6th Clinical Microbiology Conference

October 20-22, 2016 Rome, Italy

A pipeline to quickly and automatically characterize microbial genomes from high-throughput sequencing data

Segolene Caboche¹, D Hot¹ and C Audebert² ¹University of Lille, France ²Genes Diffusion, France

Recent advances in high-throughput sequencing (HTS) technologies open new perspectives in diagnostic microbiology (management of infected patients) and in epidemiology of infectious diseases. The Whole-Genome Sequencing application has significantly increased our knowledge of microorganism genomes and has facilitated investigations in clinical microbiology but from concept to routine use, some bottlenecks still need to be unlocked. One of these is the step of data analysis which requires bioinformatic expertise, needs to be reproducible and is often time consuming. Here we present MICRA, a freely available on-line pipeline to quickly and automatically characterize microbial genomes from HTS data. MICRA requires only the raw sequencing data (FASTQ) as input and then automatically finds the most closely related genomes based on sequence similarity (called reference genomes) and establish lists of discrepancies in term of genes and plasmids content and in term of variations compared to reference genomes. Optional modules allow the user to obtain Venn diagrams to quickly compare the gene content and/or the variations list between several strains. Another module identifies drug susceptibility and resistance genes and generates an e-antibiogram of the analyzed strain. All the steps included in MICRA were optimized in time. For example, for a PGM Ion Torrent run containing more than 2 million of 200-bases reads, the pipeline requires only 15 minutes to generate a full analysis. Diagnostic microbiology and epidemiology of infectious diseases will benefit from such pipeline as it defines a new standard of genome-based diagnosis combining results from a former time-consuming and heterogeneous multiple step method.

Biography

Segolene Caboche has completed her PhD in Computer Sciences in 2009 from Lille University and Postdoctoral studies from European Bioinformatics Institute, UK in 2009-2010. Her research interest is focused on high-throughput sequencing data. She is the author of 11 international publications and a regular Reviewer for several high-impact international journals. She has joined the Transcriptomics and Applied Genomics team in 2011 and co-advised PhD students in bioinformatics.

segolene.caboche@pasteur-lille.fr

Notes: