimagem e outros, dos quais a psicologia irá se ocupar, devendo, por sua formação comprometida com a ética, desenvolver uma ação responsável perante o atendido, levando em consideração seu sofrimento psíquico, suas necessidades e sua singularidade.

Nesse cenário, há um aumento no diagnóstico e classificação de transtornos de cunho psíquico. Mais sutis e menos mensuráveis em termos diagnósticos, estes conseqüentemente apresentam-se mais tênues no que diz respeito aos limites do normal-patológico. Por exemplo, há alguns anos crianças muito agitadas geralmente eram simplesmente agitadas; posteriormente, passam a ser questionados seus comportamentos e a enquadrá-los nos critérios para o diagnóstico de Transtorno de Déficit de Atenção e Hiperatividade (TDAH). Recentemente esses diagnósticos se elevaram a ponto de chegar quase ao nível epidêmico (AYCNP, 2012). Parte do aumento decorre do desenvolvimento de tecnologia diagnóstica, parte é devida ao maior conhecimento e divulgação dos sintomas. Contudo, questiona-se o quanto destes diagnósticos são resultantes da necessidade, imposta pelo biopoder, de se ter crianças saudáveis e produtivas em termos acadêmicos, ou quiçá ainda crianças que permitam um convívio pacífico em espaços limitados como apartamentos.

Esse tipo de movimentação acaba por gerar vezos de transtornos diversos e, por conseguinte, tratamentos que prometem melhoras ao indivíduo, levando-o a ser medicamentosamente menos triste, mais tranquilo ou mais capaz. A medicalização deixou de atuar apenas na esfera do bem-estar físico e passa a se responsabilizar por padronizar sensações e estados mentais, inclusive frustrações causadas por ela mesma, na busca do indivíduo por uma perfeição utópica. Retroalimentada e biotecnologicamente

embasada, ela atinge indivíduos e profissionais da saúde que passam a tratar com medicamentos, estados outrora naturais.

Ao longo de suas formações, ao psicólogo, tanto quanto ao psiquiatra – profissionais da saúde mental – faz-se necessária a inclusão de discussões de aspectos e de conflitos éticos, viabilizando atitudes seguras em suas práticas, desenvolvendo assim a competência para uma intervenção responsável no que concerne à relação com aqueles que os procuram em busca de auxílio, numa ética do cuidado que não pode deixar de estar presente, pois “o sofrimento não cuidado pode levar a processos autodestrutivos.” (KOVÁCS, 2008, p. 163) Estes profissionais, diante do sofrimento psíquico de seus atendidos, deverão ter sua atuação pautada em uma atitude tão humanizada quanto ética, capaz de proporcionar ao mesmo, bem como a seus familiares, uma atenção especial, capaz de atenuar ou mesmo extinguir esse sofrimento. Contudo, é preciso lembrar que:

os avanços da ciência tornam a relação do profissional de saúde mental com seus pacientes mais complexa do que em outros tempos. A busca pelo respeito à dignidade e integridade do outro são aspectos básicos no fazer de todo psicólogo (DIAS et al., 2007, p. 133).

Ao longo da psicoterapia, o paciente tenderá a expor sua intimidade ao terapeuta, o que novamente vai remeter este último às questões relativas ao sigilo. Este fato ressalta a importância da reflexão profissional acerca dos dilemas éticos, visto que pode mobilizar conteúdos internos do atendido, com os quais este pode ou não estar disposto ou preparado para entrar em contato. No processo da psicoterapia, tanto quanto na Bioética, é preciso um respeito à autonomia e um cuidado em não provocar o mal. Ainda, uma especial atenção a responsabilidade para com o atendido em suas necessidades, restaurando-lhe o bem-estar e apontando-lhe os valores que lhe parecerem mais adequados. O processo psicoterapêutico apresenta-se permeável às questões intra e extra sujeito; em tempos de biopoder, é necessário estar atento às influências externas, como da mídia, e a repercussão dessas em cada atendido.

Nesse sentido, a Bioética torna-se um lugar privilegiado para o exercício de uma crítica, pautada em valores éticos, os quais deverão estabelecer os limites entre aquilo que é possível fazer e aquilo que efetivamente deverá ser feito, numa tentativa de oferecer a indivíduos e comunidades o melhor para todos.

Considerações finais

Ao interpretar a vida e a saúde como bens de consumo, o biopoder se mostra capaz de propor e ampliar desigualdades, aumentando a complexidade das questões éticas envolvidas. Aquilo que é ofertado pela biotecnologia como o melhor dos bens de consumo passa a ser objeto de desejo. Contudo, como não está acessível a toda população, amplia-se a desigualdade, fragilizando mais ainda as parcela(s), extremamente vulnerável(is), em meio a essa relação multifacetada de política, poder, tecnologia e saúde.

Muitos sofrem ao ter negado o exercício de sua autonomia, seus desejos e aspirações, principalmente em se tratando de pacientes com prognóstico reservado. Estes são muitas vezes sujeitos ao privilégio terapêutico – ocasião em que a equipe de saúde pensa saber o que é melhor para o paciente – ou a medidas desproporcionais de prolongamento da vida, ou quiçá prolongamento da morte e do sofrimento.

O texto em tela se propõe a refletir acerca do cuidado e da responsabilidade dos profissionais de saúde em seus contatos com os atendidos. No torvelinho de relações entre equipe de saúde, paciente e família, o biopoder se apresenta por vezes de forma sutil. Em que pesem as diversas características inerentes a cada uma das profissões que compõem a área, o psicólogo precisa estar particularmente atento aos apelos midiáticos, que podem trazer sofrimento psíquico àqueles que colocam como objetivo o alcance de parâmetros quase sempre inalcançáveis de beleza, felicidade ou padrão de vida. Ao profissional da psicologia cabe zelar pela saúde mental daqueles que o procuram, auxiliando-os a redesenhar suas vidas e garantindo-lhes o sigilo das informações, o atendimento adequado e o acolhimento. Um maior conhecimento da Bioética será de grande valia no manejo dessas questões, juntamente com o auxílio de outros profissionais da área de saúde, pois em se tratando do biopoder, garantir a ética, minimizando desigualdades desde a pesquisa básica até o uso de mercado é, sem dúvida, uma tarefa que exige uma reflexão cuidadosa e multidisciplinar.

REFERÊNCIAS

AYCNP. ASSOCIATION FOR YOUTH, CHILDREN AND NATURAL PSYCHOLOGY **Superando el Trastorno por Déficit de Atención con Hiperactividad (TDAH) Sin Medicación: Guía para Padres y Educadores.** NJ, Newark: Education y Psychology Publications, 2012.

BAUMAN, Z. **A arte da vida.** Trad. de Carlos Alberto Medeiros. Rio de Janeiro: J. Zahar, 2009.

BEZERRA JR., B. Biopolítica e biologia, hélices do DNA contemporâneo. **IHUOn-Line**. São Leopoldo, a. XV, n. 472, p. 43-47, 14 set. 2015. Entrevista a Márcia Junges e Ricardo Machado. Disponível em: <<http://www.ihuonline.unisinos.br/media/pdf/IHUOnlineEdicao472.pdf>>. Acesso em: jan. 2017.

CONRAD, P. **The Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders**. Baltimore: The Johns Hopkins University Press, 2007. 204p.

D'ESPÍNDULA, T. S. **O paradoxo do cuidado da vida e do controle do patológico em Michel Foucault**. 2011. 145f. Dissertação (Mestrado em Filosofia) – Pontifícia Universidade Católica do Paraná. Curitiba: PUCPR, 2011.

D'ESPÍNDULA, T. S. Judicialização da medicina no acesso a medicamentos: reflexões bioéticas. **Revista bioética**, v. 21, n. 3, p. 438-447, 2013.

DIAS, H. Z. J. et al. Psicologia e bioética: diálogos. **Psic Clín**. Rio de Janeiro, v. 19, n. 1, p. 125-135, 2007.

FOUCAULT, M. **O nascimento da clínica**. Trad. de Roberto Machado. 5. ed. Rio de Janeiro: Forense Universitária, 2003.

KNOPPERS, B. M.; ISASI, R. M. Regulatory approaches to reproductive genetic testing. **Hum Reprod**, n. 19, p. 2695-2701, 2003.

KOVÁCS, M. J. Em cuidados psicológicos e psicoterápicos. In: SIQUEIRA, J. E.; ZOBOLI, E.; KIPPER, D. J. **Bioética clínica**. São Paulo: Gaia, 2008.

MORI, M. O controle da transmissão de vida. **Centro de Bioética do CREMESP** Entrevistas. 05 set. 2003. Disponível em: <<http://www.bioetica.org.br/?siteAcao=EntrevistaIntegra&id=21>>. Acesso em: 25 fev. 2015

PRADO FILHO, K.; TRISOTTO, S. Psicologia, ética e bioética. **Psychologies Argumento**. Curitiba, v. 24, n. 47 p. 45-48, out./dez. 2016.

PURDY, L. Medicalization, Medical Necessity, and Feminist Medicine. **Bioethics**, v. 15, n. 3, p. 248–261, 2001.

ROSE, N. Biopolítica e complexidade – Da cidadania biológica à ética somática. **IHUOn-Line**. São Leopoldo, a. XIV, n. 456, p. 9-13, 20 out. 2014. Entrevista concedida a Márcia Junges e Andriolli Costa. Trad. Isaque Gomes Correa. Disponível em: <http://www.ihuonline.unisinos.br/index.php?option=com_content&view=article&id=5714&secao=456>. Acesso em: abr. 2015.

SCLIAR, M. **Do mágico ao social: uma trajetória da saúde pública**. Porto Alegre, L&PM, 1987.

UNIVERSIDADE DO VALE DO RIO DOS SINOS, Instituto Humanitas Unisinos, **Cadernos IHU ideias**. São Leopoldo, a. 1, n. 1, 2003. Disponível em:

<<http://www.ihu.unisinos.br/images/stories/cadernos/ideias/229cadernosihuideias.pdf>>.
Acesso em: 05 abr. 2016.

WHO. OMS - Organização Mundial de Saúde. **Constitution of the World Health Organization.** 18p. Disponível em:
<<http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf>> Acesso em: 06 abr. 2016.

**ANEXO IV: MANUSCRITO "THE PRATA VILLAGE: AN ESTIMATE OF THE
INCUBATION PERIOD AND THE PROPORTION OF NATURALLY SUSCEPTIBLE
INDIVIDUALS TO LEPROSY IN A HUMAN POPULATION"**

Insights in elusive aspects of leprosy epidemiology: Prata Village model

Ciane Cristina de Oliveira Mackert^{1,*}, Fernando Panissa Lázaro¹, Márcia Olandowski¹, Helena Regina Salomé D'Espindula¹, Priscila Verchai Uaska Sartori¹, Rafael Saraiva de Andrade Rodrigues¹, Geison Cambri¹, Marília Brasil Xavier², Erwin Schurr³, Alexandre Alcais⁴, Marcelo Távora Mira^{1,5}

* In memoriam

¹ Core for Advanced Molecular Investigation, Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brazil;

² Core for Tropical Medicine, Federal University of Pará, Belém, Pará;

³ McGill Centre for the Study of Host Resistance, McGill University, Montreal, Canada;

⁴ Laboratory of Human Genetics of Infectious Diseases, Institut National de la Santé et de la Recherche Médicale and University Paris René Descartes, Necker Medical School, Paris, France;

⁵ School of Health and Biosciences, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brazil.

Reprints and Correspondence: Dr. Marcelo Távora Mira, Programa de Pós-Graduação em Ciências da Saúde, Escola de Medicina, Pontifícia Universidade Católica do Paraná. Rua Imaculada Conceição, 1155, Prado Velho; CEP: 80215-901, Curitiba, Paraná, Brazil. Email: m.mira@pucpr.br

Abstract

Introduction: It has been widely accepted that only a proportion of individuals exposed to an infectious agent develop the clinical disease after a period of incubation. However, the exact estimation of the incubation period and the proportion of individuals naturally susceptible to an infectious disease in a human population can be an epidemiological challenge, particularly when the pathogen is fastidious. That is the case for leprosy, caused by a bacteria that didn't grow in artificial media. Here we infer some epidemiological parameters on leprosy from a hyper-endemic, former leprosy colony, the Prata Village. The Village is located on the outskirts of the Amazon forest in the north of Brazil and it was a site for isolation of leprosy-affected individuals in the 1920s and remains isolated since then.

Methods: The entire Prata Village population was interviewed to obtain the clinical, socio-economical and demographical data. The clinical variables were cross-checked using three independent information sources. Validated data was used for a population-based, descriptive epidemiological analysis, estimation of the incubation period and the proportion of individuals naturally susceptible.

Results: From a total of 2,005 inhabitants by the time of the enrollment, 1,084 (56.19%) were born at the Village and therefore, have been under constant, heavy and lifelong exposure to the bacteria. The gender distribution in affect people was homogeneous. Among the people born at the Village, 5.9% have been affected by leprosy, which we propose as an estimate of the proportion of leprosy susceptible individuals. They also had a median age of diagnosis of 15 years old what we propose as an estimate of the incubation period of the disease.

Conclusion: Due to the unique characteristics of the Prata Village population, we believe we were able to obtain, an accurate estimate for the incubation period and the proportion of naturally susceptible individuals to leprosy in a human population.

Keywords: leprosy; population; epidemiology; incubation period; innate susceptibility

Introduction

It has been classically accepted that only a proportion of individuals exposed to an infectious agent develops clinical disease, usually after an incubation period that can range from hours to years¹. However, when the disease is caused by a fastidious pathogen the estimation of some epidemiological parameters, like the incubation period and the proportion of individuals naturally susceptible, can be a difficult task.

It is obvious that a controlled experiment involving human inoculation and lifelong follow-up is unethical. Thus, estimates of incubation period and natural resistance to human infections are often based on experiments, animal models, epidemics in isolated populations or on the observation of rare natural occurrences or accidental inoculations (for example, the Lübeck accident in 1930)².

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, an obligatory, highly specialized intracellular parasite of the skin macrophages and Schwann cells of the peripheral nervous system³. So far, it can't be cultivated in artificial media. Clinical manifestation of leprosy distributes on a spectrum displaying two polar – tuberculoid and lepromatous – and three borderline forms⁴. The disease is treated through multidrug-therapy (MDT) following the World Health Organization (WHO) protocol applied accordingly to the number of lesions and bacillary load⁵. According to the last WHO report, in 2018, the global incidence of leprosy was 208,619 cases⁶.

So far we know that humans, armadillos, red squirrels, and nonhuman primates are naturally infected with *M. leprae*⁷⁻⁹. Transmission of leprosy, though still poorly understood, is assumed to occur from a multibacillary patient to a contact through aerial nasal secretions, over a long period of exposure^{10,11}. More recently, additional sources of infection have been proposed, including soil¹², water¹³, squirrels⁷, hematophagous insects¹⁴ and infected armadillos^{9,15}. Defining exposure is challenging since serological and molecular tests underestimate the real proportion of exposed individuals mainly due to lack of sensitivity or specificity to identify low bacillary patients^{16,17}.

The incubation period of leprosy has been studied in the 30s using different study designs that generated controversial results¹⁸. Later, studies from 1967 and the 1980s estimated the incubation period of the disease from secondary cases of household contacts and from military veterans, respectively^{19,20}. Today, WHO considers the incubation period of leprosy as five years, but until now, there is no strong scientific evidence to support or refute this hypothesis⁵. Similarly, studies involving spouses or household contacts of leprosy patients have claimed to describe the proportion of individuals naturally susceptible to the disease (5%)⁵; however, these studies often suffer from severe biases or methodological limitations, such as the need for extremely long follow-up.

Here we present a population-based epidemiological study involving the entire population of the Prata Village, with the objective to explore the population's unique characteristics and obtain accurate estimations for incubation period and proportion of naturally susceptible individuals to leprosy.

The Prata Village is a former leprosy colony located on the outskirts of the Amazon forest in the Northern Brazilian state of Pará. Franciscan monks have founded the Village in 1898. In the 1920s the site became a venue for compulsory isolation of leprosy patients diagnosed across the Northern and Northeastern states of Brazil. Isolation was mandatory until 1962²¹; however, the strong social stigma associated with the disease has been preventing both emigration of affected and immigration of non-affected families to the Village. As a result, the Prata Village remains socially and geographically isolated up to date, and despite the efforts to control the disease, leprosy is still highly prevalent and homogeneously distributed across the entire community²². A genetic epidemiology study performed using the Prata families revealed the presence of a strong genetic component controlling susceptibility to leprosy in the Prata population, a finding compatible with the hypothesis that, due to the history of the Village, genetic risk factors for leprosy are enriched within this population²².

Subjects and Methods

Complete enrollment of the entire population of the Prata Village has been performed between 2006 and 2007 for a population-based Complex Segregation Analysis, and a detailed description of the enrollment strategy has been published elsewhere²². In brief, a member of the research team has visited every household at the Village and all adult members of the resident families have been interviewed. The parents have provided information about individuals younger than 18 years old. Socio-economic data has been collected through a questionnaire. All individuals self-reported their present and previous leprosy status as affected or non-affected. All self-reported leprosy cases, active and historical, have been checked using three independent sources available at the Village's health unit: (i) the original medical records, (ii) copies of the compulsory notification forms and (iii) a logbook containing registries of all cases for treatment follow-up. Self-reported cases confirmed at all three sources were tagged as a leprosy case. Upon confirmation of a leprosy case, clinical information was obtained from the medical records. All subjects enrolled in the study agreed to participate and signed a written informed consent that has been approved along with the research project by the Research Ethics Committee of the *Pontifícia Universidade Católica do Paraná* (51/07/CEP-PUCPR) and the Brazilian National Board for Ethics in Research (222/2006; 25000.001992/2006-17).

By the time of recruitment, all inhabitants, the vast majority descendants of the original isolated leprosy-affected individuals, shared the same limited environment; for example, the Village has only one church, two elementary schools, one social club, and one large central square, used by the entire population for leisure activities such as soccer games and outdoor celebrations²². In this context, it is reasonable to assume that all individuals are constantly and heavily exposed to *M. leprae*. For the analyses, we compare two sub-populations: (i) the individuals that were born at the Village and (ii) the individuals that were not-born at the Village. We understand that the sub-population that has been born at the Village; these individuals have been living their entire lives in a homogenous environment of constant, high exposure to leprosy.

We highlight that the entire population of the Village was recruited for this study, so it would be inappropriate to use inferential statistics. Thus, descriptive statistical analysis was used to calculate the proportion of affected individuals, given as a percentage of the total and of the sub-populations that has been born and not born at the Village. Median and mean age at diagnosis was estimated using the Kaplan-Meyer and box plot.

Results

Out of 2,005 individuals enrolled, 1,672 self-reported as non-affected, 319 self-reported as affected by leprosy, and 14 individuals don't inform if they are or not affected by leprosy. Data checking led to the confirmation of 257 out of the 319 self-reported cases. A prevalence rate (active and historical cases) of 1,282/10,000 was estimated for the total population (257 cases/2,005 individuals). A total of 76 individuals with unknown leprosy status (14, above mentioned + 62, unavailable medical reports) were removed from subsequent analyses (Tab. Suppl. 01). A complete demographical, socio-economic and clinical characteristics description of the population is provided as supplementary material (Tab. Suppl. 01 to 03) and in a previously published article ²².

About half of the Prata population – 1,084 individuals (56.19%) were born at the Village; among these, 64 cases of leprosy were confirmed, resulting in a cumulative leprosy prevalence of 5.9%. The same estimate reached 22.9% among the individuals not born at the Village (193 confirmed cases/844 individuals) (Figure 1 and Tab. Suppl. 02).

The parameter “age at diagnosis” also behaved differently across the two sub-populations (Tab. Suppl. 03): the median age of diagnosis among the population born at the Village was 15 years old (mean = 18.17, ranging from 5-75 years old and standard deviation of ± 12.95) in contrast with 28 years old observed for the not born at the Village (mean = 30.19, ranging from 2-74 years old and standard deviation of ± 15.25) (Figure 2). Due to the high prevalence of leprosy, isolation and few areas of social life in the Village is it's reasonable to assume that all individuals born at the Prata Village are exposed to the disease since their birth. Therefore, the median age of diagnosis among this sub-population (individual born at Village) is likely to be, to our knowledge, the best estimate of the disease's incubation period.

Also, our finding that 5.9% of these Prata-born individuals, in fact, became affected provides a rare estimate of the proportion of naturally susceptible individuals to leprosy. It is interesting to note that in the context of Prata Village's homogeneous social factors, there was practically no difference in infection in men (51.5%) and women (48.5%) born at Village and men (48.9%) and women (51.1%) not-born at Village.

Discussion

One interesting outcome of this study is the finding that although several leprosy reviews and studies present incubation period and natural susceptibility data (often 5 years and 5%, respectively), authors commonly cite textbooks or reviews and rarely, the original studies. The incubation period is often based on a book chapter²³ describing it as being between 2-5 years; nevertheless, careful analysis of the references used in the book reveals that the information is based on studies of extremely different experimental designs²⁴ resulting in findings ranging from 2-30 years.

An early attempt (1936) to estimate the incubation period of leprosy reached 8.4 years, though little information about how the study was conducted is reported and period calculations were made only based on patient's self-reports¹⁸. Data from secondary cases of leprosy household contacts reveal a mean time of leprosy onset of 4.3 (± 3.3) years independently of clinical form¹⁹; the authors considered incubation as the time between the appearance of an index leprosy patient in the household and the appearance of a leprosy secondary case in a household contact. Nevertheless, the study was performed in a hyperendemic region and the incubation period is probably reduced by the assumption that exposure occurs only through the index case living in the same household. A study on 35 military veterans revealed a median time of onset of 2-5 years for tuberculoid and 8-12 for lepromatous leprosy; however, the finding lacks statistical significance²⁰. Lastly, the incubation period estimative in animal models is based on chimpanzees, a non-consensus model for leprosy studies²⁵. Similar, previous attempts to define the proportion of individuals naturally susceptible to leprosy also led to inconsistent results, ranging from less than 1% to 12%^{26,27}.

We are aware of the limitations of our study. For our estimates to be exact, one has to assume equal and permanent exposure of all newborn individuals to *M. leprae* throughout their entire lives, which is obviously impossible to prove. For example, our strategy does not take into account the exact nature of each leprosy case; therefore, well-known differential risk factors for infection have not been considered, such as the presence or not of an index case within the household of a new case²⁸, as well as the clinical form of disease of the index case. However, stratification of all Prata patients by its clinical form revealed that 104 (40.5%) cases were lepromatous, 53 (20.6%) tuberculoid, 53 (20.6%) borderline and 47 (18.3%) of the indeterminate clinical form (Suppl. table 3). Therefore, a high proportion of the Prata patients are of higher bacillary load; considering that the observed 12.82% cumulative prevalence of leprosy in the Village is, to our knowledge, the highest ever reported for this infectious disease, these observations favor our assumption that the Prata population is under high exposure. Thus, despite the limitations, we believe that the unique historical, geographical, socio-economic and epidemiological characteristics of the Prata Village conspire to create a setting that is the closest possible to an ideal, experimental-like situation; our estimates, if not exact, are likely to be the best possible approximation for a human population.

This high prevalence of leprosy in the Village is partly a historical consequence of an artificial selection of sick individuals forced to live in there and their descendants. So, it seems reasonable to assume that the Prata population is particularly susceptible to leprosy, due to an enrichment of genetic risk factors²². This

is compatible with the young age of diagnosis among the sub-population of the Prata-born cases: it has been demonstrated that strong genetic effects controlling susceptibility to leprosy are more clearly detected among cases with early age of onset ^{29,30}. If this increased genetic effect is true, the proportion of naturally susceptible individuals in the Prata is likely to be overestimated. On the same note, one could speculate that the increased natural predisposition would result in a shorter incubation period of disease, a lower rate of naturally susceptible people and higher chances of reinfection and reactivation of the disease; intriguingly, though, the incubation period of leprosy in the Prata seems to be much higher than what has been usually reported ^{19,20,24}. Therefore, it is possible that, in the general population, the leprosy incubation period is even longer than 15 years. This may have an impact on public health, leprosy control programs, since such a large incubation period may reflect on a much longer persistence of incidence rates, as indeed observed for leprosy worldwide ³¹.

To conclude, here we use leprosy as a model to estimate the proportion of naturally susceptible individuals, as well as the incubation period of the disease. The finding may have extended impact, beyond leprosy, and may be of value to the advance on the understanding of the epidemiology of human infectious diseases in general.

Contributors:

CCOM, FPL, and MBX: collected the data.

MO: performed the statistical analysis.

HRSD, PVUS, RSAR, and GC: did the statistical analysis and wrote the article.

ES, AA, and MTM: designed the study, provided senior supervision, reviewed and wrote the article.

Potential Conflict of Interests: none reported.

Funding: (i) Unicef/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), (ii) MCT/CNPq/MS-SCTIE-DECIT 35/2005 – “Estudos da Hanseníase” and (iii) MCT/CNPq/MS-SCTIE-DECIT 25/2006 – “Estudo de Doenças Negligenciadas”.

Acknowledgments

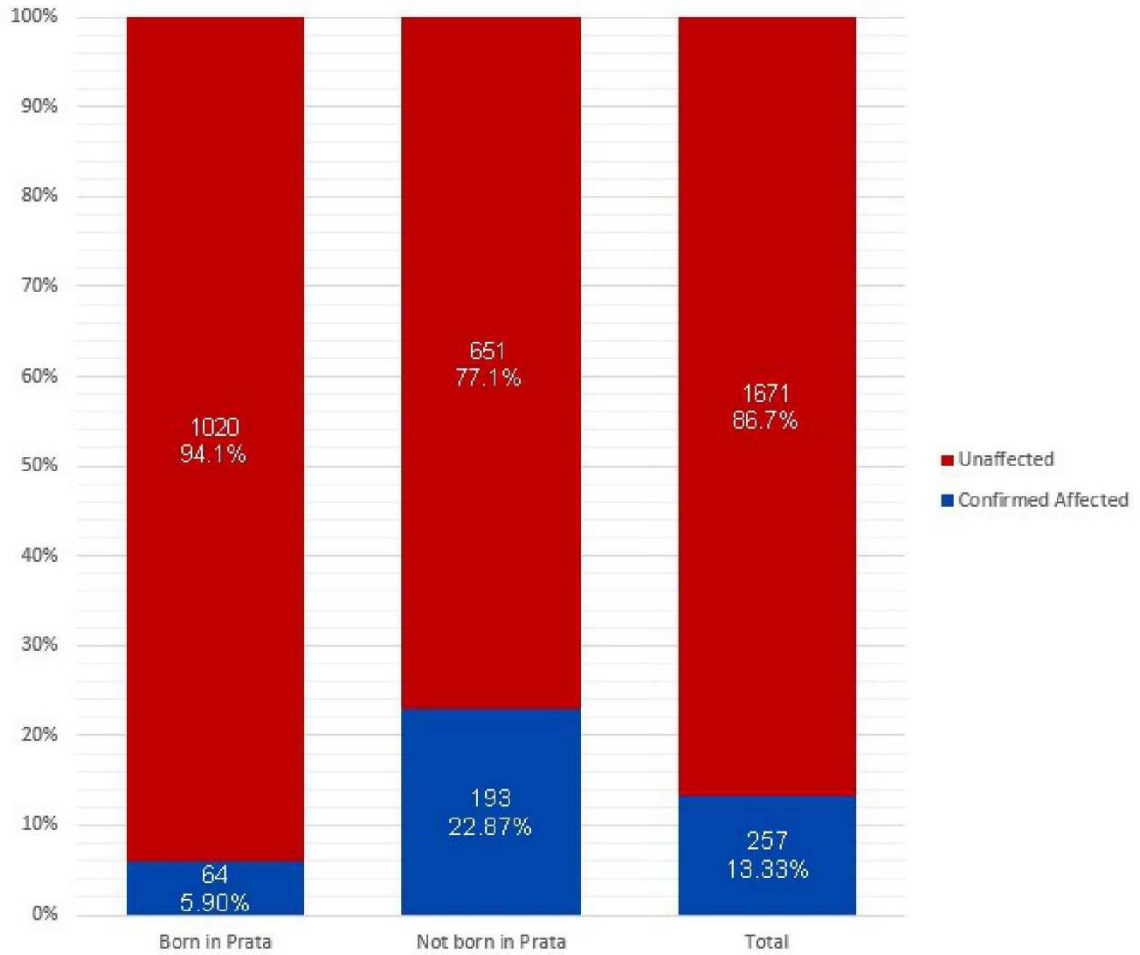
We thank all Prata Village populations for acceptance in participating in the study. This study was supported by grants of Unicef/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), MCT/CNPq/MS-SCTIE-DECIT 35/2005 – “Estudos da Hanseníase” and MCT/CNPq/MS-SCTIE-DECIT 25/2006 – “Estudo de Doenças Negligenciadas”.

References

1. National Institutes of Health (US). NIH Curriculum Supplement Series: Understanding Emerging and Re-emerging Infectious Diseases. In:2007.
2. Fox GJ, Orlova M, Schurr E. Tuberculosis in Newborns: The Lessons of the "Lubeck Disaster" (1929-1933). *PLoS pathogens*. 2016;12(1):e1005271.
3. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, Williams DL. The continuing challenges of leprosy. *Clinical microbiology reviews*. 2006;19(2):338-381.
4. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1966;34(3):255-273.
5. World Health Organization. Leprosy. <https://www.who.int/en/news-room/fact-sheets/detail/leprosy>. Published 2019. Updated 10 September 2019. Accessed 2019.
6. World Health Organization. *Global leprosy update, 2018: moving towards a leprosy-free world*. 2019.
7. Avanzi C, Del-Pozo J, Benjak A, et al. Red squirrels in the British Isles are infected with leprosy bacilli. *Science*. 2016;354(6313):744-747.
8. Honap TP, Pfister LA, Housman G, et al. Mycobacterium leprae genomes from naturally infected nonhuman primates. *PLoS neglected tropical diseases*. 2018;12(1):e0006190.
9. Truman R. Leprosy in wild armadillos. *Leprosy review*. 2005;76(3):198-208.
10. Torres P, Camarena JJ, Gomez JR, et al. Comparison of PCR mediated amplification of DNA and the classical methods for detection of Mycobacterium leprae in different types of clinical samples in leprosy patients and contacts. *Leprosy review*. 2003;74(1):18-30.
11. Davey TF, Rees RJ. The nasal discharge in leprosy: clinical and bacteriological aspects. *Leprosy review*. 1974;45(2):121-134.
12. Lavania M, Katoch K, Katoch VM, et al. Detection of viable Mycobacterium leprae in soil samples: insights into possible sources of transmission of leprosy. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2008;8(5):627-631.
13. Matsuoka M, Izumi S, Budiawan T, Nakata N, Saeki K. Mycobacterium leprae DNA in daily using water as a possible source of leprosy infection. *Indian journal of leprosy*. 1999;71(1):61-67.
14. Neumann Ada S, Dias Fde A, Ferreira Jda S, et al. Experimental Infection of Rhodnius prolixus (Hemiptera, Triatominae) with Mycobacterium leprae Indicates Potential for Leprosy Transmission. *PloS one*. 2016;11(5):e0156037.
15. Sharma R, Singh P, Loughry WJ, et al. Zoonotic Leprosy in the Southeastern United States. *Emerging infectious diseases*. 2015;21(12):2127-2134.
16. Penna ML, Penna GO, Iglesias PC, Natal S, Rodrigues LC. Anti-PGL-1 Positivity as a Risk Marker for the Development of Leprosy among Contacts of Leprosy Cases: Systematic Review and Meta-analysis. *PLoS neglected tropical diseases*. 2016;10(5):e0004703.
17. Martinez AN, Talhari C, Moraes MO, Talhari S. PCR-based techniques for leprosy diagnosis: from the laboratory to the clinic. *PLoS neglected tropical diseases*. 2014;8(4):e2655.
18. Bechelli LM. O tempo de incubação da lepra. *Rev Bras Leprol*. 1936;4:355-360.

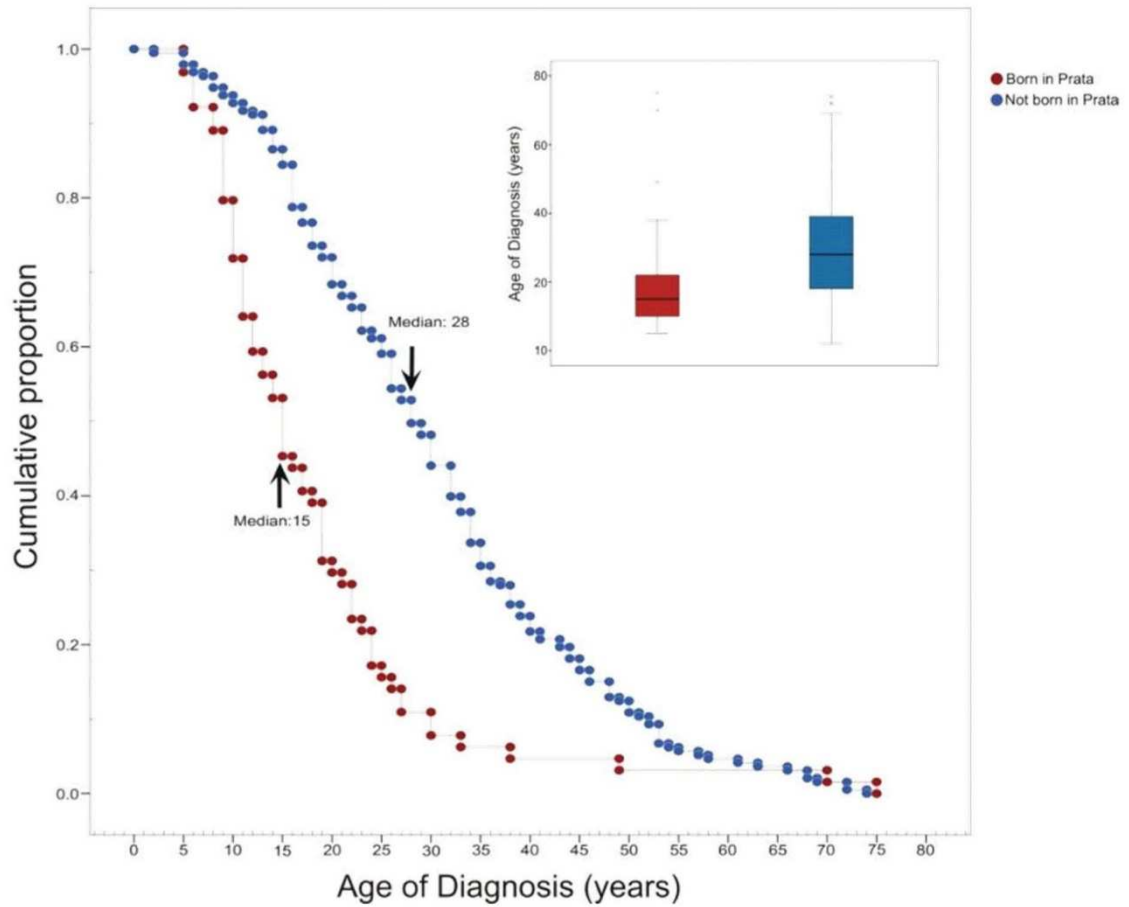
19. Prasad KV, Ali PM. Incubation period of leprosy. *The Indian journal of medical research*. 1967;55(1):29-42.
20. Fine PE. Leprosy: the epidemiology of a slow bacterium. *Epidemiologic reviews*. 1982;4:161-188.
21. Rosa Maciel L, Oliveira ML, Gallo ME, Damasco MS. Memories and history of Hansen's disease in Brazil told by witnesses (1960-2000). *Historia, ciencias, saude--Manguinhos*. 2003;10(Suppl 1):308-335.
22. Lazaro FP, Werneck RI, Mackert CC, et al. A major gene controls leprosy susceptibility in a hyperendemic isolated population from north of Brazil. *The Journal of infectious diseases*. 2010;201(10):1598-1605.
23. Noordeen SK. Elimination of leprosy as a public health problem. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 1994;62(2):278-283.
24. Bechelli LM. O tempo de incubação da lepra. *Rev Bras Leprol* 1936;4:355-360.
25. Suzuki K, Tanigawa K, Kawashima A, Miyamura T, Ishii N. Chimpanzees used for medical research shed light on the pathoetiology of leprosy. *Future microbiology*. 2011;6:1151-1157.
26. Joyce MP. Historic aspects of human susceptibility to leprosy and the risk of conjugal transmission. *Memorias do Instituto Oswaldo Cruz*. 2012;107 Suppl 1:17-21.
27. Sales AM, Ponce de Leon A, Duppre NC, et al. Leprosy among patient contacts: a multilevel study of risk factors. *PLoS neglected tropical diseases*. 2011;5(3):e1013.
28. Fine PE, Sterne JA, Ponnighaus JM, et al. Household and dwelling contact as risk factors for leprosy in northern Malawi. *American journal of epidemiology*. 1997;146(1):91-102.
29. Alter A, Fava VM, Huong NT, et al. Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Human genetics*. 2013;132(1):107-116.
30. Alcais A, Alter A, Antoni G, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nature genetics*. 2007;39(4):517-522.
31. World Health Organization. Global Leprosy Update, 2015: Time For Action, Accountability and Inclusion. *Weekly Epidemiological Record*. 2016:405-420.

Figure 1: Proportion of affected individuals in the total population and the sub-populations of individuals born and not born at the Prata Village.



Numbers inside the bars: total number of individuals (proportion); Total: the entire population of the Prata Village, as previously published on ²²; Born in Prata: sub-population of individuals that has been born at the Village; Not born in Prata: sub-population of individuals that has not been born at the Village.

Figure 2: Kaplan-Meier plot for age at diagnosis among affected individuals of the Prata Village, according to the place of birth (born at the Village; not born at the Village).



Not born in Prata (blue): sub-population of individuals that have not been born at the Prata Village; Born in Prata (red): sub-population of individuals that have been born at the Prata Village. Kaplan-Meier plot : Y-axis: the cumulative proportion of non-affected individuals; Median: median age at diagnosis. Box plot: the back bands inside the boxes are the median; the bottom and top of the boxes represent the first and third quartiles respectively; the ends of the whiskers represent the 10th and the 90th percentile.

TabSuppl01 – Epidemiological characterization of the confirmed affected and unaffected subjects in Prata Village

	Confirmed Affected Subjects		Unaffected Subjects		Total Subjects	
	n	%	n	%	n	%
Total Sample Size (n)	257		1672		1929	
Age (years)						
n ^a	257		1668		1925	
Mean	47.88		20.37		24.04	
Median	49.00		16		19.00	
Minimum	10.00		0		0.00	
Maximum	89.00		82		89.00	
Standard deviation	19.99		15.99		19.03	
Sex						
Male	153	59.5%	819	49.0%	972	50.4%
Female	104	40.5%	853	51.0%	957	49.6%
Ethnicity						
White	27	10.5%	171	10.2%	198	10.3%
Black	50	19.5%	195	11.7%	245	12.7%
Mixed Race	180	70.0%	1301	77.8%	1481	76.8%
Others	0	0.0%	4	0.2%	4	0.2%
Unknown	0	0.0%	1	0.1%	1	0.1%
Marital Status						
Single	84	48.6%	1147	68.6%	1231	63.8%
Stable Union / Married	139	80.3%	474	28.3%	613	31.8%
Divorced	2	1.2%	9	0.5%	11	0.6%
Widowed	31	17.9%	38	2.3%	69	3.6%
Others	1	0.6%	3	0.2%	4	0.2%
Unknown	0	0.0%	1	0.1%	1	0.1%
Birthplace						
Prata	64	24.9%	1020	61.0%	1084	56.2%
Others	193	75.1%	651	38.9%	844	43.8%
Unknown	0	0.0%	1	0.1%	1	0.1%
Education						
Illiterate	66	25.7%	86	5.1%	152	7.9%
Incomplete Elementary School	92	35.8%	590	35.3%	682	35.4%
Complete Elementary School	19	7.4%	71	4.2%	90	4.7%
Incomplete Middle School	39	15.2%	312	18.7%	351	18.2%
Complete Middle School	7	2.7%	71	4.2%	78	4.0%
Incomplete High School	13	5.1%	110	6.6%	123	6.4%
Complete High School	11	4.3%	111	6.6%	122	6.3%
More than High School	3	1.2%	276	16.5%	279	14.5%
Unknown	7	2.7%	45	2.7%	52	2.7%
Alcohol drink						
Yes	89	34.6%	337	20.2%	426	22.1%
No	168	65.4%	1335	79.8%	1503	77.9%
Smoking					0	
Yes	92	35.8%	292	17.5%	384	19.9%
No	165	64.2%	1380	82.5%	1545	80.1%
Drug User					0	
Yes	5	1.9%	25	1.5%	30	1.6%
No	249	96.9%	1635	97.8%	1884	97.7%
Unknown	3	1.2%	12	0.7%	15	0.8%

Abbreviations: Number: n. ^a) individuals who have the information available.

TabSuppl02 – Epidemiological characterization by the birthplace of the subjects in the study

	Born in Prata		Not born in Prata	
	n	%	n	%
Total Sample Size (n)	1084		844	
Status				
Confirmed Affected Subjects	64	5.9%	193	22.9%
Unaffected Subjects	1020	94.1%	651	77.1%
Age (years)				
n ^a	1084		840	
Mean	13.52		37.62	
Median	12		36	
Maximum	0		2	
Minimum	80		89	
Standard deviation	10.48		19.00	
Sex				
Male	558	51.5%	413	48.9%
Female	526	48.5%	431	51.1%
Ethnicity				
White	110	10.1%	88	10.4%
Black	119	11.0%	126	14.9%
Mixed Race	854	78.8%	626	74.2%
Others	0	0.0%	4	0.5%
Unknown	1	0.1%	0	0.0%
Marital Status				
Stable Union / Married	146	13.5%	466	55.2%
Single	936	86.3%	295	35.0%
Divorced	2	0.2%	9	1.1%
Widowed	0	0.0%	69	8.2%
Others	0	0.0%	4	0.5%
Unknown	0	0.0%	1	0.1%
Education				
Illiterate	14	1.3%	138	16.4%
Incomplete Elementary School	375	34.6%	307	36.4%
Complete Elementary School	38	3.5%	52	6.2%
Incomplete Middle School	213	19.6%	138	16.4%
Complete Middle School	35	3.2%	43	5.1%
Incomplete High School	65	6.0%	58	6.9%
Complete High School	59	5.4%	63	7.5%
More than High School	253	23.3%	26	3.1%
Unknown	32	3.0%	19	2.3%
Alcohol drink				
Yes	148	13.7%	278	32.9%
No	936	86.3%	566	67.1%
Smoking				
Yes	112	10.3%	272	32.2%
No	972	89.7%	572	67.8%
Recreative Drug User				
Yes	12	1.1%	18	2.1%
No	1067	98.4%	816	96.7%
Unknown	5	0.5%	10	1.2%
Treatment				
MDT	55	85.94%	133	68.91%
Unknown	9	14.06%	60	31.09%

Abbreviations: Number: n. ^a) individuals who have the information available; MDT) Multidrug therapy.

TabSuppl03 – Clinical characterization of the confirmed affected subjects in Prata Village

	Confirmed Affected Subjects
Total Sample Size (n)	257
Age of Diagnosis (years)	
n ^a	257
Mean	27.20
Median	25
Minimum	2
Maximum	75
Standard deviation	15.58
Age of Diagnosis - Born in Prata (years)	
n ^a	64
Mean	18.17
Median	15
Minimum	5
Maximum	75
Standard deviation	12.95
Age of Diagnosis - Not born in Prata (years)	
n ^a	193
Mean	30.19
Median	28
Minimum	2
Maximum	74
Standard deviation	15.25
Clinical Classification	
Tuberculoid	53 (20.6%)
Borderline	53 (20.6%)
Lepromatous	104 (40.5%)
Indeterminate	47 (18.3%)

Abbreviations: Number: n; SD: Standard Deviation; TT: Tuberculoid-Tuberculoid; LL: Lepromatous-Lepromatous; I: Indeterminate; ^a) individuals who have the information available.