

## POSTER - DATABASE AND SOFTWARE DEVELOPMENT

### **PANVITA: A PIPELINE FOR VIRULENCE AND RESISTANCE ANALYSIS**

*Diego Lucas Neres Rodrigues (diego.neresr@gmail.com)*

*Juan Carlos Ariute (jcarlosariute@gmail.com)*

*Francielly Rodrigues Da Costa (franrodriguesdacosta@gmail.com)*

*Ana Maria Benko-Iseppon (ana.iseppon@gmail.com)*

*Debmalya Barh (dr.barh@gmail.com)*

*Vasco A De C Azevedo (vascoariston@gmail.com)*

*Flavia Figueira Aburjaile (faburjaile@gmail.com)*

The rise of new sequencing technologies is an impacting factor for the world scientific scenario, but along came an unsolved problem: the huge amount of data generated and not analyzed. There are currently thousands of genome sequencing projects underway, but there is still very limited perspective on methods that can handle the analysis of these data at the same rate at which they are produced. On the other hand, another issue is the increasing number of multidrug-resistant bacterial isolates worldwide. The problem of bacterial resistance is almost incomprehensible due to the randomness with which genetic factors change and are transferred. In addition, the uncontrolled consumption of antimicrobials by the population is still a big point in the face of the problem. Another issue is the low density of scalable genomic data visualization packages and tools, i.e. big data visualization related to genomic data. Thus, there is a constant need for the development of software that can

handle bacterial resistance and virulence data, and also assist in the disposition of visual data. With this in mind, the Pan Virulence and resisTance analysis pipeline (PanViTa) was developed in 2021. The focus of this tool is the analysis of bacterial virulence and resistance genes, taking into regard the information available in known and curated databases, these being the following: Comprehensive Antibiotic Resistance Database (CARD); Virulence Factor Database (VFDB); Antibacterial Biocide and Metal Resistance Genes Database (BacMet). Through multiple comparisons against the selected datasets, the tool determines the presence and absence of interest genes in the strains used as input and returns as main outputs: (I) a clustermap containing the gene presence information for each strain, as well as the degree of identity of each gene when compared against the selected database; (II) a bar graph representing the number of genes involved in resistance to certain antimicrobials or heavy metals; (III) a bar graph representing the number of genes related to certain antimicrobial resistance mechanisms; (IV) tables indicating qualitatively and quantitatively the presence of each gene in each strain. The tool was also compared to others with the same purpose, such as ABRicate and ResFinder. The results indicate that PanViTa was superior in analysis time, being able to handle a dataset of 100 *Acinetobacter baumannii* genomes with an average genome size of 4.08 Mb in 2 minutes, while the closest tool handled the same dataset in double the time and it not being able to generate a visual output. With this in mind, we can say that PanViTa is a scalable tool for genome analysis and visualization of resistance data, being able to quickly and accurately handle large datasets. The tool is currently available on the GitHub repository and is registered by the Coordenadoria de Transferência e Inovação Tecnológica da Universidade Federal de Minas Gerais.