Taxonomic Note

Modest proposals to expand the type material for naming of prokaryotes

William B. Whitman

Department of Microbiology, University of Georgia, Athens, GA 30602-2605, USA

Correspondence
William B. Whitman
whitman@uga.edu

Gene sequences are herein proposed to be suitable type material for the description of prokaryotic species. This proposal follows from the principles described in the International Code of Nomenclature of Prokaryotes. This simple change in the Code will allow for the stability of naming of *Candidatus* taxa, endosymbionts and uncultivated prokaryotes and will meet an important need within microbiology. In addition, modern molecular techniques allow the identification of genera even when the species remain obscure. The Code should be modified to allow gene sequences to serve as the type material for genera in the absence of described species. This simple change will unite the nomenclature of the cultured and uncultured prokaryotes into a single, robust system.

A century ago Winslow *et al.* (1917) identified the two central questions in prokaryotic biology as how prokaryotes live and how they are related. They recognized that progress on both of these questions was interdependent and required for complete understanding. In the last century, enormous progress has been made on both questions, and our understanding of the relationships among prokaryotes and their systematics has profited enormously from the wealth of knowledge now available on how they live. Similarly, the nomenclature of prokaryotes as embodied in the International Code of Nomenclature of Prokaryotes (hereafter referred to as the Code; Parker *et al.*, 2016) is intimately dependent upon our understanding of prokaryotic biology, and the Code only remains useful when it reflects our current understandings of these central questions. From this perspective, proposals to broaden the nomenclatural type material for prokaryotes are necessary to incorporate our current understanding of their systematics into the Code.

Before considering the details of these proposals, it is important to review the purpose of names and type material as well as the principles of priority in the Code (Parker *et al.*, 2016). As put forth in principle 4 of the Code, the primary purpose of naming is to supply a means of referring to specific prokaryotes. Names are especially important for novices and members of different scientific communities. Experts with great familiarity of a field can usually understand the meanings of other experts. However, for the non-expert, it is often impossible to sort through the large numbers of synonymous or highly similar names that naturally form in the absence of the Code or some other set of nomenclatural rules.

The Code possesses three mechanisms to ensure uniqueness and stability of names. Firstly, it gives priority to the earliest name of an entity. Thus, if a taxon already has a valid name, it is not possible under ordinary circumstances to rename it. Secondly, because the names of the higher taxonomic ranks follow logically from the genus name, priority applies to all of the taxonomic ranks up to class, which is the highest taxonomic rank currently recognized in the Code. Thirdly, each name is associated irrevocably with some type material. The only name that can be used to include this type material is the name with priority. The relationship of the name to the type material is further determined by the circumscription, which indicates the relatedness of other members of a taxon to the type material. In more concrete terms, strain ATCC 6051 is the type material of the species *Bacillus subtilis*, and *B. subtilis* is the type species of the genus *Bacillus*. Its priority is determined by the date of its original description, in this case by Ehrenberg in 1835. By the rules of priority, any species that is described after Ehrenberg's description that includes strain ATCC 6051 must be named *Bacillus subtilis*. Similarly, any genus that includes *B. subtilis* must be named *Bacillus*, and any family that includes *Bacillus* must be named *Bacillaceae*. Likewise, any species that does not include strain ATCC 6051 cannot be named *B. subtilis*. However, whether or not a strain related to strain ATCC 6051 is a representative of *B. subtilis*, a novel species of the genus *Bacillus*, or a new genus of the family *Bacillaceae* is determined by the circumscription of each of these taxa and its relatedness to the type material. A strain very similar to ATCC 6051 might be named *B. subtilis*. A distantly related strain might be classified in another genus in the family. However, the circumscription is not determined by the Code but by taxonomic theory, which is not regulated by the Code (Tindall, 1999).

This system provides a unique name to each prokaryote. Because it prevents the creation of synonyms and changes to the name by later authors, it is essential to the usefulness

of names. Prior to the adoption of the Code, it was common for prokaryotic species to have multiple names. For instance, the 6th edition of *Bergeys's Manual* reported 31 likely synonyms for *Bacillus* and 27 likely synonyms for *B. subtilis* (Breed *et al.*, 1948). Similarly, the family containing the genus *Bacillus* had been called by various authors *Bacillaceae*, *Bacillidae*, *Bacteriaceae* and *Endosporaceae* (Breed, 1948). Only the application of strict guidelines has allowed investigators to develop a globally shared vocabulary for prokaryotes. From this perspective, the Code is an algorithm designed to ensure the orderliness of names. It is not a manual for good practices, which is more properly the domain of journal reviewers and editors and inherently relative (Whitman, 2015).

Essential to this nomenclatural system, the type material must unambiguously identify the taxon. If the type material is poorly defined, it is impossible to determine the circumscription of the taxon and the relatedness to other strains. Although cultures are the commonest type material, early versions of the Code allowed for specimens and detailed descriptions to serve as type material (Buchanan *et al.*, 1948; Lapage *et al.*, 1992). For example, *Pasteuria ramosa* is an obligate parasite of nematodes which has never been cultured, and the type material is the extensive description by Metchnikoff (1888). Although the naming of fossil prokaryotes is not common, it was recognized in the original Bacteriological Code (Buchanan *et al.*, 1948). Similarly, naming of fungal and protist fossils remains common, and descriptive type material is recognized by the Botanical Code. In principle, any descriptive material with sufficient unambiguity could serve as the type material.

In 2001, the Code was revised so that only organisms deposited in culture collections could serve as type material. At the time, this rule was justifiable because experience had shown that other descriptive material often lacked the required specificity for the unambiguous identification of species. However, this rule had some unintended consequences. Most importantly, it precluded applying the Code to the names of taxa which could not be deposited in culture collections, such as symbionts cultivated in exotic hosts and fastidious prokaryotes which required unusual or toxic culture media. Ironically, at the same time that the enormous diversity and abundance of uncultured prokaryotes was becoming widely recognized, this rule prevented application of the Code to their names.

It is now widely recognized that sequence data provides an unambiguous identification of prokaryotic taxa. For instance, genome sequences have been proposed to replace DNA–DNA hybridization for circumscription of species (Richter & Rosselló-Móra, 2009; Meier-Kolthoff *et al.*, 2013; Chun & Rainey, 2014; Li *et al.*, 2015). For this reason, I recently proposed that the Code be revised to include sequence data as type material for the species (Whitman, 2015). This proposal is fully consistent with the principles if not the word of the Code.

However, even the complete genome sequence is not necessary, and in many cases multilocus sequence typing also provides sufficient specificity to identify novel species (Glaeser & Kämpfer, 2015, Rosselló-Móra & Amann, 2015). In this regard, the key issue is whether or not the species is unambiguously identified, which is an absolute and not quantitative distinction. Either the evidence is sufficient to identify the species or it is not. To require more is to dictate best practices, which the Code has never done. Best practices are relative and based upon taxonomic theory. For instance, many organisms are easily cultured, and draft genome sequences can be determined conveniently and inexpensively. Given their enormous value for improving descriptions, best practices but not the Code demand their inclusion. However, other organisms of great interest are difficult to culture, and the determination of their genome sequences may not be technically feasible. Nevertheless, these organisms need to be named. If these organisms can be unambiguously identified with less than a draft genome sequence, they should be.

Accepting sequence data as the type material does not have to be limited to only species, but it should also be extended to genera. In this regard, there is a strong empirical basis and taxonomic theory for recognizing genera on the basis of 16S rRNA gene sequence similarity (Yarza *et al.*, 2014). Even in prokaryotes that contain multiple alleles for the 16S rRNA gene, above a certain level of sequence similarity this gene sufficiently identifies genera for establishment of priority. Because the nomenclature of the higher taxa follows from that of the genus, recognition of genus names provides stability to the nomenclature of the names of uncultured families, orders and classes. If a recent proposal to include the rank of phylum in the Code is adopted, the name of phylum will also follow from the genus name (Oren *et al.*, 2015). Moreover, given a validly published genus name, it would be a simple matter to add species as type strains are isolated or more sequence information becomes available.

While this proposal allows naming a genus without a species, it is not proposing that genera exist in the absence of a species. Rather, it allows the assertion that one or more species representing a genus exists in nature but that the evidence to unambiguously identify the species is lacking. The intention is to allow the nomenclature to more accurately reflect the state of our knowledge.

Currently, uncultured organisms are commonly either informally named or given *Candidatus* names. The *Candidatus* nomenclature was designed to recognize phylogenetic groups identified by 16S rRNA gene sequences and fluorescence *in situ* hybridization (Murray & Schleifer, 1994; Murray & Stackebrandt, 1995). According to the List of Prokaryotic Names with Standing in Nomenclature (LPSN), more than 360 *Candidatus* names have been published (Parte, 2014). Similarly, many novel phyla of uncultured prokaryotes have been proposed on the basis of single cell genomics and metagenomic analyses of

environmental samples (Hedlund *et al.*, 2015). Currently, none of these names have standing in nomenclature and the protection of priority offered by the Code. Nevertheless, these names are common in the literature, which provides a strong argument for the need to include the uncultured in the Code. The Code has proven necessary to ensure the stability of names for cultured prokaryotes, and there is every reason to believe it will ensure the stability of nomenclature of the uncultured. Because microbiologists believe that these entities should be named, the rules should exist to grant the names priority and stability.

Importantly, this proposal also allows for unification of the nomenclature of the uncultured and cultured into a single taxonomy for all prokaryotes. A single taxonomy will facilitate communication between organismal and environmental microbiologists and sharing of insights between these fields (Yarza *et al.*, 2014; Hedlund *et al.*, 2015). Thus, one might anticipate a qualitative if not quantitative understanding of phenotypes of the uncultured taxa by analogy to the phenotypic diversity of cultured taxa of comparable taxonomic ranks. Likewise, recognition of the uncultured relatives will enable a fuller understanding of the evolution and ecology of cultured taxa.

Because the names of many phyla and the botanical equivalent divisions have already been proposed (Rinke *et al.*, 2013; Brown *et al.*, 2015; Spang *et al.*, 2015), it could be argued that the same goals could be achieved by incorporating these names into the Code. There are two difficulties with this approach. Firstly, the circumscriptions of the higher taxa are inherently ambiguous and become more ambiguous the higher the taxonomic rank. Even the rank of species is to some extent ambiguous, and there is no widely accepted distinction between subspecies and species. Although a consensus is developing for the definition of a genus, there is little agreement for the higher taxonomic ranks of family, order and above. Secondly, the rules of priority only work from lower to higher taxonomic ranks, from the more specific to the more general. For instance, if priority is given to a phylum, there is no mechanism to ensure that lower taxa will be assigned to that phylum. Lastly, it is unnecessary. Because the names of all higher taxa follow from the genus names, it is only necessary to designate the genus.

For these reasons, the following revisions of the Code (Parker *et al.*, 2016) are proposed (new text is underlined):

Rule 15. The only changes in this rule are in the first two rows of Table 3.

For the type of the subspecies and species, replace 'Designated strain; in special cases the place of the type strain may be taken by a description, preserved specimen, or an illustration [see Rule 18a(1)]' with 'Designated strain, sequence of genomic DNA, description, preserved specimen or illustration as described in Rule 18a'.

For the type of the subgenus or genus, replace 'Designated species' with 'Designated species or sequence of one or more genes as described in Rule 20a'.

Note: these changes allow sequences to serve as the type material for species and genus descriptions.

Rule 18a. The type material of a species or subspecies must unambiguously identify the taxonomic group and is a designated strain or other material. Whenever possible, the type of a species or subspecies is a designated strain.

The type strain is made up of living cultures of an organism, which are descended from a strain designated the nomenclatural type. The strain should have been maintained in pure culture and should agree closely in its characteristics with those of the original description (see Chapter 4C). The type strain may be designated in various ways (see Rules 18b, 18c and 18d).

(1) Until December 2000, for a species (or subspecies) which has not so far been maintained in laboratory cultures or for which a type does not exist, a description, preserved specimen or illustration (see also Rule 18f) may serve as the type.

Example: non-cultivated *Oscillospira guilliermondii* Chatton and Perard 1913.

(2) As from January 2001, a description, preserved (non-viable) specimen or illustration may not serve as the type.

(3) As from January 2016, sequence of genomic DNA may also serve as the type material when it unambiguously identifies the species. When possible, it should be a high quality draft or better genome sequence. As new methods are developed, they may serve as the type material so long as they unambiguously identify the species or subspecies and can be readily archived and compared.

Note: these changes allow sequences to serve as the type material for species. In addition, they identify the criteria that sequences or any alternative data must fulfil to qualify as type material.

Rule 18f. If a sequence of genomic DNA, description or illustration constitutes, or a dead preserved specimen has been designated, the type of a species [Rules 18a(1) and 18a(3)] and a later strain of this species is cultivated, then the type strain may be designated by the person who isolated the strain or by a subsequent author. This type strain shall then replace the sequence of genomic DNA, description, illustration or preserved specimen as the nomenclatural type. The designation of a type strain in this manner must be published in the IJSEM, the authorship and date of priority of publication being determined by the effective and valid publication of the name by the original authors (Rule 24b).

Note: these changes allow replacement of a type sequence of genomic DNA with a type strain.

Rule 20a. The nomenclatural type (see Rule 15) of a genus or subgenus is the type species or the sequence of one or more genes that unambiguously identifies the genus or subgenus. The type species is the single species or one of the species included when the name was originally validly

published. Only species whose names are legitimate may serve as types.

Note: these changes allow a sequence to be the type material of a genus in the absence of a species.

Rule 20e (2). However, a generic name for which no identifiable <u>type</u> ['species' is deleted] can be selected in accordance with the Rules might have come into use for identifiable species which were subsequently named. In this case, one of these later species may be selected as the type species and established as such by an Opinion of the Judicial Commission. The generic name is then ascribed to the author of the name of the species selected as the type species.

Note: this change allows for application of the rule to genera where the type material is a sequence. The example is unchanged.

Rule 20g

A genus and its type subgenus share the same type ['species' is deleted].

Note: this change allows for application of the rule to genera where the type material is a sequence. The example is unchanged.

Rule 27 (3). The type of the taxon must be designated (see Rules 15, 16, 18a, 18b, 18f, 20a–c, 21a and 22). In the case of species or subspecies including new combinations, the type ['strains' is deleted] must be <u>designated and/or deposited</u> according to Rule 30.

Note: this change allows for application of the rule where the type material is a sequence.

Rule 29

For a generic or subgeneric name to be validly published it must comply with the following conditions.

(1) It must be published in conformity with Rules 27 and 28b.

(2) The genus or subgenus named must include one or more described or previously described species <u>or a designated type sequence</u>.

Note: the remainder of Rule 29 is unchanged.

Rule 30 (3) (c). [New rule] <u>When a sequence is the type material, the accession number in a publicly available database or the sequence must be given. It is recommended that, when possible, a sample of the DNA be deposited in at least two publicly accessible service collections in different countries and the catalogue numbers be indicated.</u>

Rule 39c. The principles of Rules 39a and 39b apply when a subgenus is divided into two or more subgenera, the original subgeneric name being retained for that subgenus which contains the type material ['species' is replaced by 'material'].

Rule 40b. The specific epithet must be retained for the species or subspecies which includes the type material

['strain' is replaced by 'material']. When no type was designated, one must be chosen (see Note to Rule 39b).

Rule 40c. The principles of Rules 40a and 40b apply when a subspecies is divided into two or more subspecies, the original subspecies name being retained for that subspecies which contains the type material ['strain' is replaced by 'material'].

Rule 43

When several genera are united as subgenera of one genus, the subgenus which includes the type *material* ['species' is replaced by 'material'] of the genus under which union takes place must bear the same name as that genus.

Note: the remainder of Rule 43 is unchanged.

Rule 44

If two or more species of different genera are brought together to form a genus, and if these species include the type species of one or more genera, the name of the genus is that associated with the type species having the earliest legitimate generic name.

<u>If one or more species of different genera are brought together with a genus where the type material is a sequence, the name of the genus is that associated with the type material having the earliest legitimate generic name.</u>

If no type species is placed in the genus, a new generic name must be proposed and a type species selected.

Note: the addition to this rule is to allow the integration of cultured organisms into genera where the type material is a sequence without loss of priority of the genus. The example for Rule 44 is unchanged.

Rule 45. When several species are united as subspecies under one species, the subspecies which includes the type material ['species' is replaced by 'material'] of the species under whose name they are united must be designated by the same epithet as the species.

References

Breed, R. S. (1948). Historical survey of classifications of bacteria, with emphasis on outlines proposed since 1923. In *Bergey's Manual of Determinative Bacteriology*, pp. 5–38. Edited by R. S. Breed, E. G. D. Murray & A. P. Hitchens. Baltimore, MD: Williams & Wilkins.

Breed, R. S., Murray, E. G. D., Hitchens, A. P. (editors) (1948). *Bergey's Manual of Determinative Bacteriology*. Baltimore, MD: Williams & Wilkins.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I, Castelle, C. J., Singh, A., Wilkins, M. J., Wrighton, K. C., Williams, K. H. & Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211.

Buchanan, R. E., St John-Brooks, R. & Breed, R. S. (1948). International bacteriological code of nomenclature. *J Bacteriol* **55**, 287–306.

Chun, J. & Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. *Int J Syst Evol Microbiol* **64**, 316–324.

Glaeser, S. P. & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* **38**, 237–245.

Hedlund, B. P., Dodsworth, J. A. & Staley, J. T. (2015). The changing landscape of microbial biodiversity exploration and its implications for systematics. *Syst Appl Microbiol* **38**, 231–236.

Lapage, S. P., Sneath, P. H. A., Lessel, E. F., Skerman, V. B. D., Seeliger, H. P. R. & Clark, W. A. (editors) (1992). *International Code of Nomenclature of Bacteria (1990 Revision). Bacteriological Code.* Washington, DC: American Society for Microbiology.

Li, X., Huang, Y. & Whitman, W. B. (2015). The relationship of the whole genome sequence identity to DNA hybridization varies between genera of prokaryotes. *Antonie van Leeuwenhoek* **107**, 241–249.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60.

Metchnikoff, E. (1888). *Pasteuria ramosa*, un représentant des bactéries à division longitudinale. *Ann Inst Pasteur (Paris)* **2**, 165–170.

Murray, R. G. & Schleifer, K. H. (1994). Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes. *Int J Syst Bacteriol* **44**, 174–176.

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

Oren, A., da Costa, M. S., Garrity, G. M., Rainey, F. A., Rosselló-Móra, R., Schink, B., Sutcliffe, I., Trujillo, M. E. & Whitman, W. B. (2015). Proposal to include the rank of phylum in the International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* **65**, 4284–4287.

Parker, C. T., Tindall, B. J. & Garrity, G. M. (2016). International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol*.

Parte, A. C. (2014). LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* **42** (D1), D613–D616.

Richter, M. & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**, 19126–19131.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., Darling, A., Malfatti, S., Swan, B. K. & other authors (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437.

Rosselló-Móra, R. & Amann, R. (2015). Past and future species definitions for *Bacteria* and *Archaea*. *Syst Appl Microbiol* **38**, 209–216.

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L. & Ettema, T. J. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179.

Tindall, B. J. (1999). Misunderstanding the Bacteriological Code. *Int J Syst Bacteriol* **49**, 1313–1316.

Whitman, W. B. (2015). Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Syst Appl Microbiol* **38**, 217–222.

Winslow, C. E., Broadhurst, J., Buchanan, R. E., Krumwiede, C., Rogers, L. A. & Smith, G. H. (1917). The families and genera of the *Bacteria*: preliminary report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. *J Bacteriol* **2**, 505–566.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R. & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**, 635–645.