INTERNATIONAL SEVEN JOURNAL
OF HEALTH RESEARCH
ISSN: 2764-9415

Complete genome sequencing of *Mycobacterium Tuberculosis* in the identification of gene mutations associated with antimicrobial resistance

Sequenciamento completo do genoma do *Mycobacterium Tuberculosis* na identificação de mutações gênicas associadas a resistência antimicrobiana

DOI: 10.56238/isevjhv2n2-007
Receiving the originals: 06/03/2023
Acceptance for publication: 28/03/2023

Iago Prina Rocha
Mestrando em Ciências da Saúde
Instituição: Universidade Federal do Rio Grande (FURG)
Endereço: Campus Carreiros. Av. Itália, km 8, bairro Carreiros, Rio Grande, RS, Brasil.
E-mail: iagoprina@hotmail.com

Caroline Valois Cunha
Enfermeira Especialista
Instituição: Universidade Estadual do Sudoeste da Bahia (UESB)
Endereço: Av. José Moreira Sobrinho, s/n - Jequiezinho, Jequié - BA

Ana Paula Furtado Carneiro da Fontoura
Doutoranda em Ciências da Saúde
Instituição: Universidade Federal do Rio Grande (FURG)
Endereço: Campus Carreiros. Av. Itália, km 8, bairro Carreiros, Rio Grande, RS, Brasil.

Rafaela Sinnott Dias
Mestranda em Ciências da Saúde
Instituição: Universidade Federal do Rio Grande (FURG)
Endereço: Campus Carreiros. Av. Itália, km 8, bairro Carreiros, Rio Grande, RS, Brasil.

Ivana da Cruz Goulart
Doutoranda em Ciências da Saúde
Instituição: Universidade Federal do Rio Grande (FURG)
Endereço: Campus Carreiros. Av. Itália, km 8, bairro Carreiros, Rio Grande, RS, Brasil.

Laura Fontoura Perim
Mestranda em Enfermagem
Instituição: Universidade Federal do Rio Grande (FURG)
Endereço: Campus Carreiros. Av. Itália, km 8, bairro Carreiros, Rio Grande, RS, Brasil.

Ana Claudia Klein de Almeida de Chaves
Enfermeira Especialista
Instituição: Universidade Federal do Rio Grande (FURG)
Endereço: Campus Carreiros. Av. Itália, km 8, bairro Carreiros, Rio Grande, RS, Brasil.

ABSTRACT

The development and application of new molecular tools has shown the real complexity involved in *Mycobacterium Tuberculosis* infection. The sequencing of the complete Koch's Bacillus genome can reveal information about the true extent of the point mutations that confer resistance

to treatment to these bacteria. Complete genome sequencing techniques and their availability in public databases, have been responsible for the exponential growth of computational tools to extract information about important features encoded in the genomes of Mycobacterium complex species. This is a literature review study, from scientific research platforms, with the aim of discussing how computational tools used for analysis of the complete genome sequencing of M. tuberculosis can identify the genetic relationship between strain mutations and antimicrobial resistance. Whole genome sequencing is considered a surveillance marker of transmission and drug resistance, as well as a strong predictor of the emergence of possible resistance mechanisms for first- and second-line drugs used in the treatment of pulmonary tuberculosis.

Keywords: Genome, Antibiotics, Tuberculosis

1 INTRODUCTION

With the advent of new scientific and technological development processes after the end of World War II in the year 1945, more complex and precise techniques began to be used by pharmaceutical industries, health researchers, academic researchers, and other public and private sectors for the refinement and discovery of new substances and molecules with biological activity (CHAKRAVORTY, 2017).

The completion of the Human Genome Project, starting in 1970, brought new information about molecular pathways and cellular responses to various diseases, this evolution of knowledge of molecular and cell biology, revolutionized when cell receptors became adjuncts to strategies based on molecular targets for the discovery of new therapeutic substances (CHAKRAVORTY, 2017; DA COSTA, 2017).

The computational resources used for the identification of new substances also became part of the drug development process in pharmaceutical industries, the virtual *in sílico* methodologies, increased the tactical productivity of research and decreased the time and cost during the development phases. With the improvement of previous knowledge of the histological, physiological and pathological mechanisms responsible for some diseases, new promising therapeutic compounds could be identified for treatment and from there, develop more suitable *in sílico* strategies with the available data on the structure of the selected molecular target and known ligands (DOMÍNGUEZ, 2018; LIMA, 2017; MEEHAN, 2019).

Since the availability of the Koch's Bacillus genome sequencing in 1998, the knowledge and understanding of this bacterium has increased significantly, but the expectation that the post-genomic era would lead to new and effective therapeutic interventions still remains utopian. The extreme complexity of the mycobacterial cell envelope, which confers low permeability, is one of the factors that could explain such failure. Current drugs or candidates for Tuberculosis (TB)

treatment have been identified using the drug-to-target approach in mycobacteria. This approach relies on screening small molecules in cultured bacteria (whole-cell screening) or infected cells (high-content screening) (DOMÍNGUEZ, 2018; PEDELACQ, 2020; TUNSTALL, 2020).

The research in bacterial genomics has conquered space in the sciences due to the objective of improving the current treatment. The development of new TB drugs follows some general goals, which would be: the shortening in the duration of treatment and possibly decrease the dosing frequency; improvement in the bactericidal effect in replicative and non-replicative populations; therapeutic effect against drug-resistant strains; reduction of drug interactions with other drugs and adverse effects; better oral bioavailability and efficient pharmacokinetic pattern with greater tissue penetration and finally the low cost ensuring universal accessibility (IKETLENG, 2018).

The development and application of new molecular tools has shown the real complexity involved in *Mycobacterium Tuberculosis* infection. Sequencing the complete Koch's Bacillus genome can reveal information about the true extent of the point mutations that confer resistance to treatment to these bacteria. Analysis of genomic information has begun to be used as an epidemiological and clinical marker in the study of tuberculosis (DA COSTA, 2017; MUNIR, 2020; PÉREZ-LAGO, 2014; AMBROSETTI, 2020).

2 METHODOLOGY

This is a literature review study, based on scientific research platforms. For the literature survey, the electronic bibliographic scientific databases were consulted in the months from July 2020 to November 2021: PUBMED Portal, VHL and Google Academic.

2.1 BIBLIOGRAPHIC SURVEY

The first stage of the study consisted of defining the topic and the research question, namely: How can computational tools used to analyze the whole genome sequencing of M. tuberculosis identify the genetic relationship between strain mutations and antimicrobial resistance? In the second step, the following keywords were selected: "analysis"; "tools"; "whole genome"; "*Mycobacterium Tuberculosis*"; "resistance". The Boolean operator AND was used to cross-reference the keywords in the PUBMED platform, Google Scholar and the Virtual Health Library (VHL), as shown in table 1 below.

To search the articles, the following inclusion criteria were followed: articles published between 2014 and 2021, complete, available in electronic media, in Portuguese, English and/or Spanish, and that directly covered the subject. Studies that were not related to the theme and monographs, dissertations, theses, and legislation were excluded.

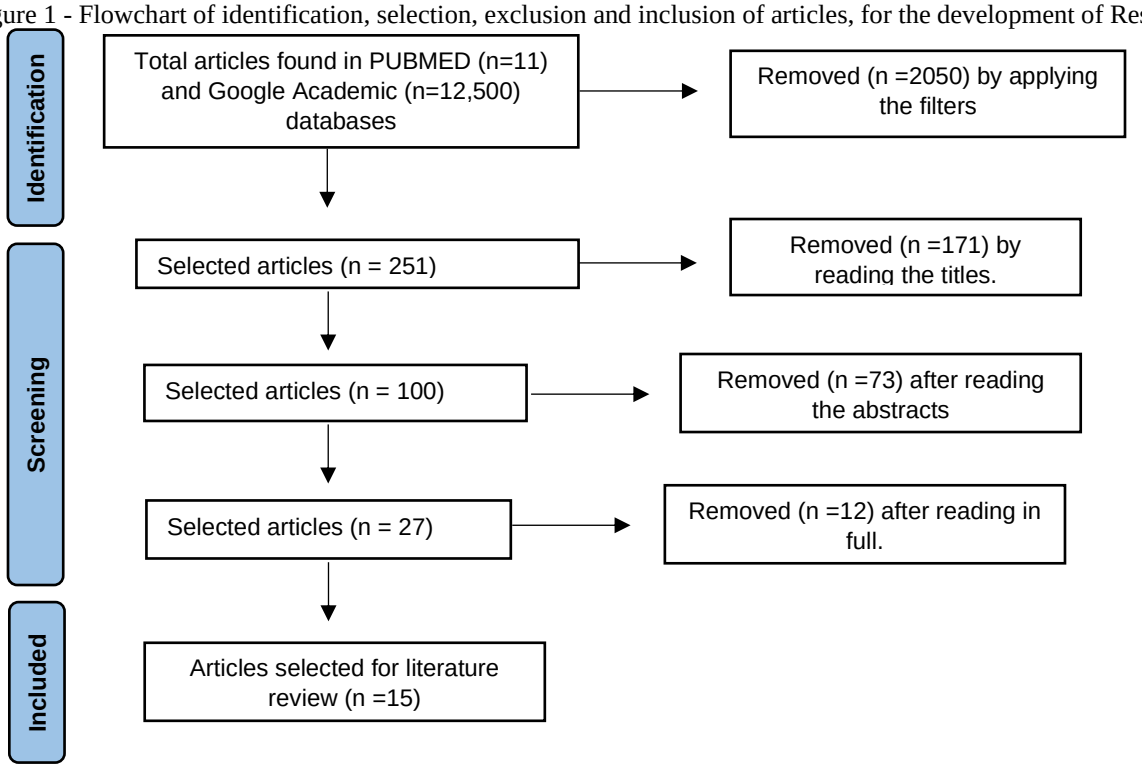
Table 1 - Cross-referencing of key words according to database and search period

Database/search keys	Any period	2017 - 2021
PubMed/ (((Drug Resistance) AND (Bacterial)) AND (Anti-Bacterial Agents)) AND (Genetic Phenomena)	34064	8785
Google Scholar/ (((Drug Resistance) AND (Bacterial)) AND (Anti-Bacterial Agents)) AND (Genetic Phenomena)	13200	5440
VHL/ (((Drug Resistance) AND (Bacterial)) AND (Anti-Bacterial Agents)) AND (Genetic Phenomena)	124	32

Source: Own elaboration (2022)

In PUBMED 34,064 articles were found, in Google Academic 13,200 were found, in BVS 124 studies were found, the filters of time were applied excluding the articles published in the period prior to 2017, types of research and articles published in full, after reading the title 87 articles were selected. After identifying the titles, the abstracts were read and the diagonal reading was performed, which would be a superficial analysis of the introduction and results of the research, resulting in 27 articles at the end of the reading. Finally, 15 of the selected articles were read in full, according to the flowchart, adapted from the PRISMA format for the preparation of systematic reviews, shown in **figure 01**:

Figure 1 - Flowchart of identification, selection, exclusion and inclusion of articles, for the development of Research.



Source: Own elaboration (2022).

In the end, the study sample consisted of 15 articles for analysis and discussion, only the articles that brought Complete Genome Sequencing within the methodologies or objectives were listed for preparation of the tables of presentation of results. After reading the studies in full, 8 articles were selected to compose the tool for analysis of results, the remaining 7 articles read in full helped in the discussion and preparation of the study.

A total of 08 (eight) articles were used to present the data, as shown in chart 02, where they were identified by title, year of publication, and authors.

Table 2- Articles included in the results according to article identification, title, authors' names, year of publication in the period from 2015 to 2020.

AUTHOR'S NAME	TITLE	YEAR
Chakravorty, S., Simmons, A. M., Rowneki, M., Parmar, H., Cao, Y., Ryan, J., ... & Alland, D.	The new Xpert MTB/RIF Ultra: improving detection of *Mycobacterium Tuberculosis* and resistance to rifampin in an assay suitable for point-of-care testing	2017
Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., ... & Van Rie, A	Whole genome sequencing of *Mycobacterium Tuberculosis*: current standards and open issues	2019
Munir, A., Vedithi, S. C., Chaplin, A. K., & Blundell, T. L.	Genomics, computational biology and drug discovery for mycobacterial infections: fighting the emergence of resistance	2020
Tunstall, T., Portelli, S., Phelan, J., Clark, T. G., Ascher, D. B., & Furnham, N	Combining structure and genomics to understand antimicrobial resistance	2020
Iketleng, T., Lessells, R., Dlamini, M. T., Mogashoa, T., Mupfumi, L., Moyo, S., ... & de Oliveira, T.	*Mycobacterium Tuberculosis* next-generation whole genome sequencing: opportunities and challenges	2018
Waman, V. P., Vedithi, S. C., Thomas, S. E., Bannerman, B. P., Munir, A., Skwark, M. J., ... & Blundell, T. L.	Mycobacterial genomics and structural bioinformatics: opportunities and challenges in drug discovery	2019
do Carmo Guimarães, J. D. L., von Groll, A., Unis, G., Dalla-Costa, E. R., Rossetti, M. L. R., Vianna, J. S., ... & Silva, P. E. A	Whole-genome sequencing as a tool for studying the microevolution of drug-resistant serial *Mycobacterium Tuberculosis* isolates	2021
Gomez-Gonzalez, P. J., Andreu, N., Phelan, J. E., de Sessions, P. F., Glynn, J. R., Crampin, A. C., ... & Clark, T. G.	An integrated whole genome analysis of *Mycobacterium Tuberculosis* reveals insights into relationship between its genome, transcriptome and methylome	2019

Source: Own elaboration (2022).

In the fourth step, the information to be extracted from the selected articles was listed. To gather and synthesize the key information of the study, an instrument was prepared that contained the following variables: title, country and year of publication, authors' names, objective, methodology, and main results.

In the fifth stage of the research, the results were reached after extraction and interpretation of the information obtained in the previous stage of the study. Finally, in the sixth stage, the

synthesis of the knowledge extracted about what has been published on Complete Genome Sequencing Analysis was presented.

The collected data were systematically analyzed through the final analysis of the articles, which were organized in two tables, aiming to capture a certain theme. Since this is a literature review using public domain articles, the present study does not require evaluation by the Research Ethics Committee.

3 RESULTS AND DISCUSSION

In 1952 the role of Deoxyribonucleic Acid (DNA) as a molecule that encodes genetic information was validated, it was DNA and not protein that was picked up and transmitted by bacterial cells infected by a bacteriophage. After the advent of bioinformatics and the recognition of the language of DNA as a genetic code, there have been advances in the field of molecular biology, especially with the advent of genome projects, which have generated intense volumes of information obtained from sequencing DNA fragments (CHAKRAVORTY, 2017; MUNIR, 2020).

The techniques of complete genome sequencing and their availability in public databases have been responsible for the exponential growth of computational tools to extract information about important characteristics encoded in the genomes of species of the Mycobacterium complex. Genomic analyses of M. tuberculosis can characterize the virulence and pathogenicity of different strains, contributing to the establishment of accurate epidemiological data on pathogenesis processes (IKETLENG, 2018; WAMAN, 2019; GÓMEZ-TANGARIFE, 2018).

The 08 selected productions brought Complete Genome Sequencing within the methodologies or objectives, are represented in table 03 and represent the synthesized sample according to the article identification, title, authors' names, objective, method and year of publication.

Table 3 - Summary of the articles included in the review according to the identification of the objective, method used in the research.

AUTHOR'S NAME	OBJECTIVE	METHOD
Chakravorty, S., Simmons, A. M., Rowneki, M., Parmar, H., Cao, Y., Ryan, J., & Alland, D.	Demonstrate the design features and operational characteristics of an enhanced Xpert Ultra trial	Frozen, prospectively collected clinical samples from patients with suspected TB, with and without culture-confirmed TB, were also tested.
Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A.,	We describe the current landscape of WGS pipelines and applications and establish	They did direct genome sequencing of some clinical specimens from tuberculosis patients. They analyzed the genomic variants of

Ezewudo, M., ... & Van Rie, A	best practices for *M. tuberculosis* WGS.	tuberculosis, including the clinical specimens, and to predict drug susceptibility profiles.
Munir, A., Vedithi, S. C., Chaplin, A. K., & Blundell, T. L.	Describe structure-guided approaches to understanding the impacts of mutations that give rise to antimycobacterial resistance and the use of this information in new drug design.	They evaluated the tools raised by researchers in a previous paper, which provided an update on recent developments in the TB drug development pipeline (including new and repurposed antimicrobials and host-targeted drugs) as they are applied to new regimens to shorten and improve TB treatment outcomes.
Tunstall, T., Portelli, S., Phelan, J., Clark, T. G., Ascher, D. B., & Furnham, N	Present several of the major computational tools and methods currently available for measuring mutational consequences, with a focus on those tools that have been used to analyze variation within a pathogen genome.	They presented some sequence and structure based tools that predict the effect of pathogen mutations. In the methodological pathway they used the table with an updated list of currently available tools. The type of method for each tool is specified using the following code; **S**: sequence-based, **St:** structure-based, **SA: sequence** alignment, **SS**: sequence and structure, **(St)**: structure if available
Iketleng, T., Lessells, R., Dlamini, M. T., Mogashoa, T., Mupfumi, L., Moyo, S., ... & de Oliveira, T.	Discuss the tremendous opportunities that next generation WGS presents in terms of understanding the molecular epidemiology of tuberculosis and the mechanisms of drug resistance.	They presented the current challenges for implementing WGS in low- and middle-income settings. Used molecular methods such as Xpert MTB/RIF
Waman, V. P., Vedithi, S. C., Thomas, S. E., Bannerman, B. P., Munir, A., Skwark, M. J., ... & Blundell, T. L.	Describe the impact of the rapid expansion of genome sequencing and genome/pathway annotations that have greatly enhanced the progress of structure-guided drug discovery.	They used rapid annotation of genomic data, leading to the development of general-purpose and specialized resources, providing important and pertinent information on sequence, structure, function, metabolic pathway, taxonomy, and drug resistance mutations.
do Carmo Guimarães, J. D. L., von Groll, A., Unis, G., Dalla-Costa, E. R., Rossetti, M. L. R., Vianna, J. S., ... & Silva, P. E. A	Describe the microevolution of drug resistance in serial isolates from six previously treated patients.	The research had 6 participants, they performed samples from the six patients, culture and DNA extraction, then performed drug sensitivity testing such as catalase assays, drug resistance was initially investigated using phenotypic methods, followed by genotypic approaches and Genotyping of the strains.
Gomez-Gonzalez, P. J., Andreu, N., Phelan, J. E., de Sessions, P. F., Glynn, J. R., Crampin, A. C., ... & Clark, T. G.	Integrate sequence-based genomic, transcriptomic and methylation characterization in a common sample set	*Mtb* was isolated from 22 sputum samples from 22 different tuberculosis patients collected between 2003 and 2009 in Karonga, a district in northern Malawi. Most of the individuals were HIV positive (16/22). Genomic DNA was extracted and sequenced using PacBio single molecule real-time sequencing (SMRT) and Illumina technologies. *Mtb isolates* were grown by liquid culture (in the absence of antimicrobial drugs) from frozen Lowenstein-Jensen stocks or liquid cultures derived from already isolated patient sputum samples. The samples were sequenced at the Genome Institute of Singapore.

Source: Own elaboration (2022).

The different methods of analysis of the samples used in the studies found demonstrated how the complete genome sequencing of *Mycobacterium Tuberculosis* has progressed rapidly from a research tool to a tool for clinical application that is applicable both for the diagnosis and treatment of tuberculosis as well as for public health surveillance.

Chart 02 shows the main results and discussion of the studies on Complete Sequencing of the *Mycobacterium Tuberculosis* Genome and the identification of Antimicrobial Resistance Mechanisms.

Table 04- Summary of the results presented in the journals published from 2015 to 2020.

AUTHOR	RESULTS AND DISCUSSION
Chakravorty, S., Simmons, A. M., Rowneki, M., Parmar, H., Cao, Y., Ryan, J., ... & Alland, D.	Ultra and Xpert limits of detection (LOD), dynamic ranges and RIF-R rpoB mutation detection were tested on DNA or sputum samples of *Mycobacterium Tuberculosis* with known numbers of M. tuberculosis H37Rv or Mycobacterium bovis BCG CFU. All M. tuberculosis rpoB associated with RIF-R mutations tested were identified by Ultra. Testing in clinical sputum samples, Ultra versus Xpert, resulted in an overall sensitivity of 87.5% (95% confidence interval [CI], 82.1/91.7) versus 81.0% (95% CI, 74.9/86.2) and a sensitivity in sputum smear-negative samples of 78.9% (95% CI, 70.0/86.1) versus 66.1% (95% CI, 56.4/74.9). Both tests had a specificity of 98.7% (95% CI, 93.0/100), and both had comparable accuracies for detecting RIF-R in these samples.
Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., ... & Van Rie, A	Whole genome sequencing (WGS) of *Mycobacterium Tuberculosis* has progressed rapidly from a research tool to a clinical application for the diagnosis and treatment of tuberculosis and in public health surveillance. This development has been facilitated by dramatic declines in costs, advances in technology, and concerted efforts to translate sequencing data into actionable information. There is a risk, however, that in the absence of international consensus and standards, widespread use of WGS technology may result in data and processes that lack harmonization, comparability, and validation.
Munir, A., Vedithi, S. C., Chaplin, A. K., & Blundell, T. L.	They discussed the application of computational tools and experimental approaches in the context of mycobacterial drug discovery and antimicrobial drug resistance with emphasis on tuberculosis and leprosy. They showed that data generated by whole genome sequencing of clinical isolates can be screened for the presence of drug-resistant mutations. A preliminary *in silico* analysis of mutations can then be used to prioritize experimental work to identify the nature of these mutations.
Tunstall, T., Portelli, S., Phelan, J., Clark, T. G., Ascher, D. B., & Furnham, N	Computational tools can quickly and inexpensively assess the effect of mutations on protein function and evolution. Subsequent insights can then inform experimental studies and direct existing or new computational methods. Combining genomic results with the biophysical effects of mutations can help reveal the molecular basis and consequences of resistance development.
Iketleng, T., Lessells, R., Dlamini, M. T., Mogashoa, T., Mupfumi, L., Moyo, S., ... & de Oliveira, T.	Direct sequencing of sputum samples without the need for culture would provide a more accurate picture of the population structure of mixed infections. The relative representation of different strains in mixed infections can be captured without overgrowth of some strains over others due to favorable culture conditions. This would better inform treatment and prevention interventions.
Waman, V. P., Vedithi, S. C., Thomas, S. E., Bannerman, B. P., Munir,	Early approaches to new antibiotic discovery relied almost entirely on phenotypic whole-cell screening of natural products, microbial extracts, and fermentation broths. Structural features can also be used to further refine target

A., Skwark, M. J., ... & Blundell, T. L.	selection and validation, including lack of structural homology with the human host to avoid mechanism-based toxicity and ligability leading to modulation of target activity.
do Carmo Guimarães, J. D. L., von Groll, A., Unis, G., Dalla-Costa, E. R., Rossetti, M. L. R., Vianna, J. S., ... & Silva, P. E. A	The researchers pointed out that the use of whole genome sequencing allowed the identification of mutations in the *katG* , *rpsL* and *rpoB genes* associated with drug resistance, including the detection of rare mutations *inkatG* and mixed strain populations. Molecular fitting simulation studies of the impact of observed mutations on isoniazid binding were also performed. Whole genome sequencing detected 266 single nucleotide polymorphisms between two isolates obtained from one patient, suggesting a case of exogenous reinfection.
Gomez-Gonzalez, P. J., Andreu, N., Phelan, J. E., de Sessions, P. F., Glynn, J. R., Crampin, A. C., ... & Clark, T. G.	The research showed that for each isolate, the raw sequence data was aligned to the H37Rv reference genome, leading to an average coverage of >100-fold. Across all samples, 9,384 unique SNPs were characterized, with ~40% of them identified in single isolates. Only 1,446 of the 9,384 SNPs were located in intergenic regions.

Source: Own elaboration (2022).

The study by Chakravorty (2017), brought the first demonstration of the design features and operational characteristics of an assay called Xpert Ultra enhanced. He evidenced in the findings excellent sensitivity in testing sputum specimens with positive sputum smears, however, he demonstrated that Xpert is slightly less sensitive when performed on sputum specimens with negative sputum smears. Ultra should significantly improve TB detection, especially in patients with paucibacillary disease, and may provide more reliable RIF-R detection. The sensitivity of the assay was also limited in results using extrapulmonary specimen types, which are known to have considered lower levels of bacilli compared to pulmonary specimens.

In the article by Iketleng (2018), the authors pointed out that the future of WGS of *M. tuberculosis* lies in the ability to apply the method directly to sputum, as this is the most commonly available clinical material in laboratories that perform strain sequencing. Waman (2019) brought that computational approaches to predict the effects of mutations on protein structure and function can be useful in understanding the mechanism of drug resistance.

For Gomes (2019) the study of lineage-specific transcriptomic profiles and the mechanisms that regulate gene expression in *M. tuberculosis* strains may provide information about the mechanisms underlying these biological differences. Such mechanisms will be useful for identifying potential targets for the development of new drugs used in the treatment of Tuberculosis or anti-tuberculosis vaccines.

The results scored that the data generated by whole genome sequencing of species isolated from clinical cases can serve as a screener for possible drug-resistant mutations. This combination of genome read results with the biophysical effects of existing mutations in strains can help investigate the molecular basis and consequences of the development of resistance to single drugs or even resistance to multiple drugs.

All studies concluded that the technologies applied for genome sequencing can detect rare point mutations related to resistance to drugs used in treatment, as well as identify subpopulations of resistant strains and analyze various causes related to exogenous reinfection and multiple resistance, contributing to the control of TB treatment, guiding the early implementation of appropriate and effective clinical and therapeutic interventions in various situations.

Genotyping methods can contribute to the differentiation and understanding of the transmission chains of the M. tuberculosis complex, as well as help in addressing clinical issues such as virulence and pathogenicity factors. However, classical genome sequencing methods are limited only to a set of polymorphic regions contained in genome fragments (CHAKRAVORTY, 2017).

3.1 COMPLETE GENOME SEQUENCING OF *MYCOBACTERIUM TUBERCULOSIS* AND THE IDENTIFICATION OF ANTIMICROBIAL RESISTANCE MECHANISMS

In the year 1998 the genome sequence of the Mtb H37Rv laboratory strain was published and revised in 2002 and 2010. Whole Genome Sequencing (WGS) approaches use DNA sequencing platforms to reconstruct the complete DNA sequence of an organism's genome. WGS-based approaches are rapidly changing research laboratory methods to clinical care and public health applications. The World Health Organization has been using WGS for drug resistance surveillance (DO CARMO GUIMARÃES, 2021; GOMEZ-GONZALEZ, 2019).

The standard workflow for WGS analysis of *Mycobacterium Tuberculosis* (Mtb) strains involves culture of Lowenstein sputum samples or liquid DNA extraction from cells, library preparation, and sequencing using short read technologies. Genomic data for a set of isolates can also be used for surveillance and transmission investigations (TUNSTALL, 2020; AMBROSETTI, 2020; DO CARMO GUIMARÃES, 2021).

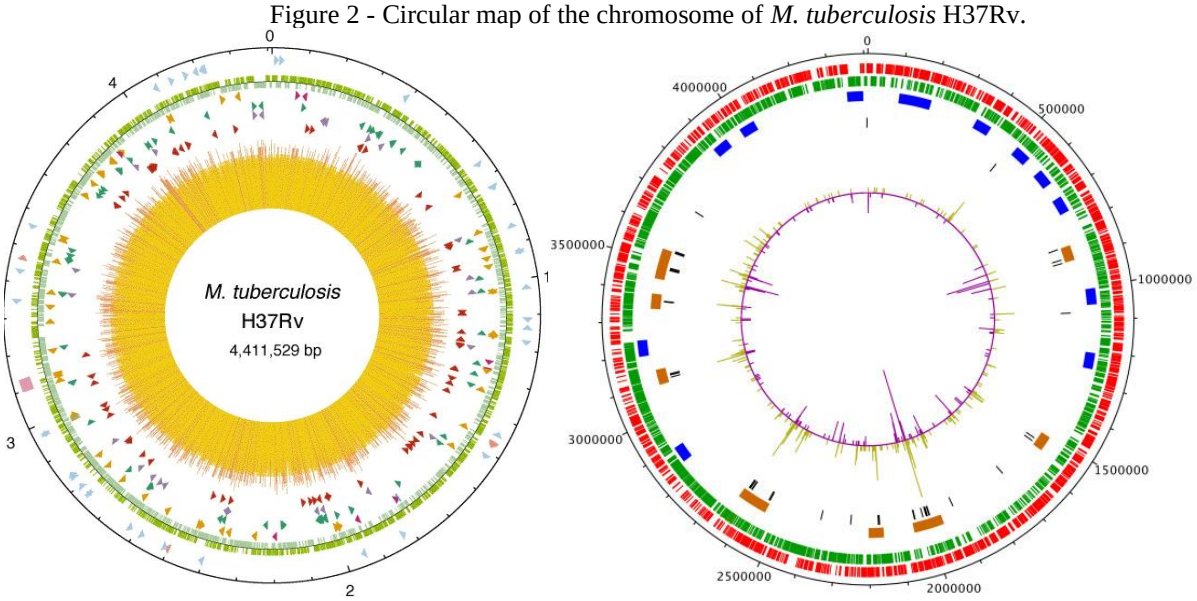
The clinical sample (usually sputum) is first cultured for up to 6 weeks, followed by genomic DNA extraction and sequencing. The resulting sequencing results can be deposited online in public repositories and also performed via *pipelines*. Segmentation of elements of these *pipelines* occurs initially with reads of a reference gene (often *M. tuberculosis* strain H37Rv) and then the genomic variants. Lists of the genomic sequencing reads can then be used for a variety of analyses, such as strain typing, transmission clustering, and drug resistance profiling (TUNSTALL, 2020).

M. tuberculosis strain H37Rv has recently been sequenced by several research centers. Continuously updated annotations and genome-derived information about H37Rv can be found on

the TubercuList and Mycobrowser servers and in the TB database. The first protein structure from an *M. tuberculosis* deposited in the Protein Data Bank (PDB) was that of iron-dependent superoxide dismutase. It was followed by the structure of the enoyl-acyl carrier protein (ACP) reductase InhA, for which 80 structures are now available. At the time the genome sequence of Mtb strain H37Rv was first published, there were only eight Mtb protein structures in the PDB (TUNSTALL, 2020).

In the year 2018, a total of 2124 Mtb structures were deposited in the PDB, representing 585 unique structures, the remaining structures mostly correspond in different complexes of the single structures with small molecules. The programs used by structural biology and genomics are designed to establish a *continuum* between individual sequences, structures and functions, thus leading to a better understanding of Mtb promoting rational structure-based design of new TB drugs. Several structural bioinformatics initiatives have also been developed to increase the structural coverage of the Mtb proteome (DOMÍNGUEZ, 2018).

Figure 2 - Circular map of the chromosome of *M. tuberculosis* H37Rv.



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

The genome of the M. tuberculosis H37Rv strain has 4,411,529 base pairs in its composition. Its content is relatively constant at 65.6% guanine and cytosine throughout the DNA strand. The complex of bacteria belonging to the Mycobacteria group presents in a certain way, a likelihood when analyzed by gene sequencing by 16S rRNA, the differences are noticeable when comparing the morphological characteristics and pathogenicity in the host (TUNSTALL, 2020; IKETLENG, 2018; WAMAN, 2019; DO CARMO GUIMARÃES,2021).

M. tuberculosis is characterized as a human pathogenic bacterium presented in the form of bacilli, with morphology presented as follows: diameter between 0.3 and 0.5 µm, and a very impermeable cell wall. Unlike other species of bacteria, Mycobacteria cannot be classified as Gram positive or negative organisms, due to the characteristics of their cell wall, the Gram staining method is ineffective, leading these bacteria to be classified as alcohol acid resistant bacteria by the Ziehl-Neelsen technique (GÓMEZ-TANGARIFE, 2018).

The complex cell wall of the Mycobacteria shows similar characteristics to the two groups of bacteria classified as Gram positive and negative, while presenting a lipid rich structure, it is unable to retain Gram staining. Features similar to Gram-negative individuals are the presence of porins in the outer lipid layer and a periplasmic space between the inner membrane and the peptideoglycan layer (GÓMEZ-TANGARIFE, 2018).

M. tuberculosis is an aerobic organism that efficiently performs its biological functions in oxygen-rich tissues, which explains its affinity for lung tissue; however, they can survive in tissues with reduced amounts of oxygen. However, the resistance patterns related to the bacterial genome are not affected by the concentrations of the corresponding antibiotics, in the blood or tissue, oxygen availability, or even those that have resistance mechanisms specific to the studied agent to which there was not an adequate clinical response when used as treatment (WAMAN, 2019).

Gómez-Tangarife (2018) points out that with the advancement in technologies responsible for whole genome sequencing it was possible to identify mutations that would possibly influence the occurrence of resistance and raise possibilities of selecting a specific target such as a gene or protein, which presents some property or characteristic that is essential for the growth and/or survival of the bacteria under the conditions of the infectious process, making it possible to identify the target to inhibit the bacterial activity and/or response.

Identifying mutations in genes through whole genome genotyping can qualify and show constitutive and drug resistance genes. By identifying strong predictors of resistance mechanisms through genome sequencing the drug can be selective, to the point where a compound recognizes effective bacterial proteins for an effective pharmacological response (GÓMEZ-TANGARIFE, 2018).

However, the availability of genome sequences from different strains has made it possible to compare the entire genome, variations in gene copy number, variation in DNA sequence, insertions or deletions of multiple nucleotides, and point mutations known as single nucleotide polymorphism can occur (CHAKRAVORTY, 2017).

Tunstall (2020) says that the process of genome sequencing analysis occurs by aligning the DNA sequence with a reference genome, usually the M. tuberculosis strain H37Rv. Sequence-based analysis methods rely only on the analyzed gene or protein sequence of the strain, they are often used when there is no known protein structure or when modeling is not possible.

The predictions made by sequence analysis tools are generally based on alignments of the predicted secondary structures. Sequence alignment methods reflect the view of two models the global alignment/similarity that considers similarity along the entire length of the sequences and the local similarity alignment that constitutes a fraction of the length of the sequences. Local alignment methods search for local similarities in short regions of the complete sequence leading to more biologically meaningful and specified results, different than those obtained by alignment along the entire sequence length (CHAKRAVORTY, 2017).

Chakravorty (2017), brings that one of the biggest challenges of bioinformatics applied to genomics is the complete analysis of genomes, i.e., it is necessary to improve the identification of computationally predicted genes and associate them with a function and provide subsidies for the design of experiments that can test these predictions and compare them with others that already exist. There are several programs that perform the genome assembly process: Phred/Phrap/Consed, CAP/PCAP, Celera Assembler, *Genome Analyzer* (Illumina) and the *GS De Novo Assembler* (Roche). Each program uses different algorithms to obtain the contiguous DNA sequence.

Tunstall (2020) further states that structure-based methods can provide a three-dimensional presentation of the molecules involved in the process of mutations, which may not be evident by sequence analysis alone. Structure-based methods include the analysis of the protein's structural and functional consequences of mutations, including those on coiling, stability, dynamics, and changes in interactions with ligands.

After complete genome sequencing, it is necessary to identify all genes encoding proteins (proteome) and the function of the encoded proteins by searching for similarities in databases. The selection of bioinformatics software for the analysis of genome sequencing data will be determined by the objective of the study. The study by Gomes (2019), used Geneious 9.1.8 (Biomatters), a desktop software that analyzes sequence data. Geneious according to the authors provides an intuitive and easy-to-use interface that transforms raw sequence data into meaningful visualizations, and can be used to map FASTQ sequence reads to a reference genome available in Bioinformatics databases. FASTQ is the text format used to store a DNA sequence and its corresponding quality indices.

Whole genome sequencing is considered a surveillance marker of transmission and drug resistance, as well as, a strong predictor of the emergence of possible resistance mechanisms for first- and second-line drugs used in the treatment of Pulmonary Tuberculosis (TB) (WAMAN, 2019; GOMEZ-GONZALEZ, 2019).

Several studies on a global scale have been showing targets in genes involved in the phenotypic resistance of Koch's Bacillus, also identifying mutations that confer different virulence and pathogenesis. Researchers point to whole genome sequencing as the standard method to detect transmission chains in real time. This sequencing can facilitate the identification of resistant strains, bringing new molecular knowledge of resistance, enabling effectiveness for TB control programs (WAMAN, 2019; GOMEZ-GONZALEZ, 2019).

With the genomic analyses of M. tuberculosis, it was possible to identify genes and genomic regions that are involved in the development of different types of resistance. For example, resistance to Isoniazid may be related to katG (catalase-peroxidase), inhA (enoyl acyl reductase), ahpC (alkyl hydroxyperoxide reductase) and, more recently, the genes kasA (ketoacyl acyl synthetase) and ndh (NADH dehydrogenase) (WAMAN, 2019; GOMEZ-GONZALEZ, 2019).

4 CONCLUSION

The study of genome pairs or multiple genomes contributes to a variety of studies, which encompasses the elucidation of basic questions involved in the evolutionary biology of microorganisms and can be used to study very specific clinical issues, such as the identification of genetic polymorphisms of disease-causing bacteria.

The use of bioinformatics analysis tools is one of the relevant factors when discussing the complete genome sequencing of M. tuberculosis, since, in addition to contributing to optimal clinical use, it can solve problems within the analysis of mutations and gene expression. Understanding the emergence of resistance of these bacteria to the drugs that treat M. tuberculosis has been facilitated by the availability of complete genome sequencing in recent years.

The analysis process that begins with the alignment of DNA sequence data to a genome reference, usually the M. tuberculosis strain H37Rv, can increase the reliability of identifying new genotypes, contributing to the prediction of strain resistance. The data generated by whole genome sequencing of clinical isolates can be used as tools to track the presence of mutations resistant to various types of drugs. There are several contributions of genomic discoveries and phenotypic approach when it comes to the identification of new targets for possible candidates for anti-TB drugs and vaccines.

Understanding about the emergence of acquired resistance to the drugs that treat M. tuberculosis has been facilitated by the availability of whole genome sequencing in recent years. The data generated by whole genome sequencing of clinical isolates can be used as tools to screen for the presence of mutations resistant to various types of drug associated or isolated drugs.

REFERENCES

CHAKRAVORTY, Soumitesh et al. The new Xpert MTB/RIF Ultra: improving detection of *Mycobacterium Tuberculosis* and resistance to rifampin in an assay suitable for point-of-care testing. **MBio**, v. 8, n. 4, p. e00812-17, 2017.

DA COSTA, Anderson Luiz Pena; JUNIOR, Antonio Carlos Souza Silva. Resistência bacteriana aos antibióticos e Saúde Pública: uma breve revisão de literatura. **Estação Científica (UNIFAP)**, v. 7, n. 2, p. 45-57, 2017.

DOMÍNGUEZ, José; LATORRE, Irene; SANTIN, Miguel. Diagnosis and therapeutic approach of latent tuberculosis infection. **Enfermedades infecciosas y microbiologia clinica (English ed.)**, v. 36, n. 5, p. 302-311, 2018.

LIMA, Camila Correa; BENJAMIM, Sandra Cristina Calixto; SANTOS, Rosana Francisco Siqueira dos. Mecanismo de resistência bacteriana frente aos fármacos: uma revisão. **CuidArte, Enferm**, p. 105-113, 2017.

MEEHAN, Conor J. et al. Whole genome sequencing of *Mycobacterium Tuberculosis*: current standards and open issues. **Nature Reviews Microbiology**, v. 17, n. 9, p. 533-545, 2019.

MUNIR, Asma et al. Genomics, Computational Biology and Drug Discovery for Mycobacterial Infections: Fighting the Emergence of Resistance. **Frontiers in Genetics**, v. 11, p. 965, 2020.

PEDELACQ, Jean-Denis et al. A Comprehensive Review on *Mycobacterium Tuberculosis* Targets and Drug Development from a Structural Perspective. **Structural Biology in Drug Discovery: Methods, Techniques, and Practices**, p. 545-566, 2020.

TUNSTALL, Tanushree et al. Combining structure and genomics to understand antimicrobial resistance. **Computational and Structural Biotechnology Journal**, 2020.

IKETLENG, Thato et al. *Mycobacterium Tuberculosis* next-generation whole genome sequencing: opportunities and challenges. **Tuberculosis research and treatment**, v. 2018, 2018.

PÉREZ-LAGO, Laura et al. Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium Tuberculosis*: potential impact on the inference of tuberculosis transmission. **The Journal of infectious diseases**, v. 209, n. 1, p. 98-108, 2014.

AMBROSETTI, Francesco et al. Modeling antibody-antigen complexes by information-driven docking. **Structure**, v. 28, n. 1, p. 119-129. e2, 2020.

WAMAN, Vaishali P. et al. Mycobacterial genomics and structural bioinformatics: opportunities and challenges in drug discovery. **Emerging microbes & infections**, v. 8, n. 1, p. 109-118, 2019.

GÓMEZ-TANGARIFE, Verónica J. et al. Resistencia a Medicamentos en *Mycobacterium Tuberculosis*: contribución de mecanismos constitutivos y adquiridos. **Revista de Salud Pública**, v. 20, p. 491-497, 2018.

DO CARMO GUIMARÃES, Jaciara de Lourdes et al. Whole-genome sequencing as a tool for studying the microevolution of drug-resistant serial *Mycobacterium Tuberculosis* isolates. **Tuberculosis**, v. 131, p. 102137, 2021.

GOMEZ-GONZALEZ, Paula J. et al. An integrated whole genome analysis of *Mycobacterium Tuberculosis* reveals insights into relationship between its genome, transcriptome and methylome. **Scientific reports**, v. 9, n. 1, p. 1-11, 2019.