

Towards a unifying system for the naming of cultured and uncultured prokaryotes

Most prokaryotes cannot easily be grown in the laboratory and distributed as pure cultures. Thus, these organisms could not be officially named. A code of nomenclature – the SeqCode – provides paths to name such organisms on the basis of genomic data, aiming to unify field and laboratory studies in microbiology.

This is a summary of:

Hedlund, B. P. et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-022-01214-9> (2022).

Published online:

Published online: 22 September 2022

Publisher's note

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

The problem

The rules for naming prokaryotes require deposition of nomenclatural type strains in culture collections¹. Nomenclatural types or simply 'types' are reference samples of taxa to which unidentified samples are compared in order to determine whether they should be classified in the same taxa and have the same names. Types are fundamental in systematics (the classification and nomenclature of organisms). For strains to serve as types, they must grow routinely in the laboratory and be available for distribution to the scientific community. This system works well for taxonomic studies and other purposes; however, it is challenging or impossible for fastidious microorganisms that require specialized growth conditions. Moreover, when methods were developed to sequence prokaryotic genomes directly from DNA isolated from the environment, it became clear that most of the prokaryotic diversity had never been grown in the laboratory. Thus, it is necessary to develop a system for naming these uncultured microorganisms as well^{2,3}.

The solution

In 2018–2021, we organized a series of international workshops to discuss the status quo in the nomenclature of prokaryotes and explore paths forward. The community overwhelmingly supported the development of an alternative code of nomenclature that would accept DNA sequence data as types. Like cultured type strains, genome sequences can serve as references to unambiguously identify cultured species, but they are easier and cheaper to store and distribute than cultured type reference strains, and they also enable expansion of the identification and naming efforts into the vast diversity of uncultivated prokaryotes. The community also strongly supported the idea of an online registry that would capture taxonomic names and associated metadata to serve as a resource for research.

In response, we developed the Code of Nomenclature of Prokaryotes Described from Sequence Data – informally called the SeqCode – along with genome quality standards that are required for a genome to serve as a nomenclatural type. This project was a logical progression of current practice because genome sequences have already been widely used in prokaryotic systematics, and hundreds of thousands of genome sequences are already available in public databases. Moreover, the SeqCode rules for generating names and creating hierarchical taxonomies are similar to those previously used for cultured type strains, enabling

integration of the existing names into a unified taxonomy incorporating both strains and genomes as types.

In parallel, the SeqCode Registry was developed to register new taxonomic names. Currently, two paths are available (Fig. 1). The recommended path, Path 1, involves the pre-registration of taxonomic names and metadata before publication to enable automated checks and curator input to guide researchers during the naming process. For example, sequences are downloaded from databases and checked for quality; curators examine the name and etymology for grammatical correctness. Path 2 enables the registration of existing names, including *Candidatus* names (that is, temporary names that are outside the normal rules of nomenclature, usually because the bacteria are uncultured). This path is deliberately flexible because all crucial information is captured and available within the Registry. A third path is also anticipated, in which the publication and registration would occur simultaneously.

Future directions

The SeqCode and SeqCode Registry provide paths to communicate about the vast diversity of prokaryotes. Combined with ever-improving technologies to assemble genomes from uncultivated prokaryotes – namely metagenomics and single-cell genomics – and robust resources to guide decisions on taxonomy^{4,5}, the key pieces are in place for an expansion in the next decade of prokaryotic systematics to the entire prokaryotic tree of life. This expansion will serve the community by providing a common naming system for all prokaryotes that is well ordered and supported by data-rich genomic datasets and associated metadata.

The success of the SeqCode will depend on community engagement. The SeqCode needs to be routinely and widely used, and, in parallel, a system of governance with statutes, committees and membership must be developed. Further developments in the SeqCode Registry will facilitate this process and should focus on streamlining registration, expanding existing automated checks, and improving interoperability with informatics and biodiversity datasets. Finally, a legacy of *Candidatus* names and other informal names should be registered under the SeqCode to stabilize and order the existing taxonomic framework.

Brian P. Hedlund¹ and William B. Whitman²

¹School of Life Sciences, University of Nevada, Las Vegas, NV, USA. ²Department of Microbiology, University of Georgia, Athens, GA, USA.

EXPERT OPINION

"The SeqCode addresses an urgent need within the community of molecular microbial ecologists, caused by a lack of consistent naming for microbial species detected by sequencing. The long-term success of this tool will depend on how robust its supporting infrastructure is, along with

support from the community. Teething problems are probable, such as confusion due to inconsistency with the naming of cultured species, but this work is a good start in the right direction." **Thomas Clavel, University Hospital of RWTH Aachen, Aachen, Germany.**

FIGURE

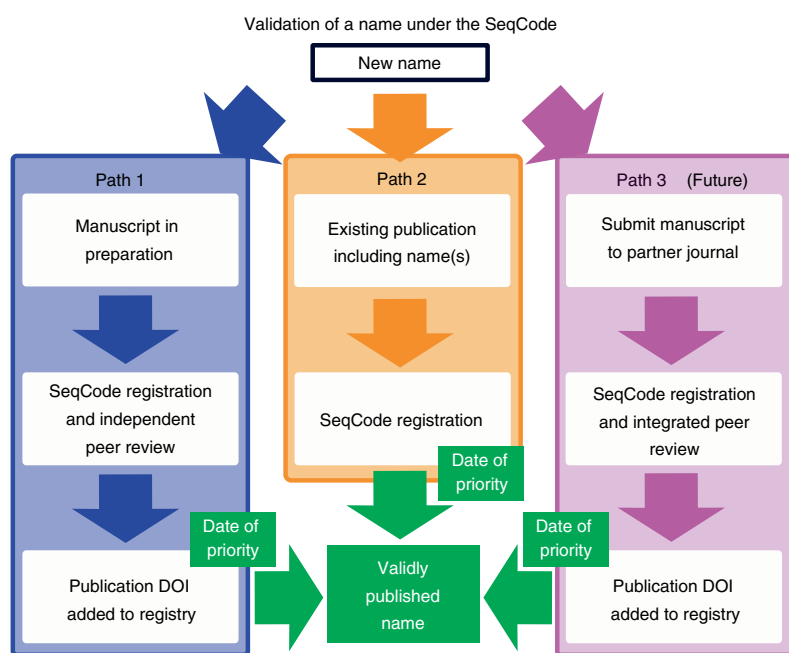


Fig. 1 | How to register prokaryote names with the SeqCode. In Path 1, names are pre-registered in the SeqCode Registry, enabling automated checks and curator review to guide the process. Names are validated once the digital object identifier (DOI) for the accepted publication is entered. In Path 2, existing names from the literature are registered and validated if the names and genome quality standards are compliant. In the future, Path 3, a system for coordinated registry and manuscript review, could be developed. © 2022, Hedlund, B. P. et al., [CC BY 4.0](#).

BEHIND THE PAPER

For B.P.H. and me, this project started in 2014, when Ramon Rossello-Mora (a co-author in the present study) invited us to contribute to a special issue of *Systematics and Applied Microbiology* to 'shake-up' prokaryotic systematics. We had both been editors of *Bergey's Manual of Systematics of Archaea and Bacteria* and were familiar with the current system of prokaryotic nomenclature. In the following years, metagenomics began generating large numbers of reliable genome sequences from uncultured organisms, and the need

to name these organisms was pointed out by co-authors Konstantinos T. Konstantinidis and Ramon Rossello-Mora, along with Professor Rudolf Amann. Afterwards, a number of researchers were eager to join the project. Importantly, B.P.H. and co-author Anna-Louise Reysenbach obtained funding from the US National Science Foundation to organize a series of workshops, including two large international online workshops in 2021. Overall, many people came to the table with energy and different perspectives and skills. **W.B.W.**

REFERENCES

- Whitman, W. B. Modest proposals to expand the type material for naming of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **66**, 2108–2012 (2016).
This paper proposes changes to the International Code of Nomenclature of Prokaryotes (ICNP) that would enable DNA sequence data to serve as nomenclatural types, in addition to strains.
- Konstantinidis, K. T., Rossello-Mora, R. & Amann, R. Uncultivated microbes in need of their own taxonomy. *ISME J.* **11**, 2399–2406 (2017).
This opinion article discusses the need for a system to name uncultivated prokaryotes to better communicate biodiversity.
- Murray, A. E. Roadmap for naming uncultivated archaea and bacteria. *Nat. Microbiol.* **5**, 987–994 (2020).
A consensus statement by 121 authors from 22 countries proposing either emendation of the ICNP or development of a new code that enables DNA sequence data to serve as nomenclatural types.
- Chaumeil, P.-A. et al. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
This manuscript discusses the Genome Taxonomy Database Toolkit (GTDB-Tk), an example of a modern system to classify prokaryotes on the basis of genomic data.
- Rodriguez-R, L. M. et al. 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 (2018).
A tool for the taxonomic classification of genome-assembled and metagenome-assembled genome sequences.

FROM THE EDITOR

"The resource presented by Hedlund et al. stands out because it is a community effort that provides a path forward for the naming of uncultivated bacteria and archaea, which has been an increasingly important issue for microbiologists. The use of the SeqCode and its platform to register names of previously unreported prokaryotes will hopefully promote best practices and enable consistency in nomenclature." **Emily White, Senior Editor, Nature Microbiology.**