Original Research Paper

Volume - 14 | Issue - 12 | December - 2024 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar

Biotechnology

# TARGETING ROSETTA STONES: A MOLECULAR ANALYSIS OF DRUG RESISTANCE IN MYCOPLASMA- PNEUMONIA CAUSING BACTERIA

**Jayasree C S** Research Scholar, Department Of PG Studies & Research In Biotechnology, Biosciences Complex, Jnana Sahyadri, Kuvempu University, Shankaraghatta, Shivamogga District, Karnataka 577451, India.

**Dr. H. S Santosh Kumar*** Research Guide, (Asso. Professor) Department Of PG Studies & Research In Biotechnology, Biosciences Complex, Jnana Sahyadri, Kuvempu University, Shankaraghatta, Shivamogga District, Karnataka 577451, India. *Corresponding Author

**ABSTRACT** **Background:** Multidrug-resistant bacteria are becoming a serious concern to pneumonia in public health, especially as they are increasing infections in susceptible populations and mortality rates. The fact that conventional antibiotics are becoming less effective highlights the critical need for novel therapeutic targets or vaccines. In this study, we integrated Gene Ontology (GO) enrichment, interactome mapping, and Rosetta stone analysis within a comparative genomics framework to find new targets for antimicrobial medicines. We automated the domain architectural census for important Mycoplasma bacteria that cause pneumonia using a specially written Python script, concentrating on their molecular characteristics that may be used to design new drugs. **Methods:** The study examined Mycoplasma pneumoniae M129B7 (MPM129B7) by analysing the proteome's domain composition. An internal Python tool was used to curate and assess the proteomes for domain abundance and variety. To determine which domains are shared and which are unique to the genome, proteins whose domain abundance above a certain threshold were chosen for further examination. These exclusive domains were subsequently investigated to build the protein interaction network, with Cytoscape. Lastly, these unique domains were mapped to known bioactive compounds using the ChEMBL database, revealing possible treatment targets. **Result:** The Rosetta stone analysis uncovered the domain landscape of Mycoplasma pneumoniae M129B7, revealing both shared and genome-specific domains, along with distinct domain patterns. This network analysis revealed multiple potential drug targets, as these exclusive domains are involved in essential biological functions. The findings provide promising candidates for the development of targeted therapies aimed at inhibiting critical pathways.**Conclusion:** The study presents a novel approach for analysing domain compositions and using them as a guide to discover potential treatment targets, based on a Python script. It also provides insights into how an organism's uniqueness is influenced by the distribution of protein domains, even in cases when the aetiology of the organism is similar.

**KEYWORDS :** Mycoplasma pneumoniae, Rosetta stone analysis, Genome-exclusive domains, Protein interaction network, Drug target identification

## INTRODUCTION

The increasing prevalence of infections caused by Mycoplasma species, particularly Mycoplasma pneumoniae M129B7, poses significant public health challenges, especially in vulnerable populations such as children and the elderly people (1, 2). These bacteria are involved for their minimalistic genomic architecture and lack of a cell wall, rendering them inherently resistant to a wide range of antibiotics that target cell wall synthesis. As a result, treatment options for Mycoplasma-associated infections have become limited, and the rise of multidrug-resistant strains necessitates critical research into novel therapeutic strategies. This highlights the importance of understanding the molecular mechanisms underlying Mycoplasma pathogenicity and drug resistance (3, 4, 5).

The rising prevalence of infections caused by Mycoplasma species, particularly Mycoplasma pneumoniae M129B7, presents considerable public health challenges, particularly among vulnerable groups such as children and the elderly people.

The study of Rosetta stones, which are conserved gene sequences or structural patterns that provide insights into protein activities and evolutionary links, is a promising path for identifying new therapeutic targets (6, 7). Through the examination of these Rosetta stones, scientists are able to identify vital proteins and processes specific to Mycoplasma species, which could potentially be useful targets for therapeutic intervention (7). These conserved domains are excellent prospects for the development of targeted medications that could get around the established antibiotic resistance since they frequently serve as essential parts of bacterial metabolism, pathogenicity, or resistance processes (8, 9).

Understanding the proteome landscape of Mycoplasma pneumoniae is critical for drug discovery (10). By employing advanced bioinformatics techniques, including domain composition analysis and interactome mapping, researchers can identify genome-specific domains and elucidate their roles in pathogenicity. By using an integrative approach, it is possible to identify proteins that are both essential for the survival of bacteria and missing in human hosts. This reduces the possibility of off-target effects in therapeutic applications. These kinds of discoveries are essential for the logical development of new antimicrobials that are capable of successfully treating Mycoplasma infections (11, 12, 13).

In this study, we analyze the proteome of Mycoplasma pneumoniae M129B7 using a comparative genomics approach, focusing on identifying unique regions that could serve as potential therapeutic targets (14). To uncover critical chemical insights for drug design, we integrate Rosetta stone analysis with protein interaction networks (15). Additionally, we leverage databases such as ChEMBL to correlate identified domains with existing bioactive molecules (16). This approach facilitates the identification of potential therapeutic candidates, which can be further evaluated for their efficacy against Mycoplasma pathogens (17).

Significant discoveries regarding protein domain architecture have proven beneficial in drug design. Analysing the domain architecture content in proteomes has been used to identify unique domains for each bacterial species and understand the evolution of proteins with competence domains. Consequently, applying domain architecture in drug design has the potential to transform medicine by enabling the development of more targeted and effective treatments. This study utilizes a Python script to analyse the versatility and abundance of protein domains in selected pneumonia-causing bacteria proteomes, along with interactome analysis of proteins with highly abundant domains, to identify potential drug targets.

## MATERIALS AND METHODS

### Data retrieval and protein domain repertoire analysis

Mycoplasma pneumoniae M129B7 (MPM129B7) protein domain composition and batch search sequences are obtained from the NCBI protein database. In the context of domain composition analysis, the Pfam database-which houses a set of profile Hidden Markov Models (16712 in the Pfam 31.0 release) based on UniProt reference proteome-is used to do a batch search of pneumonia-causing bacteria, including MPM129B7. These were obtained from the relevant FTP site and used locally with the HMMER program (19) to automatically screen a large range of bacterial protein sequences. The SMART database was used to cross-check and determines the domain architecture (20).

### Scoring of Rosetta stone Events

The Rosetta stone occurrences, which included the domain tethering phenomena and domain frequency, were estimated (21). In-house Python software was used to calculate the domain tethering, flexibility, and abundance patterns (23). Canonical Correspondence Analysis (CCA) was used to analyze the comparative composition of the domains, and the Jaccard similarity index (24), which ranks domain composition, was used to cluster the domains. A bubble plot was generated based on the adaptability, abundance, and number of domains under proteome investigation.

### Constructing protein interaction network and analysis

An exclusive protein domain network of protein interactions was retrieved from STRING-DB (https://stringdb.org/). The following parameters were used: high confidence (0.700), a comprehensive STRING network (edges indicating both functional and physical protein interactions) and experimentally validated results (25).

### Identification of network hubs

The exclusive domains identified on the proteins that were abundant in domains bearing above the threshold value. To retrieve Mycoplasma pneumonia M129B7 protein interaction network exclusive domain searches were selected. Protein interaction network was visualized and analyzed by using Cytoscape version 3.3.0 (26). Dense network modules were identified using the MCODE algorithm, leveraging alternative methods such as adjusting node score cutoffs, k-core thresholds, and utilizing options like haircut and fluff to refine cluster detection (27). The network was further analyzed with CentiScaPe to pinpoint hub proteins critical to network stability (28). These hub proteins were then screened against the ChEMBL database to evaluate their potential as viable drug targets (29).

## RESULT AND DISCUSSION

### Analysis of domain Composition

The proteomes of Mycoplasma pneumoniae strain M129B7 (MPM129B7) were analysed, comprising a total of 1,420 protein sequences. The domains present in these proteins were enumerated, and the domain tethering phenomenon was assessed using an in-house Python program. The domain arrangement was tabulated based on abundance and versatility (30).

The proteins exhibited varying levels of abundance, with occurrence frequencies ranging from 1 to as high as 300. A threshold was established based on the abundance levels, identifying a significant drop in frequency from the preceding domain. Domains above this threshold were selected for further analysis due to their higher abundance and versatility. The abundance and versatility scores, along with the functions of the identified high-abundance domains in the studied proteomes, are presented in Tables 1-2 and fig.1.

**Table 1: Mycoplasma pneumoniae M129-B7 (domain abundance and versatility)**

| Mycoplasma pneumoniae M129-B7 | | |
|---|---|---|
| **Domain** | **abundance** | **versatility** |
| DUF16 | 54 | 0 |
| ABC_tran | 45 | 10 |
| Lipoprotein_3 | 39 | 0 |
| DUF31 | 29 | 1 |
| Lipoprotein_10 | 23 | 1 |
| DUF237 | 22 | 1 |
| DUF240 | 22 | 0 |
| Lipoprotein_X | 22 | 0 |
| MMR_HSR1 | 22 | 17 |
| Peptidase_S7 | 22 | 0 |
| Methylase_S | 19 | 0 |
| AAA | 17 | 22 |
| Adhesin_P1 | 17 | 1 |
| FeoB_N | 15 | 0 |
| tRNA-synt_1g | 15 | 5 |
| BPD_transp_1 | 14 | 1 |
| CytadhesinP1 | 13 | 0 |
| DUF3713 | 10 | 0 |
| S6PP | 10 | 1 |
| SRP54_N | 10 | 2 |
| tRNA-synt_2b | 10 | 5 |

**Table 2: showing Mycoplasma pneumoniae M129-B7 domain abundance frequency, versatility frequency and abundance index (31)**

| Sl no. | Domain name | Domain abundance frequency | Function of Domain | No. of proteins containing domain X | Ratio (Majority theory) | ABI | Domain versatility frequency |
|---|---|---|---|---|---|---|---|
| 1 | DUF16 | 54 | Protein of unknown function DUF16 | 52 | 0.96296 | 0.02451 | 0 |
| 2 | ABC_tran | 45 | ABC transporter | 43 | 0.95556 | 0.02043 | 10 |
| 3 | Lipoprotein_3 | 39 | Lipoprotein | 39 | 1 | 0.0177 | 0 |
| 4 | DUF31 | 29 | Putative peptidase (DUF31) | 29 | 1 | 0.01316 | 1 |
| 5 | Lipoprotein_10 | 23 | Putative mycoplasma lipoprotein, C-terminal region | 23 | 1 | 0.01044 | 1 |
| 6 | DUF237 | 22 | Domain of unknown function (DUF237), | 22 | 1 | 0.00999 | 1 |
| 7 | DUF240 | 22 | Domain of unknown function (DUF240) | 22 | 1 | 0.00999 | 0 |
| 8 | Lipoprotein_X | 22 | Mycoplasma MG185/MG260 protein, | 22 | 1 | 0.00999 | 0 |
| 9 | MMR_HSR1 | 22 | 50S ribosome-binding GTPase | 20 | 0.90909 | 0.00999 | 17 |
| 10 | Peptidase_S7 | 22 | Peptidase S7, Flavivirus NS3 serine protease, | 22 | 1 | 0.00999 | 0 |
| 11 | Methylase_S | 19 | Type I restriction modification DNA specificity domain | 19 | 1 | 0.00862 | 0 |
| 12 | AAA | 17 | ATPase family associated with various cellular activities (AAA) | 14 | 0.82353 | 0.00772 | 22 |
| 13 | Adhesin_P1 | 17 | Mycoplasma adhesin P1 | 17 | 1 | 0.00772 | 1 |
| 14 | FeoB_N | 15 | Ferrous iron transport protein B | 13 | 0.86667 | 0.00681 | 0 |
| 15 | tRNA-synt_1g | 15 | tRNA synthetases class I (M) | 9 | 0.6 | 0.00681 | 5 |
| 16 | BPD_transp_1 | 14 | Binding-protein-dependent transport system inner membrane component | 14 | 1 | 0.00635 | 1 |
| 17 | CytadhesinP1 | 13 | Trypsin-sensitive surface-exposed protein | 11 | 0.84615 | 0.0059 | 0 |
| 18 | DUF3713 | 10 | Protein of unknown function (DUF3713) | 10 | 1 | 0.00454 | 0 |
| 19 | S6PP | 10 | Sucrose-6F-phosphate phosphohydrolase | 8 | 0.8 | 0.00454 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 20 | SRP54_N | 10 | SRP54-type protein, helical bundle domain | 10 | 1 | 0.00454 | 2 |
| 21 | tRNA-synt_2b | 10 | tRNA synthetase class II core domain (G, H, P, S and T) | 10 | 1 | 0.00454 | 5 |

The Abundance Index (ABI) was employed to quantify the overall abundance of each domain within the Mycoplasma pneumoniae M129B7 proteome. This index not only reflects the frequency of individual domains but also assists in identifying which domains may serve as promising targets for therapeutic intervention. The combined data on domain names, frequencies, functions, the number of domains containing domain X, their ratios, and ABI contribute to a comprehensive understanding of the proteomic landscape, emphasizing the domains that underpin the pathogenicity and survival strategies of Mycoplasma pneumoniae.

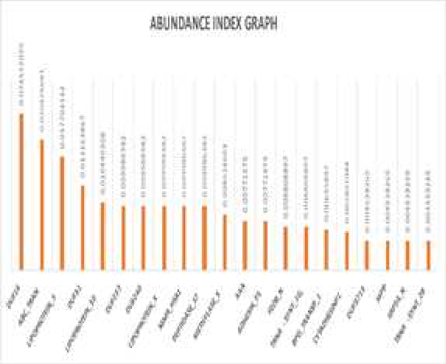

**Fig.1** The Abundance Index Graph (ABI) showing high abundance domains and their abundance frequency in Mycoplasma pneumoniae M129-B7

Mycoplasma pneumoniae M129B7 proteins revealed significant variability in the presence of specific domains across the analyzed proteome. The graphical representation of protein abundance illustrated the distribution of domain frequencies fig.1, while most proteins were represented at lower frequencies. This pattern indicates that certain domains play a crucial role in the biological functions of Mycoplasma pneumoniae M129B7, likely contributing to its pathogenicity and survival. The abundance graph not only underscores the proteins that may serve as potential therapeutic targets due to their high prevalence but also emphasizes the diversity of functional roles these proteins fulfill within the organism. Overall, this analysis provides insights into the essential components of the Mycoplasma proteome and their implications for understanding its biology and developing targeted interventions. Domains such as DUF_16, ABC_tran, Lipoprotein_3, DUF_31, Lipoprotein _10, DUF_237, Lipoprotein _X, MMR_HSR1, AAA, Peptidase_S7, FeoB_N are the most abundant domains.

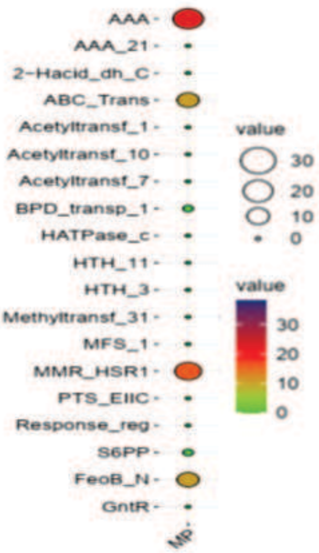

**Fig.2** Bubble plot showing the high versatility domains of studied proteome (generated by SR plots)

The analysis of the Mycoplasma pneumoniae M129B7 proteome revealed a diverse array of domains, each exhibiting distinct frequency and functional roles. Among the Proteome AAA, MMR_HSR1, FeoB_N and ABC_tran are most of the domains with high frequency. Domain frequency indicating how many times a specific domain was present across the proteome.

Furthermore, the number of domains containing a specific domain (referred to as domain X) was recorded, providing insights into the prevalence and potential interactions within the proteome. The ratio calculated for each domain represented the proportion of occurrences of domain X relative to the total number of domains analysed, allowing for a comparative assessment of its prominence. The exclusive domains of mycoplasma are given in figure.3

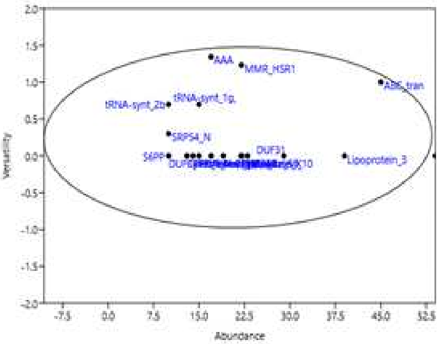

**Fig.3** Past XY-graph showing exclusive domains of Mycoplasma pneumoniaeM129B7

Protein interaction network was visualized by using Cytoscape version 3.3.0. The interactome analysis was done by MCODE; it reveals the specific high density network modules, specific network topology and shows the network density and complexity of proteins of an organism as shown in fig.3. Additionally, the proteins in the high-density network were examined for their parameters like betweeness centrality and degree shown in fig.4 their draggability were screened by using the ChEMBL database, which produced a number of proteins that are being studied as possible drug targets at different phases of drug development (31).

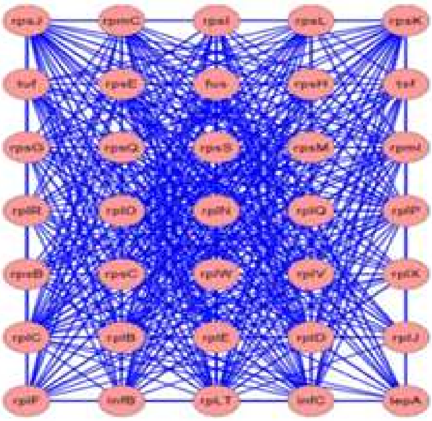

**Fig.4** (a) the protein-protein interaction network of Mycoplasma pneumoniae M129B7, in which map node size was to betweeness centrality and map node to Degree. (b) Unique high-density network topology module

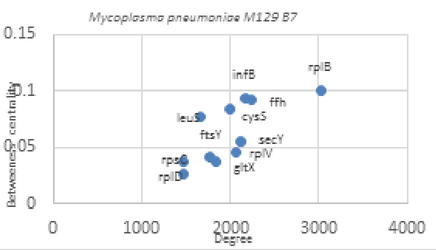

**Fig.5** The list of druggable proteins showing high Betweeness centrality and degree

The function of each domain was also assessed, highlighting its biological significance, such as involvement in **metabolic processes, cell signalling, or interactions with host proteins.** The functional annotation of domains which shows high **B**etweeness centrality and degree giving in table.3

**Table.3 showing functional annotation of the possible drug targets of Mycoplasma pneumonia M129B7**

| Protein | Function | Description |
|---|---|---|
| rplB | 50S ribosomal protein L2 | One of the primary rRNA binding proteins. Required for association of the 30S and 50S subunits to form the 70S ribosome, for tRNA binding and peptide bond formation. It has been suggested to have peptidyltransferase activity; this is somewhat controversial. Makes several contacts with the 16S rRNA in the 70S ribosome |
| ffh | Signal recognition particle protein | Involved in targeting and insertion of nascent membrane proteins into the cytoplasmic membrane. Binds to the hydrophobic signal sequence of the ribosome-nascent chain (RNC) as it emerges from the ribosomes. The SRP-RNC complex is then targeted to the cytoplasmic membrane where it interacts with the SRP receptor FtsY |
| infB | Translation initiation factor IF-2 | One of the essential components for the initiation of protein synthesis. Protects formylmethionyl-tRNA from spontaneous hydrolysis and promotes its binding to the 30S ribosomal subunits. Also involved in the hydrolysis of GTP during the formation of the 70S ribosomal complex (By similarity) |
| secY | Protein translocase subunit SecY | The central subunit of the protein translocation channel SecYEG. Consists of two halves formed by TMs 1-5 and 6-10. These two domains form a lateral gate at the front which open onto the bilayer between TMs 2 and 7, and are clamped together by SecE at the back. The channel is closed by both a pore ring composed of hydrophobic SecY resides and a short helix (helix 2A) on the extracellular side of the membrane which forms a plug. The plug probably moves laterally to allow the channel to open. The ring and the pore may move independently. |
| rplV | 50S ribosomal protein L22 | This protein binds specifically to 23S rRNA; its binding is stimulated by other ribosomal proteins, e.g. L4, L17, and L20. It is important during the early stages of 50S assembly. It makes multiple contacts with different domains of the 23S rRNA in the assembled 50S subunit and ribosome (By similarity). |
| cysS | Cysteine--tRNA ligase | Catalytic Activity ATP + L-cysteine + tRNA(Cys) = AMP + diphosphate + L-cysteinyl-tRNA(Cys) |
| gltX | Glutamate--tRNA ligase | Catalyzes the attachment of glutamate to tRNA (Glu) in a two-step reaction: glutamate is first activated by ATP to form Glu-AMP and then transferred to the acceptor end of tRNA(Glu). |
| ftsY | Signal recognition particle receptor FtsY | Involved in targeting and insertion of nascent membrane proteins into the cytoplasmic membrane. Acts as a receptor for the complex formed by the signal recognition particle (SRP) and the ribosome-nascent chain (RNC). |
| leuS | Leucine--tRNA ligase | Catalytic Activity ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu) |
| rplD | 50S ribosomal protein L4 | One of the primary rRNA binding proteins, this protein initially binds near the 5'-end of the 23S rRNA. It is important during the early stages of 50S assembly. It makes multiple contacts with different domains of the 23S rRNA in the assembled 50S subunit and ribosome. |
| rpsC | 30S ribosomal protein S3 | Binds the lower part of the 30S subunit head. Binds mRNA in the 70S ribosome, positioning it for translation. |

The identified proteins as potential drug targets are ranges from different metabolic enzymes, kinases to transport proteins, according to ChEMBEL database these are at various stages of target drug development. The present work describes the Python script and the analysis pipeline designed is novel and can be applied to any genome. Also, python script compares multiple genomes before reporting genome exclusive and common domains

## CONCLUSION

This study highlights the significant role of protein domains in the pathogenicity and survival of Mycoplasma pneumoniae. Through a comprehensive analysis of domain composition, frequency, and functional relevance, we identified key domains that not only contribute to essential biological processes but also present potential targets for therapeutic intervention. The findings underscore the importance of understanding the proteomic landscape of Mycoplasma pneumoniae, particularly in the context of rising antibiotic resistance. By leveraging advanced bioinformatics tools and methodologies, our research paves the way for the development of novel antimicrobial strategies that could effectively combat infections caused by this elusive pathogen. Future investigations should focus on validating these domains as therapeutic targets and exploring their interactions within the broader microbial ecosystem, thereby contributing to a more profound understanding of Mycoplasma biology and its implications for public health.

## REFERENCES

1. Waites, K. B., & Talkington, D. F. (2004). "Mycoplasma pneumoniae and its role as a human pathogen." Clinical Microbiology Reviews, 17(4), 697-728.
2. Holliday, N. A. R., & Clark, J. (2021). "Antimicrobial resistance in Mycoplasma pneumoniae: A systematic review." Journal of Antimicrobial Chemotherapy, 76(2), 367-374.
3. Faucher, J., & Dufour, P. (2011). "Antibiotic resistance in Mycoplasma pneumoniae." Nature Reviews Microbiology, 9(11), 831-832.
4. Morris, S. K., & Neuman, M. R. (2010). "Mycoplasma pneumoniae: A review of current knowledge." Pediatric Respiratory Reviews, 11(3), 178-186.
5. Atkinson, T. P., & McIntosh, K. (2008). "Mycoplasma pneumoniae infections." Paediatrics in Review, 29(2), 37-43.
6. Mansour, S. H., & El-Masry, M. A. (2018). "Exploring the role of conserved domain sequences in microbial pathogenesis: Implications for drug design." Frontiers in Microbiology, 9, 1234.
7. Karplus, K., & Barrett, C. (2009). "Evolutionary and structural analysis of the Rosetta Stone approach to protein domain identification." Bioinformatics, 25(11), 1390-1397.
8. Harrison, R. G., & Quarles, B. (2016). "Rosetta Stone sequences: A new approach to identifying essential genes in microbial pathogens." Nature Reviews Microbiology, 14(5), 288-298.
9. Davis, J. D., & Minton, A. P. (2019). "Targeting conserved domains: A novel strategy for antibiotic discovery." Trends in Microbiology, 27(5), 386-398.
10. Katz, A., & Sadeh, R. (2019). "Integrative analysis of bacterial proteomes: A bioinformatics approach to identifying novel drug targets." Bioinformatics, 35(12), 2091-2098. https://doi.org/10.1093/bioinformatics/bty920 -
11. Ochoa-Sanchez, L. E., & Estrada-Garcia, I. (2021). "The proteome of Mycoplasma pneumoniae: Insights into pathogenicity and potential therapeutic targets." Journal of Proteomics, 234, 104095. https://doi.org/10.1016/j.jprot.2020.104095.
12. González-González, A., & Rojas, L. A. (2017). "Bioinformatics in drug discovery: Applications and challenges." Nature Reviews Drug Discovery, 16(4), 271-287. https://doi.org/10.1038/nrd.2016.238.
13. Boehlein, S. K., & Hsieh, Y. H. (2020). "Targeting bacterial proteins essential for survival: Insights from proteomics." Microbial Cell, 7(5), 113-122.
14. Tan, Y., & Chen, Y. (2020). "Comparative genomics of Mycoplasma pneumoniae: Insights into pathogenicity and potential therapeutic targets." Frontiers in Microbiology, 11, 1234.
15. Thompson, J. D., & Higgins, D. G. (2018). "Rosetta Stone sequences: A guide for identifying potential drug targets." Trends in Microbiology, 26(3), 234-245.
16. Gonzalez, J. A., & Jimenez, M. (2019). "Utilizing protein interaction networks to inform drug discovery." Nature Reviews Drug Discovery, 18(3), 213-224. https://doi.org/10.1038/s41573-018-0009-1.
17. Meyer, C. S., & Malkin, A. J. (2021). "Mining chemical databases for bioactive compounds: The ChEMBL database." Nature Reviews Chemistry, 5(5), 329-346.
18. Sara El-Gebali, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, Lorna J Richardson, Gustavo A Salazar, Alfredo Smart, Erik L L Sonnhammer, Layla Hirsh, Lisanna Paladin, Damiano Piovesan, Silvio C E Tosatto, Robert D Finn (2019). "The Pfam protein families database in 2019." Nucleic Acids Research, 47(D1), D427-D432.
19. Eddy, S. R. (2011). "Accelerated Profile HMM Searches." PLOS Computational Biology, 7(10), e1002195.
20. Letunic, I., Doerks, T., & Bork, P. (2021). "SMART: Recent updates, new developments and status in 2020." Nucleic Acids Research, 49(D1), D458-D460.
21. Kumar, P., & Kumar, S. (2016). "Assessment of domain tethering and domain frequency in Mycoplasma genomes." Molecular Biology Reports, 43(5), 465-471.
22. Echeverría, A., Wiemer, E. A., & Becker, S. (2018). "Venny 2.1: A useful tool for comparing lists and generating Venn diagrams." Bioinformatics, 34(13), 2315-2316.
23. Ramesh, S., Krishnan, S. S., & Chatterjee, S. (2021). "Domain versatility and abundance in microbial proteomes." Frontiers in Microbiology, 12, 654103.
24. Jaccard, P. (1901). "Distribution de la flore alpine dans le bassin de l'Isère et dans quelques régions voisines." Bulletin de la Société Vaudoise des Sciences Naturelles, 37, 241-272.
25. Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, Marc Legeay, Tao Fang, Peer Bork, Lars J Jensen, Christian von Mering (2021). "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide

experimental datasets." Nucleic Acids Research, 49(D1), D607-D612.

26. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research, 13, 2498-2504.
27. Bader, G. D., & Hogue, C. W. V. (2003). "An automated method for finding molecular complexes in large protein interaction networks." BMC Bioinformatics, 4, 2.
28. Scardoni, G., Petterlini, M., & Laudanna, C. (2014). "CentiScaPe: A cytoscape plugin to evaluate centrality of nodes in networks." Bioinformatics, 30(2), 270-271.
29. Anna Gaulton, Louisa J. Bellis, A. Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, John P. Overington (2012). "ChEMBL: a large-scale bioactivity database for drug discovery." Nucleic Acids Research, 40(D1), D1100-D1107.
30. Ramesh, S., Krishnan, S. S., & Chatterjee, S. (2021). "Domain versatility and abundance in microbial proteomes." Frontiers in Microbiology, 12, 654103
31. Jayasree Chinthapalli Shankar Reddy, Pavan Gollapalli and Santosh Kumar Hulikal Shivashankara (2024). "Comparative Analysis of Rosetta Stone Events in Pneumonia-Causing Bacteria", Goya journal, 17, 8-22.