

Opinion

Divorcing Strain Classification from Species Names

David A. Baltrus^{1,*}

Confusion about strain classification and nomenclature permeates modern microbiology. Although taxonomists have traditionally acted as gatekeepers of order, the numbers of, and speed at which, new strains are identified has outpaced the opportunity for professional classification for many lineages. Furthermore, the growth of bioinformatics and database-fueled investigations have placed metadata curation in the hands of researchers with little taxonomic experience. Here I describe practical challenges facing modern microbial taxonomy, provide an overview of complexities of classification for environmentally ubiquitous taxa like *Pseudomonas syringae*, and emphasize that classification can be independent of nomenclature. A move toward implementation of relational classification schemes based on inherent properties of whole genomes could provide sorely needed continuity in how strains are referenced across manuscripts and data sets.

Confusion Abounds in Modern Bacterial Taxonomy

Communication between researchers is a foundation of all scientific disciplines, and clarity of the message is therefore essential for moving science forward. Alternatively, confusion of underlying messages leads directly to systemic problems and disagreements. For modern microbiologists perhaps the best example of how systemic confusion can slow research progress involves ongoing disagreements about bacterial classification and nomenclature, a confusion which is only amplified by the traditional entwining of these two activities. The advent of 'big data' has placed microbiology at a crossroads where we can either systematically change the way we think about describing strains or suffer within an ever expanding cloud of uncertainty. Now is the time to divorce classification of strains from any discussions about nomenclature based on species concepts and transition to a system based on genomic information, at least for metadata entry and to ensure continuity across manuscripts.

The root of this article lies in a frustration that many researchers deal with every day – a frustration born out of the clashes between how taxonomy should proceed in theory and the realities of how it proceeds in practice. Although nomenclatural confusion has always inconvenienced microbiology, the speed and focus of research, as well as the dedication of large numbers of taxonomists, previously enabled back and forth dialogues to smooth over ongoing disagreements. However, traditional taxonomic schemes have not efficiently dealt with the rapid influx of genomic data and were not designed to account for intrinsic challenges that arise when nontaxonomists publish metadata. Researchers have been arguing about bacterial species concepts since the dawn of microbiology [1–3], but the intent of this article is not to get caught up in discussions about what constitutes a bacterial species nor is it to suggest that the perfect mechanism for classification has been uncovered. My goal is to call attention to conflicts between the philosophy and practice of bacterial taxonomy, which generate much confusion for the classification of environmentally ubiquitous taxa like *P. syringae*.

Trends

Traditional classification schemes for bacteria were not designed to deal with an influx of large amounts of genomic data or with an eye toward metadata curation by nontaxonomists.

Taxonomy for environmentally ubiquitous bacteria, such as *Pseudomonas syringae*, is particularly prone to confusion given their lifestyle as facultative pathogens.

Implementation of classification schemes based upon whole-genome comparisons could provide much needed continuity across databases and publications.

¹School of Plant Sciences, University of Arizona, Tucson, AZ, USA

*Correspondence:
baltrus@email.arizona.edu
(D.A. Baltrus).

The Philosophy of Bacterial Classification and Nomenclature

Taxonomy is a branch of microbiology that consists of three fundamental and often intertwined activities: identification, classification, and nomenclature of strains [4]. While these words can often be thought of as synonymous, important yet subtle distinctions can be drawn between them. Whereas classification provides a means to index strains logically, it can exist independently of studies of how to accurately identify or name particular groups of strains.

One main reason for applying Linnean species names to microbes is that these names represent a shorthand for describing underlying aspects of the organism's biology while also informing about classification. In an academic context, although names are important for identification's sake, they also provide continuity throughout manuscripts, and enable researchers to investigate their own systems and make general predictions from other work. As science and technology progress, and many researchers become beholden to large curated databases, metadata provides a way to rapidly access and apply information across taxa and studies. For instance, if metadata were perfectly applied, one need only search the words '*Pseudomonas syringae*' and all relevant references and datasets would be retrieved. The importance of proper curation of metadata to research is illustrated by the development of text mining algorithms to discover emergent phenomena across systems [5,6]. How this metadata is curated significantly affects continuity and communication within manuscripts and across research groups.

In a perfect world, one would be able to see a name, garner something about the biology of that particular strain, and know exactly how others are related. For instance, *Helicobacter pylori* represents a cluster of strains which largely only lives within human stomachs and which are causative agents of gastric ulcers and cancer [7–9]. Classification and nomenclature for *H. pylori* strains is straightforward, despite extensive diversity across strains, because the unique and specialized niche that this organism inhabits enables tight grouping of phenotypic and phylogenetic clusters [10]. In other cases, phenotypic traits of interest exist outside those normally used for classification so that additional modifiers are added. *Escherichia coli* is a common inhabitant of vertebrate gastrointestinal tracts, but certain modifiers are added to this species name to reflect important data about biology [such as uropathogenic *E. coli* (UPEC)] [11]. Naming schemes are by no means perfect, and there have been situations where nomenclature and classification have pointed in opposite directions. These cases highlight different requirements for microbial nomenclature between professional and academic settings, and demonstrate how classification schemes can function in one context but not in the other. Even though strains of *Shigella* cause similar disease symptoms in humans, there have been multiple independent evolutionary origins due to convergent evolution of phenotypes [12]. This underlying evolutionary convergence might not matter much in a hospital, because the resolution of phylogeny does not affect treatment. However, conflation of different evolutionary lineages under the same name could affect how results are interpreted across comparative genomic studies unless researchers explicitly incorporate these phylogenetic nuances into their designs. Many other cases likely exist where nomenclature is wrong or misleading, but which will languish in obscurity due to the lack of widespread interest or medical relevance.

Nomenclature and Classification Schemes in Practice

When thinking about bacterial taxonomy, one cannot set aside historical momentum generated by the requirement of cultureability of strains in the early days of microbiology. The first step for any nomenclatural decision is, traditionally, the establishment of a 'type' strain that is used to set a foothold for new species designations [13]. Following from cultureability, bacterial types are binned by observable properties at microscopic and macroscopic scales. One of the better known schemes today involves grouping of *E. coli* strains based on their O and H antigens (e.g., O157:H7)[14], and *Salmonella* strains have been typed according to their phage sensitivities

Download English Version:

<https://daneshyari.com/en/article/6137808>

Download Persian Version:

<https://daneshyari.com/article/6137808>

[Daneshyari.com](https://daneshyari.com)