Syst. Biol. 73(2):486–494, 2024
Published by Oxford University Press on behalf of the Society of Systematic Biologists 2023. This work is written by (a) US Government employee(s) and is in the public domain in the US.
https://doi.org/10.1093/sysbio/syad068
Advance Access Publication November 13, 2023

Improving the gold standard in NCBI GenBank and related databases: DNA sequences from type specimens and type strains

Susanne S. Renner[1,*], Mark D. Scherz[2,*], Conrad L. Schoch[3,*], Marc Gottschling[4,*], and Miguel Vences[5,*]

[1]*Department of Biology, Washington University, Saint Louis, MO 63130, USA*
[2]*Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, Copenhagen 2100, Denmark*
[3]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA*
[4]*Faculty of Biology, GeoBio-Center, Ludwig-Maximilians-University, Munich 80333, Germany*
[5]*Division of Evolutionary Biology, Zoological Institute, University of Technology, Mendelssohnstr. 4, 38106 Braunschweig, Germany*
[*]*Correspondence to be sent to: Department of Biology, Washington University, Saint Louis, MO 63130, USA. E-mail: srenner@wustl.edu; Natural History Museum of Denmark, Universitetsparken 15, Copenhagen 2100, Denmark. E-mail: mark.scherz@gmail.com; National Center for Biotechnology Information, Bethesda, MD 20894, USA. E-mail: schoch2@ncbi.nlm.nih.gov; Faculty of Biology, Ludwig-Maximilians-University, Munich 80333, Germany. E-mail: gottschling@biologie.uni-muenchen.de; Zoological Institute, Technical University Braunschweig, Braunschweig 38106, Germany. E-mail: mvences@mvences.de, m.vences@tu-braunschweig.de*

Received 16 May 2023; reviews returned 21 August 2023; accepted 11 November 2023
Associate Editor: Alexandre Antonelli

Abstract.—Scientific names permit humans and search engines to access knowledge about the biodiversity that surrounds us, and names linked to DNA sequences are playing an ever-greater role in search-and-match identification procedures. Here, we analyze how users and curators of the National Center for Biotechnology Information (NCBI) are flagging and curating sequences derived from nomenclatural type material, which is the only way to improve the quality of DNA-based identification in the long run. For prokaryotes, 18,281 genome assemblies from type strains have been curated by NCBI staff and improve the quality of prokaryote naming. For Fungi, type-derived sequences representing over 21,000 species are now essential for fungus naming and identification. For the remaining eukaryotes, however, the numbers of sequences identifiable as type-derived are minuscule, representing only 739 species of arthropods, 1542 vertebrates, and 125 embryophytes. An increase in the production and curation of such sequences will come from (i) sequencing of types or topotypic specimens in museum collections, (ii) the March 2023 rule changes at the International Nucleotide Sequence Database Collaboration requiring more metadata for specimens, and (iii) efforts by data submitters to facilitate curation, including informing NCBI curators about a specimen's type status. We illustrate different type-data submission journeys and provide best-practice examples from a range of organisms. Expanding the number of type-derived sequences in DNA databases, especially of eukaryotes, is crucial for capturing, documenting, and protecting biodiversity. [Best-practice examples; curation; data submission; GenBank; museomics; nomenclatural types; taxonomy.]

DNA sequences have become increasingly important in identifying unnamed or ambiguously named specimens. This is most often achieved by comparisons with sequences available in the European Nucleotide Archive (Burgin et al. 2023), the DNA Databank of Japan (DDBJ; Tanizawa et al. 2023), and GenBank, the genetic sequence database of the National Center for Biotechnology Information (NCBI; Sayers et al. 2023). All three databases are part of the International Nucleotide Sequence Database Collaboration (INSDC; Arita et al. 2021) and exchange data on a daily basis. Only NCBI, however, employs a team of taxonomy curators who maintain a standalone set of resources, collectively referred to as NCBI Taxonomy (Schoch et al. 2020). All INSDC partners utilize this taxonomic resource, together with versions of the basic local alignment search tool, BLAST (Altschul et al. 1990), and support the principle that data should be findable, accessible, interoperable, and reusable, a goal known as the FAIR data principle (Wilkinson et al. 2016).

Reliable BLAST-based taxon identification in these databases is crucial for modern biology, especially given the metabarcoding approaches that are now employed routinely for organisms from insect traps, water, and soil samples (Pawlowski et al. 2012; Hausmann et al. 2016; Miller et al. 2016). In many cases, however, BLAST searches will match sequences in the databases whose taxonomic assignment is incomplete, incorrect, or outdated, thus propagating taxonomic error. An increase in the quality of taxon identification achieved with the INSDC databases is possible in principle by examining voucher or strain material cited by the authors who initially submitted the sequence. However, this is rarely feasible because of time and funding constraints.

The only other approach for improving the reliability of identification via BLAST searches is to focus on the addition and curation of high-value reference sequences to the database provided users can recognize such reference sequences via a flagging system. In terms of taxonomy and nomenclature, the highest-value DNA sequences are those coming from nomenclatural name-bearing type material or material seen or examined by the original author(s) of a name. A sequence derived from a type is by definition correctly named even when changing taxonomic views requires transferring a name into a different genus or ranking a taxon differently. Having high-value,

curated—that is, recognizable and literature-supported—sequences from types as references is important for all clades in the Tree of Life. In prokaryotes, this typically means sequencing full genomes from cultured-type strains. Types of eukaryotes, however, are usually preserved specimens (as stipulated by relevant Codes of Nomenclature) and unlikely to yield high-quality DNA; most type-derived sequences for eukaryotes, therefore, are short barcode sequences from mitochondrial DNA, plastid DNA or intronless nuclear markers, such as ribosomal RNA (rRNA) including internal transcribed spacer (ITS) sequences.

In this study, we analyze how users and NCBI staff have historically flagged and curated type-derived sequences and how procedures might be improved to increase the production and curation of such sequences. We begin with a brief history of the curation of type-derived sequences, an option introduced in 1998, and then explain the processes involved in curating type-derived sequences, with specific examples from Fungi, animals, plants and dinophytes/dinoflagellates, one of several lineages with species treated variously under the botanical and the zoological codes of nomenclature.

The curation of type-derived sequences in GenBank: initial focus on prokaryotes

The importance of identifying and labeling (flagging) GenBank sequences derived from type material has long been recognized (Chakrabarty 2010, 2013; Harrison et al. 2011; Federhen 2015; Robbertse et al. 2017; Kannan et al. 2023). Beginning in 2018, NCBI Taxonomy introduced and expanded a set of taxonomic resources in which type metadata can be added and the synonymy of names handled more comprehensively (Schoch et al. 2020). To ensure validation by data curators and thereby high reliability, the assignment of a sequence to type material involves several steps (Fig. 1): Submitters need to provide voucher modifiers, such as the museum (or other physical repository) catalog number and, ideally, a reference to the publication that includes the new scientific name, the original description or diagnosis of the species, and the type designation (known as protologue to botanists but not zoologists), preferably along with the respective portable document file (PDF) or link to the place of publication. If the voucher details and assigned taxonomic name match, the type material will be displayed on the relevant pages in the NCBI TaxBrowser and in the typematerial.dmp file as part of the taxdump ftp files (https://ftp.ncbi.nlm.nih.gov/pub/taxonomy/new_taxdump/; Schoch et al. 2020). Additional type metadata might be sent to the GenBank team later (gb-admin@ncbi.nlm.nih.gov).

NCBI further provides a separate set of reference resources, RefSeq, which is curated by NCBI staff and selected from GenBank records (O'Leary et al. 2016). This includes targeted marker resources under RefSeq Targeted Loci; curated sets of markers, which include collections for Bacteria, Archaea, Oomycota, and Fungi

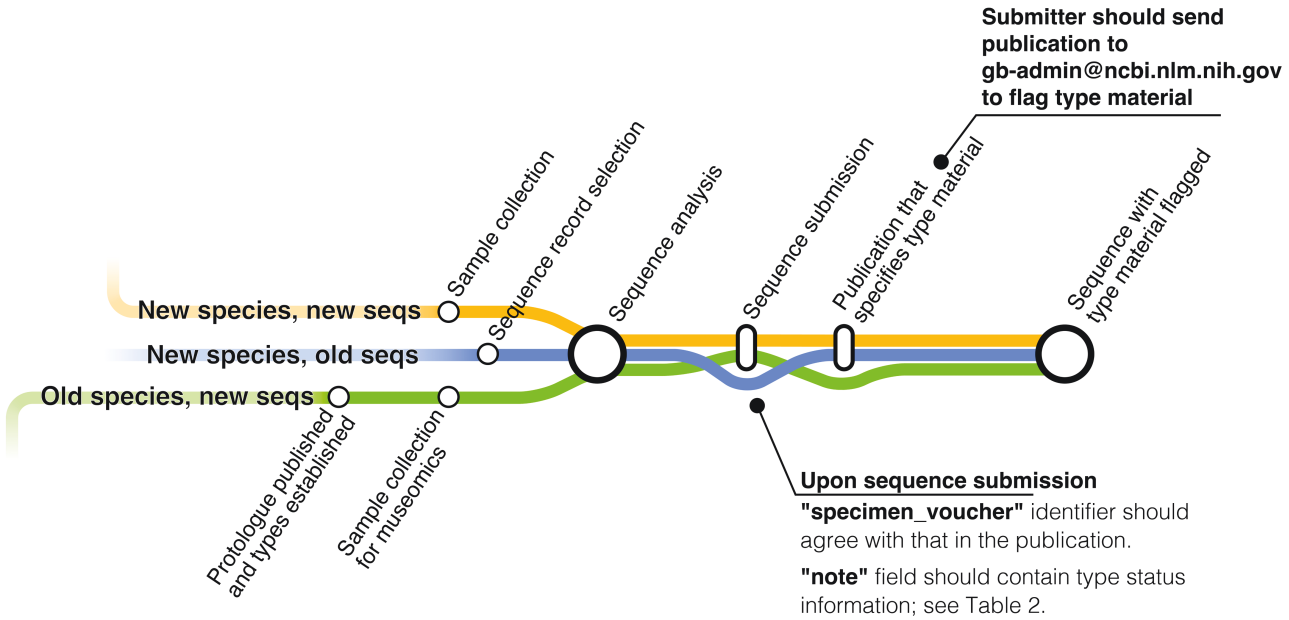

Figure 1. Data journeys for annotation of type-derived sequences from different sources. Regardless of their origins, sequence data from type specimens proceed through some shared steps towards final annotation as types in INSDC databases. The topmost track indicates the procedure for newly produced sequences from newly described taxa; the middle track indicates sequences already in INSDC that belong to types included in the description of a new taxon; and the lowermost track indicates newly produced sequences from already-published type specimens. In all cases, the publication that specifies the type material, usually the first, valid description or diagnosis of the taxon, must be sent to gb-admin@ncbi.nlm.nih.gov, so that the "type_material" qualifier can be added by NCBI Taxonomy team. Note that, even for historical names, it can be useful to provide the species description to the NCBI Taxonomy team.

(https://www.ncbi.nlm.nih.gov/bioproject/224725). This system is particularly suitable for prokaryotes whose taxonomy relies on living cell cultures or strains maintained in culture collections. Already in the early 2010s, staff at NCBI began curating type-based DNA sequences, a project for which NCBI's taxonomy curators focused on prokaryotes because "the description of a new species [of a prokaryote…] must include the designation of a type strain (see Rule 18a), and a viable culture of that strain must be deposited in at least two publicly accessible culture collections in different countries from which subcultures must be available" (Parker et al. 2019). The community of prokaryote researchers is, therefore, advanced in terms of using genetic data for taxonomy and nomenclature (Federhen 2015) to the point of accepting genome sequences as nomenclatural types by part of the community (Hedlund et al. 2022). The SeqCode Registry for "type genomes" is operational, and descriptions formulated under either the Code of Nomenclature of Prokaryotes (ICNP) or the SeqCode are accepted by a relevant journal, *Systematic and Applied Microbiology*. At the time of this writing, NCBI Taxonomy still only treats prokaryote names under the ICNP, but RefSeq has begun to include selected genome assemblies obtained from environmental samples without explicitly stored vouchers.

As a result of the years-long curation effort, NCBI has established methods and procedures to keep track of different kinds of types, which led to the introduction of new terms to the INSDC-controlled vocabulary (https://www.insdc.org/submitting-standards/controlled-vocabulary-typematerial-qualifer/). Since at least 2018, NCBI curators have improved taxon assignments of prokaryote genomes by comparing them to data from type strains (Ciufo et al. 2018). For this, they utilize the original descriptions, facilitated in the case of prokaryotes by bacterial names having to be published in a designated journal (Federhen 2015). Additionally, they rely on open data from external databases, such as BacDive (Reimer et al. 2022). By early 2023, 21,000 genome assemblies from prokaryote-type strains had been verified and used to update the taxonomy of over 1.1 million GenBank genomes, which led to over 1800 existing genomes in GenBank being assigned a different species name (Kannan et al. 2023). The curated data permit submitters to verify the accuracy of taxonomic assignments of prokaryote genome data before submission or detect contamination from foreign organisms.

The situation for type-derived sequences of eukaryotes

Among eukaryotes, Fungi are the group with by far the largest and best-curated set of type-derived DNA sequences. Thus, a curated set of fungal ITS records (Robbertse et al. 2017) already contains more than 16,000 entries with verified type-derived sequences for a similar number of species (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA177353/). It also includes downloadable ftp files and BLAST interfaces. Specimen records can be linked from GenBank via hotlinks using URL rules curated as part of NCBI BioCollections (Sharma et al. 2018); for instance, a sequence record of *Wickerhamiella versatilis* links directly to the strain, from which the metabolically inactive type has been prepared, at the Japan Collection of Microorganims (https://www.ncbi.nlm.nih.gov/nuccore/NG_063437.1). Another option is third-party controlled LinkOuts. For example, a DNA record of *Cortinarius wiebeae* (https://www.ncbi.nlm.nih.gov/nuccore/KF732479) links directly to the holotype specimen record of MyCoPortal (https://www.mycoportal.org/portal/).

No matter the future of possibly permitting DNA sequences as nomenclatural types (a possibility currently being discussed, with proposals for plants and Fungi due by 31 December 2023; Thiele et al. 2023a, 2023b, 2023c), the reliable detectability of type-derived sequences is crucial for taxonomy and identification. For Fungi, matters are facilitated by the existence of three official registries of fungal names, MycoBank (Robert et al. 2013; https://www.mycobank.org/); Index Fungorum (http://www.indexfungorum.org), and FungalNames (Wang et al. 2023; https://nmdc.cn/fungalnames). As of 1 January 2019, any new type designation for a fungal taxon at or below the rank of species must be registered and the identifier cited with the typification act (May et al. 2019), which should further facilitate automatic (algorithm-based) detection of nomenclatural types. So far, however, this step is not optimized and relies on informal contacts by NCBI curators with data managers at these registries.

The Entrez query "(sequence from type[filter] OR sequence from synonym type[filter]) AND Fungi[organism]" will list all fungal taxonomic entries (from all genes and genetic markers) with type material attributes (Schoch et al. 2014). On 6 March 2023, this revealed 21,607 different Fungi with type-derived sequences, usually from cultured material and mostly from the nuclear ITS region, the universal DNA barcode marker for Fungi (Schoch et al. 2012; our Table 1).

Type-derived sequences from non-culturable eukaryotes have been much slower to start appearing in GenBank. For example, sequences from only 2305 non-fungal eukaryotes, including 1059 animals, 550 green plants, and 108 red algae were flagged as originating from types as of January 2015 (Federhen 2015). Part of the reason for this is that until recently it was difficult to obtain good DNA from dried or liquid-preserved specimens, which limited type-derived sequences to new species based on more or less fresh material. With improved methods for obtaining DNA from museum specimens, an approach now often called museomics (Raxworthy and Smith 2021), the number of DNA sequences originating from nomenclatural types for taxa other than prokaryotes is beginning to increase.

Table 1 provides the numbers of published names and type-derived DNA sequences representing different

Table 1. Species names in the Catalogue of Life (https://www.catalogueoflife.org/) compared to the numbers of species in the NCBI taxonomy database with information on type material, with or without sequences from such material as of 6 March 2023.

Taxon (NCBI Taxonomy name)	Species numbers in the "Catalogue of Life"	Species with information on type material, with or without sequences from types (types of taxa considered heterotypic synonyms by NCBI Taxonomy in parentheses)	Species with sequences from type material (sequences from types of taxa considered heterotypic synonyms by NCBI Taxonomy in parentheses)	Total sequence records from type material in INSDC (GenBank Nucleotide)
Fungi	303,640	24,465 (423)	21,607 (272)	299,735
True yeasts (Saccharomycotina)	3381	1252 (39)	1203 (31)	51,379
Filamentous ascomycetes (Pezizomycotina)	187,142	14,596 (289)	13,166 (180)	157,668
Basidiomycetes (Basidiomycota)	110,543	7925 (89)	6661 (192)	28,563
Plants (Viridiplantae)	1389,061	4217 (43)	440 (24)	106,607
Algae; numbers are only for Green algae (Chlorophyta)	Not calculated	432 (38)	299 (24)	106,109
Embryophytes (Embryophyta)	1,388,880	3761 (5)	125 (0)	467
Animals (Metazoa)	2,683,449	12,134 (33)	3185 (11)	16,899
Vertebrates (Vertebrata)	212,654	9752 (27)	1542 (9)	8720
Amphibians (Amphibia)	22,539	815	565	4106
Squamates (Squamata)	28,028	6006	564	2312
Bony fishes (Actinopterygii)	92,000	2360	284	1519
Mammals (Mammalia)	22,182	148	101	672
Turtles (Testudines)	1724	263	11	54
Cartilaginous fishes (Chondrichthyes)	4232	112	10	37
Birds (Aves)	41,496	13	6	14
Invertebrates (defined as Metazoa NOT Vertebrata)	Not calculated	2382 (5)	1673 (2)	8212
Arthropods (Arthropoda)	2,007,822	1000	739	3717
Insects (Hexapoda)	1,739,089	654	503	2426
Mollusks (Mollusca)	262,542	511	343	1704
Gastropods (Gastropoda)	205,471	446	307	1538
Annelid worms (Annelida)	28,798	294	183	1176
Cnidarians (Cnidaria)	28,874	175	125	599
Flatworms (Platyhelminthes)	34,516	150	83	462
Sponges (Porifera)	20,603	145	105	280
Echinoderms (Echinodermata)	18,190	30	23	127
Ribbon worms (Nemertea)	2500	17	33	71
Nematodes (Nematoda)	25,659	33	9	25
Other eukaryotes (defined as Eukaryota NOT (Metazoa OR Viridiplantae OR Fungi))	Not calculated	887 (2)	511 (2)	32,762
Oomycota	1674	295	236	4696

Note: We could not determine the species number for "algae" due to the polyphyly of this group, which encompasses Chlorophyta (green algae [for which we give numbers]), Phaeophyta (brown algae), Rhodophyta (red algae), Chromista, Alveolata, and other deeply diverged branches of the Tree of Life. Alveolata include Dinophyceae, a group for which we have updated NCBI information on holotypes and epitypes (Supplementary Data Set 1).

species for various groups of eukaryotes. For most groups, the number of type-derived sequences (of different species) has increased around 10-fold since 2015 when Federhen (2015) first counted them, although birds, turtles, and cartilaginous fishes have lagged conspicuously. The near absence of bird-type sequences is particularly conspicuous, especially since many bird sequences in GenBank appear to be misidentified (Van den Burg and Vieites 2023).

For invertebrates, the current numbers of type-derived DNA sequences (Table 1) constitute proportionally enormous increases compared to 2015, when there were 141 type-derived sequences for arthropods, 15 for cnidarians, 12 for flatworms, 4 for echinoderms, 2 for mollusks, 1 from a horsehair worm, and 1 from an annelid (Federhen 2015). Relative to the described species diversity in many of these groups, however, the number of sequences from types remains extremely low. The situation for green plants is even worse than that for animals because plants' cellulose walls until recently made it extremely difficult to obtain good DNA from old specimens, which includes most type specimens (Table 1).

Finding type-derived GenBank sequences can be challenging, as is obvious from the higher numbers of names with type information in NCBI (Table 1, column 2) vs. names with sequences from type material (Table 1, column 3). These discrepancies, as well as the low numbers of sequences flagged as type-derived in some taxa appear mostly due to three factors: (i) A scarcity of curated databases of names and type specimens that prevents the automatic flagging of type sequences by NCBI taxonomy curators, a problem that is exacerbated by the different ways in which mycologists, zoologists, and botanists cite their "specimen_voucher" and "isolate" fields (a topic taken up further in the next section). (ii) Submitters failing to inform NCBI taxonomy curators that a voucher specimen corresponds to a type, perhaps due to ignorance about the proper filling-out of relevant data fields (see Chakrabarty et al. 2013 for a test case focusing on fish taxonomy papers with DNA data). And (iii) imprecise or incorrect use of organism names during submission, a problem that already affects thousands of names (Garg et al. 2019; next section).

Increasing the Number of High-Value Type-Derived Eukaryote Sequences in GenBank

Taxonomy was first indexed in GenBank's Entrez tool in 1993—at the time registering just over 5000 species with scientific names (Federhen 2015). It took some 10 years to reach 100,000 named species, 5 more years to reach 200,000, and another 5 years to reach 300,000 (Federhen 2015). Currently, the total number of named species approaches 550,000, roughly one-eighth of the ca. 4.5 million non-viral, specific and infraspecific names present in the Catalogue of Life, of which roughly half are considered synonyms. Obtaining DNA sequences from the remaining hundreds of thousands of named species will increasingly depend on using specimens already in collections because of the legal, logistic, and financial difficulty of collecting organisms in tropical countries, where much of the known and unknown biodiversity resides. This is not to disregard the fact that alpha-taxonomic studies of vertebrates, plants, and insects to date rely mostly on non-molecular data (Miralles et al. 2020) and will continue to do for the foreseeable future. DNA sequence-based identification, however, for many (notably microscopic) taxa is quicker and easier than morphology-based identification and will increase in importance because it can be automated or done *en masse*, for example, through metabarcoding, by non-experts on the taxa to be identified. Strategies now need to be developed to increase the number of available DNA sequences annotated with validated and reliable type information in the INSDC databases. This will require activities at different levels.

Firstly, sequencing efforts need to increasingly target name-bearing types or other material annotated by the original author. Botanists, phycologists, and mycologists also have the option of designating epitypes (e.g., Tillmann et al., 2021) or paratypes (a paratype is any specimen cited in the protologue that is neither the holotype nor an isotype; paratypes are also used in zoological nomenclature) and then linking these specimens to DNA sequences. The option of epitypification might fruitfully be added to the *Code of Zoological Nomenclature* (Schrödl and Haszprunar 2014; Scherz et al. 2021). To obtain DNA from "original material" in collections, many approaches are now available that require small volumes of input DNA sequences, including genome and targeted hybrid enrichment sequencing (Yeates et al. 2016; Rancilhac et al. 2020; Raxworthy and Smith, 2021; Straube et al. 2021). In some taxa, nondestructive DNA extraction methods have proven useful (Gilbert et al. 2007; Thomsen et al. 2009; Shepherd 2017). Even short sequences of single markers, if originating from types, can have great relevance as references in metabarcoding and environmental DNA analysis (Pawlowski et al. 2012; Hausmann et al. 2016; Miller et al. 2016). A prime example for the immense utility of short sequences is the Barcode of Life Database (BOLD), which represents a persistent, species-level taxonomic registry for the animal kingdom based on the analysis of patterns of nucleotide variation in the barcode region of the cytochrome c oxidase I (COI) gene (Ratnasingham and Hebert 2013). This database already contains 12,143 type-derived COI sequences (4299 from holotypes, 6670 from paratypes, 635 from syntypes, 261 from allotypes, and 278 from lectotypes: pers. comm. Sujeevan Ratnasingham, Director of Informatics, BOLD, 14 August 2023).

Secondly, sequence submitters need to provide metadata for sequences they have produced from type material to enable NCBI curators and other database users to find these high-value data (Fig. 1). A main prerequisite is that the sequence be linked to a permanently stored and retrievable voucher specimen. Ideally, this is done via a unique catalog number from a collection documented in the NCBI BioCollections or GRSciColl lists, but the numerous idiosyncrasies by which specimen vouchers are cited in taxonomy require flexibility. Many if not the majority of types of insects are unnumbered, and this is also true of many botanical and protist-type collections. Furthermore, botanical, but not zoological, types are traditionally cited with the collector's name, followed by a parenthetical reference to the herbarium where the collection is housed. Online tools to find specimens based on their specimen code, such as Roderic Page's https://material-examined.herokuapp.com/, are, therefore, animal biased. Other type material exists in private collections, and such specimens may be uncatalogued, yet uniquely identifiable by a field number (see Table 2). Nevertheless, large-scale projects such as the Integrated Digitized Biocollections (iDigBio) projects funded by the National Science Foundation in the United States hold much promise to improve access to specimen and type voucher information. Many institutions have started to implement stable, digitally retrievable specimen identifiers (Guralnick et al. 2015; Güntsch et al. 2018; Hardisty et al. 2021), which then

can be linked to other specimen information in taxonomic monographs (Mabry et al. 2022), but it is going to take substantial time and effort until all type specimens will be provided with such identifiers.

Informing the respective INSDC team at the time of sequence submission that a particular voucher specimen (and associated sequences) corresponds to a name-bearing type, along with a PDF of the respective publication, will facilitate annotation (Table 2). How to handle the updating of informal (candidate species) names used in original submissions, once a name has been formally published, is a so-far unsolved problem and one that may already affect thousands of names. Thus, an analysis of GenBank data found some 1300 reptile names of the form "*Pelomedusa* sp. A CK-2014," creating a disconnect between sequences and names (Garg et al. 2019). Systematists wanting to help improve the annotation of existing type-derived sequences—whether submitted by themselves or by others—may contact database curators (by emailing gb-admin@ncbi.nlm.nih.gov in the case of GenBank) who will then modify incomplete entries. Two of us (MG and CLS) tested this for freshwater Dinophyceae (Alveolata), a group comprising the manageable number of some 350 known species (Moestrup and Calado 2018), with sequences associated with type material for 22 species and eight subspecific taxa. Based on curated voucher lists for metabarcoding studies (Gottschling et al. 2020), MG and CLS updated NCBI GenBank with information on holotypes or epitypes, living strains from which the corresponding type material had been prepared, and a doi for the typification act that is linked directly from the NCBI TaxBrowser (Supplementary Data Set 1).

Thirdly, scientific journals publishing nomenclatural acts and species databases with information on type specimens could increasingly become sources for updating the NCBI Taxonomy database, especially if data extraction could be automatically parsed. Examples of such databases are Amphibian Species of the World (https://amphibiansoftheworld.amnh.org) and Reptile Database (http://www.reptile-database.org). Improvements in the Application Programming Interfaces of such databases or downloadable readouts from them with parseable information related to type identifiers and the kinds of types would facilitate the adding of type metadata to NCBI. At some point, it should become possible to harvest metadata on type material directly from the literature—the Pensoft journals have already developed TaxPub, an extensible markup language (XML) linking to the National Library of Medicine journal archiving standards that allows detailed markup of taxonomic descriptions (Penev et al. 2010; Federhen 2015). This would enable NCBI taxonomy curators

Table 2. Examples of sequence source modifiers that flag type-derived sequences (all examples can be found in GenBank).

NCBI modifier	Note	Example
Culture_collection	Format for cultures in culture collections: "institution-code:culture-id." culture-id and institution-code are mandatory. When possible, use code documented in NCBI BioCollections or WFCC.	/culture_collection="CBS:1752"
Isolate	Use this for lab numbers/field numbers of the specific specimen/strain from which this sequence was obtained.	/isolate="JT13209"
Note	Add any pre-publication information on potential type material here. Suggested syntax: "submitter reports sequence is from type material" to indicate that the submitted sequence is from type material. Additional information might be added: (i) kind of type (holotype, isotype, lectotype, epitype, neotype, syntype, paratype); (ii) original binomen (only of names already published at the time of submission).	*as generic information or if name is not yet published at time of submission* /note="submitter reports sequence is from type material" OR *if name is already published at time of submission* /note="submitter reports sequence is from type material: holotype of *Mantella inexistens*"
Specimen_voucher	DwC format for preserved specimens: "institution-code: internal-code:specimen-id." specimen-id is mandatory. When possible use code documented in NCBI BioCollections or GRSciColl; see http://www.insdc.org/controlled-vocabulary-specimenvoucher-qualifier If specimens are deposited but not yet catalogued in a collection, give field number or similar identifier preceded by institution code. Note: botanical specimens often have date-based specimen numbers that may not be easily processed.	/specimen_voucher="ZSM:422/2016" OR /specimen_voucher="UADBA:ZCMV 15236" OR /specimen_voucher="M:S.S. Renner 2816" [traditional in botany would have been S.S. Renner 2816 (M)]
Strain	Use this for strain numbers of pure strains, that is, those not deposited in culture collections.	/strain="ABC 1234"
Type_material	**This field is not user submitted**—it is automatically updated only after the publication or nomenclature database entry is verified by NCBI Taxonomy curators. Please provide the full publication as a pdf to gb-admin@ncbi.nlm.nih.gov. (Do NOT use the "Type" modifier for this information). See http://www.insdc.org/controlled-vocabulary-typematerial-qualifer	/type_material="holotype of Mantella inexistens"

Note: Supplementary Fig. 1 illustrates the processes involved in annotating sequences as type-derived in GenBank and Supplementary Data Set 2 provides further exemplary records for a range of organisms.

to *post facto* annotate type-derived sequences that are already in GenBank, but that are not recognizable because their submitting authors did not enter the required metadata in Entrez.

Last but not least, the possibility of more user-friendly options for submitters to annotate sequences as derived from type material in the INSDC databases requires discussion. As explained above and in Fig. 1, at present, users cannot directly annotate specimens as types, a strategy initially introduced to reduce the introduction of errors. To help submitters, we here propose a simple, but standardized syntax (Table 2) that submitters can use to enter data in the NCBI "Note" source modifier. In the future, NCBI and its INSDC partners should introduce improvements, such as automatically reminding submitting authors to enter specimen voucher data and provide information on whether sequences come from nomenclatural types.

Conclusion

Why Start Now, Who Might Do the Work and With Which Funding?

The timing of our study of how users and NCBI staff have flagged and curated type-derived sequences, and how the relevant procedures might be improved, does not coincide with any major strides or breakthroughs in museomics, genomics, taxonomy, or databasing. Those fields continue to improve every year, and we have no doubt that the coming decades will bring still better extraction or sequencing methods, perhaps requiring even less tissue. However, the increase in the number of new taxa discovered and in need of DNA-based identification or naming means that sequence producers and database curators need to deal with the flagging of type-derived sequences right now. Type specimens are so valuable that data derived from them, whether morphological or genetic, deserve to be findable, accessible, interoperable, and reusable (i.e., FAIR).

Flagging DNA sequences will not cost extra money, although it does require extra effort for sequence submitters, including an email exchange with NCBI taxonomy curators about a specimen's type status (Fig. 1). This minimal inconvenience will have major benefits for taxonomic stability and downstream users. The amount of environmental sequence data, especially from water and soil samples, will increase in the foreseeable future in a way that can hardly be managed, but such environmental sequences will be of limited use for evolutionary and biogeographic studies as long as they cannot be classified in an interoperable way. The essential importance of type material in this regard is not always clear to all involved; for example, it is not considered at all in the EukRef initiative (https://pr2-database.org/eukref/about/) for the phylogenetic curation of ribosomal RNA to "enhance understanding of eukaryotic diversity and distribution" (del Campo et al. 2018).

Only DNA sequences from type specimens remove ambiguity about taxonomic assignment of sequences from narrow-scope studies of single taxa to broad-scope metabarcoding projects (e.g., Rancilhac et al. 2020). In addition, sequences from type material in NCBI and INSDC partner databases can decrease pressure to send on loan precious material, clustered in Europe and North America (Miralles et al. 2020) and can also become a digital insurance against the loss or irrevocable damage of the physical type material itself, for instance, due to fire or other disasters (Tyler et al. 2023), by helping identify the most suitable neotype or epitype material.

Following the highly successful type-sequence-data curation and improved type-material annotations for prokaryotes and Fungi (Robbertse et al. 2017; Kannan et al. 2023), it is a worthwhile and achievable goal to now expand NCBI's set of reference resources, RefSeq, with type-derived sequences from animals and plants. This effort could be part of ongoing and planned phylogenetic projects but will require funding and commitment from biorepository directors and curators. Such funding could be secured by a combination of individual scientists (via grant applications), collections-based organizations, and national funding agencies. If the work were done as part of taxon-centered or Tree-of-Life-type project in which the sequencing of type specimens would be just one additional target, the risk of failure would be small and the potential gain high.

Sequences from type material constitute an enormously valuable resource for all of biology now and in the future because they improve the chance of obtaining correct names in BLAST searches. Given this importance, and in view of the role of artificial intelligence in genomic diagnostics, it is clear that the production and curation of DNA sequences from type material over the next years will need to grow and improve.

Supplementary Materials

Supplementary material is available at *Systematic Biology* online.

Acknowledgments

We thank Isabel Sanmartin, Alexandre Antonelli, Roderic Page, and Thomas Pape for insightful comments on our manuscript, and Sujeevan Ratnasingham and Paul Hebert for the numbers of type-derived COI sequences in BOLD. CLS additionally thanks numerous NCBI colleagues for help and suggestions before submission. His work was supported by the Intramural Research Program of the National Library of Medicine at the NIH. MDS, SSR, and MV were supported by DFG SPP 1991 "Taxon-Omics," projects SCHE 2181/1-1, RE 603/29-1, and VE 247/20-1.

References

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Arita M., Karsch-Mizrachi I., Cochrane G. 2021. The international nucleotide sequence database collaboration. Nucleic Acids Res. 49:D121–D124.

Burgin J., Ahamed A., Cummins C., Devraj R., Gueye K., Gupta D., Gupta V., Haseeb M., Ihsan M., Ivanov E., Jayathilaka S., Balavenkataraman Kadhirvelu V., Kumar M., Lathi A., Leinonen R., Mansurova M., McKinnon J., O'Cathail C., Paupério J., Pesant S., Rahman N., Rinck G., Selvakumar S., Suman S., Vijayaraja S., Waheed Z., Woollard P., Yuan D., Zyoud A., Burdett T., Cochrane G. 2023. The European Nucleotide Archive in 2022. Nucleic Acids Res. 51:D121–D125.

Chakrabarty P. 2010. Genetypes: a concept to help integrate molecular systematics and traditional taxonomy. Zootaxa 2632:67–68.

Chakrabarty P., Warren M., Page L., Baldwin C. 2013. GenSeq: an updated nomenclature and ranking for genetic sequences from type and non-type sources. ZooKeys 346:29–41.

Ciufo S., Kannan S., Sharma S., Badretdin A., Clark K., Turner S., Brover S., Schoch C.L., Kimchi A., DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int. J. Syst. Evol. Microbiol. 68:2386–2392.

Del Campo J., Kolisko M., Boscaro V., Santoferrara L.F., Nenarokov S., Massana R., Guillou L., Simpson A., Berney C., de Vargas C., Brown M.W., Keeling P.J., Wegener Parfrey L. 2018. EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. PLoS Biol. 16:e2005849.

Federhen S. 2015. Type material in the NCBI Taxonomy Database. Nucleic Acids Res. 43:D1086–D1098.

Garg A., Leipe D., Uetz P. 2019. The disconnect between DNA and species names: lessons from reptile species in the NCBI Taxonomy Database. Zootaxa 4706:401–407.

Gilbert M.T.P., Moore W., Melchior L., Worobey M. 2007. DNA extraction from dry museum beetles without conferring external morphological damage. PLoS One 2:e272.

Gottschling M., Chacón J., Žerdoner Čalasan A., Neuhaus S., Kretschmann J., Stibor H., John U. 2020. Phylogenetic placement of environmental sequences using taxonomically reliable databases helps to rigorously assess dinophyte biodiversity in Bavarian lakes (Germany). Freshw. Biol. 65:193–208.

Güntsch A., Groom Q., Hyam R., Chagnoux S., Röpert D., Berendsohn W., Casino A., Droege G., Gerritsen W., Holetschek J., Marhold K., Mergen P., Rainer H., Smith V., Triebel D. 2018. Standardised globally unique specimen identifiers. Biodivers. Inf. Sci. Stand 2:e26658.

Guralnick R.P., Cellinese N., Deck J., Pyle R.L., Kunze J., Penev L., Walls R., Hagedorn G., Agosti D., Wieczorek J., Catapano T., Page R. 2015. Community next steps for making globally unique identifiers work for biocollections data. ZooKeys 494:133–154.

Hardisty A., Addink W., Glöckler F., Güntsch A., Islam S., Weiland C. 2021. A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Res. Ideas Outcomes 7:e67379.

Harrison I.J., Chakrabarty P., Freyhof J., Craig J.F. 2011. Correct nomenclature and recommendations for preserving and cataloguing voucher material and genetic sequences. J. Fish Biol. 78:1283–1290.

Hausmann A., Miller S.E., Holloway J.D., deWaard J.R., Pollock D., Prosser S.W.J., Hebert P.D.N. 2016. Calibrating the taxonomy of a megadiverse insect family: 3000 DNA barcodes from geometrid type specimens (Lepidoptera, Geometridae). Genome 59:671–684.

Hedlund B.P., Chuvochina M., Hugenholtz P., Konstantinidis K.T., Murray A.E., Palmer M., Parks D.H., Probst A., 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. Nat. Microbiol. 7:1702–1708.

Kannan S., Sharma S., Ciufo S., Clark K., Turner S., Kitts P.A., Schoch C.L., DiCuccio M., Kimchi A. 2023. Collection and curation of prokaryotic genome assemblies from type strains at NCBI. Int. J. Syst. Evol. Microbiol. 73:005707.

May T.W., Redhead S.A., Bensch K., Hawksworth D.L., Lendemer J., Lombard L., Turland N.J. 2019. Chapter F of the international code of nomenclature for algae, fungi, and plants as approved by the 11th international mycological congress, San Juan, Puerto Rico, July 2018. IMA Fungus 10:1–14.

Mabry M.E., Zapata F., Paul D.L., O'Connor P.M., Soltis P.S., Blackburn D.C., Simmons N.B. 2022. Monographs as a nexus for building extended specimen networks using persistent identifiers. BSSB. 1:8323.

Miller S.E., Hausmann A., Hallwachs W., Janzen D.H. 2016. Advancing taxonomy and bioinventories with DNA barcodes. Philos. Trans. R. Soc. London, Ser. B 371:20150339.

Miralles A., Bruy T., Wolcott K., Scherz M.D., Begerow D., Beszteri B., Bonkowski M., Felden J., Gemeinholzer B., Glaw F., Glöckner F.O., Hawlitschek O., Kostadinov I., Nattkemper T.W., Printzen C., Renz J., Rybalka N., Stadler M., Weibulat T., Wilke T., Renner S.S., Vences M. 2020. Repositories for taxonomic data: where we are and what is missing. Syst. Biol. 69:1231–1253.

Moestrup Ø., Calado A.J. 2018. Dinophyceae. Berlin: Springer.

O'Leary N.A., Wright M.W., Brister J.R., Ciufo S., Haddad D., McVeigh R., Rajput B., Robbertse B., Smith-White B., Ako-Adjei D., Astashyn A., Badretdin A., Bao Y., Blinkova O., Brover V., Chetvernin V., Choi J., Cox E., Ermolaeva O., Farrell C.M., Goldfarb T., Gupta T., Haft D., Hatcher E., Hlavina W., Joardar V.S., Kodali V.K., Li W., Maglott D., Masterson P., McGarvey K.M., Murphy M.R., O'Neill K., Pujar S., Rangwala S.H., Rausch D., Riddick L.D., Schoch C., Shkeda A., Storz S.S., Sun H., Thibaud-Nissen F., Tolstoy I., Tully R.E., Vatsan A.R., Wallin C., Webb D., Wu W., Landrum M.J., Kimchi A., Tatusova T., DiCuccio M., Kitts P., Murphy T.D., Pruitt K.D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733–D745.

Parker C.T., Tindall B.J., Garrity G.M., eds. 2019. The international code of nomenclature of prokaryotes. Int. J. Syst. Evol. Microbiol. 69:S1–S111.

Pawlowski J., Audic S., Adl S., Bass D., Belbahri L., Berney C., Bowser S.S., Cepicka I., Decelle J., Dunthorn M., Fiore-Donno A.M., Gile G.H., Holzmann M., Jahn R., Jirků M., Keeling P.J., Kostka M., Kudryavtsev A., Lara E., Lukeš J., Mann D.G., Mitchell E.A.D., Nitsche F., Romeralo M., Saunders G.W., Simpson A.G.B., Smirnov A.V., Spouge J.L., Stern R.F., Stoeck T., Zimmermann J., Schindel D., de Vargas C. 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 10(11):e1001419.

Penev L., Kress W.J., Knapp S., Li D.-Z., Renner S.S. 2010. Fast, linked, and open—the future of taxonomic publishing for plants: launching the journal PhytoKeys. PhytoKeys 1:1–14.

Rancilhac L., Bruy T., Scherz M.D., Pereira E.A., Preick M., Straube N., Lyra M., Ohler A., Streicher J.W., Andreone F., Crottini A., Hutter C.R., Randrianantoandro J.C., Rakotoarison A., Glaw F., Hofreiter M., Vences M. 2020. Target-enriched DNA sequencing from historical type material enables a partial revision of the Madagascar giant stream frogs (genus *Mantidactylus*). J. Nat. Hist. 54:87–118.

Ratnasingham S., Hebert P.D.N. 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. PLoS One 8(8):e66213.

Raxworthy C.J., Smith B.T. 2021. Mining museums for historical DNA: advances and challenges in museomics. Trends Ecol. Evol. 36(11):1049–1060.

Reimer L.C., Sarda Carbasse J., Koblitz J., Ebeling C., Podstawka A., Overmann J. 2022. BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. Nucleic Acids Res. 50:D741–D746.

Robbertse B., Strope P.K., Chaverri P., Gazis R., Ciufo S., Domrachev M., Schoch C.L. 2017. Improving taxonomic accuracy for fungi in public sequence databases: applying "one name one species" in well-defined genera with *Trichoderma/Hypocrea* as a test case. Database (Oxford) 2017:1–14.

Robert V., Vu D., Amor A.B.H., van de Wiele N., Brouwer C., Jabas B., Szoke S., Dridi A., Triki M., Ben Daoud S., Chouchen O., Vaas L., de Cock A., Stalpers J.A., Stalpers D., Verkley G.J.M., Groenewald M., Dos Santos F.B., Stegehuis G., Li W., Wu L., Zhang R., Ma J., Zhou

M., Gorjón S.P., Eurwilaichitr L., Ingsriswang S., Hansen K., Schoch C.L., Robbertse B., Irinyi L., Meyer W., Cardinali G., Hawksworth D.L., Taylor J.W., Crous P.W. 2013. MycoBank gearing up for new horizons. IMA Fungus 4:371–379.

Sayers E.W., Bolton E.E., Brister J.R., Canese K., Chan J., Comeau D.C., Farrell C.M., Feldgarden M., Fine A.M., Funk K., Hatcher E., Kannan S., Kelly C., Kim S., Klimke W., Landrum M.J., Lathrop S., Lu Z., Madden T.L., Malheiro A., Marchler-Bauer A., Murphy T.D., Phan L., Pujar S., Rangwala S.H., Schneider V.A., Tse T., Wang J., Ye J., Trawick B.W., Pruitt K.D., Sherry S.T. 2023. Database resources of the National Center for Biotechnology Information in 2023. Nucleic Acids Res. 51:D29–D38.

Scherz M.D., Schmidt L., Crottini A., Miralles A., Rakotoarison A., Raselimanana A.P., Köhler J., Glaw F., Vences M. 2021. Into the chamber of horrors: a proposal for the resolution of nomenclatural chaos in the *Scaphiophryne calcarata* complex (Anura: Microhylidae), with a new species-level phylogenetic hypothesis for Scaphiophryninae. Zootaxa 4938:392–420.

Schoch C.L., Ciufo S., Domrachev M., Hotton C.L., Kannan S., Khovanskaya R., Leipe D., McVeigh R., O'Neill K., Robbertse B., Sharma S., Soussov V., Sullivan J.P., Sun L., Turner S., Karsch-Mizrachi I. 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford) 2020:1–21.

Schoch C.L., Robbertse B., Robert V., Vu D., Cardinali G., Irinyi L., Meyer W., Nilsson R.H., Hughes K., Miller A.N., Kirk P.M., Abarenkov K., Aime M.C., Ariyawansa H.A., Bidartondo M., Boekhout T., Buyck B., Cai Q., Chen J., Crespo A., Crous P.W., Damm U., De Beer Z.W., Dentinger B.T.M., Divakar P.K., Duenas M., Feau N., Fliegerova K., Garcia M.A., Ge Z.W., Griffith G., Groenewald J.Z., Groenewald M., Grube M., Gryzenhout M., Gueidan C., Guo L.D., Hambleton S., Hamelin R., Hansen K., Hofstetter V., Hong S.B., Houbraken J., Hyde K.D., Inderbitzin P., Johnston P.R., Karunarathna S.C., Koljalg U., Kovacs G.M., Kraichak E., Krizsan K., Kurtzman C.P., Larsson K.H., Leavitt S., Letcher P.M., Liimatainen K., Liu J.K., Lodge D.J., Luangsa-ard J.J., Lumbsch H.T., Maharachchikumbura S.S.N., Manamgoda D., Martin M.P., Minnis A.M., Moncalvo J.M., Mule G., Nakasone K.K., Niskanen T., Olariaga I., Papp T., Petkovits T., Pino-Bodas R., Powell M.J., Raja H.A., Redecker D., Sarmiento-Ramirez J.M., Seifert K.A., Shrestha B., Stenroos S., Stielow B., Suh S.O., Tanaka K., Tedersoo L., Telleria M.T., Udayanga D., Untereiner W.A., Uribeondo J.D., Subbarao K.V., Vagvolgyi C., Visagie C., Voigt K., Walker D.M., Weir B.S., Weiss M., Wijayawardene N.N., Wingfield M.J., Xu J.P., Yang Z.L., Zhang N., Zhuang W.Y., Federhen S. 2014. Finding needles in haystacks: linking scientific names, reference specimens and molecular data for fungi. Database (Oxford) 2014:1–21.

Schoch C.L., Seifert K.A., Huhndorf S., Robert V., Spouge J.L., Levesque C.A., Chen W., Bolchacova E., Voigt K., Crous P.W., Miller A.N., Wingfield M.J., Aime M.C., An K.D., Bai F.Y., Barreto R.W., Begerow D., Bergeron M.J., Blackwell M., Boekhout T., Bogale M., Boonyuen N., Burgaz A.R., Buyck B., Cai L., Cai Q., Cardinali G., Chaverri P., Coppins B.J., Crespo A., Cubas P., Cummings C., Damm U., de Beer Z.W., de Hoog G.S., Del-Prado R., Dentinger B., Dieguez-Uribeondo J., Divakar P.K., Douglas B., Dueñas M., Duong T.A., Eberhardt U., Edwards J.E., Elshahed M.S., Fliegerova K., Furtado M., García M.A., Ge Z.W., Griffith G.W., Griffiths K., Groenewald J.Z., Groenewald M., Grube M., Gryzenhout M., Guo L.D., Hagen F., Hambleton S., Hamelin R.C., Hansen K., Harrold P., Heller G., Herrera G., Hirayama K., Hirooka Y., Ho H.M., Hoffmann K., Hofstetter V., Hognabba F., Hollingsworth P.M., Hong S.B., Hosaka K., Houbraken J., Hughes K., Huhtinen S., Hyde K.D., James T., Johnson E.M., Johnson J.E., Johnston P.R., Jones E.B., Kelly L.J., Kirk P.M., Knapp D.G., Kõljalg U., Kovács G.M., Kurtzman C.P., Landvik S., Leavitt S.D., Liggenstoffer A.S., Liimatainen K., Lombard L., Luangsa-Ard J.J., Lumbsch H.T., Maganti H., Maharachchikumbura S.S., Martín M.P., May T.W., McTaggart A.R., Methven A.S., Meyer W., Moncalvo J.M., Mongkolsamrit S., Nagy L.G., Nilsson R.H., Niskanen T., Nyilasi I., Okada G., Okane I., Olariaga I., Otte J., Papp T., Park D., Petkovits T., Pino-Bodas R., Quaedvlieg W., Raja H.A., Redecker D., Rintoul T., Ruibal C., Sarmiento-Ramirez J.M., Schmitt I., Schussler A., Shearer C., Sotome K., Stefani F.O., Stenroos S., Stielow B., Stockinger H., Suetrong S., Suh S.O., Sung G.H., Suzuki M., Tanaka K., Tedersoo L., Telleria M.T., Tretter E., Untereiner W.A., Urbina H., Vágvölgyi C., Vialle A., Vu T.D., Walther G., Wang Q.M., Wang Y., Weir B.S., Weiß M., White M.M., Xu J., Yahr R., Yang Z.L., Yurkov A., Zamora J.C., Zhang N., Zhuang W.Y., Schindel D. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc. Natl. Acad. Sci. USA 109:6241–6246.

Schrödl M., Haszprunar G. 2014. Do we need epitypes in zoology? Spixiana 39:199–201.

Sharma S., Ciufo S., Starchenko E., Darji D., Chlumsky L., Karsch-Mizrachi I., Schoch C.L. 2018. The NCBI BioCollections Database. Database (Oxford) 2019:1–8.

Shepherd L.D. 2017. A non-destructive DNA sampling technique for herbarium specimens. PLoS One 12:e0183555.

Straube N., Lyra M.L., Paijmans J.L.A., Preick M., Basler N., Penner J., Rödel M.O., Westbury M.V., Haddad C.F.B., Barlow A., Hofreiter M. 2021. Successful application of ancient DNA extraction and library construction protocols to museum wet collection specimens. Mol. Ecol. Resour. 21:2299–2315.

Tanizawa Y., Fujisawa T., Kodama Y., Kosuge T., Mashima J., Tanjo T., Nakamura Y. 2023. DNA Data Bank of Japan (DDBJ) update report 2022. Nucleic Acids Res. 51:D101–D105.

Thiele K.R., Applequist W.L., Renner S.S., May T.W., Dönmez A.A., Groom Q., Lehtonen S., Maggs C.A., Malécot V., Yoon H.S. 2023a. 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. doi: 10.1002/tax.12931

Thiele K.R., Groom Q., Renner S.S., Applequist W.L., Lehtonen S. 2023b. Proposals to permit DNA sequences to serve as types of names in prescribed circumstances. Taxon 72:1143–1145.

Thiele K.R., Groom Q., Renner S.S., Lehtonen S. 2023c. Proposals to permit DNA sequences to be used for fixing the application of names in prescribed circumstances. Taxon 72:1146–1148.

Thomsen P.F., Elias S., Gilbert M.T.P., Haile J., Munch K., Kuzmina S., Froese D.G., Sher A., Holdaway R.N., Willerslev E. 2009. Non-destructive sampling of ancient insect DNA. PLoS One 4:e5048.

Tillmann U., Bantle A., Krock B., Elbrächter M., Gottschling M. 2021. Recommendations for epitypification of dinophytes exemplified by *Lingulodinium polyedra* and molecular phylogenetics of the Gonyaulacales based on curated rRNA sequence data. Harmful Algae 104:101956.

Tyler M.J., Fucsko L.A., Roberts J.D. 2023. Calamities causing loss of museum collections: a historical and global perspective on museum disasters. Zootaxa 5230:153–178.

Van den Burg M.P., Vieites D.R. 2023. Bird genetic databases need improved curation and error reporting to NCBI. Ibis 165:472–481.

Wang F., Wang K., Cai L., Zhao M., Kirk P.M., Fan G., Sun Q., Li B., Wang S., Yu Z., Han D., Ma J., Wu L., Yao Y. 2023. Fungal names: a comprehensive nomenclatural repository and knowledge base for fungal taxonomy. Nucleic Acids Res. 51:D708–D716.

Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., 't Hoen P.A.C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M.A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018.

Yeates D.K., Zwick A., Mikheyev A.S. 2016. Museums are biobanks: unlocking the genetic potential of the three billion specimens in the World's biological collections. Curr. Opin. Insect Sci. 18:83–88.