



## Classifying the uncultivated microbial majority: A place for metagenomic data in the *Candidatus* proposal



Konstantinos T. Konstantinidis<sup>a,b,\*</sup>, Ramon Rosselló-Móra<sup>c,\*\*</sup>

<sup>a</sup> School of Civil and Environmental Engineering, Georgia Institute of Technology, 311 Ferst Dr. NW, Atlanta, GA 30332, USA

<sup>b</sup> School of Biology, Georgia Institute of Technology, 311 Ferst Dr. NW, Atlanta, GA 30332, USA

<sup>c</sup> Marine Microbiology Group, Institut Mediterrani d'Estudis Avançats (IMEDEA; CSIC-UIB), E-07190 Esporles, Spain

### ARTICLE INFO

#### Keywords:

*Candidatus*  
Species concept  
Taxonomy  
Metagenomics  
16 rRNA gene  
Uncultivated diversity

### ABSTRACT

Microbial taxonomists have generally been reluctant to accept the valid publication of names of uncultured taxa given that only pure cultures allow for a thorough description of the genealogy, genetics and phenotype of the putative taxa to be classified. The classification of conspicuous uncultured organisms has been considered into the *Candidatus* provisional status, but this is only possible with organisms for which it is possible to retrieve basic data on phylogeny, morphology, ecology and some metabolic traits that unequivocally identify them. The current developments on modern sequencing techniques, and especially metagenomics, allow the recognition of discrete populations of DNA sequences in environmental samples, which can be considered to belong to individual closely related populations that may be identified as members of yet-to-be described species. The recognition of such populations of (meta)genomes allow the retrieval of valuable taxonomic information, i.e. genealogy, genome, phenotypic coherence with other populations, and ecological relevant traits. Such traits may be included in the *Candidatus* proposals of environmentally occurring, yet uncultured species not exhibiting exceptional morphologies, phenotypes or ecological relevancies.

© 2015 Elsevier GmbH. All rights reserved.

### The uncultured diversity and the *Candidatus* status for putative taxa

It is nowadays a commonplace to say that the vast majority of the microbial diversity has never been brought to culture in the laboratory. This realization is primarily due to the continuous development of molecular techniques applied to, in the first place taxonomy, and in the second place microbial molecular diversity studies of environmental samples. The taxonomy of prokaryotes has empirically developed in parallel to the technological advances that have allowed the retrieval of valuable genetic and phenotypic information beyond simple morphological traits [29]. However, the real success in the application of taxonomic methods directed to the identification of microorganisms has occurred in environmental microbiology. There, culture-independent technologies have

been extensively used to characterize naturally occurring communities. Microbial molecular ecology studies have overwhelmingly surpassed in efforts, scientific production and funding those on taxonomy. This is well reflected in the larger number of journals with higher citation indexes publishing microbial molecular ecology than taxonomic studies. The extensive interest in the discovery of environmental taxonomic novelty has resulted in the generation of millions of sequences that are deposited in public repositories, providing a better appreciation of the extent of genetic diversity on earth. In this regard, the number of deposited 16S rRNA gene sequences is currently close to 4 million [26]. On the other hand, the fraction of these sequences that corresponds to cultured organisms is far below 1% of the total entries, and that of the type strains of the hitherto classified species even smaller ( $n = 12,000$ ) [7,50].

There are neither official rules for the classification of prokaryotic taxa, nor an official classification [36]. Only the nomenclature of taxa is regulated under the International Code of Nomenclature of *Bacteria* [18]. Actually, for most taxonomists, the official recognition that a species is accepted as a new taxon is the valid publication of its name in the official journal of the International Committee for Systematics of Prokaryotes (ICSP). The effective publication of a name (i.e. the description and classification of the new taxon) occurs by either its publication in the International

\* Corresponding author at: School of Civil and Environmental Engineering, Georgia Institute of Technology, 311 Ferst Dr., Atlanta, GA 30332-0512, USA. Tel.: +1 404 385 3628.

\*\* Corresponding author. Tel.: +34 971 611 826.

E-mail addresses: [kostas@ce.gatech.edu](mailto:kostas@ce.gatech.edu) (K.T. Konstantinidis), [rossello-mora@uib.es](mailto:rossello-mora@uib.es) (R. Rosselló-Móra).

Journal of Systematic and Evolutionary Microbiology (IJSEM; published and listed in the “Notification Lists”), or when published in other non-official journals, the name should appear (after request of the authors) in the IJSEM “Validation List” [43]. However, in order to validly publish a name, its effective publication should meet several requirements raised in the Bacteriological Code. Some of them are formalities related to the etymological correctness of the proposed name, and other related to the extent of phenotypic and genetic information provided in the protologue. However, the most important cornerstone for the valid publication of a specific name under the Bacteriological Code, which is covered by its rule 18a, is the requirement for a pure culture to be designated as the type strain [18]. This requirement is one of the most controversial issues among microbiologists as it seems to hamper the “classification” of uncultured organisms or microbial consortia (the latter covered by the rule 31b of the code). The benefits of having an organism in pure culture in the laboratory are obvious for microbiological studies. In addition, the special requirement covered by rule 30a of the code, which is the deposition of the type strain in at least two international collections [43] guarantees access and reproducibility of the results in taxonomic studies. On the other hand, the need to isolate pure cultures has several major limitations. At first, and given that most of the environmentally occurring microorganisms are resistant to cultivation [3], the extent of known classifiable taxa surpasses that currently classified by orders of magnitude. Second, the isolation and characterization of organisms needs skills, time and effort, which currently leads to a rate of classification of about 700 species yearly [29], a speed that given the vast diversity awaiting to be classified, is slow. Finally, culture techniques tend to recover a tiny part of the real diversity of a given sample, but also very rarely the most abundant taxa that may play the key roles [24].

Given the difficulties in culturing most of the environmentally occurring organisms, there is an urgent need to classify the extent of diversity discovered by molecular techniques. Actually, the development of methods that allow the identification, observation, quantification, and (in some extent) assessment of metabolic properties, have brought taxonomists to introduce the *Candidatus* status [21,22]. *Candidatus* is a category with no standing in the Bacteriological Code, thus cannot be considered a rank, but a status that is permitted to be listed in the “notification or validation” lists, and denotes that the putative taxon is awaiting for a formal validation once its representatives are brought to pure culture and extensively described. Given that a *Candidatus* name can be formally recorded as susceptible to be validly published, the condensed description should meet some minimal etymological and protologue formalities [22].

*Candidatus* is the closest status to species with a name that stands in nomenclature. It is generally (and probably wrongly) understood as an incomplete species classification. It is indeed incomplete for the requirements of the Bacteriological Code to recognize the name as validly publishable; yet, it is the closest recognition that a species is formally classified within a hitherto unofficial taxonomy. As long as the data provided in the description (i.e. molecular, morphological, metabolic and ecological traits), unequivocally identifies it, the classification of the organism should be understood as acceptable. It is a somewhat a separate issue whether or not the ICSP considers the name to be provisional.

The activities of environmental microbiologists using molecular methods to describe uncultured taxa has led to the description of almost 360 *Candidatus* taxa (<http://www.bacterio.net/-candidatus.html>), but in all cases the organisms exhibited very conspicuous traits (morphology, size, environmental uniqueness and isolation, extraordinary metabolic properties, etc.) that allowed their unequivocal recognition. Good examples are the giant size of the nitrate oxidizer “*Candidatus* Thiomargarita joergensenii” [33]; or the magnetic body inclusions of “*Candidatus* Magnetobacterium

**Table 1**

Numbers of validly published names listed in the LPSN, putatively detected taxa in the public repositories (using the thresholds calculated for species (98.7%), genus (94.5%), family (86.5%), order (82.0%), class (78.5%) and phylum (75.0%)) and ambiguous sequences that according to SILVA 114 (December 2012) showed a pinstail score below 75% [51]. The data has been calculated with the entries up to year 2012.

	Validly published	SILVA REF 114 (2012)	Ambiguous
Sequences	–	1,306,670	31,469
Species	10,015	241,254	28,983
Genera	2029	80,939	19,607
Families	317	14,369	4803
Orders	137	5366	1148
Classes	88	2573	408
Phyla	29	1356	84

bavaricum” [37]. However, most of the organisms inhabiting environmental niches do not exhibit such conspicuous properties that simplify their recognition. Standard molecular data such as 16S rRNA gene sequence, rRNA targeted probes to identify and quantify intact cells, and even sometimes the assessment of metabolic properties, do not produce enough discriminative data for the unequivocal identification of the majority of uncultivated taxa. However, as exemplified below, the new –omics technologies can contribute substantially in this direction and facilitate the recognition of new taxa as an extension of the parameters to be provided in *Candidatus* proposals.

### The achievability of a global classification system

16S rRNA gene sequences may be one of the most, if not the most, extensively deposited information in public repositories, currently accounting for almost 4 millions entries [26]. A survey based on sequence identity thresholds has been performed to enumerate putative taxa discovered by environmental studies [51]. The thresholds were calculated based on the Living Tree Project (LTP) database that contains only high-quality, curated sequences of type strains with validly published names [49]. The numbers of detected taxa using the SILVA REF 114 database released in December 2012 [26], considering only good quality sequences longer than 900 nucleotides, were surprising, and revealed numbers close to 250,000 species (with the conservative threshold of 98.7%; [38]), or over 1300 phyla (Table 1). Based on the discovery rates and current sequence deposits, it was calculated that the total number of putative prokaryotic species inhabiting the biosphere would range between  $5 \times 10^5$  and  $2 \times 10^6$ , but in no case would exceed  $10^7$  [51]. These numbers were quite encouraging, as the global classification of prokaryotic taxa seemed like an unachievable task previously.

The vast, but finite, extent of taxonomic diversity hitherto detected in the public repositories requires classification for obvious reasons. Among them, it is of paramount relevance to facilitate the understanding of the uncultured diversity using common criteria with cultured organisms. With this in mind, the “*Candidatus* Taxonomic Unit” (CTU) was proposed [51] as a way to categorize environmental sequences, which, by means of taxonomic thresholds and phylogenetic uniqueness (monophyly), would be reminiscent of the categories recognized in the bacteriological code. A CTU was considered as “a biological entity delineated by a monophyletic set of sequences with a minimum identity that stays within, or very close to, the taxonomic threshold proposed for a given rank”. Moreover, the CTU was proposed to fit in a hierarchical system with the same categories as that of the cultured organisms, but with an alphanumerical nomenclature that would be computer readable. Applying these criteria, a hierarchical layout was proposed for most of the environmental clades, which could also facilitate the reclassifications of phyla such as the *Spirochaetes* [51].

Download English Version:

<https://daneshyari.com/en/article/2062965>

Download Persian Version:

<https://daneshyari.com/article/2062965>

[Daneshyari.com](https://daneshyari.com)