CellPress

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 entwinement 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 ubiquitious 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.

[1]School 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

[15]. Gradually, diagnostic assays have become increasingly specialized and technologically advanced, enabling finely grained classification of strains [16].

Although phenotyping schemes work well in many cases, they can be time consuming and often require specialized equipment and expertise. Classification methods that use compositional properties of each genome, such as DNA–DNA hybridization, were developed to overcome these challenges [4]. The creation of genotyping schemes, such as Rep-PCR, was further welcomed because genomes were thought to be less prone to convergence or experimental variability, and assays could be carried out at a lower cost and with less expertise required than for phenotypic methods [4]. The DNA sequencing revolution sparked innovation of multiple layers of sequence-based characterization, which ultimately focused on targeted sequencing of the 16S rRNA gene conserved throughout bacteria [17,18]. Paralleling the development of more specialized phenotyping methods, finely grained comparisons of DNA sequences, such as multilocus sequence analysis (MLSA), promised increased resolution [19]. Finally, decreased costs and increased reliability of whole-genome sequencing now provides an unprecedented amount of information for classification [3,4].

A crucial step for any formal taxonomic scheme involves some level of oversight, which today takes the form of dedicated groups of researchers that diligently curate and vet bacterial names [e.g., 20]. When a new bacterial species is proposed, vetting typically takes place by publishing a characterization of, or reference to, the type strain in the *International Journal of Systematic and Evolutionary Microbiology* [13]. Although there have been countless disagreements over what constitutes a bacterial species, it has been possible for these experts to take the time to reconcile the nomenclatural record. This system has held up quite well to scrutiny when phenotypes were of dire interest (as is the case of human pathogens), when there were numerous paid taxonomic positions whose purpose was to vet new entries into the nomenclatural arena, and when researchers could spend the time to learn 'a feel' for their organisms.

Blindspots in Microbial Taxonomy

There exists a legacy where many firmly believe that phenotypic characterization should figure prominently in bacterial taxonomy [1,3,4]. This belief has culminated in the polyphasic approach to classification and nomenclature, where phenotypic characterizations are blended with genotypic and phylogenetic information to place strains into nomenclatural groups. This reliance on phenotypes for taxonomic purposes persists even though phenotyping is impossible for certain strains, such as those that cannot be cultured [18]. When necessary, exceptions to nomenclatural rules have been created to account for outliers. For instance, when type strains cannot be cultured, a 'Ca.' is appended before the Linnean name, representing the idea that these are only candidate taxa that have not been validated according to the established rules [13,21]. Furthermore, the ease of sequencing has democratized the taxonomic process, and anyone who can perform PCR can generate new data. New 'species' are discovered by researchers who are not necessarily focused on taxonomy and have not been trained to classify these organisms. In these cases, when strains have not been fully vetted and assigned proper species names, or where there is no type strain, it is commonplace to simply append 'sp.' after the genus [13]. Lastly, with the increasing number of strains that are sequenced and cultured, the larger the number of exceptions that arise to established phenotypic rules for any particular group. What may have traditionally fit into a nice taxonomic box based on phenotyping methods may falter now, even for well-described taxa, because exceptions or complications frequently arise with greater sampling [22]. With greater scrutiny, some already established phenotyping schemes fall apart due to inherent variability of the phenotypes themselves within single strains [23].

Taxonomists continue to struggle with how to incorporate phenotypic and genotypic methods for bacterial classification and to how to name bacterial species [1,3,4,24]. This reliance on

human subjectivity builds in an inherent weakness for continuity across data sets and manuscripts. Every time a new phenotype of interest has emerged, or a lineage is split due to differences in evolutionary signals, there is a possibility that names of species or even genera change. While this practice is no doubt important, the plasticity of nomenclature presents a significant obstacle for chasing down relevant data in historical literature because information about strains of interest is often obscured by such changes. Advances in publishing technologies also work against the traditional nomenclature schemes since the proliferation of modern journals has made it easier than ever to publish papers regardless of conclusions. Reviewers' time is limited, and many may be unaware of current nomenclatural challenges, so that papers are published where strain descriptions conflict with previously agreed-upon taxonomic standards. Such situations amplify confusion in the literature because lineages are then officially referred to in multiple ways without a clear path to delineate the official taxonomic status.

The *P. syringae sensu lato* Species Complex as an Example

There is no better system to illustrate the nomenclatural challenges of modern day microbiology, and to give examples for confusion arising from the points mentioned above, than *P. syringae* [24]. Pseudomonads can be found ubiquitously across environments, but are also well known as pathogens of humans, animals, insects, fungi, and plants [25]. Therein lies one of the greatest challenges of pseudomonad taxonomy, that nomenclature within this group has often been biased by relying on phenotypes related to pathogenicity across hosts. It is possible, and even likely, that pathogenic contexts represent only a small sample of the selective pressures faced by strains in nature [26]. Therefore, phenotypes that might be of interest in defining species names may not be those that play a large role in the lifestyle of the strains and might therefore not lead to the most cohesive groupings. In a way, *Pseudomonas* is the taxonomic opposite of *H. pylori*.

The name *Pseudomonas syringae* began as a formal way to classify pseudomonad pathogens of plants, with the name 'syringae' coined because the original type strain was a pathogen of lilacs within the genus *Syringa* [24]. Eventually, the LOPAT diagnostic scheme was created to classify strains as *P. syringae* based upon five different phenotypes: levan production, oxidase, potato soft rot, arginine dihydrolase, and tobacco hypersensitivity. Species level nomenclature was layered onto this scheme, usually defined by pathogenic reactions and symptoms on host plants. Given the intrinsically changeable nature of species names, strains broadly classified within *P. syringae* are currently informally grouped into the designations *sensu lato* (the largest sense) and *sensu stricto* (the smallest sense). *P. syringae sensu lato* includes all species that could be classified as part of *P. syringae* according to the LOPAT scheme and genomic comparisons but in the absence of other phytopathogenic information. *P. syringae sensu stricto* includes only those strains that are closely related to, and phytopathogenically similar to, the original type strain. After bacterial species nomenclature was overhauled in 1980, this system was modified to include 'pathovar (pv.)' or 'pathogenic variety' status as a way to reflect previously established boundaries between strains with distinct phytopathogenic profiles [27]. Although the causes are outside the bounds of this piece, strains within *P. syringae sensu lato* have been split up into multiple species, with subsets assigned pathovar designations. Some pathovars themselves are polyphyletic due to convergence in host range (i.e., pv. *avenellae*), whereas others are tightly grouped (i.e., pv. *phaseolicola*) [28,29]. There are even some lineages known to cause disease and which are model systems for host–pathogen interactions, but which are not part of a monophyletic species cluster (i.e., *P. syringae* pv. *tomato* DC3000) [24]. Strains within the group *P. syringae* have further been categorized into nine genomospecies based upon DNA–DNA hybridization [30]. Strains have also been classified into 13 different groups by multilocus sequence type (MLST) and multiple subgroups, with a reduced yet accurate scheme including the loci *rpoD*, *gltA*, *cit*, and *gyrB* [31,32]. Whole-genome phylogenies can now be built using a variety of methods and with draft genomes, all approximate the earlier classification schemes (although there are notable exceptions) [30–35]. Given the

current state of the field, one can reasonably determine where within *P. syringae sensu lato* their strains of interest fall given a sequence for only the *gltA* locus (Figure 1, Key Figure, and [31]).

Perhaps the greatest challenge with taxonomic classification of *P. syringae* is the broad phenotypic versatility and variability that strains display. This bacterium is an environmentally ubiquitous hemibiotroph that can be isolated from environmental sources, including rivers, lakes, leaf litter, and clouds [36–38]. For any given strain, host range can be difficult to catagorize because the occurrence and severity of symptoms is dependent on environmental conditions [39,40]. Adding to the confusion, numerous strains have been isolated from environmental resources and from asymptomatic plants [22]. Many of these environmental strains are closely related to known pathogens of plants and may be capable of causing disease under laboratory conditions, but these are not given pathovar designations because they are nonpathogenic or their host range is undetermined [41,42]. To these points, the main reason that genomospecies have not been granted official taxonomic status is that there are no consistent phenotypes that differentiate them.

Since *P. syringae sensu lato* has become a model system for environmental genomics, the genome sequences of hundreds of individual strains have been deposited in publicly accessible databases [43,44]. The inconsistency of metadata across these strains makes it difficult to mine much of the information that would be useful for comparative genomic analyses. Numerous cases currently exist within Genbank of strains that could be classified in the *P. syringae* species complex that are arguably mischaracterized as an incorrect or obsolete species name. Therefore, if one is interested in accessing genomes for all of these strains, they either have to search and find which species name the genome is listed under or perform repetitive sequence-based searches to gauge how strains are related. These inconsistencies hamper research because they prevent discussion about overlapping data sets and lead to wasteful redundancy when multiple groups are unaware that they are independently investigating the same strains.

This already confusing taxonomic situation has been compounded by a variety of published papers that refer to strains with multiple species names because of gray areas in nomenclature. One of the best examples involves a strain from pathovar *phaseolicola*, which is a pathogen of green beans [28]. This strain has variably been referred to as *P. syringae* pv. *phaseolicola* or *Pseudomonas savastanoi* pv. *phaseolicola* even though it could properly be named as *Pseudomonas amygdali* pv. *phaseolicola* according to current rules [24,28,45]. There are also numerous cases where strains currently exist in taxonomic limbo as *Pseudomonas* sp. because they have not been formally described (i.e., UB246 in Figure 1) [22,31]. The widespread challenges of properly referring to strains within this species complex are demonstrated in Figure 1. I have no doubts that the situation is equally as confusing across other environmentally ubiquitous groups that contain phytopathogens (i.e., *Burkholderia, Erwinia*, and *Rhizobium* [46–48]), or across groups like *Vibrio*, where conflicts arise between traditional nomenclatural schemes and in-depth analyses of population genetics signals [49].

Divorcing Strain Classification from Species Names

The ever-increasing flood of genomic data will lead to an increase in nomenclatural confusion across taxa. New DNA sequencing technologies are continuing to emerge and mature so that, very soon, direct sequencing of nucleotides and single-cell genomics may be possible under field conditions [50]. Complete genome sequencing will eventually be cost efficient and straightforward enough to use for rapid classification across all taxa, even the uncultureable majority. Along these lines, it is worth noting that genome sequencing has potentially overtaken phage phenotyping as the preferred method for *Salmonella* classification in the United Kingdom [51]. Although phenotyping will certainly still be useful for more thorough studies, these tests will always be likely used in combination with genotypic data. Furthermore, it is inevitable that
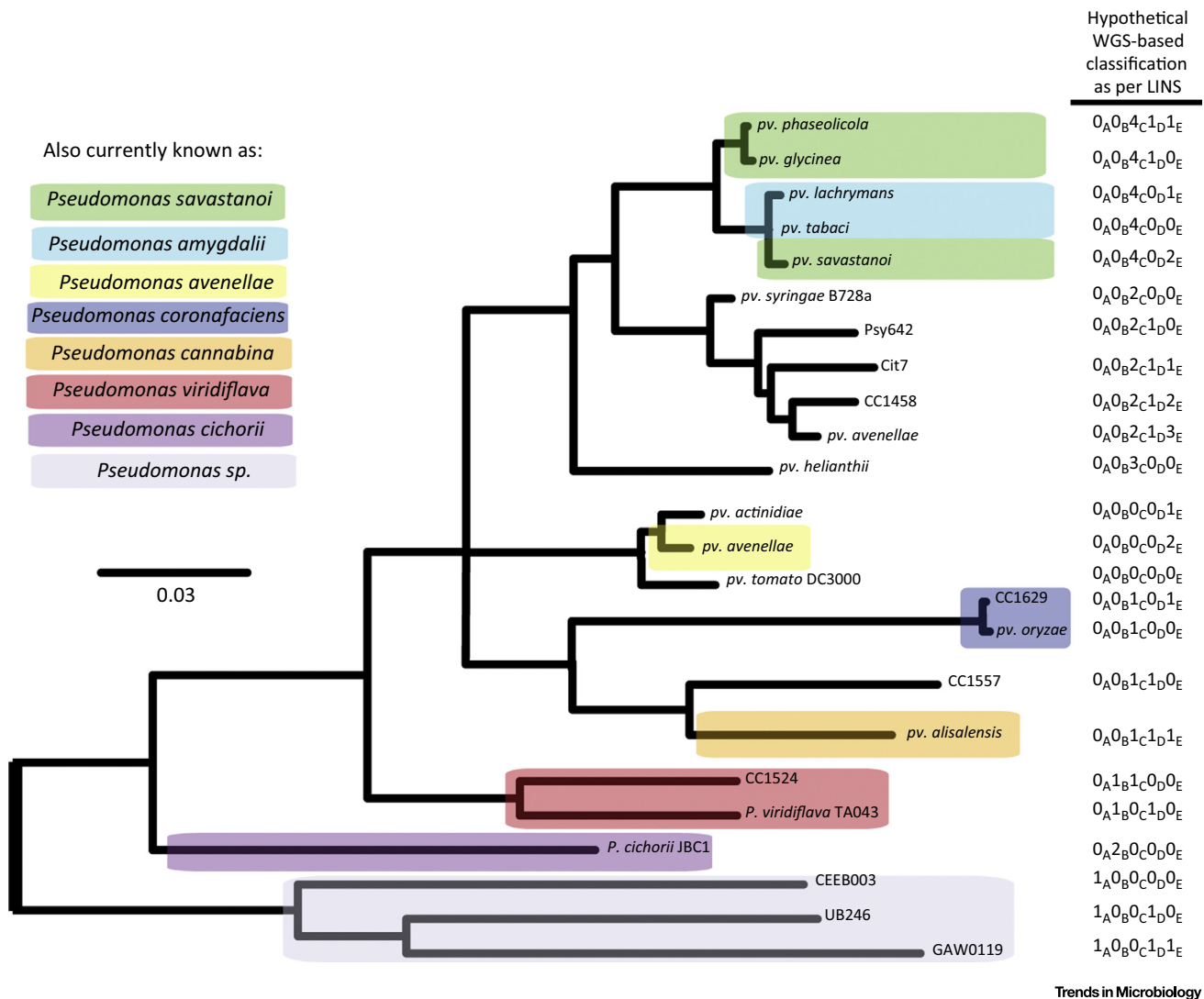
Key Figure

Current and Potential Classification of *Pseudomonas syringae sensu lato*

Figure 1. A Bayesian phylogeny was built using the full-length sequence of *gltA* from a diverse subset of *P. syringae sensu lato* genomes. Although each of these strains can be referred to as *P. syringae* in published papers and in metadata, strains that are currently referred to by species names other than *P. syringae* are highlighted in different colors. Strains which are known pathogens are labelled as either 'pv.' for pathovar or with alternative species names in the absence of pathovars. Strains which have been isolated from environmental sources or from plants in the absence of official disease classification have been labeled only with the strain number. To the right of each strain name is a mock numerical classification, based upon an example scheme described in [52], but in the absence of defined ANI thresholds or actual genome comparisons. Each column within this identifier represents an increasing threshold for ANI between that particular strain and the most similar genome already assigned a LIN. For instance, column A could represent an ANI of 40%, followed by column B at 50%, C at 75%, D at 90%, E at 99%. Possessing different numbers within these columns represents how strains differ from each other at certain threshold ANIs. Abbreviations: WGS, whole-genome sequencing; LINS, Life Identification Numbers; ANI, Average nucleotide identity.

researchers who are only casually interested in proper taxonomy will continue to submit sequences to databases and to publish papers. In this light, what can be done to minimize future confusion?

A key to clearing these challenges is the realization that taxonomic classification and nomenclature are not commutative. Although the ability to name strains inherently requires classification, classification itself can take place without the need to group strains together under potentially subjective nomenclatural umbrellas. All that is necessary for such a system is agreement as to an algorithm for classification, and the ability to identify the most closely related strain according to this scheme. Furthermore, although discussions about how to name bacterial strains inevitably deteriorate into disagreements about how to define bacterial 'species', establishment of an independent classification scheme would separate these discussions and disagreements from the actual indexing of these strains in practice for posterity.

There will never be a perfect classification system, but we can develop a scheme based on whole-genome comparisons which ensures continuity of research messages across databases and papers. To achieve any level of success this scheme must be relational in that new designations are assigned solely through genotypic similarity to previously indexed strains, whether or not they have been formally characterized as a type strain, and it must exist independently of disagreements about species concepts. An added benefit of relational classification is that identifiers provide information about relationships across strains and species. This system must be expandable so that discoveries of new clades do not force a complete rewriting of previous classifications, and must be applicable to uncultured strains. Classification by this system must be automatable and function well regardless of the level of taxonomic expertise, and could even be retroactively applied across all published genome sequences and manuscripts. Ideally, these classifications would already be implementable given the infrastructure of Genbank and could be added to keywords of journal articles to streamline search strings for text-mining software.

Numerous options for such a classification scheme already exist, they just need to be universally implemented. MLSA/MLST or 16S rRNA schemes work well and reflect phylogenetic relationships, but can be confounded by recombination, can be biased in the subsets of loci used, and ultimately contain a lower number of informative sites than whole genomes [3,4,18,19]. Moreover, assigning group IDs based on static numbers significantly decreases the ability to efficiently add intermediate taxa and limits the capability to provide information about strain relatedness through relational classification. For instance, MLST group 2 in *P. syringae* has already been expanded to include MLST groups 2a/2b/2c/2d [31], but what would happen if group 2a were further split? Likewise, *P. syringae* MLST group 10 is more closely related to group 5 than to group 9 according to the current typing scheme [31]. Life Identification Numbers (LINS) have been suggested as an alternative scheme that fits this bill, whereby the average nucleotide identity (ANI) of genomes of interest are sequentially compared [52,53]. After one genome is designated as a reference point (denoted by inclusion of 0s at all positions in its identifier), new genomes are compared and assigned standard numerical classifications based on overall genome similarity to the most closely related genome already assigned a LIN. Each position within the LINS identifier (denoted by subscript letters as in Figure 1) represents an increasing threshold of similarity for ANI between two strains. According to this scheme, strains that differ in number at position 'A' have a much greater divergence in ANI than strains that differ in number at position 'E'. Assigning identifiers to new genomes is akin to a first-order Markov process, because classification is memoryless and based solely on the most similar genome already classified. In order to assign LIN values to a strain of interest one only needs to identify the most closely related previously categorized strain and compare ANI values. Due to the

sequential nature of this scheme, increasing thresholds can be utilized where necessary in order to delineate closely related taxa, which would be manifest as additional positions added to the identifier (i.e., $0_A,0_B,0_C,0_D,0_E,1_F/0_A,0_B,0_C,0_D,0_E,0_F$ instead of $0_A,0_B,0_C,0_D/0_A,0_B,0_C,0_D$). Firm borders between species groups need not exist because schemes like LINS provide progressive classification where one simply includes more finely grained comparisons of genome similarity if necessary. LINS is a computational example of the goal of moving to classification based on DNA hybridization, except that classification can be automated and take a fraction of the time and expertise. Notably, this system has been applied to strains within *P. syringae* (http://www.apsnet.org/meetings/Documents/2015_meeting_abstracts/aps2015abP516.htm).

Some have highlighted the weaknesses of classification solely through genome sequences [1,3,4,54]. Substantive challenges include instances where whole-genome data do not exist, but strains can be partially assigned based on a limited set of genomic characteristics (such as MLST or 16S rRNA loci) where necessary. Furthermore, government agencies currently rely upon species classification to control the movement and distribution of microorganisms, and shifting classification could exacerbate existing regulatory challenges. I remain hopeful that such issues will be smoothed over by overlaying nomenclatural information onto whole-genome classifications, so that agencies can continue to rely upon species names. There will also be challenges for strains that have experienced extensive genomic diversification, either through selection or drift, as is the case with genome minimization across many intracellular parasites [55]. Lastly, many researchers have become comfortable with, and have grown accustomed to, assigning binomial names to new microbes. Numerical codes are a less elegant way to describe new taxa and will be inherently unpalatable to some, regardless of whether this is implemented in parallel with the binomial system.

Concluding Remarks

The overall message of this piece is not to throw out all previous taxonomic systems and start anew, but that we must move to implement a sequence-based classification system that is unambiguous. We can create a retroactive and expandable system that could be used by regulatory agencies, in publication keywords, and with metadata that exists independently of species nomenclature or concepts. These systems can be expandable, with algorithms that can be created to automatically classify or group genomes by similarity when they are submitted to databases or to journals. This suggestion parallels both the call for universal DNA Barcoding across species [56] as well as the creation of ORCID identifiers (orcid.org) to trace publications by individual authors. Researchers from across disciplines will likely admit that something has to be done to change the way strains are classified, challenges which are exemplified quite well by *P. syringae sensu lato* (Figure 1). There may not be one single answer to the problem of nomenclatural confusion in microbes, but today's challenges will only worsen unless we rethink taxonomic strategies (see Outstanding Questions).

References

1. Oren, A. and Garrity, G.M. (2013) Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie Van Leeuwenhoek* 106, 43–56
2. Shapiro, B.J. and Polz, M.F. (2015) Microbial speciation. *Cold Spring Harb. Perspect. Biol.* 7, 1–22
3. Zhi, X-Y. *et al.* (2011) Prokaryotic systematics in the genomics era. *Antonie Van Leeuwenhoek* 101, 21–34
4. Bull, C.T. and Koike, S.T. (2015) practical benefits of knowing the enemy: modern molecular tools for diagnosing the etiology of bacterial diseases and understanding the taxonomy and diversity of plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* 53, 157–180
5. Blaby-Haas, C.E. and de Crécy-Lagard, V. (2011) Mining high-throughput experimental data to link gene and function. *Trends Biotechnol.* 29, 174–182
6. Salhi, A. *et al.* (2016) DESM: portal for microbial knowledge exploration systems. *Nucleic Acids Res.* 44, D624–D633
7. Linz, B. and Schuster, S.C. (2007) Genomic diversity in *Helicobacter* and related organisms. *Res. Microbiol.* 158, 737–744
8. Baltrus, D. *et al.* (2009) *Helicobacter pylori* genome plasticity. *Genome Dynamics* 6, 75–90
9. Correa, P. and Piazuelo, M.B. (2008) Natural history of *Helicobacter pylori* infection. *Dig. Liver Dis.* 40, 490–496

Outstanding Questions

What is the most efficient method for whole-genome classification?

Who is in charge of establishing benchmarks for or curating whole-genome-based classification?

Is it possible to retroactively label, in keywords of articles or with genomic databases, all bacterial taxa within one numerical classification scheme?

Can a whole-genome-based classification system account for taxa where only partial data exist, such as 16S rRNA gene sequences?

Will a numerical, non-Linnean classification system be widely accepted by microbiologists?

Would regulatory agencies and non-specialists, such as doctors, transition from traditional nomenclature to a newly implemented system?

10. Solnick, J.V. *et al.* (2006) The genus *Helicobacter*. In *The Prokaryotes* (vol. 7) (Dworkin, M. *et al.*, eds), In pp. 139–177, Springer

11. Welch, R.A. (2006) The genus *Escherichia*. In *The Prokaryotes* (vol. 6) (Dworkin, M. *et al.*, eds), In pp. 60–71, Springer

12. Sahl, J.W. *et al.* (2015) Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J. Clin. Microbiol.* 53, 951–960

13. Lapage, S.P. *et al.* (1992) *International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision,* ASM Press

14. Chui, H. *et al.* (2015) Rapid, sensitive and specific *E. coli* H antigen typing by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)-based peptide mass fingerprinting. *J. Clin. Microbiol.* 53, 2480–2485

15. Baggesen, D.L. *et al.* (2010) Phage typing of *Salmonella Typhimurium* – is it still a useful tool for surveillance and outbreak investigation? *Euro. Surveill.* 15, 1–3

16. Fournier, P-E. *et al.* (2015) From culturomics to taxonomogenomics: A need to change the taxonomy of prokaryotes in clinical microbiology. *Anaerobe* 36, 73–78

17. Tringe, S.G. and Hugenholtz, P. (2008) A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.* 11, 442–446

18. Yarza, P. *et al.* (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645

19. Almeida, N.F. *et al.* (2010) PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. *Phytopathology* 100, 208–215

20. Bull, C.T. *et al.* (2010) Comprehensive list of names of plant pathogenic bacteria, 1980-2007. *J. Plant Pathol.* 92, 551–592

21. Murray, R.G. and Stackebrandt, E. (1995) Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes. *Int. J. Syst. Bacteriol.* 45, 186–187

22. Diallo, M.D. *et al.* (2012) *Pseudomonas syringae* naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic. *ISME J.* 6, 1325–1335

23. Bartoli, C. *et al.* (2014) The *Pseudomonas viridiflava* phylogroups in the *P. syringae* species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. *Environ. Microbiol.* 16, 2301–2315

24. Young, J.M. (2010) Taxonomy of *Pseudomonas syringae*. *J. Plant Pathol.* 92, S1.5–S1.14

25. Peix, A. *et al.* (2009) Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect. Genet. Evol.* 9, 1132–1147

26. Morris, C.E. *et al.* (2008) The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J.* 2, 321–334

27. Young, J.M. *et al.* (1992) Changing concepts in the taxonomy of plant pathogenic bacteria. *Annu. Rev. Phytopathol.* 30, 67–105

28. Arnold, D.L. *et al.* (2011) *Pseudomonas syringae* pv. *phaseolicola*: from "has bean" to supermodel. *Mol. Plant Pathol.* 12, 617–627

29. Marcelletti, S. and Scortichini, M. (2015) Comparative genomic analyses of multiple *Pseudomonas* strains infecting *Corylus avellana* trees reveal the occurrence of two genetic clusters with both common and distinctive virulence and fitness traits. *PLoS ONE* 10, e0131112

30. Gardan, L. *et al.* (1999) DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas cannabina* sp. Nov. (ex Sutic and Dowson 1959). *Int. J. Syst. Bacteriol.* 49, 469–478

31. Berge, O. *et al.* (2014) A user's guide to a data base of the diversity of *Pseudomonas syringae* and its application to classifying strains in this phylogenetic Complex. *PLoS ONE* 9, e105547

32. Hwang, M.S.H. *et al.* (2005) Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 71, 5182–5191

33. Baltrus, D.A. *et al.* (2011) Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog.* 7, e1002132

34. Baltrus, D.A. *et al.* (2014) Ecological genomics of *Pseudomonas syringae*. In *Genomics of Plant-Associated Bacteria* (Gross, D.C. *et al.*, eds), pp. 59–77, Springer

35. Baltrus, D.A. *et al.* (2014) Incongruence between multi-locus sequence analysis (MLSA) and whole-genome-based phylogenies: *Pseudomonas syringae* pathovar *pisi* as a cautionary tale. *Mol. Plant Pathol.* 15, 461–465

36. Hirano, S. and Upper, C. (2000) Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae* – a pathogen, ice nucleus, and epiphyte. *Microbiol. Mol. Biol. Rev.* 64, 624

37. Morris, C.E. *et al.* (2009) Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. *PLoS Pathog.* 5, e1000693

38. Morris, C.E. *et al.* (2010) Inferring the evolutionary history of the plant pathogen *Pseudomonas syringae* from its biogeography in headwaters of rivers in North America, Europe, and New Zealand. *MBio* 1, e00107–10–e00107–20

39. Mazarei, M. and Kerr, A. (1990) Distinguishing pathovars of *Pseudomonas syringae* on peas: nutritional, pathogenicity and serological tests. *Plant Pathol.* 39, 278–285

40. Elvira-Recuenco, M. *et al.* (2003) Differential responses to pea bacterial blight in stems, leaves and pods under glasshouse and field conditions. *Eur. J. Plant Pathol.* 109, 555–564

41. Bartoli, C. *et al.* (2014) A framework to gauge the epidemic potential of plant pathogens in environmental reservoirs: the example of kiwifruit canker. *Mol. Plant Pathol.* 16, 137–149

42. Mohr, T.J. *et al.* (2008) Naturally occurring nonpathogenic isolates of the plant pathogen *Pseudomonas syringae* lack a type III secretion system and effector gene orthologues. *J. Bacteriol.* 190, 2858–2870

43. O'Brien, H.E. *et al.* (2011) Evolution of plant pathogenesis in *Pseudomonas syringae*: a genomics perspective. *Annu. Rev. Phytopathol.* 49, 269–289

44. McCann, H.C. *et al.* (2013) Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. *PLoS Pathog.* 9, e1003503

45. Ramos, C. *et al.* (2012) *Pseudomonas savastanoi* pv. *savastanoi*: some like it knot. *Mol. Plant Pathol.* 13, 998–1009

46. Pritchard, L. *et al.* (2016) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24

47. Ormeño-Orrillo, E. *et al.* (2015) Taxonomy of *Rhizobia* and *Agrobacteria* from the *Rhizobiaceae* family in light of genomics. *Syst. Appl. Microbiol.* 38, 287–291

48. Gupta, R.S. (2014) Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front. Genet.* 5, 429

49. Preheim, S.P. *et al.* (2011) Merging taxonomy with ecological population prediction in a case study of *Vibrionaceae*. *Appl. Environ. Microbiol.* 77, 7195–7206

50. Loman, N.J. and Pallen, M.J. (2015) Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* 13, 787–794

51. Ashton, P. *et al.* (2015) Revolutionising public health reference microbiology using whole genome sequencing, *Salmonella* as an exemplar. *bioRxiv* Published online November 29, 2015. http://dx.doi.org/10.1101/033225

52. Marakeby, H. *et al.* (2014) A system to automatically classify and name any individual genome-sequenced organism independently of current biological classification and nomenclature. *PLoS ONE* 9, e89142

53. Weisberg, A.J. *et al.* (2015) Similarity-based codes sequentially assigned to ebolavirus genomes are informative of species membership, associated outbreaks, and transmission chains. *Open Forum Infect. Dis.* 2, ofv024

54. Staley, J.T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 361, 1899–1909

55. Klasson, L. (2004) Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* 12, 37–43

56. Goldstein, P.Z. and DeSalle, R. (2010) Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *Bioessays* 33, 135–147