

## Toward DNA-based taxonomy of prokaryotes and microeukaryotes

Trends in Genetics

Tedersoo, Leho; Geisen, Stefan; Chang, Ying; Nilsson, Henrik

<https://doi.org/10.1016/j.tig.2025.07.009>

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.


This publication is distributed using the principles as determined in the Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. According to these principles research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact [openaccess.library@wur.nl](mailto:openaccess.library@wur.nl)

## Opinion

## Toward DNA-based taxonomy of prokaryotes and microeukaryotes

Leho Tedersoo <sup>1,2,\*</sup>, Stefan Geisen<sup>3</sup>, Ying Chang<sup>4</sup>, and R. Henrik Nilsson<sup>5</sup>

The current nomenclatural rules regulating the naming of microorganisms are too conservative from the perspective of recent developments in molecular genetics tools and organism discovery. The taxonomy of microorganisms would greatly benefit from a conceptual shift toward DNA-based approaches. Current informal practices of DNA-based taxonomy include the use of DNA sequences for species description and type material. Here, we analyze the pros and cons of DNA-based taxonomic approaches and propose guidelines and examples for their appropriate use in descriptions of species and higher-ranking taxa. To facilitate taxon description and communication, we call for a broader use of DNA samples, genome and genetic marker sequences in typifying and diagnosing species and higher-ranking taxa if physical voucher strains or specimens are unavailable.

## Rules of nomenclature and classification

Biologists use **classification** (see [Glossary](#)) and **nomenclature** to structure and communicate species and higher taxonomic levels of organisms from genera to kingdoms and beyond. The principal rules of naming and typification of taxa are established in the International Code of Nomenclature (ICN) for algae, fungi, and plants (ICNAFP); ICN for animals (ICZN); ICN for prokaryotes (ICNP); and ICN for viruses (ICNV) [1]. Because of historical legacies, Cyanobacteria as well as photosynthetic and fungus-like protists are treated within ICNAFP, whereas heterotrophic protists and microsporidians (i.e., endoparasitic fungi) are treated by ICZN [1]. Independent of ICNs, the PhyloCode regulates the description and naming of phylogenetic clades of all organisms above the species level [2]. PhyloCode names groups strictly on the basis of evolutionary clades, ensuring more biologically meaningful, monophyletic classifications. PhyloCode allows taxa to be defined on the basis of shared ancestry rather than morphological characteristics, improving classification accuracy. These principles and benefits of the PhyloCode, particularly the definition of higher-level taxa based on least inclusive clades, have yet to be adopted in the ICNs.

Although the ICNs have maintained a transparent and strictly regulated naming system over decades, many biologists argue that these are too conservative in light of the opportunities provided by novel and constantly developing molecular genetics tools, impeding the naming of microorganisms that are unculturable or cryptic. This nomenclatural bottleneck inhibits taxonomic communication and classification of species and higher-ranking taxa, with far-reaching negative ramifications in legislation and biological conservation [3,4]. Given the great importance of DNA sequence data in modern biology, including biodiversity monitoring, phylogenetics, and species recognition, there is an increasing need to use DNA features in various aspects of **taxonomy** and the formalization of taxa, and regulate them by ICNs accordingly. A focus on DNA sequence data is particularly relevant for prokaryotic and eukaryotic microorganisms that are commonly known exclusively from their DNA sequences [3,5].

## Highlights

DNA-based taxonomy opens new avenues for describing and communicating organisms that cannot be named appropriately on the basis of classical taxonomic approaches because of a lack of distinguishing morphological characters and establishment in culture.

DNA-based and sequence-based types have lower physical space requirements and costs of storing and exchanging than physical organism specimens and living cultures, and they partly avoid the impediments of the Nagoya protocol.

DNA sequence-based taxonomic diagnoses are practical, because they do not require detailed images, time-consuming morphological or physiological descriptions, and visits to repositories in other countries to study physical types.

DNA-based taxonomy facilitates molecular biodiversity monitoring, particularly quality-filtering of sequence data and taxonomic and functional assignments.

We provide recommendations for the use of DNA-based taxonomy and propose changes to the international codes of nomenclature to formalize DNA-based typification and diagnoses.

Description of *Paranuclearia slavaukraini* sp. nov. (*Paranucleariidae* fam. nov.) in nucleariid amoebae serves as an example of the use of DNA-based taxonomy.

<sup>1</sup>Mycology and Microbiology Center, University of Tartu, Liivi 2, Tartu 50400, Estonia

<sup>2</sup>Department of Zoology, College of Science, King Saud University, PO Box 2455, Riyadh 11451, Saudi Arabia

<sup>3</sup>Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, Wageningen 6708PB, The Netherlands

Research toward DNA-based taxonomy has been pursued independently in bacteriology [6], zoology [7], botany [8], and mycology [9]. Each discipline has been developing approaches tailored to its unique challenges and priorities. Therefore, the meaning of the DNA-based taxonomy term varies by discipline. Its applications range from the use of genomes and nucleotide signatures in genetic markers for **species diagnosis** [6,10,11] through **alphanumeric coding** of species (e.g., SH0240158.10FU corresponds to *Saccharomyces cerevisiae* [12]) to the use of DNA samples or DNA sequences as **type material** [7–9,13]. In its most strict and contentious interpretation, DNA-based taxonomy refers to the use of DNA sequences in typification under prescribed circumstances.

Over decades, the ICNs have generally evolved along similar trajectories, with specific differences in the scientific terminology and requirements for typification, taxonomic descriptions, and registration of novel taxa [1,14,15]. Compared with other ICNs, ICNV handles changes in classification and has more strongly implemented DNA-based taxonomy, so all viruses, viroids, and satellites can be formally described on the basis of metagenomic and metatranscriptomic sequences without designating types [16]. In bacteriology, genome-level DNA-DNA hybridization assays to distinguish species were implemented in the early 1960s [11,17] and remained the principal approach until genome sequencing became a routine roughly a decade ago [18]. ICNP has remained conservative in regulating DNA-based nomenclature in spite of multiple attempts to implement sequence-based typification over three decades. To overcome these formal hurdles, the Candidatus system and SeqCode were established for prokaryotes (see below). ICNAFP and particularly ICZN primarily support researchers studying macroscopic organisms, where classification relies on observable traits. This focus makes them less adaptable to groups requiring molecular data, slowing progress in integrating genetic features. Mycologists have presented multiple proposals to allow sequence-based taxonomic typification to the ICNAFP, but these have received insufficient support [19–21].

In this overview, we use the term ‘DNA-based taxonomy’ broadly to cover its entire range of interpretations, both in nomenclature and classification. We provide an in-depth analysis of the pros and cons of various aspects of DNA-based taxonomy, covering the three principal codes of nomenclature for microorganisms – ICNAFP, ICZN, and ICNP – and offer recommendations for its use. Taken together, we propose a novel approach for DNA-based description and diagnosis of species and higher-ranking taxa, with examples in nuclearioid protists.

### Brief history of biological classification and nomenclature

Modern taxonomy of living organisms is rooted in Carl von Linné’s (Carolus Linnaeus) binomial taxonomic nomenclature that initially covered plants and fungi combined, plus animals. Although alternative fragmentary treatments existed in Ancient Greece, the Roman Empire, and Medieval Europe, von Linné’s system thrived because of its simplicity, use of Latin (the scientific and religious language in much of Europe at that time), and relatively high coverage of various organisms [1]. Unfortunately, von Linné’s system inadvertently pushed the classification of plant-like and animal-like organisms to evolve in their own ways, which has resulted in the accumulation of multiple homonyms and exceptional botanical islands within the zoological world (e.g., dinoflagellates) and vice versa (e.g., microsporidians).

The simple observational tools available at the Linnaean times facilitated species descriptions based on morphological characters visible to the naked eye and a low-magnification light microscope. These technological possibilities set deeply rooted standards for taxonomic description and identification for two centuries. Because many nineteenth-century researchers did not follow the Linnaean standards, botanists established an early set of rules for plant naming in the

<sup>4</sup>Department of Biological Sciences and Yale-NUS College, National University of Singapore, 16 College Avenue West, Singapore, 138527, Singapore

<sup>5</sup>Gothenburg Global Biodiversity Centre, Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden

\*Correspondence: [leho.tedersoo@ut.ee](mailto:leho.tedersoo@ut.ee) (L. Tedersoo).

International Botanical Congress (IBS) in 1867. IBSs refined the nomenclature rules, leading to the International Rules of Botanical Nomenclature in 1905 and Botanical ICN in 1935, which was renamed as ICNAFP in 2011 [2]. The first edition of the ICZN was published in 1905 [22]. Access to electron microscopes in the mid-twentieth century allowed for more detailed ultrastructural studies and led to the establishment of ICNP in 1947 and ICNV in 1966 [14]. The latest revisions of ICZN, ICNAFP, and ICNP were published in 1999, 2025, and 2025, respectively.

### Species concepts and species delimitation

There are multiple ways to recognize and distinguish species [23]. Traditionally, taxonomists use mainly the phenetic **species concept** for distinguishing and describing species of both macroorganisms and microbes. DNA-based taxonomy puts a stronger emphasis on the phylogenetic species concept, which advocates for species monophyly and/or coalescence of alleles [23]. The main benefits of the phylogenetic species concept include applicability to all organisms, recognition of cryptic species, and ease of use when DNA sequence data are available. The phylogenetic species concept is particularly popular in microbiology because of the lack of species-specific discriminatory phenotypic characters in many groups. Relatively easily obtainable genetic marker sequences can be used both in inferring phylogenies and in discriminating species.

The most commonly used genetic marker in prokaryotes is the nuclear **16S rRNA gene (SSU)**. To improve species- and strain-level resolution, the **rRNA internal transcribed spacer (ITS)** and **23S rRNA gene (LSU)**, as well as the protein-encoding RNA polymerase  $\beta$ -subunit (*rpoB*) and DNA gyrase subunit B (*gyrB*), are additionally used. In eukaryotes, the nuclear **18S rRNA gene (SSU)** and **28S rRNA gene (LSU)**, along with the intercalary ITS region, constitute the main genetic markers. For algae and non-photosynthetic protists, the plastid large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) gene and its introns and mitochondrial cytochrome oxidase I (*CO1*) may offer greater species-level resolution [24]. In fungi, the ITS region is commonly supplemented by the protein-encoding genes RNA polymerase II subunit 1 (*RPB1*), RNA polymerase II subunit 2 (*RPB2*), and translation elongation factor 1- $\alpha$  (*TEF1- $\alpha$* ) for better discrimination among closely related species [25]. Of these genetic markers, rRNA SSU, ITS, and LSU occur in tandem repeats that can be coamplified and sequenced using long-range PCR and long-read sequencing, which is particularly valuable for testing taxonomic hypotheses at low and high taxonomic levels simultaneously [26,27].

**Species delimitation** can be performed using various phylogenetic and sequence similarity-based tools [28,29]. It is most useful to assess the concordance of phylogenetic information from multiple unlinked genes for species discrimination [30], but this is not possible for eDNA metabarcoding studies. The General Mixed Yule Coalescent (GMYC) analysis determines species boundaries using both speciation and coalescence (within-species gene tree branching), offering flexibility in distance thresholds [31], but it requires a time-calibrated phylogenetic tree and is sensitive to the accuracy of divergence times estimates. The Bayesian Poisson Tree Processes (bPTP) method infers species on the basis of maximum likelihood phylogenies, incorporating Bayesian prior probabilities and indicating probabilistic support for alternative classifications [32], but it is particularly sensitive to sequence data and alignment quality. The Assemble Species by Automatic Partitioning (ASAP) approach delimits putative species on the basis of genetic distances in multiple sequence alignments and provides ranked partitions that allow selecting the best species delimitation scenario based on score values [33], but it is less sensitive due to not including evolutionary history and low flexibility with varying distance thresholds. All these methods are suitable for single-marker data, and GMYC and bPTP are also suitable for genomes. In bacterial genomes, the genomic distance-based measures Average Nucleotide Identity (ANI) and digital DNA-DNA hybridization (dDDH; i.e., comparison of genome fragment sequence

### Glossary

**16S rRNA gene (SSU):** gene encoding rRNA of ribosomal small subunit in prokaryotes as well as plastids and mitochondria of eukaryotes; the most widely used marker gene in prokaryote taxonomy.

**18S rRNA gene (SSU):** nuclear gene encoding rRNA of ribosomal small subunit in eukaryotes; the most widely used marker gene in certain protists.

**23S rRNA gene (LSU):** gene encoding rRNA of ribosomal large subunit in prokaryotes as well as plastids and mitochondria of eukaryotes; sometimes used for taxonomy.

**28S rRNA gene (LSU):** nuclear gene encoding rRNA of ribosomal large subunit in eukaryotes; commonly used for taxonomy in certain groups of fungi and protists.

**Alphanumeric coding:** labeling species and higher-ranking taxa with names composed of arbitrary letters and numbers (i.e., barcodes), which are machine-readable but not necessarily easy for humans to memorize.

**Barcoding gap:** measure of intraspecific variation relative to interspecific variation based on genetic marker sequence analysis; indicates how well species can be distinguished from each other based on DNA sequences.

**Classification:** process of organizing organisms into hierarchical taxonomic groups.

**Corpotype:** a term proposed for a physical specimen that is added post-description to the amendment of a DNA-based species description if there is no such physical specimen designated as type material in the original description; intended to serve as a specific epitype in DNA-based taxonomy to link DNA-based descriptions to morphophysiological features carried by the physical type specimen.

**Epitype:** a concept in ICNAFP depicting a specimen or illustration selected to serve as an interpretative type when the holotype and other name-bearing types associated with a validly published name are demonstrably ambiguous and cannot be critically identified for purposes of the precise application of the name to a species.

**Holotype:** the one specimen, illustration, or culture indicated by the original author as the nomenclatural type of a species. As long as the holotype exists, it fixes the application of the particular taxon name.

similarity) are more commonly used for species delimitation. The dDDH threshold values of 70% and ANI threshold values of 95%–96% are typically used to circumscribe bacterial species [28]. For yeasts, ANI thresholds ranging from 94% to 96% are suitable for delimiting species [34].

### DNA-based taxonomy: current practices

DNA sequence analysis reveals that microorganisms are enormously rich in species numbers and deep phylogenetic lineages, most of which have not been cultivated and observed using classical techniques. Many uncultivable organism lineages are broadly distributed and play critical roles in ecosystem functioning [35], necessitating their formal description and communication, which could only be reached using DNA-based taxonomy or radical improvements in our ability to cultivate biotrophic microorganisms. However, the codes of nomenclature for various organisms were established decades before the potential importance of DNA characters in taxonomy was recognized. Partly due to the historical legacy, the committees of ICNP, ICNAFP, and ICZN have been reluctant to integrate DNA and its features into the respective codes of nomenclature, although the use of DNA – either as characters or specimens – was proposed several decades ago [7,36]. The vague handling of DNA and other material samples has left ICNAFP and ICZN open to varying interpretation. For example, mycologists interpret the ICNAFP differently from botanists regarding the potential use of material samples, related DNA samples, and sequence illustrations as type materials [1,4,37].

Because culture-independent molecular identification methods were unable to provide living cultures for formal taxonomic description, bacteriologists implemented a Candidatus taxon system to name and characterize uncultured species and higher-ranking lineages that cannot be properly described on the basis of ICNP [6]. The Candidatus taxon system was intended to describe species provisionally in potentially novel genera for subsequent upgrades to fully ICNP-compliant species upon proper culture-based descriptions. The Candidatus species were initially required to be bundled with an SSU gene sequence of at least 1000 bases, some information about gross morphological, physiological, and ecological features, as well as a distinguishing nucleotide signature sequence for SSU. As a seminal contribution, Murray and Stackebrandt described four Candidatus species based on this approach [6], of which only one (*Lawsonia intracellularis* = Candidatus *intracellularis*) has been formally named since then. It took >30 years for the ICNP to accept and adopt the Candidatus taxon system, but the names of candidates still have lower priority than classical names [38]. However, so-called pro- validly described Candidatus names should be used if the newly described ICNP-compliant taxa represent the same taxon. According to the latest amendment to ICNP, all living strains, preserved specimens, genomes, and single gene sequences may constitute types of Candidatus species [38]. Using computer algorithms and Latin-compatible syllables, Pallen and colleagues [39] generated hundreds of thousands of artificial names that were randomly assigned to 65 000 Candidatus taxa in the Genome Taxonomy Database (<https://gtadb.ecogenomic.org/>). These artificial names (e.g., *Fedrifrarcha*) resemble the alphanumeric codes of UNITE **Species Hypotheses** and remain to be accepted by microbiologists. As of May 11, 2025, the List of Prokaryotic Names with Standing in Nomenclature included 211 Candidatus phyla, 1188 Candidatus genera and 2822 Candidatus species of prokaryotes (<https://lpsn.dsmz.de/>). Despite the notable advancement in recent years, the Candidatus taxon system was too restrictive for environmental microbiologists who adopt shotgun metagenomics approaches. This is because metagenomics (except single-cell methods) does not enable direct linking of sequences to specific morphological or physiological observations of individual cells. To address this limitation, an alternative nomenclatural code, the SeqCode, was established for uncultured prokaryotes [13,40]. SeqCode allows the nomination of a high-quality genome as type material to represent species. Such type genomes should be at least 90% complete and <5% contaminated, with recommended >75% complete SSU

**Internal Transcribed Spacer (ITS):** a transcribed, rapidly evolving marker linking SSU to LSU. In prokaryotes and eukaryotes, the ITS region contains tRNA gene and 5.8S rRNA gene, respectively; used for identification mostly in fungi, plants, and certain protists.

**Nomenclature:** internationally agreed systems regulating the naming and typification of taxa.

**Species concept:** principles of how to recognize species. The main species concepts are phenetic (morphology and ecophysiology), biological (mating barriers), and phylogenetic (common ancestor and monophyly).

**Species delimitation:** practical methods by which species are distinguished from each other; these methods depend on species concepts.

**Species diagnosis:** distinguishing species from closely related species on the basis of specific phenetic and/or molecular characters when describing new species.

**Species Hypothesis (SH) concept:** Species Hypotheses are formulated on the basis of stepwise changes in nucleotide sequence dissimilarity in the ITS marker at 0.5%–3% thresholds in a nested manner. The alphanumeric SH codes change in each UNITE database release but are linked to all previous versions to facilitate communication of species-level taxa.

**Taxonomy:** scientific field focused on describing, classifying, identifying, and communicating species and higher-ranking taxa.

**Type material:** the most important anchor for the species and its name; depending on code or treatment, it can be a living strain (in prokaryotes), genomic DNA sequence (viruses, SeqCode), voucher specimen, metabolically inactive culture, or illustration (other taxa). Ideally, the type should be available as a reference for comparison.

gene, >80% of transfer RNAs present, high genome integrity with the largest contig >100 kb, >10-fold sequence coverage and available raw and assembly data [13]. For each species described under SeqCode, the following information is recommended: etymology, pronunciation, physiological/ecological information, other metadata, and other genomic sequences. As of May 11, 2025, SeqCode has accumulated 850 taxon names of different ranks, including 436 species names ([www.seqco.de](http://www.seqco.de)). Despite its utility, SeqCode has received heavy criticism for unnecessarily duplicating the prokaryotic nomenclature code and generating confusion about the validity of taxon names [41]. Several instances of parallel taxonomy have emerged (e.g., [42]) because of simultaneous work on the same taxonomic group of organisms by several research teams, some of which rely on ICNP and some on SeqCode. These instances render subsequent merging of the two codes difficult due to the arising nomenclatural priority issues.

Mycologists and phycologists have applied DNA-based taxonomy at different levels, including species description, species delimitation, typification, and taxonomic diagnosis, although the ICNAFP does not mention the potential use of DNA for these purposes. In a pioneer study, Marin and colleagues determined diagnostic nucleotides in the SSU secondary structure for genera, families and orders of *Euglenophyceae* [10]. In 2012, Kirk presented a description of the chytrid fungus *Piromyces cryptodigmaticus*, referring to the least inclusive clade of three GenBank sequences as diagnosis and nominating a vouchered manure sample as type material [43]. An ascomycete species *Hawksworthiomyces sequentia* was described by nominating the least inclusive group of organisms that corresponded to two ITS sequences as diagnosis and including an illustration (i.e., a text string of a 494-base ITS sequence) of one of the GenBank sequences as type [44]. Lücking and Moncada described seven species of the endolichenic fungus *Lawreyomyces* gen. nov., using images of ITS sequences for illustrations and diagnostic nucleotide sequences for diagnoses [45]. Fifteen species of *Mucoromycota* were described on the basis of eDNA samples as **holotypes** and diagnostic nucleotide sequences in diagnoses, interpreting that DNA is a part of organisms therein [46]. All these species were dismissed by the ICNAFP committee, arguing that DNA samples and DNA sequences are not suited as type material and that DNA sequences do not conform to the requirements of an illustration, because they represent abstract letters rather than species-specific biological traits [4,19]. Following these arguments, *P. cryptodigmaticus* was validly redescribed on the basis of diagnostic nucleotide signatures and a fecal sample as a type specimen 10 years later [37], but the other aforementioned species remain invalid. Largely in parallel with these developments, the **Species Hypothesis (SH) concept** was released in the UNITE database in 2009 [12,47]. SHs enable communication of species-level taxa using flexible sequence dissimilarity thresholds and alphanumeric naming. Subsequently, a conceptually similar Taxon Hypothesis system was proposed for the alphanumeric coding of higher-ranking taxa to facilitate communication across present and future versions of fungal classification [48]. An alternative coding system was established for species complexes of the arbuscular mycorrhizal *Glomeromycota* based on the SSU V3-V4 fragment [49], which is commonly used in DNA metabarcoding but not in taxonomy.

Under ICNZ, DNA-based taxonomy is primarily applied to hexapods and scarcely in protists. In 2013, Jörger and Schrödl validly typified microscopic mollusks based on their DNA, arguing that these animals lack distinguishing phenetic characters, DNA analysis consumes the entire specimen, and DNA is a part of an individual organism [50]. Later, diagnostic nucleotide sequences were proposed to delimit species, genera, and families in oligotrich ciliates [51]. Meierotto and coworkers proposed that certain hyperdiverse insect groups should predominantly be described on the basis of their DNA barcodes to facilitate the classification of thousands of congeneric species [52]. In the following years, hundreds of insect species were published in large batches based on holotype specimens with photographs but using DNA consensus

barcodes for diagnosis [53]. However, such diagnoses referring to the full-length consensus barcodes failed to consider intraspecific and interspecific variation or indicate critical nucleotide signatures for distinguishing among species, being of limited utility for further sequence-based comparisons.

### Type material in DNA-based taxonomy

Type material carries taxon names, and its use is strictly regulated by the ICNs (except in viruses, where types are not used). Traditionally, type materials include living strains in prokaryotes and vouchered specimens (including nonliving strains) or illustrations of microorganisms within the botanical and zoological realms. However, the high standards for traditional type materials often pose significant challenges for minute species, especially the unculturable microorganisms that are being increasingly discovered through high-throughput sequencing data from environmental studies. These limitations highlight the need for alternative approaches that complement or expand on traditional practices.

It is open to interpretation of the codes whether environmental material samples, such as soil, plant tissues, and dung, conform to the ICNs' requirements for type specimens if collected and vouchered properly (i.e., a single collection with metadata deposited in a compliant repository). Contrary to the intention of the ICNs, these environmental material samples typically represent mixtures of multiple organisms and viruses. It can, however, be argued that most classical specimens – plant inflorescences, whole insects, lichen thalli, fruiting bodies, and cultures – are either holobionts or contain multiple organisms, including pathogenic microbes, endocellular bacterial symbionts, and the virome [54]. Still, the target organisms tend to dominate by biomass in these specimen samples but not necessarily in complex environmental samples. This makes it challenging to morphologically distinguish pieces and cells of the focal organisms from those of co-occurring species unless using DNA-based visualization techniques, such as fluorescence *in situ* hybridization (FISH) or similar methods for proteins and fatty acids. Despite these challenges, environmental samples offer unique advantages. They preserve the ecological context of species, providing insights into their roles within natural ecosystems. Moreover, environmental samples retain the potential for future characterization as technologies and analytical tools continue to advance. This forward-looking aspect allows researchers to revisit archived environmental samples and extract new insights into focal organisms that were previously inaccessible due to technological limitations.

Although 'DNA types' have been effectively used for describing cryptic species of microscopic slugs [50], DNA samples derived from specimens and living strains are currently not accepted as type materials according to ICNAFP and ICNP. Although DNA samples lack quantifiable morphological characters compared with physical specimens and cultures, they retain the genomic features that can be used as diagnostic characters or defining species features. This is particularly valuable when the physical specimens may have been depleted for molecular analysis in the case of minute organisms or nongrowing, stagnant cultures. We advocate that the DNA sample, independent of source material, could effectively represent type material in these cases. However, the use of DNA samples as type material should be preceded by rigorous molecular analyses to confirm the novelty of the taxon in question. In addition, the continuously evolving technologies allow the sequencing and analysis of other genetic markers or the entire genome, including DNA modifications such as methylations, providing a more comprehensive genomic foundation for taxonomic classification.

DNA sequences, particularly whole-genome sequences, offer another powerful alternative for physical type materials. Currently, DNA sequences may act as type materials for prokaryotes

based on the Candidatus system and SeqCode but not under the ICNs. Whole-genome sequences, as required in SeqCode, offer the maximum amount of diagnostic taxonomic characters [55]. They can additionally predict certain aspects of morphology (e.g., the presence of a flagellum), physiology (e.g., motility), and functioning (e.g., nutrition). Critics argue that DNA sequences are unsuited for types, because they are virtual data rather than physical specimens, and images of four-letter nucleotide sequences do not fulfill the illustration requirement of the ICNs [4,19]. However, sufficiently long, high-quality DNA sequences yield much more species-diagnostic information than an illustration or degraded specimen that has lost characteristic features and DNA integrity [50,56].

Furthermore, low contamination with other organisms is a clear benefit of high-quality whole-genome DNA sequences over DNA samples. Because DNA sequences do not degrade over time as DNA and physical specimens do, DNA sequence-based types thus offer an aspect of scientific democratization that physical types cannot match: the types become instantly digitally available to everyone rather than being kept in herbaria and zoological museums, without export possibilities, and where permission for partly destructive sampling (e.g., DNA extraction) is typically not given.

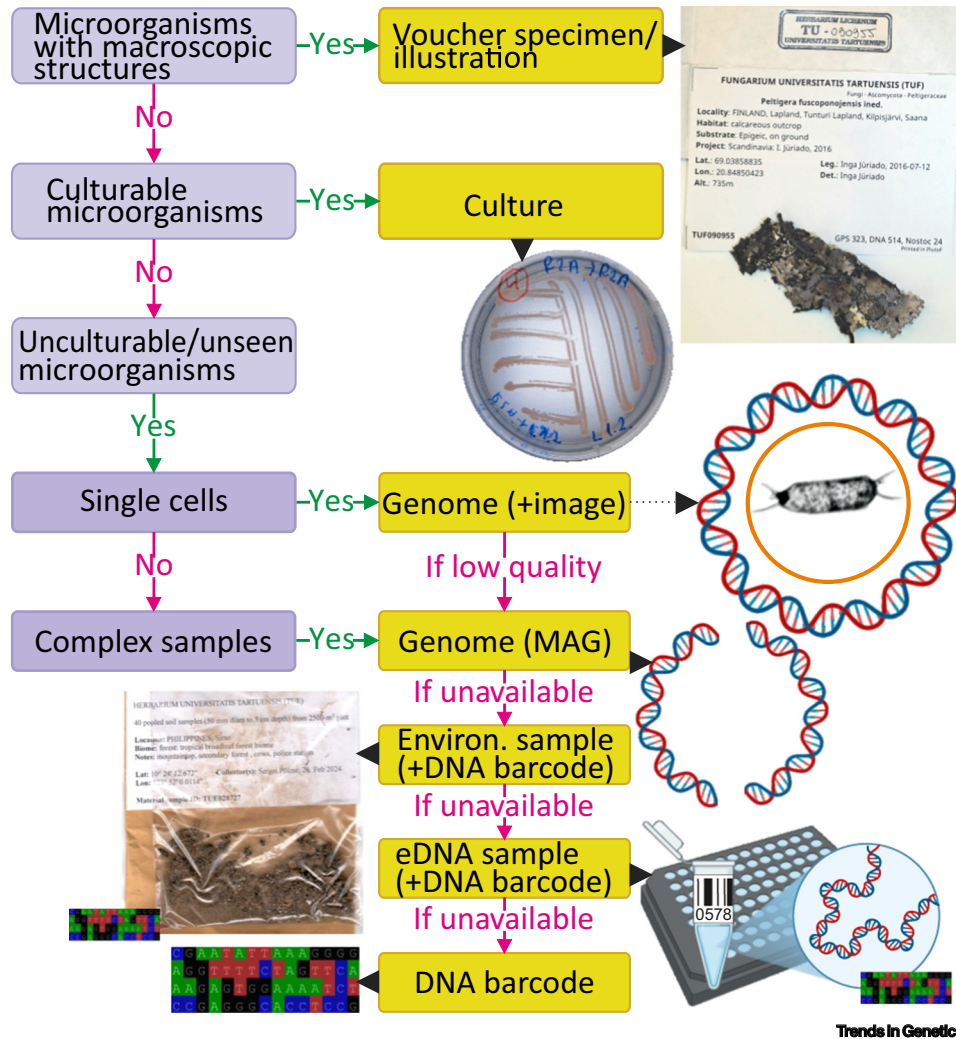
The DNA sequences of genetic markers, such as the ITS region, may also be suited for type materials on certain occasions. They may carry sufficient information for diagnostic differentiation among species, but not in all situations of closely related species [57]. In particular, nearly full-length rRNA genes might be used as an alternative for the description of novel diversity, including assessing their phylogenetic placement, intraindividual and intraspecific variability, and **barcoding gap** relative to the closest species.

In DNA-based taxonomy, **epitypes** may play an important role in bridging the taxa described based on molecular and phenotypic characters. Epitypes represent interpretationally relevant material that may be used to complement shortcomings of molecular or morphology-based species descriptions [1]. For example, genome sequences could be designated as epitypes for the species in which the holotype and other type material constitute voucher strains or specimens with poorly discriminating morphophysiological features. Conversely, physical voucher specimens could also be designated as epitypes for the species described on the basis of DNA and DNA sequences. So far, epitypes are used exclusively under ICNAFP for replacing poor-quality specimens or cultures with ones that have better-developed morphological features [1], but we argue that it is both timely and useful to extend this concept to other ICNs.

To speed up species description and classification of life, these alternative typification options based on environmental samples, DNA samples, and DNA sequences should be allowed, considering the multiple new ways of taxon discovery using molecular identification tools. From the historical and analytical perspectives, we propose to adopt a tiered approach that prioritizes physical, vouchered samples as types where possible but incorporates environmental samples, DNA samples, and DNA sequences as complementary alternatives. Considering the information content and potential further analyses, genome sequences, environmental samples, DNA samples, and genetic marker sequences could be considered, in that order, if the voucher specimens or strains have been lost, depleted, or are unavailable for legitimate reasons (Figure 1).

### Benefits of DNA-based taxonomy

Although the ICNAFP, ICZN, and ICNP claim to cover all organisms in their respective taxonomic groups, this is not effectively true for species that are uncultivable or unobserved based on morphology. DNA-based taxonomic approaches offer the following alternative solutions for describing these taxa.



Trends in Genetics

Figure 1. Decision map for choosing the taxonomic type (yellow boxes) depending on a type of organism (violet boxes). Abbreviation: MAG, metagenome-assembled genome. Photo credits: Jane Lees and Flora Vincent (unicellular eukaryote image from a single-cell analysis workflow).

First, DNA-based taxonomy opens new avenues for communicating organisms that cannot be named properly on the basis of classical taxonomic regulations and approaches. This includes microscopic organisms that lack distinguishing morphological characters and that cannot be established or maintained in culture. SeqCode ameliorated this shortfall of the ICNP by allowing genome sequences as types for prokaryotes. Species Hypotheses of UNITE enable communication of fungi and other eukaryotes [12].

Second, DNA sequence-based taxonomic diagnoses are easier to perform, because they do not require detailed images, time-consuming morphological or physiological descriptions, and visits to repositories in other countries to study type material [58]. DNA-based approaches do not necessarily require in-depth taxonomic expertise of the particular organism group, which would take several years of specific education. Nonetheless, including a specialist taxonomist team member is strongly recommended for interpretation and avoiding

mistakes. [Box 1](#) provides an example of a species and genus description based on a physical sample as a type and DNA features for diagnosis, following the ICZN, integrated with the rationale provided here.

**Box 1. Example of a new family, genus, and species description in nucleariid protists (key distinctive nucleotides in diagnostic nucleotide signatures are underlined)**

**Paranucleariidae Tedserso, fam. nov.** [Zoobank 7C441DCC-263B-4279-8E2B-DE519A51A1ED]

**Type genus.** *Paranuclearia* Tedserso

**Diagnosis.** Distinguishable from other species of *Nucleariidea* based on diagnostic nucleotide signature in LSU D6 (tcggrgragt; positions 1747–1757 in *N. slavaukraini*, 1657–1667 in *Saccharomyces cerevisiae*; one mismatch allowed). Phylogenetically delimited as the least inclusive, monophyletic clade in *Nucleariidea*, covering sequences EUK1155152, EUK1109526, EUK1218300, EUK1197978, EUK1206000, EUK1197970=PX091260 and EUK1205180=PX091252 ([Figure I](#) and [Figure S1](#) in the supplemental information online).

**Notes.** Recognized on the basis of eDNA sequences only, with no morphological information. The 315 records indicate occurrence in soil (93.7%), sediments (4.4%), and freshwater (1.6%). Found in subarctic tundra to tropical biomes worldwide, except mainland Antarctica. Expected to comprise around 200 species based on [Figure S1](#) in the supplemental information online.

**Paranuclearia Tedserso, gen. nov.** [Zoobank 3D154D32-EE0F-4266-9F7A-C8F9AD50DB56]

**Type species.** *Paranuclearia slavaukraini* Tedserso

**Diagnosis.** Distinguishable from other genera of *Paranucleariidae* based on diagnostic nucleotide signatures in SSU V9 (tctgcttgaagt; reverse strand positions 89–76 in *P. slavaukraini*, 88–75 in *S. cerevisiae*; one mismatch allowed), 5.8S (tagcaaattgcgataag; positions 47–63 in *P. slavaukraini* and *S. cerevisiae*; one mismatch allowed) and LSU D2 (tgattttaagatgttgacgg; positions 633–652 in *P. slavaukraini*, 606–625 in *S. cerevisiae*; one mismatch allowed). Phylogenetically delimited as the least inclusive, monophyletic clade in *Paranucleariidae*, covering sequences EUK1198008=PX091246, EUK1197998=PX091247, EUK1197994=PX091245, EUK1205453=PX091248, and EUK1205180=PX091252 ([Figure I](#) and [Figure S1](#) in the supplemental information online).

**Notes.** Recognized based on eDNA sequences only, with no morphological information. Expected to comprise five species based on [Figure I](#).

**Paranuclearia slavaukraini Tedserso, sp. nov.** [Zoobank 524810F5-3FC1-4822-B5CB-4472A89CD2FB]

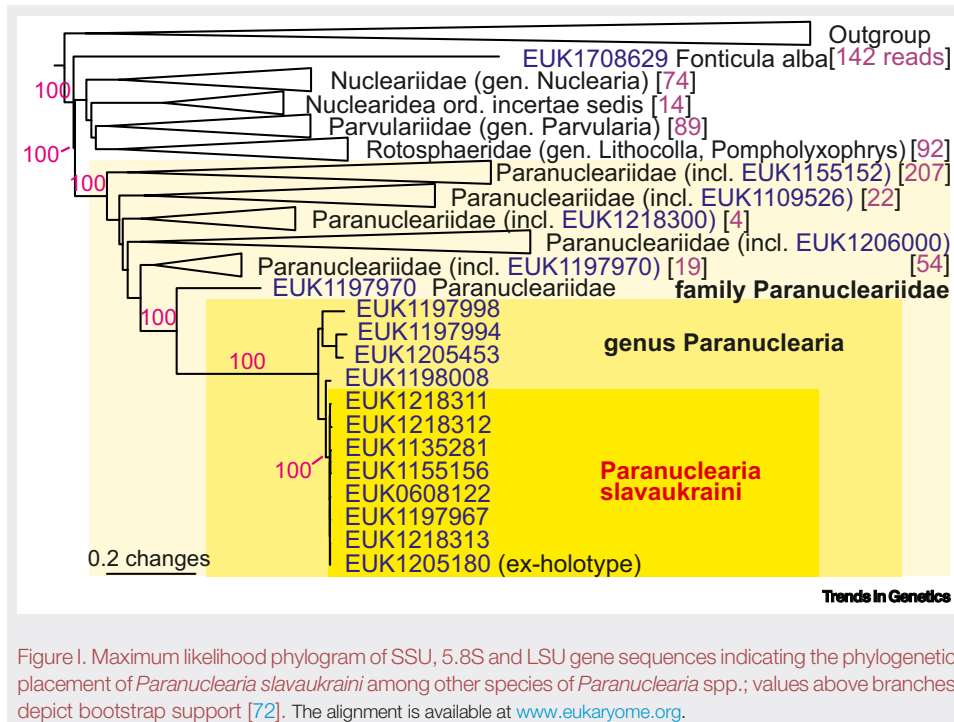
**Diagnosis.** Separation from other species of *Paranuclearia* based on ITS2 (ttgaaaaaggaccaatt; positions 209–227 in ex-holotype sequence; one mismatch allowed), SSU V9 (tgggtttgattgctgaaacg; reverse strand positions 130–109; no mismatch allowed) and LSU (tttgaggtcttaaagcata; positions 501–520; one mismatch allowed) as indicated in [Figure II](#). Intraspecific variation up to 3.0% in the ITS region, 0.6% in SSU V3–V9 and 0.5% in LSU D1–D2. Interspecific distance at least 11.3% in ITS2, 1.5% in SSU V3–V9 and 1.2% in LSU D1–D2.

**Type.** Topsoil sample TUE000403 (holotype); eDNA sample TUE100403; eDNA sequence EUK1205180=PX091252; locality: mixed boreal forest Selytska, Udmurtia, 57.0987 °N, 53.1584 °E. Coll. Kaarin Parts, 25.06.2012. TUE, repository of the University of Tartu.

**Other materials.** Sequences EUK1135281=PX091257 (*Alnus incana* forest soil in Haaslava, Estonia, 58.3280 °N, 26.8188 °E); EUK1155156=PX091249 (*Nothofagus* dominated forest soil in New Zealand, –45.9173 °N, 169.4830 °E); EUK1197967=PX091258 (mixed deciduous forest soil in Jägala, Estonia, 59.4742 °N, 25.1638 °E); EUK1218311=PX091256 (mixed deciduous forest soil in Lüütre, Estonia, 58.1444 °N, 25.2628 °E); EUK1218312=PX091259 (*Salix triandra* wetland soil in Prangli, Estonia, 59.61505 °N, 24.9871 °E); EUK1218313=PX091251 (sediment in Wilmurt Lake, NY, USA, 43.4268 °N, –74.7240 °E); and EUK0608122=PX091250 (sediment in the Olimar River, Uruguay, –33.2431 °N, –54.4017 °E).

**Etymology.** *Paranuclearia* refers to the phylogenetic links to the genus *Nuclearia*; *slavaukraini* (Ukrainian) refers to the meme in the Ukrainian defense against the Russian imperialistic war.

**Notes.** Found in wet forest soils and sediments (n=8 records), with potentially global distribution.



Third, designating epitypes facilitates linking of morphological and molecular taxonomy. In particular, epitypes facilitate discrimination among species that are difficult to identify and delimit on the basis of morphophysiological and DNA sequence data alone.

Fourth, physical space requirements and costs of storing and exchanging DNA, particularly DNA sequences, are much smaller than physical organism specimens and living cultures. DNA pellets can be stored at room temperature in physical repositories; the digital DNA sequence data are stored and routinely exchanged on servers worldwide.

Fifth, the ICNP requirement of depositing living strains in at least two culture collections in different countries that can be accessed by the scientific community conflicts with the interpretation of the Nagoya protocol and legislation of several countries, such as Brazil and South Africa [59]. SeqCode enables describing bacterial species without exporting living microorganisms [60]. Unlike biological samples and DNA samples, the DNA sequences can be easily accessed via international online repositories.

Sixth, delimiting and naming new species improves molecular biodiversity surveys for conservation and research purposes [61]. Besides the unseen taxa, it would offer names to species undescribed merely because they do not possess remarkable morphological features or the rare species found only once or twice, hence not considered priorities for description.

Seventh, classification systems accounting for DNA-based taxa offer greater resolution at higher taxonomic levels and facilitate phylogeny-free evolutionary ecological analyses across broad taxonomic groups [62]. For example, these methods indicate that soil fungal community assembly is driven more by dispersal limitation relative to selection compared with bacterial community assembly [63].

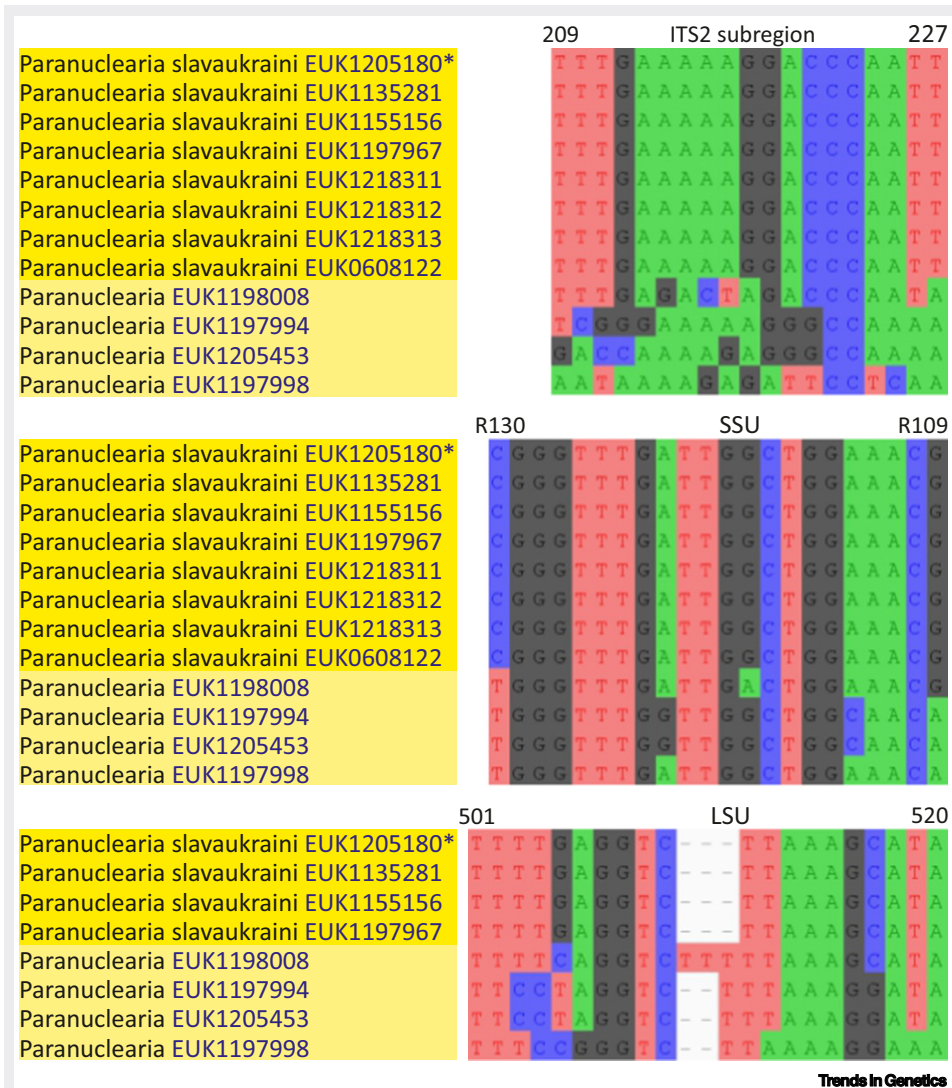


Figure II. Diagnostic nucleotide signatures of *Paranuclearia slavaukraini* in ITS, SSU, and LSU relative to other *Paranuclearia* species. Nucleotide positions refer to the forward strand (ITS and LSU) or the reverse strand of the ex-holotype sequence (asterisk).

Eighth, residual chimeric DNA sequences can be efficiently detected using taxonomy-aware software [64] and manual screening. This is required to remove artefacts from the amplicon sequencing data intended for DNA-based taxonomic descriptions or biodiversity analyses.

### Shortcomings and risks of DNA-based taxonomy

Multiple limitations of DNA-based taxonomy have been raised for prokaryotes [41], insects [65], and fungi [66]. These issues apply to all organisms governed by different ICNs, and there are ways to ameliorate these (Table 1; see Outstanding questions).

First, there is a risk of parallel taxonomy (i.e., the same species or higher-ranking taxa are independently described on the basis of phenotypic and molecular features) [50,65,66]. This risk

Table 1. Proposed requirements for DNA-based classification of prokaryotes, fungi, and microeukaryotes

| Subject                                     | Bacteria   | Fungi   | Protists and algae  |
|---|--|---|---|
| Representative or type sequence             | High-quality genome and SSU sequences, recommended ITS and LSU sequences                       | Genome or high-quality SSU, ITS, and LSU sequences; Dikarya and Mucoromyceta subkingdoms do not require an SSU sequence           | Genome or high-quality SSU, ITS, and LSU sequences; SSU coverage may vary by taxonomic group                                      |
| Biological samples and technical replicates | At least two biological samples or two technical replicates of one sample                      | At least two biological samples or two technical replicates of one sample; for genetic markers, at least three biological samples | At least two biological samples or two technical replicates of one sample; for genetic markers, at least three biological samples |
| Metadata                                    | Coordinates, sample/substrate source; preferably host, date, site and sampling details         | Coordinates, sample/substrate source; preferably host, date, site and sampling details  | Coordinates, sample/substrate source; preferably host, date, site and sampling details  |
| Sequence deposition                         | Vetted genome and rRNA genes: INSDC; raw data: INSDC; both associated with metadata            | Vetted genome and rRNA genes: INSDC; raw data: INSDC; both associated with metadata   | Vetted genome and rRNA genes: INSDC; raw data: INSDC; both associated with metadata   |
| Physical sample deposition                  | DNA and material samples in an ICN-approved repository   | DNA and substrate samples in an ICN-approved repository   | DNA and substrate samples in an ICN-approved repository   |
| Publication                                 | Peer-reviewed, preferably in community-approved journals                                       | Peer-reviewed, preferably in community-approved journals  | Peer-reviewed, preferably in community-approved journals  |
| Diagnosis of species                        | Diagnostic genome and SSU gene features; intraspecific variability estimates and barcoding gap | Diagnostic barcoding gene features, preferably for two or more markers; intraspecific variability estimates and barcoding gap     | Diagnostic barcoding gene features, preferably for two or more markers; intraspecific variability estimates and barcoding gap     |
| Diagnosis of higher-ranking taxa            | Diagnostic genome and SSU gene features; least inclusive phylogenetic lineage                  | Diagnostic barcoding or phylogenetically informative genetic marker features; least inclusive phylogenetic lineage                | Diagnostic barcoding or phylogenetically informative genetic marker features; least inclusive phylogenetic lineage                |
| Phylogenetic placement                      | Phylograms for genomes and SSU separately  | Phylograms for the ITS marker and phylogenetically informative genetic marker(s) combined or separately                           | Phylograms for the barcoding marker and phylogenetically informative genetic marker (s) combined or separately                    |
| Application restrictions                    | Uncultivable bacteria and archaea; taxa with strains unavailable due to legislation            | Uncultivable fungi that do not produce fruiting bodies or other tangible morphological structures                                 | Organisms that produce no macroscopic structures or colonies  |

can be reduced by defining the taxonomic groups for which DNA-based taxonomy is feasible (e.g., uncultivable lineages or groups that lack clear morphological distinguishing characters, or morphological structures altogether) and for which it would be unfeasible and undesirable (e.g., macroscopic organisms; taxa where multiple species are morphologically described but not sequenced). Greater efforts are needed to sequence DNA barcodes or genomes of existing type materials to provide reference data of the species defined on the basis of morphological characters and to move toward integrated taxonomy [3,58].

Second, should the SeqCode and ICNP eventually merge, as intended by the SeqCode developers [13,67], priorities of names may become a source of conflict [41]. Such confusion could be limited if the DNA-based names are retained but with amendments to the authors of taxon names, taxon descriptions, and types when ICN-compliant voucher specimens or cultures emerge. Such newly designated physical vouchers could be added as additional types (i.e. **corpotypes**; *corpotypus* in Latin) in emendations of taxonomic descriptions of DNA-based species.

Third, DNA-derived characters may be prone to artefacts, such as low read quality, chimera formation, unresolved repeats, or missing A/T or G/C-rich regions and chromosome ends. The quality issues can be alleviated by requiring sequencing at least two or three independent

samples or technical replicates, stringent quality control, and public deposition of raw and quality-filtered reads.

Fourth, on many occasions, genetic marker sequences offer insufficient taxonomic resolution by lumping closely related species [66,68]. Here, additional taxon-specific genetic markers should be used for genomic data, or DNA-based classification in such taxa should be avoided. This is mainly an issue of the SSU marker for most prokaryotes and eukaryotes and the ITS region for many fungal genera.

Fifth, the multiple chromosomes, mitochondria, and sometimes plastids in eukaryotic genomes pose unique challenges for genome reconstruction from environmental DNA or single-cell samples, with a high risk of misassembly. However, advancements in sequencing technologies and assembly tools are rapidly improving the accuracy and feasibility of eukaryotic genome reconstruction [69].

Sixth, updates to the Nagoya protocol emphasize fair sharing of benefits from digital genetic information, including for taxonomic purposes (<https://www.cbd.int/dsi-gr>). This may encourage countries to request certain benefits for describing new taxa on the basis of sequences originating from areas under their jurisdiction and potentially even demand withdrawal of publicly available sequence data.

Seventh, DNA-based taxonomy may increase the gap in opportunities among researchers in rich and developing countries due to differential access to modern sequencing facilities [65,66]. However, well-financed laboratories generate and openly release orders of magnitude more biodiversity data to test ecological hypotheses and address pressing issues in health. Because these data are not generated with a focus on taxonomic descriptions, all research teams gain equal access to data reanalysis from various taxonomic perspectives.

Eighth, DNA-based descriptions may flood the research community with hundreds of thousands of new names and taxa [66]. Indeed, massive algorithm-driven naming efforts have been conducted in the Candidatus taxon system of prokaryotes [39] as described above. The use of artificial or arbitrary Latinized words for names has sometimes been practiced since Linnaean times, and it is permitted, although not encouraged, in the ICNs. To prevent potential issues with overcomplicated, meaningless names and low-quality taxonomic descriptions, computer algorithms could be banned from preparing the names or suggesting alternative names related to, for example, material collectors, type localities (including native languages), and host organisms and avoiding homonyms. Proper peer review for publishing DNA-based taxa should be mandatory and limited to community-nominated, reliable journals [13,70]. DNA-based taxonomic descriptions warrant more stringent quality control than morphology-based approaches, because the DNA taxonomists may have limited expertise in the taxa under discussion (Box 2; Table 1).

Ninth, there may be a loss of motivation and intent to perform classical taxonomic descriptions and maintain culture collections and specimen repositories [66]. Researchers generally compromise between what they are fascinated with and what is feasible and profitable for their scientific career, considering their deep interests and the expectations of funders and other stakeholders. Should complex materials, such as soil and DNA samples, become more common as type material, specimen repositories would increasingly store these rather than discard them after analyses or publishing, adding additional value to these samples. Herbaria and culture collections would simply continue to offer invaluable sources for biotechnology, biomedicine, conservation, and restoration biology [71].

### Box 2. Overall requirements for DNA-based taxonomy

We propose the following requirements for the DNA-based description of taxa that may somewhat differ across taxonomic groups (Table 1) due to the taxonomic resolution and legacies of early molecular analysis.

1. High-quality genome and at least a nearly complete SSU (at least 1200 bp) for prokaryotes; much of SSU (V3–V9 regions), ITS region and LSU (D1–D9) for eukaryotes. In animals, the 'BOLD' barcode of CO1 might only be used for diagnosis (i.e., not as type material).
2. At least three independent records (marker gene and metagenome studies) or two technical replicates (for partial genomes) of the taxon in question (Table 1). The records should preferably stem from different datasets. Other biological or technical replicates may have lower specific genetic marker or genome coverage than the type material.
3. Rich available, standards-compliant metadata, including geographic coordinates and sample source or substrate at a minimum. If material or DNA samples serve as types, these specimens should be vouchered in a repository accepted by a relevant ICN.
4. The quality-filtered sequence data must be available in any of the International Nucleotide Sequence Databases Consortium (INSDC) member repositories, with accession numbers indicated. Raw sequences should be in fasta or fastq format in INSDC to support further efforts to refine the genome or marker sequences, phase haplotypes and estimate intraindividual or intrapopulation variability [13].
5. Publication exclusively in peer-reviewed journals approved by the research communities to avoid low-quality descriptions [70].
6. Species descriptions relying on molecular diagnoses that explicitly define differences to any closely related species in the primary DNA barcoding gene or at the level of the genome (Table 1).
7. Determination of intraspecific variation and interspecific distance (i.e., barcoding gap) to the closest species to improve classification of partly matching DNA sequences [56].
8. Maximum likelihood or Bayesian phylogenetic analyses indicating the relative placement of the newly described species and higher-ranking taxa among the sister taxa. Simple neighbor-joining analyses, sequence clustering and BLAST statistics are insufficient for this purpose.
9. In descriptions of higher-ranking taxa, the inclusion of molecular synapomorphies, phylogenetically informative genes and genomes to distinguish the described taxa from close relatives at the same taxonomic level [10,73]. The least inclusive clades should also be defined, referring to phylograms and following the standards of the PhyloCode [2].

### Proposal for ICNs

To promote taxonomy and, indirectly, classification of unculturable organisms, we propose the following general amendments to the ICNs, which should clearly state the following:

1. Physical samples, DNA and relevant genetic marker and genome sequences are permitted as type materials in addition to vouchered specimens and cultures in certain situations.
2. Diagnostic characters of genetic markers or entire genomes can be used to describe new species and higher-ranking taxa. For DNA-based taxonomic descriptions, detailed molecular diagnoses are mandatory.
3. Epitypes and corpotypes can be used as interpretative types to link phenetic and molecular species delimitation.

Relevant amendments to ICNAFP, ICZN, and ICNP are proposed in Supplementary Item 1 in the supplemental information online.

### Concluding remarks

Describing novel taxa for unculturable microorganisms is feasible and desirable if subjected to regulations and rigorous quality control (see Outstanding questions). It is most beneficial for taxonomic groups where morphology-based taxonomy is impractical or where most species have already been characterized by DNA features. Species with names improve our capacity to refer to particular organisms and facilitate biodiversity surveys; conservation planning; and assessment of toxic, pathogenic, and mutualistic organisms directly. Furthermore, well-structured classifications would offer additional possibilities for using phylodiversity and evolutionary ecological methods without performing phylogenetic analyses, and they would improve taxonomy-aware

### Outstanding questions

How can we best integrate DNA-based taxonomy into the international codes of nomenclature?

How can we reduce the risks of generating parallel taxonomies, where morphologically defined but unsequenced species and sequenced species with undefined morphology are conspecific?

Should the specialist research community nominate the taxonomic groups where DNA-based taxonomy is feasible?

How can we ensure stringent quality control of type sequences and other diagnostic sequences intended to be used in taxonomy to avoid artefacts, such as low read quality, chimera formation, unresolved repeats, or missing A/T or G/C-rich regions and chromosome ends?

How can we avoid the massive algorithm-based generation of taxonomic descriptions and fully artificial, meaningless names?

What are the objective criteria for nominating and accepting journals for publishing DNA-based taxonomic research?

How can we minimize the risk that the implementation of DNA-based taxonomy increases the gap in opportunities among researchers in rich and developing countries due to differential access to modern sequencing facilities and expertise in bioinformatics and molecular taxonomy?

How can we increase interest in, and support of, classical morphology-based taxonomy when the DNA-based taxonomy is booming?

chimera filtering for reference-based methods in eDNA metabarcoding analyses. It is important to emphasize that the goal is not to replace the traditional type concept or the expertise of traditional taxonomists, especially in organism groups where the tools and approaches of traditional taxonomy work well. Instead, the DNA-based taxonomy offers complementary pathways to better address the challenges of studying microorganisms that are unculturable, cryptic, or trapped in legislation.

### Acknowledgments

L.T. acknowledges funding for preparing this article from a European Research Council support grant (MOBERC116).

### Declaration of interests

The authors declare no competing interests.

### Supplemental information

Supplemental information associated with this article can be found online at <https://doi.org/10.1016/j.tig.2025.07.009>.

### References

- Turland, N. (2019) The code decoded. *Adv. Books* 1, e38075
- de Queiroz, K. and Cantino, P. (2020) *International Code of Phylogenetic Nomenclature (PhyloCode)*, CRC Press
- Hedlund, B.P. (2015) The changing landscape of microbial biodiversity exploration and its implications for systematics. *Syst. Appl. Microbiol.* 38, 231–236
- Lücking, R. *et al.* (2021) Fungal taxonomy and sequence-based nomenclature. *Nat. Microbiol.* 6, 540–548
- Nilsson, R.H. *et al.* (2016) Top 50 most wanted fungi. *MycKeys* 12, 29–40
- Murray, R.G. and Stackebrandt, E. (1995) Taxonomic note: implementation of the provisional status Candidatus for incompletely described prokaryotes. *Int. J. Syst. Evol. Microbiol.* 45, 186–187
- Tautz, D. *et al.* (2003) A plea for DNA taxonomy. *Trends Ecol. Evol.* 18, 70–74
- Renner, S.S. (2016) A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. *Syst. Biol.* 65, 1085–1095
- Hibbett, D.S. *et al.* (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. *Fungal Biol. Rev.* 25, 38–47
- Marin, B. *et al.* (2003) Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist* 154, 99–145
- Rossello-Mora, R. (2006) DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In *Molecular Identification, Systematics and Population Structure of Prokaryotes* (Stackebrandt, E., ed.), pp. 23–50, Springer
- Köljalg, U. *et al.* (2013) Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277
- Hedlund, B.P. *et al.* (2022) SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* 7, 1702–1708
- Oren, A. *et al.* (2023) International code of nomenclature of prokaryotes. Prokaryotic code (2022 revision). *Int. J. Syst. Evol. Microbiol.* 73, 005585
- Rheindt, F.E. *et al.* (2023) Tightening the requirements for species diagnoses would help integrate DNA-based descriptions in taxonomic practice. *PLoS Biol.* 21, e3002251
- Adams, M.J. *et al.* (2017) 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Arch. Virol.* 162, 1441–1446
- Schidkraud, C.L. *et al.* (1961) The formation of hybrid DNA molecules and their use in studies of DNA homologies. *J. Mol. Biol.* 3, 595–617
- Chun, J. *et al.* (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466
- Thiele, K.R. *et al.* (2023) DNA sequences as types: a discussion paper from the Special-purpose Committee established at the XIX International Botanical Congress in Shenzhen, China. *Taxon* 72, 965–973
- May, T.W. and Hawksworth, D.L. (2024) Proposals for consideration at IMC12 to modify provisions related solely to fungi in Chapter F of the International Code of Nomenclature for algae, fungi, and plants. *IMA Fungus* 15, 25
- Turland, N. *et al.* (2024) Synopsis of Proposals on Nomenclature–Madrid 2024: a review of the proposals to amend the International Code of Nomenclature for algae, fungi, and plants submitted to the XX International Botanical Congress. *Taxon* 73, 325–404
- International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature* (Fourth edition), The International Trust for Zoological Nomenclature, London
- de Queiroz, K. (2007) Species concepts and species delimitation. *Syst. Biol.* 56, 879–886
- Pawlowski, J. *et al.* (2012) CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10, e1001419
- Schoch, C.L. *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6241–6246
- Martijn, J. *et al.* (2019) Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. *Environ. Microbiol.* 21, 2485–2498
- Wurzbacher, C. *et al.* (2019) Introducing ribosomal tandem repeat barcoding for fungi. *Mol. Ecol. Res.* 19, 118–127
- Ferraz Helene, L.C. *et al.* (2022) New insights into the taxonomy of bacteria in the genomic era and a case study with rhizobia. *Int. J. Microbiol.* 2022, 4623713
- Ahrens, D. (2024) Species diagnosis and DNA taxonomy. In *DNA Barcoding: Methods and Protocols* (DeSalle, R., ed.), pp. 33–52, Springer
- Taylor, J.W. *et al.* (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* 31, 21–32
- Pons, J. *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609
- Zhang, J. *et al.* (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876
- Puillandre, N. *et al.* (2021) ASAP: assemble species by automatic partitioning. *Mol. Ecol. Res.* 21, 609–620
- Lachance, M.A. *et al.* (2020) Delineating yeast species with genome average nucleotide identity: a calibration of ANI with haplontic, heterothallic *Metschnikowia* species. *Antonie Van Leeuwenhoek* 113, 2097–2106

35. Hug, L.A. *et al.* (2016) Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ. Microbiol.* 18, 159–173
36. Reynolds, D.R. and Taylor, J.W. (1991) DNA specimens and the 'International code of botanical nomenclature'. *Taxon* 40, 311–315
37. Kirk, P.M. and Griffith, G.W. (2021) Nomenclatural novelties: *Piromyces cryptodigmaticus* Fliegerova, K. Voigt & P.M. Kirk, sp. nov. *Ind. Fungal* 467, 1
38. Oren, A. (2025) Addition of Section 10, Rules 66–73 for further integration of Candidatus names into the International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 75, 006638
39. Pallen, M.J. *et al.* (2022) Naming the unnamed: over 65,000 Candidatus names for unnamed Archaea and Bacteria in the Genome Taxonomy Database. *Int. J. Syst. Evol. Microbiol.* 72, 005482
40. Whitman, W.B. *et al.* (2022) Development of the SeqCode: a proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. *Syst. Appl. Microbiol.* 45, 126305
41. Arahai, D. *et al.* (2024) The best of both worlds: a proposal for further integration of Candidatus names into the International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 74, 006188
42. Imachi, H. *et al.* (2024) *Promethearchaeum syntrophicum* gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage 'Asgard' archaea, and proposal of the new archaeal phylum *Promethearchaeota* phyl. nov. and kingdom *Promethearchaeati* regn. nov. *Int. J. Syst. Evol. Microbiol.* 74, 006435
43. Kirk, P.M. (2012) Nomenclatural novelties: *Piromyces cryptodigmaticus* Fliegerová, K. Voigt P.M. Kirk. *Ind. Fungal* 1, 1
44. de Beer, Z.W. *et al.* (2016) *Hawksworthiomyces* gen. nov. (*Ophiostomatales*), illustrates the urgency for a decision on how to name novel taxa known only from environmental nucleic acid sequences (ENAS). *Fungal Biol.* 120, 1323–1340
45. Lücking, R. and Moncada, B. (2017) Dismantling *Marchandiomphalina* into *Agonimia* (*Verrucariaceae*) and *Lawreymyces* gen. nov. (*Corticaceae*): Setting a precedent to the formal recognition of thousands of voucherless fungi based on type sequences. *Fungal Divers.* 84, 119–138
46. Tedersoo, L. *et al.* (2024) Phylogenetic classification of arbuscular mycorrhizal fungi: new species and higher-ranking taxa in *Glomeromycota* and *Mucoromycota* (class *Endogonomycetes*). *MycKeys* 107, 249–271
47. Abarenkov, K. *et al.* (2024) The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Res.* 52, D791–D797
48. Kõljalg, U. *et al.* (2020) The Taxon Hypothesis paradigm—on the unambiguous detection and communication of taxa. *Microorganisms* 8, 1910
49. Öpik, M. *et al.* (2014) DNA-based detection and identification of *Glomeromycota*: the virtual taxonomy of environmental sequences. *Botany* 92, 135–147
50. Jörger, K.M. *et al.* (2013) How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front. Zool.* 10, 1–27
51. Ganser, M.H. *et al.* (2022) Molecular signature characters complement taxonomic diagnoses: a bioinformatic approach exemplified by ciliated protists (*Ciliophora*, *Oligotrichea*). *Mol. Phylogenet. Evol.* 170, 107433
52. Meierotto, S. *et al.* (2019) A revolutionary protocol to describe understudied hyperdiverse taxa and overcome the taxonomic impediment. *Deutsche Entomol. Z.* 66, 119–145
53. Sharkey, M.J. *et al.* (2021) Minimalist revision and description of 403 new species in 11 subfamilies of Costa Rican braconid parasitoid wasps, including host records for 219 species. *ZooKeys* 1013, 1–665
54. Kelliher, J.M. *et al.* (2023) The endohyphal microbiome: current progress and challenges for scaling down integrative multi-omic microbiome research. *Microbiome* 11, 192
55. Palmer, M. *et al.* (2022) It is time for a new type of type to facilitate naming the microbial world. *New Microbes New Infect.* 47, 100991
56. Dayrat, B. (2005) Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85, 407–417
57. O'Donnell, K. (2015) DNA sequence-based identification of *Fusarium*: current status and future directions. *Phytoparasitica* 43, 583–595
58. Riedel, A. *et al.* (2013) Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Front. Zool.* 10, 15
59. da Silva, M. *et al.* (2023) How legislations affect new taxonomic descriptions. *Trends Microbiol.* 31, 111–114
60. Van Lill, M. *et al.* (2024) SeqCode facilitates naming of South African rhizobia left in limbo. *Syst. Appl. Microbiol.* 47, 126504
61. Ritter, C.D. *et al.* (2019) Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. *Genome* 62, 147–159
62. Tedersoo, L. *et al.* (2018) High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers.* 90, 135–159
63. Zhang, G. *et al.* (2021) Dispersal limitation plays stronger role in the community assembly of fungi relative to bacteria in rhizosphere across the arable area of medicinal plant. *Front. Microbiol.* 12, 713523
64. Nilsson, R.H. *et al.* (2010) An open source chimera checker for the fungal ITS region. *Mol. Ecol. Res.* 10, 1076–1081
65. Zamani, A. *et al.* (2022) The future of zoological taxonomy is integrative, not minimalist. *Syst. Biodivers.* 20, 1–4
66. Thines, M. *et al.* (2018) Ten reasons why a sequence-based nomenclature is not useful for fungi anytime soon. *IMA Fungus* 9, 177–183
67. Venter, S.N. *et al.* (2024) Options and considerations for validation of prokaryotic names under the SeqCode. *Syst. Appl. Microbiol.* 47, 126554
68. Hongsanan, S. *et al.* (2018) Can we use environmental DNA as holotypes? *Fungal Divers.* 92, 1–30
69. Hu, J. *et al.* (2024) NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* 25, 107
70. Nilsson, R.H. *et al.* (2023) How, not if, is the question mycologists should be asking about DNA-based typification. *MycKeys* 96, 143
71. Johnson, K.R. *et al.* (2023) A global approach for natural history museum collections. *Science* 379, 1192–1194
72. Minh, B.Q. *et al.* (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534
73. Rosling, A. *et al.* (2011) *Archaeorhizomycetes*: unearthing an ancient class of ubiquitous soil fungi. *Science* 333, 876–879