PERSPECTIVE ARTICLE

DNA sequences as types: A discussion paper from the Special-purpose Committee established at the XIX International Botanical Congress in Shenzhen, China

Kevin R. Thiele,[1] Wendy L. Applequist,[2] Susanne S. Renner,[3] Tom W. May,[4] Ali A. Dönmez,[5] Quentin Groom,[6] Samuli Lehtonen,[7] Christine A. Maggs,[8] Valéry Malécot[9] & Hwan Su Yoon[10]

1 *School of Biological Sciences, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia*
2 *William L. Brown Center, Missouri Botanical Garden, St. Louis, Missouri 63110, U.S.A.*
3 *Department of Biology, Washington University, Saint Louis, Missouri 63130, U.S.A.*
4 *Royal Botanic Gardens Victoria, Melbourne, Victoria 3004, Australia*
5 *Botany Section, Department of Biology, Faculty of Science, Hacettepe University, 06800 Beytepe, Ankara, Turkey*
6 *Meise Botanic Garden, Meise, Belgium*
7 *Biodiversity Unit, University of Turku, 20014 Turku, Finland*
8 *School of Biological Sciences, Queen's University Belfast, Chlorine Gardens, Belfast BT9 5DL, United Kingdom*
9 *Institut Agro, Université d'Angers, INRAE, IRHS, SFR QUASAV, 49000 Angers, France*
10 *Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, Republic of Korea*
Address for correspondence: *Kevin R. Thiele, kevin.thiele@uwa.edu.au*

DOI https://doi.org/10.1002/tax.12931

Abstract A special-purpose Committee on DNA Sequences as Types was established at the XIX International Botanical Congress (IBC) in Shenzhen, China, in 2017, with a mandate to report to the XX IBC in Madrid in 2024 with recommendations on a preferred course of action with respect to potential amendments of the *International Code of Nomenclature for algae, fungi, and plants* to allow DNA sequences as types. This is the first in an expected series of papers from the Special-purpose Committee on this issue. We set out the background to the establishment of the Committee, explore key issues around typification that are pertinent to the question of DNA sequences as types, enumerate pros and cons of allowing DNA sequences as types, and foreshadow options for future discussion and potential recommendations.

Keywords DNA sequences; eDNA; typification

■ BACKGROUND

DNA sequences are becoming increasingly important in many of the core activities of modern taxonomy, including identifying specimens, delimiting species and other taxa, inferring evolutionary origins, estimating the phylogenetic structuring of life, and establishing taxonomic classifications (Blaxter & al., 2022). DNA's universality and the fact that some parts of genomes are highly variable while others are highly conserved makes it an important source of information for understanding the patterns of variation derived from evolution, at multiple scales. DNA is also chemically relatively stable in many situations, allowing DNA fragments to be sequenced from many preserved, and even some ancient, biological samples, and from the environment (Orlando & al., 2021). DNA is ubiquitous, informative and highly useful.

Perhaps not surprisingly given these properties, several proposals (e.g., Hawksworth & al., 2016 and Hawksworth, 2018 for Fungi; Renner, 2021 for Viridiplantae; Hedlund & al., 2022 for prokaryotes) have been made to use DNA sequences for another of taxonomy's core activities – typifying the names of species and infraspecific taxa[1]. Types are fundamental to a well-regulated nomenclature in the service of taxonomy, and DNA's fundamental role in living organisms and high utility for taxonomy has made it, for some, a good candidate for inclusion in a modern system of typification. These proposals have come especially from workers in taxonomic groups who believe that historical procedures of typification are inadequate for their taxonomic domain.

One such area where conventional typification is being questioned is in groups such as fungi and microorganisms

[1]Taxa above the rank of species (and at or below the rank of family; Art 7.1) are typified by the name of a species (Art. 10.1) except if by conservation a specimen or illustration is chosen as the type of a name of a genus (Art. 10.4). The current discussion on using DNA sequences as types does not concern typification of taxa at these ranks. Typification at ranks below species is the same as for species (Art. 8.1). In this paper, use of the word *species* implies *species and infraspecific taxa*.

Article history: Received: 10 Dec 2022 | accepted: 19 Jan 2023 | published online: 17 May 2023 | **Associate Editor:** Dirk C. Albach |

where increasing numbers of species are being discovered and delimited solely on the basis of assembled DNA sequences derived from environmental samples (Wu & al., 2019; Hedlund & al., 2022). However, under current rules of typification in the *International Code of Nomenclature for algae, fungi, and plants* (herein, the *Code*; Turland & al., 2018), a type must be a physical sample preserved in a herbarium or equivalent institution or, in some cases, an illustration (Art. 8.1). A DNA sequence is not a physical sample, and many of the species that are being discovered and characterised in this way cannot be cultured or purified into a sample using available technologies. Hence, the rules of the *Code* currently preclude the formal naming of significant portions of the Earth's biodiversity.

Despite the lack of any provisions under the *Code* that would allow taxa to be typified on DNA sequences, several publications have attempted to do exactly that, partly at least to demonstrate the problem. Lücking & Moncada (2017) attempted to describe novel species typified using "illustrations" of DNA sequences, while De Beer & al. (2016) attempted to typify a species using a letter-by-letter representation of a DNA sequence. The names so designated, however, are invalid under the current *Code* (see Art. 40.5, Ex. 6).

Given this issue, two proposals to amend the *Code* were published (Hawksworth & al., 2016) and presented to the Nomenclature Section of the XIX International Botanical Congress at Shenzhen, China, in 2017. These were as follows, as listed by Turland & Wiersema (2017):

Art. 8. Prop. O (308 – Hawksworth & al. in Taxon 65: 900) Insert a new paragraph after Art. 8.5 as follows:

"*8.6*. In fungi, when DNA sequence data corresponding to a new taxon have been detected, but no physical specimen has been found to serve as the type of the name of the new taxon (Art. 8.1–8.4), the type may be composed of DNA sequence data deposited in a public repository."

Rec. 8C. Prop. A (309 – Hawksworth & al. in Taxon 65: 900) Add a new Recommendation 8C:

"*8C.1*. When the type is composed only of DNA sequence data (Art. 8.6), the new taxon should be described with reference to a published phylogenetic analysis; both the phylogenetic tree and the DNA sequence alignment that was used to create the phylogenetic tree should be deposited in a publicly accessible repository."

"*8C.2*. A new taxon typified only by DNA sequence data should be represented by multiple sequences obtained in independent studies, of which one is designated as the holotype."

"*8C.3*. DNA sequence data used for typification should be drawn from the molecular regions that are appropriate for delimiting species, based on prevailing best practices as determined by the relevant taxonomic communities."

Art. 9. Prop. A (310 – Hawksworth & al. in Taxon 65: 900) Amend Art. 9.1 as follows:

"*9.1*. A holotype of a name of a species or infraspecific taxon is the one specimen, or sequence (Art. 8.6), or illustration (but see Art. 40.4) used by the author, or designated by the author as the nomenclatural type. As long as the holotype is extant, it fixes the application of the name concerned (but see Art. 9.15)."

While Art. 8 Prop. O was originally intended to deal with fungi only, a friendly amendment from the floor at the Section meeting (approved by one of the proposers) expanded it to cover all organismal groups covered by the *Code* (Lindon & al., 2020).

In the Rapporteur's comments to these proposals (Turland & Wiersema, 2017), published prior to the Congress, their significance was explained as follows:

"Prop. O, together with Rec. 8C Prop. A and Art. 9 Prop. A, would allow DNA sequence data to serve as the type of a fungal name when no physical specimen has been found to serve as the type. This is an issue that arises from sequencing environmental samples, where a sequence can indicate the existence of an apparently new taxon, but the organism itself cannot be found in the sample. The *Code* currently provides no means to name that taxon, because a type must be indicated (Art. 40.1) and that type can only be a specimen (Art. 40.4) or else an effectively published illustration "if there are technical difficulties of preservation or if it is impossible to preserve a specimen that would show the features attributed to the taxon by the author of the name" (Art. 40.5). Illustration is defined as "a work of art or a photograph depicting a feature or features of an organism" (Art. 8.1 footnote 1). One could argue that a DNA sequence is analogous to an illustration, in that it depicts the features of an organism, but it might be harder to claim it as a work of art (although, what is art?). The point is that, in principle, it would not be a great change in the *Code* to allow DNA sequences as types. However, the Rapporteurs are concerned about a practical issue: the lack of control as to the type sequence being an informative sequence. Many taxa could have the same sequence. The Recommendations of Rec. 8C Prop. A should certainly help in this respect, but they are only Recommendations and can, and no doubt will by some, be ignored. The Nomenclature Committee for Fungi does not support Prop. O, Rec. 8C Prop. A, and Art. 9 Prop. A (votes 2:7:1), with 8 voting for a Special Committee to examine the matter."

The Nomenclature Section meeting, after discussing this set of proposals, referred further deliberation to a Special-purpose Committee with a mandate to report to the XX IBC (Lindon & al., 2020). The Special-purpose Committee on DNA Sequences as Types was formally established by the General Committee in 2019 (Wilson, 2019).

The Shenzhen meeting also formally separated the provisions of the *Code* that only affect organisms treated as fungi (Chapter F of the *Shenzhen Code*) and established an independent governance mechanism for those provisions through a Nomenclature Session of an International Mycological Congress (IMC). The proposals that had been discussed at

the IBC in Shenzhen, and referred to the Special-purpose Committee, were also submitted to the San Juan meeting of the IMC (IMC11) in 2018 (Hawksworth, 2018), accompanied by a discussion of the "promises and pitfalls" of the "formal description of sequence-based, voucherless Fungi" (Lücking & al., 2018). A commentary prepared by the Secretaries (the fungal equivalent of the Rapporteurs for the IBC Nomenclature Section) on proposals to be considered by the IMC Nomenclature Session (May & Redhead, 2018) noted that there was strong opposition to the proposals by the Nomenclature Committee for Fungi and the International Commission for the Taxonomy of Fungi (ICTF, an organisation established under the International Mycological Association and International Union of Microbiological Societies). The Secretaries also noted that within the ICTF a number of options were considered that were not part of the formal proposals:

"In discussion of the formal proposal in the ICTF two procedural options for handling DNA-based names were raised: that names typified on DNA sequences alone must be: (1) published in specified journals (that adhere to a taxonomic code of practice); or (2) approved post-publication by a Subcommittee on DNA-based names. Further refinements discussed included: DNA-based names are allowed but (3) do not compete for priority with names based on specimens; (4) must be based on at least a certain number of sequences; (5) cannot be introduced in known genera unless all known species are sequenced; and/or (6) have epithets or authorship indicating the nature of their type, such as by including "DNA-" at the start of the epithet. All these options or refinements received less than 50% support within the ICTF, except for the last, for which there was a 50% Yes vote. A few ICTF members indicated that combinations of the options and refinements might be acceptable." (May & Redhead, 2018)

The proposals were the subject of a number of opinion pieces in the lead up to the San Juan IMC (Lücking & al., 2018; Thines & al., 2018; Zamora & al., 2018). They attracted a strong negative vote in the pre-Congress Guiding vote (May & Miller, 2018). In the IMC Nomenclature Session the proposals failed to attract the required support for amendment of Chapter F. Another Special-purpose Committee was established under the IMC to report to the 2024 IMC in Maastricht, The Netherlands (May & al., 2018). The present paper does not necessarily represent the views of that mycological Special-purpose Committee.

This paper is the first of several that will be prepared by the members of the IBC-established Special-purpose Committee. This first contribution is not intended as a considered opinion, but rather as an initial statement of the problem, exploration of the issues, summary of the pros and cons of DNA sequences as types, and brief outline of potential options. The intent is to be neutral and not definitive, and to lay out important issues that the Special-purpose Committee will consider in formulating its recommendations.

■ ITEMS OUT OF SCOPE FOR THE SPECIAL-PURPOSE COMMITTEE

The Special-purpose Committee on DNA Sequences as Types was formed with a narrow remit: to consider, and potentially propose, amendments to the *Code* that would govern the use of DNA sequences as types. Modern DNA sequencing is revolutionising taxonomy in many ways, some of which have been or are currently controversial, but the purpose of the Special-purpose Committee is to consider DNA sequences for the purposes of nomenclature (that is, the naming of taxa), not taxonomy (the delimitation and classification of taxa).

For this reason, the following issues are not relevant to the remit of the Committee:

(1) the utility or otherwise of different markers, particularly the so-called "barcoding" markers, for taxonomic delimitation or phylogeny;
(2) the advisability or otherwise of basing large numbers of taxonomic delimitations on DNA data alone;
(3) the merits or otherwise of algorithmic methods for species delimitation based on sequences, and
(4) the merits or otherwise of basing classifications and high-level taxa on analyses based purely or largely on DNA sequences.

In summary, every attempt has been made in this paper to separate issues around using DNA sequences for taxonomy (out of scope) from issues around using DNA sequences as types (in scope). The Special-purpose Committee recognises that, in the minds of many, these two issues will be intermixed: individuals who are enthusiastic about DNA taxonomy may be more likely to be comfortable with the idea of DNA sequences serving as types, and vice versa. But the *Code* and the Nomenclature Section of the IBC (to which the Committee will report) deal with issues of nomenclature, not taxonomy, so only nomenclatural matters pertaining to DNA sequences as types are in scope. Rigorously separating the two domains of taxonomy and nomenclature is important if we are to gain clarity on the issues in this space.

It is important also to be clear that the Special-purpose Committee has a remit to consider issues concerning DNA *sequences* as types. A sequence is the string of letters indicating the order of nucleotides (along with any uncertainties). In some circumstances, it may be possible to isolate a *sample* of genomic DNA, and designate this sample as a type. Lodging of DNA samples as types is a separate issue and is out of scope here, although it may be re-visited in future discussions.

■ TYPES AND TYPIFICATION

Typification is used to formally connect names with taxa with minimum ambiguity. Typification has a long and complex history (much of which is beyond the scope of this paper). At the starting point of our modern system of nomenclature

(1 May 1753 for vascular plants and many other taxa, later for various other groups), there was no formally agreed type concept or mechanism. The general notion of typification (at least for genera) had been stated earlier by Linnaeus (1736) but the formalisation of using type species to anchor the application of genus names was not put forward until 1843 in a set of naming rules for zoology known as the Strickland Code (a text developed by a committee of which Charles Darwin was a member; see Witteveen, 2016). Despite these early precursors, the first formal steps towards regulating the application of names of plants (as then used to cover algae, fungi and plants) through types only occurred in 1907 with the so-called American Code (Nomenclature Commission, 1907). The provisions established there were only formally adopted in 1952 with the publication of the Stockholm Code (see Nicolson, 1991, for discussion).

During the long history of development of the type concept, many European botanists remained sceptical of its utility, preferring instead to focus on the importance of descriptions (which can accommodate variation), an idea for which A.P. de Candolle was the lead proponent (Nicolson, 1991). Early American botanists, on the other hand, were convinced of the utility of the type method and became its key proponents, with comments such as "[a]ffixing of a name to a sufficient specimen in distributed collections […] [will] more surely identify the genus or species than might a brief published description" (Gray, 1864: 279) and "[when] the application of a name is determined by reference to its nomenclatorial type […] a specific (or subspecific) name stands or falls according to the disposition of the type specimen […] [providing] stability, uniformity and convenience in the designation of plants […]" (Hitchcock, 1905: 828–829). One of the clearest expositions of the advantages of the type-based approach to nomenclature is still that of Hitchcock (1921).

This early controversy highlights important differences in opinion concerning the utility of *information* (e.g., a description) versus *a sample* (e.g., a specimen) for the purpose of connecting names with taxa. This difference of opinion is in many ways being reprised in the modern discussions that form the basis and context of this paper.

■ WHAT ITEMS CAN SERVE AS A TYPE UNDER THE CURRENT CODE?

In the current *Code*, a "nomenclatural type" is defined as "that element to which the name of a taxon is permanently attached" (Art. 7.2). "Element" is not formally defined, but for the name of a new species or intraspecific taxon published on or after 1 January 2007, the type of a name can only be a specimen (Art. 40.4) or, in the case of microscopic algae or microfungi, an illustration "if there are technical difficulties of specimen preservation or if it is impossible to preserve a specimen that would show the features attributed to the taxon by the author of the name" (Art. 40.5).

A specimen is defined as "a gathering, or part of a gathering, of a single species or infraspecific taxon, disregarding admixtures, mounted either as a single preparation or as more than one preparation with the parts clearly labelled as being part of the same specimen or bearing a single, original label in common." Type specimens must be permanently preserved and may not be living (Art. 8.4). An illustration is defined as "a work of art or a photograph depicting a feature or features of an organism, e.g. a drawing, a picture of a herbarium specimen, or a scanning electron micrograph" (Art. 6.1 footnote 1). DNA sequences are excluded from this definition in Art. 40.5 Ex. 6 on the basis that sequences "are not illustrations under Art. 6.1 footnote because they are not depictions of features of the organism".

■ THE PURPOSE OF TYPES

The principal utility of types is that they reduce, and often effectively eliminate, nomenclatural ambiguity and uncertainty, and this reduction or elimination is prospective (that is, it is more or less future-proof). This is best illustrated by considering the life cycle of many taxa and their names.

Consider a newly published taxon for which adequate specimens can be preserved. A type specimen is designated for the name (as required under Art. 40.4), and in addition a diagnosis or description is provided for the taxon (as required under Art. 39.1). At the time of publication, if the author has done adequate taxonomic due diligence and follows best practice, ambiguity as to the application of the name will be low.

However, a future taxonomist may recognise two or more taxa where the first taxonomist recognised one. Ambiguity as to the correct names of the segregate taxa would be high except that the *Code* unambiguously requires (Principle II) that the name remains with the segregate taxon that includes its type (with some limited exceptions as defined in Art. 14, which allows in some circumstances a change of type or priority of names via conservation). The type is the mechanism by which this potential ambiguity or uncertainty is resolved.

Consider now a counterfactual where descriptions or diagnoses[2] are used for determining the application of names, rather than types. At the time that the second taxonomist subdivides the original taxon into two or more taxa, the original description or diagnosis may or may not have utility for reducing ambiguity: if the traits used by the second taxonomist to delimit the new taxa were not included in the original description or diagnosis, the correct application of the name could not be readily determined. Thus, descriptions and diagnoses have lower utility than types at reducing future uncertainty.

The reason why specimens or samples have more utility than descriptions or diagnoses for reducing ambiguity is that they comprise potentially *unbounded sources of information*,

[2]A description is a more or less complete description of the traits of an organism. A diagnosis is "a statement of that which in the opinion of its author distinguishes the taxon from others." (Art. 32.2)

whereas descriptions and diagnoses comprise *bounded sets of information*. A type specimen may, if well-preserved, be subject to review using new methods of observation or analysis (e.g., closer microscopic scrutiny, scanning electron microscopy, CT-scanning, DNA extraction and sequencing) that may not have been available when it was first designated. Descriptive information, by contrast, cannot be subject to new methods: with a description or diagnosis, "what you see is all you have". It is in this sense that descriptions and diagnoses are bounded. New interpretations of descriptive information are possible, but new information *per se* cannot be created from them.

It is important to note in the context of this discussion paper that the *Code* currently includes an exception to the argument developed above. As noted, microscopic algae and fungi may be typified by illustrations (Art. 40.5) if there are valid reasons why typification on a specimen is impractical. This provision was introduced at the time that typification was otherwise restricted to specimens (Art. 40.4), to accommodate the practical problem that adequate preservation of many microscopic algae and fungi is technically difficult, and preserved type specimens in some cases are less effective at reducing ambiguity than a good illustration. An illustration, however, is also a fully bounded set of information ("what you see is all you have") which may be inadequate in the face of future uncertainty. While it may be subject to closer scrutiny (and may capture important traits that were not considered diagnostic at the time), it cannot be re-analysed. Hence, for practical purposes, the *Code* already includes an important exception from specimen-based typification in some cases.

Of course, there are cases where even a designated type is not fit for the purpose of eliminating ambiguity. The type may be lost or destroyed, or be inadequate due to either mishandling or the generally poor state of preservation achievable for some organismal groups. In such cases the *Code* provides mechanisms for neotypification (Art. 9.8) or epitypification (Art. 9.9) of a name. In essence, these provisions enhance the utility of the type concept in certain boundary cases, but do not extend the concept of a type itself.

■ ADVANTAGES OF TYPIFYING SPECIES ON DNA SEQUENCES

Some members of the Special-purpose Committee consider that allowing DNA sequences to serve as types would be advantageous to taxonomy and nomenclature. The advantages are as follows:

1. *Species could be named that cannot currently be named*. Allowing DNA sequences as types would enable the naming of many microscopic species for which the culturing, isolation, or preparation of physical specimens, or preparation of adequate typifying illustrations, is either technically challenging or impossible with current technologies. Amongst organisms treated as algae, fungi and plants (and hence covered by the *Code*) these include some groups of soil-living and endophytic fungi, symbiotic fungi that co-occur with lichens, some cyanobacteria, some chytrids, and certain microscopic green algae and photosynthetic protists (e.g., Benites & al., 2019). Comparisons of environmental DNA (eDNA) sequencing with traditional culturing techniques using the same samples has shown that many species in these groups are discoverable only through sequencing, as they fail to culture (Seeleuthner & al., 2018; Kalsoom Khan & al., 2020). Such species cannot currently be readily named under the *Code*.

2. *Taxonomic studies of hyperdiverse groups could be accelerated*. Some of the taxonomic groups discussed above (particularly fungi, cyanobacteria and protists) are likely to be, or are known to be, hyperdiverse and very poorly known: that is, they comprise large numbers of species, most of which remain un-named (Lücking & al., 2021). Taxonomy – discovering, delimiting, naming and classifying organisms – is clearly important, including for these poorly studied and hyperdiverse groups. However, for several reasons (including the relatively low numbers of taxonomists who specialise in them), the current rate at which species in these groups are discovered, delimited and named is low compared with the magnitude of the task. One impediment to an acceleration in their documentation is the current requirement in the *Code* for a physical type or illustration, because even if culturing or obtaining adequate pure samples is possible, it is often laborious and time-consuming.

3. *DNA is information-rich*. DNA sequences are richly informative (if sufficiently long and with adequate levels of sequence variation); it is largely for this reason that they have revolutionised the science of taxonomy. Good DNA sequences are often better than morphology at recovering evolutionary and phylogenetic relationships of species and at delimiting species in some taxonomic groups where many species are morphologically cryptic (that is, morphologically conservative) or where a substantial degree of parallel or convergent morphological evolution has occurred (Blaxter & al., 2022, and references therein). It seems sensible to utilise this richness in information content for the purpose of typification as well as for delimitation and classification.

4. *DNA is less plastic than morphology.* DNA provides the foundational information that determines core aspects of a phenotype. A phenotype, however, is influenced by environmental factors and exhibits a degree of plasticity that complicates, and sometimes obscures, relationships and identities of organisms. Types are foundational to taxonomy; it seems sensible then to typify species on foundational traits, such as DNA sequences, rather than to rely on a physical specimen of an individual organism, the phenotype of which has been shaped by the environment in complex, sometimes unpredictable, and sometimes *ad hoc* ways.

5. *DNA sequences are permanent, distributable records*. Being physical objects, type specimens are perishable and may be destroyed, either through poor handling or misadventure. DNA sequences, by contrast, are imperishable (as long as the information media on which they are stored exist). Furthermore, DNA sequences can be readily copied, backed up and distributed freely and electronically anywhere and can

be readily accessed from digital repositories. Type specimens, by contrast, cannot be replicated; while it is good practice to ensure that physical types comprise duplicate series, it is usually impracticable to distribute these to more than a handful of institutions. This creates a barrier to their access and study, particularly for workers in institutions in the Global South and others who are physically distant from major herbaria and other institutions, and have limited resources for travel.

■ DISADVANTAGES OF TYPIFYING SPECIES ON DNA SEQUENCES

Some members of the Special-purpose Committee consider that allowing DNA sequences to serve as types would be disadvantageous to taxonomy and nomenclature. The disadvantages are as follows:

1. *A physical specimen can have only one identity; a DNA sequence has no physical identity.* A physical specimen (excluding admixtures) can only have one identity under any system of circumscription or classification: that is, a physical specimen may belong to one taxon and one taxon only (at the same rank). This is the cardinal value of a type specimen, and it allows the unambiguous correlation of a type with a name (and hence a name with a taxon). Note that the same cannot necessarily be said of an illustration, but an illustration at least depicts a physical organism, which could have had only one identity. By contrast, a DNA sequence is an ordered list of base-pairs; it comprises information and is not an object. This means that the same DNA sequence may be found in multiple organisms comprising multiple taxa, and a sequence may not be unambiguously assignable to one taxon. This is particularly true of nothotaxa or hybrids, which always share portions of their DNA with two or more other taxa, and cannot necessarily be distinguished by any single sequence. It is also true of many widely accepted, short barcode regions. While such sequences may successfully distinguish taxa in many lineages, they may not be discriminatory across multiple species in other lineages (e.g., Pei & al., 2015; Liu & al., 2022). Hence, if a taxon whose name was typified by a sequence, which was believed at the time to be unique, is later split into two taxa, there can be no guarantee that the sequence can be used to unambiguously determine to which segregate the name should apply.

2. *Sequences may contain errors generated by the sequencing method*. Because they comprise information and are not physical entities, DNA sequences are always generated using a method (and usually one or more algorithms that interpret the primary signal generated by the method). Information always carries the risk of misinformation, and any method carries the risk of errors. While many sequencing methods have been shown to be robust and to have low error rates (Stoler & Nekrutenko, 2021), this is not necessarily the case with all. In particular, the methods of most interest for this paper (sequence assembly from environmental DNA samples) have a potentially high, and difficult-to-quantify, risk of misassembly, including the assembly of hybrid or chimeric sequences (e.g., Salomaki & Lane, 2017; Arroyo Mühr & al., 2020). If such errors were to occur, a retrieved sequence may correspond to *no* actual organism or taxon. Typifying taxa and their names using such error-prone sequences runs the risk of accruing "garbage" taxa, with little opportunity for correction or amelioration.

3. *Sequences as types may be prone to scientific fraud and other taxonomic malpractice*. A sequence, in the form of a digital string of base-pairs, can be readily edited; a physical type cannot be. For this reason, using sequences as types opens the possibility of taxonomic malpractice: a dishonest taxonomist may edit diagnostic differences into a type sequence before publishing the sequence and the taxon. This may be done for reasons considered harmless by the author (e.g., "correcting" potentially mis-called bases) or as truly dishonest taxonomic vandalism (e.g., an author downloading a sequence, editing it, then claiming the discovery of a new species). While, of course, the vast majority of taxonomists do not and would not engage in such malpractice, some may. Such edits would be undetectable without considerable work by others (and may indeed be completely undetectable if there is no possibility of independently re-sequencing part of the original sample).

4. *Sequences (particularly long ones) cannot be preserved in physical form*. Most physical types are long-lasting (in the order of multiple centuries at least) with relatively low dependence on technology if minimal archival conditions are maintained. Short DNA sequences could potentially be durably preserved by printing on paper, but long sequences, especially genomic sequences, cannot reasonably be printed and must be stored as digital files. This means their persistence for long periods is uncertain and is dependent upon the persistence, continued dedication to that purpose, and continued broad accessibility of sophisticated digital resources that are not under the control of taxonomists and their institutions. Of course, this risk can be ameliorated by digital duplication and backup, which cannot be applied to physical specimens; under any scenario in which DNA sequences are allowed as types, lodgment standards (e.g., deposition in a registered, secure sequence repository) will be needed to reduce the risk of inadvertent loss.

5. *A mixed typification model that allows both DNA and specimen types would be problematic*. Many of the taxonomic groups in which DNA sequences might be used as types, if allowed, include at least some species that can be, and are, typified conventionally. If DNA sequences were allowable as types in such groups, but sequences were not mandated as types, it is likely that some species would be typified using a sequence (and no specimen type) while some would be typified using a specimen (with no sequence). In such cases, it would become impossible to ascertain whether a new taxon (typified using a sequence or a specimen) has already been named. A core principle of the *Code* (Pre. 1) is that nomenclature must reduce ambiguity, but ambiguity would be a necessary consequence, at least at times, of any mixed typification model.

6. *DNA methods are not universally available to all taxonomists*. Sequencing is a costly technology, particularly long-read and genomic sequencing. While technological improvements are progressively reducing the cost per base-pair, new and improved (and more expensive) methods that result in longer reads are continually being developed, so costs at the cutting edge remain high. Cheap technologies that provide short-read sequences are more widely available, but short sequences are more problematic as types than long sequences. By contrast, typification using physical specimens is available to all (as long as a physical specimen can be obtained) and the later comparison by other workers of a taxon in hand to a physical type specimen (or an image of a physical type) requires no special equipment. There is a danger that the wide adoption of DNA sequences as types may exacerbate the digital divide and further disadvantage already disadvantaged sections of the taxonomic community. If DNA sequences were to be accepted both as types and as diagnoses, and become widespread (particularly if used, for convenience of authors, for species that could be typified and diagnosed more conventionally) then taxa will become increasingly inaccessible to workers without access to sequencing facilities or methods, particularly those in institutions in the Global South. (Note, however, that, for the types of organisms for which DNA typification appears most appropriate, especially unculturable or difficult-to-culture microbes, the technology and cost barriers for creating pure samples for specimen typification are already high, and a technology divide already exists.)

■ OPTIONS FOR DISCUSSION

Two broad classes of option are available to the Special-purpose Committee: *make no recommendations* and continue with business-as-usual nomenclature under the current provisions of the *Code* (which would continue a situation where many discovered species cannot be named), or *recommend amendments* which, if adopted, would change the rules of nomenclature, leading to a change in the way that taxonomy of some or all groups is carried out and potentially bringing about a rapid increase in the number of DNA-typified species.

A subsidiary set of options is to recommend changes for some organismal groups or some circumstances (such as organisms for which physical types are impossible) but not for others (such as organisms for which designating a physical type is straightforward).

The Committee notes that the governance of provisions in the *Code* that apply only to the names of organisms treated as fungi (dealt with in Chapter F of the *Code*) is delegated to the International Mycological Congress (IMC). At the San Juan IMC in 2018, a parallel Special-purpose Committee was established to consider these issues; this Committee will report to the 2024 IMC in Maastricht, The Netherlands. Hence, an outcome may be that amendments are made to Chapter F following the 2024 IMC but not to the remainder of the *Code*. This would result in taxonomic groups that are microscopic and unable to be typified by specimens, but are not fungi, remaining with the *status quo*.

The International Commission on the Taxonomy of Fungi has published a discussion paper (Lücking & al., 2021) that outlines five potential options for that community (see their table 1). Options 1, 2 and 5 make no recommendations to alter the *Code*, with option 1 a business-as-usual option, option 5 essentially the establishment of a parallel nomenclature outside the *Code* for fungal taxa that cannot be typified under the *Code*, and option 2 a recommendation to use existing provisions under the *Code* that allow for images as types. Only options 3 and 4 involve amendments to the *Code*: option 3 would allow formal binomials to be established, with priority, based on sequence types, while option 4 would establish a system of provisional names, without priority, analogous to Candidatus names as used in the *International Code of Nomenclature of Prokaryotes* (Parker & al., 2019).

As discussed in the Types and typification section of this paper, a key choice needs to be made as to whether there is utility in revisiting the agreement struck in the early history of the formal *Code* that specimens and not information are obligatory for typification. That discussion contrasted physical types with morphological descriptions, and determined that physical types are superior. However, while DNA sequences are analogous to descriptions in many ways (they both comprise information only), sequences potentially have much richer information content than descriptions; also, at the time of the decision to require a physical type there was little comprehension of the existence of vast numbers of "dark taxa" (unculturable micro-organisms), and, of course, no technology platforms for sequencing DNA.

This paper is the first of two or three planned by the Special-purpose Committee on DNA Sequences as Types. The next will provide and discuss candidate recommendations for amendment of the *Code*; if the view of the Committee is that some of these should proceed to a formal vote as part of the Madrid IBC, then a subsequent paper will make formal recommendations.

■ AUTHOR CONTRIBUTIONS

KRT led the drafting of the manuscript, and all other authors contributed text and comments. — KRT, https://orcid.org/0000-0002-6658-6636; SSR, https://orcid.org/0000-0003-3704-0703; AAD, https://orcid.org/0000-0001-6695-6177; QG, https://orcid.org/0000-0002-0596-5376; SL, https://orcid.org/0000-0001-6235-2026; CAM, https://orcid.org/0000-0003-0495-7064; VM, https://orcid.org/0000-0002-9350-3306; HSY, https://orcid.org/0000-0001-9507-0105

■ ACKNOWLEDGMENTS

Quentin Groom is funded by the BiCIKL project (European Union's Horizon 2020 Research and Innovation Action under grant agreement No 101007492) Open access publishing facilitated by The University of Western Australia, as part of the Wiley - The University of Western Australia agreement via the Council of Australian University Librarians.

■ LITERATURE CITED

Arroyo Mühr, L.S., Lagheden, C., Hassan, S.S., Kleppe, S.N., Hultin, E. & Dillner, J. 2020. De novo sequence assembly requires bioinformatic checking of chimeric sequences. *PLoS One* 15(8): e0237455. https://doi.org/10.1371/journal.pone.0237455

Benites, L.F., Poulton, N., Labadie, K., Sieracki, M.E., Grimsley, N. & Piganeau, G. 2019 Single cell ecogenomics reveals mating types of individual cells and ssDNA viral infections in the smallest photosynthetic eukaryotes. *Philos. Trans., Ser. B.* 374: 20190089. https://doi.org/10.1098/rstb.2019.0089

Blaxter, M., Archibald, J.M., Childers, A.K., Coddington, J.A., Crandall, K.A., Di Palma, F., Durbin, R., Edwards, S.V., Graves, J.A.M., Hackett, K.J., Hall, N., Jarvis, E.D., Johnson, R.N., Karlsson, E.K., Kress, W.J., Kuraku, S., Lawniczak, M.K.N., Lindblad-Toh, K., Lopez, J.V., Moran, N.A., Robinson, G.E., Ryder, O.A., Shapiro, B., Soltis, P.S., Warnow, T., Zhang, G. & Lewin, H.A. 2022. Why sequence all eukaryotes? *Proc. Natl. Acad. Sci. U.S.A.* 119(4): e2115636118. https://doi.org/10.1073/pnas.2115636118

De Beer, Z.W., Marincowitz, S., Duong, T.A., Kim, J.J., Rodrigues, A. & Wingfield, M.J. 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(11): 1323–1340. https://doi.org/10.1016/j.funbio.2016.07.004

Gray, A. 1864. Nomenclature. *Amer. J. Sci. Arts*, ser. 2, 37: 278–281.

Hawksworth, D.L. (ed.) 2018. Proposals for consideration at IMC11 to modify provisions solely related to fungi in the *International Code of Nomenclature for algae, fungi, and plants*. *IMA Fungus* 9: (i)–(vii). https://doi.org/10.1007/BF03449481

Hawksworth, D.L., Hibbett, D.S., Kirk, P.M. & Lücking, R. 2016. (308–310). Proposals to permit DNA sequence data to serve as types of names of fungi. *Taxon* 65: 899–900. https://doi.org/10.12705/654.31

Hedlund, B.P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K.T., Murray, A.E., Palmer, M., Parks, D.H., Probst, A.J., Reysenbach, A.-L., Rodriguez-R., L.M., Rossello-Mora, R., Sutcliffe, I.C., Venter, S.N. & Whitman, W.B. 2022. SeqCode: A nomenclatural code for prokaryotes described from sequence data. *Nature, Microbiol.* 7: 1702–1708. https://doi.org/10.1038/s41564-022-01214-9

Hitchcock, A.S. 1905. Nomenclatural type specimens of plant species. *Science* 21: 828–832. https://doi.org/10.1126/science.21.543.828

Hitchcock, A.S. 1921. The type concept in systematic botany. *Amer. J. Bot.* 8: 251–255. https://doi.org/10.1002/j.1537-2197.1921.tb05622.x

Kalsoom Khan, F., Kluting, K., Tångrot, J., Urbina, H., Ammunet, T., Sahraei, S.E., Rydén, M., Ryberg, M. & Rosling, A. 2020. Naming the untouchable – Environmental sequences and niche partitioning as taxonomical evidence in fungi. *IMA Fungus* 11: 23. https://doi.org/10.1186/s43008-020-00045-9

Lindon, H.L., Hartley, H., Knapp, S., Monro, A.M. & Turland, N.J. 2020. XIX International Botanical Congress, Shenzhen: Report of the Nomenclature Section, 17th to 21st July 2017. *PhytoKeys* 150: 1–276. https://doi.org/10.3897/phytokeys.150.50687

Linnaeus, C. 1736. *Fundamenta botanica*, 1st ed. Amstelodami [Amsterdam]: apud Salomonem Schouten. https://bibdigital.rjb.csic.es/idurl/1/11535

Liu, Z.-F., Ma, H., Zhang, X.-Y., Ci, X.-Q., Li, L., Hu, J.-L., Zhang, C.-Y., Xiao, J.-H., Li, H.-W., Conran, J.G., Twyford, A.D., Hollingsworth, P.M. & Li, J. 2022. Do taxon-specific DNA barcodes improve species discrimination relative to universal barcodes in Lauraceae? *Bot. J. Linn. Soc.* 199(4): 741–753. https://doi.org/10.1093/botlinnean/boab089

Lücking, R. & Moncada, B. 2017. Dismantling *Marchandiomphalina* into *Agonimia* (*Verrucariaceae*) and *Lawreymyces* gen. nov. (Corticiaceae): Setting a precedent to the formal recognition of thousands of voucherless Fungi based on type sequences. *Fungal Diversity* 84: 119–138. https://doi.org/10.1007/s13225-017-0382-4

Lücking, R., Kirk, P.M. & Hawksworth, D.L. 2018. Sequence-based nomenclature: A reply to Thines et al. and Zamora et al. and provisions for an amended proposal "from the floor" to allow DNA sequences as types of names. *IMA Fungus* 9: 185–198. https://doi.org/10.5598/imafungus.2018.09.01.12

Lücking, R., Aime, M.C., Robbertse, B., Miller, A.N., Aoki, T., Ariyawansa, H.A., Cardinali, G., Crous, P.W., Druzhinina, I.S., Geiser, D.M., Hawksworth, D.L., Hyde, K.D., Irinyi, L., Jeewon, R., Johnston, P.R., Kirk, P.M., Malosso, E., May, T.W., Meyer, W., Nilsson, H.R., Öpik, M., Robert, V., Seifert, K.A., Stadler, M., Thines, M., Vu, D., Yurkov, A.M., Zhang, N. & Schoch, C.L. 2021. Fungal taxonomy and sequence-based nomenclature. *Nature, Microbiol.* 6: 540–548. https://doi.org/10.1038/s41564-021-00888-x

May, T.W. & Miller, A.N. 2018. XI International Mycological Congress: Guiding Vote on nomenclature proposals to amend Chapter F of the *International Code of Nomenclature for algae, fungi, and plants*. *IMA Fungus* 9: (xv)–(xxi). https://doi.org/10.1007/BF03449447

May, T.W. & Redhead, S.A. 2018. Synopsis of proposals on fungal nomenclature: A review of the proposals concerning Chapter F of the *International Code of Nomenclature for algae, fungi, and plants* submitted to the XI International Mycological Congress, 2018. *IMA Fungus* 9(1): (ix)–(xiv). https://doi.org/10.1007/BF03449482

May, T.W., Redhead, S.A., Lombard, L. & Rossman, A.Y. 2018. XI International Mycological Congress: Report of Congress action on nomenclature proposals relating to fungi. *IMA Fungus* 9: (xxii)–(xxvii). https://doi.org/10.1007/BF03449448

Nicolson, D.H. 1991. A history of botanical nomenclature. *Ann. Missouri Bot Gard.* 78: 33–56. https://doi.org/10.2307/2399589

Nomenclature Commission 1907. American Code of Botanical Nomenclature. *Bull. Torrey Bot. Club* 34: 167–178. https://doi.org/10.2307/2479237

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C. & Warinner, C. 2021. Ancient DNA analysis. *Nat. Rev., Meth. Primers* 1: 14. https://doi.org/10.1038/s43586-020-00011-0

Parker, C.T., Tindall, B.J. & Garrity, G.M. (eds.) 2019. International Code of Nomenclature of Prokaryotes: Prokaryotic Code (2008 Revision). *Int. J. Syst. Evol. Microbiol.* 69(1A): S1–S111. https://doi.org/10.1099/ijsem.0.000778

Pei, N., Erickson, D., Chen, B., Ge, X., Mi, X., Swenson, N.G., Zhang, J.-L., Jones, F.A., Huang, C.-L., Ye, W., Hao, Z., Hsieh, C.-F., Lum, S., Bourg, N.A., Parker, J.D., Zimmerman, J.K., McShea, W.J., Lopez, I.C., Sun, I-F., Davies, S.J., Ma, K. & Kress, W.J. 2015. Closely-related taxa influence woody species discrimination via DNA barcoding: Evidence from global forest dynamics plots. *Sci. Rep.* 5: 15127. https://doi.org/10.1038/srep15127

Renner, S.S. 2021. Proposals to permit nuclear DNA sequences as nomenclatural types when preservation of specimens is not feasible. *Taxon* 70(6): 1380–1381. https://doi.org/10.1002/tax.12607

Salomaki, E.D. & Lane, C.E. 2017. Red algal mitochondrial genomes are more complete than previously reported. *Genome Biol.* 9(1): 48–63. https://doi.org/10.1093/gbe/evw267

Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., Leconte, J., Mangot, J.-F., Poulain, J., Labadie, K., Logares, R., Sunagawa, S., de Berardinis, V., Salanoubat, M., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Tara Oceans Coordinators, Pesant, S., Poulton, N., Stepanauskas, R., Bork, P, Bowler, C.,

Hingamp, P., Sullivan, M.B., Iudicone, D., Massana, R., Aury J.-M., Henrissat, B., Karsenti, E., Jaillon, O., Sieracki, M., de Vargas, C. & Wincker, P. 2018. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nature, Commun.* 9: 310. https://doi.org/10.1038/s41467-017-02235-3

Stoler, N. & Nekrutenko, A. 2021. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics Bioinf.* 3(1): lqab019. https://doi.org/10.1093/nargab/lqab019

Thines, M., Crous, P.W., Aime, M.C., Aoki, T., Cai, L., Hyde, K.D., Miller, A.N., Zhang, N. & Stadler, M. 2018. Ten reasons why a sequence-based nomenclature is not useful for fungi anytime soon. *IMA Fungus* 9: 177–183. https://doi.org/10.5598/imafungus.2018.09.01.11

Turland, N.J. & Wiersema, J.H. 2017. Synopsis of Proposals on Nomenclature – Shenzhen 2017: A review of the proposals concerning the *International Code of Nomenclature for algae, fungi, and plants* submitted to the XIX International Botanical Congress. *Taxon* 66(1): 217–274. https://doi.org/10.12705/661.36

Turland, N.J., Wiersema, J.H., Barrie, F.R., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T.W., McNeill, J., Monro, A.M., Prado, J., Price, M.J. & Smith, G. F. (eds.) 2018. *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books. https://doi.org/10.12705/Code.2018

Wilson, K.L. 2019. Report of the General Committee: 21. *Taxon* 68(1): 160–162. https://doi.org/10.1002/tax.12011

Witteveen, J. 2016. Suppressing synonymy with a homonym: The emergence of the nomenclatural type concept in nineteenth century natural history. *J. Hist. Biol.* 49: 135–189. https://doi.org/10.1007/s10739-015-9410-y

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X. & Xiang, M. 2019. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. *Mycology* 10(3): 127–140. https://doi.org/10.1080/21501203.2019.1614106

Zamora, J.C., Svensson, M., Kirschner, R., Olariaga, I., Ryman, S., Parra, L.A., Geml, J.. Rosling, A., Adamčík, S. […] Ekman, S. 2018. Considerations and consequences of allowing DNA sequence data as types of fungal taxa. *IMA Fungus* 9: 167–175. https://doi.org/10.5598/imafungus.2018.09.01.10