environmental microbiology

Environmental Microbiology (2020) **00**(00), 00–00 doi:10.1111/1462-2920.14934

Correspondence

Advantages outweigh concerns about using genome sequence as type material for prokaryotic taxonomy

Konstantinos T. Konstantinidis[1*]
Ramon Rosselló-Móra[2*] **and Rudolf Amann**[3*]
[1]*School of Civil and Environmental Engineering and School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA.*
[2]*Marine Microbiology Group, Institut Mediterrani d'Estudis Avançats (IMEDEA; CSIC-UIB), E-07190, Esporles, Spain.*
[3]*Max Planck Institute for Marine Microbiology, D-28359, Bremen, Germany.*

About two years ago, we pointed out the importance of reconciling the taxonomy of cultivated organisms with that of the uncultivated taxa toward a single, standardized nomenclatural system that will encompass all Prokaryotes (Konstantinidis *et al*., 2017). We (Konstantinidis *et al*., 2017, 2018), and others (Hedlund *et al*., 2015; Whitman, 2015, 2016; Whitman *et al*., 2019), believe that this is a feasible task and, in fact, it would only require two straightforward changes to the International Code of Nomenclature for Prokaryotes [ICNP; (Parker *et al*., 2019). That is, (i) to give priority to *Candidatus* names, and thus, treat them similarly to the names of organisms isolated in pure culture (Konstantinidis and Rosselló-Mora, 2015; Whitman *et al*., 2019), and (ii) to qualify genome sequences as an alternative type material (voucher) for taxonomic descriptions (Whitman, 2015, 2016). It is important to note that this proposal was not meant to substitute the deposition of isolated type strains in cases where those are available. In fact the intention of the scientists proposing genome sequences as alternative type material was not to weaken the high standards of prokaryotic taxonomy that are in place, but rather, to bring compatible taxonomic standards to the genomic information of uncultivated taxa retrieved over and over again from all important habitats, being it environmental or of medical relevance.

The proposal that DNA could serve as alternative type material (Whitman, 2015) (Konstantinidis *et al*., 2017) has raised strong concerns (Oren and Garrity, 2018; Bisgaard *et al*., 2019; Overmann *et al*., 2019) some of which we would like to address herein. In particular, we doubt that '*the motivation for researchers to cultivate and preserve strains and to attempt to investigate phenotypes will decrease*' (Bisgaard *et al*., 2019). We argue here that the proposed changes to ICNP are unlikely to result in a lower focus on isolation efforts since isolating an organism in the laboratory has important advantages for its study and use in downstream applications. In fact, we can offer at least one example from our own work where the culture-independent discovery of an abundant bacterial halophile in salterns (Anton *et al*., 2000) led to its cultivation (Anton *et al*., 2002). There is also a more recent case of an ubiquitous oil-degrading organism that was first observed based on metagenomes and was subsequently isolated in pure culture due to its apparent important role in oil biodegradation (Karthikeyan *et al*., 2019). There are many more examples in the literature where the knowledge of uncultured and ecologically relevant microorganisms led to their isolation (Stott *et al*., 2008; Harbison *et al*., 2016; Henson *et al*., 2018; Lee *et al*., 2019), and the successful application of novel metagenome-guided cultivation methods (Tyson *et al*., 2005; Karthikeyan *et al*., 2019; Zhang *et al*., 2019). We are convinced that the in-depth taxonomic description of yet-uncultured prokaryotic clades will actually make continued isolation efforts more relevant due to potential economic benefits (Keller and Zengler, 2004), and the personal satisfaction from cultivating not any bacterium, but a 'missing' candidate taxon (Pandit and Rahalkar, 2019).

Furthermore, we do not anticipate an overwhelming increase in the number of sloppily described candidate taxa, especially if the classification standards would require multiple high-quality genome sequences from different sites or sampling times. A standing committee for

Received 25 January, 2020; accepted 27 January, 2020. For correspondence. *E-mail kostas@ce.gatech.edu. Tel. 001-404-385-3628; Fax 001-404-894-8266. **E-mail rossello-mora@uib.es. Tel.+34 971 611 826; Fax +34 971 611 761. ***E-mail ramann@mpi-bremen.de. Tel.(+49) 421 2028 930; Fax (+49) 421 2028 580.

the taxonomy of the uncultured could effectively discourage descriptions based on single genomes (e.g., single single-cell amplified genomes or SAGs), much as single strain descriptions of novel species of Bacteria and Archaea should be avoided. The best genome sequence available should serve as alternative type material, and additional information on diversity, occurrence and the partly predicted, partly measured phenotypic information should be part of the description (Konstantinidis and Rossello-Mora, 2015; Konstantinidis *et al*., 2017). That is, the description of such taxa would require a substantial effort on behalf of the authors and thus, only yet-uncultivated microorganisms of interest would be taxonomically classified among the 'great majority' of uncultivated taxa that exist in nature. Accordingly, we expect the increase in the number of described uncultivated taxa to be modest. It would, in all likelihood, not overwhelm the review process and publishing resources available.

Next, we would like to consider the criticism that metagenome-assembled genomes (MAGs) are of insufficient quality to serve as a stable type material. Vouchers have to be sufficiently detailed and stable to allow for unequivocal identification. The information content of MAGs and SAGs is today routinely used to reveal the genealogy of the microorganisms. The information content is also sufficient for identification purpose. Bioinformatics predictions following the community standards recently proposed (Field *et al*., 2008) can serve as a minimum description of the functional potential. Metatranscriptomics, metaproteomics or isotope-based approaches (e.g., NanoSIMS) can be used to confirm, at least part of the bioinformatics predictions and/or reveal the *in situ* functions carried by the organisms, if desired. These 'environmental' data are potentially even more relevant than some of the phenotypic tests enforced on isolated organisms in the laboratory, especially when the laboratory growth conditions deviate from the *in situ* conditions, as it is often the case.

Several scientists have argued that MAG/SAG-based information is less detailed than the information derived from isolate-based experiments, and they presented examples where the MAG/SAG quality is lower compared with what the currently available bioinformatics pipelines for quality estimation predict and thus, does not represent well the organisms under investigation (Bisgaard *et al*., 2019; Overmann *et al*., 2019). One should acknowledge here that journals publishing taxonomic studies have made compulsory the deposition of genome sequences for novel taxa. Journals as Systematic and Applied Microbiology (compulsory since 2014), Archives of Microbiology and Current Microbiology (both since 2017) or the International Journal of Systematic and Evolutionary Microbiology (since 2019) have adopted this policy in an effort to improve prokaryotic taxonomy. Therefore, the value of genomic information should not be challenged *per se*. While it is true that MAGs and SAGs could generally be of lower quality compared with isolate genomes (note that isolate genome sequences could also be of low quality or contaminated), this is not critical enough to prevent progress towards cataloguing the taxonomic diversity of uncultivated organisms, for several reasons. First, prokaryotic taxonomy has always relied on imperfect methods; MAGs/SAGs are not an exception to this. Take, for instance, the DNA–DNA hybridization (DDH) method, the 'gold standard' for species demarcation (more precisely, genomospecies demarcation). The genome-aggregated average nucleotide (ANI) value of shared genes between two related genomes (Konstantinidis and Tiedje, 2005) has been shown to correlate well with their DDH values, and deviations in the values were common and largely attributable to the experimental noise of the former as opposed to the latter method (Goris *et al*., 2007). Second, there are approaches to assess quality beyond reasonable doubt such as visual examination of read-recruitment pots (Rodriguez-R and Konstantinidis, 2016) in combination with the quality checking pipelines (Parks *et al*., 2015; Rodriguez *et al*., 2018), and in our view, only genomes of high enough quality based on these tests should be taxonomically described (Konstantinidis *et al*., 2017) [for some possible exceptions to the latter, please see (Konstantinidis *et al*., 2017)]. Third, the standards to use have been outlined already previously by us (Konstantinidis *et al*., 2017) and others (Bowers *et al*., 2017), and are of similar stringency to those used for isolate genomes; the reader is referred to these publications for further details. Furthermore, long-read sequencing for routine taxonomic descriptions, even on environmental samples, is coming up soon (e.g., Andersen *et al*., 2019), and is strongly expected to circumvent several of the low-quality issues reported for MAGs and SAGs in the literature, e.g. provide complete genome of similar quality to the isolate genomes, and/or help to identify and fix genome sequences that may be chimeric. It has been argued that when DNA sequence type material is replaced by new versions due to new sequencing technologies and/or tools for genome assembly, the species descriptions would have to be consequently revised, resulting in an unstable classification (Bisgaard *et al*., 2019). However, this is unlikely to be true for most—if not all—taxa because such new versions will mostly affect only a small number of genes or nucleotide substitution positions in the genome as analysis of mock datasets of known composition has revealed (Sczyrba *et al*., 2017) or the sequencing of the isolated '*Candidatus* Macondimonas diazotrophica' that was almost identical to its corresponding MAG (e.g., ANI >99.9%)

(Karthikeyan *et al*., 2019). It is even less likely that the affected genes by new genome versions would represent the species-diagnostic traits because these genes are often the hypothetical, mobile or prophage-associated genes found in multiple copies (and short contigs) in the genome (Pena-Gonzalez *et al*., 2019). It is also important to realize that for two genomes to accumulate ~1% difference in their ANI value, more than 20 000 years of evolution would be required (Lawrence and Ochman, 1998), which represents a too long time to affect current taxonomy practice. Hence, the genealogy of the genome and thus, its nomenclature and classification, will remain unaffected in the great majority of cases where new versions of the genome become available. In a few cases that the new genome version will include major changes in gene content, the old version could be replaced by the new version in a process analogous to replacing the (usually lost) type strain of a (named) species by a neotype strain for isolated organisms. Related to the latter, it is important to note that close to 60% (15 out of 27) of requests for an opinion addressed to the Judicial Commission of the International Committee for Systematics of Prokaryotes since 2007 is dealing with the rejection of names or the establishment of a neotype strain due to the lack of an authenticated living culture as type material. Cases like the wrong isolate was deposited by the authors (Pukall *et al*., 2008), lack of depositing to culture collections and/or loss of the original culture (Podkopaeva *et al*., 2009), distributed isolates do not match the nomenclatural type (Oggerin *et al*., 2011), loss of a culture deposit (Duncan and Flint, 2008) or deposits of contaminated isolates (Urdiain *et al*., 2008) are the causes for these requests. We argue that if the genome sequence was to serve as alternative type material, the link to the authenticity of the nomenclature type would not have been lost and several of the problems mentioned above with cultures would not apply (such as losing the culture or culture viability over time). Replacing versions of genome sequence can be digitalized as part of the major public genome databases and thus, would be easier to manage and update compared with living cultures as well. How much ecological or phenotypic information can be routinely provided for a yet-uncultivated species is of secondary importance for prokaryotic taxonomy, but there is little doubt that such information can be retrieved as well, e.g. by single-cell methods.

Overall, we strongly believe that the advantages of adopting genome sequence as alternative type material for uncultivated and fastidious taxa far outweigh potential threats. We suggest to form an expert committee discussing and organizing the steps required for unifying the taxonomy of cultured and uncultured microorganisms, and assuring that this is done without compromising the high quality standards and stability of prokaryotic taxonomy.

References

Andersen, M.H., McIlroy, S.J., Nierychlo, M., Nielsen, P.H., and Albertsen, M. (2019) Genomic insights into Candidatus Amarolinea aalborgensis gen. nov., sp. nov., associated with settleability problems in wastewater treatment plants. *Syst Appl Microbiol* **42**: 77–84.

Anton, J., Oren, A., Benlloch, S., Rodriguez-Valera, F., Amann, R., and Rossello-Mora, R. (2002) Salinibacter ruber gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. *Int J Syst Evol Microbiol* **52**: 485–491.

Anton, J., Rossello-Mora, R., Rodriguez-Valera, F., and Amann, R. (2000) Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Appl Environ Microbiol* **66**: 3052–3057.

Bisgaard, M., Christensen, H., Clermont, D., Dijkshoorn, L., Janda, J.M., Moore, E.R.B., *et al*. (2019) The use of genomic DNA sequences as type material for valid publication of bacterial species names will have severe implications for clinical microbiology and related disciplines. *Diagn Microbiol Infect Dis* **95**: 102–103.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., *et al*. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**: 725–731.

Duncan, S.H., and Flint, H.J. (2008) Proposal of a neotype strain (A1-86) for *Eubacterium rectale*. Request for an opinion. *Int J Syst Evol Microbiol* **58**: 1735–1736.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., *et al*. (2008) The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* **26**: 541–547.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.

Harbison, A.B., Carson, M.A., Lamit, L.J., Basiliko, N., and Brauer, S.L. (2016) A novel isolate and widespread abundance of the candidate alphaproteobacterial order (Ellin 329), in southern Appalachian peatlands. *FEMS Microbiol Lett* **363**: 1–8.

Hedlund, B.P., Dodsworth, J.A., and Staley, J.T. (2015) The changing landscape of microbial biodiversity exploration and its implications for systematics. *Syst Appl Microbiol* **38**: 231–236.

Henson, M.W., Lanclos, V.C., Faircloth, B.C., and Thrash, J. C. (2018) Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J* **12**: 1846–1860.

Karthikeyan, S., Rodriguez, R.L., Heritier-Robbins, P., Kim, M., Overholt, W.A., Gaby, J.C., *et al*. (2019) "Candidatus Macondimonas diazotrophica", a novel gammaproteobacterial genus dominating crude-oil-contaminated coastal sediments. *ISME J* **13**: 2129–2134.

Keller, M., and Zengler, K. (2004) Tapping into microbial diversity. *Nat Rev Microbiol* **2**: 141–150.

Konstantinidis, K.T., and Rossello-Mora, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomic data in the Candidatus proposal. *Syst Appl Microbiol* **38**: 223–230.

Konstantinidis, K.T., Rossello-Mora, R., and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* **11**: 2399–2406.

Konstantinidis, K.T., Rossello-Mora, R., and Amann, R. (2018) Reply to the commentary "uncultivated microbes-in need of their own nomenclature?". *ISME J* **12**: 653–654.

Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**: 2567–2572.

Lawrence, J.G., and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**: 9413–9417.

Lee, J., Kwon, K.K., Lim, S.I., Song, J., Choi, A.R., Yang, S. H., *et al*. (2019) Isolation, cultivation, and genome analysis of proteorhodopsin-containing SAR116-clade strain Candidatus Puniceispirillum marinum IMCC1322. *J Microbiol* **57**: 676–687.

Oggerin, M., Rubio, V., Marin, I., and Arahal, D.R. (2011) The status of the species Beijerinckia fluminensis Dobereiner and Ruschel 1958. Request for an opinion. *Int J Syst Evol Microbiol* **61**: 1757–1759.

Oren, A., and Garrity, G. (2018) Uncultivated microbes – in need of their own nomenclature? *ISME J* **12**: 309–311.

Overmann, J., Huang, S., Nubel, U., Hahnke, R.L., and Tindall, B.J. (2019) Relevance of phenotypic information for the taxonomy of not-yet-cultured microorganisms. *Syst Appl Microbiol* **42**: 22–29.

Pandit, P.S., and Rahalkar, M.C. (2019) Renaming of 'Candidatus Methylocucumis oryzae' as Methylocucumis oryzae gen. nov., sp. nov., a novel type I methanotroph isolated from India. *Antonie Van Leeuwenhoek* **112**: 955–959.

Parker, C.T., Tindall, B.J., and Garrity, G.M. (2019) International code of nomenclature of prokaryotes. *Int J Syst Evol Microbiol* **69**: S1–S111.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Pena-Gonzalez, A., Soto-Giron, M.J., Smith, S., Sistrunk, J., Montero, L., Paez, M., *et al*. (2019) Metagenomic signatures of gut infections caused by different *Escherichia coli* Pathotypes. *Appl Environ Microbiol* **85**: pii:e01820-19.

Podkopaeva, D., Grabovich, M., Kuever, J., Lysenko, A.M., Tourova, T.P., Kolganova, T.V., and Dubinina, G. (2009) Proposal of Spirillum winogradskyi sp. nov., a novel microaerophilic species, an emended description of the genus Spirillum and request for an opinion regarding the status of the species Spirillum volutans Ehrenberg 1832. *Int J Syst Evol Microbiol* **59**: 2916–2920.

Pukall, R., Schumann, P., Clermont, D., and Bizet, C. (2008) Bacillus aeolius DSM 15084T (=CIP 107628T) is a strain of Bacillus licheniformis. *Int J Syst Evol Microbiol* **58**: 1268–1270.

Rodriguez, R.L., Gunturu, S., Harvey, W.T., Rossello-Mora, R., Tiedje, J.M., Cole, J.R., and Konstantinidis, K.T. (2018) The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* **46**: W282–W288.

Rodriguez-R, L.-M., and Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints*. e1900v1.

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Droge, J., *et al*. (2017) Critical assessment of metagenome interpretation-a benchmark of metagenomics software. *Nat Methods* **14**: 1063–1071.

Stott, M.B., Crowe, M.A., Mountain, B.W., Smirnova, A.V., Hou, S., Alam, M., and Dunfield, P.F. (2008) Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environ Microbiol* **10**: 2030–2041.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005) Genome-directed isolation of the key nitrogen fixer Leptospirillum ferrodiazotrophum sp. nov. from an acidophilic microbial community. *Appl Environ Microbiol* **71**: 6319–6324.

Urdiain, M., Lopez-Lopez, A., Gonzalo, C., Busse, H.J., Langer, S., Kampfer, P., and Rossello-Mora, R. (2008) Reclassification of Rhodobium marinum and Rhodobium pfennigii as Afifella marina gen. nov. comb. nov. and Afifella pfennigii comb. nov., a new genus of photoheterotrophic Alphaproteobacteria and emended descriptions of Rhodobium, Rhodobium orientis and Rhodobium gokarnense. *Syst Appl Microbiol* **31**: 339–351.

Whitman, W.B. (2015) Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Syst Appl Microbiol* **38**: 217–222.

Whitman, W.B. (2016) Modest proposals to expand the type material for naming of prokaryotes. *Int J Syst Evol Microbiol* **66**: 2108–2112.

Whitman, W.B., Sutcliffe, I.C., and Rossello-Mora, R. (2019) Proposal for changes in the international code of nomenclature of prokaryotes: granting priority to Candidatus names. *Int J Syst Evol Microbiol* **69**: 2174–2175.

Zhang, A.N., Mao, Y., Wang, Y., and Zhang, T. (2019) Mining traits for the enrichment and isolation of not-yet-cultured populations. *Microbiome* **7**: 96.