Systematic and Applied Microbiology 38 (2015) 217–222

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Genome sequences as the type material for taxonomic descriptions of prokaryotes

William B. Whitman*

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

ARTICLE INFO

Keywords:
Genomics
Type strain
Bacteriological Code

ABSTRACT

Genome sequencing of type strains promises to revolutionize prokaryotic systematics by greatly improving the identification of species, elucidating the functional properties of taxonomic groups, and resolving many of the ambiguities in the phylogeny of the higher taxa. Genome sequences could also serve as the type material for naming prokaryotic taxa, which will greatly expand the nomenclature governed by the Bacteriological Code to include many fastidious and uncultured organisms and endosymbionts of great biological interest.

The genome sequencing of type strains promises great advances in the systematics of prokaryotes. In addition to improving the general understanding of prokaryotic biology, these advances include improved: (1) identification of prokaryotic species, (2) identification of functional characteristics useful for resolving taxonomic groups, and (3) resolution of the phylogeny of higher taxa. For many prokaryotic species, the genome sequence could also replace live cultures as the type material. This practice would be especially useful for prokaryotes that are difficult to cultivate or maintain in culture collections.

A major goal in prokaryotic systematics is to delineate the relationships of new isolates with the type strains that serve as the basis for taxonomic classification. The focus on type strains follows from the development of the Bacteriological Code [28] and the Approved List [51]. Prior to the Approved List, tens of thousands of bacterial names were present in the literature [52]. However, the descriptions associated with many of these names were so vague that it was impossible to know to what the names referred. Many of the names were also redundant, with the some organisms possessing multiple names. The Approved List discarded all names that were ambiguous, retaining about 2500 names that were either clearly associated with a biological specimen, i.e. a culture, or detailed and unambiguous descriptions. The Bacteriological Code then insured that all future names would possess clear descriptions, usually by deposition of a representative culture in a public culture collection. It also introduced a system for naming the higher taxonomic ranks based upon the genus names of the type strains and a system of priority. For instance, strain ATCC 6051 is the type 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 that includes strain ATCC 6051 must be named *B. subtilis*. Similarly, any genus that includes *B. subtilis* must be named *Bacillus*. Because the Bacteriological Code specifies that the name of the higher taxonomic ranks is determined by the name of the genus, the higher taxonomic ranks are similarly constrained. Thus, the family and order that include strain ATCC 6051 must be named *Bacillaceae* and *Bacillales*, respectively, unless they include a species with greater priority, i.e. validly described at an earlier date. This system allows for naming novel species by inserting them into the existing taxonomy. For instance, a new species similar to *B. subtilis* might be named *Bacillus*. A new species less similar to *B. subtilis* might be given a unique genus name but included in the family *Bacillaceae*. An even less similar species might be given unique genus and family names but included in the order *Bacillales*. This clever system insures the stability of names by preventing subsequent authors from overwriting the established nomenclature with their own names.

Two decisions are paramount in this system. One, is the isolate a new species? Two, if an isolate is a new species, what higher taxonomic classifications are appropriate? Genomics will play important roles in addressing both of these questions.

Genomics for species delineation

In the original proposal for the delineation of species based upon genome similarity [59], two measures of genetic relatedness were proposed to set the boundary for prokaryotic species. The first

* Tel.: +1 706 542 4219; fax: +1 706 542 2674.
E-mail address: whitman@uga.edu

http://dx.doi.org/10.1016/j.syapm.2015.02.003
0723-2020/

measure was the change in the melting temperature (or ΔT_m) of heteroduplex DNA formed upon annealing the DNAs from the two strains to be tested. The ΔT_m is directly related to the sequence identity of the DNAs, and a ΔT_m of about 5 °C, the cutoff proposed for prokaryotic species, corresponds to an average sequence identity of about 92% between the hybridizing DNA [7,18]. A second measure was also suggested to be of equal importance, the extent of DNA–DNA hybridization (or DDH). This would be the fraction of DNA capable of forming heteroduplexes under optimal conditions, usually 25 °C below the melting temperature of the homoduplexes. Importantly, it was the DNA sequence itself and not the method used for determining relatedness that was proposed as the ultimate standard for prokaryotic species delineation [59].

These criteria can now be replaced by Overall Genome Relatedness Indices (or OGRI) derived from the genome sequences [9]. ΔT_m and DDH are laborious to determine and prone to errors [21,50]. With the availability of many new genome sequences, it is now possible to calculate surrogates for both ΔT_m and DDH with a very high precision directly from the genome sequence [3,14,46]. This approach will provide the highest possible resolution and much higher reproducibility in delineating species. The average nucleotide identity (or ANI) is a good surrogate for the ΔT_m because it only compares the sequence identity of DNAs that meet a certain threshold of similarity, usually defined by a BLAST score [23]. ANI is readily determined at EzGenome or JSpecies, which calculates the ANI based upon either the BLAST algorithm or the rapid alignment tool MUMmer [9,46]. The Genome Blast Distance Phylogeny tool (GBDP) offers multiple ORGIs to estimate the DDH and genome sequence identity [2,3]. For closely related strains, these genome-based tools yield values highly correlated with DDH and other measures of genome relatedness [3,14,46]. Lastly, specI is a species identification tool developed to form species clusters based on 40, universal, single-copy phylogenetic marker genes [36].

Recent work suggests that criteria based upon surrogates either of the ΔT_m or DDH may yield substantively different results depending upon the taxon [32]. Because they measure very different properties of DNA, each of the cutoffs have very different implications for genetic relatedness [17,47]. ANI, formula d_4 of GBDP and specI measure sequence identity. They are similar to the standard measures of phylogenetic relatedness and measure the diversity acquired during the accumulations of substitutions and deletions by neutral and other evolutionary processes. In contrast, surrogates of DDH, such as formula d_6 of GBDP, measure the fraction of DNA that is homologous between two strains or the shared gene content and should reflect the prevalence of horizontal gene transfer (HGT) and other processes that insert or remove genes. Given that these are very different evolutionary processes, it is possible for the DNA of two stains to exceed the species cut off by one criterion but not the other. In fact, the ratio of the ANI and ORGIs based upon formula d_6 of GBDP, a surrogate of DDH, varies about two-fold among different prokaryotic lineages [32].

There are several advantages for using measures of sequence identity, whether or not they are based upon the entire genome, such as ANI or formula d_4 of GBDP, or small groups of genes, such as specI or multilocus sequence analyses [12,34,36]. Sequence identity is widely used in phylogenetic studies and is supported by a solid theoretical understanding of the evolutionary processes and a wealth of experimental evidence. Second, sequence identity has clock-like properties and provides the promise of correlation with times of divergence [25]. Discovery of divergence times of prokaryotic groups will enable correlation of the formation of species with the fossil record, the established evolutionary record of eukaryotes and the geological record [4,6]. Third, sequence identity has been used to proposed thresholds for higher taxa in addition to species [29,62]. Thus, classification can proceed by a uniform set of criteria from ancient to modern groups.

In contrast, surrogates of DDH possess many disadvantages. One of the major arguments for surrogates of DDH is that it extends a tradition of DDH as the major genetic criterion for prokaryotic species delineation [34]. This argument neglects the practical and theoretical difficulties of DDH. Although the DDH has been widely used, the accuracy of values reported in the literature is generally quite low [17]. DDH is not symmetrical. Thus, the DDH of strain A to strain B may be different from that of strain B to strain A. When this occurs, there is no theoretical basis for choosing the lessor or greater value or an average of the values. The DDH is also sensitive to the genome size, which is known to vary within species. Thus, even though DDH has been widely used to delineate prokaryotic species, it lacks a precise physical and chemical interpretation. While the 'average' DDH may provide a good sense of prokaryotic diversity, the particular values for any lineage are suspect.

Genomics for identification of functional properties of taxonomic groups

In addition to setting the criteria for delineation of species, genomics can play an important role in how thresholds are applied. While thresholds are necessary to maintain uniformity in taxonomic ranks among phylogenetic lineages, there are many reasons why they should be applied flexibly [12,35]. First, no matter what threshold is chosen, there will be certain groups that fall just below or above the threshold and would be inappropriately subdivided or grouped, respectively (Fig. 1). A related problem is the difficulty in applying thresholds to all strains in a species. For instance, if the threshold is 95% similarity, strains A and B may both possess 95% similarity to the type strain but <95% similarity to each other. In these cases, there may be little value in grouping these strains as separate species. Lastly, there will likely be some lineages where the evolutionary processes are so complex that comparisons of sequence similarity are of little value (Fig. 1D). Genomics will help recognize these lineages and avoid creation of superfluous species. Thus, thresholds are necessary but not sufficient for classification, and other factors such as the physiology and ecology of the groups being classified will have to be considered [49].

Because genome sequences provide enormous insights into the biology of organisms, they are an excellent tool for identifying features that will assist in the final classification [45]. For complex processes, such as development, stress response, quorum sensing, more will be inferred from the genome sequence than ever directly measured for most species. Likewise, for many fastidious organisms, more will probably be known about their physiology and metabolism from their genome sequence than it will ever be possible to observe directly. The genome sequence also provides enormous insights into the evolutionary processes within a group. By revealing deep insights into the biology of the organisms, genome sequencing will reveal the functional criteria most appropriate for creating biologically relevant classifications.

Historically, polyphasic taxonomy has served this role. Polyphasic taxonomy analyzes the relationships among prokaryotes by combining many types of evidence, from ecological to molecular, and often includes sequence as well as growth and chemotaxonomic data [10,19,56,58]. However, many of the growth tests and chemotaxonomic analyses are time-consuming and expensive to perform [57]. Because of their low reliability, the type strains must often be reanalyzed each time a new isolate is added to a group [56]. Importantly, these methods often provide very limited information about the biologically relevant properties of an organism or those properties likely to play a significant role in an organism's persistence in the environment or evolution. For instance, growth experiments are typically conducted in laboratory media with enormous quantities of single substrates. By their very nature,

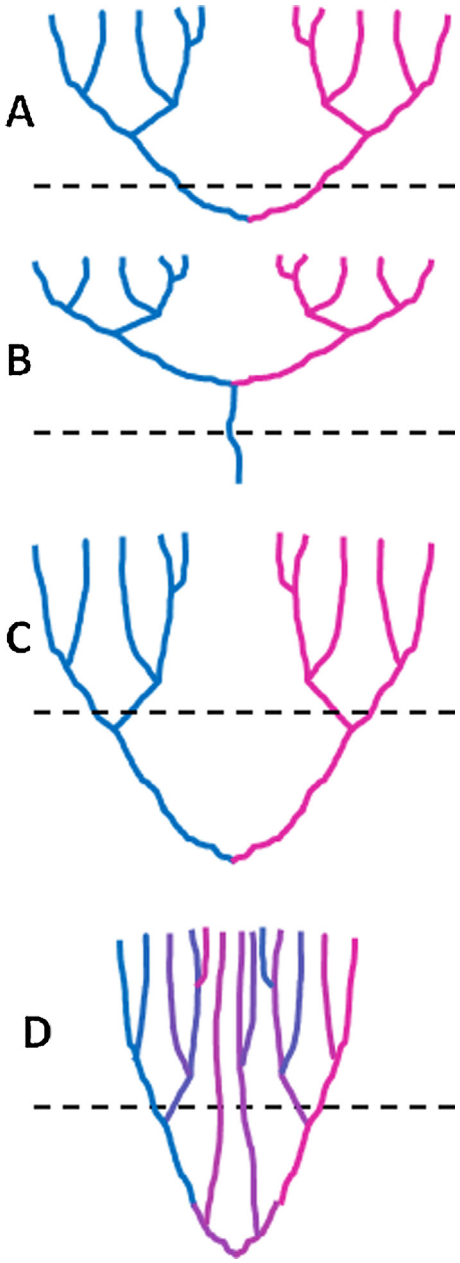

Fig. 1. Thresholds are necessary but not sufficient for species delineation. Thresholds based upon an Overall Genome Relatedness Index or OGRI is indicated by the dashed line in four hypothetical situations. Lineages with different biological properties are indicated by different colors. In (A) the threshold correctly distinguishes between two lineages with different properties. In (B) the threshold fails to distinguish between the two lineages, and they are inappropriately grouped together. In (C) the threshold inappropriately separates the two lineages into four by splitting the two groups with similar biological properties. In (D) the important biological properties of the strains are not distributed according to the phylogeny and it is inappropriate to differentiate specific lineages based upon thresholds.

these conditions are artificial with little relevance to environmental conditions where low concentrations of mixtures of substrates are common. Not surprisingly, growth rates common in the laboratory are seldom observed in nature. Similarly, while the chemical compositions of membrane, cell walls, and other cellular components are of occasional interest, they seldom provide insight into differences in the functional properties of closely related organisms. While polyphasic taxonomy was once good practice, genomics provides so much more genetic and functional information that this approach is unlikely to provide additional taxonomic insight. As the database of genome sequences becomes more complete, the usefulness of many chemotaxonomic tests will continue to diminish. For this reason, it is recommended that they be discontinued for most descriptions of species.

Genomics for delineation of higher taxa

The same concerns that exist at the species level also occur at the higher taxonomic ranks. In addition, the phylogeny of the taxonomic groups becomes critical at the higher taxonomic ranks where phylogenetic trees based upon single genes often fail to resolve the tree. For instance, 16S rRNA gene trees fail to resolve branching patterns of most of the orders within the class *Actinobacteria* [33]. Complete or draft genome sequences can overcome many of these limitations. Because many more genes are available for analysis, paralogous genes and genes with evidence for horizontal genetic transfer can be avoided. Likewise, it would be possible to correct for compositional differences resulting from changes in the G + C content or other factors and concatenate protein sequences to improve the phylogenetic signal. Thus, genomics may resolve many of the deep branches in the phylogenetic tree of prokaryotes.

Genomics in the description of novel species

The opportunity genomics provides to microbial systematics is enormous, and we should expect that genome sequences will become common in descriptions of new species. The sequences are just too valuable not to determine. Moreover, the value of the sequences increases with the size of the data base of genome sequences. According to a search of the NCBI database, at the end of 2014 complete or draft genome sequences were available for about 3600 type strains of prokaryotes. The DOE-JGI initiatives, such as the Genomic Encyclopedia of Bacteria and Archaea (GEBA) and the Thousand Microbial Genome (KMG) projects, seek to increase this valuable resource [24,26,61]. Thus, we should envision and plan for a future in which genome sequences are available for the majority if not all 12,000+ species type strains. For this reason, it is recommended that whenever possible the description of new species include a high quality draft or better genome sequence [8].

The genome sequence could also serve as the type material for species descriptions. There is no inherent conflict with this approach and the principles if not the rules of the Bacteriological Code. The principles of the Code only require that species be unique and completely identified prior to naming [60]. In 1990, the Bacteriological Code allowed for descriptions and preserved specimens to serve as the type [see Rule 18a [28]]. 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 [37].

The requirement for type strains composed of live cultures is relatively modern and follows from two related changes in the rules. The first created the designation of *candidatus* for taxa of uncultured prokaryotes where the phylogeny was known by sequencing of single genes from environmental DNA and the morphology was known based upon fluorescence in situ hybridization or similar techniques [13,27,42]. These changes were adopted to prevent the creation of large numbers of poorly described taxa based upon clones of 16S rRNA or other genes from environmental samples. Because *candidatus* names do not possess priority, a second nomenclatural system was created. The reasons given for this special category were the following. The sequence of a single gene is not representative of the organism and that the "whole capability of the organism must be considered in taxonomic arrangements" [41]. In addition, polyphasic taxonomy is necessary for a proper classification [42]. In summary, the molecular methods used at the time were

not sufficiently precise to uniquely identify a prokaryotic species. However, these arguments did not anticipate the wealth of information available in the genome sequence, which is representative of the entire organism and greatly exceeds the requirements for *candidatus* status.

Currently, descriptions of new species also require the deposition of type strains in two public culture collections [27,40]. A major goal for this rule is to allow for an unequivocal identification of the species. Prior to widespread genome sequencing, additional phenotypic and genotypic tests were necessary to insure the uniqueness of isolates. Thus, a living culture of the type strain had to be preserved to serve as the standard for the description of future isolates. However, when genome sequences are available, comparison of these sequences can serve this purpose. Thus, while there may be many other excellent reasons to deposit cultures in publically available collections, the requirements for systematics can be met by the genome sequence alone.

The requirements for deposition of live cultures strongly biases prokaryotic nomenclature towards organisms that are easily cultivated by public culture collections. For instance, in the last ten years pure or highly enriched cultures have been obtained for a number of nitrifying archaea related to *Candidatus* Nitrosopumilus maritimus [5,11,16,22,30,31,39,43,44]. These fastidious prokaryotes are key components of the N-cycle and among the most abundant archaea in the ocean and soil. With one exception [53], most of the species remain *candidatus* in spite of extensive characterization and genome sequences. Thus, they are much better characterized than anticipated by the original *candidatus* proposal [42]. Similarly, many endosymbionts require complex growth conditions that are difficult to replicate in the laboratory. For instance, *Candidatus* Criblamydia sequanensis was isolated from river water by co-culture in the amoeba *Acanthamoebia castellanii* [55]. Given these exacting growth requirements, it is not reasonable to expect that viable cultures will ever be maintained by public culture collections. By preventing the formal description of fascinating and important prokaryotes like these, the rules encourage the development of an informal nomenclature unregulated by the Bacteriological Code and threaten to undue the decades of effort necessary to create the modern system [15].

Lastly, preservation of cultures of type strains comes at enormous expense and requires maintaining large culture collections dedicated to this purpose. Using genome sequences as the type material will provide an opportunity to reallocate resources in culture collections to the study of the most important and valuable cultures. The current system of type strains was developed when the number of known species was in the thousands. The number of described species is now about 12,000, and the number of species on earth is probably in the millions [62]. The current system where cultures of type strains must be preserved would require expansion of the culture collections at least a hundred-fold to fully describe the richness of prokaryotic life on earth. An alternative is to use genome or other sequences, which can be easily stored on electronic media, as the nomenclatural type material for new taxa. Culture collections can then focus of preserving the most biologically interesting cultures that are likely to be of the greatest scientific and economic value.

Genome sequence as type material

Therefore, it is recommended that the genome sequence be allowed serve as the type material when it has been obtained from either a clonal population or a single cell. In this case, samples of the DNA should be submitted to public culture collections to insure the authenticity of the sequence. When the genome sequence is not possible, as might be the case for obligate endosymbionts, multilocus sequence analyses would be a suitable alternative if it can be shown to provide an unambiguous identification of the species. However, metagenomes or other genome sequences inferred by combining sequences from many closely related but not identical strains are inherently ambiguous and are not suitable type material. This recommendation should be implemented by revision of parts of Rules 18a and 27 [3], which describe the requirement of the type of a species or subspecies [28], as below:

> [Rule 18a]. The type material of a species or subspecies must unambiguously identify the taxonomic group and is a designated strain or other material. A type strain is made up of living cultures of an organism which are descended from a strain designated as the nomenclatural type. The strain should have been maintained in pure culture and should agree closely in its characters with those of the original description (see Chapter 4C). The type strain may be designated in various ways (see Rule 18b, c, and d). 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.
>
> [Rule 27 [3]]. The type of the taxon must be designated. In the case of species or subspecies the culture collection numbers of at least two publically accessible service collections in different countries where a subculture of the type strain has been deposited must be indicated. In the case of sequences the catalog numbers of at least two publically accessible service collections in different countries where a sample of the DNA that has been deposited must be indicated.

In this regard, it is important to distinguish between the minimal practice which is demanded by the Bacteriological Code and good practice which is dependent on the specific research field [54,57]. The system of nomenclature envisioned in the Bacteriological Code requires that the description of a novel species be sufficiently unambiguous to assign priority. For many taxonomic groups, the sequences of a relatively small number of genes, such as those used in multilocus sequence analyses, fulfill this requirement as well as the genome sequence [12,19,57], and this topic has been reviewed in depth in another article in this issue [48]. Similarly, under some conditions genome sequences with very low coverage are able to distinguish novel species [3,46]. Lastly, for patent or other strains whose distributions are limit by law or contract, genome sequences with the commercially valuable sequences deleted would also satisfy the minimum requirements for taxonomy. In the future, other technologies may be discovered that also meet this standard and should also be suitable type material for nomenclatural purposes.

In contrast, good practices are relative, and criteria considered crucial to some fields are often irrelevant or impossible for others. For instance, the taxonomic descriptions of thermophiles often include detailed growth responses to temperature. The descriptions of anaerobes often include fermentation profiles, and descriptions of plant pathogens often include host ranges. Descriptions of novel taxa from these groups would poorly serve their communities if they lacked these data. Good practices should be enforced by reviewers, editors, and granting agencies during peer review and not the Bacteriological Code.

For the reasons given above, it is neither possible nor desirable to define universal good practices; it is possible to describe a reasonable set of criteria that could be considered by authors, reviewers and editors [54,57]. If the draft genome sequence is the type material, it should comprise >30-fold coverage. For genomes less than 5 mb, the number of contigs should be <150. For larger genomes, the number of contigs should be <400. These goals would be readily

achieved by Illumina 100 bp pair-end sequencing for most prokaryotes. At this sequencing depth, the mol% G + C of the draft sequence would be close to the true value, which is evidence that large gaps are absent [20]. Moreover, >97% of the genes would be identified, and ANIs would be close to the true values (J. Chun, personal communication). The sequence should also be accompanied with the appropriate metadata [1]. As appropriate, this would include basic information about the original sample from which the strain was isolated, such as geographic location (longitude and latitude), time and date of collection, depth or altitude, and any special conditions. It would also include basic properties of the organism, such as the temperature, pH and salinity ranges and optima, doubling time under optimal conditions, relationship to O_2, carbon sources, and terminal electron acceptors. The cell shape, motility, and ability to form endospores or other resting cells should be reported. Recommendations should also be provided for good growth media, storage conditions, and availability, i.e. deposits in culture collections.

Similarly, distribution of cultures plays integral roles in scientific communication and reproducibility and is invaluable to the progress of many fields of bacteriology [38]. Some journals such as those published by the American Society for Microbiology require sharing of biological materials as a condition of publication. Thus, deposition of type strains in culture collections or sharing with other investigators is good practice in many fields of microbiology. However, in some fields it is clearly impossible or impractical. Sometimes the growth conditions are so extreme that culture collections cannot maintain them. Sometimes legal restrictions on the distribution of the strains preclude deposition, such as patent strains. Whatever the reason, decisions the suitability of depositions of type strains of novel species are best made by specialists in the taxonomic group.

Recommendations

It is proposed that the sequence identity of homologous genes be accepted as the new criterion for delineation of species. Sequence identity could be based upon measures of the entire genome such as ANI and distance formula d_4 of the GGDC package, selections of specific genes such as specI or multilocus sequence analysis, or similar criteria deemed to be appropriate for each prokaryotic group. The similarity of gene content or DDH will then become a describable property of the species, as important to describing the properties of the group as its habitat, cell envelope structure, metabolism and other biological properties.

Given the enormous value of genome sequences to understanding the systematics and other biological properties of a species, the genome sequence of the type strain should be included as part of the description of all new taxa whenever possible.

The genome sequence may serve as the type material for description of novel taxa. However, the sequence must be derived from either a clonal population or a single cell, as is currently the case for axenic cultures. When the genome sequence is not possible, as might be the case for obligate endosymbionts, multilocus sequence analyses would be a suitable alternative if it can be shown to provide an unambiguous identification of the species. To verify the authenticity of the sequence, samples of the DNA should be submitted to public culture collections.

Acknowledgements

This work was supported in part by a NSF Dimensions of Biodiversity grant OCE-1342694. I also thank Jongsik Chun, Michael Galperin, Radhey S. Gupta, Kuo Hong, Hans-Peter Klenk, Nikos Krypides, Haiwei Luo, Amrita Pati, and Tanja Woyke for helpful discussions.

References

[1] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000) Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.
[2] Auch, A.F., Klenk, H.-P., Göker, M. (2010) Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand. Genom. Sci. 2, 142.
[3] Auch, A.F., von Jan, M., Klenk, H.-P., Göker, M. (2010) Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genom. Sci. 2, 117.
[4] Battistuzzi, F.U., Feijao, A., Hedges, S.B. (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol. Biol. 4, 44.
[5] Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., Quake, S.R. (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. PLoS ONE 6, e16626.
[6] Blank, C.E. (2011) An expansion of age constraints for microbial clades that lack a conventional fossil record using phylogenomic dating. J. Mol. Evol. 73, 188–208.
[7] Caccone, A., Desalle, R., Powell, J.R. (1988) Calibration of the change in thermal-stability of DNA duplexes and degree of base pair mismatch. J. Mol. Evol. 27, 212–216.
[8] Chain, P.S., Grafham, D.V., Fulton, R.S., Fitzgerald, M.G., Hostetler, J., Muzny, D., Ali, J., Birren, B., Bruce, D.C., Buhay, C., Cole, J.R., Ding, Y., Dugan, S., Field, D., Garrity, G.M., Gibbs, R., Graves, T., Han, C.S., Harrison, S.H., Highlander, S., Hugenholtz, P., Khouri, H.M., Kodira, C.D., Kolker, E., Kyrpides, N.C., Lang, D., Lapidus, A., Malfatti, S.A., Markowitz, V., Metha, T., Nelson, K.E., Parkhill, J., Pitluck, S., Qin, X., Read, T.D., Schmutz, J., Sozhamannan, S., Sterk, P., Strausberg, R.L., Sutton, G., Thomson, N.R., Tiedje, J.M., Weinstock, G., Wollam, A., Detter, J.C. (2009) Genomics. Genome project standards in a new era of sequencing. Science 326, 236–237.
[9] 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.
[10] Colwell, R.R. (1970) Polyphasic taxonomy of bacteria. In: Iizuka, H., Hasegawa, T. (Eds.), Culture collections of Microorganisms, University of Tokyo Press, Tokyo, pp. 421–436.
[11] De La Torre, J.R., Walker, C.B., Ingalls, A.E., Könneke, M., Stahl, D.A. (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environ. Microbiol. 10, 810–818.
[12] Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J. (2005) Opinion: re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3, 733–739.
[13] Goodfellow, M. (1995) International Committee on Systematic Bacteriology XVIth International Congress of Microbiology: Minutes of the Meetings, 2, 3, and 5 July 1994, Prague, Czech Republic. Int. J. Syst. Bacteriol. 45, 613–615.
[14] 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.
[15] Greub, G. (2010) International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of the Chlamydiae: Minutes of the inaugural closed meeting, 21 March 2009, Little Rock, AR, USA. Int. J. Syst. Evol. Microbiol. 60, 2691–2693.
[16] Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., Wagner, M. (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. Proc. Natl. Acad. Sci. U.S.A. 105, 2134–2139.
[17] Johnson, J.L., Whitman, W.B. (2007) Similarity analysis of DNAs. In: Reddy, C.A. (Ed.), Methods for General and Molecular Microbiology, third ed., ASM Press, Washington, DC, pp. 624–652.
[18] Johnson, J.L., Whitman, W.B. (2007) Similarity analysis of DNAs. In: CA., R. (Ed.), Methods for General and Molecular Microbiology, 3rd ed., ASM Press, Washington, DC, pp. 624–652.
[19] Kampfer, P., Glaeser, S.P. (2012) Prokaryotic taxonomy in the sequencing era – the polyphasic approach revisited. Environ. Microbiol. 14, 291–317.
[20] Kim, M., Park, S.C., Baek, I., Chun., J. (2014) Large-scale evaluation of experimentally determined DNA G + C contents with whole genome sequences of prokaryotes. Syst. Appl. Microbiol. 37, 95–99.
[21] Klenk, H.-P., Göker, M. (2010) En route to a genome-based classification of Archaea and Bacteria? Syst. Appl. Microbiol. 33, 175–182.
[22] Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437, 543–546.
[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.
[24] Krypides, N.C., Woyke, T., Eisen, J.A., Garrity, G., Lilburn, T.G., Beck, B., Whitman, W.B., Hugenholtz, P., Klenk, H.P. (2014) Genomic encyclopedia of type strains. Phase I: the one thousand microbial genomes (KMG-I) project. Stand. Genom. Sci. 9, 1278–1284.

[25] Kuo, C.H., Ochman, H. (2009) Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol. Dir. 4, 35.
[26] Kyrpides, N.C., Hugenholtz, P., Eisen, J.A., Woyke, T., Goker, M., Parker, C.T., Amann, R., Beck, B.J., Chain, P.S.G., Chun, J., Colwell, R.R., Danchin, A., Dawyndt, P., Dedeurwaerdere, T., DeLong, E.F., Detter, J.C., De Vos, P., Donohue, T.J., Dong, X.-Z., Ehrlich, D.S., Fraser, C., Gibbs, R., Gilbert, J., Gilna, P., Glockner, F.O., Jansson, J.K., Keasling, J.D., Knight, R., Labeda, D., Lapidus, A., Lee, J.-S., Li, W.-J., Ma, J., Markowitz, V., Moore, E.R.B., Morrison, M., Meyer, F., Nelson, K.E., Ohkuma, M., Ouzounis, C.A., Pace, N., Parkhill, J., Qin, N., Rossello-Mora, R., Sikorski, J., Smith, D., Sogin, M., Stevens, R., Stingl, U., Suzuki, K.-I., Taylor, D., Tiedje, J.M., Tindall, B., Wagner, M., Weinstock, G., Weissenbach, J., White, O., Wang, J., Zhang, L., Zhou, Y.-G., Field, D., Whitman, W.B., Garrity, G.M., Klenk, H.-P. (2014) Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol. 12, e1001920.
[27] Labeda, D. (2000) International Committee on Systematic Bacteriology: IXth International (IUMS) Congress of Bacteriology and Applied Microbiology. Int. J. Syst. Evol. Microbiol. 50, 2245–2247.
[28] 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, DC.
[29] Lasher, C., Dyszynski, G., Everett, K., Edmonds, J., Ye, W., Sheldon, W., Wang, S., Joye, S.B., Moran, M.A., Whitman, W.B. (2009) The diverse bacterial community in intertidal, anaerobic sediments at Sapelo Island, Georgia. Microb. Ecol. 58, 244–261.
[30] Lebedeva, E.V., Hatzenpichler, R., Pelletier, E., Schuster, N., Hauzmayer, S., Bulaev, A., Grigor'eva, N.V., Galushko, A., Schmid, M., Palatinszky, M., Le Paslier, D., Daims, H., Wagner, M. (2013) Enrichment and Genome Sequence of the Group I.1a Ammonia-Oxidizing Archaeon "*Ca* Nitrosotenuis uzonensis" Representing a Clade Globally Distributed in Thermal Habitats. PLoS ONE 8, e80835.
[31] Lehtovirta-Morley, L.E., Stoecker, K., Vilcinskas, A., Prosser, J.I., Nicol, G.W. (2011) Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. Proc. Natl. Acad. Sci. U.S.A. 108, 15892–15897.
[32] Li, X., Huang, Y., Whitman, W.B. (2014) The relationship of the whole genome sequence identity to DNA hybridization varies between genera of prokaryotes. Antonie van Leeuenhoek 106, xxx-xxx.
[33] Ludwig, W., Euzeby, J., Schumann, P., Busse, H.J., Trujillo, M.E., Kaempfer, P., Whitman, W.B. (2012) Road map of the Actinobacteria, in: Goodfellow, M., Kaempfer, P., Busse, H.J., Trujillo, M.E., Suzuki, K.-S., Ludwig, W., Whitman, W.B. (Eds.), Bergey's Manual of Systematic Bacteriology, vol. 5, second ed., Springer, New York, NY, pp. 1–28.
[34] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Goker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform. 14, 60.
[35] Meier-Kolthoff, J.P., Klenk, H.P., Goker, M. (2014) Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. Int. J. Syst. Evol. Microbiol. 64, 352–356.
[36] Mende, D.R., Sunagawa, S., Zeller, G., Bork, P. (2013) Accurate and universal delineation of prokaryotic species. Nat. Methods 10, 881–884.
[37] Metchnikoff, E. (1888) *Pasteuria ramosa*, un representant des bacteries a division longitudinale. Ann. Inst. Pasteur (Paris) 2, 165–170.
[38] Moore, E.R., Mihaylova, S.A., Vandamme, P., Krichevsky, M.I., Dijkshoorn, L. (2010) Microbial systematics and taxonomy: relevance for a microbial commons. Res. Microbiol. 161, 430–438.
[39] Mosier, A.C., Allen, E.E., Kim, M., Ferriera, S., Francis, C.A. (2012) Genome Sequence of "Candidatus Nitrosopumilus salaria" BD31, an Ammonia-Oxidizing Archaeon from the San Francisco Bay Estuary. J. Bacteriol. 194, 2121–2122.
[40] Murray, R.G.E. (1996) Taxonomic note: a rule about the deposition of type strains. Int. J. Syst. Bacteriol. 46, 831.
[41] Murray, R.G.E., Schleifer, K.H. (1994) Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes. Int. J. Syst. Bacteriol. 44, 174–176.
[42] 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.
[43] Park, S.-J., Kim, J.-G., Jung, M.-Y., Kim, S.-J., Cha, I.-T., Ghai, R., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., Rhee., S.-K. (2012) Draft genome sequence of an ammonia-oxidizing archaeon, "Candidatus Nitrosopumilus sediminis" AR2, from Svalbard in the Arctic Circle. J. Bacteriol. 194, 6948–6949.
[44] Park, S.-J., Kim, J.-G., Jung, M.-Y., Kim, S.-J., Cha, I.-T., Kwon, K., Lee, J.-H., Rhee, S.-K. (2012) Draft genome sequence of an ammonia-oxidizing archaeon, "Candidatus Nitrosopumilus koreensis" AR1, from Marine Sediment. J. Bacteriol. 194, 6940–6941.
[45] Ramasamy, D., Mishra, A.K., Lagier, J.C., Padhmanabhan, R., Rossi, M., Sentausa, E., Raoult, D., Fournier, P.E. (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int. J. Syst. Evol. Microbiol. 64, 384–391.
[46] 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.
[47] Rosselló-Móra, R. (2006) DNA–DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt, E. (Ed.), Molecular Identification, Systematics, and Population Structure of Prokaryotes, Springer-Verlag, Berlin, pp. 23–50.
[48] Rossello-Mora, R., Amann, R. (2015) Past and future species descriptions for *Bacteria* and *Archaea*. Syst. Appl. Microbiol. 38, 209–216.
[49] Rossello-Mora, R., Amann, R. (2001) The species concept for prokaryotes. FEMS Microbiol. Rev. 25, 39–67.
[50] Schleifer, K.H. (2009) Classification of Bacteria and Archaea: past, present and future. Syst. Appl. Microbiol. 32, 533–542.
[51] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. 1989 Approved Lists of Bacterial Names, American Society for Microbiology, Washington, DC.
[52] Sneath, P.H.A., Brenner, D.J. (1992) Official" nomenclature lists. ASM News 58, 175.
[53] Stieglmeier, M., Klingl, A., Alves, R.J.E., Rittmann, S.K.-M.R., Melcher, M., Leisch, N., Schleper, C. (2014) *Nitrososphaera viennensis* gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum Thaumarchaeota. Int. J. Syst. Evol. Microbiol. 64, 2738–2752.
[54] Sutcliffe, I.C., Trujillo, M.E., Goodfellow, M. (2012) A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. Antonie Van Leeuwenhoek 101, 13–20.
[55] Thomas, V., Casson, N., Greub, G. (2006) Criblamydia sequanensis, a new intracellular Chlamydiales isolated from Seine river water using amoebal co-culture. Environ. Microbiol. 8, 2125–2135.
[56] Tindall, B.J., Rossello-Mora, R., Busse, H.J., Ludwig, W., Kampfer, P. (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. Int. J. Syst. Evol. Microbiol. 60, 249–266.
[57] Vandamme, P., Peeters, C. (2014) Time to revisit polyphasic taxonomy. Antonie Van Leeuwenhoek 106, 57–65.
[58] Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., Swings, J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol. Rev. 60, 407–438.
[59] Wayne, L., Brenner, D., Colwell, R., Grimont, P., Kandler, O., Krichevsky, M., Moore, L., Moore, W., Murray, R., Stackebrandt, E. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.
[60] Whitman, W.B. (2011) Intent of the nomenclatural Code and recommendations about naming new species based on genomic sequences. Bull. BISMiS 2, 135–140.
[61] Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., Hooper, S.D., Pati, A., Lykidis, A., Spring, S., Anderson, I.J., D'Haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J.F., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E.M., Kyrpides, N.C., Klenk, H.P., Eisen, J.A. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462, 1056–1060.
[62] Yarza, P., Yilmaz, P., Prüße, 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 by means of SSU rRNA gene sequences. Nat. Microbiol. Rev. 12, 635–645.