

## Genome sequence data: the key driver in shaping prokaryotic systematics

Dr Praveen RAHI<sup>1</sup>, Dr Federica PALMA<sup>1</sup>, Dr Dominique CLERMONT

<sup>1</sup>*Institut Pasteur*

Names are important for all organisms, including bacteria and archaea. It is often the first thing we want to know and form judgments instantly once the name is known. The description of a taxon of bacteria and archaea requires several experiments. Traditionally, microbial classification was based on the physical and biochemical characteristics of organisms, such as their morphology, physiology, and metabolism. However, these methods were often subjective and unreliable, which lead to the misclassification of several prokaryotes. Microbial biological resource centres (mBRCs) ensure access to authenticated and quality-controlled microbial resources. Therefore, ensuring the correct taxonomy of microbial strains is a must for an mBRC, as its resources are often used to support new discoveries and follow-up studies. With the advent of high-throughput sequencing technologies, it is now possible to obtain a comprehensive and objective view of the genetic makeup and evolutionary relationships of microorganisms, allowing a more accurate and systematic classification. At the Collection of Institut Pasteur (CIP), we are using genome sequencing data for the accurate identification of bacterial strains. Based on the core-gene phylogeny inference, organisms with shared evolutionary histories were accurately placed into monophyletic clades. Monophyletic clades were also found for the members of complex groups, for which the 16S rRNA gene-based identification was inconclusive and 16S rRNA gene phylogeny resulted in polyphyletic clades. Genome sequencing data has therefore led to the description of many new bacterial taxa, which were previously misclassified. Additionally, the genome sequence-based analyses, such as phylogenomics, average nucleotide identity, and average amino acid identity helped resolve the taxonomic conflicts and provided evidence for several reclassifications. All these advanced bioinformatics are accessible to everyone. At Institut Pasteur, we host BIGSdb-Pasteur, a genomic taxonomy and nomenclature platform for bacterial strains available at [bigsdb.pasteur.fr](http://bigsdb.pasteur.fr). This system allows non-bioinformatician users to perform species identification and whole genome typing based on (ribosomal, core genome) multilocus sequence typing (MLST) scheme, in addition to antimicrobial resistance or virulence characterization. In conclusion, genome sequencing and analysis have become an essential tool for prokaryotic systematics (providing a more accurate, objective, and comprehensive view of microbial diversity and evolution) and can be easily accessed by the scientific community through dedicated service platforms.