Contribution ID: **12**   Type: **Presentation**

Implementing High-Throughput Sequencing in bacterial foodborne pathogen surveillance: The INNUENDO Platform

Background: Outbreak investigations and pathogen surveillance are crucial tasks to control transmission of foodborne transmitted diseases. The decreasing costs of High-Throughput Sequencing (HTS) are boosting application of HTS for molecular typing in routine surveillance and outbreak investigation, maximizing discriminatory power in outbreak detection. However, lack of standardized bioinformatics infrastructures for data processing and integration, together with limited bioinformatics skills, continue to be major hurdles of HTS routine implementation, specially when analysing large datasets where the required computational needs are not available to most of the groups. To overcome these limitations, we developed the INNUENDO platform, an infrastructure that provides a user-friendly interface and the required framework for data analysis, from raw data quality assurance to integration of epidemiological data and visualization of the final analyses, providing the tools for the use of HTS techniques in everyday surveillance and outbreak investigation.

Methods: The INNUENDO platform is composed of two main applications that interact with each other using a REST API. The first application provides a graphical user interface that allows the user to define their own projects (groups of sequencing data and their associated metadata), control which protocols (software and their parameters) are applied to the data, and visualize the results. The second application controls the job submissions and status by using Nextflow as workflow engine and SLURM or any other job scheduler supported by Nextflow to control the available resources. Each software is available as a docker image and are loaded if required depending on the submitted job. The INNUENDO Platform includes the INNUca pipeline for automatic QC from reads to draft genome assemblies, which ultimately aims at producing consistently high-quality and comparable genomic data. The curated genome assemblies are then analysed following a gene-by-gene typing based approach. The chewBBACA software is used to perform the allele calling for whole genome MLST (wgMLST) profile definition. The wgMLST profiles generated for each isolate of interest can then be compared with profiles already stored in the platform's database. The wgMLST profiles of the isolates of interest, together with a selection of the closest ones in the database, are then filtered to produce a core genome MLST (cgMLST) and the data sent to PHYLOViZ Online for the construction of a minimum spanning tree annotated with metadata, allowing the exploration of possible epidemiological scenarios.

Results and conclusion: INNUENDO platform was developed with a modular design allowing the incorporation of different bioinformatic tools for the characterization of specific pathogens, and the capacity of being run in High Performance Computers clusters, which can greatly reduce the analysis time for large datasets. The modular nature of the platform implementation also allows for scalability in terms of computing needs. It also aims to facilitate data sharing and communication between different institutions, promoting cooperation in surveillance and outbreak investigation. The use of open source tools and standardized protocols will allow a future accreditation of the INNUENDO platform.

References and acknowledgements: More information on INNUENDO project (co-funded by EFSA) and the platform can be found in http://www.innuendoweb.org/. A working prototype of the INNUENDO platform was produced with the support of INCD funded by FCT and FEDER under the project 22153-01/SAICT/2016".

Primary authors: RIBEIRO-GONÇALVES, Bruno (Instituto de Microbiologia and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal); SILVA, Diogo (Instituto de Microbiologia and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal); P. MACHADO, Miguel (Instituto de Microbiologia and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal); SILVA, Mickael (Instituto de Microbiologia and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal); HALKILAHTI, Jani (National Institute for Health and Welfare, Helsinki, Finland); JAAKKONEN, Anniina (Microbiology Research Unit, Finnish Food Safety Authority, Evira, Finland); RAMIREZ, Mario (Instituto de Microbiologia and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portuga); ROSSI, Mirko (Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland); ANDRÉ CARRIÇO, João

(Instituto de Microbiologia and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal)

Track Classification: R&D for computing services, networking, and data-driven science at the Iberian level.