Comment received 20th February, posted February 25th.

Dear colleagues,

I´m still following the discussion related to the use of DNA as type material for bacterial species descriptions and I have some additional remarks on the role of type strains.

The role of the deposition of type strains.

I now have worked over 30 years in the field of taxonomy of Actinomycetes especially members of the genus Streptomyces. If you go back to the golden area of antibiotics, many novel species were described and as it was not necessary to deposit them in an open collection, so in many cases every new antibiotics producer was described as a novel species. Basing on the huge number of species within this genus until today not all the taxonomic positions of these different species has been clarified. The members of the genus Streptomyces have also very large genomes with a size of 8 to 10.000 kb. There are many reports about the horizontal gene transfer within this genus so also within one species the different isolates show differences in their genomes. If the species will only be defined by their genome than we will came back to a similar situation as we had during the use of antibiotic production as the only taxonomic marker.

PD Dr. Joachim Wink (Working Group Microbial Strain Collection, Helmholtz Centre for Infection Research, Braunschweig, Germany

>>

Comment received February21st, posted February 25th

Comment 1

I would like to provide some comments regarding the proposed changes to the ICNP. Please add my statement below to the official record being circulated by E-mail. Additionally, please add me to all other E-mails circulated on the subject. Thank you.

The proposed changes to the ICNP recommend that sequences of DNA may serve as type material if it “unambiguously identifies” the taxon.

Reliance upon the phrase “unambiguously identify” is shortsighted and willfully disregards the progressive revisionism that is key to science. Decades ago, Bacillus anthracis could be “unambiguously” differentiated from near neighbors by its pathogenicity, i.e. its ability to cause the disease anthrax. This definition ultimately proved incorrect, as we now know that the virulence genes are plasmid-borne, and that B. cereus strains with these genes can cause essentially the same disease. Similarly, a gene or set of genes can be thought to be specific to a certain taxon and incorrectly used as definitive identification of that taxon. One such example is the botulinum neurotoxin (BoNT) originally thought to be specific to and therefore indicative of Clostridium botulinum. Since then, BoNT genes or homologs have been found in C. argentinense, C. baratii, C. butyricum, Weisella oryzae, Chryseobacterium piperi, and Enterococcus faecium.[1] Genes or sets of genes can “unambiguously identify” a taxon... until they no longer do.

If these genes are mere identification criteria published in a standard (non-IJSEM) journal, they can simply be a) nullified by a subsequent publication proving the genes do not unambiguously identify the taxon, or b) supplanted by a newer set of genes with better specificity. However, if these genes are officially designated as type sequences, it is unclear how a retraction of type status would occur. In case a), it would seem destined for referral to the Judicial Commission in order to allow a type sequence to be "undesignated". In case b), the proposed changes to Rule 18f allow for replacement of a sequence of genomic DNA with later cultivated type strain; however, they do not allow for replacement of a type sequence with a different sequence (or more generally, "material"). If sequence is to be allowed to serve as type, a protocol must exist for the inevitable situation where sequence must be replaced with sequence. An additional proposal to modify the ICNP should be made to this effect. At the very least, the current proposals should be tabled until such time as a coherent implementation can be evaluated in toto.

[1]. Poulain B and Popoff MR. Why Are Botulinum Neurotoxin-Producing Bacteria So Diverse and Botulinum Neurotoxins So Toxic? *Toxins (Basel)* **11**, doi:10.3390/toxins11010034 (2019).

Comment 2

I am submitting a second response on a separate issue I have with the proposed changes.

It is universally acknowledged that science faces a reproducibility crisis. The proposed changes to the ICNP threaten to exacerbate this crisis.

An essential foundation of prokaryotic taxonomy is the availability of type strains to the entire scientific community. Currently, type strains of novel taxa must be deposited in two culture collections in different countries [Rule 30(3)(b)].[1] This ensures that scientists around the world can order the same strain, reproduce experiments, verify results, and build upon the science relating to this organism.

By allowing valid publication of taxa without making a viable culture available, researchers will be unable to reproduce research related to this organism. If the type sequences proposal is accepted, yes, researchers will be able to download the type sequence and analyze it. However, because the input sequence will be identical, the results will almost certainly be identical as well. The ANI results I generate on my computer will not differ from those anyone else generates. This is not proper scientific reproducibility; this is simply running the same thing multiple times.

Under the current system, criteria subsequently found to be insufficient or ambiguous (as addressed in my previous response) can be ameliorated by returning to the preserved type strain and determining new criteria. However, this will not be possible if non-biological criteria (e.g., sequences) are accepted as type. This is the primary advantage of the current culture-based system. The type strain is the definition of the species; it is a *specific organism* that is the taxonomic reference point. As new technologies are developed, scientists can return to the type culture to reexamine it using the latest techniques. Thus, the current system allows the taxonomy to evolve and adapt to the future. On the other hand, the type sequence (or other material) will exist as a fixed snapshot in time. As new criteria or novel technologies develop (e.g., a metabolome, or some as yet undiscovered [futuretech]-ome), one cannot return to the type DNA sequence to identify new

criteria under the new system (with the possible exception of extracting a hypothetical proteome via translation of the gene sequences). The DNA sequence will always be the DNA sequence. Thus, despite the comments expressing the proposals as bringing the nomenclatural system into the future, dissociating nomenclatural types from viable cultures would in fact have the opposite effect.

In order to preserve the adaptability of our systematic scheme, nomenclatural types should continue to be viable cultures, as currently required. The proposals under consideration should be rejected.

[1] Parker CT, Tindall BJ, and Garrity GM. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* **69**, S1-S111, doi:10.1099/ijsem.0.000778 (2019).

Marco Riojas (ATCC, USA)

>>

Comment posted February 25th

Dear Colleagues,

I try to address here the specific proposal for one of the proposed modifications (underlined), i.e., addition of a third clause (3) to Rule 18a:

"(3) As from 1 April 2020*, sequences of genomic DNA may also serve as the type when it unambiguously identifies the species.

When possible, it should be a high quality draft or better genome sequence".

While most discussion has concerned the usefulness of WGS data for characterising and identifying bacteria, the purpose of rule 18a of The Code has been to define what should be the reference and Rule 30(3-b) insures (since 2001) that the references serving taxonomy and systematics are available to the scientific community.

I emphasise that WGS data are not the reference of a taxon – they are the result of an analysis of the reference – as already pointed out by Ulrich Nübel (Feb. 07).

As such, any particular WGS data are dependent upon varying factors – none of which have been defined by proponents of the new rules.

So, it is important to try to address this issue of the proposed Rule 18a(3), i.e., what should serve as nomenclatural type material as the 'ultimate' reference for prokaryote taxa.

More specifically, the implications of implementing the new rule 18a(3), as Markus Göker (Feb 08) stressed.

As Nübel (Feb 07) pointed out, the reproducibility of analyses for proposing and validly publishing new taxonomic names cannot be insured with WGS electronic data.

Yet, reproducibility of analyses is essential (required?) for reliable science, including reliable taxonomy.

Being able to reproduce the analyses of research is generally accepted as essential for publication.

This goes to the crux of the argument in considering Rule 18a(3), regardless of any issues of what should be done for characterising taxa.

And, this has not been addressed by proponents of the proposed new rules.

Furthermore, it was not addressed in 2 ICSP meetings (2017 and 2019), although I and others raised this question.

Do proponents of the new rules not believe that it is necessary to be able to reproduce the characterisations of bacterial taxa?

Do proponents of the new rules not believe that it is necessary to safe-guard the reference material for bacterial taxa?

The so-called, 'chain-of-custody' of the WGS data cannot be confirmed, beyond the expertise and the word of the depositor of the WGS data into a public database.

Given the overall levels of 'crap' genomic data in the public databases, I submit that such trust would not be sensible.

It would be good to receive a discussion from any of the proponents for changing the rules about how you see these issues.

I think these points have been somewhat lost in the discussions about how bacteria should be analysed.

Btw – Noone, I think, suggests that WGS data does not provide for more comprehensive analyses.

Another issue, another mail.

Thank you for your consideration.

Edward Moore, PhD
CCUG - Avd. Klinisk Bakteriologi
Sahlgrenska Universitetssjukhuset
Göteborgs Universitet
Tel: 031-342-4696
e-mail: erbmoore@ccug.se

>>

Comment posted 25th February.

With respect to the proposed change of the Code of Nomenclature of Prokaryotes, I would like to add the following suggestion (whatever the result of the proposed changes):

When defining 'the sequence of one or more genes that unambiguously identifies the genus or subgenus' a minimal set of NOMENCLATURE-SPECIFIC METADATA' must be included with the sequence according to the following principles:

'FAIRDATA-2'

Findability (where has the sequence been deposited?)

Authentication (when, where and by which department and institution the sample was taken, isolated and sequenced?)

Interoperable (is the sequence derived from an OTU or from whole genome/gene sequencing ?)

Reusable (is their clinical material or cultured bacterial isolate available for further studies?)

Depth (what is the minimum sequencing depth for the published sequence?)'

Association (what is the current most closely related genus, species or subspecies related to the new sequence ? – Include information on how this was determined)

Technology (which manufacturer and sequencing technology was used and which version?)

Algorithm (which software package and version was used to obtain the sequence?)

-2 (the sequence has been confirmed from a minimum of 2 different independent sample sources and/or scientific institutions).

A letter could be added to the name or strain identifier to indicate that FAIRDATA-2 information is available and would act as a potential marker of quality of the sequence.

Dr. John Hays, *Associate Professor* Medical Microbiology & Infectious Diseases, Erasmus University Medical Center Rotterdam

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>