Systematic and Applied Microbiology 42 (2019) 22–29

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Relevance of phenotypic information for the taxonomy of not-yet-cultured microorganisms

Jörg Overmann[a,b,c,*], Sixing Huang[a], Ulrich Nübel[a,b], Richard L. Hahnke[a], Brian J. Tindall[a]

[a] Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen, Inhoffenstraße 7B, 38124 Braunschweig, Germany
[b] Deutsches Zentrum für Infektionsforschung (DZIF), Standort Braunschweig-Hannover, Braunschweig, Germany
[c] German Center for Integrative Biodiversity Research (iDiv) Jena Halle Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

ARTICLE INFO

Keywords:
Not-yet-cultured microorganisms
Taxonomy
Nomenclature
Candidatus
International Code of Nomenclature of Prokaryotes

ABSTRACT

To date, far less than 1% of the estimated global species of *Bacteria* and *Archaea* have been described and their names validly published. Aside from these quantitative limitations, our understanding of phenotypic and functional diversity of prokaryotes is also highly biased as not a single species has been described for 85 of the 118 phyla that are currently recognized. Due to recent advances in sequencing technology and capacity, metagenomic datasets accumulate at an increasing speed and new bacterial and archaeal genome sequences become available at a faster rate than newly described species. The growing gap between the diversity of *Bacteria* and *Archaea* held in pure culture and that detected by molecular methods has led to the proposal to establish a formal nomenclature for not-yet-cultured taxa primarily based on sequence information. According to this proposal, the concept of *Candidatus* species would be extended to groups of closely related genome sequences and their names validly published following established rules of bacterial nomenclature. The corresponding sequences would be deposited in public databases as the type. The suggested alterations of the International Code of Nomenclature of Prokaryotes raise concerns regarding (1) the reliability and stability of nomenclature, (2) the technological and conceptual limitations as well as availability of reference genomes, (3) the information content of *in silico* functional predictions, and (4) the recognition of evolutionary units of microbial diversity. These challenges need to be overcome to arrive at a meaningful taxonomy of not-yet-cultured prokaryotes with so far poorly understood phenotypes.

Current challenges in understanding microbial diversity

Despite decades of cultivation and isolation attempts, only about 14,000 species of *Bacteria* and *Archaea*, probably representing 0.1–0.001% of the estimated global species number, have so far been validly published [47] (Fig. 1). Even more significantly, laboratory cultivation has returned predominantly isolates from four bacterial phyla (*Proteobacteria*, *Actinobacteria*, *Firmicutes*, and *Bacteroidetes*), many of which are easily cultivated on existing, commercially available media. On the opposite, only 10% of the described species are affiliated to the other 29 other bacterial and archaeal phyla that contain cultivated representatives, and no single isolate is currently available for 85 phyla [47]. One important, but often grossly underestimated, challenge is the repeated failure of subculturing, purifying and preserving interesting candidates from promising primary enrichments. Despite recent advances in cultivation technology, these latter key steps of the retrieval of novel microorganisms have remained particularly tedious and time consuming. In addition, a comparatively small number of laboratories is actively involved in systematic cultivation trials targeting novel types of bacteria. As a result, progress in the description of novel species has been rather slow, i.e. on the order of 600–800 novel species year^{-1}, and the severe bias towards the cultivation of only four major bacterial phyla has remained and even become more pronounced over the past years [47].

Due to advances in high-throughput sequencing technology, new 16S rRNA gene sequence types have been accumulating exponentially in the databases over the past 15 years and new species-level operational taxonomic units (OTUs) were recognized at a rate that exceeded the rate of species description almost 100

* Corresponding author at: Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen, Inhoffenstraße 7B, 38124 Braunschweig, Germany.
E-mail address: joerg.overmann@dsmz.de (J. Overmann).

https://doi.org/10.1016/j.syapm.2018.08.009
0723-2020/

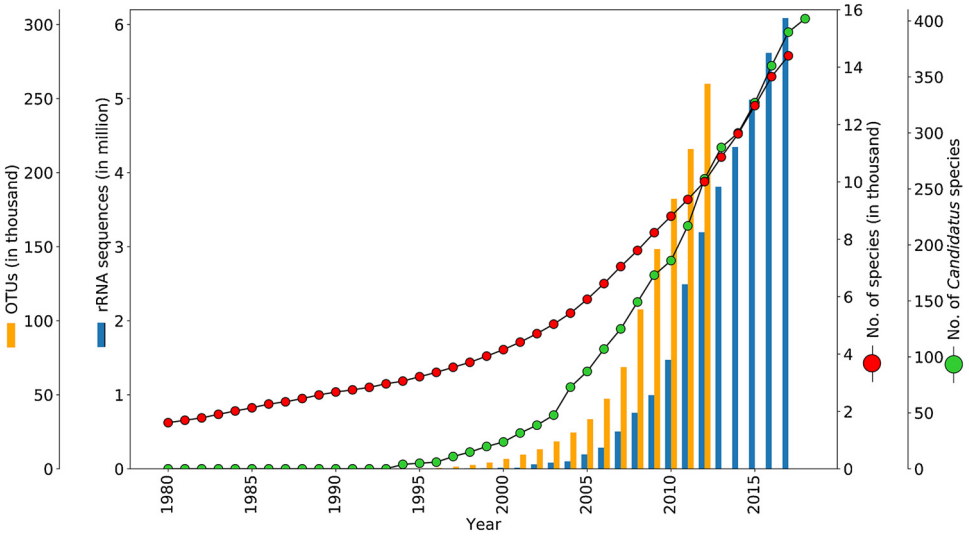

Fig. 1. Increase in numbers of rRNA sequences (blue columns) and OTUs (orange columns, defined as OTUs with 98.7% similarity of their 16S rRNA genes) in SILVA REF 114 [76], as well as of validly published species names (red dots), and *Candidatus* species (green dots) between 1980 and September 2017. Numbers of rRNA sequences were compiled from https://www.arb-silva.de/documentation/release-132/and their "Former statistics", accessed in March 2018. The numbers of species were extracted from "Prokaryotic Nomenclature Up-to-date" (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Germany, accessed in September 2017). *Candidatus* species were searched with the keyword "candidatus" in LSPN (http://www.bacterio.net/), IJSEM (http://ijs.microbiologyresearch.org/content/journal/ijsem), Sciencedirect (https://www.sciencedirect.com) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed), accessed in May 2018. From all the returned articles, publication years and the names of the *Candidatus* were extracted from the article titles, followed by manual curation to remove duplicates. Only *Candidatus* species and subspecies were included in the analysis. Numbers of OTUs in SILVA REF 114 were taken from Yarza et al. [76].

fold (Fig. 1; [76]). Meanwhile, even full bacterial genome sequences become available at a faster rate than newly described species [46]. Phylogenetically distant bacterial species are expected to feature novel biosynthetic pathways and unknown biochemical characteristics. This has been substantiated for *Tectomicrobia* [74], *Chloroflexi* [23,40], *Acidobacteria* [48], *Thaumarchaea* [35] and others. However, a better understanding of microbial diversity still depends on the availability of laboratory cultures for assessing their biochemical and physiological traits [47]. The pronounced bias of cultivation approaches towards a few taxonomic groups together with the persisting limitations of culture-independent omics approaches (detailed below) severely limit our current understanding of bacterial diversity. Eventually, both must be overcome.

The concept of *Candidatus* or "*U*species" to classify not-yet-cultivated microbes

In order for the name of a bacterial species to be validly published, the International Code of Nomenclature of Prokaryotes (ICNP) requires that a type strain, isolated in pure culture, is designated [50] and is deposited in two internationally recognized culture collections [66]. While axenic cultures are preferred there is some degree of leeway with regards, e.g., to cultures of defined symbionts.

Recognizing that a considerable number of microorganisms remain difficult or may even be impossible to isolate in pure culture (such as obligate symbionts), the category *Candidatus* was established as a means to unequivocally identify and name them. As originally proposed, the *Candidatus* status is provisional, dealing with putative taxa, and can be assigned to not-yet-cultivated microorganisms which represent unique 16S rRNA gene sequence types that have been authenticated through *in situ*-hybridization [38]. In its subsequent recommendation, the International Committee on Systematic Bacteriology (ICSB) stated that all additional information so far available such as the morphology, physiology and metabolism, reproductive features, and the natural habitat should be included in the *Candidatus* description [39]. Since 2002, the category *Candidatus* is mentioned in the International Code of Nomenclature of Prokaryotes ([9,50]). However, as putative taxa, *Candidatus* names are not validly published under the Code and hence have no status under the Code and do not have priority in nomenclature. It has to be pointed out, that the original proposal emphasized the need for subsequent cultivation attempts, based on the information gathered for a *Candidatus* [38].

So far, the *Candidatus* category has been used rather rarely for the description of novel types of bacteria. Only 402 *Candidatus* species and subspecies have been described as of May 2018 (Fig. 1). The time course of new *Candidatus* descriptions reveals that, after introduction of this category in 1994, numbers increased more or less exponentially for about 10 years but then proceeded at a rather constant rate (Fig. 1). The overall low number of descriptions has been attributed to the lack of general standards and the lack of priority of *Candidatus* names in official nomenclature [26]. However, the results of a more detailed analysis suggest that the limited use of the *Candidatus* concept likely is due to different reasons.

The category *Candidatus* was established to accomodate putative taxa of organisms such as those that had not been cultivated. This includes symbionts and parasites of eukaryotic cells or certain bacteria with conspicuous morphological features; in fact the category is still mainly used for this purpose (e.g., [36]). As an example, of the 115 *Candidatus* taxa published in recent years since 2014 (Fig. 1), 80 were described as endobionts/pathogens of protozoa, animals or plants, and at least 7 as constituents of bacterial biofilms or consortia. Another 12 *Candidatus* species belong to phylogenetically and ecologically highly divergent groups that can thus be more easily distinguished, like the *Thaumarchaeota*, anammox planctomycetes, *Verrucomicrobia*, *Spirochaetae* or *Nitrospira*. Thus, it becomes evident that *Candidatus* descriptions are considered worth the effort only if the bacteria described are significant for human health or plant protection, or have a particularly interesting biology. Secondly, the often specific association and localization of these bacteria facilitated the analyses of their DNA sequences and phenotypic characteristics, and permitted a rather advanced and detailed description.

Molecular ecological studies of bacterial communities across a multitude of environments have yielded 16S rRNA gene sequences

of so far about 260,000, often free-living, bacterial species (defined as OTUs with 98.7% similarity of their 16S rRNA genes) [76] (Fig. 1). In an attempt to promote a more systematic classification of this vast diversity of not-yet-cultivated bacteria and to render it compatible with the established bacterial taxonomy, "Candidatus Taxonomic Units" (CTUs) were introduced [76]. CTUs are defined as bacterial groups that represent unique monophyletic sequence clusters. A hierarchical system was proposed, in which within-cluster 16S rRNA gene sequence similarities on each level of classification are based on a statistical analysis of the 16S rRNA gene sequence similarities between taxa with validly published names (for instance by lower thresholds of 98.7, 94.5, 86.5, 82,0, 78.5, 75% and medians of 98.6, 95, 92, 89, 86 and 83% for species, genera, families, orders, classes and phyla, respectively; [26,76]).

A subsequent development considered the rapidly increasing amount of metagenomic data that is currently becoming available through high throughput sequencing, and it has been suggested that sequences obtained from metagenomic studies serve as the basis for the classification of candidate families, genera, and species [25,26]. This suggestion is based on several rationales. Firstly, metagenomic analysis of natural microbial communities indicate that distinct populations of closely related bacterial genome sequence types exist. Since the sequence average nucleotide similarities (ANI) between individual genomes within each population is typically >95% (similar to the ANI between cultivated bacteria of one species) but sufficiently discrete (ANI <90%) to distinguish them from other co-occurring populations, they match the current species definition. Secondly, these genomically defined populations often do not display the distinct morphology or phenotypic properties that are required for the description of a *Candidatus*. Thirdly, many of these populations are abundant, may have key functions in the ecosystem, and therefore should be classified for future reference. Fourthly, it has been argued that the information content of genome sequences retrieved from natural communities is frequently as good as, or even better than, that of the characterization of strains associated with the valid publication of a name, since phenotypic characterizations of laboratory pure cultures have been deemed unreliable and traits determined were suggested to be uninformative [25,26,65].

Based on the above considerations, it has been concluded that classification of not-yet-cultured microorganisms based on metagenomic sequences represents the only realistic approach to classify existing bacterial diversity within the years to come. The alphanumeric identifiers that are currently in use for phylogenetic clades of not-yet-cultivated bacteria (e.g., "JS1", "SAR11", "SUP05") do not inherently reflect the taxonomic rank and also do not inform about phenotypic or ecological features of environmental taxa. Therefore, a proposal has been put forward to implement a second nomenclatural system just for not-yet-cultivated taxa, in parallel and independent of the ICSB but instead governed by a novel official expert committee under the auspices of an international microbiological society or organization. It was suggested to apply this nomenclatural system under the ICNP to enable later convergence with established prokaryotic nomenclature. The nomenclatural system for not-yet-cultivated taxa should follow standards that rely on comparative genome sequence analyses [26] and would lead to a list of validly published names (denoted by a *U* superscript as a prefix of the taxon name) that is entirely separate from the existing list of cultivated bacterial species with validly published names. Originally it had been suggested to deposit gene sequence data that are sufficient to unambigously identify a species as type in databases [73]. In the most recent proposal, a mosaic consensus sequence of related genomes of an environmental "^{U}species" would then serve as type and be deposited [26].

Systematics is key to bacterial diversity research since it serves to delineate, order, and name the evolutionary units of diversity and to provide information on their distinctive (phenotypic) characteristics. Without a coherent systematics, progress in the understanding of bacterial diversity can hardly be envisioned. An entirely independent, exclusively (meta)genome-based, nomenclatural system for not-yet-cultivated prokaryotes faces major challenges: (1) the reliability and stability of nomenclature, (2) technological and conceptual limitations as well as long-term availability of reference genomes, (3) the limited information content of *in silico* functional predictions, and (4) the recognition of evolutionary units of microbial diversity. These challenges need to be addressed and inherent problems must be solved prior to formal establishment of a new systematics for prokaryotes.

Reliability and stability of nomenclature

As has already been pointed out in a comment to the proposal by Konstantinidis et al. [26], confusion would be unavoidable if a second nomenclatural system operated in parallel and essentially independent of the Bacteriological Code [44]. Even if comparable rules would be employed and information on names of not-yet-cultured prokaryotes are made readily available through public databases, no mechanism would exist to prevent the generation of synonyms (different names for the same species), homonyms (same name for different species) and an incongruence of taxonomic units (e.g. the same formally described species only partially overlap) recognized by the two systems. This is what has happened in the systematics of cyanobacteria [45]: LPSN [31] lists just 6 species names which are validly published under the Prokaryotic Code, whereas >2000 species have been described under the Botanical Code based solely on morphological traits [21]. Attempts to devise and publish recommended standards and an approved list of cyanobacterial species names have kept different bodies and committees busy for more than 40 years but failed so far ([17,20–22,32,64]). A unification of these two nomenclatural systems is still not within sight and probably unrealistic for the next years to come.

If the priority of names for not-yet-cultivated taxa shall be recognized, the only approach that avoids confusion between the nomenclature of cultivated and not-yet-cultivated prokaryotes is a careful modification and amendment of the existing Prokaryotic Code [73]. This was also originally intended for the taxonomy of not-yet-cultivated taxa [27].

Technological and conceptual limitations, and long-term availability of reference genomes

According to the proposed novel concept for the taxonomy of not-yet-cultured prokaryotes, (mosaic consensus) genome sequences provide the basis for the delineation of taxa and at the same time serve as a long-term and reliable reference, i.e. as type.

The reconstruction of genome sequences from metagenomic data is based on the assembly of sequencing reads into contigs and the subsequent binning of these contigs into groups that are thought to derive from the same microbial population (i.e. a species). Recent tests with simulated datasets revealed that the software tools available for these tasks performed impressively well when reconstructing genomes from diverse families or genera, whereas the two processing steps were much less efficient and more error prone at the species level. The assembly quality varied substantially depending on the software and parameter settings being used [30,60]. Importantly, all assemblers performed poorly when closely related genomes (ANI $\geq$95%) were present in the simulated sample [60]. Similarly, in an analysis of a defined mixture of prokaryotic genomic DNA by using three different assemblers, strain diversity caused assembly failure despite complete repre-

sentation of genomes in short-read sequencing data [2]. None of the presently available software tools was able to simply condense such related genome sequences into mosaic consensus sequences as recently suggested [26]. The performance of binning tools on simulated data also varied widely [60]. From samples composed of unique strains with <95% ANI, bins commonly contained up to 20% incorrectly assigned contigs that were in fact derived from different genomes. When closely related strains (ANI ≥95%) were present, binning accuracy deteriorated, such that more than 30% of recovered sequences were assigned incorrectly in the majority of experiments [60]. However, the co-occurrence of multiple, closely related genomes may be common in natural environmental samples [24], and hence is a realistic hurdle to metagenomic analysis. The sequencing of genomes of single cells cannot solve this problem in all cases, since whole-genome amplification often skips a considerable fraction of the genome and also may introduce single-nucleotide errors at a rate of several percent [8]. The resulting sequence information may thus be insufficient for delineating a species in many cases.

High proportions of wrongly assigned sequence stretches in draft genome sequences may compromise the quantification of genome similarity (needed for delineation of species), and at the same time affect gene and phenotype predictions (needed for the prediction of properties that would be included in a description, see below). Evidently, the abilities and limitations of metagenomic methods require further investigation, preferably by analyses of mock communities with defined composition, but such studies have as yet been performed rarely [49]. While various software tools have been developed for validation of genomic assemblies, most of these are not appropriate in metagenomic settings [42]. Validation methods based on marker genes perform best with genomes from well-studied phylogenetic groupings, for which high-quality reference sequences are available [51]. The most accurate and universal tools for metagenome assembly validation currently are based on the detection of inconsistencies between the *de-novo* assembly and a subsequent alignment (mapping) of the same sets of sequencing reads, but these methods are plagued with high false-discovery rates [42]. As yet, the choice and parameterization of assembly tools requires a trade-off between contig length and contig purity. Undoubtedly, bioinformatic tools for metagenome assembly and validation will continue to be devolped rapidly, and many of these obstacles may be overcome in the future. Arguably though, effective approaches for the quality assessment of genome assemblies from unknown organisms will be required most urgently [42].

Some of these limitations maybe overcome by complementing the standard short-read by additional long-read sequencing approaches. While short-read sequencing has become truly high throughput, low in error, and cost effective (US$ 0.012 per Megabase; [72]), available long read technologies are still more error-prone, require more sample input, higher sample quality, and are more costly (e.g., US$ 0.1 per Megabase for the PacBio Sequel technology; Dr. Ralph Vogelsang, pers. communication) which may hinder their widespread application. Given the current situation in sequencing costs and the existing limitations of bioinformatic tools, many species present in the natural environment may not be recogenized in a sufficiently reliable manner for more years to come.

A second point of concern is that a metagenomic sequence obtained from an environmental sample represents the genetic makeup of a certain cell at a particular point in time. In addition, all (meta)data obtained (e.g., for temperature, salinity, pH, carbon substrates, or a possible host) that are used for the formal description of a genomically defined taxon are linked to this particular sequence. Bacterial genomes evolve rapidly in their natural habitats and even under constant environmental conditions in the laboratory. *Escherichia coli* populations acquire sufficient non-synonymous mutations, causing different adaptations and distinguishable fitness phenotypes, within two decades [29]. Genetic changes mediated by transposons, prophages and plasmids can confer more conspicuous phenotypic differences over rather short time periods [4,10,56,75]. Relevant phenotypic characteristics therefore may appear or disappear rather quickly in natural populations, hence the original description of a genomically defined species may not match the same populations under the same selection pressures when assessed at a later point in time. Similar to cultures of type strains that are preserved in liquid nitrogen to minimize genomic changes and to allow future analysis of the originally isolated type, only sufficient, defined amounts of well preserved metagenomic DNA deposited at two independent collections in two countries will enable future comparative studies of the metagenomic sequence types that were originally described. In addition to a unique identifier, accompanying information on the other taxa present in these DNA samples should be provided.

Taken together, the technical limitations of obtaining high quality genome sequences for bacteria from complex samples, the continuous evolution of microorganisms in the environment, and the need for being able to validate and reproduce their genome sequence data, currently limit the recognition of species and also render it mandatory to physically deposit sufficient genomic material in a public repository for future reference before a genomically defined species can be described and formally named. In the future, high quality genome sequences (for instance with an error level of 10^{-6} that can presently be attained for pure cultures) may become feasible and affordable also for mixed assemblages. Until then, depositing genomic DNA as a voucher would ensure later reassessment by improved sequencing technologies. Sneath [62] had already proposed the deposition of dead material that may be used as a source of DNA, but in both cases (complex samples or DNA extracts) there would be limited access to a defined quantity of physical material.

A third concern pertains to the applicability of a rather narrow range for numerical threshold values for genome sequence similarity to delineate different bacterial taxa. The emergence of different bacterial groups is due to different evolutionary mechanisms. Rates of homologous recombination (relative to mutation rates) vary widely in different bacterial species [71]. For example, lineages within *Helicobacter pylori* exchange 1.5% of their genome per year through highly frequent recombination, resulting in a panmictic population structure with a particularly low sequence similarity of orthologous genes (92.6%; [11]); every third nucleotide is polymorphic [1]. This is clearly no exemption, since recombination is the main driver of the evolution also of the free-living marine "*Candidatus* Pelagibacter ubique" or the marine opportunistic pathogen *Vibrio parahaemolyticus*. In sharp contrast, numerous other bacteria such as the marine *Microcoleus chthonoplastes* show relative recombination rates that are two orders of magnitude lower [71]. Based on a recent detailed population genomic analysis, recombination within marine *Phaeobacter* species is very limited or even not detectable; in addition the divergence of genome sequences differs considerably between different species of the same genus (i.e. ranges between 6.2 and 0.01% in the core genome, far beyond the rather narrow range for numerical threshold proposed so far; [13]). Furthermore, specific parts of bacterial genomes are subject to lateral gene transfer, leading to lineages with similar core genome sequences that can occupy different niches and become genetically isolated. The *E. coli* laboratory strain MG1655 and the outbreak strain O157:H7 share 4.1 Mb of homologous backbone genome with only 1.8% of variable positions, but contain 0.53 and 1.34 Mb of lineage specific genome segments, respectively, mostly acquired by horizontal gene transfer [53]. Because of this varying importance of the different evolutionary mechanisms, the numerically defined

ranges of sequence similarity alone are probably not sufficient to delineate bacterial taxa, if the latter were to represent coherent, evolutionary units.

The current definition of a species – a coherent group of microorganisms that can be identified as a unit (monophyletic group) through polyphasic studies and that can be clearly discriminated from other such entities [57] – is flexible enough to allow for these different population structures of bacterial species. In contrast, a narrow threshold of core genome similarity for the delineation of taxa does not account for the heterogeneity in the evolutionary trajectories of different bacteria and hence may entice researchers to erroneously dismiss or overemphasize biological differences.

Information content of in silico *functional predictions*

For a genome-based taxonomy of not-yet-cultivated prokaryotes, it has been suggested to provide functional predictions for a minimum of 33% of the genes present in an environmental genome [26].

Unknown and hypothetical genes are significantly more numerous for genomes of not-yet-cultured or poorly covered prokaryotic phyla where they can account for >75% of all open reading frames [47]. In fact, the fraction of genes which can be annotated for a given genome is positively correlated with the fraction of cultivated representatives in the corresponding phylum [47]. Hence, functional predictions to date largely rely on the availability of annotated genomes from closely related cultured representatives [7,33]. In a genome-based taxonomy of not-yet-cultivated prokaryotes, functional predictions thus will be confined a low fraction of annotated genes, sometimes even less than the 33% suggested. At the same time, this fraction typically provides information on general housekeeping genes and genes of central metabolic pathways [3,5], whereas only few genes can actually be linked to putative taxon-specific, distinguishing functional traits [41,61,68,74]. In addition, traits deduced from metabolic reconstructions were sometimes subsequently falsified by physiological and biochemical analyses of the living organisms (e.g., [6,67]). Physiological traits inferred from genome analysis alone may not correspond to the observed phenotype, as shown for glycoside hydrolysis [18,69], and hence require biochemical verification and elucidation of the enzymatic mechanism [41]. Finally, functional traits that distinguish certain prokaryotic species from their relatives are often too complex to be easily deducible from genome sequences alone. Examples are the autotrophic growth via the oxidation of phosphite [59] or arsenite [43], or the functions of novel types of photosynthetic antenna in *Proteobacteria* [16,52]. It is particularly challenging to infer unusual characteristics like the high affinity ammonium oxidation of the Thaumarchaeon '*Candidatus* Nitrosopumilus maritimus' [35] or the unprecedented low maintenance energy requirement of an extremely low-light-adapted anoxygenic phototroph [34] from genome sequence analyses. Thus, metagenomic sequences will allow to delineate prokaryotic taxa based on sequence comparisons, but often will provide only limited information on their specific functional characteristics which is too unspecific and widespread for a meaningful taxonomic description of species.

Based on the above considerations, *in silico* predictions of functional traits are not equivalent or even inherently superior to phenotypic characterizations as has been argued by some [26,65]. Chemotaxonomic or phenotypic data are not inherently prone to errors or even meaningless as has been criticized [65], but have a limited information content only if tests are applied under conditions of low reproducibility or are even not appropriate for testing a particular target microorganism (e.g., when relying solely on commercially available kits developed for other types of microorganisms). Quite often, physicochemical data characterizing the natural environment of an isolate are not obtained or at least not considered once a bacterium has been transferred to the artificial environment of a laboratory test tube. Evidently, the selective phenotypic properties that enable a particular type of bacterium to occupy its ecological niche and which therefore distinguish it most from others must be considered more systematically in future systematics.

Thus, genomic analysis of prokaryotic traits and phenotypic characterizations potentially provide complementary data and hence are most powerful if combined in taxonomical investigations — particularly of the not-yet-cultivated prokaryotes.

The role of taxonomy for understanding the evolutionary units of microbial diversity

A meaningful and long-lasting classification system should reflect the course and outcome of evolution and should be able to identify the key properties of members of a taxon. Nomenclature assigns unique labels to the different taxa. Binominal names given to a species refer to its classification and characterization and hence to its distinctive properties.

In the currently employed polyphasic approach, *Bacteria* and *Archaea* are characterized based on the study of pure cultures. However, the enrichment and isolation of previously uncultivated types of microorganisms often takes considerable time and effort. One prominent example is "*Candidatus* Pelagibacter ubique" that could be isolated [54] more than a decade after its discovery (originally, as the SAR11 sequence cluster; [15]). Tedious, year-long cultivation and isolation attempts finally yielded representatives of phylogenetically deep-branching *Actinobacteria* [37], and the first cultured representatives of subdivision 4 and subdivision 6 *Acidobacteria* that dominate in soils worldwide [12,19]. In an even more extreme case, it took 92 years after the first description of symbiotic phototrophic consortia [28] until the first stable enrichment ("*Chlorochromatium aggregatum*") was reported and characterized [14], and exactly 100 years until the phototrophic partner of one of these consortia ("*C. aggregatum*") could be described [70]. It is self-evident that such long-term commitments are not feasible for the taxonomic description of the vast majority of so-far-uncultivated bacterial species.

It is this inherently slow pace of species descriptions that has provoked the recent proposal to base the taxonomy of not-yet-cultivated bacteria solely on genome sequence comparisons with a functional annotation of 33% of the genes per genome plus information on their habitat and abundance as a minimum and mandatory requirement [26]. Valuable information on the phenotypic traits of not-yet-cultivated bacteria can be acquired through an entire suite of elegant culture-independent methods that have become available over the recent years. High-resolution analysis (e.g. by microelectrodes) of a large number of physicochemical parameters provides initial insights into the possible niche of a microorganism. Molecular approaches like fluorescent *in situ* hybridization (FISH) allows the study of cellular morphology and the distribution of such bacteria in the environment. In certain cases, a particular morphology of the target microorganisms allows the identification of specific subcellular features (e.g. by 3D cryo-electron tomography or super resolution light microscopy) that are highly informative regarding the lifestyle of the not-yet-cultivated bacteria or archaea [3]. Metatranscriptomics and metaproteomics provide information on the expression of particular genes under specific conditions *in situ* that can be related to the physicochemical conditions. Metabolic properties can be studied on the single cell level for a selected 16S rRNA sequence type by microautoradiographic tracing of radioactively labelled compounds coupled to FISH (MAR-FISH), or by secondary ion mass spectroscopic tracing of labelled compounds coupled to FISH (Nano-SIMS). While some of

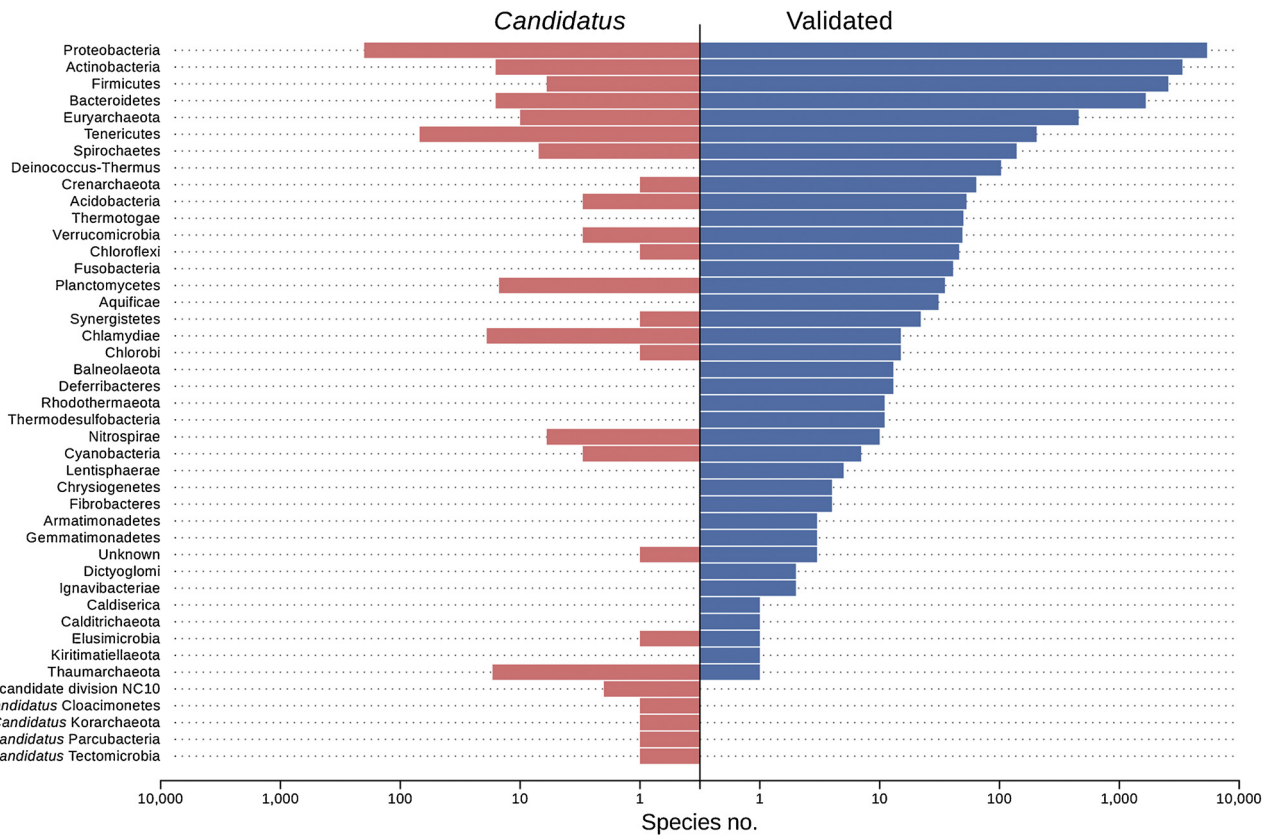

Fig. 2. Numbers of validly named species and of *Candidatus* species described for different phyla. Numbers of validly published species names were extracted from "Prokaryotic Nomenclature Up-to-date" (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Germany, accessed in September 2017). Numbers of *Candidatus* species were compiled from bacterio.net (http://www.bacterio.net/-candidatus.html) and from ijs.microbiologyresearch.org(http://ijs.microbiologyresearch.org, accessed in April 2018).

the above approaches have been recommended for the description of [U]species, they were not regarded as mandatory [26]. However, a more data-rich description would be essential to enable a clear delineation particularly of those [U]species (e.g., from uncultivated phyla; see above) for which only a low fraction of the genome can currently be annotated.

Evidently, and similar to cultivation-based approaches, the above methods can be much more time consuming than automated high throughput metagenomics and will require an iterative process, beginning with metagenomic analysis and correlative analysis of ecological data to infer potential adaptations and phenotypic traits, followed by expression analysis and culture-independent metabolic assays, returning to more specific metagenomic analysis etc. [47]. For some groups of prokaryotes, determining distiguishing phenotypic properties will actually be rather efficient since their key genes can already be reliably predicted from the genome sequence. This is the case for the *Thaumarchaeota* which meanwhile encompass more *Candidatus* species than species with validly published names (Fig. 2).

So far, such phenotypic data have been relevant for the description of a *Candidatus* taxon but required a considerable additional experimental effort. Of the mere 402 resulting *Candidatus* species and subspecies descriptions (Fig. 1), 327 fall into 7 bacterial phyla with numerous cultivated representatives (i.e., those containing ≥100 species with validly published names; Fig. 2); whereas only six single *Candidatus* descriptions (corresponding to 1.5% of all *Candidatus* descriptions; Fig. S1) became available for phyla that contained no cultivated representative (candidate phyla; Fig. 2). Thus the introduction of the *Candidatus* category did not result in a significantly accelerated recognition and analysis particularly of the phylogenetically distant novel species. Notably, the constantly low overall rate of *Candidatus* descriptions over the past 12 years (<25 per year; Fig. 1) suggests that the efforts for an in-depth systematic description of unknown bacterial taxa remained even much lower than the efforts to isolate and describe novel type of bacteria (>600 per year).

Based on this previous experience, a genome-based taxonomic system that considers experimental data on key metabolic functions only as non-mandatory information [26] at present is unlikely to lead to the generation of sufficient phenotypic data for many not-yet-cultivated species. Rather, high throughput sequencing and automated annotation most probably will result in an avalanche of thousands of taxon names assigned to metagenomic bins which are often characterized by rather generic functional predictions of limited discriminatory power and lack ecologically relevant information. So far, it has been the task of systematics to identify, classify and name evolutionary units of diversity and to provide information on their distinct characteristics. However, it is not the primary goal of systematics, and specifically taxonomy, to provide as rapidly as possible species tags just for future reference and without further substantial information.

In order to avoid the mere nomenclatural stamp collecting excercise that conventional taxonomy has been accused of [65], a taxonomy suitable for the not-yet-cultivated majority of *Bacteria* and *Archaea* must be based on more comprehensive data than a limited annotation of genes, and should include detailed information on the physicochemical conditions in these habitats (e.g., temperature, pressure, pH, salinity, water content, concentrations of major ions, organic carbon and nitrogen). Obtaining, assembling, providing and categorizing such key phenotypic information on the major

fraction of so-far-uncultured microorganisms, such an extended concept of *Candidate* taxa (or ^{U}species) would then indeed become suitable to address a key challenge of current microbial diversity research. Actually, the original proposal for the *Candidatus* category [38] also included the specific recommendation to establish an appropriate computer-linked database to deposit the sequence information together with the salient features of the *Candidatus*. The combined phenotypic characteristics of validly names *Candidate* taxa will become most useful for the scientific community if deposited in a standardized format and in a dedicated database. Recently databases like Bac*Dive* and the Digital Protologue have been established for cultivated bacteria [55,58,63] and partially also for not-yet-cultured taxa [58], and thus may serve as models for this enterprise.

A taxonomy of not-yet-cultivated that encompasses sufficient phenotypic and ecological information would certainly stimulate and facilitate the functional analysis of not-yet-cultured prokaryotic taxa, may actually boost diversity research and rejuvenate innovative cultivation attempts in the near future. A better understanding of microbial diversity could also generate novel solutions for public health, biotechnology, and agriculture.

Conclusions

- The growing gap between the limited number of described, taxa with validly published names and the vast majority of not-yet-cultivated *Bacteria* and *Archaea* calls for novel approaches for integrating the latter in prokaryotic systematics.
- A second nomenclatural system independent of the Prokaryote Code is not suited to establish consistent, stable, and non-redundant identifiers for not-yet-cultivated prokaryotic taxa and hence does not solve the problem.
- The raison d'etre of taxonomy is not to describe taxa at the fastest possible rate at the expense of cytological, biochemical and physiological scrutiny.
- Validly published names of not-yet-cultivated prokaryotes will contribute to the understanding of microbial diversity if they relate to sufficient morphological-cytological, metabolic, and ecological traits that clearly distinguish the named taxon from the others.
- Descriptions of not-yet-cultivated prokaryotes therefore must not only include the abundance and range of specific habitats, but also the morphological characteristics, the precise physicochemical conditions in these habitats, and key phenotypic properties; this information is mandatory in contrast to the current concept [26].
- *Candidatus* descriptions could become part of the established nomenclatural system, including the priority of names, if accompanied by this substantial information on distinct traits.
- Such biologically informative *Candidatus* descriptions will also promote novel, more successful cultivation attempts toward the cultivation of the corresponding target organisms.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2018.08.009.

References

[1] Achtmann, M. (2016) How old are bacterial pathogens?Proc. R. Soc. Lond. B 283, 20160990.
[2] Awad, S., Irber, L., Brown, C.T. (2017) Evaluating metagenome assembly on a simple defined community with many strain variants. bioRxiv, http://dx.doi.org/10.1101/155358.
[3] Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., Land, M.L., VerBerkmoes, N.C., Hettich, R.L., Banfield, J.F. (2010) Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl. Acad. Sci. U. S. A. 107, 8806–8811.
[4] Beres, S.B., Sylva, G.L., Barbian, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., Campbell, D.S., Smith, T.M., McCormick, J.K., Leung, D.Y., Schlievert, P.M., Musser, J.M. (2002) Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. U. S. A. 99, 10078–10083.
[5] Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., Banfield, J.F. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523, 208–211.
[6] Carini, P., Steindler, L., Beszteri, S., et al. (2013) Nutrient requirements for growth of the extreme oligotroph Candidatus *Pelagibacter ubique* HTCC1062 on a defined medium. ISME J. 7, 592–602.
[7] Choi, J., Yang, F., Stepanauskas, R., Cardenas, E., Garoutte, A., et al. (2016) Strategies to improve reference databases for soil microbiomes. ISME J. 11, 829–834.
[8] de Bourcy, C.F., De Vlaminck, I., Kanbar, J.N., Wang, J., Gawad, C., Quake, S.R. (2014) A quantitative comparison of single-cell whole genome amplification methods. PLoS One 9, e105585.
[9] De Vos, P., Trüper, H.G., Tindall, B.J. (2005) Judicial Commission of the International Committee on Systematics of Prokaryotes Xth International (IUMS) Congress of Bacteriology and Applied Microbiology. Minutes of the meetings, 28, 29 and 31 July and 1 August 2002, Paris, France. Int. J. Syst. Evol. Microbiol. 55, 525–532.
[10] Draper, J.L., Hansen, L.M., Bernick, D.L., Abedrabbo, S., Underwood, J.G., Kong, N., Huang, B.C., Weis, A.M., Weimer, B.C., van Vliet, A.H.M., Pourmand, N., Solnick, J.V., Karplus, K., Ottemann, K.M. (2017) Fallacy of the unique genomes: sequence diversity within single *Helicobacter pylori* strains. mBio 8, e02321–16.
[11] Falush, D., Kraft, C., Taylor, N.S., Correa, P., Fox, J.G., Achtman, M., Suerbaum, S. (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates recombination size, and minimal age. Proc. Natl. Acad. Sci. U. S. A. 98, 15056–15061.
[12] Foesel, B.U., Rohde, M., Overmann, J. (2013) *Blastocatella fastidiosa* gen. nov., sp. nov., isolated from semiarid savanna soil — the first described species of Acidobacteria subdivision 4. Syst. Appl. Microbiol. 36, 82–89.
[13] Freese, H.M., Sikorski, J., Bunk, B., Scheuner, C., Meier-Kolthoff, J.P., Spröer, C., Gram, L., Overmann, J. (2017) Trajectories and drivers of genome evolution in surface-associated marine Phaeobacter. Genome Biol. Evol. 9, 3297–3311.
[14] Fröstl, J.M., Overmann, J. (1998) Physiology and tactic response of the phototrophic consortium *Chlorochromatium aggregatum*. Arch. Microbiol. 169, 129–135.
[15] Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., Field, K.G. (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345, 60–63.
[16] Glaeser, J., Overmann, J. (1999) Selective enrichment and characterization of *Roseospirillum parvum*, gen. nov. and sp. nov., a new purple nonsulfur bacterium with unusual light adsorption properties. Arch. Microbiol. 171, 405–416.
[17] Hoffmann, L. (2005) Nomenclature of Cyanophyta/Cyanobacteria: roundtable on the unification of the nomenclature under the Botanical and Bacteriological Codes. Algol. Stud. 117, 13–29.
[18] Huang, S., Vieira, S., Bunk, B., Riedel, T., Spröer, C., Overmann, J. (2016) First complete genome sequence of a subdivision 6 *Acidobacterium* strain. Genome Announc. 4, e00469–16.
[19] Huber, K.J., Geppert, A.M., Wanner, G., Fosel, B.U., Wüst, P.K., Overmann, J. (2016) The first representative of the globally widespread subdivision 6 *Acidobacteria, Vicinamibacter silvestris* gen. nov., sp. nov., isolated from subtropical savannah soil. Int. J. Syst. Evol. Microbiol. 66, 2971–2979.
[20] Imhoff, J.F. (2014) International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of phototrophic bacteria. Minutes of the closed online meeting, 10–30 June 2014. Int. J. Syst. Evol. Microbiol. 64, 3910–3912.
[21] Imhoff, J.F., Madigan, M.T. (2004) International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of phototrophic bacteria. Minutes of the meetings, 27 August 2003, Tokyo, Japan. Int. J. Syst. Evol. Microbiol. 54, 1001–1003.
[22] Imhoff, J.F., Wilmotte, A. (2014) International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of phototrophic bacteria. Minutes of the meetings, 11 August 2009, Montreal, Canada. Int. J. Syst. Evol. Microbiol. 64, 3907–3909.
[23] Judger, B.E., Ertan, H., Bohl, S., Lee, M., Marquis, C.P., Manefield, M. (2016) Organohalide respiring bacteria and reductive dehalogenases: key tolls in organohalide bioremediation. Front. Microbiol. 7, 249.
[24] Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., Follows, M.J., Stepanauskas, R., Chisholm, S.W. (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 344, 416–420.
[25] Konstantinidis, K.T., Rossello-Mora, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomic data in the *Candidatus* proposal. Syst. Appl. Microbiol. 38, 223–230.
[26] Konstantinidis, K.T., Rossello-Mora, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. ISME J. 11, 2399–2406.
[27] Konstantinidis, K.T., Rossello-Mora, R., Amann, R. (2017) Reply to the commentary Uncultivated microbes — in need of their own taxonomy? ISME J. 12, 653–654.

[28] Lauterborn, R. (1906) Zur Kenntnis der sapropelischen Flora. Allg. Bot. Z. 12, 196–197.
[29] Lenski, R.E. (2017) Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J. 11, 2181–2194.
[30] Lindgreen, S., Adair, K.L., Gardner, P.P. (2016) An evaluation of the accuracy and speed of metagenome analysis tools. Sci. Rep. 6, 19233.
[31] LPSN-List of Prokaryotic names with standing in nomenclature (2018) http://www.bacterio.net/-candidatus.html. (Accessed March 2018).
[32] Madigan, M.T., Imhoff, J.F. (2007) International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of phototrophic bacteria. Minutes of the meetings, 29 August 2006, Pau, France. Int. J. Syst. Evol. Microbiol. 57, 1169–1171.
[33] Marcy, Y., Ouverney, C., Bik, E.M., Lösekann, T., Ivanova, N., et al. (2007) Dissecting biological dark matter with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc. Natl. Acad. Sci. U. S. A. 104, 11889–11894.
[34] Marschall, E., Jogler, M., Henssge, U., Overmann, J. (2010) Large-scale distribution and activity patterns of an extremely low-light-adapted population of green sulfur bacteria in the Black Sea. Environ. Microbiol. 12, 1348–1362.
[35] Martens-Habbena, W., Berube, P.M., Urakawa, H., de al Torre, J.R., Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. Nature 461, 976–979.
[36] Mehari, Y.T., Hayes, B.J., Redding, K.S., Mariappan, P.V., Gunderson, J.H., Farone, A.L., Farone, M.B. (2016) Description of '*Candidatus* Berkiella aquae' and '*Candidatus* Berkiella cookevillensis', two intranuclear bacteria of freshwater amoebae. Int. J. Syst. Evol. Microbiol. 66, 536–541.
[37] Monciardini, P., Cavaletti, L., Schumann, P., Rohde, M., Donadio, S. (2003) *Conexibacter woesei* gen. nov., sp. nov., a novel representative of a deep evolutionary line of descent within the class *Actinobacteria*. Int. J. Syst. Evol. Microbiol. 53, 569–576.
[38] Murray, R.G.E., Schleifer, K.-H. (1994) Taxonomic notes: a proposal for recording the properties of putative prokaryotes. Int. J. Syst. Bacteriol. 44, 174–176.
[39] Murray, R.G.E., Stackebrandt, E. (1995) Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described prokaryotes. Int. J. Syst. Bacteriol. 45, 186–187.
[40] Nett, M., Erol, Ö., Kehraus, S., Köck, M., Krick, A., et al. (2006) Siphonazole, an unusual metabolite from *Herpetosiphon* sp. Angew. Chem. Int. Ed. 45, 3863–3867.
[41] Nobu, M.K., Dodsworth, J.A., Murugapiran, S.K., Rinke, C., Gies, E.A., et al. (2016) Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. ISME J. 10, 273–286.
[42] Olson, N.D., Treangen, T.J., Hill, C.M., Cepeda-Espinoza, V., Ghurye, J., Koren, S., Pop, M. (2017) Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. Brief. Bioinform., bbx098, http://dx.doi.org/10.1093/bib/bbx098.
[43] Oremland, R.S., Hoeft, S.E., Santini, J.M., Bano, N., Hollibaugh, R.A., Hollibaugh, J.T. (2002) Anaerobic oxidation of arsenite in Mono Lake water and by a facultative, arsenite-oxidizing chemoautotroph, strain MLHE-1. Appl. Environ. Microbiol. 68, 4795–4802.
[44] Oren, A., Garrity, G.M. (2017) Commentary. Uncultivated microbes — in need of their own nomenclature? ISME J. 12, 309–311.
[45] Oren, A., Ventura, S. (2017) The current status of cynobacterial nomenclature under the 'prokaryotic' and the 'botanical´ code'. Ant. Van Leeuwenhoek 110, 1257–1269.
[46] Overmann, J. (2013) Principles of enrichment, isolation, cultivation, and preservation of bacteria. In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (Eds.), The Prokaryotes-Prokaryotic Biology and Symbiotic Associations, 4th edition, Springer, New York, pp. 149–207.
[47] Overmann, J., Abt, B., Sikorski, J. (2017) The significance and future of cultivation. Annu. Rev. Microbiol. 71, 711–730.
[48] Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S.C., Treusch, A.H., et al. (2003) Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. Mol. Microbiol. 50, 563–575.
[49] Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N. (2017) Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. 35, 833–844.
[50] Parker, C.T., Tindall, B.J., Garrity, J.M. (2015) International code of nomenclature of prokaryotes. Int. J. Syst. Evol. Microbiol., http://dx.doi.org/10.1099/ijsem.0.000778.
[51] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055.
[52] Permentier, H.P., Neerken, S., Overmann, J., Amesz, J. (2001) A bacteriochlorophyll *a* antenna complex from purple bacteria absorbing at 963 nm. Biochemistry 40, 5573–5578.
[53] Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., Blattner, F.R. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409, 529–533.
[54] Rappé, M., Connon, S., Vergin, K., Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418, 630–633.
[55] Reimer, L.C., Söhngen, C., Vetcininova, A., Overmann, J. (2017) Mobilization and integration of bacterial phenotypic dataöEnabling next generation biodiversity analysis through the BacDive metadatabase. J. Biotechnol. 261, 187–193.
[56] Riedel, T., Bunk, B., Thürmer, A., Spröer, C., Brzuszkiewicz, E., Abt, B., Gronow, S., Liesegang, H., Daniel, R., Overmann, J. (2015) Genome resequencing of the virulent and multidrug-resistant reference strain *Clostridium difficile* 630. Genome Anounc. 3, e00276–15.
[57] Rosselló-Móra, R., Amann, R. (2015) Past and future species definitions for Bacteria and Archaea. Syst. Appl. Microbiol. 38, 209–216.
[58] Rosselló-Móra Trujillo, M.E., Sutcliffe, I.C. (2017) Introducing a digital protologue: a timely move towards a database-driven systematics of archaea and bacteria. Ant. Van Leeuwenhoek 110, 455–456.
[59] Schink, B., Friedrich, M. (2000) Phosphite oxidation by sulphate reduction. Nature 406, 37.
[60] Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Droge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jorgensen, T.S., Shapiro, N., Blood, P.D., Gurevich, A., Bai, Y., Turaev, D., DeMaere, M.Z., Chikhi, R., Nagarajan, N., Quince, C., Meyer, F., Balvociute, M., Hansen, L.H., Sorensen, S.J., Chia, B.K.H., Denis, B., Froula, J.L., Wang, Z., Egan, R., Don Kang, D., Cook, J.J., Deltel, C., Beckstette, M., Lemaitre, C., Peterlongo, P., Rizk, G., Lavenier, D., Wu, Y.W., Singer, S.W., Jain, C., Strous, M., Klingenberg, H., Meinicke, P., Barton, M.D., Lingner, T., Lin, H.H., Liao, Y.C., Silva, G.G.Z., Cuevas, D.A., Edwards, R.A., Saha, S., Piro, V.C., Renard, B.Y., Pop, M., Klenk, H.P., Göker, M., Kyrpides, N.C., Woyke, T., Vorholt, J.A., Schulze-Lefert, P., Rubin, E.M., Darling, A.E., Rattei, T., McHardy, A.C. (2017) Critical assessment of metagenome interpretation — a benchmark of metagenomics software. Nat. Methods 14, 1063–1071.
[61] Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T., Hentschel, U. (2011) Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. ISME J. 5, 61–70.
[62] Sneath, P.H.A. (1995) Thirty years of numerical taxonomy. Syst. Biol. 44, 281–298.
[63] Söhngen, C., Podstawka, A., Bunk, B., Gleim, D., Vetcininova, A., Reimer, L., Ebeling, C., Pendarovski, C., Overmann, J. (2016) BacDive — the Bacterial Diversity Metadatabase in 2016. Nucl. Acids Res. 44 (1), D581–D585.
[64] Stanier, R.Y., Sistrom, W.R., Hansen, A., Whitton, B.A., Castenholz, R.W., Pfennig, N., Gorlenko, V.N., Kondratieva, E.N., Eimhjellen, K.E., Whittenbury, R., Gherna, R.L., Trüper, H.G. (1978) Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. Int. J. Syst. Bacteriol. 28, 335–336.
[65] Sutcliffe, I.C. (2015) Challenging the anthropocentric emphasis on phenotypic testing in prokaryotic species descriptions: rip it up and start again. Front. Genet. 6, 218.
[66] Tindall, B.J., Kämpfer, P., Euzéby, J.P., Oren, A. (2006) Valid publication of names of prokaryotes according to the rules of nomenclature: past history and current practice. Int. J. Syst. Evol. Microbiol. 56, 2715–2720.
[67] Tripp, H.J., Schwalbach, M.S., Meyer, M.M., et al. (2009) Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. Environ Microbiol. 11, 230–238.
[68] Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.
[69] Vieira, S., Luckner, M., Wanner, G., Overmann, J. (2017) *Luteitalea pratensis* gen. nov., sp. nov. a new member of subdivision 6 *Acidobacteria* isolated from temperate grassland soil. Int. J. Syst. Evol. Microbiol. 67, 1408–1414.
[70] Vogl, K., Glaeser, J., Pfannes, K.R., Wanner, G., Overmann, J. (2006) *Chlorobium chlorochromatii* sp. nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium *Chlorochromatium aggregatum*. Arch. Microbiol. 185, 363–372.
[71] Vos, M., Didelot, X. (2009) A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3, 199–208.
[72] Wetterstrand, K.A. 2018 DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) , Available at: www.genome.gov/sequencingcostsdata. (Accessed 3 April 2018).
[73] Whitman, W.B. (2016) Modest proposals to expand the type material for naming of prokaryotes. Int. J. Syst. Evol. Microbiol. 66, 2108–2112.
[74] Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., et al. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506, 58–62.
[75] Yang, L., Jelsbak, L., Marvig, R.L., Damkiær, S., Workman, C.T., Rau, M.H., Hansen, S.K., Folkesson, A., Johansen, H.K., Ciofu, O., Høiby, N., Sommer, M.O.A., Molin, S. (2011) Evolutionary dynamics of bacteria in a human host environment. Proc. Natl. Acad. Sci. U. S. A. 108, 7481–7486.
[76] 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 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635–645.