

Some thoughts and observations on “taxon calling”

George M. Garrity^{*1,2}, Phan Chuong-Vo², Daniel Bakoz¹, Charles T. Parker²

¹*Microbiology and Molecular genetics, Michigan State University, East Lansing, Michigan, USA,*

²*NamesforLife, LLC, East Lansing, MI, USA*

Abstract

In the life sciences, the formal Latinized names that are applied to different ranks in contemporary taxonomies are deeply rooted in both custom and practice. The formation and application of names are governed by international codes of nomenclature that have evolved over more than two centuries and function much like contemporary international standards. Conceptually, a formal name in a Linnaean system of nomenclature signifies a particular entity (a taxon) or group of entities (taxa) with a given circumscription, position and rank in a published taxonomic proposal. Not only does the publication establish a taxonomic hypothesis, it also establishes the priority of a name, a clear definition of the taxon composition and taxon boundaries (circumscription), and information about where the reference material (the type) is held and permanently available for comparative purposes. The codes also establish numerous rules regarding procedures for revision of prior taxonomies, the orderly emendation of existing taxa and the introduction and naming of new taxa. When done correctly, the complete historical record can be followed both backward and forward in time; revealing not only differing points of view, but the changes in methodology that gave rise to those differing views.

These codes of nomenclature act much like the international standards that are widely used in many other settings; however, a common misconception within the larger microbiology community is that the codes govern taxonomy. With the exception of viruses, that is not the case. In fact, freedom of taxonomic thought is protected to ensure that outdated taxonomic hypotheses do not become rigidly fixed in place. But this freedom also imposes requirements for a higher degree of scientific rigor, transparency and reproducibility that does not exist in other areas of biology. Nowhere is this more obvious than in contemporary microbiology, where gene and genome sequencing have rapidly replaced older methods of classification and identification. Sequencing has democratized taxonomy because of its universal applicability, ease of use and low cost. It is not uncommon to encounter ten or more proposals for newly named taxa in the literature on any given day or revisions of entire phylogenetic lineages, but only a small number of those proposals are done with sufficiently rigor to be included in the formal nomenclatural record. The application of different methodologies, different data sets of varying quality, different sample types and naïve reasoning can lead to very surprising conclusions that can have significant downstream effects when those names and accompanying assertions are incorporated into subsequent studies.

For more than two decades, we have been following the changes to the taxonomy and nomenclature of bacteria and archaea with validly published names. As part of an effort to understand the evolution of a well-regulated subject language terminology (prokaryotic nomenclature), we have tracked the changes in the underlying gene and genome sequences that serve as the principal definition of each taxon. A benefit of this effort is that it allows us to pose a wide range of questions about methodological variation and hidden biases that can impact classification, identification and naming including: the effect of different alignments, the effect of different marker genes, the effect of varying sequence quality and length, the effect of different algorithms and the effect of the application of a single arbitrarily derived cut-off point to define the taxonomic boundaries that serve as the foundation of our current taxonomy of prokaryotes. The focus of this lecture will be to demonstrate the value of a well-curated and carefully annotated reference database that can be used to evaluate existing and new methods of identifying and assigning names to prokaryotic taxa which can serve as a standard and be used for routine re-annotate and updating of existing metagenomes and microbiomes at a much finer grain of resolution that is currently used.