




Authors need to be prudent when assigning names to microbial isolates

Erko Stackebrandt  · Juan A. Mondotte · Luigi Lembo Fazio · Mike Jetten

Published online: 7 January 2022

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Isolating new microbial strains and submitting the corresponding manuscript to a scientific journal for peer-review is an exciting step for any microbiologists. After a series of dedicated experiments to characterize these new isolates, writing sessions and feedback iterations with your co-authors, you are ready to enter the editorial system of the Journal of choice. Scientific journals provide their authors with clear advice/directives on aims and scope, style and format requirements, permissions, etc. as ‘Submission Guidelines’. These Instructions may differ among journals due to different focus on content and the history of the journal. Regular updates are necessary to follow new developments in methodologies, concepts,

analytic software programs, and publication formats. Despite this assistance to match a journal’s style and to facilitate the work of editors and reviewers these guidelines are not always followed consistently by authors. As a consequence too many manuscripts are returned for revision by handling editors before they can be sent to reviewers to evaluate their scientific merit, thus resulting in delays in the reviewing process. In order to avoid this time-consuming ‘unnecessary’ iteration steps, editors of a few Springer Nature microbiology journals summarize some of the main omissions and recommend advice which should be read as an explanatory note to the ‘Guidelines’.

The following short text is meant for authors attempting to affiliate environmental isolates or clone sequences to species with validly published names. In contrast to the procedure of formal naming of novel strains at the rank of species [International Code of Nomenclature of Prokaryotes (Parker et al. 2019)] and guidance through minimal standards for the description of new taxa (<https://lpsn.dsmz.de/text/minimal-standards>) the aspect of assigning a name to other microbial resources is not regulated. This aspect, however, deserves as much attention in order not to create a database in which a defined name is pinned to a genetic entity that may differ significantly from the one whose name it bears.

By end of August 2021, the List of Prokaryotic names with Standing in Nomenclature (LPSN, Parte et al. 2020) contains 21.261 species with validly

This Editorial was simultaneously published in Archives of Microbiology (2021a) and Current Microbiology (2021b).

E. Stackebrandt (✉)
Ampleben, Germany
e-mail: est@dsmz.de

Present Address:
J. A. Mondotte
Paris, France

L. L. Fazio
Rome, Italy

M. Jetten
Microbiology, Radboud University, Nijmegen,
Netherlands
e-mail: m.jetten@science.ru.nl

published names. The recent release of the All-species Living Tree Project (LTP) (www.imedea.uib-csic.es/mmg/ltp; Ludwig et al. 2021) contains more than 17,500 curated complete or almost complete 16S rRNA gene sequences of archaeal and bacterial type strains (indicating that such sequences are available for comparison of > 80% of the type strains). In contrast, the number of all 16S rRNA sequences in GenBank-NCIB (the majority being less than 300 nucleotides of short variable regions) deposited are as high as 40 million (Marcela Borba, pers. communication).

1. In manuscripts giving names to bacterial and archaeal isolates a major issue is the unfamiliarity of the meaning of a ‘type strain’ or the nescience to use this category for comparative purposes. A type strain is the strain on which the description is based, the name bearing strain, hence the point of reference (Lapage et al. 1990; Parker et al. 2019). As the intra-species variability in phenotypic and genomic differences among strains is often wide, the type strain—and not the species/strain name with the highest similarity, e.g. BLAST score—is the most important one to be used. Hence, new isolates must be compared to the phenetic and genetic properties of the type strain when attempts are made to affiliate strains to species, i.e. giving a name to a hitherto unassigned strain. While this is a mandatory aspect of any new prokaryotic species description, the affiliation of environmental isolates is in most cases not done against a type strain but to the one that shows the highest similarity 16S rRNA gene similarity value. However, as the vast majority of sequences are named without reference to type strains the credibility of such species affiliation is not granted. Comparison to type strains is mandatory for any methodological approach as it is the only strain which is available from public resource centers for comparative purposes while for the vast majority of strains the gene sequence accession number but not the strain itself is available.

Advice to authors: when using BLAST-NCBI tick the option ‘Limit to sequences from type material’ or, in the case of EzBioCloud (Yoon et al. 2017a), restrict the search only to type strains. Also, the type strain should be denoted with a with a superscript ^T after the strain number.

2. Wet-lab DNA-DNA hybridization (DDH) has been used for several decades as a basis for the recognition of a genomospecies (above a 70% threshold value), originally defined by Wayne et al (1987).

With the introduction of ribosomal RNA gene sequencing the measure of relatedness was footed at a more objective basis. For complete or almost complete 16S rRNA sequences (> 1,300 nt) a 97% similarity threshold value between an isolate and type strains has been originally recommended for recognizing a novel genomospecies (Stackebrandt and Goebel 1994). Above this value, due to the evolutionary constraint nature of the 16S rRNA gene sequence (Woese et al 1990), DNA hybridization still needed to be performed to determine whether the isolate can be affiliated to a species with validly published names or whether it represents a novel species. This value was later increased to species cut-off values around 98.7% 16S rRNA gene similarity (Stackebrandt and Ebers 2006).

An even more advanced step in the definition of a genomospecies was reached some years ago by in-silico methods based on the comparison of genome sequences that made laborious wet-lab DNA-DNA hybridization studies between an isolate and most closely related type strains unnecessary. A range of 98.2 and 99.0 % cut-off values of digital (d)DDH values were recommended by Auch et al. (2010) and Meier-Kolthoff et al. (2013) for in-silico genome-to-genome comparison which showed a better correlation with 16S rRNA gene sequence distances than the wet-lab DDH values. For an alternative measure of relatedness the Average Nucleotide Identity (ANI/OrthoANI) values of genome sequences around 95% were defined (Konstantinidis and Tiedje 2005; Rodriguez-R and Konstantinidis, 2014; Yoon et al. 2017b; Lee et al. 2016; Jain et al 2018). Thus, in groups of species, once described on the basis of lower than 70% DNA hybridization values but sharing higher than about 98.7% 16S rRNA gene sequence similarity among each other (most prominent examples [but not exclusively] are members of the *Bacillus cereus* or *B. subtilis* group, many species of *Streptomyces*, *Vibrio*, *Rhizobium*, or *Pseudomonas*) an isolate cannot be assigned a species name without additional genomic evidence.

Advice to authors: Additional evidence must be provided before a name is given to an isolate, such as generated by DDH, dDDH, or ANI analysis between an isolate and type strain of the most highly related species. Also, MALDI-TOF (Shah et al. 2020) or multilocus sequence analysis -MLSA- (Glaeser and Kämpfer 2015) are helpful to assign an isolate to a validly named species. In the absence of any of these data authors should be prudent to actually name an isolate but either use the name of a species group or the Genus spp category.

3. The situation is more complex in mycology as the 18S rRNA gene does not discriminate between closely related species and the size of the genome sequencing makes its routine analysis hardly feasible. DNA barcode sequences of ITS regions (ITS1, ITS2), 5.8S rRNA, internal described spacer and flanking regions of the 18S (SSU)—and 28S rRNA (LSU) regions are the preferred choice. While for yeast the combination of LSU and ITS discriminates well among species (with a threshold value of 98.4–99.5), this approach cannot be applied to the entire fungal kingdom as the intraspecific variability of ITS regions varies considerably and a similar range of threshold value cannot be applied for the species level (Raja et al. 2017). The underlying results and problems associated with sequence-based methods in mycology are well outlined in the article by Borman and Johnson (2020). As the species-resolution is often poor in many fungal taxa the discrimination power of regions of protein-coding genes were compared (Stielow et al. 2015) to select specific barcodes for species identification, among which TEF 1 α , but also COX1, PBR2, ACT, TUB2, or RPL 10 ranked high in certain fungal lineages. Genes and primer sequences for DNA-based identification are summarized, among others, by Borman and Johnson (2020) and Tekpinar and Kalmer (2019).

Advice to authors: As ITS regions alone are insufficient to unambiguously name a new fungal isolate in a given genus or genus complex the literature must be searched for more discrimination barcoding regions to support the naming. In the absence of such data the authors must be refrain of applying an explicit name to an isolate.

4. Yet another source of premature assignment of species names to 16S rRNA gene sequences derives from the attempts to categorize Operational Taxonomic Units (OTUs) generated by next-generation sequencing of variable regions of this gene (usually V3 and V4). Due to the massive generation of such short sequences, different protocols were published (see e.g. Caporaso et al. 2010) allowing to subsampled open-reference OTUs. Going back to an early paper by Stackebrandt and Göbel (1994) who proposed a 97% threshold value of almost complete (> 1300 nt) 16S rRNA gene sequence similarities to separate different species (see paragraph 2), this value has been applied to define species and genera on the basis of short-sequence OTUs (250–500 nt) of metagenomic studies (e.g. Schloss and Handelsmann 2005; Rideout et al., 2014). This value has been re-assessed by analysis of high quality 16S rRNA gene sequences to species delineation threshold optimization at about 99% for full length sequences (in accord with the revised data; see paragraph (2) and 100% for the V4 region (Edgar, 2018). Regrettably, these more recent values are ignored in the majority of submissions and species names are still affiliated to short stretch-OTUs sharing higher than 97% 16S rRNA gene similarity—in most studies without providing intra-OUT ('species') similarity values. Along the same line is the application of linking functions to previously sequenced genomes of 97%-defined OTUs (e.g. Langille et al. 2013; PICRUST), generating a mélange of genomic properties from a group of species with often significantly different genomic makeup. In the absence of cultured organisms the evaluation of such genomic properties of environmental microorganisms is indeed helpful to understand the function of an ecosystem—though the threshold values for OTUs should be narrowed down to 99% to make sure that only the genomes of the closest relatives at the intra- and infraspecific levels are included.

Advice to authors: In combination with arguments listed under 1. and 2. the names given today for OTU-defined 'species' are highly speculative, not re-assessed by different approaches but once published, taken for granted. The threshold value of short-stretch

environmental OTUs should be increased from 97% to 99% to narrow the species diversity. As now species name-attached sequences enter databases the likelihood for a future even higher erroneous identification is predictable, hence authors should refrain from naming species on the basis of such shaky scientific grounds.

We hope these advice help to design your studies, ease the preparation of manuscripts and make the submission process more agreeable. Fingers crossed for a constructive peer-review process and good luck with your next isolation attempt.

Authors contributions All authors participated in the concept, writing and editing. All authors approved the submitted manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Auch AF, Von Jan M, Klenk HP, Göker M (2010) Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2:117–134. <https://doi.org/10.4056/sigs.531120>
- Borman AM, Johnson EM (2020) Sequence-based identification and classification of fungi. In: *Trends in the Systematics of Bacteria and Fungi*. Bridge P, Smith D, Stackebrandt E. (eds). CABI UK, 198–216. ISBN: 9781789244984
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD et al (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Meth* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>
- Edgar R C (2018) Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* 15;34:2371–2375. <https://doi.org/10.1093/bioinformatics/bty113>
- Glaeser SP, Kämpfer P (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *System Appl Microbiol* 38:237–245. <https://doi.org/10.1016/j.syapm.2015.03.007>
- Jain C, Rodriguez-R LM, Phillippy AM et al (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
- Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 102:2567–2572. <https://doi.org/10.1073/pnas.0409727102>
- Langille MGI, Zaneveld J, Caporaso J, McDonald D, Knights D et al (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*. <https://doi.org/10.1038/nbt.2676>
- Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR et al (1990) International code of nomenclature of bacteria. ASM Press, Washington
- Lee I, Ouk Kim Y, Park SC, Chun J (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100–1103. <https://doi.org/10.1099/ijsem.0.000760>
- Ludwig W, Viver T, Westram R, Gago JF, Bustos-Caparrós E, Knittel K, Amann R, Rossello-Mora R (2021) Release LTP_12_2020, featuring a new ARB alignment and improved 16S rRNA tree for prokaryotic type strains. *Syst Appl Microbiol* 40:126218. <https://doi.org/10.1016/j.syapm.2021.126218>
- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinfo*. <https://doi.org/10.1186/1471-2105-14-60>
- Parker CT, Tindall BJ, Garrity GM, eds. (2019) International Code of Nomenclature of Prokaryotes Prokaryotic Code (2008 Revision). *Int J Syst Evol Microbiol*. <https://doi.org/10.1099/ijsem.0.000778>
- Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M (2020) List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 70:5607–5612
- Raja HA, Miller AN, Pearce CJ, Nicholas H, Oberlies NH (2017) Fungal identification using molecular tools: a primer for the natural products research community. *J Nat Prod*. <https://doi.org/10.1021/acs.jnatprod.6b01085>
- Rideout JR, He Y, Navas-Molina JA, Walters WA, Luke K, Ursell LK et al (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *Peer* 2:e545. <https://doi.org/10.7717/peerj.545>
- Rodriguez R, Konstantinidis K (2014) Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters avoid cultivation bias, and provide true insights into microbial species. *Microbe Magazine*. <https://doi.org/10.1128/MICROBE.9.111.1>
- Schloss PD, Handelsman J (2005) Introducing dotur, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol*. <https://doi.org/10.1128/AEM.71.3.1501-1506.2005>
- Shah HN, Shah AJ, Belgacem O, Ward M, Dekio I et al. (2020) MALDI-TOF MS and currently related proteomic technologies in reconciling bacterial systematics. In: Bridge P, Smith D, Stackebrandt E. (eds), *Trends in the Systematics of Bacteria and Fungi*. CABI UK, 93–118. ISBN: 9781789244984
- Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 33:152–155
- Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Evol Microbiol* 44:846–849. <https://doi.org/10.1099/00207713-44-4-846>
- Stackebrandt E, Mondotte JA, Fazio LL, Jetten M (2021a) Authors need to be prudent when assigning names to

- microbial isolates. *J Arch Microbiol* 203:5845–5848. <https://doi.org/10.1007/s00203-021-02599-7>
- Stackebrandt E, Mondotte JA, Fazio LL, Jetten M (2021b) Authors Need to be Prudent When Assigning Names to Microbial Isolates. *Curr Microbiol*. <https://doi.org/10.1007/s00284-021-02678-4>
- Stielow JB, Lévesque CA, Seifert KA, Meyer W, Irinyi L et al (2015) One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia* 35:242–263
- Tekpinar AD, Kalmer A (2019) Utility of various molecular markers in fungal identification and phylogeny. *Nova Hedwigia*. https://doi.org/10.1127/nova_hedwigia/2019/0528
- Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O et al (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* 37:463–464. <https://doi.org/10.1099/00207713-37-4-463>
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci* 87:4576–4579. <https://doi.org/10.1073/pnas.87.12.4576>
- Yoon S-H, Ha S-M, Kwon S et al (2017a) Introducing EzBio-Cloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol*. <https://doi.org/10.1099/ijsem.0.001755>
- Yoon S H, Ha S M, Lim J, Kwon S, Chun J (2017b) A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek*, 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.