Systematic and Applied Microbiology 45 (2022) 126305

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.com/locate/syapm

Development of the SeqCode: A proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type

William B. Whitman [a,*], Maria Chuvochina [b], Brian P. Hedlund [c], Philip Hugenholtz [b], Konstantinos T. Konstantinidis [d], Alison E. Murray [e], Marike Palmer [c], Donovan H. Parks [b], Alexander J. Probst [f], Anna-Louise Reysenbach [g], Luis M. Rodriguez-R [h], Ramon Rossello-Mora [i], Iain Sutcliffe [j], Stephanus N. Venter [k]

[a] Department of Microbiology, University of Georgia, Athens, GA, USA
[b] The University of Queensland, School of Chemistry and Molecular Biosciences, Australian Centre for Ecogenomics, Australia
[c] School of Life Sciences, University of Nevada, Las Vegas, NV, USA
[d] School of Civil and Environmental Engineering, Georgia Tech, Atlanta, GA, USA
[e] Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, NV, USA
[f] Department of Chemistry, Environmental Microbiology and Biotechnology (EMB), Group for Aquatic Microbial Ecology and Centre of Water and Environmental Research (ZWU), University of Duisburg-Essen, Essen, Germany
[g] Biology Department, Portland State University, Portland, OR, USA
[h] Department of Microbiology and Digital Science Center (DiSC), University of Innsbruck, Innrain 15 / 01-05, Innsbruck 6020, Austria
[i] Marine Microbiology Group, Department of Animal and Microbial Diversity, Mediterranean Institute of Advanced Studies (CSIC-UIB), Esporles, Illes Balears, Spain
[j] Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, UK
[k] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

ARTICLE INFO

Article history:
Received 9 November 2021
Revised 30 January 2022
Accepted 31 January 2022

Keywords:
Nomenclatural type
Metagenome
SeqCode
Nomenclatural code
Candidatus

ABSTRACT

Over the last fifteen years, genomics has become fully integrated into prokaryotic systematics. The genomes of most type strains have been sequenced, genome sequence similarity is widely used for delineation of species, and phylogenomic methods are commonly used for classification of higher taxonomic ranks. Additionally, environmental genomics has revealed a vast diversity of as-yet-uncultivated taxa. In response to these developments, a new code of nomenclature, the *Code of Nomenclature of Prokaryotes Described from Sequence Data* (SeqCode), has been developed over the last two years to allow naming of Archaea and Bacteria using DNA sequences as the nomenclatural types. The SeqCode also allows naming of cultured organisms, including fastidious prokaryotes that cannot be deposited into culture collections. Several simplifications relative to the *International Code of Nomenclature of Prokaryotes* (ICNP) are implemented to make nomenclature more accessible, easier to apply and more readily communicated. By simplifying nomenclature with the goal of a unified classification, inclusive of both cultured and uncultured taxa, the SeqCode will facilitate the naming of taxa in every biome on Earth, encourage the isolation and characterization of as-yet-uncultivated taxa, and promote synergies between the ecological, environmental, physiological, biochemical, and molecular biological disciplines to more fully describe prokaryotes.

Contents

* Corresponding author.
E-mail address: whitman@uga.edu (W.B. Whitman).

https://doi.org/10.1016/j.syapm.2022.126305
0723-2020/

Introduction

Genomics has profoundly influenced prokaryotic systematics over the last fifteen years. Led by Nikos Kyrpides and Tanja Woyke at the U.S. Department of Energy's Joint Genome Institute, in collaboration with Hans-Peter Klenk and later Markus Göker at the DSMZ, the systematic sequencing of microbial genomes began with the Genomic Encyclopedia of Bacteria and Archaea (GEBA) project [24,66]. It was joined in 2018 by the World Data Centre for Microorganisms at the Chinese Academy of Sciences in Beijing in a project called GCM2.0, led by Juncai Ma and Linhuan Wu [67]. These large-scale genome sequencing projects targeted type strains, the taxonomic reference material of prokaryotic species, and together have sequenced over 4,000 type strains to date. In parallel, most prokaryotic systematics journals now require or strongly recommend that genome sequences be included in descriptions of new species [5]. As a result of these and other efforts, the genome sequences of >14,500 of the approximately 18,000 type strains are now available (https://gctype.wdcm.org/). The remaining as-yet-unsequenced type strains were mostly described prior to the advent of widespread genome sequencing and include fastidious prokaryotes from a wide variety of habitats.

The availability of inexpensive genome sequencing has also led to a key role for genomics in the establishment of modern species definitions. Previously, DNA:DNA hybridization was the consensus method to delineate and distinguish prokaryotic species [60]. However, this method was difficult to execute and prone to experimental error [51]. Genome sequence similarity was shown to correlate with DNA:DNA hybridization values [13] but is more precise, repeatable, and less prone to experimental error, leading to the development of genomic methods, such as Average Nucleotide Identity (ANI) [23,42,49] and digital DNA:DNA hybridization (dDDH) [27]. These *in silico* techniques have allowed the broad application of sequence similarity and improved the fidelity of prokaryotic species delineation. As a result, genome comparisons are now central to prokaryotic species descriptions and provide greater resolution than 16S rRNA gene-based comparisons, which have long been widely used.

In parallel with culture-based genomics, high throughput sequencing has increasingly been applied to microbial communities (metagenomics) [48], culminating in the ability to obtain complete and nearly complete bacterial and archaeal genomes directly from environmental sequence data, so called metagenome-assembled genomes (MAGs) [4,35,54]. However, since most MAGs are still draft assemblies of environmental contigs, and each contig is often a mixture of sequences from sympatric strains, the fidelity of such genomes has been questioned [28,58]. Continuing improvements in both laboratory and bioinformatic approaches, including the use of long-read sequencing, suggest that MAG quality can often, but not always, be on par with that of isolate genomes and will likely soon be on par in all cases when long-read sequencing becomes more accessible and reliable [30]. The cost-effective generation of MAGs has resulted in sequences of tens of thousands of genomes across dozens of ecosystems, which has substantially increased our understanding of prokaryotic diversity [46].

Apart from using genome sequences for the delineation of species, the availability of genomes representing higher taxonomic groups [24], as well as intra-species diversity, has changed the classification of Bacteria and Archaea in a profound way. Genomic data have not only been used to address the systematics and evolution of specific genera [3,8,12,53,69] but also several higher taxa [14,18]. Genome sequences have also been used to develop a standardized bacterial and archaeal classification based on phylogenetic analyses and sequence divergence [45,50], as is currently reflected in the Genome Taxonomy Database (GTDB) [46,50]. The incorporation of nearly 260,000 genomes, including >53,000 MAGs has significantly improved the state of knowledge of the phylogeny and classification of Bacteria and Archaea [46].

The extensive use of genomic data in taxonomy and the stability and insight it has provided has led microbiologists to propose better integration of genomic data into formal systems to name and classify taxa, including those without representative pure cultures. Various proposals were made to incorporate DNA sequence data into formal systems of nomenclature. Hedlund et al. [15] and Konstantinidis and Rossello-Mora [20] proposed revision of the *Candidatus* concept from the International Code of Nomenclature of Prokaryotes (ICNP) to incorporate SAGs and MAGs regardless of the ability to visualize these organisms in natural samples. Whitman [61,62] further proposed revision of the ICNP to incorporate high-quality genome sequences as nomenclatural types. That proposal would have paved the way for validation of *Candidatus* names and improved the stability of their nomenclature [64].

In 2017, Konstantinidis et al. [21] articulated the need for a stable system for naming uncultivated microbes and for the first time proposed the possibility of a separate nomenclatural system based on high-quality MAGs and SAGs. The following year, Hedlund and Reysenbach obtained funding from the US National Science Foundation's Systematic and Biodiversity Science Cluster to develop a strategy for *Microbial Systematics for the Next Decade*. Several online meetings and in-person workshops in October 2018 and April 2019 resulted in the publication of a Consensus Statement [34] endorsed by 120 prominent microbiologists from around the globe. The first goal ('Plan A') was the ratification of Whitman's proposal [62] by the ICNP to allow DNA sequences as nomenclatural types, whilst the alternative goal ('Plan B') was a separate code based on genome sequences as types, as proposed earlier [21]. In anticipation of using genome sequences as type, criteria for selecting type genomes have been proposed [6,21]. After debating 'Plan A' in an e-mail discussion forum, the International Committee on Systematics of Prokaryotes (ICSP) rejected the proposal by Whitman [62] that DNA could serve as types under the ICNP [57]. In response, 'Plan B' was initiated. The first draft of the Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode) was prepared in the summer of 2020 and discussed at a series of international online workshops held during February 2021 under the banner of the International Society for Microbial Ecology (ISME), which engaged 848 registrants from 42 countries. The focus of the workshops was to communicate 'Plan B' and gather stakeholder feedback on important issues, which were incorporated into revisions of the draft SeqCode (https://www.

isme-microbes.org/seqcode-workshops). As a result of this process, the first edition of the SeqCode has now been published [16].

Below we review some of the context and considerations underpinning the design and proposed operation of the SeqCode, particularly in relation to lessons learned from the workings of the ICNP.

Some general principles of biological nomenclature

Naming is an essential component of scientific investigations, including those in biology. Without the ability to assign precise and unambiguous names, it is impossible to communicate effectively. For instance, prior to the implementation of the ICNP and formation of the Approved Lists [55], *Mycobacterium tuberculosis*, the causative agent of tuberculosis, had at least nine scientific names [2]. Not only did the multiple names make following the literature difficult, but the exact meanings of the names were uncertain, which caused confusion and was perilous in a medical/veterinary microbiology context. Precise naming is also critical for the meaningful implementation of large databases and for large-scale analyses made possible by computing. Without a single, relatively stable name, databases require frequent curation, which becomes increasingly difficult as their content grows. Thus, a precise naming system is needed for the effective application of modern bioinformatics tools.

In biology, most codes of nomenclature have been designed to create precise names by establishing identifiers that have a one-to-one correspondence with natural entities [65]. Their goals are often described in sections entitled "General considerations" or "Principles". Chief among their goals is to ensure that each entity has only a single, unique formal name in a particular taxonomic context or 'position'. This goal is challenging because naming proceeds concurrently with biodiversity discovery, which occurs independently by investigators all over the world. Investigators often have different taxonomic opinions and philosophies, and naming must accommodate these differences. Nomenclatural codes seek to knit these divergent efforts into a single, shareable, robust nomenclature.

To accomplish this goal, nomenclatural codes must deal with four fundamental components of naming. The first is the hypothesis that there exists a biological entity deserving of a name. From the viewpoint of a taxonomist, the biological entity is defined as a taxon. Like other codes, the ICNP is purposely ambiguous regarding the nature of these entities and defines them as "any group of organisms treated as a named group in a formal taxonomy" (General Consideration 7(3)) [44]. This ambiguity is necessary because codes must be useable by investigators with very different taxonomic philosophies. However, the ICNP does exclude certain things, such as fossils, which are named under other codes.

The second component is evidence for the taxon, which is the nomenclatural type (or just 'type') under most codes. Types have three essential functions. First, they prove the existence of the taxon. Second, they provide a reference standard for comparison with a new specimen to determine whether it belongs to the same or a different taxon. As a consequence of the latter, the third function of types is their role in the application of names. Presence of a type in a taxon demands that its name be based upon the name of the type. Exclusion of all known types from a taxon warrants the creation of a new name. The system of types ensures the precise meaning of the names [17]. For instance, a specimen named *Mycobacterium tuberculosis* must be included in a taxon that encompasses the type of the species to warrant the same name.

There are two fundamentally different kinds of types. Firstly, the types for species and subspecies are the experimental evidence for the taxon and in prokaryotic biology have historically been strains or, in some cases prior to 1 January 2001, detailed descriptions, preserved (non-viable) specimens or illustrations [26,43]. In the SeqCode, this kind of type is a genome sequence. Secondly, the types for genera and higher taxa are their subordinate taxa (for example, a genus has a type species and a family has a type genus). As a consequence, every taxon is associated with some experimental evidence of its existence, either directly as for species and subspecies or indirectly as for taxa above the rank of species. Likewise, the precise meaning of each name is derived from the knowledge that the taxon includes a particular type.

The third component is the name itself, which is arguably the least important component. Although Linnaean tradition favours binomial names formed from Latin and Greek, in principle the name could be formed from any source and even arbitrarily [39]. Latinization of names was understandable when naming was primarily a European activity and Latin was a common component of their scientists' education [65]. It is less justified in the modern global context, and the International Code of Virus Classification and Nomenclature has recently been emended to permit names that are not Latin binomials (https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code).

Ideally, names should be understandable and easy to form. In practice, latinization of names often requires extensive curation by experts and lacks scalability. Thus, while it is not difficult to create a few names during a research project, generating large numbers of names is time-consuming, tedious, and prone to grammatical error. Fortunately, there are a number of ways to increase the scalability of naming even within the ICNP. Pallen et al. [41] recently created the 'Great Automatic Nomenclator' or GAN, which can combinatorically generate large numbers of names from a small number of Latin roots. This approach may be generalized and improved in the near future and eliminate the problem of scalability while providing even more options for creating linguistically correct names. An alternative is to create a system of simplified latinization with relaxed rules to simplify the naming process for the research community and decrease the burden on curators. Some biological codes of nomenclature have also relaxed the rules for naming so that strict adherence to Latin is not required. A more complete discussion of the process of creating names is given by Pallen [39].

Fourthly, a code should describe a process or system that allows names to be added as new data or new perspectives are acquired. This process resembles an algorithm, which is a simple set of rules that enables solving of complex problems. By systematically combining the discoveries from a large pool of investigators, a consensus nomenclature can be achieved even in the absence of a unified classification. The algorithm must allow for creation of new names as taxa are discovered or when the classification of existing taxa change. The process of union, transfer, and division of taxa must be straightforward so that it is easily understood and readily adopted; and when conflicts arise, there must be means to objectively resolve them.

The principle of priority is a key element of this process. This principle states that the earliest validly published name for a taxon is the correct name for the taxon. This principle ensures the stability of names and that, in a given taxonomic position, each taxon has only one correct name. At the rank of species and subspecies, names are directly associated with a nomenclatural type, so the earliest name that includes a particular type has priority over any subsequent name and cannot be changed. For instance, if a species is moved to another genus, its species epithet remains the same even though the genus name changes. If a species is united with another species, the earliest validly published name remains the correct name. If a species is divided, the taxon that retains the type must retain the original name.

Because the nomenclature is binomial, a genus name and the epithet of its type species have to be proposed at the same time and, thus, have the same date of priority. Any genus that includes that type species must have the earliest validly published genus name of that species. Likewise, if a taxonomist unites two genera, the type species of the resulting merged taxon will be the one with the earliest, validly published name. If a genus is divided, the name must be retained by the taxon that retains the type species.

Most nomenclatural codes attempt to be philosophically neutral on the great issues in taxonomy. Principle 1(4) of the ICNP states: "Nothing in this Code may be construed to restrict the freedom of taxonomic thought or action" [44]. However, in practice, absolute neutrality is not possible [52]. For instance, both the ICNP and the SeqCode are designed to create names for a hierarchical classification, and thus both assume that a hierarchy is both desirable and reflective of nature. Likewise, they also assume that entities called 'species' exist in prokaryotic biology. Implicit in these assumptions is that every species belongs to a genus, and every genus belongs to one of each of the higher ranks used in this hierarchy, i.e. family, order, class, and phylum. Even if these founding assumptions are incorrect, it is still possible to create a useful nomenclature.

Nomenclatural codes also strive to be independent of taxonomic philosophy. This is necessary because perspectives change with discoveries and the application of new techniques. As an example, prior to the mid-1980s, prokaryotic taxonomy was largely determinative and did not necessarily seek to identify natural relationships among taxa described at the higher taxonomic ranks. Consequently, the ICNP placed great importance on the identification of diagnostic properties of the lower taxa. With the development of robust methods to determine genetic relationships, prokaryotic taxonomy transitioned to a phylogenetic approach, resulting in major changes in the nomenclature of individual taxa [10,11,60]. The ICNP provided the framework that guided this process, and many of the names created under the determinative taxonomy survived the transition to a phylogenetic taxonomy.

Comparison of the ICNP and the SeqCode

The SeqCode was developed over the last two years to allow the naming of prokaryotes based on their genome sequences as nomenclatural types. Thus, it will allow the naming of the enormous uncultivated biodiversity discovered through the aid of metagenomic sequencing and single-cell genomics as well as fastidious prokaryotes that cannot reasonably be maintained and distributed by culture collections [34]. However, the SeqCode will accomplish other things as well: (i) The naming rules will be largely consistent with those of the ICNP [44], and it is hoped that the nomenclatures formed under both codes will eventually be merged. (ii) The ICNP is difficult to read, and some parts are contradictory. The SeqCode strives to resolve ambiguities without contradicting the nomenclature developed under the ICNP. (iii) By being more plainly written, less apt to alternative interpretations and easier to understand, it will reduce controversy and confusion. (iv) The SeqCode will provide an online system for registration and curation of names. (v) The SeqCode will provide means to generate names with priority for taxa that can presently only be named under the ICNP with the provisional *Candidatus* status.

Major differences between the SeqCode and ICNP are summarized in Table 1. In addition, a detailed comparison of the two codes is provided in the Supplementary Material.

In accomplishing these goals, the SeqCode is evolutionary and not revolutionary. It adds no new ideas to the canon of biological naming. For instance, the most controversial aspect of this code is that it does not require physical evidence for the taxon in the form of viable cultures. However, neither did the 1990 International Code of Nomenclature of Bacteria [26], which allowed detailed descriptions and illustrations to serve as nomenclatural types until January 2001 [44]. Similarly, the International Code of Virus Classification and Nomenclature (https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code) does not require deposition of physical samples, and none of the other codes requires living samples. The goals of the SeqCode are practical and not philosophical, and it seeks to make naming easy and yet clearly regulated. Ease is an important virtue for nomenclatural codes because it encourages wide use. The diversity of prokaryotes is enormous and far exceeds the capacity of contributors to any specialized discipline to fully understand. Hence it is hoped that the SeqCode will encourage participation by those working in diverse disciplines.

Nomenclatural types

In addition to providing guidelines for naming, nomenclatural codes also provide standards for the evidence that can be considered as type. While recognized standards are necessary to prevent

Table 1
Comparison of the SeqCode and ICNP.

Category	ICNP	SeqCode	Comments
Type	Viable culture	DNA sequence and viable culture	The SeqCode recognizes ICNP names with priority
Criteria of authenticity	Viable culture representing type is deposited in two culture collections in different countries.	Sequence representing type is deposited in an INSDC database and accession number is cited in effective publication and online Registry.	
Registration of names	Publication in IJSEM or Validation Lists	Online SeqCode Registry	SeqCode registration requires an effective publication
Compatibility of names with databases	None, requires manual curation	Database structure of registration system avoids a need of third parties and allows easy access, search and retrieval of information.	
Governing body	International Committee on the Systematics of Prokaryotes (ICNP)	Committee on the Systematics of Prokaryotes Described from Sequence Data (ICSPSDS)	
Number of taxonomic ranks above genus	9	4	SeqCode does not recognize many rarely utilized taxonomic ranks
Type for taxonomic ranks above genus	Genus except class, which is order	Genus for all higher taxa	
Priority for taxonomic ranks above genus	Depends on date the name is validly published	Depends on the date the type genus is validly published	
Number of rules	65	50	Omission of rules related to unused taxa and names prior to the Approved List allowed reduction of SeqCode
Start date	1947	January 1, 2022	

the creation of imprecise names and other abuses, they necessarily restrict the freedom of taxonomic thought. For instance, since 2001, Rule 30 of the ICNP has required deposition of viable type strains into two service collections [7,25,31,44]. The deposition of axenic cultures is a restrictive methodological standard and prevents the expansion of nomenclature to the large numbers of uncultured prokaryotes identified by metagenomic sequencing. Many fastidious prokaryotes are also excluded because service collections lack the facilities to cultivate them. Lastly, some organisms are excluded by international laws preventing the export of biological material from their countries, which include several recognized biodiversity hotspots [57]. Although these organisms may be readily culturable, at the present they cannot be freely distributed through culture collections and so are disbarred from formal, valid naming. The restrictive nature of the ICNP is discordant with the more inclusive nature of all other codes of nomenclature, and by restricting names to a small range of the full diversity of Archaea and Bacteria, it fails to serve the greater microbiology research community [43].

Candidatus names were proposed to ameliorate some of the consequences of Rule 30 and are described in an appendix of the ICNP [32,33,40]. The *Candidatus* provisional category was intended for taxa known originally by 16S rRNA gene data and other data, principally morphology, ecology, or physiology, but has since been co-opted to include MAGs and SAGs [20] as well as pure cultures which could not be deposited in culture collections. Although over a thousand *Candidatus* names were proposed before 2019 [37,38], these names lack priority and other protections offered by the legislative section of the ICNP. Therefore, *Candidatus* names fail to accomplish the major goals of nomenclatural codes and have not replaced many of the alphanumeric identifiers widely used for decades for many uncultured taxa. Most of these alphanumeric codes are imprecisely defined and many suffer from a high degree of synonymy, lack any concept of rank, and are difficult to remember because they are memorized and recalled as strings of words, letters, and numbers rather than a single word for a formal taxonomic name, which strains memory and easy recall [22,29,43].

The major difference between the SeqCode and the ICNP is that the SeqCode allows DNA sequences, primarily genome sequences to serve as types. Genome sequences satisfy all the criteria necessary for types [61]. They prove the evidence of the existence of taxa and allow ready comparisons to the types of other taxa. Moreover, they offer many practical advantages over type strains [22,43]. They are suitable for both uncultivated and cultivated taxa. They are inexpensive to store and easy to share. They create a permanent record, whereas culture collections require regular maintenance, and the viability of archival material is uncertain. A mechanism is also proposed to create minimum standards to ensure data quality of the SeqCode. The minimum standards are themselves outside the code to allow for flexibility as experimental methods change. Thus, the SeqCode also requires a working group to review and approve minimum standards. Previously proposed standards for isolate genomes, MAGs and SAGs served as guides for these standards [1,5,6,9,21]. In addition, less restrictive minimum standards could be proposed by taxonomists with expertise in specific taxa. For instance, it may not be possible at the present time to generate genome sequences for certain obligate endosymbionts. If experts familiar with these organisms have experimentally validated alternative, sequence-based procedures such as multilocus sequence analyses, they will be acceptable types under the SeqCode.

Priority of higher taxa

Elements of the ICNP are poorly suited for a phylogenetic classification. When most of the backbone of the ICNP was written in the second half of the last century, the extent of prokaryotic diversity was greatly underestimated. Many genera were not assigned to higher taxonomic ranks, and class was the highest recognized rank. The Approved Lists from 1980 contained only seven named classes and 58 names in total above the rank of family [55]. Moreover, there was no standardization in the formation of class names. At that time, the ICNP heavily emphasized naming species and genera, and rules for naming the higher taxonomic ranks were rarely discussed and remained ambiguous. In contrast, today there are more than five hundred validly published names of taxa above the rank of family [47]. More than a hundred classes have been named, and the rank of phylum has become widely used even though it has only recently been incorporated into the ICNP. As a consequence of the ambiguities in the ICNP regarding higher taxa, many of these names were not created in a standard manner, resulting in contradictions in the nomenclature.

With the exception of classes, the types of the higher taxonomic ranks in the ICNP are genera, and the genus names serve as the roots of their names. Although the genus with the earliest name may be chosen as the type of a family or order, the ICNP does not provide clear direction in this regard and the fact that the type of a class is one of the contained orders introduces a discontinuity into the system. Moreover, the priority of the names of higher taxa depends on the date of validation of the name. Since naming higher taxa was not a common practice prior to Garrity et al. [11], these rules created the potential for instability and confusion. For instance, it has been argued that a validly published family name has priority even when its type genus is illegitimate [59,63]. It has also led to the creation of confusing names. For instance, under the ICNP, if the genera *Alcaligenes* Approved Lists 1980 and *Burkholderia* Yabuuchi et al. 1993 are united in the same family, the family is named *Alcaligenaceae* De Ley *et al.* 1986 and not *Burkholderiaceae* Garrity *et al.* 2006 because of the priority of *Alcaligenaceae* De Ley *et al.* 1986. However, because an order "*Alcaligenales*" has never been validly published, the family is classified in the order *Burkholderiales* Garrity *et al.* 2006.

In the SeqCode, the priority of the names of higher taxa depends on the priority of the genus name. The rationale is that, in a hierarchical classification, the creation of a new genus implies the potential for a new family, order, class, and phylum even if they are not named at the time. By using the earliest named genus as the type, future unions of genera are unlikely to change the name, thus ensuring the stability of the names of higher taxa. Moreover, if higher taxa are united, the name would be chosen from the one whose genus name had priority. If a higher taxon is divided, the branch that includes the type would retain the name. The newly recognized branch would acquire a name based upon the taxa immediately below it in rank. An entirely new higher taxon could only occur upon discovery of a novel genus. Importantly, the SeqCode accepts all names validly published under the ICNP before January 1, 2022. As a consequence, this rule has no effect on the names of higher taxa validly published before that date, and it only affects the names of higher taxa proposed after that date. For instance, any order that included the genus *Burkholderia* would be named *Burkholderiales* in both the ICNP and the SeqCode.

Simplification of the code

The ICNP distinguishes five categories of names: legitimate, illegitimate, effectively published, validly published, and correct. Legitimate names are formed according to the rules of the code, including being validly published. In the ICNP, validly published means registered either in the Approved Lists or by subsequent publication in the IJSEM or its Validation Lists. All legitimate names must be validly published. While in principle all validly published

names should be legitimate, this is not always the case. Sometimes names become illegitimate due to breach of the principle of priority when applied to merged taxa or changes in the rules of the Code. Lastly, correct names must be legitimate but are also those to be used in a particular classification. These distinctions are also used in the SeqCode. However, unlike the ICNP, the SeqCode uses an online registration system, the SeqCode Registry. In this system, the registration is completed by the authors of new names, preferably prior to publication. The names will then be reviewed and approved by the list curators. Upon completion of an entry with effective publication and type sequence details, the names will become validly published. This system ensures priority of names based on valid publication through the SeqCode Registry and not based on the date of publication in journals, which equalizes the opportunities of publishing names in any journal and do not favor one over others. Moreover, the SeqCode Registry makes provisions for names undergoing the proposal process, which are not yet validly published, reserving these names for up to a year to avoid synonymy with future publications. It also allows correction of names early in the validation process, which will decrease confusion and instability.

The goal of the SeqCode Registry is just to provide a list of names that are connected with metadata and compliant with the data standards and orthography of the SeqCode, and it is outside the mandate of nomenclature committees to dictate taxonomic principles to the user community. Nevertheless, recommendations for good practices for naming uncultivated taxa have been proposed by Chuvochina et al. [6]. Similarly, there are only minimal requirements for validation of names in the ICNP. As a consequence, the SeqCode will not prevent publications of papers with thousands of computer-generated names with little associated information should such papers be effectively published. However, we feel that such naming is of little value and should be discouraged. While thresholds based upon sequence similarities have proven to be invaluable tools, they are also well known to be unreliable as the sole criterion for taxon circumscription. Moreover, replacing alphanumeric labels with Latin names in the absence of further analyses misrepresents the state of knowledge of a taxon and leads to the proliferation of confusing names. In general, taxa should only be named when there is something to say about them.

Nevertheless, this concern is balanced by the consideration that providing names or labels are important steps that facilitate communication about taxa when they are further studied. While in an ideal situation naming would be concurrent with detailed descriptions, these two steps can be performed by different parties or the same party at different times. In either case, naming helps track knowledge as it is added and can serve as the starting rather than the end point of the search for understanding.

The SeqCode proposes two other major departures from the ICNP. The ICNP currently recognizes thirteen taxonomic ranks, while the SeqCode is restricted to seven canonical ranks: subspecies, species, genus, family, order, class, and phylum. The ranks of subgenus, subtribe, tribe, subfamily, suborder, and subclass, which are recognized in the ICNP, are not included in the SeqCode. These ranks are rarely used in the modern literature, and their conceptual and experimental bases are ambiguous [36]. Secondly, many of the Rules of the ICNP deal with naming prior to adoption of the Approved Lists and Validation Lists. Now that all the current names are incorporated into these Lists, these Rules are no longer necessary.

Administration of the SeqCode

The SeqCode also proposes the creation of two administrative bodies to facilitate its implementation. The Committee on the Systematics of Prokaryotes Described from Sequence Data (colloquially,the SeqCode Committee) will be responsible for the content of the code and administering the infrastructure required for maintenance of the SeqCode Registry. Its functions are similar to those of the ICSP. The SeqCode Committee will also oversee the election of a SeqCode Reconciliation Commission with the authority to resolve disputes regarding the interpretations of the rules and to grant exceptions in unusual circumstances. Its functions are similar to those of the Judicial Commission of the ICSP. As noted above, a working group will be maintained to advise on appropriate standards and quality control of sequence data. Oversight and organizational support for these will be provided by ISME.

Conclusions

By creating the SeqCode, our intention is to produce a practical solution to the nomenclature of the huge number of taxa currently excluded from formal naming by the specific insistence in the ICNP that types must be viable cultures. We hope that this will provide a service to the wider community working on prokaryotic diversity and that, in due course, this will allow the formal nomenclature of all taxa to be brought under a unified framework.

The SeqCode is not intended to discourage isolation and collection of strains, and investigators are strongly encouraged to deposit strains into culture collections whenever feasible to promote resource sharing and reproducibility. The study of living cultures is essential to progress in prokaryotic biology and obtaining a full understanding of life on Earth. Currently, it is not possible to infer the complete potential of Archaea and Bacteria from their genome sequences, and strains remain important tools for fully understanding their pathogenicity, biotechnological applications, physiology, ecology, lifestyle, and other properties of great practical and theoretical importance. For these reasons, strains will retain their value even when they are not nomenclatural types. This principle is well illustrated by the recent isolation and characterization of *Prometheoarchaeum syntrophicum* by Imachi et al. [19]. This archaeon possesses the signature features of the Asgard archaea. Described initially from MAGs, these deep sea microorganisms are likely descendants of the ancestors of the archaeal component of the first eukaryotes [56,68], and the availability of a representative culture provides key insights into the evolution of the first eukaryotes [19]. Moreover, naming uncultivated microorganisms facilitates all subsequent investigations whatever their scope and nature. Importantly, it allows the creation of classifications unifying cultured and uncultured taxa, identifying key taxa for environmental and other processes, and encourages the isolation and characterization of representative uncultivated taxa. In this manner, the SeqCode will encourage cultivation of species of prokaryotes known now only from their genome sequences. Of great importance, it will foster the development of synergies between the microbial disciplines of molecular biology, biochemistry, physiology, systematics, environmental science and ecology. Already evident, the benefits of this emergent understanding will continue to have enormous consequences on the study of life.

Acknowledgements

Funding was provided by the US National Science Foundation (DEB 1841658 and EAR 1516680), the US National Institute of General Medical Sciences (P20 GM103440) from the National Institutes of Health, the Spanish Ministry of Science, Innovation and Universities (PID2021-126114NB-C42), the Australian Research Council (FL150100038), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, SFB 1439/1 2021 – 426547801) also supported with European Regional Development Funds (FEDER), and the International Society for Microbial Ecology (ISME). We also

thank all participants in the SeqCode workshops, especially guest speakers who graciously shared their expertise: Jongsik Chun, Nicole Dubilier, Emiley Eloe-Fadrosh, Chris Lane, Juncai Ma, Edward Moore, Aharon Oren, Jörg Overmann, Susanne Renner, Vincent Robert, Conrad Schoch, Scott Tinghe, Linhuan Wu, and Arvind Varsani.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.syapm.2022.126305.

References

[1] Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., The Genome Standards Consortium, , Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35, 725–731. https://doi.org/10.1038/nbt.3893.

[2] Buchanan, R.E., Holt, J.G., Lessel Jr., E.F. Index Bergeyana. Williams & Wilkins, Baltimore, p. 733.

[3] Chan, J.Z.-M., Halachev, M.R., Loman, N.J., Constantinidou, C., Pallen, M.J. (2012) Defining bacterial species in the genomic era: insight from the genus *Acinetobacter*. BMC Microbiol. 12, 302. https://doi.org/10.1186/1471-2180-12-302.

[4] Chen, L.X., Anantharaman, K., Shaiber, A., Eren, A.M., Banfield, J.F. (2020) Accurate and complete genomes from metagenomes. Genome Res. 30, 315–333. https://doi.org/10.1101/gr.258640.119.

[5] Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.W., De Meyer, S., Trujillo, M.E. (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int. J. Syst. Evol. Microbiol. 68, 461–466. https://doi.org/10.1099/ijsem.0.002516.

[6] Chuvochina, M., Rinke, C., Parks, D.H., Rappe, M.S., Tyson, G.W., Yilmaz, P., Whitman, W.B., Hugenholtz, P. (2019) The importance of designating type material for uncultured taxa. Syst. Appl. Microbio. 42, 15–21. https://doi.org/10.1016/j.syapm.2018.07.003.

[7] De Vos, P., Trueper, H.G. (2000) Minutes of the judicial commission of the international committee on systematic bacteriology. Int. J. Syst. Evol. Microbiol. 50, 2239–2244.

[8] Estrada-de los Santos, P., Palmer, M., Chavez-ramirez, B., Beukes, C., Steenkamp, E.T., Briscoe, L., Khan, N., Maluk, M., Lafos, M., Humm, R.E., Arrabit, M., Crook, M., Gross, E., Simon, M.F., Bueno do Reis Jr., F., Whitman, W. B., Shapiro, N., Poole, P.S., Hirsch, A.M., Venter, S.N., James, E.K. (2018) Whole genome analyses suggests that *Burkholderia senu lato* contains two additional novel genera (*Mycetohabitans gen. nov*., and *Trinickia gen. nov*.): Implications for the evolution of diazotrophy and nodulation in the *Burkholderiaceae*. Genes 9, 389. https://doi.org/10.3390/genes9080389.

[9] Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., Allen, M.J., Angiuoli, S.V., Ashburner, M., Axelrod, N., Baldauf, S., Ballard, S., Boore, J., Cochrane, G., Cole, J., Dawyndt, P., De Vos, P., dePamphilis, C., Edwards, R., Faruque, N., Feldman, R., Gilbert, J., Gilna, P., Glöckner, F.O., Goldstein, P., Guralnick, R., Haft, D., Hancock, D., Hermjakob, H., Hertz-Fowler, C., Hugenholtz, P., Joint, I., Kagan, L., Kane, M., Kennedy, J., Kowalchuk, G., Kottmann, R., Kolker, E., Kravitz, S., Kyrpides, N., Leebens-Mack, J., Lewis, S.E., Li, K., Lister, A.L., Lord, P., Maltsev, N., Markowitz, V., Martiny, J., Methe, B., Mizrachi, I., Moxon, R., Nelson, K., Parkhill, J., Proctor, L., White, O., Sansone, S.-A., Spiers, A., Stevens, R., Swift, P., Taylor, C., Tateno, Y., Tett, A., Turner, S., Ussery, D., Vaughan, B., Ward, N., Whetzel, T., San Gil, I., Wilson, G., Wipat, A. (2008) The minimum information about a genome sequence (MIGS) specification. Nat. Biotechnol. 26, 541–547. https://doi.org/10.1038/nbt1360.

[10] Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., Woese, C.R. (1980) The phylogeny of prokaryotes. Science 209, 457–463. https://doi.org/10.1126/science.6771870.

[11] Garrity, G.M., Bell, J.A., Lilburn, T.G. (2001) Taxonomic outline of the prokaryotes. Bergey's manual of systematic bacteriology, vol. 1. . . . . .

[12] Gomila, M., Peña, A., Mulet, M., Lalucat, J., García-Valdés, E. (2015) Phylogenomics and systematics in Pseudomonas. Front. Microbiol. 6, 214. https://doi.org/10.3389/fmicb.2015.00214.

[13] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91. https://doi.org/10.1099/ijs.0.64483-0.

[14] Hahnke, R., Meier-Kolthoff, J.P., López, M.G., Mukherjee, S., Huntemann, M., Ivanova, N.N., Woyke, T., Kyrpides, N.C., Klenk, H.-P., Göker, M. (2016) Genome-based taxonomic classification of *Bacteroidetes*. Front. Microbiol. 7, 2003. https://doi.org/10.3389/fmicb.2016.02003.

[15] Hedlund, B.P., Dodworth, J.A., Staley, J.T. (2015) The changing landscape of microbial diversity exploration and its implications for systematics. Syst. Appl. Microbiol. 38, 231–236. https://doi.org/10.1016/j.syapm.2015.03.003.

[16] Hedlund, B.P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K.T., Murray, A. E., Palmer, M., Parks, D.H., Probst, A.J., Reysenbach, A.-L., Rodriguez-R, L.M., Rossello-Mora, R., Sutcliffe, I., Venter, S.N., Whitman, W.B. (2022) SeqCode, a nomenclatural code for prokaryotes described from sequence data. Microbiol. Nat. (in press)

[17] Hitchcock, A.S. (1921) The type concept in systematic botany. Am. J. Botany 8, 251–255.

[18] Hördt, A., López, M.G., Meier-Kolthoff, J.P., Schleuning, M., Weinhold, L.-M., Tindall, B.J., Gronow, S., Kyrpides, N.C., Woyke, T., Göker, M. (2020) Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of *Alphaproteobacteria*. Front. Microbiol. 11, 468. https://doi.org/10.3389/fmicb.2020.00468.

[19] Imachi, H., Nobu, M.K., Nakahara, N., et al. (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. Nature 577, 519–525. https://doi.org/10.1038/s41586-019-1916-6.

[20] Konstantinidis, K., Rosselló-Móra, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomic data in the Candidatus proposal. Syst. Appl. Microbiol. 38, 223–230.

[21] Konstantinidis, K.T., Rosselló-Móra, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. ISME J. 11, 2399–2406. https://doi.org/10.1038/ismej.2017.113.

[22] Konstantinidis, K.T., Rossello-Mora, R., Amann, R. (2020) Advantages outweigh concerns about using genome sequence as type material for prokaryotic taxonomy. Env. Microbiol. 22, 819–822. https://doi.org/10.1111/1462-2920.14934.

[23] Konstantinidis, K.T., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U.S.A. 102, 2567–2572. https://doi.org/10.1073/pnas.0409727102.

[24] Kyrpides, N.C., Woyke, T., Eisen, J.A., Garrity, G., Liburn, T.G., Beck, B.J., Whitman, W.B., Hugenholz, P., Klenk, H.-P. (2014) Genomic Encyclopedia of Type Strains, Phase I: The one thousand microbial genomes (KMG-I) project. Stand. Genomic Sci. 9, 1278–1284. https://doi.org/10.4056/sigs.5068949.

[25] Labeda, D.P. (2000) Minutes of the international committee on systematic bacteriology. Int. J. Syst. Evol. Microbiol. 50, 2245–2247.

[26] Lapage, S.P., Sneath, P.H.A., Lessel, E.F., Skerman, V.B.D., Seeliger, H.P.R., Clark, W.A. (1992) International Code of Nomenclature of Bacteria. American Society for Microbiology, Washington, D.C.

[27] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform. 14, 60. https://doi.org/10.1186/1471-2105-14-60.

[28] Meziti, A., Rodriguez-R, L.M., Hatt, J.K., Peña-Gonzalez, A., Levy, K., Konstantinidis, K.T. (2021) The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. Appl. Environ. Microbiol. 87, e02593–e02620. https://doi.org/10.1128/AEM.02593-20.

[29] Miller, G. (1956) The magical number seven, plus or minus two-the magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63 (2), 81–97.

[30] Moss, E.L., Maghini, D.G., Bhatt, A.S. (2020) Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat. Biotechnol. 38, 701–707. https://doi.org/10.1038/s41587-020-0422-6.

[31] Murray, R.G.E. (1996) Taxonomic note: a rule about the deposition of type strains. Int. J. Syst. Bacteriol. 46, 831.

[32] Murray, R.G.E., Schleifer, K.-H. (1994) Taxonomic notes: a proposal for recording the properties of putative procaryotes. Int. J. Syst. Bacteriol. 44, 174–176.

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

[34] Murray, A.E., Freudenstein, J., Gribaldo, S., Hatzenpichler, R., Hugenholtz, P., Kämpfer, P., Konstantinidis, K.T., Lane, C.E., Papke, R.T., Parks, D.H., Rossello-Mora, R., Stott, M.B., Sutcliffe, I.C., Thrash, J.C., Venter, S.N., Whitman, W.B., Acinas, S.G., Amann, R.I., Anantharaman, K., Armengaud, J., Baker, B.J., Barco, R. A., Bode, H.B., Boyd, S.B., Brady, C.L., Carini, P., Chain, P.S.G., Colman, D.R., DeAngelis, K.M., de los Rios, M.A., Estrada-de los Santos, P., Dunlap, C.A., Eisen, J.A., Emerson, D., Ettema, T.J.G., Eveillard, D., Girguis, P.R., Hentschel, U., Hollibaugh, J.T., Hug, L.A., Inskeep, W.P., Ivanova, E.P., Klenk, H.-P., Li, W.-J., Lloyd, K.G., Löffler, F.E., Makhalanyane, T.P., Moser, D.P., Nunoura, T., Palmer, M., Parro, V., Pedrós-Alió, C., Probst, A.J., Smits, T.H.M., Steen, A.D., Steenkamp, E.T., Spang, A., Stewart, F.J., Tiedje, J.M., Vandamme, P., Wagner, M., Wang, F.-P., Yarza, P., Hedlund, B.P., Reysenbach, A.-L. (2020) Roadmap for naming uncultivated Archaea and Bacteria. Nat. Microbiol. 5, 987–994. https://doi.org/10.1038/s41564-020-0733-x.

[35] Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T.B.K., Nielsen, T., Kirton, E., Faria, J.P., Edirisinghe, J.N.,

Henry, Jungbluth, S.P., Chivian, D., Dehal, P., Wood-Charlson, E.M., Arkin, A.P., Tringe, S.G., Visel, A., IMG/M Data Consortium, Woyke, T., Mouncey, N.J., Ivanova, N.N., Kyrpides, N.C., Eloe-Fadrosh, E.A. (2021) A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 39, 499–509. https://doi.org/10.1038/s41587-020-0718-6.

[36] Oren, A. (2019) Proposal to modify the Rules of the International Code of Nomenclature of Prokaryotes to abolish the taxonomic categories Subfamily, Subtribe and Kingdom. Int. J. Syst. Evol. Microbiol. 69, 1524–1525. https://doi.org/10.1099/ijsem.0.003317.

[37] Oren, A., Garrity, G.M. (2021) Candidatus List No.2. List of names of prokaryotic *Candidatus* taxa. Int. J. Syst. Evol. Microbiol. 71, 004671. https://doi.org/10.1099/ijsem.0.004671.

[38] Oren, A., Garrity, G.M., Parker, C.T., Chuvochina, M., Trujillo, M.E. (2020) List of names of *Candidatus* taxa. Int. J. Syst. Evol. Microbiol. 70, 003789. https://doi.org/10.1099/ijsem.0.003789.

[39] Pallen, M.J. (2021) Bacterial nomenclature in the era of genomics. New Microbe New Infect. 44, 100942. https://doi.org/10.1016/j.nmni.2021.100942.

[40] Pallen, M.J. (2021) The status *Candidatus* for uncultured taxa of Bacteria and Archaea: SWOT analysis. Int. J. Syst. Evol. Microbiol. 71, 005000. https://doi.org/10.1099/ijsem.0.005000.

[41] Pallen, M.J., Telatina, A., Oren, A. (2020) The next million names for Archaea and Bacteria. Trends Microbiol. https://doi.org/10.1016/j.tim.2020.10.009.

[42] Palmer, M., Steenkamp, E.T., Blom, J., Hedlund, B.P., Venter, S.N. (2020) All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. Int. J. Syst. Evol. Microbiol. 68, 1396–1407. https://doi.org/10.1099/ijsem.0.004124.

[43] Palmer, M., Sutcliffe, I., Venter, S.N., Hedlund, B.P. (2022) It is time for a new type of type to facilitate naming the microbial world. New Microbes New Infect. 47, 100991. https://doi.org/10.1016/j.nmni.2022.100991.

[44] Parker, C.T., Tindall, B.J., Garrity, G.M. (2019) International code of nomenclature of prokaryotes. Int. J. Syst. Evol. Microbiol. 69, S1–S111. https://doi.org/10.1099/ijsem.0.000778.

[45] Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., Hugenholz, P. (2020) A complete domain-to-species taxonomy for Bacteria and Archaea. Nat. Biotechnol. 38, 1079–1086. https://doi.org/10.1038/s41587-020-0501-8.

[46] Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholz, P. (2021) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucl. Acid Res. https://doi.org/10.1093/nar/gkab776.

[47] Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C., Göker, M. (2020) List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. Int. J. Syst. Evol. Microbiol. 70, 5607–5612. https://doi.org/10.1099/ijsem.0.004332.

[48] Quince, C., Walker, A., Simpson, J., Lomans, N.J., Segata, N. (2017) Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35, 833–844. https://doi.org/10.1038/nbt.3935.

[49] Richter, M., Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U.S.A. 106, 19126–19131. https://doi.org/10.1073/pnas.0906412106.

[50] Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.-A., Davin, A.A., Waite, D.W., Whitman, W.B., Parks, D.H., Hugenholz, P. (2021) A standardized archaeal taxonomy for the Genome Taxonomy Database. Nat. Microbiol. 6, 946–959. https://doi.org/10.1038/s41564-021-00918-8.

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

[52] Rossello-Mora, R., Whitman, W.B. (2018) Dialogue on the nomenclature and classification of prokaryotes. Syst. Appl. Microbiol. 42, 5–14. https://doi.org/10.1016/j.syapm.2018.07.002.

[53] Sangal, V., Goodfellow, M., Jones, A.L., Schwalbe, E.C., Blom, J., Hoskisson, P.A., Sutcliffe, I.C. (2016) Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. Sci. Rep. 6, 38392. https://doi.org/10.1038/srep38392.

[54] Sharon, I., Banfield, J.F. (2013) Microbiology. Genomes from metagenomics. Science 342, 1057–1058. https://doi.org/10.1126/science.1247023.

[55] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. Approved Lists of Bacterial Names. American Society for Microbiology, Washington, D.C..

[56] Spang, A., Saw, J., Jørgensen, S., et al. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173–179. https://doi.org/10.1038/nature14447.

[57] Sutcliffe, I.C., Dijkshoorn, L., Whitman, W.B. (2020) Minutes of the International Committee on Systematics of Prokaryotes online discussion on the proposed use of gene sequences as type for naming of prokaryotes, and outcome of vote. Int. J. Syst. Evol. Microbiol. 70, 4416–4417. https://doi.org/10.1099/ijsem.0.004303.

[58] Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43. https://doi.org/10.1038/nature02340.

[59] Tindall, B.J. (2014) The family name Solimonadaceae Losey 2013 is illegitimate, proposals to create the names ' *Sinobacter soli*' comb. nov. and ' *Sinobacter variicoloris*' contravene the Code, the family name *Xanthomonadaceae* Saddler and Bradbury 2005 and the order name *Xanthomonadales* Saddler and Bradbury 2005 are illegitimate and notes on the application of the family names *Solibacteraceae* Zhou *et al.*, 2008, *Nevskiaceae* Henrici and Johnson 1935 (Approved Lists 1980) and *Lysobacteraceae* Christensen and Cook 1978 (Approved Lists 1980) and order name *Lysobacteriales* Christensen and Cook 1978 (Approved Lists 1980) with respect to the classification of the corresponding type genera *Solibacter* Zhou *et al.* 2008, *Nevskia* Famintzin 1892 (Approved Lists 1980) and *Lysobacter* Christensen and Cook 1978 (Approved Lists 1980) and importance of accurately expressing the link between a taxonomic name, its authors and the corresponding description/circumscription/emendation. Int. J. Syst. Evol. Microbiol. 64, 293–297.

[60] Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, L., Moore, L.H., Moore, W.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Trüper, H.G. (1987) Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.

[61] Whitman, W.J. (2015) Genome sequences as the type material for taxonomic descriptions of prokaryotes. Syst. Appl. Microbiol. 38, 217–222. https://doi.org/10.1016/j.syapm.2015.02.003.

[62] Whitman, W.B. (2016) Modest proposal to expand the type material for naming of prokaryotes. Int. J. Syst. Evol. Microbiol. 66, 2108–2112. https://doi.org/10.1099/ijsem.0.000980.

[63] Whitman, W.B., Lawson, P.A., Losey, N.A. (2015) Response to Tindall (2014) on the legitimacy of the names *Solimonadaceae* Losey *et al.* 2013, *Xanthomonadaceae* Saddler and Bradbury 2005 and *Xanthomonadales* Saddler and Bradbury 2005. Int. J. Syst. Evol. Microbiol. 65, 1086–1087. https://doi.org/10.1099/ijs.0.000061.

[64] Whitman, W.B., Sutcliffe, I.C., Rossello-Mora, R. (2019) Proposal for changes in the International Code of Nomenclature of Prokaryotes: granting priority to *Candidatus* names. Int. J. Syst. Evol. Microbiol. 69, 2174–2175. https://doi.org/10.1099/ijsem.0.003419.

[65] Winston, J.E. (2018) Twenty-first century biological nomenclature - the enduring power of names. Integr. Comp. Biol. 58, 1122–1131. https://doi.org/10.1093/icb/icy060.

[66] Wu, D., Hugenholtz, P., Mavromatis, K. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462, 1056–1060. https://doi.org/10.1038/nature08656.

[67] Wu, L., McCluskey, K., Desmeth, P., Liu, S., Hideaki, S., Yin, Y., Moriya, O., Itoh, T., Kim, C.Y., Lee, J.-S., Zhou, Y., Kawasaki, H., Hazbón, M.H., Robert, V., Boekhout, T., Lima, N., Evtushenko, L., Boundy-Mills, K., Bunk, B., Moore, E.R.B., Eurwilaichitr, L., Ingsriswang, S., Shah, H., Yao, S., Jin, T., Huang, J., Shi, W., Sun, Q., Fan, G., Li, W., Li, X., Kurtböke, I., Ma, J. (2018) The global catalogue of microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species. GigaScience 7, 1–4. https://doi.org/10.1093/gigascience/giy026.

[68] Zaremba-Niedzwiedzka, K., Caceres, E., Saw, J., et al. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353–358. https://doi.org/10.1038/nature21031.

[69] Zheng, J., Wittouck, S., Salvetti, E., Franz, C.M.A.P., Harris, H.M.B., Mattarelli, P., O'Toole, P.W., Pot, B., Vandamme, P., Walter, J., Watanabe, K., Wuyts, S., Felis, G.E., Gänzle, M.G., Lebeer, S. (2020) A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. Int. J. Syst. Evol. Microbiol. 70, 2782–2858. https://doi.org/10.1099/ijsem.0.004107.