COMMENTARY

Microbial taxonomy with DNA sequence data as nomenclatural type: How far should we go?

Li-Wei Zhou[1,*]

[1]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

*Correspondence: liwei_zhou1982@im.ac.cn

Received: May 20, 2023; Accepted: June 30, 2023; Published Online: July 28, 2023; https://doi.org/10.59717/j.xinn-life.2023.100017

Citation: Zhou L.W. (2023). Microbial taxonomy with DNA sequence data as nomenclatural type: How far should we go? The Innovation Life 1(2), 100017.

Taxonomy aims to recognize all living beings and understand their evolutionary relationships. The discipline of taxonomy using binomial nomenclature as a system for naming species is generally considered to start with Linnaeus's publication *Species Plantarum* in 1753. As the most basic discipline, taxonomy is of benefit to other disciplines via facilitating scientific communication; it also uses data from other disciplines such as morphology, anatomy, biochemistry, physiology, and molecular biology as evidence to delimit the circumscription of taxa. The data from these disciplines provides different weighted evidence as technology has advanced over the past 270 years. At the moment, whatever discipline the data comes from, the nomenclatural type of a species or infraspecific taxon is required to be a single specimen, a metabolically inactive culture or an illustration, namely a real material that is permanently attached to the name of the taxon. However, recently, this requirement of nomenclatural type has been challenged in the taxonomy of microorganisms. High-throughput sequencing technology and bioinformatics tools have revealed innumerable microorganisms that may possess crucial ecological functions but are unculturable under the current methodology from various environments.[1] Due to a lack of real materials, these microorganisms cannot currently be formally named under the framework of any classical code of nomenclature, which thus blocks the convenient scientific communication that is the fundamental purpose of taxonomy.

In the last ten years, many microbiologists tried to emend codes to permit DNA sequence data as a category of nomenclatural type. However, even if DNA sequences have been commonly used as the most powerful diagnostic features, this kind of proposals has not been accepted by the codes of nomenclature related to microorganisms, viz. the International Code of Nomenclature of Prokaryotes (ICNP) and the International Code of Nomenclature for algae, fungi, and plants (ICNafp). After the final rejection of the proposal by the International Committee on Systematics of Prokaryotes, an independent code, viz. SeqCode was newly initiated for prokaryotes.[2] More than one hundred microbiologists from 22 countries and six continents advocate the new code. The main difference between SeqCode and ICNP is that the former code uses whole-genome sequence data as the nomenclatural type for prokaryotes regardless of cultivability. Two paths for validation of names under the SeqCode are available: one is for naming sequences in manuscripts in preparation, and the second is for naming sequences in existing publications. In future, the third possible path will be available by cooperating with potential partner journals. Whatever the path is, the core requirement is to deposit high-quality whole-genome sequences through the SeqCode Registry (https://disc-genomics.uibk.ac.at/seqcode/) as the nomenclatural type.

SeqCode definitely will facilitate scientific communication about innumer-

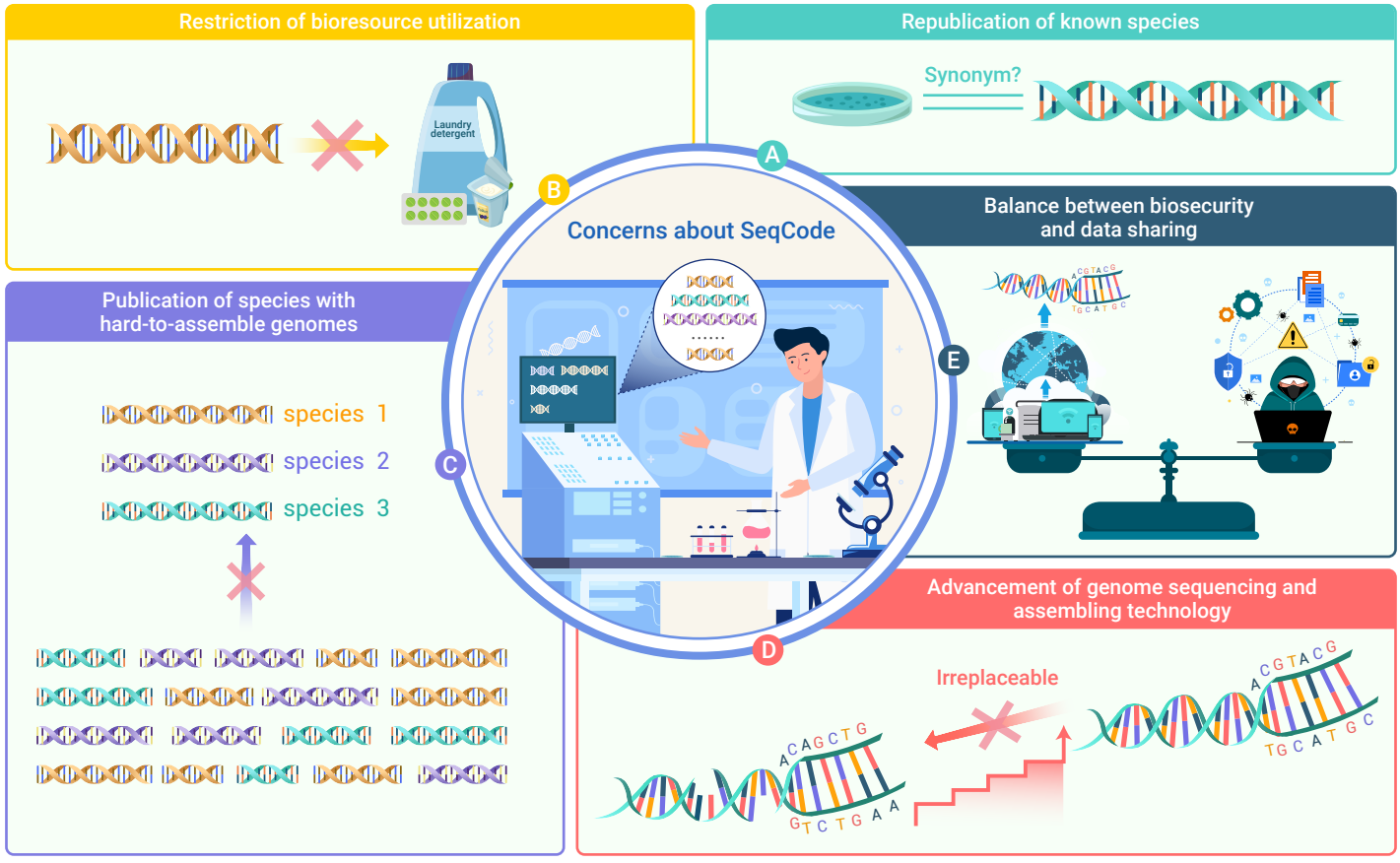

Figure 1. The concerns about SeqCode using whole-genome sequence data as the nomenclatural type for prokaryotes.

able unculturable prokaryotes. However, like other new things, some concerns come along with SeqCode (Figure 1), although it had been conceived for several years before being formally engaged. Then, how to deal with at least the five concerns shown below is crucial for further utilizing SeqCode in my opinion.

A. Republication of known species. As introduced above, taxonomic evidence comes from other disciplines that have changeable weights from time to time. Therefore, the evidence for microorganisms published in different times sometimes shows no consistency. For example, the current dominant taxonomic evidence of DNA sequences is lacking for most microorganisms previously published based on morphology, biochemistry and physiology during the first 200 years of taxonomy. This phenomenon may lead to the republication of known species that lack of available DNA sequences. In the case that this kind of prokaryotes has a large number, it is impossible to sequence them all at a high quality in a short period of time. So, how to judge whether the newly generated whole-genome sequences really come from previously unknown species rather than known but unsequenced species is a huge challenge.

B. Restriction of bioresource utilization. In classical microbiology, pure cultures are essential materials for studying microorganisms. Under the framework of ICNP, pure cultures are required to name prokaryotes, which facilitates efforts to isolate cultures and thus their utilization. From the progress of molecular sequencing technology, it is known that more than 80% of prokaryotes are unculturable according to the concept of phylogenetic species. However, with the advancement of culture omics, pure cultures have been successfully isolated from more and more previously unculturable microorganisms.[3] Therefore, some of the so-called unculturable microorganisms are not really unculturable but need to be cultured with special methodology: media and conditions. Although it is not against culture omics, SeqCode may, to some extent, impair the passion to isolate pure cultures, and thus indirectly restrict the bioresource utilization of prokaryotes.

C. Publication of species with hard-to-assemble genomes. It seems that the generation of whole-genome sequences is easier than isolating pure cultures. However, to generate high-quality genomes, the methodology of metagenome assembly from complex microbial communities is equally, if not more, important compared with genome sequencing technology. This methodology still needs to be extremely improved, especially for the generation of high-quality metagenome-assembled genomes from microorganisms with complex genome structures, e.g., those with a high proportion of repeated sequences and variable GC content. It is not really known how many percent of unculturable microorganisms have hard-to-assemble genomes. Therefore, the percent of prokaryotes that SeqCode help name may not be high enough to trade off extreme instability for a well-established code of nomenclature. Then, should the requirement of genome quality for naming prokaryotes decrease? Or, radically, should one more new code be initiated for naming unculturable species with hard-to-assemble genomes? There are no easy answers.

D. Advancement of genome sequencing and assembling technology. The current genome sequencing and assembling technology is considered to meet, at least partially, the requirement for naming unculturable prokaryotes with high-quality genomes. Yet, the related technology will definitely be improved. Then, in the near future, many more intact genomes may be generated for unculturable microorganisms. In that time, how to treat the currently named prokaryotes with relatively 'low-quality' genomes as nomenclatural types? The better way is to resequence and reassemble them, but the digital nomenclatural types cannot provide this possibility.

E. Balance between biosecurity and data sharing. Data sharing is a crucial way to facilitate scientific communication and progress. Meanwhile, according to the Nagoya Protocol, the fair and equitable sharing of benefits should be guaranteed for the party providing genetic resources. Unlike the specimens or metabolically inactive cultures, the genome sequence data itself of microorganisms carries critical genetic information. After the publication of new prokaryotes under the framework of SeqCode, their whole-genome sequences will be freely accessible to anyone. It is reasonably postulated that the deep mining of innumerable genome sequence data can result in huge knowledge for commercial utilization related to these prokaryotes and their hosts. Then, it will be hard, if possible, to guarantee the benefits of microbiologists publishing new prokaryotes. Beyond the economic value, some prokaryotes that come from special environments, like the human gut and oral cavity, may be directly related to national biosecurity. So, how to balance biosecurity and data sharing is one of the most critical concerns about long-term running SeqCode.

Compared with prokaryotes, the debate as to whether accept DNA sequence data as a category of nomenclatural type for fungi or not is ongoing.[4] Nevertheless, in the official journal of the International Mycological Association, viz. *IMA Fungus*, *Archaeorhizomyces secundus* and *A. victor* have been newly introduced with rDNA sequences as types.[5] Although these two names are invalid, their authorities are indexed permanently according to the current version of ICNafp. Moreover, some mycologists began to talk about how to proceed DNA sequence data as nomenclature types, instead of discussing if it should be or not. The next round of voting on this article may be at the 12th IMC in 2024 that was postponed in 2022 because of Covid-19. Let us look forward to the voting result, and especially to what will happen if the vote goes against using DNA sequence data as a category of nomenclatural type, i.e., whether a fungal version of 'SeqCode' will be initiated. Noteworthily, fungi as eukaryotic microorganisms have much more complex and larger genomes than prokaryotes. The quality standard of whole-genome sequences required for naming new prokaryotes under the framework of SeqCode may be not suitable for fungi in the foreseeable future.

In summary, from my point of view, the old taxonomy needs to embrace new technology as it has over the ages. However, it should find a suitable methodology to face the emendation of codes and adjust the method continuously, especially when the emendation will challenge the taxonomic foundation that has run for 270 years.

REFERENCES

1. Jiao, J.Y., Liu, L., Hua, Z.S., et al. (2021). Microbial dark matter coming to light: Challenges and opportunities. Natl. Sci. Rev. **8**: nwaa280. DOI: 10.1093/nsr/nwaa280.
2. Hedlund, B.P., Chuvochina, M., Hugenholtz, P., et al. (2022). SeqCode: A nomenclatural code for prokaryotes described from sequence data. Nat. Microbiol. **7**: 1702–1708. DOI: 10.1038/s41564-022-01214-9.
3. Lewis, W.H., Tahon, G., Geesink, P., et al. (2021) Innovations to culturing the uncultured microbial majority. Nat. Rev. Microbiol. **19**: 225–240.
4. Lücking, R., Aime, M.C., Robbertse, B., et al. (2021). Fungal taxonomy and sequence-based nomenclature. Nat. Microbiol. **6**: 540–548. DOI: 10.1038/s41564-021-00888-x.
5. Khan, F.K., Kluting, K., Tångrot, J., et al. (2020). Naming the untouchable – environmental sequences and niche partitioning as taxonomical evidence in fungi. IMA Fungus **11**: 23. DOI: 10.1186/s43008-020-00045-9.

FUNDING AND ACKNOWLEDGMENTS

Ms. Xue-Wei Sherry Wang (Institute of Microbiology, Chinese Academy of Sciences) is thanked for her assistance in the preparation of Figure 1. The research was financed by the National Key Research and Development Program of China (No. 2022YFC2601200). The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

DECLARATION OF INTERESTS

The author declares no competing interests.