Computational Analysis of N-Acetyltransferase-Mediated Antimicrobial Resistance in Escherichia coli

Ansh Nitin Agarwal[1], Shreosee Ghosh[2], Rupak Roy[3]
[1]*Cygnus International School, Vadodara, Gujarat.*
[2]*SHRM Biotechnologies Pvt. Ltd., Kolkata, West Bengal.*

***Abstract*—Antimicrobial Resistance (AMR) is an extensively studied field of biology that aims to identify, justify, and eliminate bacterial organisms' antibiotic-resistant properties, which allows them to survive for a protracted period. This study aims to detect the AMR properties in bacterial organisms using computational tools and analysis. Specifically, the research studies *Escherichia coli,* a Gram-negative bacterium that exhibits both intrinsic and acquired AMR traits, such as the outer cell membrane and the β-lactamase, respectively.**

This study analysed the N-acetyltransferase gene from *E. coli*, which was obtained from the official National Center for Biotechnology Information (NCBI) website. This gene was compared to 3 other genes that code for N-acetyltransferase in 3 different bacterial species: *Staphylococcus aureus, Streptococcus pneumoniae, and Mycobacterium tuberculosis.* Clustal Omega was used to conduct this Multiple Sequence Alignment (MSA). Next, the study uses the Basic Local Alignment Search Tool protein (BLASTp) computational tool to identify proteins that are highly aligned to the N-acetyltransferase gene found in *E. coli*. Finally, the study uses the Resistance Gene Identifier (RGI) on CARD to discern the resistance properties and genes that *E. coli* possesses. The accession number used to conduct this computational analysis was HQ451074.1.

The results obtained from the computational analysis and research methodology provide insights into the AMR properties of *E. coli* and their similarity to those of other bacterial organisms. Multiple sequence alignments in the N-acetyltransferase genes from the 4 organisms highlight the gene's sequence conservation. BLASTp analysis of the N-acetyltransferase protein from *E.coli* identifies a highly similar protein sequence in *Acinetobacter baumannii*, indicating the presence of a homologous gene encoding a protein with conserved amino acid sequence (Bitscore: 42, E- value: 4.3e-05). This similarity depicts the potential of a possible horizontal gene transfer, which led to the propagation of microbial resistance properties amongst bacterial organisms during evolution.

Finally, RGI explains the resistance properties of *E. coli*. Using pictorial representations and a summary table, the tool explains and presents *E. coli's* resistance genes and properties that help it to sustain antibiotic threats. Specifically, the tool detected 7 perfect and 1strict RGI criteria hits.

***Index Terms*—Antimicrobial Resistance (AMR); Escherichia coli; N-acetyltransferase; Multiple Sequence Alignment (MSA); BLASTp; Resistance Gene Identifier (RGI); Horizontal Gene Transfer; CARD database.**

I. INTRODUCTION

Antimicrobial Resistance (AMR) is a major problem in clinical medicine, which allows pathogens to evade antibiotics and other forms of medication. Although several advancements have been made to combat AMR, such as gene databases, the high mutation rates in pathogenic organisms result in the emergence of resistance genes that are not known to microbial biologists and researchers.

Hence, this study aims to investigate the role of N-acetyltransferase in antimicrobial resistance in *Escherichia coli* and compare its similarity across different bacterial species. This research uses the CARD database and computational tools such as the Basic Local Alignment Search Tool (BLASTp), Multiple Sequence Alignment (MSA), and the Resistance Gene Identifier (RGI) to answer the research question.

E. coli is a gram-negative, facultative anaerobic, rod-shaped, coliform bacterium of the genus *Escherichia* that is commonly found in the lower intestine of warm-blooded organisms. These bacteria are mostly harmless or even beneficial to humans.

E. coli has a unique property of intrinsic, or natural, resistance. Due to its cell structure and physiology, *E. coli* can naturally defend itself against antibiotics such as Penicillin G, Vancomycin, and Macrolides. Subsequently, the organism also possesses Antimicrobial Resistance Genes (ARGs), such as β-lactam resistance genes, which confer additional resistance properties.

This research uses the N-acetyltransferase enzyme to study and compare the antimicrobial resistance properties of *E. coli* with those of other bacterial organisms. Specifically, the N- acetyltransferase transfers an acetyl group ($–COCH_3$) from acetyl-CoA onto a drug molecule, thus reducing its antimicrobial efficacy, benefiting the bacterial molecule's survival.

II. METHODOLOGY

This research made use of the following bioinformatic tools to understand and link *E. coli*'s DNA to various organisms:

- Multiple Sequence Alignment (MSA)
- Basic Local Alignment Search Tool (BLAST)
- Resistance Gene Identifier (RGI)

The gene sequence of the N-acetyltransferase for all five bacterial organisms, including E.coli, was extracted from published data on the National Centre for Biotechnology Information (NCBI) website.

The genomic sequences were obtained by filtering the database to show only proteins. The N-acetyltransferase of each organism was obtained using the search syntax shown in the image given below. All genomic sequences were the designated protein's FASTA sequence.

Figure 1: NCBI protein database search syntax used to retrieve FASTA sequences for the N- acetyltransferase enzyme

1. Multiple Sequence Alignment

Multiple sequence alignment, or MSA, was used to identify similarities between genes from different organisms that performed similar functions. The genes were procured from the National Centre for Biotechnology Information (NCBI) website. The genes included the N-acetyltransferase enzyme from *E. coli*, *S. aureus*, *S. pneumoniae*, and *M. tuberculosis*. The Clustal Omega online software was used to conduct the MSA process.

2. Basic Local Alignment Search Tool

The Basic Local Alignment Search Tool, or BLAST, is used to find genes that are similar to each other. The sample gene used was the N-acetyltransferase protein of *E. coli*. The process was conducted using the CARD, and BLASTp (BLAST protein) was used.

3. Resistance Gene Identifier

The Resistance Gene Identifier, or RGI, was used to obtain the list of genes that enabled *E. coli*'s antimicrobial resistance. The process was conducted using CARD, and the accession number used for RGI was HQ451074.1. This precise accession number was used due to its organism relevance and annotational number.

III. RESULTS

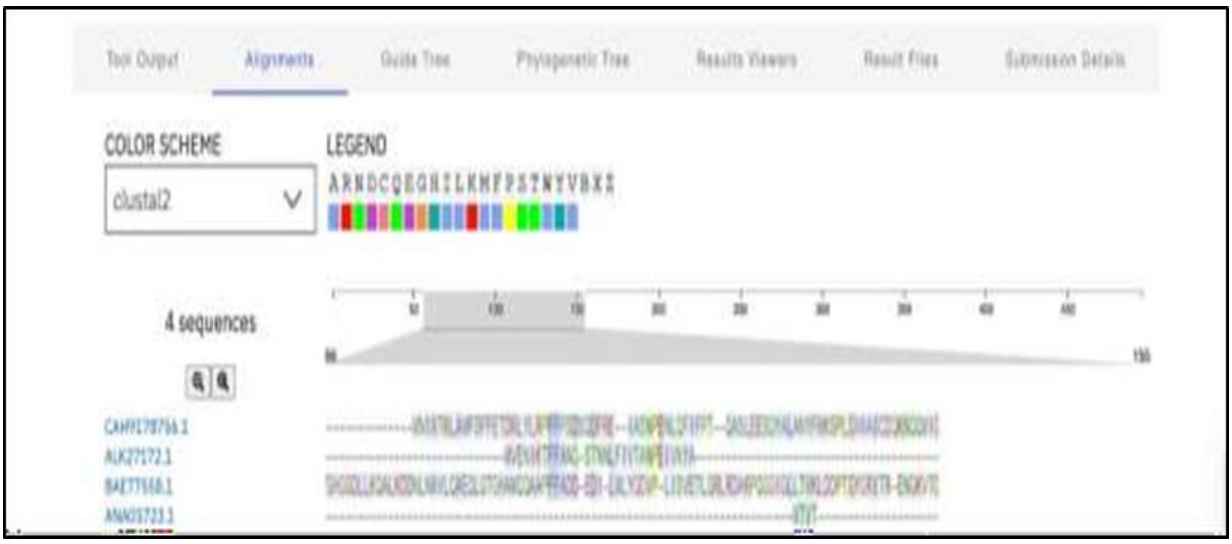

Figure 2: Multiple Sequence Alignment (MSA) of N-acetyltransferase protein sequences from E. coli, S. aureus, S. pneumoniae, and M. tuberculosis using Clustal Omega.

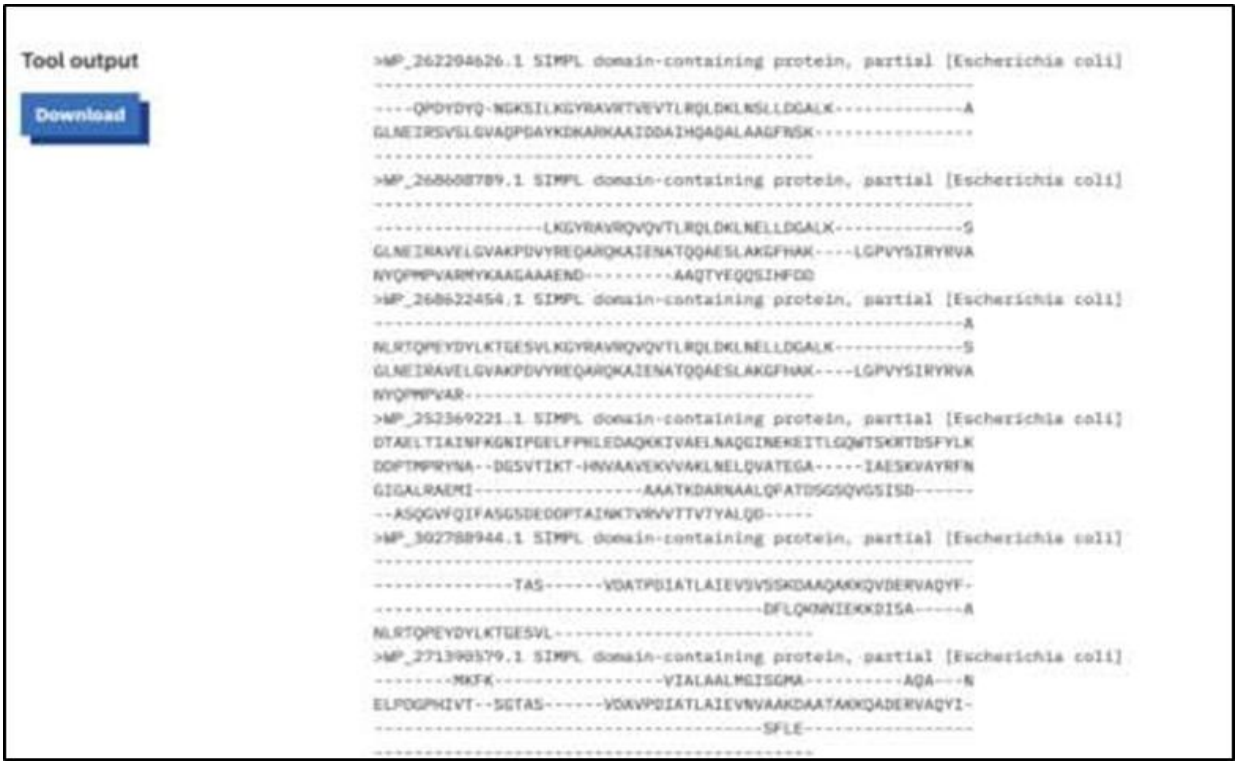

Figure 3: Visual representation of multiple sequence alignment across selected bacterial species, highlighting identical amino acid positions (e.g., Methionine) and gaps signifying non-alignment.

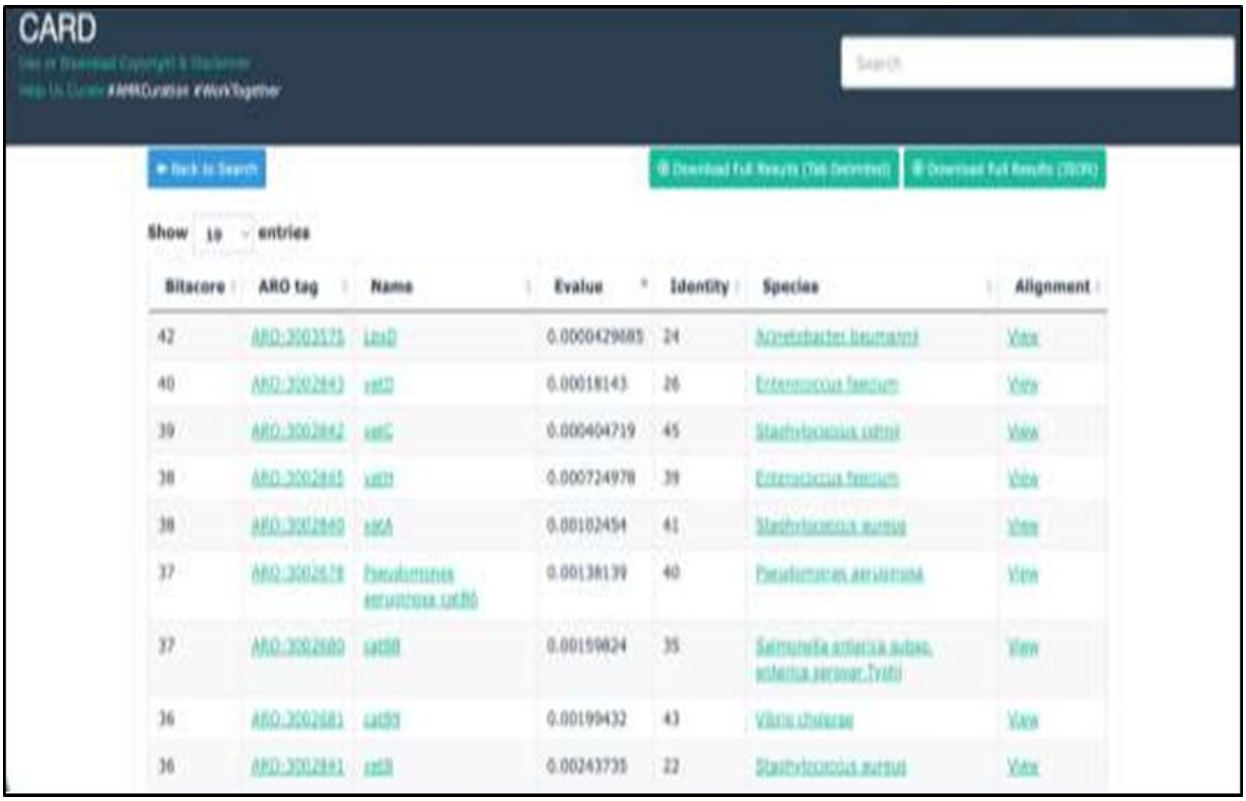

Figure 4: BLASTp results for E. coli N-acetyltransferase, showing high sequence similarity with Acinetobacter baumannii (Bitscore: 42, E-value: 4.3e-05)

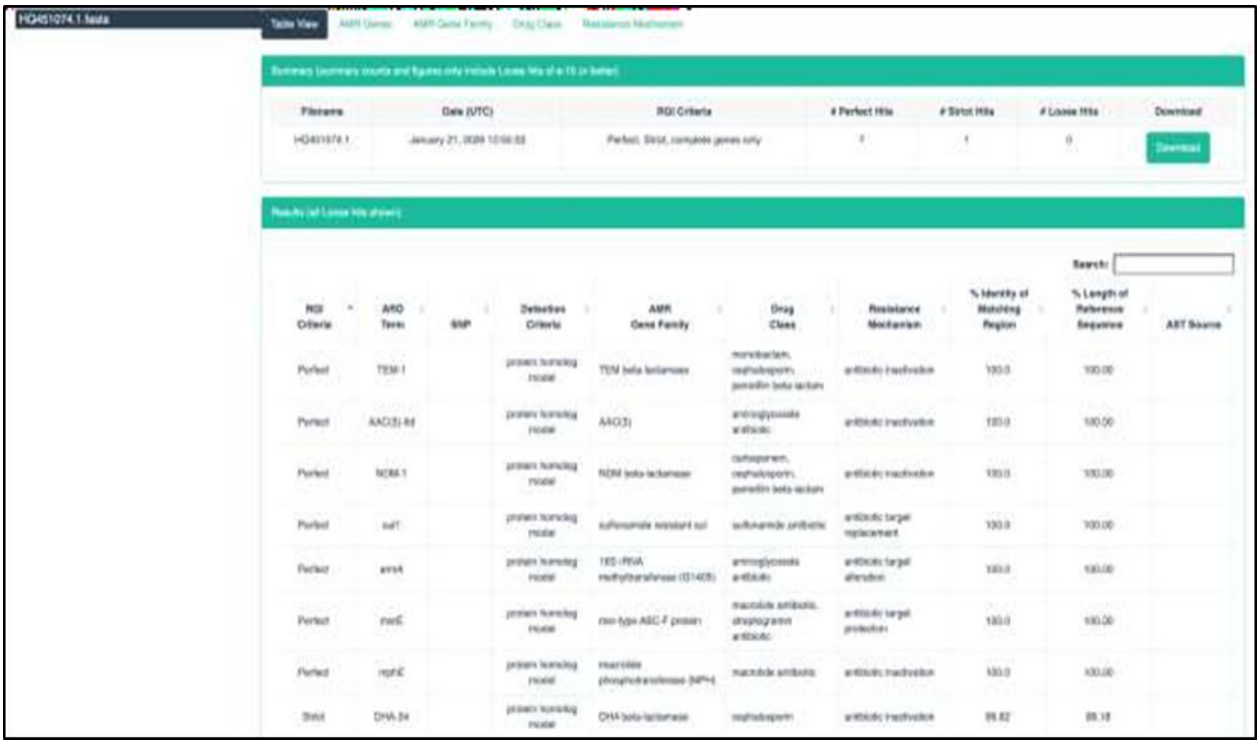

Figure 5: Summary of Antimicrobial Resistance Genes (ARGs) in E. coli identified via RGI and filtered for 'Perfect' and 'Strict' hits from the CARD database.

Figure 6: Pictorial representation of the Antibiotic Resistance Ontology of AMR genes, filtered by perfect and strict hits only

Figure 7: Pictorial representation of the AMR gene families of corresponding AMR genes, filtered by perfect and strict hits only

Figure 8: Pictorial representation of drug classes on which the proteins coded by AMR genes act, filtered by perfect and strict hits only

Figure 9: Pictorial representation of the resistance mechanism of AMR genes, filtered by perfect and strict hits only

IV. DISCUSSION

The results obtained from the methodology highlight several aspects of *E. coli*'s genes and resistance mechanisms.

The results from MSA highlight the similarity between *E. coli*'s gene coding for N-acetyltransferase and the genes of other organisms performing a similar function. The two images suggest that several commonalities exist in the amino acid sequences between two or more genes at the same sequence position. For example, each gene has methionine as its first amino acid in the sequence, following the universal genetic code.

The first picture uses specific colours to highlight each amino acid, while the second picture uses a simpler format to signify sequence alignment, containing dashes to represent non-alignment.

The results obtained from *E. coli*'s N-acetyltransferase enzyme gene's BLAST express the gene's similarity, in terms of the protein sequence, to other genetic sequences. The table in picture 3 shows the species with high gene alignment to *E. coli*, with the highest being the gene found in *Acinetobacter baumannii*, which has a high bitscore of 42 and a low E-value of 0.0000429685, highlighting the improbability of any chance.

The transfer of this gene between the two organisms can be attributed to the process of horizontal gene transfer, where genomic sequences from one species are transferred to an organism of a different species, leading to genetic diversity. The alignment between the N-acetyltransferase gene of *E. coli* and *A. baumannii* indicates that the gene in both organisms was highly conserved, meaning that it endured a minimal number of base mutations over the course of evolution and subsequent natural selection. This conservation of genomic sequences suggests that the N- acetyltransferase plays a crucial role in organism survival and protection against antibiotics and other antibacterial medications.

Finally, the results highlighted by *E. coli*'s RGI show the various genes that provide the organism with antimicrobial resistance.

Figure 5 combines all the data types into a table. It restricts the RGI criteria to perfect and strict hits only, excluding any loose hits. Further, the table mentions the names of the antimicrobial resistance genes, discussing their gene families, drug classes, the resistance mechanisms, % identity to matching regions, and the % identity to the length of reference sequences.

The high number of antimicrobial genes in *E. coli* reflects the organism's ability to resist microbial threats and stress.

Figures 6, 7, 8, and 9 represent the attributes of the table in Figure 5 in a graphical, pie chart format.

The existence of several AMR genes in *E. coli* explains the importance of AMR genes during an organism's evolution and the process of natural selection. As AMR genes allow organisms to survive against dedicated antimicrobial medications, the

process of natural selection disproportionately prefers retaining AMR genes over other genes with lesser functionalities.

This research possesses several limitations. The data used in this research is entirely computational and thus secondary in nature. This research does not incorporate the development of any new AI models that could yield authentic primary data, thus relying on existing, public computational tools and bioinformatics resources present on the internet. The use of a single accession number during the RGI also limits the scope of this research study.

V. CONCLUSION

By applying MSA, BLASTp, and RGI to the N-acetyltransferase enzyme gene of *E. coli*, this research successfully found similarities between the same gene in different organisms. It also highlighted the genes similar to the research's reference gene and the antimicrobial resistance genes in *E. coli*. Using RGI, the research found 8 resistance genes in the organism, with 7 having perfect alignment and one with a strict alignment constraint. These analytics highlight the exhaustive resistance mechanisms of *E. coli* and its relevance to the same in other microbial organisms.

Although significant, this research study possesses substantial prospects. Researchers can develop AI algorithms and new computational tools to acquire new, authentic information about AMR genes, which can be tested under the same computational tools to obtain unique data that can be analysed to infer new knowledge about AMR. Finally, researchers can also expand the scope of this research by using larger amounts of genomic data, such as obtaining more FASTA sequences for BLASTp and MSA, or running RGI using multiple gene accession numbers.

These findings indicate that N-acetyltransferase plays a significant role in antimicrobial resistance in E. coli and is evolutionarily conserved across species.

To conclude, this research study, using publicly available genetic resources, databases, and computational tools, highlights *E. coli*'s high resistance gene count and suitable resistance traits, helping it sustain the perpetual danger of antimicrobial damage.

REFERENCES

[1] Nsubuga, M., Galiwango, R., Daudi Jjingo, & Mboowa, G. (2024). Generalizability of machine learning in predicting antimicrobial resistance in E. coli: a multi-country case study in Africa. *BMC Genomics*, *25*(1). https://doi.org/10.1186/s12864-024-10214-4

[2] Lund, D., Parras-Moltó, M., Inda-Díaz, J. S., Ebmeyer, S., Larsson, D. G. J., Johnning, A., & Kristiansson, E. (2025). Genetic compatibility and ecological connectivity drive the dissemination of antibiotic resistance genes. *Nature Communications*, *16*(1). https://doi.org/10.1038/s41467-025-57825-3

[3] Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., Belk, K. E., Noyes, N. R., & Morley, P. S. (2020). MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic acids research*, *48*(D1), D561-D569. https://doi.org/10.1093/nar/gkz1010

[4] Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, *6*(1). https://doi.org/10.1186/s40168-018-0401-z

[5] Olatunji, I., Bardaji, D. K. R., Miranda, R. R., Savka, M. A., & Hudson, A. O. (2024). Artificial intelligence tools for the identification of antibiotic resistance genes. *Frontiers in Microbiology, 15*, 1437602. https://doi.org/10.3389/fmicb.2024.1437602

[6] Kim, J. I., Maguire, F., Tsang, K. K., Gouliouris, T., Peacock, S. J., McAllister, T. A., McArthur, A. G., & Beiko, R. G. (2022). Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. *Clinical Microbiology Reviews*, *35*(3). https://doi.org/10.1128/cmr.00179-21

[7] Ardila, C. M., González-Arroyave, D., & Tobón, S. (2025). Machine learning for predicting antimicrobial resistance in critical and high-priority pathogens: A systematic review considering antimicrobial susceptibility tests in real-world healthcare settings. *PLoS ONE*, *20*(2), e0319460–e0319460.

https://doi.org/10.1371/journal.pone.0319460

[8] Ali, T., Ahmed, S., & Aslam, M. (2023). Artificial Intelligence for Antimicrobial Resistance Prediction: Challenges and Opportunities towards Practical Implementation. *Antibiotics*, *12*(3), 523. https://doi.org/10.3390/antibiotics12030523

[9] Anahtar, M. N., Yang, J. H., & Kanjilal, S. (2021). Applications of machine learning to the problem of antimicrobial resistance: An emerging model for translational research. *Journal of Clinical Microbiology*, *59*(7). https://doi.org/10.1128/JCM.01260-20

[10] Giacobbe, D. R., Zhang, Y., & de, J. (2023). Explainable artificial intelligence and machine learning: novel approaches to face infectious diseases challenges. *Annals of Medicine*, *55*(2). https://doi.org/10.1080/07853890.2023.2286336

[11] Rannon, E., Shaashua, S., & Burstein, D. (2025). DRAMMA: a multifaceted machine learning approach for novel antimicrobial resistance gene detection in metagenomic data. *Microbiome*, *13*(1). https://doi.org/10.1186/s40168-025-02055-4