Genealogical concordance and other lines of evidence for the recognition and description of bacterial species

SN Venter[1], WY Chan[1], M Palmer[1], G Shin[1], CW Beukes[2], ET Steenkamp[2]

[1] Department of Microbiology and Plant Pathology, University of Pretoria, South Africa
[2] DST-NRF Centre of Excellence in Tree Health Biotechnology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa

Bacterial species are commonly defined by applying a set of predetermined criteria, including DNA:DNA hybridization values, 16S rRNA sequence similarity and phenotypic and chemotaxonomic comparisons. Alternative genome-based similarity criteria have also been proposed to define bacterial taxa. These criteria mostly allow for the delimitation of taxa that resemble typical bacterial species. The application of these criteria is often complicated when the objective is to delineate new species that are characterized by significant population-level diversity or that have evolved recently. This is because most of the predetermined taxonomic criteria utilize species cut-off values that are often not appropriate for older or more diverse taxa. However, we believe that these complexities and limitations can be easily circumvented by recognizing that bacterial species represent unique and exclusive assemblages of diversity. Within such a framework, methods that account for the population processes involved in species evolution are used to infer species boundaries, while the resulting species hypothesis are tested using additional biological data. In practice, a method such as genealogical concordance analysis allows for the generation of species hypotheses (i.e. this method delineates a putative species). The existence of the new taxon is then interrogated using an array of characters that include the traditionally used criteria, as well as genome-based criteria such as gene content and average nucleotide identity (ANI). By making use of taxa in the genera *Escherichia, Pantoea* and *Paraburkholderia* we demonstrate how genealogical concordance can be used to delimit a bacterial taxon. Other biological criteria were used to provide independent lines of evidence for the existence of that taxon. This approach to species recognition and description is straightforward and applicable to bacterial species of all ages and diversities. It is particularly useful in the post-genomic era, with whole genome sequences now being accessible for taxa of interest.