



**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM DOENÇAS INFECCIOSAS**

Alterações genéticas em *M. tuberculosis* durante a fase de infecção latente e suas possíveis implicações biológicas.

Carlos André Cunha da Silva

**Vitória
2022**

Carlos André Cunha da Silva

Alterações genéticas em *M. tuberculosis* durante a fase de infecção latente e suas possíveis implicações biológicas.

Dissertação apresentada ao Programa de Pós-Graduação em Doenças Infecciosas do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo, como pré-requisito para a obtenção do título de mestre em Doenças Infecciosas.
Orientador: Prof. Dr. Moisés Palaci.

**Vitória
2022**

Dedicatória

Acima de tudo agradeço a Deus por me conceder esta oportunidade. Dedico a minha família, ao núcleo de doenças infecciosas da UFES, aos meus orientadores e aos meus colegas pela colaboração neste trabalho.

Agradecimentos

Primeiramente agradeço a minha mãe por ser um suporte durante esse período, nunca me deixou desistir e durante momentos difíceis foi uma luz no meu caminho, me ensinou a arte da fé, me deu sabedoria.

Ao meu pai que me motivou a estudar e me deu todo amparo necessário para continuar.

A minha companheira Elīna Andersone por sempre me motivar a conquistar coisas grandes e me manter na caminhada, sempre acreditando no que eu empreendia, *paldies!*

Agradeço imensamente a meu orientador Moisés Palaci, pelo amparo, paciência e zelo nas horas difíceis, principalmente por não ter desistido de mim quando não consegui prosseguir, foi mais que um orientador.

À Solange Vinhas, por ter me adotado desde o primeiro semestre da graduação e estar ao meu lado até hoje, sempre acreditou em mim e transcendeu o ensino, me orientou como uma mãe.

Agradeço a Renata Lyrio pela profunda amizade e carinho, me auxiliou e me ajudou a trilhar o caminho.

Agradeço a Lizânia Borges pela inspiração e conhecimento que me ajudou durante os estudos.

Agradeço a FAPES e ao NDI e ao PGDI pelo amparo durante os estudos, assim como o espaço físico e a oportunidade.

“To those like them; look at the story of Man, and come to your senses! It is not the destination, but the trip that matters, and what you do today influences tomorrow, not the other way around. Love Today, and seize All Tomorrows!”

All tomorrows, kösemen, 2006.

Alterações genéticas em *M. tuberculosis* durante a fase de infecção latente e suas possíveis implicações biológicas.

RESUMO

A tuberculose (TB) é a doença infectocontagiosa mais antiga da humanidade, e ainda hoje representa um grande desafio. Causada por espécies do Complexo *Mycobacterium tuberculosis* (CMTB) a doença possui ampla distribuição com predomínio nas populações mais pobres e em países de menor desenvolvimento socioeconômico. Um grande problema relacionado a TB é a infecção latente, em que a bactéria se encontra no hospedeiro em uma forma quiescente denominada “dormência”. Apesar de estudos recentes quebrarem o paradigma do modelo binário de interação de MTB com o hospedeiro, o qual era constituído pelas formas de doença (ativa) e infecção latente (LTBI), pouco se sabe sobre o estado fisiológico do MTB durante o período de latência. Neste contexto, este trabalho possui como objetivo estudar alterações genéticas nos casos secundários a partir de seus respectivos casos índices e investigar possíveis alterações e suas implicações biológicas. Foram utilizadas análises de bioinformática em 24 cepas, por meio de genômica comparativa entre casos índices e seus respectivos casos secundários utilizando-se os softwares RAST, Snippy e Spades. As diferenças genéticas que levaram a alguma modificação no caso secundário foram denominadas variantes ou variações. A análise dos dados indicou que a maioria das variações encontradas estavam nos subsistemas metabólicos relacionados a virulência e ao transporte de membrana, também constatamos que nos subsistemas relacionados a dormência e resposta ao stress não houveram variações. Não ocorrem alterações genéticas que levam a perda ou ganho de resistência de *M. tuberculosis* durante o período de dormência nos casos secundários.

Palavras chave: tuberculose, bioinformática, latência.

Genetic alterations in *M. tuberculosis* during the latent infection phase and their possible biological implications.

ABSTRACT

Tuberculosis (TB) is humanity's oldest infectious disease, and it still represents a major challenge today. Caused by species of the *Mycobacterium tuberculosis* Complex (CMTB), the disease has a wide distribution, predominantly in the poorest populations and in countries with lower socioeconomic development. A major problem related to TB is latent infection, in which the bacterium is in the host in a quiescent form denominated "dormancy". Despite recent studies breaking the paradigm of the binary model of interaction between MTB and the host, which consisted of the forms of disease (active) and latent infection (LTBI), little is known about the physiological state of MTB during the latency period. In this context, this work aims to study genetic alterations in secondary cases from their respective index cases and to investigate possible alterations and their biological implications. Bioinformatics analyzes were used in 24 strains, through comparative genomics between index cases and their respective secondary cases using RAST, Snippy and Spades software. The genetic differences that led to some modification in the secondary case were called variants or variations. Data analysis indicated that most of the variations found were in the metabolic subsystems related to virulence and membrane transport, we also found that in the subsystems related to dormancy and stress response there were no variations. There are no genetic alterations that lead to loss or gain of *M. tuberculosis* resistance during the dormancy period in secondary cases.

Keywords: tuberculosis, bioinformatics, latency infection.

Lista de tabelas

TABELA 1	Classificação funcional e temporal dos genes de <i>M. tuberculosis</i>	28
TABELA 2	Comparação do genoma de cepas de referência do Complexo <i>M. tuberculosis</i> em nível de nucleotídeo.....	30
TABELA 3	Comparação entre as análises de SNPs das cepas de MTB utilizadas por Colangeli e colaboradores e posteriormente por Cunha e colaboradores.....	51
TABELA 4	Distribuição das modificações nos subsistemas.....	52
TABELA 5	Análise de aquisição ou perda de resistência nos casos secundários.....	55
TABELA 6	Dados detalhados das modificações metabólicas encontradas....	81

Lista de figuras

FIGURA 1	Representação da parede celular do MTB.....	22
FIGURA 2	Filogeografia das linhagens do complexo <i>Mycobacterium tuberculosis</i>	25
FIGURA 3	Distribuição percentual das funções de proteínas produzidas por <i>M. tuberculosis</i>	29
FIGURA 4	Caracterização da população de estudo.....	36
FIGURA 5	Fluxo resumido de execução do trabalho.....	49
FIGURA 6	Distribuição das modificações metabólicas encontradas divididas ao longo dos principais subsistemas de ocorrência.....	54

SUMÁRIO

1. Introdução.....	14
1.1 A tuberculose.....	14
1.2 Epidemiologia.....	15
1.3 A patogenia.....	16
1.4 O diagnóstico.....	18
1.5 O tratamento.....	19
2. Revisão de literatura e justificativa do estudo.....	21
2.1 Características biológicas de <i>Mycobacterium tuberculosis</i>	21
2.1.1 Características fenotípicas.....	21
2.1.2 Sequenciamento do genoma completo.....	23
2.1.3 Linhagens.....	24
2.2 Resistência a agentes antimicrobianos.....	25
2.3 Características genéticas e suas possíveis implicações.....	26
2.4 Qualificação do problema e justificativa do estudo.....	30
3. Objetivos.....	33
3.1 Objetivo geral.....	33
3.2 Objetivos específicos.....	33
4. Materiais e métodos.....	34

4.1 Local e modelo do estudo.....	34
4.2 Caracterização da população de estudo.....	34
4.3 Primeira etapa do estudo.....	35
4.4 Segunda etapa do estudo.....	36
4.4.1 Cultura e identificação dos isolados selecionados para o estudo.....	36
4.4.2 Análise do tamanho dos fragmentos de restrição baseado na sequência de inserção IS6110.....	37
4.4.3 Análise da região <i>Direct Repeat (DR)– Spacer Oligonucleotide Typing</i> (<i>Spoligotyping</i>).....	40
4.5 Terceira etapa do estudo.....	42
4.5.1 Sequenciamento total do genoma.....	42
4.5.2 Ferramentas de bioinformática.....	43
4.5.2.1 FASTQC.....	43
4.5.2.2 Trimagem (Trimmomatic)	44
4.5.2.3 Montagem das amostras.....	44
4.5.2.4 RAST (Análise comparativa)	46
4.5.2.5 Identificação dos SNPs (SNIPPY)	47
4.6 Aspectos éticos.....	47
5. Resultados.....	49
5.1 Fluxo de execução do trabalho.....	49
5.2 Resultado das análises de qualidade.....	49

5.3 Distribuição dos SNPs ao longo do tempo.....	50
5.4 Distribuição das modificações nas vias ao longo do tempo distribuídos em subsistemas metabólicos.....	52
5.5 Análise preliminar dos dados das vias e funções afetadas pelas modificações.....	53
5.6 Análise de aquisição ou perda de resistência nos casos secundários.....	55
6. Discussão.....	56
6.1 SNPs.....	56
6.2 Taxas de mutação em <i>M. tuberculosis</i> durante o período de latência.....	57
6.3 Variações genéticas em <i>M. tuberculosis</i> durante o período de latência e suas possíveis implicações.....	59
6.3.1 Virulência.....	59
6.3.2 Transporte de membrana.....	61
6.3.3 Regulação e sinalização.....	61
6.3.4 Metabolismo.....	62
6.3.5 Sistemas baseados em cluster.....	62
6.3.6 Aminoácidos.....	63
6.3.7 Cofatores.....	63
6.3.8 Ácidos graxos e lipídios.....	64
6.3.9 Resistência a drogas.....	65
6.3.10 Comentários finais.....	66

7. Conclusões.....	67
8. Referências.....	68
APÊNDICE A.....	81
GLOSSÁRIO.....	84

1. Introdução

1.1 A tuberculose

A tuberculose (TB) é uma das doenças mais antigas da humanidade. O gênero do organismo causador, o *Mycobacterium*, pode ter milhões de anos. Alguns estudos apontam que *Mycobacterium tuberculosis* (MTB) provavelmente emergiu como um patógeno de nossos primeiros ancestrais, a cerca de 70.000 anos atrás, no leste da África (GALAGAN, 2014). Poucas doenças tiveram uma associação com seres humanos tão longa e impactante quanto a TB. Descrições sugestivas dessa doença datam de milênios e são encontradas em textos médicos chineses (~2.700 AC), em escritos védicos da Índia (~1.500 AC), no antigo testamento da Bíblia (~1.300–400 AC) e nos escritos de Hipócrates (460 AC) (GALAGAN, 2014).

Enquanto os humanos povoavam o globo, eles também levavam suas enfermidades, incluindo a TB. Pesquisadores investigando o passado da doença encontraram o material genético do MTB contendo bacilos álcool ácido resistente em múmias egípcias e peruanas, essas evidências mostraram que o bacilo da TB pode permanecer viável por muitos anos nos tecidos de pessoas saudáveis, e a infecção pode produzir a doença em um ser humano após décadas de dormência (SCHLOSSBERG, 2017).

A TB é uma doença infecto contagiosa causada por espécies de micobactérias pertencentes ao Complexo *Mycobacterium tuberculosis* (CMTB), que é composto pelas seguintes espécies: *M. tuberculosis*, *M. africanum*, *M. orygis*, *M. bovis*, *M. microti*, *M. canetti*, *M. caprae*, *M. pinnipedii*, *M. suricattae* e *M. mungi*. O principal agente etiológico da TB é o *M. tuberculosis* (VAN INGEN *et al.*, 2012).

A infecção se inicia pela inalação de aerossóis, partículas de 1-5 µm de tamanho constituídas de dois a três bacilos viáveis envolvidos por secreção pulmonar desidratada, denominadas núcleos de Wells. Somente esses núcleos produzidos pela tosse, espirro ou fala são capazes de atingir bronquíolos e alvéolos pelas vias aéreas

distais (DIETZE *et al.*, 2009). Devido ao tamanho diminuto, esses aerossóis podem permanecer em suspensão por minutos ou até horas. O risco de infecção depende de vários fatores tais como, o estreito contato entre indivíduos, a carga bacilar inalada, e o sistema imune do potencial hospedeiro (VINHAS, 2013).

Estima-se que apenas 30% das pessoas expostas ao contágio adquirem a infecção; desse total apenas 10% evoluem com alguma forma clínica da doença ao longo da vida, e os 90% restantes desenvolvem resposta imune suficiente para eliminar a infecção ou estabilizá-la (NORTH; JUNG, 2004).

1.2 Epidemiologia

A TB é uma doença de distribuição ampla, com predomínio nas populações mais pobres e nos países de menor desenvolvimento socioeconômico. A maior prevalência encontra-se na África, América Latina, Sudeste Asiático e em alguns países do antigo bloco soviético (WHO., 2020).

De acordo com a Organização Mundial da Saúde (WHO), a TB figura como a principal causa de morte entre pessoas vivendo com HIV. Segundo esta entidade, em 2019, havia um número estimado de 10 milhões de novos casos de TB em todo o mundo, e 1,2 milhão de pessoas morreram devido à doença. No mesmo ano no Brasil, a incidência de TB foi de 35 casos/100.000 habitantes, e a taxa de mortalidade relacionada foi de 2,3 óbitos/100.000 habitantes (Ministério da saúde., 2020).

A incidência da doença é maior entre os homens na faixa etária dos 15 aos 49 anos, comprometendo, assim, grande parcela da população produtiva de vários países; muitos pacientes se tornam incapazes em decorrência de sequelas graves; há uma significativa redução na expectativa de vida e aumento no número de dias perdidos de vida. Esses indicadores revelam o alto custo social e econômico da doença e o elevado grau de sofrimento humano por ela ocasionada (WHO., 2020).

Na população em geral o risco de infecção e de adoecimento é maior entre os contatos diretos de pacientes portadores de TB pulmonar, principalmente coabitantes domiciliares ou em instituições como prisões, albergues, asilos e similares. Isto se explica pelo fato de que a proximidade com o foco bacilífero possibilita maior chance de contágio com carga bacilar mais numerosa e virulenta. Os profissionais de saúde expostos ao contato com pacientes tuberculosos também constituem população de alto risco de infecção (LEWINSOHN *et al.*, 2017).

1.3 A patogenia

A resistência natural e a suscetibilidade do homem à TB são determinadas geneticamente, estando envolvidos provavelmente vários genes que definem um “espectro de suscetibilidades” ao bacilo (NORTH; JUNG, 2004).

A evolução da infecção após o contágio não tem tempo de incubação definido. Nos alvéolos pulmonares a multiplicação bacilar provoca uma resposta inicialmente mediada por elementos do sistema fagocitário da imunidade natural, principalmente os macrófagos alveolares (VYNNYCKY; FINE, 2000).

A atividade dessas células pode neutralizar os bacilos e, nesse caso, a infecção não progride para a doença. Contudo, os bacilos podem resistir à fagocitose ou à destruição no interior dos macrófagos, permanecendo em multiplicação dentro ou fora deles. A atividade metabólica bacteriana e a morte bacilar produzem substâncias com propriedades antigênicas. O processamento dessas substâncias e a sua exposição na superfície celular dos macrófagos, acopladas a moléculas do complexo de histocompatibilidade (*MHC / major histocompatibility complex*), bem como a produção, pelos macrófagos, de interleucinas 12 e 18, produzem estimulação de linfócitos T, principalmente dos tipos CD4+ e CD8+, resultando na formação de clones celulares com resposta imune específica, capazes de produzir outras citocinas, como a interleucina 2 e interferon gama, potentes ativadores da atividade macrófágica (VYNNYCKY; FINE, 2000).

A produção de linfócitos da linhagem Th1 parece ser o mecanismo fundamental na imunidade protetora contra a TB (HOWARD; ZWILLING, 1999). Há também formação de clones de linfócitos B e produção de anticorpos, mas, ao que parece, esses não desempenham papel relevante na imunidade antituberculosa (NORTH e JUNG, 2004; ELLNER, 1997; SCHLUGER; ROM., 1998).

A multiplicação bacilar no interior dos macrófagos ocasiona a destruição destas células, possivelmente pela ação de linfócitos citotóxicos (WALLIS *et al*, 2001; SCHLUGER e ROM, 1998); a morte celular resulta na liberação de enzimas líticas que produzem necrose tecidual.

A introdução de clones de linfócitos T na resposta imune ao bacilo da TB, e sua interação com os macrófagos, origina os granulomas, estruturas microscópicas caracterizadas pela presença de células gigantes multinucleadas, resultantes da fusão de macrófagos, em meio a áreas de necrose caseosa, circundadas por halos de linfócitos que parecem circunscrever a lesão, limitando, assim, a disseminação dos bacilos. Histiócitos e colágeno também estão presentes neste local, assim como graus variáveis de reação fibrogênica (NORTH; JUNG, 2004; FENTON; VERMEULEN, 1996; SCHLUGER; ROM, 1998).

Evidências experimentais sugerem que a produção de fator de necrose tumoral alfa (TNF- α) pelos macrófagos desempenha papel importante na formação dos granulomas, mas que também poderia favorecer a progressão da doença devido à indução de necrose caseosa (SCHLUGER; ROM, 1998; FENTON; VERMEULEN, 1996).

A formação de conglomerados de granulomas dá origem a estruturas nodulares, visíveis macroscopicamente, denominadas de tubérculos pelos antigos patologistas, em alusão à sua semelhança com estruturas vegetais assim denominadas. Esta característica conferiu à doença o nome de tuberculose, conforme já assinalado. A drenagem de material caseoso liquefeito através da luz brônquica origina lesões cavitárias pulmonares, em cujas paredes os bacilos se multiplicam intensamente no

meio extracelular favorecidos pelo pH alcalino e pela grande oferta de oxigênio (FENTON; VERMEULLEN, 1996).

O processo infeccioso que se instala após o contágio pelo MTB pode permanecer restrito a um segmento do pulmão, mas durante o desenvolvimento da resposta imune é comum acontecer a disseminação de bacilos por via linfática, hematogênica ou brônquica, dando origem à formação de focos secundários da infecção no próprio pulmão ou em outros órgãos, principalmente em linfonodos hilares e mediastinais. Na maioria dos casos a resposta imune consegue conter a multiplicação bacilar e os focos granulomatosos se curam, sem manifestações clínicas da doença. Assim se caracteriza a primo-infecção tuberculosa (FENTON; VERMEULEN, 1996).

Os pacientes portadores de lesões pulmonares que eliminam bacilos na expectoração (bacilíferos) são os principais transmissores da infecção, e constituem o alvo preferencial para o diagnóstico e tratamento precoce da doença, pois somente a sua cura é capaz de interromper a cadeia de transmissão da infecção tuberculosa (STYBLO, 1991).

1.4 O Diagnóstico

Por se tratar de um exame rápido e de baixo custo, a baciloscopia constitui a principal forma de diagnóstico da TB, principalmente nos países de maior incidência. Esta técnica possui sensibilidade subótima (aproximadamente 50% com limite de detecção de 10^4 bacilos por mL de amostra). Mediante a impossibilidade de adoção de métodos mais sensíveis de diagnósticos em alguns centros, existem alternativas que aumentam a sensibilidade da técnica, como o uso de fluorescência e de microscópios com lâmpadas de LED (WHITELAW *et al.*, 2011).

Devido à baixa sensibilidade do método de baciloscopia, os laboratórios dispõem de uma alternativa de alto valor diagnóstico, a cultura. Esta técnica pode ser realizada por meio de inoculação direta ou após processamento do escarro em meios de cultura sólidos a base de ágar ou ovo ou em meios de cultura líquidos (FURIN; JOHNSON., 2005). A cultura em meio líquido no sistema MGIT (que apresenta limite de detecção

de aproximadamente 10 bacilos por mL) é considerada a metodologia padrão-ouro para confirmação diagnóstica da TB. É uma técnica bastante sensível, porém possui alto custo e é mais propensa a contaminação do que métodos que utilizam meio de cultura sólido (CHIEN *et al.*, 2000).

Outros testes confirmatórios utilizados para diagnóstico utilizam a amplificação de ácidos nucleicos, sendo o GeneXpert MTB/RIF® um dos mais utilizados por apresentar boa sensibilidade em amostras com baciloscopia positiva (89%), necessitar de pouco treinamento para ser implantado e ser custo-efetivo em áreas endêmicas (LANGLEY *et al.*, 2014).

Em 2017, a Organização Mundial da Saúde endossou a utilização de uma nova versão do GeneXpert MTB/RIF®, denominado ULTRA, que apresentou aumento de sensibilidade de aproximadamente 17% nos casos de baciloscopia negativa e 14% em pacientes HIV positivos (DORMAN *et al.*, 2018).

1.5 O tratamento

O tratamento atual da TB é constituído de duas fases, sendo a fase intensiva, realizada com doses diárias de rifampicina (RIF), isoniazida (INH), pirazinamida (PZA) e Etambutol (ETM) por 4 meses, seguida da fase de manutenção com doses diárias de RIF e INH por mais 2 meses (WHO, 2020; MINISTÉRIO DA SAÚDE DO BRASIL, 2020).

Embora denominado de tratamento de curta-duração, o tempo é a principal desvantagem da terapia, com taxas de abandono entre 7% e 54% após a quarta semana (KRUK; SCHWALBE; AGUIAR, 2008).

Uma das estratégias utilizadas para melhorar a adesão dos pacientes ao tratamento é a Terapia Assistida Orientada (do inglês *Directly Observed Therapy* – DOT) em que o paciente recebe orientação e supervisão do tratamento por um profissional de saúde. Entretanto, o DOT possui alto custo para o sistema público de saúde e para os pacientes. Um estudo de metanálise que comparou o resultado do DOT em relação

ao tratamento de administração própria reportou que não houve diferença entre os dois em relação à cura, recidiva e emergência de resistência (FRIEDEN; SBARBARO, 2007).

Além da evasão, outra grande preocupação em relação ao tratamento é a atividade dos fármacos utilizados em bacilos com baixa taxa metabólica, que são responsáveis pela infecção latente (KONDRATIEVA *et al.*, 2014).

Dos fármacos utilizados na terapia anti-TB, somente a pirazinamida possui conhecida atividade *in vitro* sobre o MTB dormente. No entanto outros experimentos *in vitro* revelaram que a rifampicina também possui relativa atividade sobre estes microrganismos (WAYNE; SOHASKEY, 2001), o que ajuda a sustentar o tratamento recomendado pela OMS para TB latente que inclui INH por 6 a 9 meses, RIF e INH por 3 a 4 meses ou Rifapentina (RIF) e INH por três meses em países com alta taxa de incidência da doença (WHO., 2020; MINISTÉRIO DA SAÚDE DO BRASIL, 2020). Apesar de ser recomendado para o tratamento da TB latente, nenhum dos regimes citados é capaz de eliminar o MTB completamente do seu hospedeiro, sendo este capaz de permanecer indetectável em reservatórios ainda pouco conhecidos (CAÑO-MUÑIZ *et al.*, 2018).

Outro aspecto altamente relevante relacionado ao tratamento é o fenômeno de resistência bacteriana aos fármacos. Neste caso, há a necessidade **de modificações** no esquema de tratamento incluindo drogas de segunda linha (WALLIS *et al.*, 2016). No geral, a OMS passou a recomendar a utilização de drogas de forma individual observando cada caso de resistência (WHO, 2020). Enquanto os casos de monorresistência à Isoniazida (Hr-TB) são tratados substituindo a droga por Levofloxacina, em pacientes MDR/RR-TB o esquema deve ser montado de modo que o tratamento se inicie com pelo menos 4 drogas seguindo a ordem de escolha por grupo (WHO, 2020). Casos de XDR-TB (Tuberculose Extensivamente Resistente) envolvem resistência à Isoniazida e Rifampicina, como as MDR-TB e ainda a qualquer uma das fluoroquinolonas (levofloxacina ou moxifloxacina) e a pelo menos um dos três medicamentos injetáveis de segunda linha (amicacina, capreomicina ou canamicina) (SOTGIU *et al.*, 2009).

Uma das grandes dificuldades na erradicação da doença é a não adesão ao tratamento que pode levar à falha terapêutica, recidiva e os surgimento de resistência (PAI *et al.*, 2016).

2. Revisão de literatura e justificativa do estudo.

2.1 Características biológicas de *Mycobacterium tuberculosis*.

As micobactérias pertencem ao filo *Actinobacteria*, classe *Actinobacteria*, ordem *Corynebacteriales*, família *Mycobacteriaceae* e Gênero: *Mycobacterium*. O gênero *Mycobacterium* é constituído por 188 espécies e 13 subespécies, entre as quais *M. tuberculosis* (MTB), *M. leprae* e outras denominadas micobactérias não causadoras de tuberculose (MNT) (GUPTA, 2012).

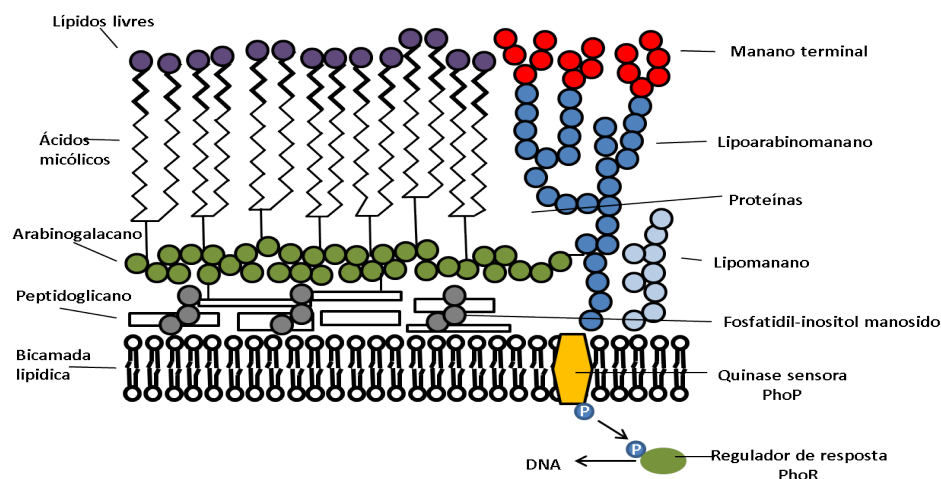
M. tuberculosis é um bacilo caracterizado, principalmente, por sua falta de motilidade, ser ácido-álcool resistente e de crescimento lento (tempo de geração em torno de 18 horas). É membro do complexo *Mycobacterium tuberculosis*, composto por *M. tuberculosis*, agente etiológico da TB humana; *M. africanum*, causador de TB em humanos somente em certas regiões da África; *M. bovis*, *M. caprae* e *M. pinnipedii*, que causam TB em mamíferos domésticos e selvagens; *M. microti*, que causa TB em roedores silvestres, *M. Canetti* que também infecta humanos (DELOGU *et al.*, 2013) e mais recentemente *M. mungi* que foi identificado em 2010 como causador de TB em mangustos na África (ALEXANDER *et al.*, 2010).

2.1.1 Características fenotípicas.

As espécies do CMTB são aeróbias estritas, intracelulares facultativas, de crescimento lento, que podem se multiplicar dentro dos macrófagos e em outras células do organismo. Tem a capacidade de permanecer vivo por anos no granuloma formado pela reação tecidual e se multiplicar quando falham as defesas do hospedeiro (NORTH, JUNG, 2004; COSCOLLA, 2013), de modo geral, as bactérias do Complexo MTB apresentam-se na forma de bastonetes curvos ou retos, com 1 – 4 µm de

comprimento por 0,3 - 0,6 μm de largura. Citológicamente, a parede celular semelhante às Gram-positivas (COSCOLLA, 2013), porém a camada de peptidoglicano é mais delgada e coberta por uma rede de ácidos micólicos e proteínas e lipídeos glicosilados (Figura 1).

FIGURA 1 - Representação da parede celular do MTB.



Fonte: (PARK E BENDELAC, 2000)

O genoma de MTB contém 250 genes envolvidos no metabolismo de ácidos graxos, sendo 39 deles responsáveis pelo metabolismo de policetídeo, precursores dos constituintes lipídicos da parede celular bacteriana. Esse grande número de genes conservados mostra a importância evolucionária do revestimento lipídico para a sobrevivência do patógeno. A abundância de ácidos graxos, cerca de 40% do total da célula, são responsáveis pela característica tintorial de resistência à descoloração por soluções álcool-ácidas apresentada pelas células bacterianas que são então designadas como bacilos álcool-ácido resistentes (BAAR) e, por esse motivo, na prática clínica e laboratorial são conhecidas e transcritas como simplesmente BAAR (GOROCICA *et al.*, 2005).

É também na parede que se encontra o composto (6-6' dimicolato de trealose) responsável pela película que o MTB forma em meio de cultivo líquido e o aspecto de corda dos bacilos em esfregaços feitos a partir de cultivos, resultante do arranjo característico dos bacilos em cadeias lineares paralelas (ou cordões de bacilos). O alto teor e diversidade de glicolipídeos na parede celular de *M. tuberculosis* tais como 6-6'dimicolato de trealose, sulfatídeos e micosídeos podem apresentar um papel

importante nos mecanismos de adesão, ingestão (fagocitose) e no destino deste microrganismo no interior das células fagocíticas. Segundo GOREN e colaboradores (1976), os sulfatídeos seriam capazes de inibir a fusão dos fagossomos com os lisossomos, já o lipoarabinomamano (ácido tuberculoesteárico), segundo outros investigadores, atuaria eficientemente (i) na quelação de radicais livres derivados da explosão respiratória do macrófago; (ii) na inibição da enzima proteína kinase C, molécula ativadora do processo oxidativo das células fagocitárias; e (iii) na inibição da transcrição dos genes que codificam o gama-interferon (CHAN; KAUFMANN, 1994).

2.1.2 Sequenciamento do genoma completo (SGC).

O sequenciamento é uma metodologia que permite detectar e analisar sequências de bases nucleicas, sua organização, disposição e sobretudo, genes e regiões não codificadoras. O sequenciamento de DNA se baseia em equipamentos que leem uma amostra de DNA e geram um arquivo eletrônico. Este arquivo contém símbolos que representam a sequência de bases nitrogenadas – A, C, G, T – presentes na amostra. As tecnologias de primeira geração, também conhecidas como sequenciamento clássico, foram as pioneiras na área. O sequenciamento de Sanger, o mais conhecido e utilizado por décadas, e o de Maxam-Gilbert são as tecnologias que constituem essa primeira geração. Apesar do sucesso no sequenciamento do genoma completo, o Sequenciamento de Sanger, é um método caro e ineficiente para projetos de grande escala. Para tarefas como essa, as técnicas de Sequenciamento de Nova Geração (do inglês, *Next-Generation Sequencing*) são mais eficientes e possuem menor custo. Essas tecnologias permitem sequenciar o DNA de forma muito mais rápida e barata em comparação com o sequenciamento de Sanger, usado anteriormente, e revolucionaram o estudo da genômica e da biologia molecular. No entanto, diferentemente do método Sanger, o método de NGS sequencia fragmentos curtos de DNA (*reads*) que costumam variar de 50 a 300 nucleotídeos de comprimento. Com o NGS, é possível sequenciar o genoma inteiro ou apenas áreas específicas de interesse.

O SGC tem sido uma ferramenta útil para descrever a epidemiologia de vários organismos patogênicos, incluindo o MTB. As potenciais aplicações do SGC visam

também as análises comparativas de genomas, gerando dessa forma conhecimentos para o surgimento de novos tratamentos, vacinas e testes diagnósticos para a TB (WEI *et al.*, 2002).

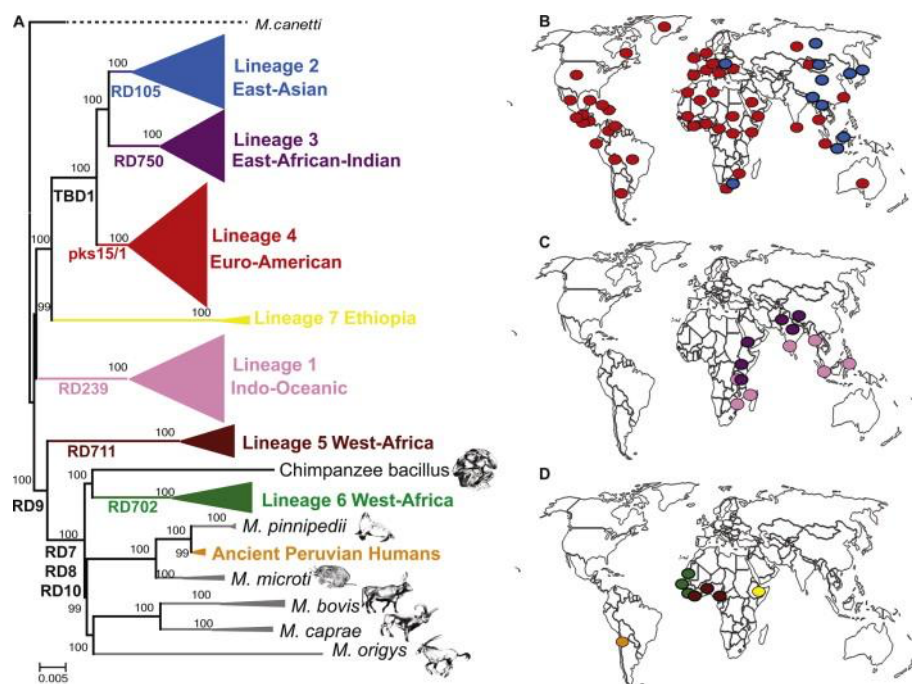
Nas últimas décadas, observamos um grande número de genes envolvidos na virulência do *M. tuberculosis*, que foram descobertos por SGC. Novos campos da ciência tais como a proteômica e a bioinformática estão permitindo estudar esses produtos gênicos e entender melhor o seu papel na patogênese da TB (FARHAT *et al.*, 2013). Por meio da identificação dos polimorfismos de base única, também chamados de “SNPs do inglês “*Single Nucleotide Polymorphism*” que estão relacionados com eventos de transmissão que anteriormente eram indiscrimináveis pelas técnicas mais antigas (WALKER *et al.*, 2013).

2.1.3 Linhagens

As técnicas de tipagem e sequenciamento gênico permitiram identificar sete linhagens do complexo *Mycobacterium tuberculosis* (CMTB) que são filogeneticamente distintas entre si e exibem diferentes distribuição geográfica e virulência para humanos. Essa diversidade pode estar ligada a características tanto do hospedeiro e do parasito, quanto aspectos sociais. Essas sete linhagens ainda podem ser divididas em dois subgrupos, em que um consiste das linhas chamadas ancestrais, L1, L5, L6 e L7 e o outro grupo de linhagens mais modernas, L2 (Beijing) amplamente distribuída pelo leste asiático sendo ainda maior no centro da Ásia, L3 e L4 (*Euro-American*).

As linhagens 1 e 3 são restritas às regiões da África ocidental, Ásia oriental, Ásia central, sul e sudeste asiático. As linhagens 5 (*M. africanum West Africa 1*) e 7 são ainda mais restritas a algumas regiões africanas. A linhagem 6 (*M. africanum West Africa 2*) e 5 aparecem em regiões da África ocidental ou podem ser encontradas em imigrantes (COSCOLLA; GAGNEUX., 2014).

FIGURA 2 - Filogeografia das linhagens do complexo *Mycobacterium tuberculosis*



Fonte: Gagneux (2014)

2.2 Resistência a agentes antimicrobianos.

A resistência a drogas antituberculose pode ser considerada, em termos conceituais, como resultado da inter-relação do fenômeno da mutação espontânea em MTB e da seleção de população predominantemente resistente, decorrente de tratamento irregular e/ou inadequado. Várias mutações gênicas pontuais que levam a modificações estruturais ou funcionais de proteínas e outros constituintes do MTB e que o torna resistente às drogas antituberculose foram identificadas ao longo das últimas décadas (COHEN; BISHAI; PYM, 2014; ISLAM *et al.*, 2017). Dentre estas mutações, destaca-se aquelas que ocorrem nos genes: (i) *katG* e *inhA* relacionados com resistência à isoniazida; (ii) *rpoB*, *rpoA* e *rpoC* associados à resistência a rifampicina; (iii) *rrs* e *rpsL* responsáveis pela resistência à estreptomicina; (iv) *pncA*, *rpsA*, *panD*, *mas* e *ppsA-E* relacionados com a resistência à pirazinamida; e (v) *embCAB* associado com a resistência ao Etambutol (COHEN; BISHAI; PYM, 2014; ISLAM *et al.*, 2017).

Embora a maioria das drogas antituberculose possua sítios alvo bem conhecidos, existem ainda cepas de MTB resistentes a uma ou mais drogas que não apresentam alterações aparentes nestes genes específicos, o que corrobora com a hipótese de que outros mecanismos não excludentes entre si, estejam envolvidos neste fenômeno (RAMASWAMY; MUSSER, 1998).

O conhecimento do perfil de sensibilidade e de resistência de isolados do MTB a drogas antituberculose permite o uso de regimes terapêuticos adequados e propicia um impacto favorável nas ações de controle da doença por evitar a propagação de cepas resistentes (PALOMINO, 2006). Os testes de detecção de resistência do MTB a antimicrobianos podem ser realizados diretamente em espécimes clínicos, ou a partir da cultura do MTB originada da amostra do paciente utilizando-se métodos fenotípicos ou genotípicos (PAI *et al.*, 2016).

2.3 Características genéticas e suas possíveis implicações.

O genoma da cepa de referência do MTB, H37Rv consiste em 4.411.529 pares de bases compreendendo 4008 genes que codificam um total de 3906 proteínas e 70 RNAs estáveis. A construção do mapa genético desta cepa revelou 91% de capacidade de codificação com cerca de 65,6% de conteúdo G + C. Das sequências gênicas codificantes, apenas 59% foram encontradas transcritas, sendo a sua maior expressão relacionada a coordenação da transcrição e da replicação o que justifica claramente sua lenta taxa de crescimento (BROSCH *et al.*, 2002; COLE *et al.*, 1998).

Por meio de comparações de banco de dados, as funções dos genes codificadores de proteínas foram classificados em 11 grupos. Aproximadamente 40% dos genes codificadores de proteínas foram atribuídos seguramente a alguma função enquanto 40% foram classificados como hipotéticos conservados. Os demais genes mostraram-se exclusivos à micobactérias em razão de não apresentarem qualquer semelhança com outras sequências gênicas microbianas (LAMICHHANE; MILIC., 2018).

O alto conteúdo G + C do genoma revelou que a composição de aminoácidos do proteoma é super-representado por aminoácidos, como Glicina, Alanina, Prolina e

Arginina, em detrimento aos nucleotídeos A + T codificadores de Lisina e Asparagina encontrados em menor quantidade. Uma análise proteômica de MTB, identificou duas grandes proteínas das famílias Prolina-Ácido Glutâmico (PE) e Prolina- Ácido prolina-glutâmico (PPE), excepcionalmente ricas em Glicina. As proteínas PPE são consideradas proteínas de armazenamento, pois contêm Asparagina como fonte de nitrogênio (LAMICHHANE; MILIC., 2018).

Pode ser observado ainda que uma parte significativa deste genoma bacteriano contém genes envolvidos no metabolismo lipídico, ou seja, potencialmente relacionados a degradação de ácidos graxos. No entanto, 7% da sequência genômica analisada mostrou-se absolutamente desconhecida. A análise proteômica também evidenciou a presença de 20 monooxigenases da superfamília de proteínas do citocromo P450. Estas proteínas estão supostamente envolvidas em processos de transformação de esteroide e metabolismo xenobiótico (LAMICHHANE; MILIC., 2018). Um resumo da análise funcional dos genes de MTB descritos acima, pode ser observado na Tabela 1 a seguir.

TABELA 1 - Classificação funcional e temporal dos genes de *M. tuberculosis*

Categorias funcionais	Número de genes		
	1998	2002	2010
0 Virulência, detoxificação, adaptação	91	99	228
1 Metabolismo de lipídio	225	233	247
2 Vias informativas	207	229	241
3 Parede e processos celulares	516	708	773
4 RNAs	50	50	71
5 Sequências de inserção e fagos	137	149	147
6 PE/PPE	167	170	168
7 Respiração e metabolismos intermediários	877	894	923
8 Regulação de proteínas	188	189	195
9 Proteínas hipoteticamente conservadas	910	1051	1081
10 Desconhecidas	606	272	16

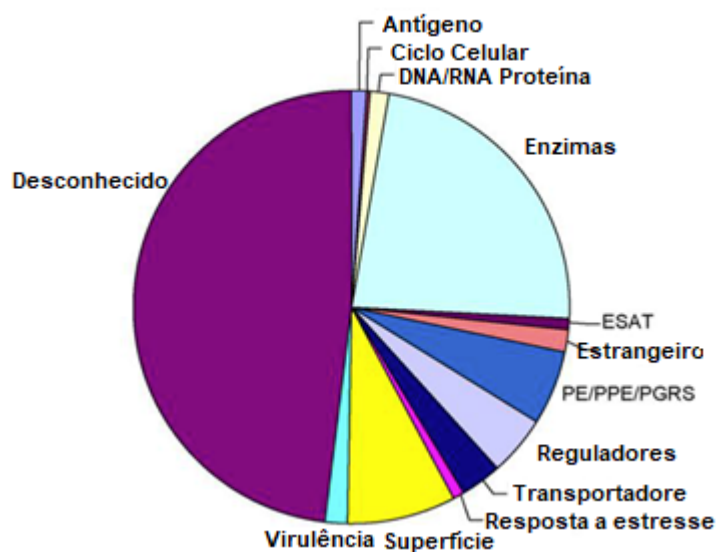
Fonte (Traduzido e adaptado): Lamichhane e Milic, 2018.

Os avanços tecnológicos das técnicas de espectrofotometria de massa associados a eletroforese e a separação cromatográfica permitiu nos últimos anos que várias proteínas de MTB fossem detectadas e identificadas em diferentes grupos funcionais. Um estudo recente publicado por Lamichhane e Milic (2018) mostra que 1678 proteínas foram identificadas por intermédio de técnicas proteômicas. O primeiro relatório de sequenciamento em 1998 do MTB, detalhou apenas oito proteínas da tuberculose que foram publicadas no *Protein Data Bank* (PDB). Dados mais recentes no PDB mostram que este número aumentou para 724 estruturas de proteínas, entre elas 129 proteínas que são consideradas como proteínas micobacterianas distintas (LAMICHHANE; MILIC., 2018).

As proteínas das famílias PE / PPE, de superfície, antigênicas e desconhecidas, desempenham um papel significativo na virulência ou estilo de vida intracelular do microrganismo. Algumas proteínas possuem um papel importante na prevenção da morte do hospedeiro e infecção abortiva outras atuam como transcriptase reversa, excisionase, proteínas de transporte, proteínas de regulação da transcrição, proteínas metabólicas, proteínas de sinalização e proteínas de choque térmico (LAMICHHANE e MILIC., 2018).

Na Figura 3 é possível observar a distribuição funcional dos principais grupos de proteínas do MTB.

FIGURA 3 - Distribuição percentual das funções de proteínas produzidas por *M. tuberculosis*



Fonte: Lamichhane (2018).

Ainda no que se refere análise genômica, Lamichhane e Milic, (2018) relatam um elevado grau de identidade de sequência nucleotídica entre cepas do Complexo MTB quando utilizadas técnicas baseadas em hibridização de ácidos nucleicos. A análise genômica comparativa entre três importantes cepas de referência do Complexo MTB é apresentada por estes autores na Tabela 2. A comparação entre as 2504 sequências codificantes presentes nos três genomas das cepas complexo MTB

revelou 1629 sequências de codificação idênticas entre *M. bovis* e *M. tuberculosis* H37Rv e 1.656 idênticas entre *M. tuberculosis* CDC1551 e *M. tuberculosis* H37Rv. Um total de 506 polimorfismos de nucleotídeo único (SNPs) sinônimos e 769 não sinônimos foram também observados em *M. bovis*. Esta análise aponta o motivo por trás da conservação de sequências de genes e da divergência entre os membros do complexo MTB embora pareçam semelhantes em nível de um único nucleotídeo. Ambos os grupos de proteínas PE-PGRS (total 29) e PPE (total 28) foram amplamente distribuídos entre as diferentes espécies analisadas.

TABELA 2. Comparação do genoma de cepas de referência do Complexo *M. tuberculosis* em nível de nucleotídeo.

Características	<i>M. bovis</i> AF2122/97	<i>M. tuberculosis</i> H37RV	<i>M. tuberculosis</i> CDC1551
Tamanho do genoma (Bp)	4,345,492	4,411,532	4,403,836
G + C, %	65.6	65.6	65.6
Genes codificadores de proteínas*	3,951	3,995	4,249
Comparado ao <i>M. bovis</i>			
SNPs	-	2,437	2,423
Transições	-	1,649	1,630
Transversões	-	788	793
Deleções (>= 1 bp)	-	205	221
Inserções (>= 1 bp)	-	177	245

*O aumento no número de potenciais genes codificadores de proteínas no CDC1551 é explicado devido a utilização de um outro algoritmo de predição.

2.4 Qualificação do problema e justificativa do estudo

Estudos recentes quebraram o paradigma do modelo binário de interação de MTB com o hospedeiro, o qual era constituído pelas formas de doença (ativa) e infecção latente (LTBI) (YOUNG *et al.*, 2009; BARRY *et al.*, 2009; CADENA *et al.*, 2017; RUSTAD *et al.*, 2008). Em vez disso, estes estudos propõem a existência de vários estágios distintos na resposta imune ao *M. tuberculosis* que produz um "ciclo de vida imunológico". Este modelo não só redefine a latência, mas também sugere que

diferentes tipos de tratamento preventivo poderiam ser mais apropriados para cada estágio da LTBI, assumindo-se que estes estágios de LTBI poderiam ser identificados nos pacientes de forma individual (RUSTAD *et al.*, 2008).

Compreender o estado fisiológico do MTB durante o período de latência é importante para potencialmente identificar pacientes com alto risco de reativação e para desenvolver tratamentos para infecção latente pelo MTB que sejam mais eficientes. Acessar o estado metabólico de MTB durante a LTBI e determinar se ele se replica e com que frequência constituem um dos grandes desafios relacionados ao tema infecção latente. Nos organismos vivos, as alterações genéticas geralmente ocorrem a uma taxa constante por geração. Por exemplo, espécies com tempos de geração mais rápidos tendem a ter maiores taxas de evolução molecular (THOMAS *et al.*, 2010). Essas observações tornaram possível o uso de taxas de mutações como um “relógio genético” que fornece uma estimativa do tempo de geração em uma espécie bacteriana (WELLER., 2015).

Vários estudos utilizaram relógios moleculares para estudar a evolução molecular do MTB (SALGAME *et al.*, 2015; HOUBEN, DODD., 2016; BARRY *et al.*, 2009; GUPTA *et al.*, 2012). Usando *Macaca fascicularis* como modelo para o estudo da TB latente, Ford e colaboradores (2011) mostraram que as taxas de mutação de *M. tuberculosis* isoladas de lesões latentes e infecções reativadas (cultivadas em média 41,7 semanas e 68,1 semanas após a infecção, respectivamente) foram comparáveis às de *M. tuberculosis* isoladas de doença pulmonar ativa (FORD *et al.*, 2011) e concluíram que *M. tuberculosis* em fase de latência possui tempo geração semelhante ao das bactérias em fase ativa e ainda que as taxas de mutação são muito maiores durante a latência.

Em contrapartida, Colangeli e colaboradores (2014) ao analisarem as taxas de mutação de cepas do MTB obtidas de quatro casos secundários de TB, originados a partir de um mesmo caso índice (2 casos com menos de dois anos de infecção latente e 2 com mais de 20 anos de infecção latente) observaram que as taxas de mutação do MTB foram substancialmente mais baixas durante o período de latência em comparação com as taxas obtidas do MTB em fase ativa da doença (COLANGELI *et*

al., 2014). As taxas médias de mutação a partir do sequenciamento completo do genoma destas 4 cepas assumindo tempos de geração de 20 horas foram $5,5 \times 10^{-10}$ mutações/bp/geração para as cepas isoladas dos casos de TB recente e $7,3 \times 10^{-11}$ mutações/pb/geração para as cepas isoladas dos casos de reativação endógena tardia

Recentemente Colangeli e colaboradores (2020) investigaram a taxa de mutação e o tempo de geração do MTB durante LTBI. Os autores sequenciaram e analisaram o genoma total (*Whole Genome sequencing - WGS*) de cepas do MTB isolados de 24 casos índices de TB e de 24 cepas isoladas dos respectivos contatos domiciliares diagnosticados com TB entre 0-63 meses após a exposição ao caso índice. Neste estudo pioneiro sobre taxas de mutação e geração de MTB durante a infecção latente, Colangeli e colaboradores (2020) verificaram que a taxa de novas mutações no genoma do *M. tuberculosis* diminui drasticamente após dois anos de infecção latente e que mutações indicativas de estresse oxidativo não aumentam com uma maior duração da latência. Observaram ainda que quando *M. tuberculosis* encontra-se em estado quiescente durante a latência, ocorre um menor risco de mutações pontuais relacionadas à resistência a drogas e o aumento do tempo de geração e tolerância bacteriana às drogas.

Estes estudos conduzidos por Colangeli e colaboradores (2020) trouxeram uma contribuição inegável para o avanço do conhecimento sobre o tema genética bacteriana e latência, contudo, neles são abordados apenas aspectos genéticos relacionados a taxas de mutações, não havendo um aprofundamento sobre em quais genes, em quais vias metabólicas ou sistemas ocorrem e suas possíveis implicações ou consequências biológicas para a bactéria.

3. Objetivos.

3.1 Objetivo geral.

Estudar alterações genéticas do *M. tuberculosis* nos casos secundários a partir de seus respectivos casos índices e investigar quais seriam as suas possíveis implicações biológicas.

3.2 Objetivos específicos.

1. Analisar as principais alterações genéticas em *M. tuberculosis* ocorridas durante a sua fase de latência e investigar quais seriam as possíveis implicações em sua estrutura, metabolismo, virulência e resistência a fármacos;
2. Analisar quais subsistemas ou elementos da morfofisiologia da célula bacteriana (MTB) são mais afetados durante a fase de latência;
3. Verificar se ocorrem mutações nos genes relacionados à resistência a fármacos durante a fase de latência de MTB.

4. Materiais e métodos.

4.1 Local e modelo do estudo.

O presente trabalho é um estudo do tipo descritivo e retrospectivo. Foi conduzido no Laboratório de Micobacteriologia do Núcleo de Doenças Infecciosas (NDI) da Universidade Federal do Espírito Santo (UFES) e no Laboratório de Micobacteriologia da Rutgers-New Jersey Medical School, Newark, New Jersey, Estados Unidos da América, no período de 2008 a 2015.

4.2 Caracterização da população de estudo.

Durante o período de 2008 a 2015 um total de 1577 pacientes (160 famílias) com TB foram triados no protocolo “Colaboração EUA-Brasil em Pesquisa Sobre Variação de Cepas em Tuberculose”. Desta população, 139 pacientes foram classificados como casos índices e 741 seus respectivos contatos, que foram arrolados no referido protocolo de acordo com os critérios de inclusão previamente descritos (JONES *et al.*, 2014).

As seguintes definições serão aplicadas a este estudo:

CASO ÍNDICE (CI): Pacientes com TB pulmonar, adultos, HIV negativos com pelo menos uma amostra com baciloscopia positiva e exame de cultura de BAAR positiva, realizados em qualquer laboratório da micro rede de laboratórios de diagnóstico de TB, que incluem os municípios de Vitória, Serra, Vila Velha e Cariacica.

CASO SECUNDÁRIO (CS): Indivíduo de qualquer idade, contato domiciliar de caso índice, com TB pulmonar, com pelo menos uma amostra com baciloscopia positiva e exame de cultura de BAAR positiva; e que satisfaça a pelo menos um dos 3 critérios abaixo de contato próximo com o caso índice por pelo menos 3 meses antes do arrolamento no estudo:

- 1 - Dormir no mesmo local que o caso índice por pelo menos cinco dias por semana;
- 2 - Compartilhar pelo menos uma refeição por dia por pelo menos, 5 dias por semana;
- 3 - Assistir à TV ou semelhante junto ao caso índice pelo menos por cinco dias por semana ou nos finais de semana.

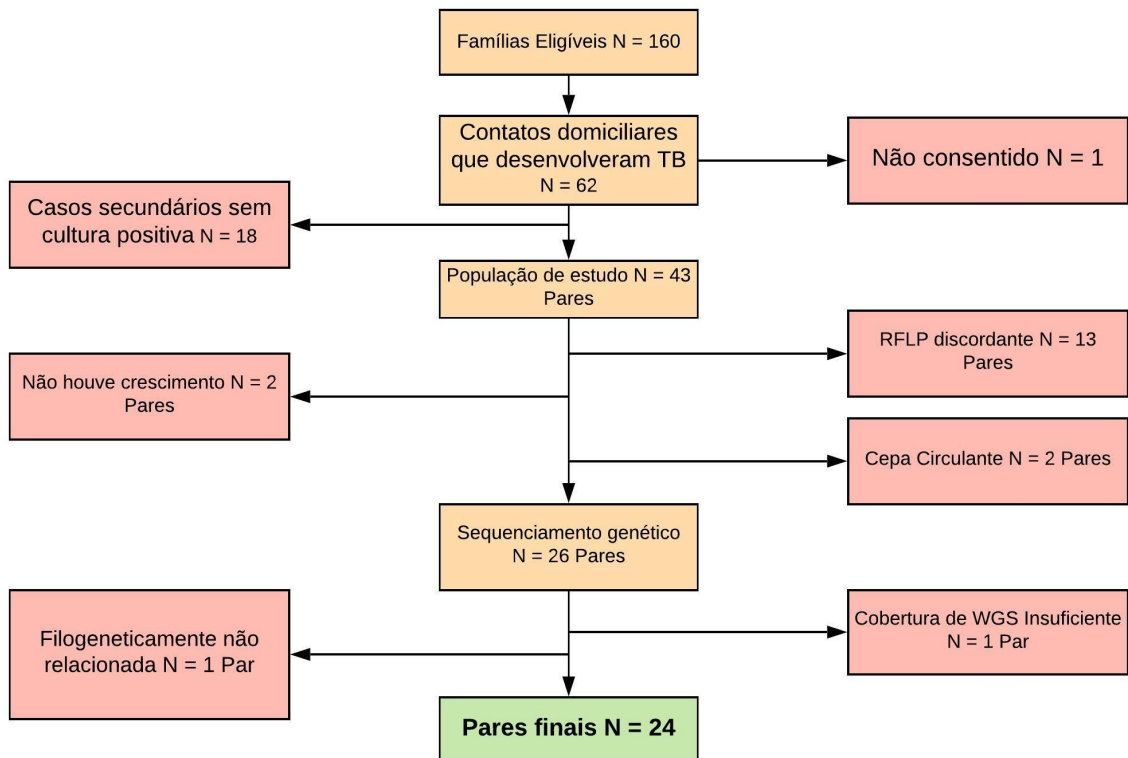
O presente estudo foi realizado em três etapas. Na primeira etapa foram selecionados os isolados de MTB dos casos índices e realizada uma busca entre seus contatos para verificar o desenvolvimento de TB ativa nesses pacientes. Foram realizados exames e culturas de escarro dos contatos, apenas aqueles com exame de cultura positivo formava par com seu caso índice.

A segunda etapa do estudo compreendeu a genotipagem molecular de todos os isolados provenientes dos CI e CS para verificar o parentesco genético entre os pares CI/CS, por meio de dois métodos moleculares, RFLP-IS6110 e *Spoligotyping*. A terceira etapa do estudo compreendeu o sequenciamento do genoma total dos isolados selecionados e o estudo de genômica comparativa.

4.3 Primeira etapa do estudo

Para o presente estudo foram selecionados 62 pares de isolados de MTB, de casos índices e seus respectivos casos secundários (figura 4).

FIGURA 4 - Caracterização da população de estudo.



4.4 Segundo etapa do estudo.

4.4.1 Cultura e identificação dos isolados selecionados para o estudo.

Os 62 pares de isolados selecionados para este estudo foram subcultivados, em meio de cultura Ogawa-Kudoh e incubados a 37°C por três semanas, quando alcançaram crescimento suficiente para extração de DNA. Todos os isolados foram identificados como MTB com base nas características fenotípicas das colônias (colônias rugosas, cremes e opacas), na sensibilidade ao ácido p-nitrobenzóico (500µg/ml) e na resistência à hidrazida do ácido tiofeno-2-carboxílico (2µg/mL) (KENT., 1985).

4.4.2 Análise do tamanho dos fragmentos de restrição baseado na sequência de inserção IS6110.

Este método foi realizado de acordo com o protocolo preconizado, com algumas modificações, como descrito a seguir (HERMANS *et al.*, 1990). Extração e Purificação do DNA: Duas alças cheias de cultura de micobactérias ressuspensas em 500µL de TE foram inativadas em banho seco a 82°C por 30 minutos e resfriadas à temperatura ambiente. A essa suspensão foram adicionados 50µL de solução de lisozima, os tubos foram invertidos 20X para homogeneizar e incubados em banho-maria a 37°C por 12 horas. Após a incubação foram adicionados 70µL de solução de SDS 10% e 6µL de proteinase K (10mg/mL), os tubos foram novamente invertidos 20X e incubados a 65°C por 10 minutos. Posteriormente, foram adicionados 100µL de NaCl 5M, os tubos invertidos 20X e adicionados 80µL de solução n-cetyl N,N,N-trimethyl ammonium bromide (CTAB).

Os tubos foram agitados em agitador tipo Vórtex até que a suspensão se tornasse leitosa quando foram então incubados a 65°C por 10 minutos. Após a incubação foram adicionados 700µL de solução clorofórmio/álcool isoamílico (24:1) e os tubos foram agitados novamente por 10 segundos e então centrifugados à temperatura ambiente por 5 minutos a 13.000xg. Os sobrenadantes foram transferidos para microtubos estéreis, foram adicionados 500µL de isopropanol e os tubos homogeneizados em agitador tipo vórtex.

A mistura foi incubada a -20°C por pelo menos 1 hora e centrifugada por 15 minutos a 13.000xg. Após a centrifugação o sobrenadante foi descartado, o sedimento lavado com 1 mL de etanol 70% gelado, centrifugado por 5 minutos a 13.000xg e o sobrenadante descartado. Alíquotas de 20µL de TE foram adicionados aos tubos e incubados a 37°C em banho-maria por 1 hora. A concentração do DNA foi obtida utilizando-se o aparelho NANODROP®. Os DNA foram diluídos para a concentração de 600ng/µL e conservados a 4°C para uso imediato ou a -20°C para uso futuro.

Digestão do DNA com Enzima Pvu II: Foram digeridos ~ 4,5µg de cada amostra de DNA a ser analisada e da cepa de referência Mt14323 com Pvu II, em um volume final

de 30µL. A mistura foi preparada adicionando-se um volume de DNA contendo a quantidade desejada, 1µL da enzima Pvu II (10U/µL), tampão de enzima (10X) e quantidade de água destilada estéril para completar 30µL. A mistura da reação foi incubada em banho-maria a 37°C por 3 horas.

Southern Blotting: A separação dos fragmentos provenientes da restrição das amostras de DNA com a enzima Pvu II foi realizada por eletroforese em gel de agarose 1,0% em tampão TBE 0,5X por 16 horas a 50 volts, e visualizado no sistema de imagem MiniBis. Em seguida o gel recebeu tratamento com 500 mL de HCL 0,25M por 10 minutos, seguido de duas lavagens com 500mL de NaOH 0,4M por 20 minutos. O gel foi colocado em um sistema *vacuum blotter apparatus* (GE Healthcare Life Sciences) para a transferência dos fragmentos de DNA para a membrana Hybond N-Plus (GE Healthcare Life Sciences). Após a transferência, a membrana contendo os fragmentos de DNA foi lavada com 500mL de solução SSC 2X (SSC 20X – NaCl 3M, citrato de sódio 0,3M pH 7).

Produção da Sonda pela PCR: Foram utilizados 20ng de DNA da cepa de referência Mt14323 (*M. tuberculosis*) e os oligonucleotídeos iniciadores INS 1 – 5' CGT GAG ATC GAG GTG GC; INS 2- 5' GCG TAG GCG TCG GTG ACA AA. O programa de amplificação consistiu de: desnaturação inicial a 95°C por 10 minutos, 30 ciclos de desnaturação a 94°C por 1 minuto, anelamento a 56°C por 2 minutos e extensão a 72°C por 1 minuto. O produto da PCR foi submetido à dosagem da concentração do DNA utilizando-se o aparelho NANODROP® e armazenado a -20°C até o uso.

Marcação da Sonda IS6110: A sonda IS6110 foi diluída com água do Kit ECL (*GE Healthcare Life Sciences*) para uma concentração de 10ng/µL, desnaturada a 100°C em banho-maria por 5 minutos e resfriada imediatamente em gelo por 5 minutos. A seguir foram adicionados volumes iguais de reagente de marcação e glutaraldeído (Kit ECL) homogeneizados cuidadosamente e, incubados a 37°C por 10 minutos. A sonda marcada foi utilizada imediatamente.

Hibridização e Detecção da IS6110: A membrana foi colocada numa garrafa de hibridização e adicionado tampão de hibridização aquecido a 42°C por 45 minutos.

Em seguida, foi adicionada a sonda à membrana e incubada overnight. Após o tempo de incubação, a membrana foi submetida a duas lavagens com 50mL de SSC/SDS a 55°C por 10 minutos, seguido de duas lavagens em 500 mL de tampão SSC 2X e uma lavagem em água destilada.

Detecção por Quimioluminescência: Finalizada a etapa de hibridização, a membrana foi tratada com a mistura de volumes iguais de reagente de detecção 1 (peroxidase) e reagente de detecção 2 (luminol - Kit ECL) e em seguida sobreposta por um Hyperfilm ECL (*GE Healthcare Life Sciences*) por um período de 30 minutos para a detecção dos perfis de RFLP- IS6110. Os filmes foram revelados com solução reveladora (Kodak®) por um minuto, lavados em água deionizada por um minuto, fixados em solução fixadora (Kodak®) por um minuto, novamente lavados em água deionizada e água corrente abundante, e secos.

Análise dos Resultados de RFLP-IS6110: Depois de secos, os filmes foram digitalizados e os perfis analisados pelo software “BioNumerics”, versão 6.5 (*Applied Maths* – Bélgica). Este programa fornece uma análise dos clusters utilizando a análise de agrupamento pelo método de médias aritméticas não ponderadas (UPGMA) baseado no coeficiente de Dice para o pareamento, com uma tolerância de 1,0% e otimização de 1,5%.

Para a análise dos resultados foram seguidas as recomendações de Tenover e colaboradores (1997). Dois ou mais isolados foram considerados pertencentes a um mesmo cluster quando seus padrões de RFLP eram indistinguíveis, isto é, apresentaram o mesmo número de bandas e as mesmas posições dessas bandas. Aqueles clusters que apresentaram diferenças de uma até três bandas foram considerados membros de uma mesma família (similaridade de 99% até 65%). Os isolados que apresentaram padrões de RFLP totalmente diferentes dos demais isolados foram considerados padrões únicos (não cluster).

4.4.3 Análise da região *direct repeat* (DR) – *Spacer Oligonucleotide Typing* (*Spoligotyping*).

Este método foi realizado de acordo com o protocolo preconizado a seguir (BRUDEY *et al.*, 2006).

Amplificação do Locus DR: Foram utilizados nas reações de PCR 20ng de DNA e as cepas *M. tuberculosis* H37Rv e *M. bovis* BCG foram utilizadas como controles. A PCR foi realizada utilizando 20pmol dos iniciadores DRa 5' CCG AGA GGG GAC GGA AAC 3' – biotilado e DRb 5' GGT TTT GGG TCT GAC GAC 3'; 0,2mM de cada dNTP, 1,5mM MgCl₂, perfazendo um volume total de 50µL. O programa de amplificação consistiu de desnaturação inicial do DNA a 96°C por 3 minutos; 35 ciclos de desnaturação a 96 °C por 1 minuto, anelamento a 55 °C por 1 minuto e extensão a 72 °C por 30 segundos e um ciclo final de 72 °C por 5 minutos, em termociclador Gene Amp PCR System 2400 (Perkin Elmer, USA).

Hibridização com a Membrana e Revelação: Trinta microlitros do produto da PCR foram diluídos em 150µL de 2x SSPE/ 0,1% SDS e desnaturados a 100°C durante 10 minutos e em seguida colocados em banho de gelo. Esta suspensão foi aplicada ao miniblatter (MN45, Isogen, Holanda), onde previamente foi montada uma membrana (Isogen, Holanda e posteriormente Ocimum Biosolutions Inc., Hyderabad India) contendo as 43 sequências correspondentes aos espaçadores conhecidos na região DR de MTB, que foi lavada com 2x SSPE/ 0,1% SDS a 60°C por 5 minutos. A hibridização da membrana foi realizada a 60°C por 60 minutos.

Em seguida a membrana foi lavada duas vezes com 2x SSPE/ 0,5% SDS a 60°C por 5 minutos cada, e incubada com um conjugado de estreptavidina-peroxidase, a 42°C por 60 minutos. A membrana foi então lavada duas vezes com 2x SSPE/ 0,5% SDS a 42 °C por 10 minutos, seguidos de 2 lavagens com 2x SSPE à temperatura ambiente. Em seguida foi tratada com a mistura de volumes iguais de reagente de detecção 1 e reagente de detecção 2 (IMMOBILON™ WESTERN - Chemiluminiscent HRP Substrate - Millipore Corporation, Billerica, MA, USA) e em seguida, sobreposta por um filme por um período de 15 minutos.

Reutilização da membrana: As membranas de Spoligotyping podem ser utilizadas por mais de uma vez. A quantidade de vezes que pode ser utilizada depende da qualidade da membrana adquirida e da qualidade do tratamento da membrana (*Stripping*) após a hibridização e detecção, em alguns casos podem ser reutilizadas por mais nove vezes. Para a reutilização numa outra hibridização, os produtos de PCR são dissociados da membrana. Para isso a membrana é lavada duas vezes em SDS 1% a 80°C por 30 minutos a cada lavagem. Em seguida a membrana é lavada em solução de EDTA 20 mM pH 8, por 15 minutos a temperatura ambiente. Transcorrido esse tempo a membrana é armazenada em solução de EDTA 20 mM pH 8 a 4°C até o uso.

Interpretação e Tabulação dos Resultados: Após revelação, os sinais obtidos com os 43 espaçadores para cada isolado foram analisados, digitalizados, transformados para formato binário (0 é falta de sinal e 1 é presença de sinal) em tabela do Microsoft Excel e comparados aos padrões no banco de dados SITVITWEB (<http://www.pasteur-guadeloupe.fr:8081/SITVITDemo/>) para identificar o Shared International Type (SIT) (DEMAY *et al.*, 2012). A partir do SIT as famílias foram identificadas de acordo com o banco internacional de espoligotipos SpoIDB4 (BRUDEY *et al.*, 2006). Os isolados que não obtiveram correspondência a um SIT no banco SpoIDB4 foram denominados padrão novo e a probabilidade de pertencerem a uma subfamília foi calculada pelo Spotclust (http://tbinsight.cs.rpi.edu/run_spotclust.html). Este software analisa os perfis com base numa mistura de modelos como o SpoIDB3 e o *Random Initialized Model* - RIM (VITOL *et al.*, 2006). Como norma do laboratório, todo perfil novo era confirmado repetindo-se a PCR e a Hibridização.

4.5 Terceira etapa do estudo.

4.5.1 - Sequenciamento total do genoma.

Esta metodologia foi realizada em colaboração com o Laboratório de Micobacteriologia da Rutgers-New Jersey Medical School, em Newark, New Jersey. Para isto os isolados selecionados foram congelados e enviados para o referido laboratório.

As plataformas de NGS (*Next Generation Sequencing*), fornecem taxas de alta transferência, gerando output de centenas de gigabytes (GB) de dados em um único experimento. Na maioria das vezes o número total de *reads* (sequências curtas de leitura) é muito maior que o número de reads necessários para cobrir o comprimento da sequência de DNA.

Esse estudo utilizou a plataforma HiSeq 2500 Illumina®, a metodologia é baseada na síntese da enzima DNA polimerase e nucleotídeos terminadores marcados com fluoróforos. A inovação dessa plataforma consiste na clonagem *in vitro* dos fragmentos, processo conhecido como PCR de fase sólida.

Resumidamente, o DNA é fragmentado aleatoriamente e, após a fragmentação, dois tipos diferentes de adaptadores são ligados às extremidades dos fragmentos. Posteriormente, esses fragmentos são desnaturados, distribuídos e imobilizados (por uma extremidade) também de maneira aleatória em uma superfície sólida, baseada na tecnologia “*flow cell*” (revestida por uma camada densa de oligonucleotídeos complementares aos dois tipos ou adaptadores dos fragmentos). Os oligonucleotídeos atuam na captura da fita molde de DNA para amplificação subsequente que ocorre para um processo denominado de amplificação em ponte (RADFORD *et al.*, 2012).

As próximas etapas da PCR geram pequenas ilhas ou clusters que servem como clones para o sequenciamento seguinte. Essa plataforma utiliza o método por síntese, onde é fornecida a cada cluster a enzima polimerase e os quatro nucleotídeos

marcados com diferentes fluorescências. A terminação 3'OH é quimicamente inativada, dessa forma garante que apenas uma base seja incorporada por ciclo. Após a incorporação de cada nucleotídeo há uma geração de uma imagem que permite a identificação de cada cluster. Existe um tratamento de desbloqueio químico que remove o radical fluorescente e a incorporação do nucleotídeo seguinte é sintetizado no próximo ciclo. A sequência de cada cluster é identificada e submetida a uma filtração para eliminar os *reads* de baixa qualidade. Essa plataforma pode produzir leituras de até 150 pb e possui como principal vantagem a elevada qualidade dos dados gerados.

As bibliotecas Illumina das culturas *in vitro* da cepa laboratorial H37Rv foram preparadas utilizando kit da Illumina *Nextera Flex Library prep*. E o Kit *Illumina's Nextera DNA CD Index* (para os adaptadores), e as amostras de DNA normalizados foram sequenciados na plataforma NextSeq 550 Illumina system no The Genome Center of Rutgers-NJMS, Newark.

4.5.2 - Ferramentas de bioinformática.

4.5.2.1 – FASTQC.

Para a etapa de verificação do controle de qualidade das reads que foram geradas pelo sequenciador, utilizamos um software conhecido como FASTQC (<http://www.bioinformatics.babraham.ac.uk/>) desenvolvido com uma linguagem de programação computacional em JAVA para verificar controle de qualidade de sequências genômicas no formato FASTQ, SAM e BAM. Este software foi executado a partir dos arquivos R1 (*Foward*) e R2 (*Reverse*) provenientes do sequenciamento pela plataforma Illumina HiSeq 2500. O software utiliza uma interface gráfica onde é possível analisar visivelmente regiões do genoma sequenciado que poderiam conter erros ou apresentar uma qualidade de sequenciamento duvidosa (ANDREWS, 2010).

4.5.2.2 – Trimagem (Trimmomatic).

Uma das etapas primordiais é aprimorar a qualidade das bases de leitura (*reads*) para evitar problemas ou erros nas etapas seguintes da montagem do genoma. Sendo assim, utilizamos uma ferramenta chamada de TRIMMOMATIC, um software do tipo “script” desenvolvido em linguagem computacional Java, exclusivo para a plataforma Illumina, com o objetivo de remover os adaptadores utilizados na construção da biblioteca genômica do sequenciamento. Os adaptadores podem ocasionar problemas quando essas sequências forem anotadas (analisadas através de outros softwares), removê-los pode tornar as sequências livres de interferências metodológicas (BOLGER *et al.*, 2014).

O programa é executado em etapas (rotinas), cada uma delas com uma determinada função, a sequência e o controle de ação do software dependem de como o usuário o executa na linha de comando (*command line – Terminal*), utilizando para isso as linhas de comando através do console, o programa foi desenvolvido para sistemas com base “UNIX”, separando as etapas com um “;” e os respectivos valores usados foram:

- Remover os adaptadores (*ILLUMINACLIP:NEXTERAPE-PE.fa:2:30:10*)
- Remover bases com baixa qualidade N bases (baixa qualidade 3) (*LEADING:3*)
- Remover “*trailing*” baixa qualidade ou N bases (baixa qualidade 3) (*TRAILING:3*)
- Escanear a *read* com as 4-bases, cortando quando a qualidade média das bases for abaixo de 15 (*SLIDINGWINDOW:4:15*)
- Cortar *reads* abaixo de 36, longas bases (*MINLEN:36*)

A Illumina utiliza diferentes adaptadores e por isso o software precisa ser configurado para o adaptador utilizado na construção da Biblioteca Genômica.

4.5.2.3 - Montagem das amostras.

O sequenciamento do genoma completo compreende vários processos, podemos destacar o processo de preparação das bibliotecas de sequenciamento (preparação

das amostras que vão ser sequenciadas formando um *input*, ou melhor dizendo, o que vai ser realmente lido na análise), o sequenciamento em si, é uma etapa complexa que pode variar dependendo da tecnologia utilizada, no geral fragmenta o genoma em vários pedaços menores que chamamos de “*reads*” e a etapa de reconstrução do genoma bem como seu ordenamento na sequência original, a essa etapa chamamos de montagem.

A montagem das amostras foi feita através do método de montagem “*de novo*” que se refere a montagem de um genoma em que não se utiliza um modelo, ou um genoma de referência. Essa abordagem tenta reconstruir os genomas explorando a sobreposição de leituras e a contiguidade das *reads*, para a análise foi utilizada a ferramenta de bioinformática “Spades”. O Spades (*St. Petersburg genome assembler*) foi desenvolvido pela Universidade de St. Petersburg na Rússia e pela Universidade da Califórnia, USA. Lançado sob licença livre (<https://github.com/ablab/spades>) e desenvolvido para a plataforma Linux e MacOS.

O algoritmo é específico para a montagem de genoma bacteriano e é recomendado apenas para genomas pequenos (BANKEVICH *et al.*, 2012) o que justifica o seu uso para este estudo.

O SPAdes pode realizar a montagem dos genomas para diferentes plataformas do NGS: Ion Torrent, PacBio, Nanopore e Illumina. Este software funciona de forma similar a um outro software bem conhecido chamado de Velvet, montando os gráficos de Brujin e em seguida removendo bolhas na identificação de erros (quando há uma descontinuação na montagem dos contigs). Após essa etapa são ajustadas as k-mers e são estimadas suas distâncias, os gráficos das montagens são construídos e por último são feitas as montagens dos *Contigs* (BANKEVICH *et al.*, 2012).

De acordo com Gurevich e colaboradores o desempenho do software Spades é melhor do que outros softwares como o conhecido Velvet nos parâmetros de cobertura do genoma e na montagem de *contigs* coesos (GUREVICH *et al.*, 2013). O Spades utiliza arquivos de entrada no formato “fastq.gz” e o K-mer precisa ser especificado

pelo usuário, os arquivos são direcionados para uma pasta de saída, neste trabalho testamos os k-mers 21, 30, 25 e 70, a melhor montagem foi utilizando o k-mer 21.

O Spades gera 25 arquivos de saída, os mesmos são organizados em pastas contendo os k-mers “setados” pelo usuário. O arquivo mais importante é “final.lib_data” contém todas as informações dos resultados da corrida: tamanho das sequências, cobertura e número de nucleotídeos.

4.5.2.4 - RAST (Análise comparativa).

Após a montagem dos genomas das cepas, utilizamos um ecossistema de softwares para fazer a análise comparativa dos mesmos, um dos principais softwares que compõem esse ecossistema é chamado de RAST. O RAST (*Rapid Annotation Using Subsystem Technology*) é um Sistema web (<http://rast.nmpdr.org>) criado para a anotação de genomas bacterianos e de Archaeae, possui algumas ferramentas que trazem informações que podem ser usadas no estudo comparativa do genoma, informações da via metabólica e alguns dados estatísticos de determinados sistemas gênicos. Este software permite uma análise com uma abordagem comparativa na própria plataforma web (AZIZ *et al.*, 2008).

O RAST recebe como entrada um “*upload*” das sequências no formato “.fasta” no site (<http://rast.nmpdr.org>) ou em linha de comando através da instalação de uma pequena ferramenta chamada de “myRAST Toolkit”, o sistema processa as informações dentro de aproximadamente um dia e o resultado pode ser acessado na plataforma. São gerados três arquivos como resultado, um com a extensão “.gbk” (formato do Genbank), outro com a extensão “.gff”. O software Artemis foi utilizado para essa visualização. A geração do arquivo “.xls” para análises contém informações essenciais tais como: tipos de genes que são ordenados em grupos e funções, identificação, codificação das proteínas e suas determinadas funções contendo o código de acesso dos bancos de dados e sua posição no genoma que pode ser muito útil para a identificação dessas regiões.

O pipeline (fluxo de trabalho) utilizado neste estudo foi executado da seguinte forma: Primeiro os dados foram analisados qualitativamente para aferirmos a qualidade das amostras, depois foram submetidos na plataforma RAST onde realizamos a anotação de genomas com bancos de dados curados, executando predição gênica e identificação de proteínas, foi realizada então uma análise para a identificação de determinados grupos com o pipeline FIGfams e depois analisados com uma abordagem comparativa de forma manual.

4.5.2.5 – Identificação dos SNPs (SNIPPY).

O Snippy (<https://github.com/tseemann/snippy>) é um software sobe licença GPL desenvolvido para plataforma Linux e MacOS usado para encontrar SNPs (*single nucleotide polymorphism*, ou em português, Polimorfismo de Base Única). O programa foi utilizado nesse estudo para identificação dos SNPs, indels (deleções e inserções), usando um genoma de referência haplóide (H37Rv).

Nesse trabalho essa análise dos SNPs foi feita usando análises de CoreSNIPPY, ou seja, foram removidas as regiões de repetição PEIPPE, que não são regiões consideradas nas análises Filogenéticas do *Mycobacterium Tuberculosis*. Na figura 10 observamos a linha de comando executada neste trabalho para o SNIPPY (SEEMANN., 2014).

4.6 Aspectos éticos

Todos os procedimentos realizados para a obtenção dos isolados utilizados neste estudo foram realizados de acordo com as diretrizes e normas regulamentadoras de pesquisa envolvendo seres humanos e com aprovação no Comitê Nacional de Ética em Pesquisa sob Registro Nº 14151, Comitê de Ética em Pesquisa do Centro de Ciências da Saúde – Universidade Federal do Espírito Santo, e pelo Comitê de Ética em Pesquisa da *Boston University Medical Center and Rutgers University—New Jersey Medical School - New Jersey*.

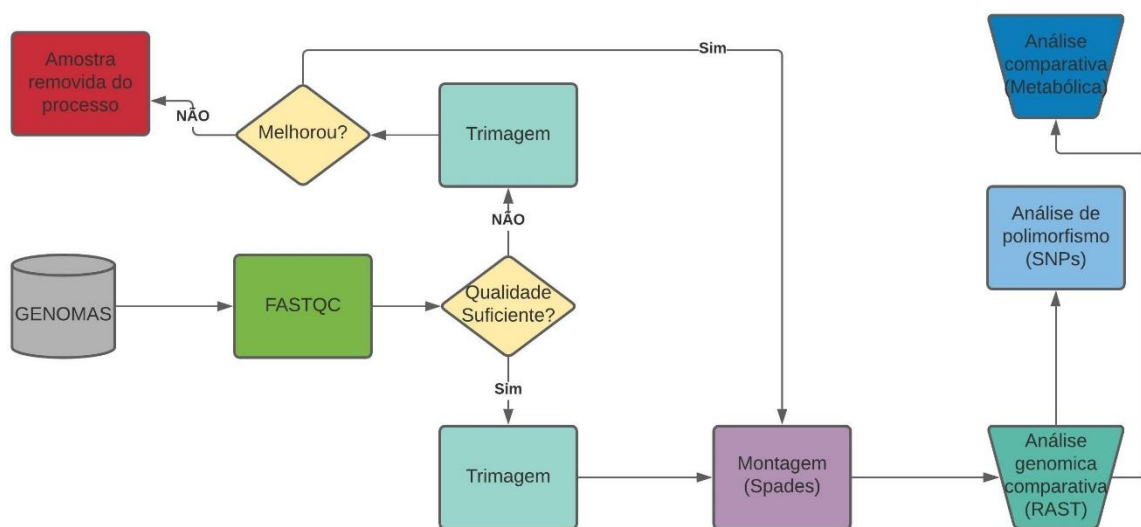
Os procedimentos de coleta de amostras e os exames bacteriológicos para o diagnóstico da TB nas Unidades de Saúde e nos Laboratórios dos Municípios participantes foram realizados de acordo com as práticas de biossegurança instituídas nesses locais. Os procedimentos bacteriológicos e de biologia molecular foram realizados de acordo com as normas de biossegurança e as Boas Práticas Laboratoriais instituídas nos locais da realização dos testes.

5. Resultados.

5.1 Fluxo de execução do trabalho.

O pipeline resumido de execução do trabalho pode ser encontrado na figura 5 abaixo. Um total de 48 amostras (24 pares) de dados genômicos foram analisados e organizados em pares.

FIGURA 5 – Fluxo resumido de execução do trabalho.



Durante as análises nenhuma amostra apresentou problemas durante a execução do pipeline, o tamanho dos arquivos somados foi de ~200Gb e todos foram devidamente checados através de criptografia de duas chaves (conferir glossário).

O processo foi realizado para certificar a integridade dos arquivos após *download* da base de dados. O resultado final mostrou que durante o processo de transporte dos dados nenhuma informação foi perdida/corrompida.

5.2 Resultado das análises de qualidade.

Todas as amostras foram submetidas a uma análise de qualidade para apuração da suficiência da cobertura dos “reads” e para identificar possível contaminação no sequenciamento. A cobertura no sequenciamento refere-se à quantidade de vezes em que uma determinada sequência foi lida pela máquina de leitura do DNA. Os dados foram testados com o software FASTQC e todas as amostras apresentaram cobertura suficiente.

A cobertura mínima no caso índice foi de 378X (amostra VV3015) e a maior cobertura foi de 606X (Amostra CA3010). Nos casos secundários a menor cobertura foi de 356X que corresponde a amostra VT3059, enquanto a maior cobertura foi de 603X (Amostra CA3038).

O tamanho do genoma teve uma média de 4.6 Milhões de pares de bases, correspondendo ao tamanho padrão do genoma do H37RV que é utilizado como uma cepa padrão (COLE *et al.*, 1998).

5.3 Distribuição dos SNPs ao longo do tempo.

Com o intuito de verificar e ratificar a quantidade de SNPs das cepas utilizadas neste estudo, comparamos os nossos resultados com aqueles obtidos por Colangeli e colaboradores (2020). Os SNPs foram quantificados pela utilização de um *pipeline* feito para o software Snippy e anotados a partir dos arquivos provenientes das análises. A menor quantidade detectada de SNPs ocorreu na amostra CA3061. A quantidade de SNPs obtidos nos casos secundários a partir de seus casos índices podem ser encontrados na tabela 3 abaixo.

O menor tempo entre o caso índice e a manifestação da doença em forma ativa no caso secundário, ocorreu no par de amostra VT3010 e o maior intervalo de tempo ocorreu no par SE3013 e foi de 63 meses.

TABELA 3 - Comparação entre as análises de SNPs das cepas de MTB utilizadas por Colangeli e colaboradores e posteriormente por Cunha e colaboradores.

Amostras		Tempo entre casos em meses	Quantidade SNPs (Cunha <i>et al</i>)	*Quantidade SNPs (Colangeli <i>et al</i>)
1	CA3001	45	3	0
2	CA3010	40	5	1
3	CA3038	60	6	3
4	CA3043	10	8	0
5	CA3046	58	4	1
6	CA3050	21	4	0
7	CA3051	39	6	7
8	CA3061	24	1	0
9	CA3064	16	3	2
10	CA3077	13	4	3
11	CA3078	26	10	13
12	SE3001	1	9	4
13	SE3010	44	3	0
14	SE3013	63	8	1
15	SE3031	54	8	0
16	SE3035	38	6	8
17	SE3063	2	5	1
18	VT3010	0	2	1
19	VT3046	27	2	1
20	VT3050	10	13	1
21	VT3059	1	10	1
22	VT3067	34	4	4
23	VV3012	45	9	0
24	VV3015	11	7	3

5.4 Distribuição das modificações nas vias ao longo do tempo distribuídas em subsistemas metabólicos.

Podemos observar na tabela 4 as modificações identificadas, distribuídas dentro de subsistemas envolvidos em algumas atividades metabólicas dentro de célula. Os dados apresentados são dos CS em comparação aos CI.

TABELA 4 - Alterações gênicas e suas possíveis implicações nos subsistemas bacterianos dos casos secundários em relação aos casos índices.

Amostras	Tempo entre CI e CS em meses	Virulência e defesa celular	Parede Celular	Transporte de membrana	Regulação e Sinalização	Dormência	Resposta a stress	Metabolismo de DNA	Proteínas	Outros
VT3010	0	1	0	1	0	0	0	0	0	0
VT3059	1	0	0	0	0	0	0	0	0	0
SE3001	1	0	0	0	1	0	0	0	0	1
SE3063	2	0	0	1	0	0	0	1	0	0
VT3050	10	0	0	0	0	0	0	0	0	0
CA3043	10	1	0	0	1	0	0	0	0	1
VV3015	11	2	0	0	0	0	0	0	1	0
CA3077	13	2	0	0	1	0	0	0	0	0
CA3064	16	0	0	1	0	0	0	0	0	0
CA3050	21	1	0	0	0	0	0	0	0	0
CA3061	24	1	0	0	0	0	0	0	0	0
TOTAL		8	0	3	3	0	0	1	1	2
MÉDIA		0.8	0	0.33	0.33	0	0	0.11	0.11	0.22
CA3078	26	2	0	0	0	0	0	0	0	0
VT3046	27	0	0	0	0	0	0	0	0	0
VT3067	34	2	0	0	0	0	0	0	0	1
SE3035	38	2	0	1	0	0	0	0	0	0
CA3051	39	1	0	1	0	0	0	0	0	0
CA3010	40	1	0	1	0	0	0	0	0	2
SE3010	44	1	0	0	1	0	0	0	0	0
CA3001	45	0	0	0	1	0	0	0	0	1

SE3031	54	0	0	0	1	0	0	1	1	2
CA3046	58	0	0	0	0	0	0	0	0	0
CA3038	60	1	0	1	0	0	0	0	0	1
SE3013	63	0	0	0	0	0	0	0	0	1
VV3012	45	0	0	1	0	0	0	0	0	0
TOTAL		10	0	5	3	0	0	1	1	8
MÉDIA		0.66	0	0.33	0.13	0	0	0.01	0.01	0.6



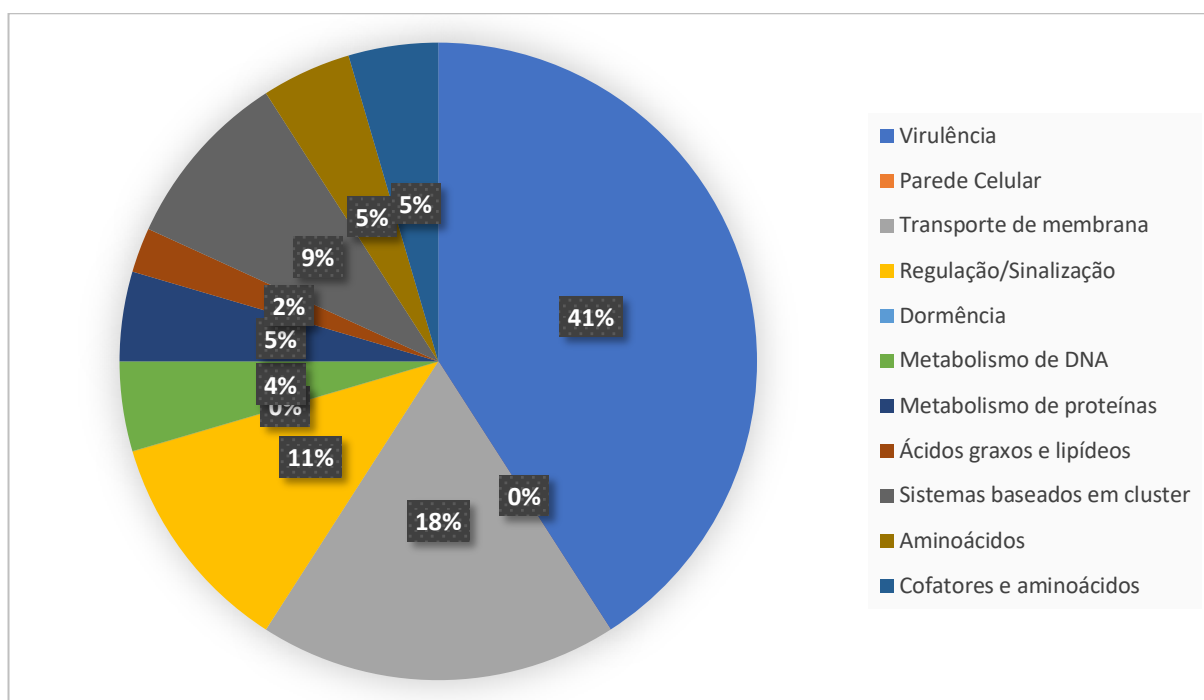
- Ganho de variação gênica apenas no caso secundário.
- Perda de variação gênica apenas no caso secundário.
- Perda e ganho de variação gênica.

Durante o estudo podemos notar de modo geral a baixa frequência na alteração nos subsistemas, no entanto, dentre as alterações detectadas foi possível registrar uma leve diminuição da média de alterações gênicas na maioria dos subsistemas bacterianos na fase tardia em relação a fase inicial de latência, destacando que os subsistemas de virulência e transporte de membrana mantiveram uma taxa de alterações semelhantes em ambas as fases, inicial e tardia e que não houveram modificações nos subsistemas de dormência e resposta ao stress.

5.5 Análise preliminar dos dados das vias e funções afetadas pelas modificações ocorridas.

A maior quantidade das modificações metabólicas ocorreu no subsistema relacionado a defesa celular e virulência, seguido do metabolismo de DNA e transporte de membrana. Não foram detectadas modificações no subsistema relacionado a dormência. Os dados encontrados podem ser observados na figura 7.

FIGURA 6 – Distribuição das modificações metabólicas encontradas divididas ao longo dos principais subsistemas de ocorrência.



Após o nome do subsistema é exibida a quantidade de modificações encontradas em relação a quantidade total, dentro de cada subsistema é mostrada a porcentagem das mesmas.

Os resultados detalhados das modificações metabólicas podem ser encontrados no apêndice A na tabela 6 (página 81).

5.6 Análise de aquisição ou perda de resistência nos casos secundários.

Os resultados da análise comparativa em regiões de resistência a um determinado fármaco mostraram que não existiu um caso em que o CS adquirisse ou perdesse a resistência. Os dados foram analisados manualmente e apenas dois pares mostraram resistência a Rifampicina e Estreptomicina. O par VT3067 foi o único que mostrou resistência a mais de um fármaco.

TABELA 5 - Análise de aquisição ou perda de resistência nos casos secundários.

Drogas	CIs/CSr	Clr/CSs	Clr/CSr	CIs/CSs	Amostras
Isoniazida	0	0	0	24	--
Etambutol	0	0	0	24	--
Rifampicina	0	0	1	23	VT3067
Bedaquilina	0	0	0	24	--
Fluoroquinolonas	0	0	0	24	--
<i>Ácido p-Aminosalicílico</i>	0	0	0	24	--
D-Cicloserina	0	0	0	24	--
Estreptomicina	0	0	2	22	VT3067, VT3059
TOTAL	0	0	3		

CIs/CSr = Caso Índice sensível e Caso Secundário resistente
 Clr/CSs = Caso Índice resistente e Caso Secundário sensível
 Clr/CSr = Caso Índice resistente e Caso Secundário resistente
 CIs/CSs = Caso Índice sensível e Caso Secundário sensível

6. Discussão

Os avanços científicos e tecnológicos nas áreas da genética e biologia molecular ocorridos nos últimos anos, particularmente relacionados ao sequenciamento completo do genoma bacteriano, contribuíram substancialmente para uma melhor compreensão das relações da TB humana e do seu patógeno. No presente estudo nos valemos destes avanços para investigar variações genômicas em cepas de MTB em seu estado de dormência no hospedeiro humano.

A discussão dos resultados por nós obtidos e de suas possíveis implicações será apresentada a seguir sob a forma de subitens, com o intuito de facilitar didaticamente a compressão das modificações ocorridas nos subsistemas bacterianos.

6.1 SNPs

O SNP (do inglês *Single nucleotide polymorphisms*) é uma variação pontual em uma única base nucleotídica que ocorre em um determinado organismo, podendo afetar ou não o subproduto de um determinado gene (BROOKES., 1999). A frequência de SNPs é geralmente maior em regiões não codificadoras do que em regiões codificadoras. Fatores tais como a taxa de mutação, recombinação genética e seleção natural podem influenciar a densidade de SNPs (Nachman, 2001; Barreiro et al., 2008).

Colangeli e colaboradores (2020) ao investigarem a taxa de mutação e tempo de geração de *M. tuberculosis* durante ILTB verificaram que os isolados de *M. tuberculosis* dos casos índices e dos contatos correlacionados diferiram um do outro por uma média de 2,25 SNPs. As análises publicadas por Colangeli e colaboradores (2020) foram executadas com a utilização de dois pipelines, o MTBseq com uma média de 2.25 SNPs e o SNPTB com uma média de 17.67 SNPs contra uma média de 5.83 SNPs encontrados em nossa análise realizada com a utilização do software Snippy.

Embora em nosso estudo tenhamos observado uma quantidade de SNPs pouco divergente daquela relatada Colangeli e colaboradores (2020), esta, no entanto, pode ser considerada justificável para este tipo de análise. De acordo com Bush e colaboradores (2020) a utilização de diferentes softwares/pipelines para um determinado conjunto de amostras pode levar a resultados diferentes, isso pode ser explicado pelo fato de que os programas executados pelos softwares podem conter algoritmos bem diferentes.

6.2 Taxas de mutação em *M. tuberculosis* durante o período de latência.

A capacidade dos microrganismos de persistirem em estados metabolicamente inativos em seu hospedeiro ou no ambiente garante a sua sobrevivência em condições desfavoráveis e, provavelmente, contribui para a diversidade microbiana ao facilitar a riqueza taxonômica em sistemas pobres em nutrientes. A persistência do MTB no hospedeiro por anos pode ser explicada pela modificação estrutural e epigenética das células bacterianas, desenvolvendo formas que são menos imunogênicas (PEDDIREDDY, DODDAM & AHMED., 2017) e com padrão de expressão gênica diferente das formas replicantes (VOSKUIL, VISCONTI & SCHOOLNIK., 2004).

Essas células entram em um estado metabólico denominado de dormência, em que frequentemente diminuem de tamanho (ANUCHIN *et al.*, 2009), apresentam espessamento da parede celular (WAYNE & SOHASKEY., 2001), modificam seu comportamento em relação ao metabolismo de macromoléculas (GENGENBACHER & KAUFMANN., 2012) e apresentam perfil de tolerância às drogas utilizadas no tratamento da TB (LORAINE *et al.*, 2016; SARATHY *et al.*, 2018).

Essas células bacterianas podem sobreviver dentro de macrófagos e células dendríticas, bem como em outros reservatórios incomuns, como células epiteliais não-fagocíticas, fibroblastos e adipócitos (EHLERS, 2009). Um bom reservatório de MTB no estado de dormência são microgranulomas presentes no pulmão de pessoas com infecção latente. Nestes indivíduos, pode ocorrer um equilíbrio estático onde há pouca ou nenhuma replicação dos microrganismos (EHLERS, 2009).

Porém, o tecido pulmonar também pode apresentar formas replicantes de MTB que estimulam células T, especialmente na área mais externa dos granulomas em estruturas neolinfóides. Ainda há muita especulação sobre esse assunto, principalmente por falta de modelos experimentais que possam acompanhar in vivo a mudança de infecção latente para uma infecção ativa. Tem sido alimentado o senso comum em que o MTB se encontra em constante transição entre os estados metabólicos ativo e latente (ULRICHS *et al.*, 2004).

Nesta segunda forma (Latente), MTB é mantido em um estado quiescente, sendo este contido pelo sistema imune do hospedeiro dentro dos granulomas (DUTTA; KARAKOUSIS., 2014).

De acordo com Colangeli e colaboradores (2020) a LTBI do inglês "*Latent tuberculosis infection*", não compõem um único estado, podendo haver na verdade um espectro entre outros estados durante a infecção latente, desta forma propõem a divisão da LTBI em dois novos estados, o primeiro chamado de "latência precoce", uma fase com duração de zero a até dois anos, onde a taxa mutacional é predominantemente mais elevada ocorrendo nesta fase a maioria dos casos de infecção secundária (BEHR *et al.*, 2018). Após esse período, chamado de latência precoce, MTB começa a transacionar para uma fase em que a taxa mutacional é mais baixa, sendo esta fase denominada pelos autores de "latência tardia".

Recentemente Colangeli e colaboradores (2020) relatam que a taxa de novas mutações no genoma de MTB diminui drasticamente após dois anos de infecção latente e que mutações indicativas de estresse oxidativo não aumentam com o aumento do tempo de latência.

Neste contexto, nossos resultados corroboram com os mesmos encontrados por Colangeli e colaboradores (2020). Como pode ser observado na tabela quatro, a média de variações genéticas distribuídas em determinados grupos funcionais ao longo dos dois primeiros anos são levemente maiores do que as encontradas no restante do tempo, ou seja, nos dois primeiros anos ocorreram a maior quantidade de variações genéticas ligadas a determinadas funções metabólicas, o que poderia

explicar a taxa mutacional mais elevada encontrada por Colangeli, de igual forma, a redução da taxa mutacional após os dois primeiros anos pode explicar a queda na média de variações genéticas e conseqüentemente e do risco de reativação endógena.

6.3 Variações genéticas em *M. tuberculosis* durante o período de latência e suas possíveis implicações.

Em razão da abordagem de nosso estudo ser pioneira, não encontramos na literatura científica artigos ou publicações detalhadas a respeito das alterações genéticas em MTB que ocorrem durante o período de latência e suas possíveis implicações biológicas. Diante deste fato, torna-se difícil estabelecer comparações e inferências com os resultados obtidos em nosso estudo.

Com base nesta premissa abordamos alguns subtópicos organizados por algumas das principais funções vias/grupos metabólicas abordadas neste estudo, onde modificações genéticas levaram a variações ativas a nível de metabolismo **nos casos secundários** quando comparados com os seus respectivos casos índices. Estes serão chamados de variantes ou variação para fins didáticos terminológicos.

6.3.1 - Virulência:

Dentro deste grupo foram identificadas dezoito variantes nos casos secundários. Em uma cepa constatamos uma variação na família de proteína PPE, uma família de proteínas que possui importante papel regulatório na geração de resposta antigênica (ABDALLAH *et al.*, 2006) e diferentes respostas no hospedeiro de acordo com seu microambiente. Vários estudos apontam que essa família de proteínas pode ser um fator importante na virulência de MTB (VOSKUIL *et al.*, 2004).

Em oito cepas foram encontradas variações relacionadas a proteínas “semelhantes” a Esat6 (*Early secreted antigenic target 6*) na região gênica RD1, ou seja, durante as análises o resultado de determinados genes que sofreram alterações no caso secundário quando comparado com seu caso índice, levaram a uma proteína muito

similar a Esat6, essa inferência é feita através de uma comparação da sequência analisada e sequências encontradas em bancos de dados genéticos e de proteínas. Ao comparar uma dada sequência X onde não sabemos sua função e uma sequência B em um banco de dados onde a função dessa sequência é bem documentada por outros pesquisadores, podemos inferir a função da sequência X a partir da sequência B quando a similaridade de ambas as sequências é alta, essa similaridade também é chamada de grau de identidade (HARDISON., 2003).

Neste contexto, as análises sugerem que em oito cepas foram encontradas alterações em uma região gênica responsável por um operon de virulência com proteínas semelhantes a Esat6. Infelizmente são necessários mais estudos para a compreensão profunda dos mecanismos envolvidos.

A quantidade de alterações encontradas sugere a importância dessa via (Esat6) para a bactéria, Derrick e Morris (2007) argumentam que esta proteína pode ter um papel importante a apoptose de macrófagos humanos, enquanto BERTHET e colaboradores (1998) demonstraram que a Esat6 compõe uma região deletada em cepas atenuadas em *Mycobacterium bovis* BCG utilizado em vacinas, sugerindo que esse sistema pode ter uma atuação na virulência (BERTHET *et al.*, 1998).

Em nove cepas foram constatadas alterações nos genes que codificam as β -lactamases, enzimas que catalisam a hidrólise do anel β -lactâmico, impossibilitando a atividade antimicrobiana. Apesar desta enzima não ser considerada um fator de virulência, mantivemos neste contexto em razão do software Rast (e seus desenvolvedores) terem alocado esta variação dentro do grupo de virulência (DIAS *et al.*, 2018).

Os antibióticos beta-lactâmicos foram usados pela primeira vez na década de 1940 para tratar infecções Gram-positivas. Esses antibióticos inibem a síntese da parede celular ligando-se a transpeptidases que catalisam a reticulação de peptidoglicanos, Estudos mostraram que uma forma alternativa de reticulação de peptidoglicano, chamada L,D-transpeptidase, é predominante em MTB e é particularmente capaz de tolerar a penicilina na fase estacionária de crescimento (WIVAGG *et al.*, 2014).

6.3.2 - Transporte de membrana:

Grande parte das variações encontradas foram relacionadas ao transporte de membrana. Em seis amostras foram identificadas variações no sistema de secreção de proteínas Esat6, em uma amostra no sistema oligopeptídeo transportador ABC e em outra amostra no sistema de aminoacilação de tRNA.

Grande parte das variações genéticas ativas (variação presente no caso secundário) e não ativas (variação presente apenas no caso índice) foram associadas à região Rv3879c, cuja expressão codifica uma proteína rica em alanina e prolina e presente em cepas virulentas do complexo MTB e de acordo com Inwald e colaboradores (2003), não possui papel importante na virulência.

6.3.3 - Regulação e sinalização:

Nas cepas CA3001, CA3043, CA3077 e SE3001 foram identificadas quatro variantes ativas e nas cepas SE3010 e SE3031 duas variantes inativas, ou seja, presente apenas no caso índice e que foram perdidas no caso secundário. Do total dessas 6 variantes, cinco estavam relacionadas ao grupo de proteínas WhiB.

Acredita-se que a proteína WhiB possua um papel importante na regulação do MTB em situações de baixo oxigênio (SOLIVERI *et al.*, 2000) atuando como um fator de transcrição se ligando ou liberando DNA de acordo com as condições do ambiente (CRACK *et al.*, 2004).

A regulação do MTB durante a fase de latência e a forma ativa da doença é crucial para a bactéria, a quantidade de oxigênio disponível pode ser uma variável determinante entre essas fases (RUSTAD *et al.*, 2009). O gene WhiB codifica um total de 7 proteínas com funções regulatórias, porém, infelizmente nossas análises não conseguiram determinar qual proteína WhiB foi o maior alvo dessas variações.

Com uma alta concentração de variações em sistemas relacionadas à virulência, esperávamos importantes variações em regiões reguladoras associadas a tal

característica, pois, sabemos que a proteína WhiB7 possui papel importante na virulência do MTB como demonstrado por Morris e colaboradores (2005), no entanto não conseguimos traçar tal paralelo devido às limitações dos softwares utilizados nas análises e falta de concordância nos pares afetados pelas variações associadas à virulência e a regulação.

6.3.4 - Metabolismo de DNA:

Foram identificadas duas variações relacionadas ao metabolismo de DNA, especificamente relacionadas à replicação. Uma variação foi encontrada na cepa SE3031 e uma foi perdida no caso secundário na cepa SE3063. A implicação dessas variações está relacionada a uma exonuclease 5' - 3' YpcP, que pode atuar em uma variedade de substratos contendo DNA e RNA, também atuando na maturação de fragmentos de okazaki, removendo o RNA desses fragmentos (RANDALL *et al.*, 2019; REUß *et al.*, 2016).

6.3.5 - Sistemas baseados em cluster:

As análises realizadas demonstraram modificações em sistemas baseados em cluster, sistemas que possuem funções que envolvem proteossomos, ribossomos e grupos relacionados a recombinação (MEYER *et al.*, 2008) de acordo com as análises realizadas no sistema RAST.

Apenas a cepa SE3001 apresentou uma variação ativa no caso secundário, outras três amostras perderam uma variação, essas foram as cepas CA3043, SE3031 e VT3067, todas as quatro alterações identificadas estão relacionadas ao tRNA, a maior parte dessas alterações foram na enzima GidA de modificação de uridina 5-carboximetil aminometil.

Acredita-se que a proteína GidA junto com a proteína de ligação TrmE possua função enzimática. A deficiência de ambas pode estar associada a uma diminuição da biossíntese de tRNA, modificado na posição cinco (SCRIMA e WITTINGHOFER., 2004). Essa modificação na região U34 nas posições dois e cinco podem

consequentemente gerar moléculas de 5-metilaminometil-2-tiouridina (YOKOYAMA e NISHIMURA., 1995) nas bactérias mutantes.

6.3.6 - Aminoácidos:

Duas variações foram encontradas relacionadas ao metabolismo de aminoácidos, uma ativa na cepa CA3038 e outra perdida na cepa CA3010, ambos nos casos secundários.

As variações encontradas foram no metabolismo da arginina, especificamente na biossíntese de arginina, afetando a via da N-acetil-gama-glutamil-fosfato redutase codificada pelo gene *argC*. Essa enzima é responsável por catalisar a terceira etapa na síntese de arginina, uma etapa de muito consumo energético pela bactéria (VARGAS-LAGUNAS *et al.*, 2017).

6.3.7 - Cofatores:

Duas variações relacionadas a cofatores foram encontradas, uma ativa na amostra CA3010 e uma perdida na amostra SE3031. Ambas as variações estão relacionadas à biossíntese de menaquinona (também conhecida como vitamina K2) e filoquinona (podendo também ser chamada de vitamina K1). A menaquinona e a filoquinona são importantes compostos transportadores de elétrons no sistema respiratório ou fotossistema bacteriano (SONG., 2014).

O estudo realizado por UPADHYAY e colaboradores (2015) demonstrou que o nocaute do gene *menJ*, responsável por sintetizar uma subunidade da menaquinona (β -isopreno), pode levar MTB a perder a eficácia no processo de transporte de elétrons, porém as micobactérias conseguem compensar e manter o nível de ATP elevado aumentando a concentração de menaquinona na membrana.

Ainda de e acordo com UPADHYAY e colaboradores (2015) o gene *menJ* é necessário para a sobrevivência da micobactéria em macrófagos humanos, e

consequentemente seria um importante alvo para a ação de compostos contra micobactérias patogênicas (UPADHYAY *et al.*, 2015).

6.3.8 - Ácidos graxos e lipídios:

Nas nossas análises identificamos apenas uma variação relacionada ao metabolismo de ácidos graxos, essa variação foi encontrada na amostra SE3013, o sistema afetado foi o da biossíntese de ácidos graxos FASII.

Este sistema juntamente com o FASI está envolvido na síntese de ácidos micólicos, tais componentes são constituintes da parede celular de MTB protegendo o patógeno contra produtos químicos danosos à célula. A presença de ácidos micólicos na parede das micobactérias pertencentes ao complexo MTB, confere um traço morfológico bruto distinto conhecido como "*cording*" (algo como "encordoamento") (TAKAYAMA *et al.*, 2005).

Muitas das características das micobactérias são relacionadas à sua parede celular distinta que é constituída principalmente por lipídeos e ácidos micólicos, que formam uma camada serosa e resistente à água. Essa camada serosa torna as micobactérias resistentes ao ressecamento, impede a entrada de fármacos na célula e potencializa a sua virulência, ao inibir a sua digestão por fagócitos (TORTORA *et al.*, 2016).

Nossas análises sugerem que o proteoma afetado pela variação encontrada seja a enzima "enoil- (proteína transportadora de acila) (ACP) redutase". Segundo MARRAKCHI e colaboradores (2003), essa enzima é um importante catalisador da reação de última etapa no alongamento do ácido graxo no sistema. Contudo, esta observação pontual, necessita de maior aprofundamento, para se avaliar as possíveis consequências na alteração desta enzima e suas implicações nas vias metabólicas onde atua.

6.3.9 - Resistência a drogas

O controle da TB é constantemente ameaçado pelo surgimento de cepas resistentes aos fármacos utilizados no tratamento desta doença. Compreender as bases moleculares e mecanismos envolvidos na resistência é de grande relevância para o diagnóstico e orientação terapêutica, pilares para o controle da doença. Considerando que mesmo em estado de dormência, MTB pode sofrer alterações em seu genoma, indagamos mesmo que de forma improvável se genes associados à resistência a fármacos poderiam também sofrer alterações na fase de latência da infecção (LTBI).

Ao realizarmos análises genômicas de 24 cepas dos casos secundários, não constatamos variações relacionadas a perda ou ganho de resistência a fármacos, o que pode ser amparado de forma plausível pelo fato de que a frequência de mutações que conferem algum tipo de resistência em MTB são muito raras, com cerca de uma mutação em INH ocorrendo em 10^6 replicações, em RIF ocorrendo em 10^8 , em EMB ocorrendo em 10^6 , em Strep 10^6 e em INH + RIF ocorrendo em 10^{14} conforme demonstrado por Ford e colaboradores (2003).

Das vinte e quatro cepas estudadas apenas três apresentaram resistência a algum tipo de droga. A cepa VT3067, tanto o caso índice como o caso secundário, apresentou resistência a Rifampicina. Era esperado que a perda da resistência ou o seu ganho no caso secundário fosse um evento incomum, pois genes responsáveis pela resistência a Rifampicina são em sua maioria domínios conservados, dessa forma, nossos resultados corroboram com TELENTI e colaboradores (1993).

De forma semelhante os casos índices e seus respectivos casos secundários das cepas VT3067 e VT3059 apresentaram resistência à Estreptomicina, um medicamento muito utilizado no tratamento da TB. A resistência mantida nos casos secundários também pode ser explicada por áreas genéticas altamente conservadas (HONORÉ e COLE., 1994).

6.3.10 – Comentários finais

A taxa de mutação do MTB pode ser modulada dentro do hospedeiro e devido a isso causas ambientais podem ser fatores mutagênicos importantes durante a infecção pelo MTB, conseqüentemente influenciando a taxa de mutação que varia de hospedeiro para hospedeiro (MCGRATH et al., 2014).

As alterações no genoma de MTB durante a LTBI, fase em que a bactéria se encontra em um ambiente adverso ocorrem em baixa frequência, conforme demonstrado em nosso estudo, principalmente após 24 meses da infecção latente. Além disto, estas alterações, como demonstrado de forma pioneira em nosso estudo não podem ser desprezadas, pelo contrário, apresentam grande complexidade e necessitam de estudo complementares utilizando-se outras "ômicas" para corroborar as implicações biológicas das alterações genômicas de MTB aventadas neste estudo, durante o período de dormência da bactéria.

7. Conclusões

1 - Alterações genéticas em *M. tuberculosis* durante o período de dormência ocorrem em baixa frequência;

2 - Alterações genéticas de *M. tuberculosis* ocorrem com maior frequência nos dois primeiros anos da infecção latente (LTBI) do que nos anos subsequentes a este período;

3 - Alterações genéticas em *M. tuberculosis* durante o período de dormência são mais comumente observadas no subsistema de virulência e de transporte de membrana;

4 - Alterações nos subsistemas de dormência e resposta ao stress não ocorrem durante o período de dormência, possivelmente para manter a integridade das funções desses genes fundamentais para sobrevivência da bactéria no hospedeiro;

5 – Não ocorrem alterações genéticas que levam a perda ou ganho de resistência de *M. tuberculosis* durante o período de dormência;

6 - As alterações genéticas em *M. tuberculosis* durante o período de dormência apresentam grande complexidade e necessitam de estudos complementares utilizando-se outras "ômicas" para corroborar as implicações biológicas das alterações genômicas de MTB aventadas neste estudo.

8. Referências

- ABDALLAH, Abdallah M. *et al.* A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. **Molecular microbiology**, v. 62, n. 3, p. 667-679, 2006.
- ALEXANDER, Kathleen A. *et al.* Novel Mycobacterium tuberculosis complex pathogen, *M. mungi*. **Emerging infectious diseases**, v. 16, n. 8, p. 1296, 2010.
- ALTMANN, André *et al.* A beginners guide to SNP calling from high-throughput DNA-sequencing data. **Human genetics**, v. 131, n. 10, p. 1541-1554, 2012.
- ANDREWS, Simon *et al.* FastQC: a quality control tool for high throughput sequence data. 2010.
- ANUCHIN, Aleksey M. *et al.* Dormant forms of *Mycobacterium smegmatis* with distinct morphology. **Microbiology**, v. 155, n. 4, p. 1071-1079, 2009.
- AZIZ, Ramy K. *et al.* The RAST Server: rapid annotations using subsystems technology. **BMC genomics**, v. 9, n. 1, p. 75, 2008.
- BANKEVICH, Anton *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology**, v. 19, n. 5, p. 455-477, 2012.
- BARRY, Clifton E. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. **Nature Reviews Microbiology**, v. 7, n. 12, p. 845-855, 2009.
- BEHR, Marcel A.; EDELSTEIN, Paul H.; RAMAKRISHNAN, Lalita. Revisiting the timetable of tuberculosis. *Bmj*, v. 362, 2018.
- BERTHET, Francois-Xavier *et al.* A *Mycobacterium tuberculosis* operon encoding ESAT= 6 and a novel low-molecular-mass culture filtrate protein (CFP-10). **Microbiology**, v. 144, n. 11, p. 3195-3203, 1998.

BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014.

BROOKES, Anthony J. The essence of SNPs. **Gene**, v. 234, n. 2, p. 177-186, 1999.

BROSCH, Roland *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. **Proceedings of the national academy of Sciences**, v. 99, n. 6, p. 3684-3689, 2002.

BRUDEY, Karine *et al.* *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. **BMC microbiology**, v. 6, n. 1, p. 1-17, 2006.

BUSH, Stephen J. *et al.* Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines. **GigaScience**, v. 9, n. 2, p. g1aa007, 2020.

CADENA, Anthony M.; FORTUNE, Sarah M.; FLYNN, JoAnne L. Heterogeneity in tuberculosis. **Nature Reviews Immunology**, v. 17, n. 11, p. 691-702, 2017.

CAÑO-MUÑIZ, Santiago *et al.* New approaches and therapeutic options for *Mycobacterium tuberculosis* in a dormant state. **Clinical microbiology reviews**, v. 31, n. 1, 2018.

CHAN, John; KAUFMANN, Stefan HE. Immune mechanisms of protection. Tuberculosis: pathogenesis, protection, and control, p. 387-415, 1994.

CHIEN, H. P. *et al.* Comparison of the BACTEC MGIT 960 with Löwenstein-Jensen medium for recovery of mycobacteria from clinical specimens. **The International Journal of Tuberculosis and Lung Disease**, v. 4, n. 9, p. 866-870, 2000.

COMAS, Inaki *et al.* Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. **PloS one**, v. 4, n. 11, p. e7815, 2009.

COHEN, Keira A.; BISHAI, William R.; PYM, Alexander S. Molecular basis of drug resistance in *Mycobacterium tuberculosis*. **Molecular Genetics of Mycobacteria**, p. 411-429, 2014.

COLANGELI, Roberto *et al.* *Mycobacterium tuberculosis* progresses through two phases of latent infection in humans. **Nature communications**, v. 11, n. 1, p. 1-10, 2020.

COLANGELI, Roberto *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. **PLoS One**, v. 9, n. 3, p. e91024, 2014.

COLE, S. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. **Nature**, v. 393, n. 6685, p. 537-544, Jun, 1998.

COSCOLLA, Mireia *et al.* Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. **Emerging infectious diseases**, v. 19, n. 6, p. 969, 2013.

COSCOLLA, Mireia; GAGNEUX, Sebastien. Consequences of genomic diversity in *Mycobacterium tuberculosis*. In: *Seminars in immunology*. p. 431-444. **Academic Press**, 2014.

DELOGU, Giovanni; SALI, Michela; FADDA, Giovanni. The biology of mycobacterium tuberculosis infection. **Mediterranean journal of hematology and infectious diseases**, v. 5, n. 1, 2013.

CRACK, Jason; THOMSON, Andrew J.; GREEN, Jeffrey. Mechanism of oxygen sensing by the bacterial transcription factor fumarate-nitrate reduction (FNR). **Journal of Biological Chemistry**, v. 279, n. 10, p. 9278-9286, 2004.

DEMAY, Christophe *et al.* SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. **Infection, genetics and evolution**, v. 12, n. 4, p. 755-766, 2012.

DERRICK, Steven C.; MORRIS, Sheldon L. The ESAT6 protein of *Mycobacterium tuberculosis* induces apoptosis of macrophages by activating caspase expression. **Cellular microbiology**, v. 9, n. 6, p. 1547-1555, 2007.

DIAS, Marcio Vinicius Bertacine *et al.* L, D-transpeptidases de *Mycobacterium tuberculosis*: estudo das interações com antibióticos β -lactâmicos e triagem de fragmentos, 2018.

DIETZE, R., HADAD, D. J., PEREIRA, F. E. L. e RODRIGUES, R. R. Tuberculose. In: Rocha, M.O.C.; Pedroso, E.R.P. (Eds). Fundamentos em Infectologia. Ltd, Editora Rubio. Rio de Janeiro, Brasil. p. 703-38. 2009.

DORMAN, Susan E. *et al.* Xpert MTB/RIF Ultra for detection of *Mycobacterium tuberculosis* and rifampicin resistance: a prospective multicentre diagnostic accuracy study. **The Lancet infectious diseases**, v. 18, n. 1, p. 76-84, 2018.

DUTTA, Noton K.; KARAKOUSIS, Petros C. Latent tuberculosis infection: myths, models, and molecular mechanisms. **Microbiology and Molecular Biology Reviews**, v. 78, n. 3, p. 343-371, 2014.

ELLNER, Jerrold J. The immune response in human tuberculosis: implications for tuberculosis control. **The Journal of infectious diseases**, v. 176, n. 5, p. 1351-1359, 1997.

EHLERS, S. Lazy, dynamic or minimally recrudescing? On the elusive nature and location of the mycobacterium responsible for latent tuberculosis. **Infection**, v. 37, n. 2, p. 87-95, 2009.

FARHAT, Maha R. *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. **Nature genetics**, v. 45, n. 10, p. 1183, 2013.

FENTON, Matthew J.; VERMEULEN, Mary W. Immunopathology of tuberculosis: roles of macrophages and monocytes. *Infection and immunity*, v. 64, n. 3, p. 683, 1996.

FORD, Christopher B. *et al.* Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. **Nature genetics**, v. 43, n. 5, p. 482-486, 2011.

FRIEDEN, Thomas R.; SBARBARO, John A. Promoting adherence to treatment for tuberculosis: the importance of direct observation. **Bulletin of the World Health Organization**, v. 85, p. 407-409, 2007.

FURIN, Jennifer J.; JOHNSON, John L. Recent advances in the diagnosis and management of tuberculosis. **Current opinion in pulmonary medicine**, v. 11, n. 3, p. 189-194, 2005.

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

GAGNEUX, Sebastien. Ecology and evolution of *Mycobacterium tuberculosis*. **Nature Reviews Microbiology**, v. 16, n. 4, p. 202-213, 2018.

GENGENBACHER, Martin; KAUFMANN, Stefan HE. *Mycobacterium tuberculosis*: success through dormancy. **FEMS microbiology reviews**, v. 36, n. 3, p. 514-532, 2012.

GOREN, Mayer B. *et al.* Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, v. 73, n. 7, p. 2510-2514, 1976.

GOROCICA, P; JIMENEZ-MARTINEZ, M; GARFIAS, Y.; SADA, I.; LASCURAIN, R. Componentes glicosilados de la envoltura de *Mycobacterium tuberculosis* que intervienen en la patogénesis de la tuberculosis. **Revista del Instituto Nacional de Enfermedades Respiratorias**, México, v.18, n.2, p.142-153, 2005.

GUPTA, Antima *et al.* *Mycobacterium tuberculosis*: immune evasion, latency and reactivation. **Immunobiology**, v. 217, n. 3, p. 363-374, 2012.

GUREVICH, Alexey *et al.* QUASt: quality assessment tool for genome assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072-1075, 2013.

HARDISON, Ross C. Comparative genomics. **PLoS biology**, v. 1, n. 2, p. e58, 2003.

HERMANS, P. W. *et al.* Insertion element IS986 from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *Journal of clinical microbiology*, v. 28, n. 9, p. 2051-2058, 1990.

HONORÉ, Nadine; COLE, Stewart T. Streptomycin resistance in mycobacteria. *Antimicrobial agents and chemotherapy*, v. 38, n. 2, p. 238-242, 1994.

HOUBEN, Rein MGJ; DODD, Peter J. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. **PLoS medicine**, v. 13, n. 10, p. e1002152, 2016.

HOWARD, A. D.; ZWILLING, B. S. Reactivation of tuberculosis is associated with a shift from type 1 to type 2 cytokines. **Clinical and experimental immunology**, v. 115, n. 3, p. 428, 1999.

INWALD, Jacqueline *et al.* Inactivation of the *Mycobacterium bovis* homologue of the polymorphic RD1 gene Rv3879c (Mb3909c) does not affect virulence. **Tuberculosis**, v. 83, n. 6, p. 387-393, 2003.

ISLAM, Md Mahmudul *et al.* Drug resistance mechanisms and novel drug targets for tuberculosis therapy. **Journal of genetics and genomics**, v. 44, n. 1, p. 21-37, 2017.

JONES-LÓPEZ, Edward C. *et al.* Importance of cough and M. tuberculosis strain type as risks for increased transmission within households. **PloS one**, v. 9, n. 7, p. e100984, 2014.

KENT, Patricia T. Public health mycobacteriology: a guide for the level III laboratory. US Department of Health and Human Services, Public Health Service, Centers for Disease Control, 1985.

KONDRATIEVA, Tatiana *et al.* Latent tuberculosis infection: what we know about its genetic control?. **Tuberculosis**, v. 94, n. 5, p. 462-468, 2014.

KRUK, Margaret E.; SCHWALBE, Nina R.; AGUIAR, Christine A. Timing of default from tuberculosis treatment: a systematic review. **Tropical Medicine & International Health**, v. 13, n. 5, p. 703-712, 2008.

LAMICHHANE, SHREE RAM; MILIC, NATALIE. *Mycobacterium tuberculosis*: GENE AND GENOME ANALYSIS. Asian Journal of Microbiology and Biotechnology, v. 3, n. 1, p. 24-33, 2018.

LANGLEY, Ivor *et al.* Assessment of the patient, health system, and population effects of Xpert MTB/RIF and alternative diagnostics for tuberculosis in Tanzania: an integrated modelling approach. **The Lancet Global Health**, v. 2, n. 10, p. e581-e591, 2014.

LEWINSOHN, David M. *et al.* Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention clinical practice guidelines: diagnosis of tuberculosis in adults and children. **Clinical Infectious Diseases**, v. 64, n. 2, p. e1-e33, 2017.

LORAINE, Jessica *et al.* Development of an in vitro assay for detection of drug-induced resuscitation-promoting-factor-dependent mycobacteria. **Antimicrobial Agents and Chemotherapy**, v. 60, n. 10, p. 6227-6233, 2016.

MARRAKCHI, Hedia *et al.* Characterization of Streptococcus pneumoniae enoyl-(acyl-carrier protein) reductase (FabK). **Biochemical Journal**, v. 370, n. 3, p. 1055-1062, 2003.

MCGRATH, M. *et al.* Mutation rate and the emergence of drug resistance in Mycobacterium tuberculosis. **Journal of Antimicrobial Chemotherapy**, v. 69, n. 2, p. 292-302, 2014.

MEYER, Folker *et al.* The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC bioinformatics**, v. 9, n. 1, p. 1-8, 2008.

MS (2020). MINISTÉRIO DA SAÚDE. Boletim Epidemiológico de Tuberculose 2020. Disponível em: <<http://www.aids.gov.br/pt-br/pub/2020/boletim-epidemiologico-de-tuberculose-2020/>>. Acesso em 29/12/2020.

MORRIS, Rowan P. *et al.* Ancestral antibiotic resistance in Mycobacterium tuberculosis. **Proceedings of the National Academy of Sciences**, v. 102, n. 34, p. 12200-12205, 2005.

NORTH, Robert J.; JUNG, Yu-Jin. Immunity to tuberculosis. *Annu. Rev. Immunol.*, v. 22, p. 599-623, 2004.

PAI, Madhukar; SOTGIU, Giovanni. Diagnostics for latent TB infection: incremental, not transformative progress. 2016.

PALOMINO, Juan Carlos. Newer diagnostics for tuberculosis and multidrug resistant tuberculosis. *Current opinion in pulmonary medicine*, v. 12, n. 3, p. 172-178, 2006.

PARK, Se-Ho; BENDELAC, Albert. CD1-restricted T-cell responses and microbial infection. **Nature**, v. 406, n. 6797, p. 788-792, 2000.

PEDDIREDDY, Vidyullatha; DODDAM, Sankara Narayana; AHMED, Niyaz. Mycobacterial dormancy systems and host responses in tuberculosis. **Frontiers in immunology**, v. 8, p. 84, 2017.

RADFORD, Alan D. *et al.* Application of next-generation sequencing technologies in virology. *The Journal of general virology*, v. 93, n. Pt 9, p. 1853, 2012.

RAMASWAMY, S.; MUSSER, James M. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. **Tubercle and Lung disease**, v. 79, n. 1, p. 3-29, 1998.

RANDALL, Justin R. *et al.* RNase HIII is important for Okazaki fragment processing in *Bacillus subtilis*. **Journal of bacteriology**, v. 201, n. 7, p. e00686-18, 2019.

REUß, Daniel R. *et al.* The blueprint of a minimal cell: MiniBacillus. **Microbiology and Molecular Biology Reviews**, v. 80, n. 4, p. 955-987, 2016.

RUSTAD, Tige R. *et al.* The enduring hypoxic response of *Mycobacterium tuberculosis*. **PloS one**, v. 3, n. 1, p. e1502, 2008.

RUSTAD, Tige R. *et al.* Hypoxia: a window into *Mycobacterium tuberculosis* latency. *Cellular microbiology*, v. 11, n. 8, p. 1151-1159, 2009.

SALGAME, Padmini *et al.* Latent tuberculosis infection—revisiting and revising concepts. **Tuberculosis**, v. 95, n. 4, p. 373-384, 2015.

SARAH GALBRAITH-EMAMI/WHO. Global Tuberculosis Report 2015. {on line}. Disponível na Internet via http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf. Arquivo

SARATHY, Jansy P. *et al.* Extreme drug tolerance of *Mycobacterium tuberculosis* in caseum. **Antimicrobial agents and chemotherapy**, v. 62, n. 2, p. e02266-17, 2018.

SEEMANN, Torsten. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, v. 30, n. 14, p. 2068-2069, 2014.

SCHLOSSBERG, David L.; SAMUEL, Rafik. *Antibiotics manual: a guide to commonly used antimicrobials*. John Wiley & Sons, 2017.

SCHLUGER, Neil W.; ROM, William N. The host immune response to tuberculosis. **American journal of respiratory and critical care medicine**, v. 157, n. 3, p. 679-691, 1998.

SCRIMA, A.; WITTINGHOFER, A. Proteins involved in the cmnm5U-modification at the wobble-position of tRNA (2004).

STYBLO, Karel; ENARSON, Donald A. *Epidemiology of tuberculosis*. **The Hague: Royal Netherlands Tuberculosis Association**, 1991.

SOLIVERI, J. A. *et al.* Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in *Streptomyces* and other actinomycetes The GenBank accession numbers for the sequences reported in this paper are: *wbIA*, AJ239085; *wbIB*, AJ239086; *wbIE*, AJ239087; and *wbII*, AJ239088. **Microbiology**, v. 146, n. 2, p. 333-343, 2000.

SONG, Haigang. *Biochemical and structural studies of enzymes involved in the biosynthesis of menaquinone and calicheamicin*. 2014. Tese de Doutorado.

SOTGIU, Giovanni *et al.* Epidemiology and clinical management of XDR-TB: a systematic review by TBNET. **European Respiratory Journal**, v. 33, n. 4, p. 871-881, 2009.

PHILIPS, Jennifer A.; ERNST, Joel D. Tuberculosis pathogenesis and immunity. *Annual Review of Pathology: Mechanisms of Disease*, v. 7, p. 353-384, 2012.

TAKAYAMA, Kuni; WANG, Cindy; BESRA, Gurdyal S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. **Clinical microbiology reviews**, v. 18, n. 1, p. 81-101, 2005.

TELENTI, Amalio *et al.* Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. **The Lancet**, v. 341, n. 8846, p. 647-651, 1993.

THOMAS, Jessica A. *et al.* A generation time effect on the rate of molecular evolution in invertebrates. **Molecular biology and evolution**, v. 27, n. 5, p. 1173-1180, 2010.

TORTORA, Gerard J.; CASE, Christine L.; FUNKE, Berdell R. **Microbiologia-12ª Edição**. Artmed Editora, 2016.

ULRICHS, Timo *et al.* Differential T cell responses to *Mycobacterium tuberculosis* ESAT6 in tuberculosis patients and healthy donors. **European journal of immunology**, v. 28, n. 12, p. 3949-3958, 1998.

ULRICHS, Timo *et al.* Human tuberculous granulomas induce peripheral lymphoid follicle-like structures to orchestrate local host defence in the lung. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, v. 204, n. 2, p. 217-228, 2004.

UPADHYAY, Ashutosh *et al.* Partial saturation of menaquinone in *Mycobacterium tuberculosis*: function and essentiality of a novel reductase, MenJ. *ACS central science*, v. 1, n. 6, p. 292-302, 2015.

VAN INGEN, Jakko *et al.* Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. **Emerging infectious diseases**, v. 18, n. 4, p. 653, 2012.

VARGAS-LAGUNAS, Carmen *et al.* Overproduction of *S. meliloti* ArgC (N-acetyl-gamma-glutamyl phosphate reductase) promotes growth delay and inefficient nodules. **FEMS Microbiology Letters**, 2017.

VINHAS, S. ESTUDO SOBRE CARACTERÍSTICAS GENÉTICAS DE *Mycobacterium Tuberculosis* ISOLADOS DE PACIENTES COM E SEM LESÕES CAVITÁRIAS. Tese (Doutorado em doenças infecciosas) – Universidade Federal do Espírito Santo. Espírito Santo, 2013.

VITOL, Inna *et al.* Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. *Infection, Genetics and Evolution*, v. 6, n. 6, p. 491-504, 2006.

VOSKUIL, M. I. *et al.* Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis*, v. 84, n. 3-4, p. 256-262, 2004.

VOSKUIL, M. I.; VISCONTI, K. C.; SCHOOLNIK, G. K. *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. **Tuberculosis**, v. 84, n. 3-4, p. 218-227, 2004.

VYNNYCKY, Emilia; FINE, Paul EM. Lifetime risks, incubation period, and serial interval of tuberculosis. **American journal of epidemiology**, v. 152, n. 3, p. 247-263, 2000.

WALKER, Timothy M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. **The Lancet infectious diseases**, v. 13, n. 2, p. 137-146, 2013.

WALLIS, Robert S. *et al.* Inhibition of isoniazid-induced expression of *Mycobacterium tuberculosis* antigen 85 in sputum: potential surrogate marker in tuberculosis chemotherapy trials. **Antimicrobial agents and chemotherapy**, v. 45, n. 4, p. 1302-1304, 2001.

WALLIS, Robert S. *et al.* Tuberculosis—advances in development of new drugs, treatment regimens, host-directed therapies, and biomarkers. **The Lancet infectious diseases**, v. 16, n. 4, p. e34-e46, 2016.

WAYNE, Lawrence G.; SOHASKEY, Charles D. Nonreplicating persistence of *Mycobacterium tuberculosis*. **Annual Reviews in Microbiology**, v. 55, n. 1, p. 139-163, 2001.

WEI, Liping *et al.* Comparative genomics approaches to study organism similarities and differences. **Journal of biomedical informatics**, v. 35, n. 2, p. 142-150, 2002.

WELLER, Cory; WU, Martin. A generation-time effect on the rate of molecular evolution in bacteria. *Evolution*, v. 69, n. 3, p. 643-652, 2015.

WHITELAW, Andrew *et al.* Comparative cost and performance of light-emitting diode microscopy in HIV–tuberculosis-co-infected patients. **European Respiratory Journal**, v. 38, n. 6, p. 1393-1397, 2011.

WIVAGG, Carl N.; BHATTACHARYYA, Roby P.; HUNG, Deborah T. Mechanisms of β -lactam killing and resistance in the context of *Mycobacterium tuberculosis*. **The Journal of antibiotics**, v. 67, n. 9, p. 645-654, 2014.

WHO (2020). Organização Mundial da Saúde (World Health Organization). Global Tuberculosis Report 2017. Disponível em: <<https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf/>>. Acesso em 29/12/2020.

YOKOYAMA, S.; NISHIMURA, S. tRNA: structure, biosynthesis, and function. Modified Nucleosides and Codon Recognition. ASM Press, Washington, DC, p. 207-223, 1995.

YOUNG, Douglas B.; GIDEON, Hannah P.; WILKINSON, Robert J. Eliminating latent tuberculosis. **Trends in microbiology**, v. 17, n. 5, p. 183-188, 2009.

APÊNDICE A - Análise preliminar dos dados das vias e funções afetadas pelas modificações ocorridas.

Os dados das vias afetadas, assim como sua função ao longo do tempo nos casos secundários em relação aos casos índices são mostrados na tabela 6 abaixo.

TABELA 6 – Dados detalhados das modificações metabólicas encontradas

Amostra	Sistemas Afetados no caso secundário com variante ativa	Subsistemas
CA3001	Sistemas de estabilização de replicon de toxina-antitoxina	Regulação e sinalização celular
	Tiorredoxina-dissulfeto redutase	Metabolismo do enxofre
CA3010	Resistência a antibióticos e compostos tóxicos	Virulência
	Cofatores de quinona	Cofatores e aminoácidos
	Arginina; ciclo da ureia, poliaminas	Aminoácidos (Perda)
	Translocação de proteínas através da membrana citoplasmática	Transporte de Membrana (Perda)
CA3038	Arginina; ciclo da ureia, poliaminas	Aminoácidos
	Translocação de proteínas através da membrana citoplasmática	Transporte de Membrana
	Resistência a antibióticos e compostos tóxicos	Virulência
CA3043	Operon de virulência de Mycobacterium com proteínas semelhantes a Esat6	Virulência
	Modificação de RNA e cluster de partição de cromossomos	Subsistemas baseados em cluster
	Proteínas regulatórias do tipo WhiB e WhiB	Regulação e Sinalização
CA3050	Operon de virulência de Mycobacterium com família PE e proteínas semelhantes a Esat6	Virulência

CA3051	Translocação de proteínas através da membrana citoplasmática	Virulência (Perda)
	Translocação de proteínas através da membrana citoplasmática	Transporte de membrana
CA3061	Beta-lactamase	Virulência
CA3064	Sistema de secreção de proteínas ESAT-6 em Actinobactérias	Transporte de membrana
CA3077	Proteínas regulatórias do tipo WhiB e WhiB	Regulação e Sinalização
	Operon de virulência de Mycobacterium com proteínas semelhantes a Esat6	Virulência
	Beta-lactamase	Virulência
CA3078	Beta-lactamase	Virulência
	Operon de virulência de Mycobacterium com proteínas semelhantes a Esat6	Virulência
SE3001	Modificação de RNA e cluster de partição de cromossomos	Subsistemas baseados em cluster
	Proteínas regulatórias do tipo WhiB e WhiB	Regulação e Sinalização
SE3010	Beta-lactamase	Virulência
	Proteínas regulatórias do tipo WhiB e WhiB	Regulação e Sinalização
SE3013	Biossíntese de ácidos graxos FASII	Ácidos graxos e lipídeos
SE3031	Provável 5'-3' exonuclease Bsu YpcP	Metabolismo de DNA
	Modificação de RNA e cluster de partição de cromossomos	Subsistemas baseados em cluster (Perda)
	Biossíntese de Menaquinona e Filoquinona	Cofatores e aminoácidos
	Proteínas regulatórias do tipo WhiB e WhiB	Regulação e sinalização
	Aminoacilação de tRNA, Pro	Metabolismo de proteína
SE3035	Operon de virulência de Mycobacterium com proteínas semelhantes a Esat6	Virulência
	Sistema de secreção de proteínas ESAT-6 em Actinobactérias	Transporte de membrana
	Beta-lactamase	Virulência
SE3063	Provável 5'-3' exonuclease Bsu YpcP	Metabolismo de DNA
	Sistema de secreção de proteínas ESAT-6 em Actinobactérias	Transporte de Membrana

VT3010	Sistema de secreção de proteínas ESAT-6 em Actinobactérias	Transporte de Membrana
	Beta-lactamase	Virulência
VT3067	Operon de virulência de Mycobacterium com proteínas semelhantes a Esat6	Virulência
	Beta-lactamase	Virulência (Perda)
	Modificação de RNA e cluster de partição de cromossomos	Subsistemas baseados em cluster
VV3012	Sistema de secreção de proteínas ESAT-6 em Actinobactérias	Transporte de membrana
VV3015	Operon de virulência de Mycobacterium com proteínas semelhantes a Esat6	Virulência
	Beta-lactamase	Virulência
	Aminoacilação de tRNA, Pro	Metabolismo de proteína

GLOSSÁRIO

Anotação: Processo de expandir o conhecimento/dados de um determinado alvo de estudo. Em bioinformática pode se referir ao processo de extrair informações de dados brutos.

Baixa cobertura: Baixa quantidade de vezes em que a máquina de sequenciamento leu um determinado trecho de DNA.

Read: Durante o sequenciamento o DNA é fragmentado em pedaços pequenos para que a leitura das bases seja realizada, esses fragmentos são chamados de *reads*.

Trimagem: O sequenciamento completo do genoma envolve várias enzimas, produtos químicos e biológicos. Um desses produtos são chamados de adaptadores, os adaptadores auxiliam o sequenciamento, por vezes levando um determinado “*read*” a uma base de leitura. O processo de trimagem ocorre para remover esses adaptadores após o sequenciamento, dessa forma, evitando que determinadas sequências não influenciam nas análises.

k-mer: Subsequências de tamanho variável usadas durante a montagem do genoma.

Pipeline: Script de execução e uma determinada técnica. Em bioinformática pode se tratar de um script para automatizar a execução de um determinado software em amostras.

SIDA: Síndrome da imunodeficiência adquirida. Doença causada pelo vírus do HIV.

Bioinformática: Ramo da ciência que aplica a ciência da computação, matemática e estatística às ciências biológicas.

SNPs: SNPs são polimorfismos de base única (*Single nucleotide polymorphism* - SNP), são modificações pontuais que ocorrem no genoma a nível de nucleotídeo,

essas modificações podem envolver uma remoção de uma base nitrogenada, a adição de uma base ou um rearranjo complexo entre bases.

Montagem: O sequenciamento do genoma completo por vezes fragmenta a sequência de DNA de um determinado organismo para sequenciá-lo. O processo de reconstrução da sequência original do DNA é chamado de montagem.

Criptografia de duas chaves: Método utilizado para certificar a integridade de arquivos que trafegam pela internet ou em uma rede interna. O arquivo recebe uma criptografia (número de identificação criado a partir de dados binários), a criptografia é atrelada ao arquivo de modo que ele seja lido através de determinados índices. Quando o arquivo deixa a origem a integridade pode ser checada na outra ponta através do índice criado. Caso o índice não seja correspondente, o arquivo pode estar corrompido.