ORIGINAL ARTICLE

Relevance of prokaryotic subspecies in the age of genomics

S. N. Venter[1], **M. Palmer**[1,2] **and E. T. Steenkamp**[1]

1) *Department of Biochemistry, Genetics and Microbiology and Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa and* 2) *School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV, USA*

Abstract

The availability of multiple gene sequences, and in particular full genome sequence data, for microbial strains has changed how taxonomists delineate subspecies belonging to the Archaea and Bacteria. Well-defined phylogenetic lineages that share higher genome similarity values compared to the widely used species thresholds are often described as subspecies, despite clear evidence of genetic isolation between them. These well-defined lineages, reflecting notable genetic isolation of the core genome represent more recently evolved, unique and *sui generis* evolutionary units. Because they bear all of the hallmarks of species, most contemporary subspecies likely represent species in their own right. Although there is considerable value in defining intraspecies variation (e.g., pathovar, serovar and symbiovar), the discriminating properties of such units are mostly encoded on accessory subgenomic compartments. We therefore argue that the taxonomic category of subspecies has become irrelevant and propose that its use should be discontinued. This will minimize inconsistencies related to the subjective nature of species-subspecies distinctions. Formal recognition of biologically relevant variation within species based on the accessory genome information will have practical significance in fields such as clinical, industrial and agricultural microbiology.

Keywords: Accessory genome, core genome, genetic isolation, Prokaryotic taxonomy, species, subspecies, varieties
Original Submission: 17 December 2021; **Revised Submission:** 13 August 2022; **Accepted:** 22 August 2022
Article published online: 30 August 2022

Corresponding author. SN Venter, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Private Bag X 20, Hatfield, Pretoria, 0028, South Africa.
E-mail: fanus.venter@up.ac.za

Introduction

The use of subspecies as taxonomic category is not limited to prokaryotes. In ornithology, for example, it is often applied to geographically or morphologically distinct populations [1,2]. But as with prokaryotes, the nature of species is widely debated and sometimes controversial [2,3]. To resolve the issue in herpetological taxonomy, De Queiroz [3] defined subspecies as incompletely separated lineages where some inter-lineage gene exchange still exists, which is an extension of his general lineage concept of species as portions of "*separately evolving metapopulation lineages*" [4]. According to De Queiroz's definition, subspecies are units that are not completely isolated in terms of reproduction and overall genetics, a notion that is fully compatible with the semipermeable nature of species boundaries in eukaryotes [5,6]. De Queiroz therefore stressed that in evolutionary terms, subspecies should not be viewed as something less than a species as they have all the hallmarks of species units.

Due to the wide prevalence of horizontal gene transfer (HGT), prokaryotic species boundaries are often many orders of magnitude more permeable than those of most eukaryotes [7–10]. In fact, a large proportion of the genes encoded by a prokaryotic species may be dispensable, the number and identity of which can vary substantially among strains and populations of a species [11]. The flexible, HGT-prone fraction of species' genomes are referred to as the accessory genome or subgenomic compartment, while the stable fraction is denoted as the core genome [12,13]. It is this latter, well-conserved, subgenomic compartment that carries and encodes the properties used by taxonomists to delineate species [14–16].

https://doi.org/10.1016/j.nmni.2022.101024

However, no attempts have been made to reconcile prevailing interpretations of the subspecies category with currently accepted species definitions, especially in light of recent developments in taxonomic practice and the theory underpinning it.

In this article, we consider the nature and value of subspecies as taxonomic category for prokaryotes by making use of contemporary genome-based evidence. To do this, we first describe how species and subspecies are currently defined, and then discuss how genomics data have impacted our view of these taxonomic units. We then present an argument for discontinuing the use of subspecies as a defined taxonomic unit, after which we discuss the value of recognizing varieties to denote intraspecies variation as a viable and functional alternative. We conclude by highlighting how the description of varieties would align with current taxonomic practice and how it would complement and enrich existing taxonomic frameworks, thereby enhancing their value to end-users, while at the same time reflecting their nature more realistically.

What are species and subspecies?

The naming of species and subspecies is governed by the International Code of Nomenclature of Prokaryotes (ICNP). Names of these taxa are recognized as validly published when they meet all the requirements of ICNP Rules 30 to 32 [17]. Based on the original version of the Bacteriological Code [18], phenotypic variants within species could be described as either varieties or subspecies. Subspecies were recommended if the variation was sufficiently distinct and stable. This nomenclature has since changed as Rule 5c of the ICNP clearly states that a variety should be treated as a synonym of subspecies. However, species and subspecies may be delineated using a diverse set of approaches as the ICNP does not oversee the process nor restricts "the freedom of taxonomic thought or action" [17].

Most microbiologists regard an archaeal or bacterial species as a genotypically and phenotypically coherent cluster of isolates with a binomial name that enables unambiguous communication and sharing of information related to the taxon. Accordingly, all of the widely used criteria or species definitions employed for recognizing species revolves around the idea that members of a species are monophyletic and cluster together based on phenotypic and genomic similarities [19]. Of these, the level of phenotypic coherence is often most difficult to determine, but usually includes shared physiological and ecological properties. Phylogenetic coherence is typically determined using sequence information for multiple genes, and most contemporary analyses of genomic coherence are based on the proportion and similarity of shared genes [20,21]. In practice, however, taxonomic decisions are based primarily on genomic coherence where a quantitative threshold or cut-off value of 95% Average Nucleotide Identity (ANI) is used to define species [22]. The latter largely corresponds to 70% DNA/DNA hybridization (DDH), estimated experimentally or using digital DDH (dDDH) [23,24].

Subspecies are identified by a trinomial name, with *Lactobacillus salivarius* being one of the earliest species for which varieties or subspecies were formally used to differentiate metabolically distinct strains [25]. Since then numerous traits have been used to recognize subspecies (see Box 1) ranging from clinical, pathological and physiological properties through to phylogenetic monophyly [26]. The latter was introduced in 1987 by an *ad hoc* committee of the then International Committee on Systematic Bacteriology [24], which subsequently led to the development of numerous multi-locus sequence typing and analysis (MLST and MLSA) schemes for delineating subspecies [27–31]. As a result, phenotype combined with monophyly has become the most important characteristic used for the delineation of subspecies [32]. In July 2022, there were

BOX 1.

Subspecies in the *Klebsiella pneumoniae* complex

The *Klebsiella pneumoniae* complex is a common cause of nosocomial infections and antimicrobial resistant strains are considered to be a critical health threat [66]. The complex includes a range of populations associated with particular hosts and environmental niches [67]. Through the years, authors have treated its taxonomy differently and the complex thus provides an excellent example of how the delineation of varieties and subspecies changed over time. Based on early work from the 19th century, three subspecies of *K. pneumoniae* sensu stricto were formally described in the first edition of Bergey's Manual of Systematic Bacteriology [68]. The main methods of differentiation among these taxa were a set of phenotypic characteristics and their clinical manifestations. Subsequent phylogenetic and phylogenomic studies failed to provide sufficient evidence for distinguishing among *K. pneumoniae* subsp. *pneumoniae*, *K. pneumoniae* subsp. *ozaenae* and *K. pneumoniae* subsp. *rhinoscleromatis* [69] and the use of these subspecies has mainly fallen into disuse. By contrast, subspecies of two other species in the complex have been delineated using robust phylogenetic inferences (i.e., *K. quasipneumoniae* and *K. variicola*). Despite the limited phenotypic differences among the respective subspecies, they can easily be differentiated based on genetic isolation and limited gene flow of genes associated with the core genome (stable sequence differences) [66,70] and the accessory genomes [71].

		Distinguishable taxonomic units of *Klebsiella pneumoniae* complex?			
		K. pneumoniae subspecies	*K. quasipneumoniae* subspecies	*K. variicola* subspecies	Species of the *K. pneumoniae* complex
Distinguishing criteria	Phenotypic cohesion	Yes	Yes	Yes	Yes
	Phylogenetically distinguishable clades (monophyly)	No	Yes	Yes	Yes
	Notable genetic isolation of the core genome	No	Yes	Yes	Yes
		97.56 – 98.41 %	*96.27 – 96.33 %*	*96.38 – 96.65 %*	*92.36 – 94.86 %*
		ANI values between type strain of subspecies and type strain of the species			

871 validly named subspecies, of which 454 were considered as "correctly named" based on the most recent taxonomic opinion [33]. An excellent example illustrating the diversity of types of groups being recognized as subspecies is provided by the *Klebsiella pneumoniae* complex (Box 1).

Genomics have impacted our understanding of the basic taxonomic unit

The increased availability of whole genome sequences has contributed enormously to the systematics of prokaryotes. It allowed for the development of tools and procedures to streamline species recognition, which is evidenced by the range of overall genome related indexes (OGRI) that have been developed and implemented by taxonomists [15,21,34]. Also, the availability of genome data for species and even populations of species have facilitated an improved understanding of the nature of species in Archaea and Bacteria [7,8,21,35–37]. In other words, wide access to whole genome data for these taxa not only mediated improvements in taxonomic practice, but also stimulated research into the theory underlying prokaryotic species evolution.

From a practical point of view, genomics studies revealed that the distinction between species and subspecies as taxonomic units is not straightforward. For example, extensive genome comparisons showed that species boundaries likely fall in the range of 93-96 % ANI [20], which corresponds well with the previous "gold standard" of 70% DDH [38]. However, this ANI range represents a "fuzzy zone" because classification of strains into separate species depends largely on the interpretation of the taxonomist [20]. For example, certain taxonomists use 95% ANI as species threshold [22,39,40], while others designate strains with > 93% ANI as members of distinct genomovars, i.e., genome-based groups sufficiently distinct to be recognized as separate species, but lacking phenotypic differences for unambiguously differentiating them [20]. The latter could thus also be regarded as subspecies, which are comparable to species, but at lower taxonomic rank due to their strains sharing high genome-based similarity [41]. This would also be in line with proposals that dDDH values between 70% and 80% are used as a quantitative approach for the delineation of subspecies [41]. Indeed, strains belonging to a well-defined phylogenetic lineage, but with ANI/DDH/dDDH values exceeding the suggested species thresholds, are now often described as subspecies [42–44]. In some cases, species have even been lowered in rank to subspecies following the implementation of OGRI measurements [45,46].

A theoretical basis for distinguishing between species and subspecies of prokaryotes has also remained elusive. Current models hold that prokaryotic diversity represent speciation spectra [9] on which gene flow barriers, together with drift and natural selection, lead to the formation and maintenance of discrete and cohesive units [47]. As such, each of these evolutionary units is unique and *sui generis* ("of its own kind") in nature [7,8]. In other words, they are produced by distinct evolutionary processes and may differ markedly in terms of intra-unit sequence similarity, population size and evolutionary

age, as well as ecological and phenotypic characteristics [7–9]. Whether these units are recognized and described as species or as subspecies depends entirely on the taxonomist's view, because other than convention (e.g., associated with particular taxa or methodologies) a theoretical framework for such decisions are lacking.

The definitions of most contemporary subspecies, especially where genome-based evidence was used, almost exactly match those used for defining species [19]. They represent monophyletic units that are phenotypically and genomically coherent, and potentially only differ from *bona fide* species in having more recent evolutionary origins, smaller population sizes, and in spanning less phenotypic and genotypic diversity. However, we know that all of these properties are intrinsically variable among the units we recognize as species due to the unique evolutionary trajectories that gave rise to them [8]. How then can one objectively justify why strains belonging to well-defined phylogenetic clusters, with OGRI values higher than the usual species thresholds, should or should not be described as species or subspecies? Hence, we argue that subspecies as a taxonomic category has become irrelevant and of limited value to users of the taxonomic frameworks we establish.

Recognizing varieties within species has practical value

Various aspects of the ecology of prokaryotes depend on functions encoded by their accessory genome. The gene content of this subgenomic compartment largely reflects the response of strains to niche exploration, diversification and their adaptation to ecological changes [48]. This is because accessory genomes are the product of differential gene loss and conservation, together with gene gains from other genomes in the immediate environment via HGT. Although homologous recombination and neutral acquisitions of genes occur, adaptation to various selection pressures is seen as an important driver that shapes the accessory genome of a strain [49]. The accessory genome therefor provides a crucial link to observable intraspecies variation and seldomly exhibits the broad level of genetic isolation associated with species (Fig. 1).

Practically, there is immense importance in recognizing biologically relevant variation within species based on specific operational attributes or phenotypes linked to accessory genomes. For this reason, the pangenomes of human, animals, and plants are a subject of active research [50,51]. Among human

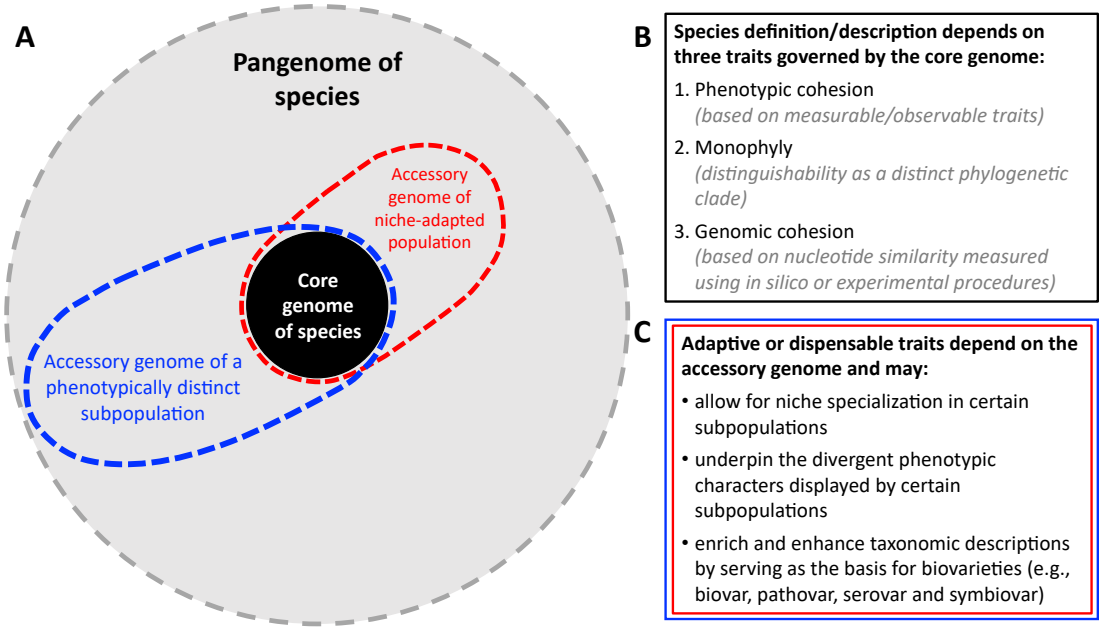

FIG. 1. The existence of species is governed by properties of core genomes, while varieties within species are mostly dictated by accessory genome features. A: The genomes of prokaryotic species are typically made-up of core and accessory subgenomic compartments, which together form the pangenome of the species [12,13]. While being genetically cohesive with other members of the species based on core genome sequences, individual strains/populations may have substantially diverged accessory genomes, which can cause them to be phenotypically or ecologically distinct (represented by the areas enclosed with red and blue dotted lines). Consequently, species definitions/descriptions are determined by traits governed by the core genome (B), while dispensable traits are determined by the accessory genome (C).

and animal pathogens, the focus has been on studying and understanding varieties and variations associated with host specificity, antigenic properties, as well as disease symptoms or clinical presentation, leading to the description of pathovar/pathotype and serovar/serotype, for example (Table 1). Similar accessory genome-derived units are also used within rhizobial species denoting symbiovars associated with particular legume hosts (Table 1). Defining subpopulations or varieties within species using these designations are only important if they convey useful and important biological information for specific user groups such as clinicians or environmental microbiologists with an interest in specific traits linked to the accessory genome. Varieties may therefore have far greater utility and practical relevance compared to many of the currently recognized subspecies, especially those that do not display any biological differences of relevance. Use of such varieties is already well embedded in the scientific literature, where the approach implemented depends on practical needs.

The recognition process for varieties associated with a unique set of accessory genes is flexible (Fig. 2). Also, the same strain may be grouped with a different set of strains depending on the variety of interest and the practical needs of different research fields. For example, the same pathovar could be linked to strains belonging to different serovars [52] or vice versa [53]. It does not require, as in the case of subspecies, that all strains in a species belong to one of the groups defining a variety (e.g. pathogenic *vs.* commensal *E. coli*). It also supports the use of typing schemes critical for epidemiological studies of pathogenic bacteria [54,55]. Another practical benefit is that variety-based groupings are transferable across species when their genetic determinants are subject to interspecies HGT [56,57].

What are the implications of abandoning subspecies as taxonomic category and promoting the use of varieties?

Discontinuing the use of subspecies as category, combined with formal recognition of diagnosable varieties within species, will enhance prokaryote taxonomy among its users (e.g., clinicians and plant pathologists). The reason for this is two-fold: (i) it will reduce taxonomic confusion and instability related to the subjective nature of species-subspecies distinctions; and (ii) the introduction of named varieties within species would have direct practical value. With regards to abandoning the subspecies category, it is well-known that confusion is caused among the users of taxonomy when important groups of organisms are not robustly delineated and/or when their naming conventions are illogical [58,59]. Such issues could also lead to the non-use or disregard of the taxa described by prokaryote taxonomists. Assigning a group of interest to a category (i.e., subspecies) that is essentially indistinguishable from another (i.e., species) would be a good example of such a practice. In terms of the pragmatism related to assigning varieties within species, designations such as pathovar, biovar or symbiovar would complement species descriptions, making their taxonomy more user-friendly and information-rich. These varieties may include any of those listed in Appendix 10 of the ICNP

TABLE 1. Illustrative examples of species/genera that contain varieties for denoting distinct clusters in which the diverged traits are governed by information encoded on the accessory genome

Variety recognized	Informative trait	Species or genera[a]	References
Pathovar	Pathological distinguishability by causing the development of distinctive symptoms on one or more plant or animal host	*Escherichia coli*, *Pseudomonas syringae**, *Xanthomonas hortorum**	[72–74]
Symbiovar	Nodulation and establishment of the nitrogen-fixing symbiosis with the same legume host, often independently of species affiliation	*Bradyrhizobium*, *Mesorhizobium*, *Paraburkholderia*, *Rhizobium*	[56,75–77]
Serovar	Serological distinguishability due to the presence of similar cell surface antigens (in certain cases, independent of species affiliations)	*Leptospira*, *Listeria monocytogenes*, *Salmonella enterica*	[57,78,79]
Morphovar	Morphological characters in culture	*Mycobacterium tuberculosis* complex*	[80]
Biovar	Various:		
	- Single physiological/biochemical character	*Bacillus cereus* sensu lato, *Lactococcus lactis*	[59,81]
	- Sets of physiological/biochemical properties	*Campylobacter sputorum*, *Corynebacterium diphtheriae*	[82,83]
	- Host association	*Corynebacterium pseudotuberculosis*	[84]

[a]Genome-based evidence showed that determinants for the respective traits are encoded by dispensable/accessory genes. In cases where the molecular basis of traits is yet unknown (indicated with *), genome and phylogenetic data together indicated that the traits' genetic determinants are not encoded by the core genome, and subject to HGT, as is typical for accessory genes.

	Hierarchy	Delineation methods	Taxon Hallmarks
Core genome derived	**Genus**	Sequence (e.g., for 16S rRNA gene or the whole genome) similarity searches against public domain databases	An arbitrarily defined monophyletic taxon above species level
Core genome derived	**Species**	Polyphasic schemes employing methodologies based on ➢ phenotypic characterization *(e.g., conventional culture- or metabolomics-based assays)* ➢ sets of multiple core genes *(e.g., MLST, MLSA, phylotyping, genealogical concordance)* ➢ whole genome information *(e.g., phylogenomics, conventional DDH, OGRI)*	Monophyletic group of prokaryotic strains, characterized by high levels of phenotypic and genomic coherence
Accessory genome derived	**Variety** (*optional*)	Characterization of one or more ecological, physiological or biochemical traits, which can include assays based on ➢ host associations *(e.g., serology, pathogenicity/virulence, symbiosis, genome-informed sequence typing schemes)* ➢ physiological/biochemical traits *(e.g., enzyme and compound/metabolite production, resistance to compounds/metabolites)* ➢ cellular/cultural characteristics *(e.g., colony morphology; growth in specific conditions)*	A group of strains within a species displaying distinct properties Groups may overlap, as membership to more than one type of group* is possible * e.g., biovar, chemovar, morphovar, pathovar, phagovar, serovar, symbiovar
	Examples: *Leptospira interrogans serovar* Grippotyphosa; *Mesorhizobium loti* symbiovar loti; *Salmonella enterica* serovar Gallinarum biovar Pullorum		

FIG. 2. The core and accessory genomes of prokaryotes inform the demarcation of taxonomic units at different levels. Genus and species units are delineated using properties inherent to the core genome, while varieties are recognized based on traits linked to the accessory genome. Because of the prevalence of HGT, three examples are indicated for cases where variety designations are transferrable across species (serovars of *Leptospira* and symbiovars of *Mesorhizobium*) [56,57], and where strains form part of multiple variety forms/types (*Salmonella* serovars and biovars) [78].

[17], or any other types or forms distinguishing groups within species. Designating varieties in this way further allows clear communication that their distinguishing properties are likely encoded on the species' accessory genomes (which is not the case for "subspecies" whose existence are dictated by core genome information).

Taxonomic implications associated with abandoning the subspecies for prokaryotic species would not be insurmountable and would provide a better reflection of current knowledge of genome organisation. Many of the subspecies described using DNA-based information, especially whole genome sequences, would need to be elevated to species-level as they represent phenotypically and genomically coherent monophyletic units, albeit with more recent evolutionary origins than their contemporaries (Fig. 2). Using this proposal for reviewing the *Klebsiella pneumoniae* complex (Box 1) for example, will result in the abandonment of *K. pneumoniae* subspecies. It would only be possible to distinguish the current subspecies as biovars based on phenotypic differences. Their specific clinical manifestations are not conclusive to justify pathovar designations [60]. At the same time the *K. quasipneumoniae* and *K. variicola* subspecies would be recognised as distinct species in their own right.

Abandoning the subspecies category also aligns well with the Genome Taxonomy Database (GTDB; https://gtdb.ecogenomic.org) where the demarcation of subspecies has been discontinued [61]. The GTDB's exclusion of subspecies could potentially have far-reaching significance because its classification system is utilized as the taxonomic framework for the current edition of Bergey's Manual of Systematics of Archaea and Bacteria [62]. Our approach of distinguishing between core and accessory genomes when demarcating taxa and varieties also resonates with the newly proposed SeqCode [63,64] where high quality metagenome assembled genomes (MAGs) will be acknowledged as suitable permanent types for the description of prokaryotic species. However, to ensure consistency with the ICNP and to accommodate community feedback during the SeqCode's development, this new taxonomic initiative also allows for the description of subspecies. Abandonment of the subspecies category would thus require amendments to both Codes.

Genomics provided us with new insights into the evolution and structure of microbial populations, their genomes and the species they belong to. It has also given us the ability to propose a more uniform classification system that will address the needs of end-users, especially those interested in specific traits such as

pathogenicity and host specificity which are often shared through HGT. The recognition of intraspecies varieties thus has direct practical value to clinicians, pathologists and industrial microbiologists. From an ecological point of view, delineation of groups within species allows for the contextualization of intraspecies variation relative to the microbial communities within which they appear [65]. Our proposed use of core genome data to define species and the use of accessory derived traits to define intraspecies variation will bring stability to the naming of biologically relevant units and taxa, one of the main principles of the ICNP [17], while at the same time also enhancing taxonomy's value to its end-users.

Conflict of interest

None.

Author contribution

Stephanus Venter: Conceptualization, Investigation, Writing – original draft. preparation, Writing – review & editing, Project administration. Emma Steenkamp: Conceptualization, Writing – original draft. preparation, Writing – review & editing, Visualization. Marike Palmer: Conceptualization, Formal analysis, Writing – review & editing

References

[1] Mayr E. Of what use are subspecies? The Auk 1982;99(3):593–5.
[2] O'Neill JP. The subspecies concept in the 1980's. The Auk 1982;99(3):609–12.
[3] de Queiroz K. An updated concept of subspecies resolves a dispute about the taxonomy of incompletely separated lineages. Herpetological Review 2020.
[4] De Queiroz K. Species concepts and species delimitation. Systematic Biology 2007;56(6):879–86.
[5] Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 2014;105(S1):795–809.
[6] Steenkamp ET, Wingfield MJ, McTaggart AR, Wingfield BD. Fungal species and their boundaries matter–Definitions, mechanisms and practical implications. Fungal Biology Reviews 2018;32(2):104–16.
[7] Doolittle WF. Speciation without species: a final word. Philosophy, theory, and practice in biology11; 2018.
[8] Palmer M, Venter SN, Coetzee MP, Steenkamp ET. Prokaryotic species are *sui generis* evolutionary units. Systematic and Applied Microbiology 2019;42(2):145–58.
[9] Shapiro BJ, Polz MF. Microbial speciation. Cold Spring Harbor Perspectives in Biology 2015;7(10):a018143.
[10] Shapiro BJ. What microbial population genomics has taught us about speciation. In: Population genomics: microorganisms. Springer; 2018. p. 31–47.
[11] McInerney JO, McNally A, O'connell MJ. Why prokaryotes have pangenomes. Nature Microbiology 2017;2(4):1–5.
[12] Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome. Proceedings of the National Academy of Sciences 2005;102(39):13950–5.
[13] Young JPW, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, et al. Genome Biology 2006;7(4):R34.
[14] Schleifer KH. Classification of Bacteria and Archaea: past, present and future. Systematic and Applied Microbiology 2009;32(8):533–42.
[15] Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 2018;68(1):461–6.
[16] Chung M, Munro JB, Tettelin H, Dunning Hotopp JC. Using core genome alignments to assign bacterial species. MSystems 2018;3(6):e00236. 18.
[17] Parker CT, Garrity GM, Tindall BJ. International code of nomenclature of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 2019;69(1A):S1–111.
[18] Buchanan RE, John-Brooks RS, Breed RS. International bacteriological code of nomenclature. Journal of Bacteriology 1948;55(3):287–306.
[19] Rossello-Mora R, Amann R. The species concept for prokaryotes. FEMS Microbiology Reviews 2001;25(1):39–67.
[20] Rossello-Mora R, Amann R. Past and future species definitions for Bacteria and Archaea. Syst Appl Microbiol 2015;38(4):209–16.
[21] Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM. Prokaryotic taxonomy and nomenclature in the age of big sequence data. The ISME Journal 2021;15(7):1879–92.
[22] Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences 2009;106(45):19126–31.
[23] Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Standards in Genomic Sciences 2010;2(1):117–34.
[24] Wayne L, Brenner D, Colwell R, Grimont P, Kandler O, Krichevsky M, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic and Evolutionary Microbiology 1987;37(4):463–4.
[25] Rogosa M, Wiseman R, Mitchell JA, Disraely M, Beaman A. Species differentiation of oral lactobacilli from man including descriptions of *Lactobacillus salivarius* nov spec and *Lactobacillus cellobiosus* nov spec. Journal of Bacteriology 1953;65(6):681–99.
[26] Petersen KD, Christensen H, Bisgaard M, Olsen JE. Genetic diversity of *Pasteurella multocida* fowl cholera isolates as demonstrated by ribotyping and 16S rRNA and partial atpD sequence comparisons the GenBank accession numbers for the 16S rRNA sequences of strain HIM 830-7T (NCTC 10204T) and 77179 of P. m. Microbiology. 2001;147(10):2739–48.
[27] Maiden MC. Multilocus sequence typing of bacteria. Annu Rev Microbiol 2006;60:561–88.
[28] Brady C, Cleenwerck I, Venter S, Coutinho T, De Vos P. Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov as *Lelliottia nimipressuralis* comb. nov and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov as *Pluralibacter gergoviae* comb. nov and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii, E. radicincitans, E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov as *Kosakonia cowanii* comb. nov., *Kosakonia radicincitans* comb. nov., *Kosakonia oryzae* comb. nov and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis, E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov and *Cronobacter pulveris* comb.

nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. Syst Appl Microbiol 2013;36(5):309–19.
[29] Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 2012;158(4):1005–15.
[30] Maiden MC, Van Rensburg Mjj, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nature Reviews Microbiology 2013;11(10):728–36.
[31] Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Systematic and Applied Microbiology 2015;38(4):237–45.
[32] Christensen H, Kuhnert P, Busse H-J, Frederiksen WC, Bisgaard M. Proposed minimal standards for the description of genera, species and subspecies of the *Pasteurellaceae*. International Journal of Systematic and Evolutionary Microbiology 2007;57(1):166–78.
[33] Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. International Journal of Systematic and Evolutionary Microbiology 2020;70(11):5607.
[34] Hayashi Sant'Anna F, Bach E, Porto RZ, Guella F, Hayashi Sant'Anna E, Passaglia LM. Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy. Critical Reviews in Microbiology 2019;45(2):182–200.
[35] VanInsberghe D, Arevalo P, Chien D, Polz MF. How can microbial population genomics inform community ecology? Philosophical Transactions of the Royal Society B 2020;375(1798):20190253.
[36] Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. Nature Reviews Microbiology 2008;6(6):431–40.
[37] Bobay L-M, Ochman H. Factors driving effective population size and pan-genome evolution in bacteria. BMC Evolutionary Biology 2018;18(1):1–12.
[38] Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 2007;57(Pt 1):81–91.
[39] Chan JZ, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ. Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. BMC Microbiology 2012;12(1):1–11.
[40] Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN. All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. International Journal of Systematic and Evolutionary Microbiology 2020;70(4):2937–48.
[41] Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Standards in Genomic Sciences 2014;9(1):2.
[42] Brady C, Hunter G, Kirk S, Arnold D, Denman S. Description of *Brenneria roseae* sp. nov. and two subspecies, *Brenneria roseae* subspecies *roseae* ssp. nov. and *Brenneria roseae* subspecies *americana* ssp. nov. isolated from symptomatic oak. Systematic and Applied Microbiology 2014;37(6):396–401.
[43] Jin H, Wang H, Zhang Y, Hu T, Lin Z, Liu B, et al. Description of *Azotobacter chroococcum* subsp. isscasi subsp. nov. isolated from paddy soil and establishment of *Azotobacter chroococcum* subsp. *chroococcum* subsp. nov. International Journal of Systematic and Evolutionary Microbiology 2020;70(3):2124–31.
[44] Garcia C, Mesnil A, Tourbiez D, Moussa M, Dubreuil C, Gonçalves De Sa A, et al. Vibrio aestuarianus subsp. *cardii* subsp. nov., pathogenic to the edible cockles *Cerastoderma edule* in France, and establishment of *Vibrio aestuarianus* subsp. *aestuarianus* subsp. nov. and Vibrio aestuarianus subsp. francensis subsp. nov. International Journal of Systematic and Evolutionary Microbiology 2021;71(2).
[45] García-López M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, Kyrpides NC, et al. Analysis of 1,000 type-strain genomes improves taxonomic classification of Bacteroidetes. Frontiers in Microbiology 2019;10:2083.
[46] Behrendt U, Wende S, Kolb S, Ulrich A. Genome-based phylogeny of the genera *Proteus* and *Cosenzaea* and description of *Proteus terrae* subsp. terrae subsp. nov. and *Proteus terrae* subsp. *cibarius* subsp. nov. International Journal of Systematic and Evolutionary Microbiology 2021;71(3):004651.
[47] Bobay L-M. The prokaryotic species concept and challenges. The Pangenome 2020:21–49.
[48] Azarian T, Huang I-T, Hanage WP. Structure and dynamics of bacterial populations: pangenome ecology. The Pangenome 2020:115–28.
[49] Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, Maclean C. The ecology and evolution of pangenomes. Current Biology 2019;29(20):R1094–103.
[50] Arnold DL, Jackson RW. Bacterial genomes: evolution of pathogenicity. Current Opinion in Plant Biology 2011;14(4):385–91.
[51] Kim Y, Gu C, Kim HU, Lee SY. Current status of pan-genome analysis for pathogenic bacteria. Current Opinion in Biotechnology 2020;63:54–62.
[52] Rasko DA, Rosovitz M, Myers GS, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. Journal of Bacteriology 2008;190(20):6881–93.
[53] Tanner JR, Kingsley RA. Evolution of *Salmonella* within hosts. Trends in Microbiology 2018;26(12):986–98.
[54] Jolley KA, Maiden MC. Using MLST to study bacterial variation: prospects in the genomic era. Future Microbiology 2014;9(5):623–30.
[55] Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. Clinical Microbiology and Infection 2018;24(4):350–4.
[56] Paulitsch F, Delamuta JRM, Ribeiro RA, da Silva Batista JS, Hungria M. Phylogeny of symbiotic genes reveals symbiovars within legume-nodulating *Paraburkholderia* species. Systematic and Applied Microbiology 2020;43(6):126151.
[57] Guglielmini J, Bourhy P, Schiettekatte O, Zinini F, Brisse S, Picardeau M. Genus-wide *Leptospira* core genome multilocus sequence typing for strain taxonomy and global surveillance. PLOS Neglected Tropical Diseases 2019;13(4):e0007374.
[58] Garnett ST, Christidis L, Conix S, Costello MJ, Zachos FE, Bánki OS, et al. Principles for creating a single authoritative list of the world's species. PLOS Biology 2020;18(7):e3000736.
[59] Carroll LM, Wiedmann M, Kovac J. Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. MBio 2020;11(1):e00034. 20.
[60] Goldstein EJ, Lewis RP, Martin WJ, Edelstein PH. Infections caused by *Klebsiella ozaenae*: a changing disease spectrum. Journal of Clinical Microbiology 1978;8(4):413–8.
[61] Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research 2021.
[62] Rosselló-Móra R, Stackebrandt E. I bridging 200 Years of bacterial classification. 2021.
[63] Murray AE, Freudenstein J, Gribaldo S, Hatzenpichler R, Hugenholtz P, Kämpfer P, et al. Roadmap for naming uncultivated Archaea and bacteria. Nature Microbiology 2020;5(8):987–94.
[64] Whitman WB, Chuvochina M, Hedlund BP, Hugenholtz P, Konstantinidis KT, Murray A, et al. Development of the SeqCode: a proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. Systematic and Applied Microbiology 2022:126305.

[65] Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nature Reviews Microbiology 2020;18(9):491–506.
[66] Wyres KL, Lam MM, Holt KE. Population genomics of *Klebsiella pneumoniae*. Nature Reviews Microbiology 2020;18(6):344–59.
[67] Bagley ST. Habitat association of *Klebsiella* species. Infection Control & Hospital Epidemiology 1985;6(2):52–8.
[68] Orskov I. Genus V. Klebsiella trevisan 1885, 105AL. Bergey's Manual of Systematic Bacteriology 1984;1:461–4.
[69] Brisse S, Verhoef J. Phylogenetic diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, gyrA and parC genes sequencing and automated ribotyping. International Journal of Systematic and Evolutionary Microbiology 2001;51(3):915–24.
[70] Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov. Research in Microbiology 2019;170(3):165–70.
[71] Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, et al. Whole-genome sequencing of human clinical *Klebsiella pneumoniae* isolates reveals misidentification and misunderstandings of *Klebsiella pneumoniae, Klebsiella variicola*, and Klebsiella quasipneumoniae. Msphere 2017;2(4):e00290. 17.
[72] Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic Escherichia coli. Nature Reviews Microbiology 2021;19(1):37–54.
[73] Gomila M, Busquets A, Mulet M, García-Valdés E, Lalucat J. Clarification of taxonomic status within the Pseudomonas syringae species group based on a phylogenomic analysis. Frontiers in Microbiology 2017;8:2422.
[74] Morinière L, Burlet A, Rosenthal ER, Nesme X, Portier P, Bull CT, et al. Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that Xanthomonas cynarae Trébaol et al. 2000 emend. Timilsina et al. 2019 is a later heterotypic synonym of Xanthomonas hortorum Vauterin et al. 1995. Systematic and Applied Microbiology 2020;43(4):126087.
[75] Laranjo M, Alexandre A, Oliveira S. Legume growth-promoting rhizobia: an overview on the Mesorhizobium genus. Microbiological Research 2014;169(1):2–17.
[76] Ramírez-Bahena MH, Flores-Félix JD, Velázquez E, Peix Á. The Mimosoid tree *Leucaena leucocephala* can be nodulated by the symbiovar genistearum of Bradyrhizobium canariense. Systematic and Applied Microbiology 2020;43(1):126041.
[77] Young JPW, Moeskjær S, Afonin A, Rahi P, Maluk M, James EK, et al. Defining the *Rhizobium leguminosarum* species complex. Genes 2021;12(1):111.
[78] Banerji S, Simon S, Tille A, Fruth A, Flieger A. Genome-based *Salmonella* serotyping as the new gold standard. Scientific Reports 2020;10(1).
[79] Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, Mariet J-F, et al. An assessment of different genomic approaches for inferring phylogeny of Listeria monocytogenes. Frontiers in Microbiology 2017;8:2351.
[80] Riojas MA, McGough KJ, Rider-Riojas CJ, Rastogi N, Hazbón MH. Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum*, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of Mycobacterium tuberculosis. International Journal of Systematic and Evolutionary Microbiology 2018;68(1):324–32.
[81] Wels M, Siezen R, Van Hijum S, Kelly WJ, Bachmann H. Comparative genome analysis of Lactococcus lactis indicates niche adaptation and resolves genotype/phenotype disparity. Frontiers in Microbiology 2019;10:4.
[82] Hennart M, Panunzi LG, Rodrigues C, Gaday Q, Baines SL, Barros-Pinkelnig M, et al. Population genomics and antimicrobial resistance in *Corynebacterium diphtheriae*. Genome Medicine 2020;12(1):1–18.
[83] Miller WG, Yee E, Chapman MH, Bono JL. Comparative genomics of all three Campylobacter sputorum biovars and a novel cattle-associated C. sputorum clade. Genome Biology and Evolution 2017;9(6):1513–8.
[84] Soares SC, Silva A, Trost E, Blom J, Ramos R, Carneiro A, et al. The pan-genome of the animal pathogen *Corynebacterium pseudotuberculosis* reveals differences in genome plasticity between the biovar ovis and equi strains. PLoS One 2013;8(1):e53818.