Systematic and Applied Microbiology 38 (2015) 209–216

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Past and future species definitions for *Bacteria* and *Archaea*

Ramon Rosselló-Móra[a,*], Rudolf Amann[b]

[a] *Marine Microbiology Group, Department of Ecology and Marine Resources, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), E-07190 Esporles, Illes Balears, Spain*
[b] *Max Planck Institute for Marine Microbiology, D-28359 Bremen, Germany*

ARTICLE INFO

Keywords:
Species concept
Species definition
Genomics

ABSTRACT

Species is the basic unit of biological diversity. However, among the different microbiological disciplines there is an important degree of disagreement as to what this unit may be. In this minireview, we argue that the main point of disagreement is the definition (i.e. the way species are circumscribed by means of observable characters) rather than the concept (i.e. the idea of what a species may be as a unit of biodiversity, the meaning of the patterns of recurrence observed in nature, and the why of their existence). Taxonomists have defined species by means of genetic and expressed characters that ensure the members of the unit are monophyletic, and exhibit a large degree of genomic and phenotypic coherence. The new technologies allowing high-throughput data acquisition are expected to improve future classifications significantly and will lead to database-based taxonomy centered on portable and interactive data. Future species descriptions of *Bacteria* and *Archaea* should include a high quality genome sequence of at least the type strain as an obligatory requirement, just as today an almost full-length 16S rRNA gene sequence must be provided. Serious efforts are needed in order to re-evaluate the major guidelines for standard descriptions.

Introduction

Species is the basic unit of biological diversity. However, microbiologists' perceptions of what a species is differ significantly, not least for semantic reasons [44]. Taxonomists, ecologists and evolutionary biologists interpret species differently and necessarily within the framework of their needs and the tools they use for identification. In this context, it is important to recall that the concept of species is different from the species definition, which is the way species are described [48]. On the other hand, the concept of species is the idea and the theoretical framework that explains what the unit can be. This idea should be as universal as possible in order to embrace all living beings. The concept explains what experts consider a species to be as a unit of biodiversity, the meaning of the patterns of recurrence observed in nature, and the why of their existence [21].

Given the relevance of the unit for different disciplines, several concepts have been proposed. These ranges from understanding species as ecotype lineages that are bound to periodic selection by means of evolutionary and environmental constraints [9], to more abstract concepts, such as the method-free species that provides scientists with the freedom to decide on the methods, thresholds and criteria for circumscribing their units [1]. The latter considers species as "metapopulation lineages" evolving separately from other such entities, that "occupy an adaptive zone minimally different from any other lineage in its range", and that "do not have to be phenotypically distinguishable, or diagnosable, or monophyletic, or reproductively isolated or ecologically divergent to be species" [1]. However, if there are no measurable phenotypic, genealogical or ecological parameters, species identification will be difficult, and it will not be possible to classify species in a universal system. The capabilities of circumscribing species depend on the measurable parameters [48]. For example, in ecology, where discrimination has hitherto mostly been carried out by means of 16S rRNA gene variation, a relaxed and more flexible circumscription might be sufficient [40], whereas among evolutionary microbiologists able to discriminate closely related lineages with distinct evolutionary fate, ecotypes would be more suitable units of diversity [28]. One way or another, what is used in taxonomy stays within these two extremes.

Contrary to other opinions [11–13], we believe that a satisfactory universal and pragmatic formulation of the species concept for *Bacteria* and *Archaea* can be achieved based on the current knowledge of taxonomic and technological developments. Actually, this

* Corresponding author. Tel.: +34 971 611 826; fax: +34 971 611 761.
E-mail address: rossello-mora@uib.es (R. Rosselló-Móra).

http://dx.doi.org/10.1016/j.syapm.2015.02.001
0723-2020/

belief is reinforced by the fact that recent (meta-)genomic studies provide strong evidence that discrete populations thrive in natural environments [5,29]. These populations are reminiscent of species circumscribed for cultured prokaryotes when the common taxonomic (i.e. genetic) thresholds are applied.

We previously proposed a phylo-phenetic species concept in which a species was "a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity with respect to many independent characteristics and is diagnosable by a discriminative phenotypic property" [49]. This was subsequently reformulated as "a category that circumscribes a (preferably) genomically coherent group of individual isolates/strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardized conditions" [53]. Here, the wording is specifically tailored for strains isolated in the laboratory, and reflects Rule 18a of the International Code of Nomenclature of Bacteria that only the names of cultured prokaryotes can be validly published [32]. As a consequence, this concept cannot be universally applied to all Bacteria and Archaea, since only a minor fraction of all microbial species thriving in the biosphere have so far been cultured and classified [68].

Based on recent technological advances in high-throughput methods, it is now possible to circumscribe biological units based on their genealogic, genomic and phenotypic coherence at relatively low cost, and most of these methods can even be applied to uncultured organisms. Therefore, we now suggest using a concept in which a species is "*a category that circumscribes monophyletic, and genomically and phenotypically coherent populations of individuals that can be clearly discriminated from other such entities by means of standardized parameters*". Such parameters should be based on genetic and phenotypic data, retrievable by methods allowing the generation of interactive databases, which permit computerized comparisons using generally used bioinformatics tools. We believe that this concept embraces evolutionary units of different size (i.e. the extent of genetic and phenotypic diversity is allowed to vary within itself) depending on the thresholds used, which are covered by the definition (i.e. the way we recognize the units by means of observable characters). We are aware that distinct disciplines in microbiology are provided with different tools for recognizing microbial populations and, depending on their resolution, the definitions may not be equivalent [48]. Monophyly is an absolute premise that guarantees that the organisms embraced by the concept have a common evolutionary history. The genomic coherence modulates the circumscription of the unit by means of understanding to which extent the amount and identity of shared genes guarantees the observation of an unequivocal phenotype that serves as a diagnostic tool for identification purposes. Finally, phenotypic coherence should demonstrate that the organisms belonging to the same taxon share physiological, structural and even ecological properties that justify considering a single taxonomic unit. The phenotype can either be predicted from the inference of the genome sequences, or from the determination of the largest set of characters possible (metabolic, chemical, morphological, and even ecological), which would tend to minimize the relevance of singular changes caused by horizontal gene transfer. We believe that the "species problem" in microbiology is caused rather by differences in the species definition than by a failure in the conceptual basis.

The definition is the way we circumscribe the unit, which is the compilation of different parameters (e.g. genomic, phenotypic) that allow its unequivocal identification. However, these parameters vary depending on the organisms under study. What is valid for animals may not be applicable to prokaryotes, as they exhibit completely different characters, evolutionary fate and ecological constraints. The simple morphological characters of microorganisms together with their expected vast diversity have promoted the development of methodologies that go far beyond the description of the phenotype. The taxonomy of prokaryotes has benefited enormously from the technological advances in molecular biology and analytical chemistry [49]. Therefore, in the following, we briefly recall the past before reviewing recent advances towards a more solid definition of bacterial and archaeal species, and finally suggest additional data that should be obligatory for future descriptions.

The past

The first descriptions of bacterial species were made based on phenotypic traits that led to fuzzy classifications [49]. Then an important breakthrough occurred in the 1960s when methodological developments permitted genome comparisons, which were initially by means of mol% G + C content and DNA-DNA hybridization (DDH). The integration of the genomic data into the species descriptions had an important influence on the definition and shaped the current view of taxonomic classifications [48]. Potentially, the most important finding of this period was when scientists observed that phenotypically coherent microorganisms could be regarded as a single species if they shared high DDH values, in general above 70%. This cutoff value was later recommended by an ad hoc committee as an approximate threshold for circumscribing species genomically, which especially reinforced that only the complete DNA sequence should be the reference standard for determining phylogeny and, therefore, taxonomy [60]. As a result, in many cases, the approximate value of 70% DDH was taken strictly as a "gold standard" for circumscribing species, which in some cases forced the division of sets of isolates that could objectively represent single species [49]. The second breakthrough in the 1970s was genealogical reconstruction based on the ribosomal RNA genes, in particular 16S rRNA [63], which became the standard for reconstructing genealogies and the backbone for a new prokaryotic taxonomy [33]. Since then, a 16S rRNA-directed classification of prokaryotes has been superimposed on the hierarchical framework provided by the rules of nomenclature [16]. This second breakthrough was responsible for the arithmetical increase of new descriptions [55], but it also resulted in a flood of species descriptions based on single strains with a 16S rRNA gene having a similarity of <98.6% compared to the validly named species [15]. However, Single Strain Species Descriptions (SSSDs) neglect the strain diversity within a species, resulting in incomplete descriptions, even though the single strain may have been studied with a detailed polyphasic approach [18].

The present

Currently, there is general agreement that circumscriptions of species for taxonomic purposes (i.e. the definition) must be founded on a wide set of parameters that guarantee the understanding of their uniqueness. This is referred to as the "polyphasic approach" which aims at obtaining a consensus classification by integrating different kinds of data into a classification of minimal contradictions [18]. For an accurate classification of a species, three major premises should be fulfilled: (i) monophyly, (ii) genomic coherence, and (iii) phenotypic coherence.

Monophyly

In general, the demonstration that the members of a new species all belong to one monophyletic lineage has been performed by means of phylogenetic inferences based on housekeeping genes, in particular by comparative sequence analysis of the 16S rRNA

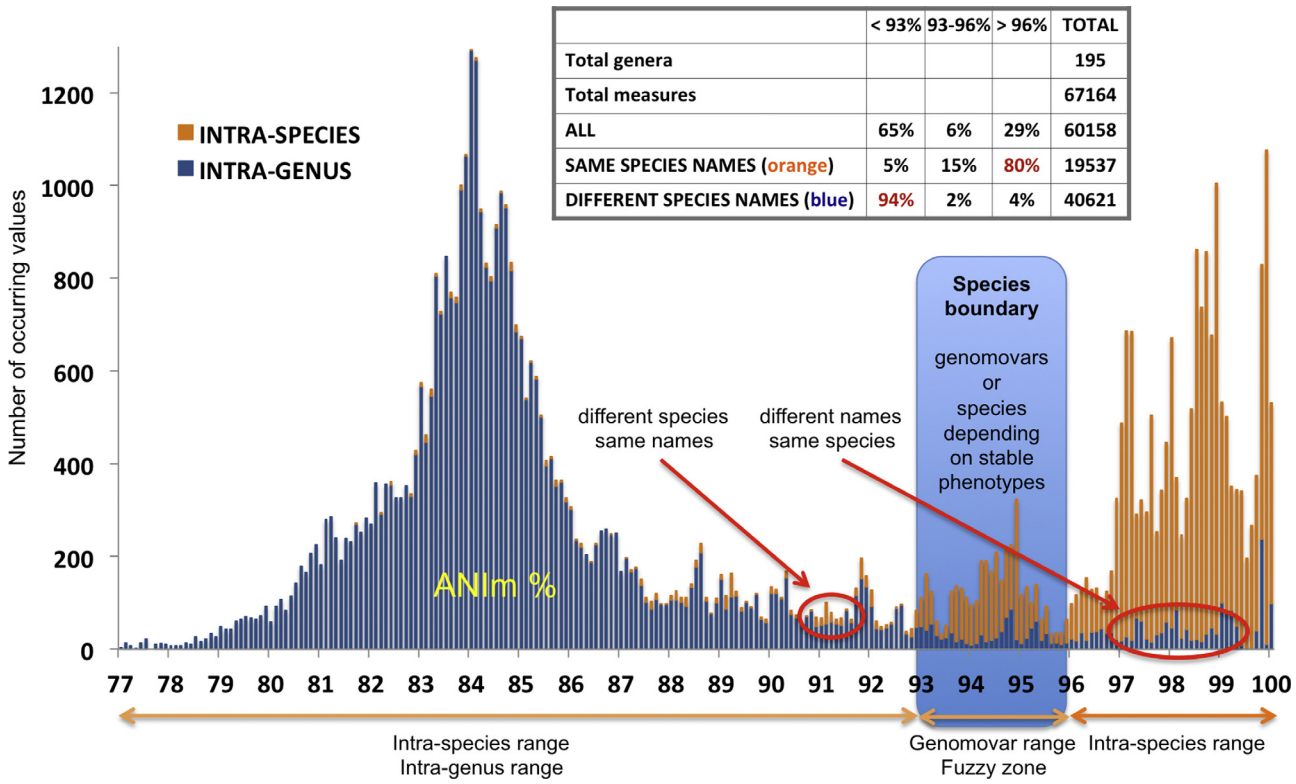
gene [33]. Different tools and databases have been developed to reconstruct the phylogenies of microorganisms [68], and recommendations have been given for achieving the best reliable topology [57]. The resolution power of the 16S rRNA gene for discriminating species has been extensively discussed. As a result, based on empirical knowledge, a conservative sequence identity threshold of 98.7% has been recommended below which no DDH experiments seem to be needed [52]. Above this value, there is a need to prove by genome to genome comparisons whether distinct isolates belong to the same species. The 16S rRNA gene has become the gold standard for phylogenetic reconstructions in all disciplines of microbiology and, because of its interest (especially in microbial ecology), the sequencing efforts of this molecule have been so intense that currently the number of entries for this gene in public repositories surpasses 4 million [43]. For pragmatic reasons, a curated database named the Living Tree Project (LTP; [65,66]) has been created in order to address the need for recognizing the best sequences of type strains of classified taxa. As part of this project, a significant number of validly named species was recognized where the 16S rRNA had not been sequenced. Therefore, as a collaborative project, several international culture collections completed a catalogue of high quality 16S rRNA gene sequences for all classified species of *Bacteria* and *Archaea* [67]. In parallel with the LTP, the EzTaxon project [6] was also initiated, which offered a highly curated and interactive 16S rRNA gene database of type strains for identification purposes. Both databases have, in only a few years, become much-used tools for identification and classification of prokaryotic taxa. The current *recommendation* (e.g. [57,68]) for proper reconstruction of a 16S rRNA-based genealogy for taxonomic purposes is the use of high quality almost complete 16S rRNA gene sequences aligned taking into account their secondary structure. All relevant type strain sequences should be included and their phylogeny should be reconstructed by means of different algorithms (usually maximum likelihood, maximum parsimony and neighbor-joining), with the application of different conservational filters (depending on the level of resolution needed hypervariable positions may be removed) and different datasets (i.e. different sets of sequences that may also include those of non-type strains). The result of such a phylogenetic reconstruction may be a consensus tree based on different observations, or one selected tree in which bootstrap values indicate branch instabilities.

As an alternative to the 16S rRNA approach, multilocus sequence analysis (MLSA; [25]) based on phylogenetic reconstructions of several housekeeping genes can be used. For taxonomic purposes, it has been *recommended* that a minimum of five genes should be used in order to overcome the problems of phylogenetic noise, but the number depends on the robustness obtained by the selection of each different group [53]. For distantly related organisms, the tree topologies have been shown to stabilize only when at least twelve genes were used in conjunction with the use of filters to remove phylogenetic noise [51]. The MLSA approach may show significantly better resolution than 16S rRNA-based phylogenies if the selection of genes is large enough. In such cases, MLSA groupings match those obtained by genome-to-genome comparisons, such as in the *Bacillus cereus* group [23]. However, not least because of the need to custom-design multiple primer sets for each group of interest, the MLSA approach has proved to be cumbersome and suffers from arbitrary gene selection [19,26].

Genomic coherence

As stated above, the use of DDH has significantly shaped the current species definition, in particular with respect to genomic coherence. As the ad hoc committee stated many years before full genome sequencing was routinely achievable: "*the complete deoxyribonucleic acid (DNA) sequence would be the reference standard to determine phylogeny and that phylogeny should determine taxonomy*" [60]. The Committee seems to have foreseen recent technological advances in next generation sequencing (NGS), which might make DDH obsolete [20,45]. DDH has been considered cumbersome and error-prone. It was technically so demanding that it was even described as a "non-portable" method, only used in a few laboratories around the world producing data that were difficult to compare with each other. Therefore, with respect to genomic coherence, MLSA was recalled in order to substitute DDH [53]. MLSA has been continuously evaluated in many different genera (for a review see [19,26]), but no common threshold for substituting DDH has been established. The literature shows that different thresholds for different gene sets may fit different genera such as, for example, 2.1% divergence for *Achromobacter* spp. based on concatenates of the seven genes *nusA*, *eno*, *rpoB*, *gltB*, *lepA*, *nuolL* and *nrdA* [58], 3% for *Burkholderia* spp. based on concatenates of the seven genes *atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC* and *trpB* [39], or an MLSA distance of 0.007 for *Streptococcus* spp. based on concatenates of the five genes *atpD*, *gyrB*, *recA*, *rpoB* and *trpB* [47]. For MLSA, neither a universal threshold nor a common number and set of genes have yet been recommended.

The rapid progress of NGS is currently facilitating genome-wide sequence analyses as an invaluable tool for substituting DDH. In addition, compared to MLSA, comparative genome analysis is more economical and less subjective than having to select genes. The high speed and low cost of draft genome sequencing opens the door to in silico comparisons that are reminiscent of DDH. In this regard, different parameters (also called Overall Genome Relatedness Indices or OGRI [7]) to compare genomes for species circumscription purposes have been developed, such as the average nucleotide identity (ANI; [20,30,45]), digital DDH (dDDH; [34], and maximal unique matches (MUM; [10]). Among the different parameters devised, ANI is perhaps the one most acknowledged by the scientific community, since it has been used in more than 30 classifications published following its recommendation [3]. ANI can be calculated by different algorithms that have been implemented in the JSpecies software package (www.imedea.uib.es/jspecies; [45]). One of the algorithms is based on BLAST searches of 1 kb genome fragments against a target genome (ANIb; [20]. A second algorithm, with much higher speed, is based on the MUMmer algorithm (ANIm) that does not require the artificial generation of 1 kb fragments. Both algorithms give nearly identical values in the high identity range (80% to 100%). However, for distant genome comparisons, both algorithms diverge in their results [45] and perhaps the former (ANIb) has better application. Nevertheless, the use of ANI for distantly related genomes is questionable because complementarity is so low that only very small parts of the genomes are compared. Due to this reason, the average amino acid identity (AAI) has been recommended as a measure for distantly related genomes (below 80%), since it provides a much more robust resolution [46]. DDH has shown that what is considered to be a genospecies (i.e. a species from the genomic point of view; [49] embraces organisms sharing DDH values higher than ± 70%, and that this DDH value is equivalent to ANI values higher than ± 94% [20,30,45]. Fig. 1 shows the distribution of values among all genomes for the 195 genera with at least two representatives present in the NCBI database in September 2014. The 60,158 values appear in a trimodal distribution showing that values of <93% ANIm correspond, in 94% of cases, to intra-genus but inter-species measures, and values of >93% ANIm correspond, in 95% of cases, to intra-species values. It should be noted that, as previously reported [45], the genome databases contain entries where there are genomes with wrong specific assignments, others with identification only at the genus level (i.e. labeled as "sp."), and others with non-existent specific names. All such cases account for the possible background noise observed in Fig. 1. In addition, there is an intermediate zone between 93% and 96% ANIm where



	< 93%	93-96%	> 96%	TOTAL
Total genera				195
Total measures				67164
ALL	65%	6%	29%	60158
SAME SPECIES NAMES (orange)	5%	15%	80%	19537
DIFFERENT SPECIES NAMES (blue)	94%	2%	4%	40621

Fig. 1. ANIm value distribution calculated for all genomes present in the NCBI database (ftp.ncbi.nih.gov/genomes/Bacteria) in September 2014 identified with the same generic name. For the calculation, only genera names with at least two genomes have been considered (see Supplementary Tables S1–S4). In the figure, 195 genera representing a total of 1883 genomes have been examined. Pairwise calculations between genomes identified with the same generic, but different, specific names are shown in blue. Pairwise calculations between genomes with the same specific names are shown in orange. The complete dataset comprised 67,167 reciprocal calculations, 7006 of which did not show any match due to the genetic divergence between the genomes. The ANIm of <93% may be considered as the intra-genus, but inter-species range. The ANIm of >96% may be considered the intra-species range, as recommended by Goris et al. [20]. The ANIm ranging between 93 and 96% can be considered as the fuzzy zone where the boundary of a species may fall [45]. The 5% of ANIm intra-species values <93% may be considered as misidentified organisms with the same specific name, as previously indicated [45]. The 4% of ANIm inter-species values >96% are due to either unidentified genomes at the species level (i.e. *Genus* sp.) or, probably, misidentified organisms at the species level. The 15% of ANIm intra-species values ranging between 93 and 96% can be considered as different genomovars of the same species [49], whereas the 2% ANIm of inter-species values in the same range may be considered as closely related species.

approximately 6% of the total measures fall. In this zone, most of the cases (76%) account for intra-species values, and the remainder (24%) reflects inter-species measures within the same genus. This range can be regarded as the zone where the classification of the organisms into different or the same species depends on the decision of the taxonomist. It can be considered as the range in which different genomovars of a species may be allocated [49]. Genomovar stands for distinct genomic groups that are sufficiently different to be classified as different species, but with phenotypes that do not show sufficient robust differences for discriminating them. Splitting different genomovars into different species often depends on the decision and skills of the taxonomist for retrieving the relevant differences. A particular advantage of the method is that there is no requirement for a complete genome sequence in order to obtain a stable ANI value. It has been demonstrated that two draft genomes with approximately 20% coverage each render stable values, and even a very basic partial genome of approximately 4% coverage yields a useful ANIm value if compared against a complete target genome [45].

In addition to ANI, a complementary parameter that has been shown to identify genomes at the species level is the tetranucleotide signature regression (TETRA; [45]. This is an alignment-free parameter exclusive to each genome that has a very fast calculation. Experience has shown that only highly similar genomes with regression values above 0.999 will correspond to ANI values of >94%. The high calculation speed of this parameter permits the screening of very large sets of genomes prior to *in silico* hybridization (Fig. 2).

For a stable taxonomic framework, there is an urgent need to complete the genome sequencing of all ~12,000 hitherto classified species of *Bacteria* and *Archaea*, since currently only 3500 type strains have an almost full or a draft genome available [14,62]. However, important efforts are already being undertaken in this respect, such as the international consortium headed by the Joint Genome Institute (USA) "Genomic Encyclopedia of *Bacteria* and *Archaea*" (GEBA; www.jgi.doe.gov/programs/GEBA) that focuses efforts on a good range of selected type strains in order to cover the diversity gaps in the phylogenetic tree as much as possible [64]. Moreover, some journals (such as Systematic and Applied Microbiology publishing this special issue) already strongly recommend that any description of a novel species needs to be accompanied by at least the almost complete genome sequence of the type strain. A complete genome catalogue of classified species will not only be the basis for a better understanding of the natural relationships of the classified organisms, but it will also enable reliable and fast identification of new isolates. In addition, it would facilitate the analysis of metagenome data [29].

As a *recommendation for genome coherence analyses*, we suggest that DDH should be abandoned. In addition, MLSA, which is currently evaluated with enlarged sets of genes, will likely prove to be uncompetitive. Instead, the genome of the type strain should be sequenced in high quality, and genome sequences should also be determined for a subset or all additional strains used for the description of the novel species. The genomes should be subsequently compared by one or several parameters, taking into account that ANI > 96% (and TETRA > 0.999) would guarantee

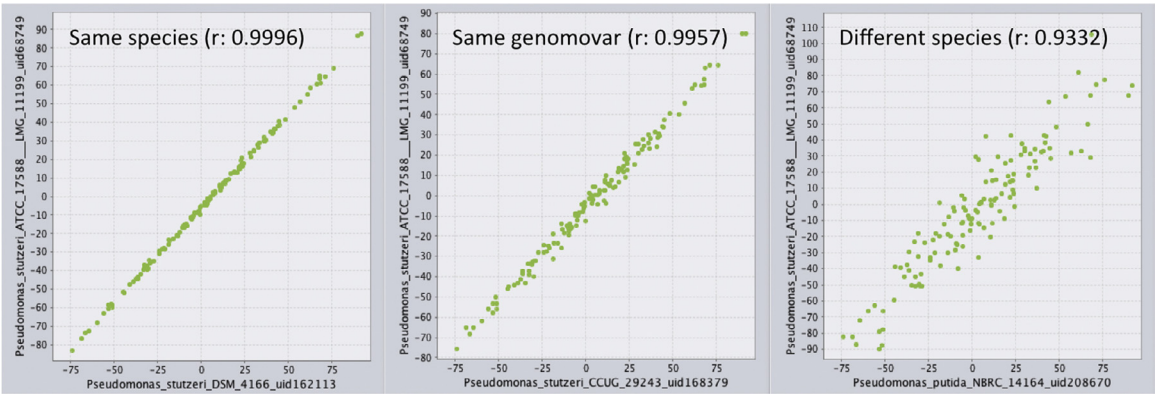


Fig. 2. Tetranucleotide charts showing the regression of signatures belonging to the same species, the same genomovar of the same species and different species.

species circumscriptions that will need to be verified by phenotypic studies (see below). An ANI range between 93 and 96% should be treated cautiously, and implies that the phenotypic diagnostic traits that would serve as a reason for splitting strains into two or more species need to be as robust as possible disregarding unstable or unclear phenotypes.

Phenotypic coherence

The phenotype is best defined as the "observable characteristics that result from the expression of genes of an organism, which can largely be modulated by environmental or other conditions" [35]. The origins of microbial taxonomy relied very much on phenotypic data to circumscribe taxa, and this information was of paramount importance for identification keys [49]. Actually, the phenotypic description of taxa links classification with any other discipline in microbiology. It is important to study the phenotype for any taxonomic description, since the general growth characteristics, colony and cell morphology, as well as physiological features, may not only be the result of the expression of genes, but also a reflection of the conditions under which the genes are expressed [25]. Based on the current knowledge of biochemistry, a significant part of the phenotype can already be predicted by automated or manual genome annotation. However, many genes still encode proteins of unknown function and, therefore, a solid phenotypic characterization will remain of paramount importance for taxonomy. In addition, the correct interpretation of coded information as a phenotype is essential for further evaluation of the putative function and interaction of microorganisms with their biotic and abiotic environment.

In general, metabolic characters have been studied by means of traditional laboratory tests, some of which can be found in commercial kits [49]. However, such tests should be used with care as many isolates exhibit requirements that prevent their adequate use. Phenotypic characters also include the composition and structure of components such as cell walls, cell membranes and the cytoplasm. These chemotaxonomic markers [57] are frequently stable and predictable [25]. Given that many of the expressed characters depend on environmental conditions (media, temperature, light, etc.), many methods suffer from variable reproducibility [35], and thus it is essential that rigorously standardized conditions are used for obtaining phenotypic data [25]. Moreover, in a similar way to what had often been expressed for methods such as DDH, some of the phenotyping methods currently in use appear to be cumbersome and are only used by a few research groups, and thus may be beyond the routine capacity of many laboratories [54].

Recently [54], the current practices in phenotyping for taxonomic purposes have been questioned given that (i) most of the new species classifications are based on SSSDs and, thus, the universality of the measured traits within the taxon is questionable; (ii) comparative studies require expensive reference strain libraries and/or also rather costly commercial tests; (iii) some fundamental parameters, such as cell morphologies, may be potentially variable; and (iv) some of the chemotaxonomic methods may be antiquated and subjective in their interpretation. Altogether, phenotyping seems to be the major barrier impeding the acceleration of species descriptions. One of the solutions to overcome this may be that large studies are retrospectively performed by specialized laboratories [54] once several strains are available, and that initial descriptions are based on meaningful minimal standards only centered on really relevant, and perhaps taxon-dependent, characters [54].

The final goal of phenotyping should be the generation of a comprehensive open access database of parameters that can be readily obtained in high throughput by standardized protocols and instruments. Future species definitions of *Bacteria* and *Archaea* would greatly benefit if such phenotype databases were used together, as occurs for the databases of 16S rRNA genes and genomes [48]. When this happens we would also anticipate a realistic chance for the acceleration of species descriptions.

Similarly to NGS improvements, taxonomy may also benefit from high-throughput phenotypic methods, such as mass spectrometry [27]. For example, the use of Matrix-Assisted Laser-Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS; [27,61]) targeting whole-cell biomass has been demonstrated to be an excellent tool for taxonomic and identification purposes. The success of this technique relies on the fact that the peptide profiles generated are mostly detecting ribosomal proteins, and these have a common evolutionary history with rRNA [61]. Actually, the studies of large sets of isolates by means of DDH and phylogenetic reconstructions based on 16S rRNA gene analyses have indicated that the clustering of MALDI-TOF MS profiles reflect species [36]. Therefore, in the near future, MALDI-TOF MS profiles will most probably be a standard parameter required when classifying new species. Alternatively, there are other powerful techniques, such as Ion Cyclotron Resonance–Fourier Transform Mass Spectrometry (ICR-FT MS), targeting small molecules (between 50 and 1000 D) that can provide metabolomic descriptions of the organisms under study [50]. This latter spectrometric technique still needs exhaustive standardization, but it may generate very valuable information for a database-based metabolic classification [48]. In all cases, different cultivation-based dependencies need to be taken into account [25]. Similarly, other techniques that some taxonomists recommend exploring and implementing in future descriptions are high-throughput Fourier transform infrared spectroscopy (HT-FTIR), high pressure liquid chromatography coupled to mass spectrometry (HPLC–MS), or gas chromatography coupled to isotope–ratio mass spectrometry (GC–IRMS), among others [27].

As a *recommendation for phenotypic coherence analyses* for taxonomic studies, it is important to extend the description as much

Table 1
Thresholds based on 16S rRNA gene identities for the different major taxonomic categories of prokaryotes. Genus and above thresholds were calculated by Yarza and colleagues [68], and the conservative threshold for species recommended by Stackebrand and Ebers [52] was used. The confidence interval is indicated for each value in brackets.

Category	Threshold	Minimum (%)	Median (%)
Species	98.7	98.7	
Genus	94.5	94.8 (94.5, 95.1)	96.4 (96.2, 96.6)
Family	86.5	87.7 (86.8, 88.4)	92.3 (91.7, 92.9)
Order	82.0	83.6 (82.3, 84.8)	89.2 (88.3, 90.1)
Class	78.5	80.4 (78.6, 82.5)	86.4 (84.7, 88.0)
Phylum	75.0	77.5 (75.0, 79.9)	83.7 (81.6, 86.0)

as possible where predictions based on genome annotation could serve as a major basis for the description, and, when possible, relevant physiological traits should be determined by standardized tests. The descriptions should be based preferably on phenotypic characters that may be evolutionarily relevant for the species under study, or at least serve as prominent traits for unequivocal identification. The environmental conditions of the isolation origin may help to decide which relevant tests need to be undertaken, rather than conducting available, yet irrelevant tests. In addition, it is of paramount importance that the methods used are standardized and reproducible in other laboratories [53]. High-throughput spectrometric and spectroscopic techniques may produce interesting and database-based taxonomic information, and may help to substitute some of the cumbersome and non-portable chemical analyses established in only a few laboratories.

Numerical thresholds for high taxa based on 16S rRNA. Only the species category is regarded as a real entity, whereas all higher categories in the taxonomic rank are considered to be abstract [22], and are therefore easily comparable between all taxonomies. However, there are no robust rules for their circumscription. This lack of criteria for circumscribing high taxa has been identified as a major problem in microbiology today [17]. In addition, although high taxa are considered to be ecologically meaningful, their coherence is inversely correlated to their taxonomic rank [41]. For this purpose, and given that the 16S rRNA gene is the one most represented in the public repositories [43], thresholds to circumscribe high taxa have been calculated on the basis of the known classification exemplified in the LTP [68]. The results have shown that quite stable thresholds can be established according to the taxonomic practices in the past (Table 1), and that by analyzing the current set of sequences for uncultured organisms the quantity of described taxa ranges from approximately 250,000 species to 1300 phyla yet to be classified [68]. Based on the median values shown in Table 1, higher taxa boundaries are spaced by steps of approximately 3%, so that pairwise identity values of almost full length 16S rRNA sequences of <95%, <92%, <89%, <86%, and <83% are indicative, respectively, of affiliation with different genera, families, orders, classes and phyla.

Whole genome pairwise comparisons will also help in circumscribing high taxa. However, as the genetic distance increases between genomes, it is frequently not possible to identify a sufficient number of true orthologues because the gene identities are low [31], and for this reason comparisons at the protein level may provide better resolution [47]. Extensive evaluations of the average amino acid identity (AAI) between genomes at higher taxonomic levels have been performed [31] and have shown it to be a very powerful parameter for improving the classification of higher taxa. In addition, the percentage of conserved proteins (POCP) has been proposed as a parameter for the circumscription of genera [42], although it still needs thorough evaluation. POCP may serve as an objective parameter where "*a prokaryotic genus can be defined as a group of species for which all pairwise POCP values are higher than approximately 50%*" [42]. Only the use of the parameters by taxonomists will prove their value, further reinforcing the relevance of genomic data for a natural classification.

A Candidatus taxa definition of uncultured organisms. The major difficulties in classifying prokaryotes arise from the need to obtain them as pure cultures in the laboratory so they can be studied in enough depth to construct a stable classification that is predictive and unequivocally identifies new isolates. The vast majority of existing prokaryotes are as yet uncultured. The 16S rRNA gene repositories already currently host ~250,000 species-level entries, which exceed the 12,000 hitherto classified species by a factor of 20× [68]. In some cases, given that certain uncultured organisms may exhibit conspicuous features, such as morphology, cell structures, and ecological uniqueness, the *Candidatus* provisional status for classification purposes was proposed [37,38T]. This issue is dealt with extensively in another chapter of this SAM special issue [29]. However, it is important to mention that given the current developments of NGS, the metagenomic approach allows the recognition of unique DNA populations that can be taxonomically analyzed by the genetic methods described above. It is now possible to track the phylogeny and the genomic coherence of distinct naturally occurring DNA populations observed in metagenomic studies [5]. In addition, the inference of metabolic properties from the (meta)genome data combined with microscopic observations using rRNA-targeted oligonucleotide probes for in situ localization [2], as well as some additional ecological data, will add further information to the classification of yet uncultured taxa.

The future

A wide application of the new high-throughput technologies to the taxonomy of *Bacteria* and *Archaea* would open the door to a new universe of parameters. However, care must be taken that this will not further accelerate the loss of traditional knowledge, which already currently has negative impacts on disciplines such as medical microbiology and diversity conservation in the culture collections. Therefore, we are convinced that the future of taxonomy lies in the generation of interactive genomic and phenotypic databases, which provide portable data in openly accessible public repositories. For accelerating the process of classification and enabling more laboratories to perform taxonomic descriptions, pragmatism must prevail in this transformation [48,54].

Looking into the future, taxonomists will need to re-evaluate the basic guidelines used for classification, as well as the minimum set of requirements and also the formats of their publications [54]. The one-species (often one strain) – one publication formula should be reconsidered, since it seems to overfill journals publishing taxonomic papers, which appear to reach a steady state due to volume and refereeing constrains [55]. The authors of this review agree that there is an urgent need to change traditional publication formats towards the development of novel templates for online publication, coupled to interactive databases [54]. Also, there is a need to find new formulas for preparing protologues, as these have become too long and often fail to indicate the truly important properties in which the new taxon differs from its relatives (comment provided by Aharon Oren). In our opinion, a modern protologue could be similar to the layout of the entries used for gene or genome sequences and be prepared in fields that can be computer-recognized, which are explicit and need to be filled in as standard procedure. Such protologues could be stored in interactive databases that could be easily managed. Whether such a layout should be embedded in the published description paper would be a decision for the journal publishing it, but it could be easily provided in a table format and removed from the main text.

NGS permits genotyping with an accuracy and data quality which is much higher than that of classical phenotyping. Due

to these reasons, it has been proposed that the description of full genome sequences may be the basis for future classifications [54,56], and that to a large extent the phenotype could be predicted from the gene sequence. Whitman proposes in this actual journal issue [62] that the genome sequences may already serve as type material, and the extracted DNA could be submitted to public culture collections, thus complementing or substituting the deposit of live type strains for classification purposes. If this radical proposal (i.e. genomic DNA as the only type material to be deposited) becomes practice, the process of authenticating type material will be faster and less expensive, by reducing the costs generated by culture collections [62], although the motivation for retrieving pure cultures might decrease even further.

Future taxonomic descriptions must therefore include high quality rRNA gene sequences (almost complete and free of errors), as well as high quality genome sequences of at least the type strain with coverage of no less than 99% of the genome size in just a few contigs. Complete or nearly complete genome sequences of microorganisms will in the near future be the basis of genealogical reconstructions [8], where different sets of many conserved genes, such as, for example, ribosomal protein gene sequences (rMLST; [24]), can be selected in an attempt to reconstruct robust genealogies [4]. Genome databases need curation because many entries with specific names either lack their corresponding type genomes for authentication, or the strains sequenced are misidentified [45]. However, important efforts to categorize the identity of the deposited genome entries are being taken [14] that will ultimately provide a reliable backbone for taxonomic purposes.

Today, phenotyping is often based on tests of questionable taxonomic value. Some of these are cumbersome and expensive due to the need for large reference strain libraries, and they have even been described as “antiquated” [54]. Given that most of the classifications are based on SSSDs, the diagnostic properties extracted may be of little value, and the validation of taxa properties may be better undertaken retrospectively once a reasonable number of strains have been collected [54]. We suggest that, in the future, viable pure cultures (or stable enrichments) should also be deposited along with pure genomic DNA, and that phenotypes should be extensively characterized, although still by standardized high-throughput techniques. The prediction of physiological traits based on genome-wide metabolic reconstructions will contribute significantly to the acceleration of classifications.

It would be better to abandon SSSDs or give them less prominence for classification purposes. Large-scale culturing techniques combined with MALDI-TOF MS examination [36,59] may help in isolating several strains of the same taxon, subsequently allowing characterization of the intra-species diversity. Finally, a challenge yet to be resolved is the reconciliation of the taxonomy of cultured organisms with the classification of the vast diversity of not yet cultured *Bacteria* and *Archaea*, which are readily detectable in the environment. The new -omics technologies already enable the classification of the status *Candidatus* among those uncultured organisms that do not exhibit conspicuous phenotypic traits [29], but for which (meta)genomes allow in-depth taxonomic comparisons and phenotype predictions. The final goal is that classifications as the candidate taxonomic units (CTUs; [68]) converge smoothly with the standard classifications of *Bacteria* and *Archaea*.

Acknowledgements

The authors acknowledge the critical comments of Kostas Konstantinidis and Peter Vandamme. R.R.M. acknowledges the support of Sara Díaz for the calculation of the ANIm values, as well as the scientific support given by the Spanish Ministry of Economy through the project CGL2012-39627-C03-03, which was also financed with European Regional Development Fund (FEDER) funds, and the preparatory phase of the Microbial Resource Research Infrastructure (MIRRI) funded by the EU (grant number 312251). In addition, the funding from competitive research groups (Microbiology) of the Government of the Balearic Islands (also co-supported with FEDER funds), is also acknowledged. R.A. acknowledges funding by the Max Planck Society and the German National Academy of Sciences Leopoldina for the working group “Taxonomy in the age of –OMICS”.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.syapm.2015.02.001.

References

[1] Achtman, M., Wagner, M. (2008) Microbial diversity and the genetic nature of microbial species. Nat. Rev. 6, 431–440.
[2] Amann, R., Ludwig, W., Schleifer, R. (1995) Phylogenetic identification and detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.
[3] Arahal, D. (2014) Whole-genome analyses: average nucleotide identity. Methods Microbiol. 41, 103–122.
[4] Boussau, B., Daubin, V. (2010) Genomes as documents of evolutionary history. Trends Ecol. Evol. 25, 224–232.
[5] Caro-Quintero, A., Konstantinidis, K.T. (2012) Bacterial species may exist, metagenomics reveal. Environ. Microbiol. 14, 347–355.
[6] Chun, J., Lee, J.-H., Jung, Y., Kim, M., Kim, S., Kim, B.K., Lim, Y.W. (2007) EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. Int. J. Syst. Evol. Microbiol. 57, 2259–2261.
[7] Chun, J., Rainey, F.A. (2014) Integrating genomics into the taxonomy and systematics of *Bacteria* and *Archaea*. Int. J. Syst. Evol. Microbiol. 64, 316–324.
[8] Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283–1287.
[9] Cohan, F.M. (2002) What are bacterial species? Annu. Rev. Microbiol. 56, 457–487.
[10] Deloger, M., El Karoui, M., Petit, M.-A. (2009) A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. J. Bacteriol. 191, 91–99.
[11] Doolittle, F., Zhaxybayeva, O. (2009) On the origin of prokaryotic species. Genome Res. 19, 755–756.
[12] Ereshefsky, M. (2010) Darwin’s solution to the species problem. Synthese 175, 405–425.
[13] Ereshefsky, M. (2010) Microbiology and the species problem. Biol. Philos. 25, 553–568.
[14] Federhen, S. (2014) Type material in the NCBI Taxonomy Database. Nucl. Acids Res., http://dx.doi.org/10.1093/nar/gku1127.
[15] Felis, G.E., Dellaglio, F. (2007) On species descriptions based on a single strain: proposal to introduce the status *species proponenda* (sp. pr.). Int. J. Syst. Evol. Microbiol. 57, 2185–2187.
[16] Garrity, G.M. 2001 Bergey’s Manual of Systematic Bacteriology, second ed., Springer, New York, NY.
[17] Garrity, G.M., Oren, A. (2012) Response to Gribaldo and Brochier-Armanet: time for order in microbial systematics. Trends Microbiol. 20, 353–354.
[18] Gevers, D., Dawyindt, P., Vandamme, P., Willems, A., Vancanneyt, M., Swings, J., De Vos, P. (2006) Stepping stones towards a new prokaryotic taxonomy. Philos. Trans. R. Soc. B 361, 1911–1916.
[19] Glaeser, S.P., Kämpfer, P. (2015) MultiLocus sequence analysis (MLSA) in prokaryotic taxonomy. Syst. Appl. Microbiol. 38, 237–245.
[20] 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.
[21] Hey, J. 2001 Genes, Categories and Species, Oxford University Press, New York, USA.
[22] Hull, D.L. (1997) The ideal species concept-and why we can’t get it. In: Claridge, M.F., Dawah, H.A., Wilson, M.R. (Eds.), Species: the Units of Biodiversity, Chapman & Hall, London, pp. 357–380.
[23] Jiménez, G., Urdiain, M., Cifuentes, A., López-López, A., Blanch, A.R., Tamames, J., Kämpfer, P., Kolstø, A.-B., Ramón, D., Martínez, J.F., Codoñer, F.M., Rosselló-Móra, R. (2013) Description of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise genome comparisons of the species of the group by means of ANI calculations. Syst. Appl. Microbiol. 36, 383–391.
[24] Jolley, K.A., Bliss, C.M., Bennett, J.S., Bratcher, H.B., Brehony, C., Colles, F.M., Wimalarathna, H., Harrison, O.B., Sheppard, S.K., Cody, A.J., Maiden, M.C.J. (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158, 1005–1015.

[25] Kämpfer, P. (2015) Continuing importance of the phenotype in the genomic era. Methods Microbiol. 41, 307–320.
[26] Kämpfer, P., Glaeser, S.P. (2012) Prokaryotic taxonomy in the sequencing era – the polyphasic approach revisited. Environ. Microbiol. 14, 291–317.
[27] Karlsson, R., Gonzales-Siles, L., Boulund, F., Svensson-Stadler, L., Skovbjerg, S., Karlsson, A., Davison, M., Hulth, S., Kristiansson, E., Moore, E.R.B. (2015) Proteotyping: proteomic characterization, classification and identification of microorganisms. Syst. Appl. Microbiol. 38, 246–257.
[28] Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, A., Ward, D.M., Rooney, A.P., Brambilla, E., Connor, N., Ratcliff, R.M., Nevo, E., Cohan, F. (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc. Natl. Acad. Sci. U.S.A. 105, 2504–2509.
[29] Konstantinidis, K., Rosselló-Móra, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomic data in the *Candidatus* approach. Syst. Appl. Microbiol. 38, 223–230.
[30] Konstantinidis, K., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U.S.A. 102, 2567–2592.
[31] Konstantinidis, K., Tiejde, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187, 6258–6264.
[32] 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 (1990 Revision), American Society for Microbiology, Washington, DC.
[33] Ludwig, W., Klenk, H.-P. (2001) Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In: Boone, D.R., Castenholz, R.W., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology, second ed., Springer, New York, NY, pp. 49–65.
[34] Meier-Kolthoff, J.P., Klenk, H.-P., Göker, 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.
[35] Moore, E.R.B., 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.
[36] Munoz, R., López-López, A., Urdiain, M., Moore, E.R.B., Rosselló-Móra, R. (2011) Evaluation of matrix-assisted laser desorption ionization-time of flight whole cell profiles for assessing the cultivable diversity of aerobic and moderately halophilic prokaryotes thriving in solar saltern sediments. Syst. Appl. Microbiol. 34, 69–75.
[37] Murray, R.G.E., Schleifer, K.-H. (1994) Taxonomic notes: a proposal for recording the properties of putative procaryotes. Int. J. Syst. Bacteriol. 44, 174–176.
[38] 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.
[39] Peeters, C., Zlosnik, J.E.A., Spilker, T., Hird, T.J., LiPuma, J.J., Vandamme, P. (2013) *Burkholderia pseudomultivorans* sp. nov., a novel *Burkholderia cepacia* complex species from human respiratory samples and the rhizosphere. Syst. Appl. Microbiol. 36, 483–489.
[40] Pernthaler, J., Amann, R. (2005) Fate of heterotrophic microbes in pelagic habitats: focus on populations. Microbiol. Mol. Biol. Rev. 69, 440–461.
[41] Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., Hallin, S. (2010) The ecological coherence of high bacterial taxonomic ranks. Nat. Rev. Microbiol. 8, 523–529.
[42] Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., Oren, A., Zhang, Y.-Z. (2014) A proposed genus boundary for the prokaryotes based on genomic insights. J. Bacteriol. 196, 2210–2215.
[43] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596.
[44] Reydon, T.A.C. (2004) Why does the species problem still persist? Bioessays 26, 300–305.
[45] 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.
[46] Rodríguez-R, M.L., Konstantinidis, K.T. (2014) Bypassing cultivation to identify bacterial species. Microbe 9, 111–118.
[47] Rong, X., Huang, Y. (2010) Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, proposal to combine 29 species and three subspecies as 11 genomic species. Int. J. Syst. Evol. Microbiol. 60, 696–703.
[48] Rosselló-Móra, R. (2011) Towards a taxonomy of *Bacteria* and *Archaea* based on interactive and cumulative data repositories. Environ. Microbiol. 14, 318–334.
[49] Rosselló-Móra, R., Amann, R. (2001) The species concept for prokaryotes. FEMS Microbiol. Rev. 25, 39–67.
[50] Rosselló-Móra, R., Lucio, M., Peña, A., Brito-Echeverría, J., López-López, A., Valens-Vadell, M., Frommberger, M., Antón, J., Schmitt-Kopplin, P. (2008) Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. ISME J. 2, 242–253.
[51] Soria-Carrasco, V., Valens-Vadell, M., Peña, A., Antón, P., Amann, R., Castresana, J., Rosselló-Mora, R. (2007) Phylogenetic position of *Salinibacter ruber* based on concatenated protein alignments. Syst. Appl. Microbiol. 30, 171–179.
[52] Stackebrandt, E., Ebers, J. (2006) Taxonomic parameter revisited: tarnished gold standards. Microbiol. Today 33, 152–155.
[53] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A.D., Kämpfer, P., Maiden, M.C.J., Nesme, X., Rosselló-Móra, R., Swings, J., Trüper, H.G., Vauterin, L., Ward, A., Whitman, W.B. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 52, 1043–1047.
[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. Ant. Van Leeuwen. 101, 13–20.
[55] Tamames, J., Rosselló-Móra, R. (2012) On the fitness of microbial taxonomy. Trends Microbiol. 20, 514–516.
[56] Thompson, C.C., Amaral, G.R., Campeao, M., Edwards, R.A., Polz, M.F., Dutilh, B.E., Ussery, D.W., Sawabe, T., Swings, J., Thompson, F.L. (2014) Microbial taxonomy in the post-genomic era: rebuilding from scratch? Arch. Microbiol., http://dx.doi.org/10.1007/s00203-014-1071-2.
[57] (a) Tindall, B.J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W., Kämpfer, P. (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. Int. J. Syst. Evol. Microbiol. 60, 249–266;
(b) Ursing, J., Rosselló-Móra, R., García-Valdés, E., Lalucat, J. (1995) Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups. Int. J. Syst. Bacteriol. 45, 604.
[58] Vandamme, P., Moore, E.R.B., Cnockaert, M., De Brandt, E., Svensson-Stadler, L., Houf, K., Spilker, T., LiPuma, J. (2013) *Achromobacter animicus* sp. nov., *Achromobacter mucicolens* sp. nov., *Achromobacter pulmonis* sp. nov., *Achromobacter spiritinus* sp. nov., from human clinical samples. Syst. Appl. Microbiol. 36, 1–10.
[59] Viver, T., Cifuentes, A., Díaz, S., Rodríguez-Valdecantos, G., González, B., Antón, J., Rosselló-Móra, R. (2015) Diversity of extremely halophilic cultivable prokaryotes in Mediterranean, Atlantic and Pacific solar salterns: evidences that unexplored sites constitute sources of cultivable novelty. Syst. Appl. Microbiol, in press.
[60] Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, L., Moore, L.H., Moore, W.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Trüper, H.G. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.
[61] Welker, M., Moore, E.R.B. (2011) Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst. Appl. Microbiol. 34, 2–11.
[62] Whitman, W.B. (2015) Taxonomic descriptions in the age of genomics. Syst. Appl. Microbiol. 38, 217–222.
[63] Woese, C.R. (1987) Bacterial evolution. Microbiol. Rev. 51, 221–271.
[64] 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., Kyripides, N.C., Klenk, H.-P., Eisen, J.A. (2009) A phylogeny driven genomic encyclopaedia of Bacteria and Archaea. Nature 462, 1056–1060.
[65] Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., Glöckner, F.O., Rosselló-Móra, R. (2010) Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. Syst. Appl. Microbiol. 33, 291–299.
[66] Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K.-H., Ludwig, W., Glöckner, F.O., Rosselló-Móra, R. (2008) The all-species living tree project: a regularly updated 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst. Appl. Microbiol. 31, 241–250.
[67] Yarza, P., Spröer, C., Swiderski, J., Mrotzek, N., Spring, S., Tindall, B.J., Gronow, S., Pukall, R., Klenk, H.-P., Lang, E., Verbarg, S., Crouch, A., Lilburn, T., Beck, B., Unosson, C., Cardew, S., Moore, E.R.B., Gomila, M., Nakagawa, Y., Janssens, D., De Vos, P., Peiren, J., Suttels, T., Clermont, D., Bizet Ch Sakamoto, M., Iida, T., Kudo, T., Kosako, Y., Oshida, Y., Ohkuma, M., Arahal, D.R., Spieck, E., Pommerening Roeser, A., Figge, M., Park, D., Buchanan, P., Cifuentes, A., Munoz, R., Euzéby, J., Schleifer, K.-H., Ludwig, W., Amann, R., Glöckner, F.O., Rosselló-Móra, R. (2013) Sequencing orphan species initiative (SOS): filling the gaps in the 16S rRNA gene sequence database for all species with validly published names. Syst. Appl. Microbiol. 36, 69–73.
[68] 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. 12, 635–645.