**E051**

**2-hour Educational Workshop**

**Whole genome sequencing - what the heck?**

## Traditional typing efforts in the era of whole genome sequencing: how disruptive is disruptive?

D. Harmsen[1]

[1]*Univ. Münster, Münster, Germany*

Infectious diseases remain a major cause of morbidity and mortality worldwide, representing major threats to human health. They spread through human populations, among animals and humans, in livestock, and *via* the food chain, making an effective trans-sector 'One Health' approach essential for their control. Whole Genome Sequencing (WGS) using Next Generation Sequencing (NGS) technology provides comprehensive information for pathogen studies of all types; however, the cross-border and cross-sector application of this approach is currently hampered by a lack of standardization and backwards compatibility with traditional typing methods.
Therefore, the next big challenge will be to harmonize analysis of WGS data so that microbiologist speak one language world-wide, i.e. 'molecular Esperanto', that allows them to compare their data quickly with data not generated by them. For establishing such a public and worldwide nomenclature the developed algorithms must scale with arbitrarily large datasets. Design requirements are that the algorithm must be able to handle 'data streams' and the compute cost to add a new genome must ≈ O(1). These requirements are best met by a genome-wide gene-by-gene analysis (core genome MLST [cgMLST] or MLST+); even though it is reductionist, loses complexity and only focuses on coding sequences it comes with a nomenclature. Backwards compatibility with some sequence-based typing techniques like MLST is already now achieved. With increasing read length of 2$^{nd}$ generation or even very long reads of 3$^{rd}$ generation sequencing technology extracting of repetitive element information like spa-typing, spoligo or MLVA/VNTR is becoming more and more successful. Even traditional typing efforts like sero-typing are amiable from WGS data although currently frequently the interpretation middleware is missing. Only band-based genotyping techniques as PFGE are especially with draft genomes challenging and might never be fully backwards compatible.