

Decoding the Microbial Spectrum: Bioinformatic Approaches for Viruses, Bacteria and Fungi

Marius SURLEAC^{a,b}, Ilda CZOBOR-BARBU^a, Simona PARASCHIV^b, Irina GHEORGHE-BARBU^a, Liviu-Iulian ROTARU^b, Dan OȚELEA^b, and Mariana Carmen CHIFIRIUC^b

^a Research Institute of University of Bucharest, Bucharest, Romania.

^b National Institute for Infectious Diseases ‘Matei Balș’, Bucharest, Romania.

E-mail (*): marius.surleac@gmail.com

* Author to whom correspondence should be addressed;

Abstract

Next-generation sequencing is one of the most used techniques for pathogen research such as that on viruses, bacteria and fungi. Bioinformatics plays a very important role for processing, analyzing and interpreting huge datasets, therefore providing complete insights into pathogen evolution, antimicrobial resistance, mobile genetic elements (MGE) composition, etc. These are key features that could contribute to disease control and surveillance or even understanding underlying mechanisms for transmission. For viral sequencing bioinformatics we use genome assembly, variant calling and phylogenetic analyses, among other approaches, for tracking outbreaks or for identifying various strains and understanding the role of some particular viral mutations. For bacterial sequencing bioinformatics, we use whole-genome, 16S- and shotgun-metagenomics sequencing approaches for subtyping, resistance genes and MGE identification, among other approaches. Comparative genomics as well as pan-genome analyses help us explain bacterial adaptations and other mechanisms that could contribute to pathogenicity. For fungal bioinformatics we face unique challenges due to larger and more complex genomes (e.g. multiple chromosomes), the novelty of the pathogen (e.g. *C. auris*) and other particularities such as carrying specific traits (e.g. copy number variations) with yet unknown roles due to scarce annotation data. Some of the methods used involve quality control filtering and trimming performed with *HTStream* and *Trimmomatic*, *de novo* assembly and reference mapping with tools like *shovill* or *Snippy*, antibiotic resistance predicted with *ResFinder* or *CARD*, pangenome analysis with *Roary*, taxonomy profiling with *Metaxa2* or huge phylogenetic analyses on fungi performed with *Pathogenwatch*.

Keywords: Bioinformatics; Sequencing; Viruses; Bacteria; Fungi.

