

Exploring Antimicrobial Resistance in the Lung Microbial Community Using Unsupervised Machine Learning

Authors : Camilo Cerda Sarabia, Fernanda Bravo Cornejo, Diego Santibanez Oyarce, Hugo Osses Prado, Esteban Gómez Terán, Belén Díaz Díaz, Raúl Caulier-Cisterna, Jorge Vergara-Quezada, Ana Moya-Beltrán

Abstract : Antimicrobial resistance (AMR) represents a significant and rapidly escalating global health threat. Projections estimate that by 2050, AMR infections could claim up to 10 million lives annually. Respiratory infections, in particular, pose a severe risk not only to individual patients but also to the broader public health system. Despite the alarming rise in resistant respiratory infections, AMR within the lung microbiome (microbial community) remains underexplored and poorly characterized. The lungs, as a complex and dynamic microbial environment, host diverse communities of microorganisms whose interactions and resistance mechanisms are not fully understood. Unlike studies that focus on individual genomes, analyzing the entire microbiome provides a comprehensive perspective on microbial interactions, resistance gene transfer, and community dynamics, which are crucial for understanding AMR. However, this holistic approach introduces significant computational challenges and exposes the limitations of traditional analytical methods such as the difficulty of identifying the AMR. Machine learning has emerged as a powerful tool to overcome these challenges, offering the ability to analyze complex genomic data and uncover novel insights into AMR that might be overlooked by conventional approaches. This study investigates microbial resistance within the lung microbiome using unsupervised machine learning approaches to uncover resistance patterns and potential clinical associations. It downloaded and selected lung microbiome data from HumanMetagenomeDB based on metadata characteristics such as relevant clinical information, patient demographics, environmental factors, and sample collection methods. The metadata was further complemented by details on antibiotic usage, disease status, and other relevant descriptions. The sequencing data underwent stringent quality control, followed by a functional profiling focus on identifying resistance genes through specialized databases like Antibiotic Resistance Database (CARD) which contains sequences of AMR gene sequence and resistance profiles. Subsequent analyses employed unsupervised machine learning techniques to unravel the structure and diversity of resistomes in the microbial community. Some of the methods employed were clustering methods such as K-Means and Hierarchical Clustering enabled the identification of sample groups based on their resistance gene profiles. The work was implemented in python, leveraging a range of libraries such as biopython for biological sequence manipulation, NumPy for numerical operations, Scikit-learn for machine learning, Matplotlib for data visualization and Pandas for data manipulation. The findings from this study provide insights into the distribution and dynamics of antimicrobial resistance within the lung microbiome. By leveraging unsupervised machine learning, we identified novel resistance patterns and potential drivers within the microbial community.

Keywords : antibiotic resistance, microbial community, unsupervised machine learning., sequences of AMR gene

Conference Title : ICMCSSE 2025 : International Conference on Mathematical, Computational and Statistical Sciences and Engineering

Conference Location : Madrid, Spain

Conference Dates : March 17-18, 2025