Towards a global list of accepted species V. The devil is in the details

Richard L. Pyle[1*], Saroj K. Barik[2], Les Christidis[3], Stijn Conix[4], Mark John Costello[5,6], Peter Paul van Dijk[7], Stephen T. Garnett[8], Donald Hobern[9], Paul M. Kirk[10], Aaron M. Lien[11], Thomas M. Orrell[12], David Remsen[13], Scott A. Thomson[14], Nina Wambiji[15], Frank E. Zachos[16,17,18], Zhi-Qiang Zhang[19,20] and Kevin R. Thiele[21,22]

[1] Bernice Pauahi Bishop Museum, Honolulu, Hawaiʻi, United States of America. https://orcid.org/0000-0003-0768-1286

[2] CSIR-National Botanical Research Institute, Lucknow, India. https://orcid.org/0000-0002-1795-9539

[3] Southern Cross University, Coffs Harbour, Australia. http://orcid.org/0000-0002-3345-6034

[4] Centre for Logic and Philosophy of Science, KU Leuven, Leuven, Belgium. https://orcid.org/0000-0002-1487-0213

[5] Faculty of Bioscience and Aquaculture, Nord Universitet, Postboks, Norway. https://orcid.org/0000-0003-2362-0328

[6] School of Environment, University of Auckland, Auckland, New Zealand.

[7] Global Wildlife Conservation, Austin, Texas, USA. https://orcid.org/0000-0001-7283-7000

[8] Research Institute for the Environment and Livelihoods, Charles Darwin University, Northern Territory 0909 Australia. https://orcid.org/0000-0002-0724-7060

[9] Species 2000, Canberra, Australia. https://orcid.org/0000-0001-6492-4016

[10] Royal Botanic Gardens Kew, RIchmond United Kingdom. https://orcid.org/0000-0002-0658-7338

[11] School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona, USA

[12] Integrated Taxonomic Information System, National Museum of Natural History, Smithsonian Institution, Washington DC, 20560. https://orcid.org/0000-0003-1038-3028

[13] Marine Biological Laboratory: Woods Hole, Massachusetts, United States of America. https://orcid.org/0000-0003-1289-0840

[14] Chelonian Research Institute, Oviedo, Florida, USA. https://orcid.org/0000-0003-1279-2722

[15] Kenya Marine and Fisheries Research Institute, Mombasa, Kenya. https://orcid.org/0000-0002-7775-5300

[16] Natural History Museum Vienna, Vienna, Austria. https://orcid.org/0000-0002-0133-6265

[17] Department of Evolutionary Biology, University of Vienna, Vienna, Austria

[18] Department of Genetics, University of the Free State, Bloemfontein, South Africa

[19] Manaaki Whenua-Landcare Research, Auckland, New Zealand https://orcid.org/0000-0003-4172-0592

[20] School of Biological Sciences, The University of Auckland, Auckland, New Zealand

[21] Australian Academy of Science, Ian Potter House, 9 Gordon Street, Acton ACT 2601 Australia. http://orcid.org/0000-0002-6658-6636

[22] The University of Western Australia, 35 Stirling Highway, Perth WA 6009, Australia.

* Corresponding Author: deepreef@bishopmuseum.org

In total, there will be six follow-up papers submitted to ODE:
Towards a global list of accepted species I: Taxonomy as map-making: the foundation for lists of accepted species (Thiele et al.)
Towards a global list of accepted species II: Consequences of inadequate taxonomic list governance (Thomson et al.)
Towards a global list of accepted species III: Independence and stakeholder inclusion (Conix et al.)
Towards a global list of accepted species IV: Overcoming fragmentation in the governance of taxonomic lists (Lien et al.)
Towards a global list of accepted species V: The devil is in the details (Pyle et al.)
Towards a global list of accepted species VI: The Catalogue of Life Checklist (Hobern et al.)

Abstract

A consensus among biologists has been growing in recent years for the development of a global list of accepted species (and other taxa). To date, much discussion has focused on visions for how such a list would benefit many scientific and societal disciplines. Less emphasis has been placed on understanding the many technical challenges of compiling and maintaining such a list. Challenges include details of implementation such as defining what each entry on the list represents, the scope (taxonomic breadth), granularity (only species, all taxonomic ranks, unnamed operational taxonomic units), and level of confidence in the status of individual list entries. The specific properties and minimum information requirements of list items need to be defined, and a process for ensuring accuracy, consistency, and noting uncertainties, needs to be adopted. Perhaps the greatest technical challenge is in developing the procedures by which the global list is created, updated and maintained. Considerations of how to incorporate obscure and newly described taxa, the extent and specific implementation of a review process, and mechanisms for arbitrating disputes or alternative taxonomic viewpoints will need to be addressed through an open and transparent process with broad engagement from multiple communities. Details concerning how the global list can be accessed, how it will be maintained, and the way in which the list and its contents are properly cited need to be determined. Many of these issues have been considered and sometimes solved by the Catalogue of Life, which should serve as the core foundation for the actual implementation of any global list of species.

Keywords: taxonomy; informatics; global list; Catalogue of Life

Declarations:

Conflicts of interest/Competing interests: No authors declare conflicts of interest or competing interests arising from this work.

Consent for publication: All authors consent to this article being submitted for publication

Data sharing: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Introduction

The primary function of taxonomy, and a key reason why it is fundamental to many other scientific disciplines (essentially all of biology), is that it helps to organize knowledge of, and serves as a basic tool for communication about, biodiversity (Costello and Wieczorek 2014, Ruggiero et al, 2015). These functions are best facilitated by a stable mapping of scientific names to taxonomic entities, but taxonomic progress necessitates that such entities and their associated names, classifications and relationships will change over time (Costello et al. 2013a, Thiele et. al. In Prep). Striking the right balance between the dynamic science of taxonomy and the broader societal need for effective communication about biodiversity has fostered a vigorous debate within the scientific and conservation communities about the need for a formalized global list of accepted species and other taxa (Garnett & Christidis 2017; Raposo et al. 2017; Holstein & Lueber 2017; Cotterill et al. 2017; Jackson et al. 2017; Hollingsworth 2017; Lamberts 2017; Funk et al. 2017; Conix 2018; Garnett & Christidis 2018a; 2018b; Thomson et al. 2018; Zachos 2018). This debate led to the formation of a Working Group on the governance of taxonomic lists, supported by the International Union of Biological Sciences (IUBS), and focused on developing a consensus view on the issues that initiated the debate (Garnett et al. 2020a; Garnett et al. 2020b). Other papers in this Special Issue discuss the taxonomic context (Thiele et. al. In Prep), governance (Lien et al. In Prep), case studies (Thomson et al. In Prep), issues related to independence and stakeholder inclusion (Conix et al. In Prep), and role of the Catalogue of Life (Hobern et al. In Prep) for creating a global list of accepted species.

Much consideration has been given to broad visions of how such lists can be generated, managed, and used in ways that satisfy diverse needs and priorities, and to the general complexity of social, societal and traditional norms across taxonomy and its dependent fields. However, many specific details remain unresolved on how to generate, populate, and maintain such a list, and these require careful attention (Costello et al. 2014, Costello et al. 2018, Kroh et al. 2019). After the broader conceptual framework of core principles, governance, and stakeholder inclusion are developed and adopted, there will remain the important task of actually

implementing the framework and generating a universal list of accepted species and other taxa. This paper addresses some of these important details. Many issues described herein have been considered and addressed during the development of the Catalogue of Life (COL). We, therefore, recommend that COL should be the foundation of implementation efforts going forward.

One list, or multiple lists?

Discussions on this issue have variously referred to "lists" of species in the plural or a single global "list" of accepted species or taxa (e.g., Garnett et al. 2020b). Although this may appear to represent an inconsistency, it is a reflection of the assumption that a *single* "global list" will emerge as an assembly and aggregation of *multiple* non-overlapping "partial lists", each with a well-defined, limited and mutually-exclusive scope (Figure 1). These partial lists are likely to be defined and delimited by higher taxonomic groupings (e.g., at the ranks of classes, orders, families, etc.), rather than by non-taxonomic parameters such as geographic distribution (but see some possible exceptions below), because a risk of using non-taxonomic parameters to subdivide the global list into partial lists is that overlap in species coverage of partial lists may result in conflict or ambiguity in the global list. In some cases (such as with birds, see McClure et al. 2020), such overlap and conflict already exist, with different subcommunities producing different lists for the same taxonomic group. Such lists with overlapping content are referred to here as "competing lists" (not to be confused with "partial lists"; see above and Figure 1).

The assumption that there should be a single global list made up of partial lists, and that partial lists should be defined taxonomically, is not necessarily absolute. For example, certain taxa are "ambiregnal", whereby a single species is classified within more than one kingdom (Patterson 1986; Corliss 1995; Nakada 2010). Such taxa are governed by more than one nomenclatural Code, and therefore could potentially be partitioned amongst more than one partial list. Another non-taxonomic distinction concerns extant *versus* extinct taxa based on fossil material. While some partial lists may cover both extant and fossil members within a single taxonomic group; others might focus exclusively on extant *or* fossil members. Some fossil groups are also difficult to allocate to a taxonomic group (such as extinct bird-like dinosaur lineages). Moreover, some taxonomic groups are represented both by many extant and extinct fossil taxa (e.g., Foraminifera, Echinodermata, some turtle groups, etc.). Establishing a single list that encompasses both extant and extinct species will be challenging, but can be achieved (Hayward et al. 2021).

Another potential complexity with dividing the global list into multiple partial lists is that there is no universally adopted or agreed-upon higher classification of life. Depending on which perspective of higher classification is accepted, there may be cases where different partial lists might have overlapping content (or, conversely, there may be gaps between partial lists where certain taxa are missing completely). Insofar as the spectrum of biodiversity is partitioned into subsets of taxa for management and governance, there will need to be some mechanism for aggregating these subsets into a single consolidated global list, such that subsets are either strictly exclusive of each other, or an unambiguous process of arbitration accommodates any discrepancies among overlapping subsets.

A list of what, exactly?

A list of names?

The system of scientific nomenclature and classification, which began over 250 years ago (Linnaeus 1753; 1758), is among the most long-standing and universally adopted standards in all of science (Pyle 2016). Governance systems for biological nomenclature have been formalized as Codes of Nomenclature (ICZN 1999; King et al. 2011; Turland et al. 2018; Parker et al. 2019). It is important, however, to recognize that two of these Codes (which together cover animals, algae, fungi and plants) confine their governance to the names themselves, and explicitly exclude rules that would affect taxonomic judgement, whereas the bacterial and virological Codes include rules for both nomenclature and taxonomy. Indeed, there is persistent confusion and conflation throughout biological sciences concerning the fundamental difference between taxonomy (the task of recognizing and classifying groups of organisms in nature) and nomenclature (the system for assigning names to those groups).

Scientific names are the key to contextualization and cross-referencing of nearly all aspects of biological information (Patterson et al. 2010). Nevertheless, as has been extensively discussed (e.g., Berendsohn 1995; 1997;Berendsohn et al. 1999; Zhong et al. 1996; Koperski et al. 2000; Le Renard 2000; Pullan et al. 2000; Gradstein et al. 2001; Ytow et al. 2001; Raguenaud 2002; Geoffroy & Berendsohn 2003; Pyle 2004; Kennedy et al. 2006; Franz et al. 2008; Franz & Thau 2010; Franz & Cardona-Duque 2013; Lepage et al. 2014), scientific names of taxa are distinct from the taxa themselves. Scientific names are text-string labels, most of which (at the rank of species and below) are rendered as multi-part names (e.g., binominal combinations of genus and species components), often with embellishments such as authorship, year, and qualifiers of various sorts or highly abbreviated (e.g., only the first letter of the genus part of a multi-part name). The taxa to which they refer are much less-well defined.

The only objective link between a scientific name and the taxon to which it refers is the type specimen (or syntype series) (see Figure 1 of Thomson et al. 2018; Figure 2 herein). But taxa are represented by many more individual organisms in nature than the type specimen(s). The full set of organisms regarded by a taxonomist as comprising a taxon depends on a variety of factors, ranging from alternative species concept definitions to purely subjective interpretations by different taxonomists (Thiele et. al. In Prep). The link between a particular scientific name and the taxon it is intended to represent is, at best, imprecise.

One simple representation of this imprecision is evident when comparing two different interpretations by two different taxonomists, one of whom is a "lumper" (who prefers a broader, more inclusive interpretation of a species), and the other a "splitter" (who prefers to partition the same set of populations into two or more species defined more narrowly). Whereas the splitter may recognize three different names for three different populations, the lumper may regard only one of the three names as accepted, encompassing all three populations. To the lumper, the other two names (regarded by the splitter as distinct species) are "subjective" or "heterotypic" synonyms. Thus, when the lumper refers to the single species name, it implies a much broader set of organisms (*sensu lato*) than the splitter's interpretation of the exact same name (*sensu*

stricto). Both taxonomists are using the name correctly in the sense that, in both cases, the same type specimen is included among the implied set of organisms. The difference is in how broad the scope is of other organisms implied to be included within the taxon referred to by the scientific name. Thus, the same scientific name can apply to more than one taxon, depending on the perspective of different taxonomists (i.e., "lumpers" *versus* "splitters").

Another problem with using scientific names as proxies for taxa arises from objective, or "homotypic" synonymy. The most common example of homotypic synonymy is when the same species epithet is placed in different genera by different taxonomists (Patterson et al. 2010). Such alternate combinations may refer to the same species circumscription (see below), or they may not, and the spelling of the species epithet may change when the gender of the alternative genera are different. Whereas heterotypic synonymy often leads to situations where the same name potentially refers to different taxa, homotypic synonymy more often leads to situations where two different names (specifically, two different combinations with the same or similar terminal epithet) may refer to the same taxon. However, in some situations the reverse may also be true. For example, in some cases of heterotypic synonymy, taking the *sensu lato* perspective of a taxon means that more than one name may have been used to represent (at least part of) the same taxon, in that the same taxon may be referenced by some authors as the accepted name, and by other authors as one of the heterotypic synonyms. Likewise, when two authors treat the same species epithet in two different genera (homotypic synonyms), it is unclear whether the referenced concept is the same, or different. In cases where both the genus combination (= classification) and the concept change, nothing about the difference in the two names reveals anything about the difference in the referenced taxa.

Heterotypic and homotypic synonymy are not the only problems with using scientific names to represent taxa. Another (though somewhat less frequent) problem is homonymy. Homonyms are examples where the exact same scientific name is applied to two completely unrelated taxa. In cases where both taxa are covered by the same nomenclatural Code, procedures are in place to resolve the homonymy and assign a different name to one of the two taxa. However, even in such cases the homonym continues to exist in historical literature and can contribute to confusion, especially when it involves the need for replacement names. Moreover, when the two different homonymous names are applied to taxa governed by different Codes of nomenclature (e.g., one animal and one plant), there are no protocols to suppress either one of them, so they both continue to remain in active use.

These are only some of the real-world problems that are encountered when trying to map scientific names to the sets of organisms they represent. Other examples include alternate spellings of the same name (emendations, changes due to gender agreement, misspellings, etc.), and treating taxa at different ranks (e.g., as a full species vs a subspecies). Yet more problems are manifest in cases where a widely-used name is found to be unavailable (or invalidly published) in the sense of the respective Code, and needs to be replaced by an available or validly published name. Similarly, names of taxa may need to change when a name with higher nomenclatural priority (e.g., published earlier) than the one in common use is discovered.

These ambiguities in mapping scientific names to the taxa they are intended to represent can cause a great deal of confusion when synthesizing biodiversity data (Patterson et al. 2010, Costello et al. 2013a). For example, Correia et al. (2018) found that between 5% and 50% of searched web pages were omitted from results when subjective synonyms were not included among the search terms. This can be particularly problematic when implementing conservation policy or applying legislation. If a single species name listed for protection may be interpreted as either a taxon with a highly restricted distribution (*sensu stricto*), or a taxon with a wide distribution (*sensu lato*), then it may be ambiguous whether the species is or is not subject to protection. Also, if a scientific name for a taxon changes (e.g., because it is reclassified or its name is replaced), or even if the spelling of its name subtly changes, serious legal ramifications may arise. Thomson et al. (In Prep) provide a more detailed discussion and specific examples of these cases.

Against this backdrop of imprecision in mapping scientific names to taxa is another more sociological conflict, which is that between the desire for scientific nomenclature to reflect the most recent taxonomic research results, and the desire to maintain a stable nomenclature. Taxonomy is, and likely will remain, a dynamic scientific endeavor just like any other scientific discipline. A necessary consequence is the need to alter nomenclature and classification to reflect new taxonomic research. But many outside the domain of active taxonomic research depend on scientific names as proxies for the taxa they routinely deal with, and such users prefer more stability. This conflict between the need to reflect current taxonomic knowledge and the need for nomenclatural stability should be carefully considered when developing policies around the maintenance of a global list of species.

In summary, although a global list will likely be composed of scientific *names*, it is important to understand that these names are not the listed *taxa* themselves; rather, they are simply text-string labels that serve as proxies for the listed taxa.

A list of taxa?

The intention of the IUBS Working Group on the governance of taxonomic lists is to establish a list of accepted *taxa* (Garnett & Christidis 2017). As explained above, a scientific name alone is an inadequate proxy for a taxon: something more is needed to uniquely represent entries on the list as taxa. But before this can be discussed, it is important to clarify exactly what taxa are.

To paraphrase U.S. Supreme Court Justice Potter Stewart's famous remark about the nature of obscenity (Gewirtz 1996), although most taxonomists have a clear conceptualization of what a taxon is and can recognize one when they see it, few would be able to provide a concise definition if pressed to do so. The term "taxonomic concept" is most often used to distinguish taxa as biological units in nature from the scientific names used to represent them. This should not be confused with terms such as "biological species concept" or "phylogenetic species concept" (e.g., sensu Wheeler & Meier 2000; see also Balakrishnan 2005), which operate at higher levels of abstraction in taxonomic discourse. The term "taxonomic concept" is here used to represent a specific set of organisms regarded as comprising a particular taxon. For example, the *sensu-*

stricto interpretation of a scientific name discussed above represents one taxonomic concept, whereas the *sensu-lato* interpretation represents a different taxonomic concept.

Some taxonomists define a taxonomic concept to represent not only a particular set of organisms, but also its complete hierarchical classification. In this sense, a species classified by two taxonomists in two different genera represent two different taxonomic concepts (because the classification differs even if the included set of represented organisms is the same). However, most taxonomists use it to explicitly represent the included set of organisms, irrespective of their hierarchical classification. For a more detailed discussion, see Ytow et al. (2001). A more explicit term often used to represent this notion is taxonomic *circumscription* (Berendsohn 1995; Pullan et al. 2000). In this discussion, the term "circumscription" refers to the circumscribed set of organisms, living, recently dead, and yet-to-be born, which collectively comprise a particular taxon (in philosophy, this is called *extension*). Within the system of Linnean nomenclature, changing the hierarchical classification of a taxon has no impact on its circumscription.

Clarifying the term taxonomic circumscription does not, unfortunately, help clarify how a taxonomic concept is actually defined. In a precise and literal sense, a taxonomic circumscription could be defined by an explicit enumeration of all of its individual members. This, of course, is a practical impossibility in almost all cases: not only would it require identifying every living organism on the planet that falls within the circumscription, it would require doing so for all antecedents back to the point where the taxon first diverged from another, and every descendant forward to the point where it may evolve into a different taxon. Thus, in almost all cases, taxonomic circumscriptions can only be defined by inference.

One common method for defining the set of organisms represented by a taxonomic circumscription is through a set of diagnostic characters. While this has long-standing precedence in taxonomic practice, it is neither precise nor stable. For example, suppose we define a new species of fish as having 10-11 dorsal-fin spines (among other diagnostic characters), distinguished from its closest relative that has 15-16 dorsal-fin spines. What happens if an individual is found with 12 dorsal-fin spines? Must it *by definition* be regarded as a distinct species, or can the original character-based species definition simply be amended? Character-based definitions of taxonomic circumscriptions, while intuitive to taxonomists, are excessively complex and imprecise for practical use in defining every individual entity on a global list of accepted taxa.

Another common way to define a taxonomic circumscription is by citing representative individual specimens (Pullan et al. 2000; Raguenaud 2002) or populations. This approach is conceptually appealing and offers a certain degree of granularity and precision when mapping scientific names to the sets of organisms the names are intended to represent; however, it is only practical for comparing similarly-defined taxonomic circumscriptions, depends heavily on the reliability of identifications of the representative organisms, and suffers the same problem as character-based definitions (when, e.g., new populations are discovered).

A list of asserted circumscriptions?

As discussed above, there are substantial logical and practical challenges to treating items on the proposed global list (and constituent partial lists) as either scientific names, or abstract taxonomic concepts. Another approach, which obviates the ambiguity of using only scientific names, but also provides a tangible unit of reference, is to focus on specific, citable, asserted circumscriptions that are themselves labelled with scientific names. Berendsohn (1995) introduced the notion of a *potential taxon*, a specific usage of a scientific name within the context of a particular treatment of that name. He proposed a standard representation of potential taxa in the form of:

Scientific name NomenclaturalAuthor **sec.** TreatmentAuthor

"NomenclaturalAuthor" is the traditionally formatted authorship of the scientific name and TreatmentAuthor is the authorship of a specific usage or treatment of a name. When a new scientific name is first published, TreatmentAuthor and NomenclaturalAuthor are the same, a special case to which Pyle (2004) applied the term *Protonym*. For all subsequent treatments, TreatmentAuthor and NomenclaturalAuthor are different. In all cases, these authorships are interpreted as citations to published bibliographic references. The "sec." qualifier is an abbreviation of the Latin word *secundum*, meaning "following" or "according to", The Latin *sensu* - "in the sense of" - is also sometimes used, but "sec." is the preferred qualifier in this context (Berendsohn 1995).

This method of representing taxa using a combination of a scientific name and a specific treatment context has been referred to by a variety of terms, including "potential taxon" (Berendsohn 1995), "taxon view" (Zhong et al. 1996), "circumscribed taxon" (Pullan et al. 2000), "taxonym" (Koperski et al. 2000), "taxonomic reference" (Le Renard 2000), "NameRecord" (Ytow et al. 2001), "assertion" (Anonymous 2002; Pyle 2004), and "taxonomic name usage" (commonly abbreviated as "TNU"; Pyle & Michel 2008; Krell & Pyle 2010; Pyle 2016; Alvarez & Luebert 2018; Senderov et al. 2018). Whatever term is used (hereinafter we use "TNU"), the unit of a specific *treatment* of a taxon (sensu Catapano 2010) is widely recognized as the most granular and direct (as well as historically *de facto*) citable entity to represent a taxon, taxonomic concept, or taxonomic circumscription [see Ytow et al. (2001) and Pyle (2004) for more detailed discussion].

One of the first publications to robustly incorporate a system of referencing specific taxonomic name usages as units of taxonomy was that of Koperski et al. 2000 (who applied the term "taxonym"). According to Geoffroy & Berendsohn 2003, this was the first time "an in depth analysis and statistical evaluation of the stability and instability of names and concepts" was produced; and led "to the somewhat depressing conclusion that for at least 45% of the taxa there have been changes in the concept over time" (Geoffroy & Berendsohn 2003). This study revealed the magnitude of the problem of how scientific names, by themselves, are imprecise labels for the taxon concepts they are intended to represent. The fact that few similar analytical studies have been published since, both underscores the challenge faced by any effort to assemble a global

list of taxa, and the vital role that TNUs (by whatever name) will play in constructing and managing such lists.

Lepage et al. (2014) described these "unique name/source combinations" (TNUs) as "shallow" representations of taxonomic concepts/circumscriptions. While they acknowledged them as suitable "concept labels", they noted that many TNUs represent "congruent" taxa (meaning they represent the same circumscription), and described an approach to collapse sets of congruent TNUs into clusters that collectively represent a "deeper" notion of a taxonomic concept. Although this approach to stacking multiple TNUs into a single entity is valuable for conducting reasoning across disparate datasets (e.g., see Chen et al. 2014; Franz et al. 2016), it has been limited to relatively small and well-studied taxonomic groups (e.g., birds, plants). Moreover, it is very difficult (and, indeed, may not be possible) to define the relationships between many TNUs any more precisely than to say that they overlap in some way. For this reason, identifying relationships among TNUs may not increase power for reasoning over data.

For the purposes of assembling, managing, and maintaining a single global list of taxa, both a label and a definition for each entry in the list need to be established. Decades of effort by the biodiversity informatics community has not yet produced a single method for organizing taxonomic units that is meaningful, practical and scalable. Each item on the global list must be understood to refer to a taxon (concept/circumscription); however a list of simple scientific names is insufficient, and a list of robustly-defined taxonomic concepts cannot be scaled to all of biodiversity within the foreseeable future. Therefore, some compromise between these two extremes must be established to empower a global list with practical value, without succumbing to practical difficulties of construction and formulation.

The core problem is the scalability of the task. There are millions of taxa on Earth, tens of millions of names for those taxa, and hundreds of millions of TNUs for those names as applied to those taxa. Mapping every TNU to every other relevant TNU would be an almost incomprehensibly large (and entirely impractical) task. However, a global list of taxa (comprising multiple partial lists) may provide an important opportunity: a milestone that is entirely achievable would be to have every entry on the global list anchored to one TNU that represents the taxonomic concept indicated for that particular list entry. Simply by qualifying each entry on the list with "sec." and a TreatmentAuthor representing a specific TNU, each list entry becomes more than a simple name: it represents a specific taxonomic concept. This would allow a reference point for other TNUs representing congruent taxa to be mapped, allowing most of the problems of ambiguity between taxa and their TNUs can be resolved.

Establishing a label for each item on the list is relatively straightforward. With more than 250 years of successful legacy, Linnean-style scientific names should clearly serve as the foundation for any labelling scheme. Although such names by themselves are insufficient, when qualified with a "sec. TreatmentAuthor" following the style of Berendsohn (1995) (as described above), the full-context TNU/treatment captures both the familiar label of a scientific name as well as the taxonomic context (and implied concept or circumscription) to dramatically reduce ambiguity of what "sense" of the name (e.g., *sensu stricto* vs. *sensu lato*) is implied for the item on the global list. Moreover, this approach only requires the addition of a single element of metadata to

accompany each entry on the list (i.e., a citation for the designated treatment of the name as represented on the list). Thus, each item on the global list of accepted taxa could be represented in the form of: "*Scientific name* NomenclaturalAuthor sec. TreatmentAuthor" (where the format and content of NomenclaturalAuthor follow the rules of the relevance nomenclatural Code, and TreatmentAuthor represents the citation for the specific asserted circumscription of the name).

A more significant challenge will be in determining which particular TNU to reference, when dozens (or even hundreds) of TNUs representing congruent published taxonomic circumscriptions are available to choose from. A related challenge is in representing circumscriptions that do not explicitly exist in published literature. A thorough discussion of these related challenges is beyond the scope of this paper. However, a few general guiding principles for crafting the global list (and constituent partial lists) can be suggested.

First, when a new species is described it will in most cases represent one of the following general scenarios:

a) a brand-new discovery that has no impact on the circumscriptions of existing entries on the list;

b) a new taxon circumscription that is taxonomically adjacent to that of an existing entry on the list (i.e., a case where a lumper might regard it as an expansion of an existing circumscription, whereas a splitter might regard it as a new discovery); or

c) cases where either a well-established circumscription is divided into two or more components (splitting), or one or more well-established circumscriptions are merged into a single circumscription (lumping).

The latter two (and especially the third) will require updates to the particular "sec." treatment indicated on the list (e.g., switching from a lumper TNU to a splitter TNU, or vice-versa). Each provider of content to the global list will need to account for cascading ramifications to other entries on the list. For example, if two distinct species (*A* and *B*) are recognized on the global list, and a decision is made to merge them into one species, then one of the names (e.g., *A*) will remain on the global list but must be amended with a different "sec. TreatmentAuthor". This change has a cascading effect on entry *B*, which had previously appeared on the list as distinct species, but is now regarded as a heterotypic synonym of *B*. Thus, a change in the TreatmentAuthor of *A* causes the removal of *B* from the list. Likewise, if *A* is split into two taxa (*A* and *B*), then the entry for *A* on the list will require a new "sec. TreatmentAuthor" to indicate that it has a different circumscription, and entry B will need to be added, with appropriate "sec. TreatmentAuthor".

Second, when selecting from among multiple TNUs for a particular entry on the list, it will be best to choose one that has a robust corresponding published treatment, including detailed geographic range information, numerous specimens as "material examined", and a complete (global, not regional) heterotypic synonymy. This last point is perhaps most important, because complete heterotypic synonymies offer a crude but effective name-based mechanism of objectively comparing congruency (or not) between two TNUs referencing the same scientific

name. Each name included among heterotypic synonyms serves as a proxy for its corresponding type specimen(s), and as such confer a taxonomic circumscription of a name that goes beyond the type specimen of the "accepted" name used to represent the taxon. Such TNUs with robust information about circumscription boundaries cannot account for changes asserted later (especially when new names that potentially infringe on those boundaries are subsequently established), and there will always be a need to balance the need for robust treatments with those reflecting the most recent research, but treatments that include complete synonymies are extremely effective for serving as reliable "anchor-points" for taxa up until the time of the TNU publication.

There are many technical issues associated with interpolating, extrapolating and cross-referencing taxa via TNUs, but these are beyond the scope of this paper. For better or for worse, most taxa on the global list (e.g., most invertebrates, protists, etc.) will not encounter the problem of multiple alternative circumscriptions or multiple congruent TNUs. Indeed, most entries on the list are likely to be protonyms (see above). Those groups of taxa for which arbitration among alternate TNUs is an issue suffer primarily from a taxonomic "embarrassment of riches". But they are also often the groups of greatest socio-economic importance, and therefore often have a more acute need for resolving potential taxonomic disputes.

Defining the Scope

Several aspects of the scope of an envisioned global list of accepted taxa need to be defined. These include breadth, granularity, "level of confidence" (for lack of a better term), and whether taxa without formal scientific names and fossil taxa should be included on the list. Each of these aspects requires careful consideration, and should be determined only through engagement with constituent communities. The extent to which these and other details are applied consistently across all groups, or are optimized for the specific circumstances of each group, also remains an open question. For purposes of discussion, we provide general commentary on these aspects to facilitate future discussion.

Breadth

The question of breadth involves how broad the scope of life-forms included on the global list will be. The global list should minimally be open to all prokaryotic and eukaryotic organisms, and some support the inclusion of viruses as well, given that both names and circumscriptions of viruses are well-organized, and they play an important role in biodiversity and medicine (as the world population of humans was overtly reminded in the year 2020).

Granularity

The question of granularity in this context involves whether the global list is limited only to species, or whether it should also accomodate scientific names and associated taxonomic circumscriptions of higher-rank taxa (e.g., families, orders, classes, phyla and kingdoms) and lower-rank taxa (e.g., subspecies, varieties, forms, etc.). We suggest that the list should be open to taxa at all ranks, both above and below species rank. Accommodating names (as represented

by TNUs) for higher taxa is necessary for embedding a classification scheme within the global list, and accommodating lower taxonomic ranks is important for certain groups (especially vertebrates and plants) where such taxonomic divisions can have important conservation implications (e.g., Thomson et al. ==In Prep.==).

Level of confidence

The question of "level of confidence" has some overlap with the granularity question, but includes some aspects that are related to how confident taxonomists are as to the status of the entities on the list. In some cases, species are understood to represent valid taxa, but their placement within a genus is uncertain. Such species could be included on the list as *incertae sedis* within a higher-rank name, or they could be provisionally included within the genus in which they were first described or most recently combined, or they could be excluded from the list pending a proper published taxonomic treatment. Similarly, newly published species that have been questioned by the relevant taxonomic community may be added only provisionally, excluded entirely, or added without qualification. Decisions of this sort will ultimately be made by the communities who maintain the partial lists, but a uniform set of principles should be developed and applied across all entries on the global list.

Names outside the scope of Linnean nomenclature

In addition to standard Linnean scientific names, there are several formal and informal systems for nomenclature used to assign labels to specific taxa. One formal class of such names is plant cultivars, which are governed by an international Code of nomenclature (Stearn 1953; Trehane 2002; Brickell et al. 2009; Spencer and Cross 2007). Similarly, certain trade names have important meaning and standing in some countries or taxonomic groups (e.g., Page et al. 2013; Callomon 2016; Chesser et al. 2017). Several projects focused on DNA sequencing data processing and the associated regulatory genomics analyses have been combined in a initiative called "SeqCode" (https://github.com/seqcode), and similar initiatives have been applied for uncultivated Archaea and Bacteria (Murray et al. 2020).

Another important class of informal names are so-called "provisional" (sometimes referred to as "temporary" or "tag" names) as applied to taxa. Most such names involve labels where one component of the name consists of a Linnean-style scientific name (e.g., genus or family), and the remainder of the name consists of non-Linnean designators (e.g., "sp.1234"). In many groups, the process of formally naming taxa lags behind the discovery of taxa (because taxonomic due diligence is necessarily work- and time-intensive). Many taxonomists, particularly in relatively species-rich groups such as arthropods, marine invertebrates and plants, make reference to species that have not yet been formally assigned a scientific name. Giving informal names to these taxa before they are formally named under the Codes of nomenclature is sometimes important (e.g. if they are rare). Such informal names may appear frequently in literature, and sometimes in legislation. While a case can be made that such names should be excluded from a global list of accepted taxa, this would limit the ability to refer to their taxa in potentially important contexts, such as conservation and management legislation. A global list of names, including

informal ones, provides an important opportunity to control and regulate such names to reduce any ambiguity in their application.

A special class of informal names are higher-rank names that, while widely used, have not yet been formally established. Examples include Bilateria, an important name for all animals that have a bilateral rather than radial body plan. Bilateria is an unranked, informal name, because the higher-rank classification of animals remains in a state of flux. Such informal names are abundant at the highest levels of our current classification system and, while they have few day-to-day consequences for users of a global list of species, they will be important for organising it.

Extinct taxa based on fossils

Another question of scope is whether or not the global list should be limited to extant (or recently extinct) taxa only, or if it should be open to all named taxa including long-extinct ones represented only by fossil specimens. Excluding all extinct taxa would result in a disorderly fringe to an otherwise ordered system. For example, should the list include recently-extinct taxa such as the dodo, and if so, how far back should it go? (The IUCN has adopted the year 1500 AD as a cut-off). On the other hand, extinct taxa are generally no longer subject to conservation or protection concerns, so their inclusion may be unnecessary in the context of the needs that led to the call for governance of the global list of taxa in the first place. However, the needs of conservationists and resource managers are not the only purposes of the proposed global list. Extinct taxa that have extant relatives are used in phylogenetics to calibrate the branching of the tree of life, so their inclusion could be useful and valuable for broader context. A compromise position could be to have the list of long-extinct taxa developed and managed separately from the list of extant (or recently-extinct) taxa for groups where there was no tradition of integration. Because many living taxa also occur in the fossil record, and some may occur in the same samples (e.g. marine sediments contain living and fossil Foraminifera), it makes sense to include both extant and extinct, fossil and non-fossil, taxa in a global species list, especially if they can be identified as such within the list.

Do all taxa even need to have a name?

Perhaps the most fundamental question of scope is whether to restrict the entries in a global list of taxa to only those taxa that *have* names. As already discussed, taxa and their names are different things, and taxa that lack standard scientific names are potentially important to include on a global list. Moreover, to be useful, a global list should include a classification (that is, it should not be a mere alphabetical list of accepted species names).

An important development in modern taxonomy is the field of phylogenetics, with its tools for classifying taxa and inferring patterns of evolution. Before phylogenies were developed, taxonomists' knowledge of the structure of life were captured in *classifications*, based first on a limited set of Linnaean ranks (kingdom, phylum, class, order, family, genus, and species) then, as knowledge increased, on an open-ended set of interpolated ranks (sub-order, super-family etc.). Much knowledge of the relationships between taxa and the structure of life are now captured in *phylogenies*. And most nodes in phylogenies are un-named. The question remains open as to

whether un-named nodes of an hypothesized phylogeny should be included as entries on the global list.

Defining the Information for each Entry on the Global List

Once the definition and scope of entities to be included on the global list of accepted taxa are determined, another series of questions concern the data and metadata that are required or recommended for each entry on the list. Several relevant data standards have been developed to represent the key properties of scientific names and their associated implied taxon concepts. Broad biodiversity data standards such as DarwinCore and ABCD (Wieczorek et al. 2012; Holetschek et al. 2012; Blum et al. 2019; Fichtmueller et al. 2019) include many relevant terms and associated definitions to capture taxonomic information. More targeted standards, such as the Taxonomic Concept Transfer Schema (TCS; Kennedy et al. 2006), Open Biomedical Ontologies Taxonomy ontology (Obo-Taxonomy; Midford et al. 2013), and Catalogue of Life Data Package (CoIDP; Döring & Ower 2019) include other important defined terms and conceptual relationships. While these and other data standards should serve as the foundation for a global list of accepted species, a subset of standard properties should be recognized as representing the minimum information necessary for each entry on the list. We suggest the following are consistent with both the existing standards and the principles for global lists (Garnett et al. 2020b), but emphasize that these are recommendations open for broader discussion.

Persistent Identifier

Each item on the global list must be branded with a globally unique identifier with unambiguous and persistent mechanisms to dereference (i.e., resolve or retrieve metadata). Several different kinds of unique identifiers (both with and without associated dereferencing mechanisms) have been adopted in the biodiversity informatics community. The form of the identifier is not suggested here, but whatever system ultimately is adopted should be robust and consistent with biodiversity information standards and best practices.

Version History

Each entry on the global list must have a robust "audit trail" so that changes to the lists and associated properties and metadata can be tracked through time.

Label

Every item on the global list must include a unique label representing either a scientific name (compliant with the relevant nomenclatural Code), or a standardized informal taxon label that is widely recognized as such. In cases of Code-governed names, the spelling used should be in full compliance with the relevant Code. Each entry on the list should include more properties than just a scientific name, such as the original authorship of the name formatted in accordance with the relevant Code, and an additional "sec." TNU qualifier, as well as some additional minimum properties as described below.

Nomenclatural Authorship

Whereas the Label as described above should be constructed of certain key elements, these elements should be managed as distinct properties. In addition to a scientific name, the nomenclatural authorship, formatted in accordance with the relevant nomenclatural Code should also be included with each entry on the global list. Such authorship information is important for disambiguating homonyms, and for helping to link original literature. The formatting of authorship should follow the recommendations of the relevant nomenclatural Code and be consistent with relevant community practice.

Source

Each entry on the list should minimally indicate the source of the entry. This should be a link or textual description of the database or working group that provided the entry to the global list, with some minimal amount of associated properties of the source, or a link to source metadata.

Nomenclatural Code

Each item on the global list should include an indication of which nomenclatural Code (or Codes) applies to the associated scientific name (if any). In most cases, each item will be represented by a name governed by a single nomenclatural Code, but in some cases (i.e., ambiregnal taxa; Patterson & Larson 1992) more than one nomenclatural Code may apply. Depending on whether the scope of the global list includes or excludes informal names, some entries may be represented by names not governed by any nomenclatural Code. A link to an associated nomenclator (e.g., ZooBank, IPNI, Index Fungorum, etc.) is very helpful, but should not be mandatory.

Original Ranks and Combinations

In cases of entries on the list represented by names below the rank of genus, information about the rank and combination for the original establishment of the name (e.g., the basionym combination, or names treated as subgenera/subspecies that were originally established as full genera/species, and vice-versa), should be included when available, but such information should not be mandatory.

Original Literature Citation

A full literature citation of the original establishment of the name (e.g., the basionym citation) should be included whenever possible, either as a full textual citation or via a standard unique identifier such as a Digital Object Identifier (DOI). It will be impractical to regard this as mandatory for all entries on the list.

Accepted Status Treatment Citation

A citation to the treatment representing the current/accepted taxonomic status of the entry (i.e., the "TreatmentAuthor" part following the "sec." qualifier of a TNU label) should be included with

each entry. This should include (or be resolvable to) a full literature citation or standard identifier such as a DOI, but minimally should include the authorship and year of the publication that is cited for representing the taxonomic circumscription. When this is not available, the source of the named taxon (i.e., the database or working group that provided the entry to the global list) could be used in place of a literature citation (although published treatments are strongly preferred). Careful thought will need to be given to standards regarding the scope of sources that can serve as "Accepted Status" treatments. Would it require that compilers of partial lists enumerate and acknowledge all possible alternatives to demonstrate a well-informed decision on the choice of specific accepted treatment? Must such treatments be limited to scholarly works published in the traditional sense, or could the source be an online database, unpublished document, or even a documented "personal communication" from a recognized expert? There are many points to consider in this context, but some guidelines will be necessary for compilers of partial lists to follow, to ensure consistency of reliability across the different partial lists.

Confidence

An indicator representing the degree of confidence for the assertion of the current status of the entry would also be valuable. At minimum, a distinction of "accepted" should be assumed, as well as some confidence indication of that status assessment, but it remains to be determined how granular such an indicator would need to be (examples include a simple "green/yellow/red" indication of confidence, or a numeric value drawn from a continuous scale). Additionally, there are different axes of confidence, such as whether a name conforms to the relevant Code, whether the taxon is broadly accepted by the relevant community or subject to contention, levels of confidence with respect to taxonomic rank, and other factors. Moreover, depending on the scope of treatments open to evaluation as the accepted status (see previous section), additional qualifiers may be needed to indicate the reliability of the accepted status treatment. Thus, there may need to be multiple different confidence indicators.

Synonymy

Each item on the list should be represented by a scientific name (or other accepted form of name) that is clearly displayed as the "accepted" representation of that name. In other words, names regarded as junior synonyms of other accepted names should not themselves constitute separate entries on the list. Instead, names representing taxonomic synonyms should be included as part of the metadata along with each name, whenever that information is available. Including full heterotypic synonymy information along with each entry on the global list will be very helpful for recognizing the circumscriptions of the taxa in the list. In all cases, synonymies should be consistent with the associated referenced treatment and should whenever possible include associated additional reference citations associated with each synonym. Homotypic synonyms, misspellings and other orthographic variants would also be helpful to include as metadata with each entry on the list.

Classification

In addition to a unique and commonly recognized label, each entry in the global list should include an indication of hierarchical classification. Minimally, all names on the list representing taxa below the rank of genus (i.e., subgenera, species, and named taxa at infraspecific ranks) should include sufficient hierarchical links to the rank of genus (e.g., no "unplaced" species should be included). Genus and other higher-rank taxa should also include information about taxonomic placement, thereby establishing a complete classification for all entries on the list. As with the taxa themselves, an indicator of confidence for the placement in the classification would be useful, because in some cases taxonomists may be confident about the circumscription of a taxon but less confident of its hierarchical placement.

Geographic Distribution

Including geographic distributions (including distributions with respect to political regions or conservation areas) for each entry on the list would be very valuable, both for defining circumscription in some cases and to allow extraction of geographically-based lists to be conveniently extracted. Geographic distributions are poorly known in some cases, and only known distributions can be included.

Images

Images of specimens can be very helpful for representing taxa on the list, but images would be best dealt with by specialist image services that *use* the list (rather than in the list itself).

Annotations

Many online data aggregation systems include support for annotations and comments to be applied to individual entries and their attributes. Such annotation systems would be extremely valuable to engage users of the global list and to help curate it by reporting errors and providing missing details. An annotation system would need to be monitored and moderated in order to be effective.

Associated Vernacular Names

Except in cases where an item on the global list is represented by an informal name (if such entries are allowed), additional vernacular names should not be included with any item on the global list.

Type Specimen(s)

Although type specimens are a fundamental and integral component of scientific nomenclature and serve as critical anchor points to objectively link scientific names with biological entities (see Figure 2), they only play a very limited role defining circumscription boundaries for taxa included on the global list, and are primarily interesting to taxonomists who focus on nomenclature *per se*. Thus, information on type specimens should remain within the domain of nomenclators and be outside the scope of a global list of accepted taxa.

Character Diagnoses

Although taxonomic circumscriptions are often defined through descriptions of diagnostic characters (both morphological and biochemical, including genetic), inclusion of detailed diagnoses of this sort are outside the scope of data to include with entries on this global list. As with images, the global list will become an extremely valuable service to other systems that manage descriptive and diagnostic information.

Species Concept

This discussion has focused on taxonomic concepts in the sense of circumscriptions of organisms, but has avoided any discussion about the relative merits of alternate "species concepts" in the sense of Wheeler & Meier (2000) and numerous others (i.e., biological vs. phylogenetic species concept, etc.). Although it might be considered valuable for entries on the list to be annotated with the species concept used in determining its circumscription, such information is unlikely to be available for the vast majority of entries on the list, and has the potential to be both contentious and unworkable. An explicit indication of the intended species concept should not be mandatory.

Creating and Maintaining the Global List

Perhaps the greatest technical challenge for implementing a global list of accepted taxa, after determining its structure, scope and properties, is establishing procedures for its creation and maintenance. The most fundamental (and arguably most important) question is whether the standards and procedures for curating the global list are applied at the level of the taxonomic communities that provide partial lists, or at the granularity of individual entries on the list.

If the former, standards and procedures would be designed around the processes by which taxonomic communities evaluate the accuracy and legitimacy of the content they provide through a partial list. Communities would ensure that the content they provide meets appropriate quality checks and principles adopted by the global list, and exclude content that fails to meet these standards. If the latter, standards and procedures would be applied individually to each entry on the list. Instead of providing an "all or nothing" contribution, taxonomic communities who maintain the partial lists would assemble and verify the necessary minimum required properties and metadata (as well as additional optional properties and metadata), and every item that meets the minimum criteria would be accepted. Placing emphasis on individual entries rather than than the entirety of individual partial lists allows more flexibility and assurances in adding and maintaining content on the global list, but also introduces additional complexities.

These two options are not mutually exclusive. For example, one mandatory requirement for individual entries could be that each item must have passed through the partial list of a taxonomic community that itself adheres to the relevant standards and procedures (e.g., review, transparency, and conflict resolution requirements). But in addition, each item added to the global list may also be required to fulfill certain minimum criteria (e.g., minimum properties and metadata, some metric of consensus achievement, etc.) In this combined approach, items

provided by a standards-compliant community may or may not be accepted onto the global list. Items that fail the acceptance criteria could be excluded from the global list completely, or could be flagged as incomplete (e.g., 'provisional' or 'pending'). A combined approach would share the burden of quality of content and allow a more granular acceptance or rejection of individual items instead of an 'all or nothing' approach.

Regardless of whether standards and procedures are applied to partial lists as a whole or to individual entries (or a mixture of both), certain key issues will need to be addressed through standards and procedures. The following issues warrant further discussion.

Identifying and Completing the Gaps

It will be difficult to find recently updated expert curated species lists for some taxa. One approach would be to exclude such taxa from the global list entirely. A more useful option for users would be to include such taxa on the list, but note that they need specialist review, perhaps because they are based on decades-old publications (though we note that just because a list may be old does not mean it will be inadequate or of poor quality).

Newly Described Species and New Taxonomic Arrangements

An estimated 15,000–20,000 new species are described each year (including fossils; Costello et al. 2013b), and many more species are taxonomically revised with changes to classification (including genus placement), change in rank, and changes in synonymy. Depending on the taxonomic group, the addition of new species and alteration of existing species will be evaluated and added to relevant partial lists (and, therefore the, global list) at different rates. New species of conspicuous, charismatic, or otherwise important taxa are likely to be incorporated quickly, but more obscure and cryptic species may languish. In the case of newly described species, one approach would be to accept all newly described species and by default, and "demote" them to junior synonym status only if warranted by subsequent evaluation. Another approach would be to adopt new species and taxonomic arrangements to the list only after a consensus of subsequent usage of the new names is established. Such an approach would require threshold criteria consensus, such as number of subsequent publications, degree of consistency of treatment, and extent of necessary or acceptable delay. The communities that maintain partial lists are likely to be best-placed to determine their own processes for incorporating newly described and revised taxa, but standard best practice should be to encourage new species and new taxonomic arrangements to be added quickly, especially in cases where conservation or management implications are high.

Peer Review

Peer review takes many forms, operates in several formal and informal layers, and has broad and complex implications. Although a foundation of the scientific method, peer review is often inconsistently applied and implemented, especially in taxonomy. Several of the Codes of nomenclature do not require peer review for newly proposed scientific names. The question of whether or not taxa become accepted emerges over time by community consensus, which is a

form of peer review. Several layers of peer review come into play in the context of global lists. The first is at the original scientific publications through which taxonomy is produced. As previously noted, there are no consistent standards for how (or even whether) such published literature is subjected to peer review. Another layer of peer review is provided by the communities that maintain the partial lists. In an ideal circumstance, these communities would apply standards to assess both publication-level peer review and the content of the published work itself. An additional layer of peer review could be applied after taxa are added to the global list, as feedback from list users and others is used to assess acceptance. Elements of peer review will most likely be applied at all of these levels, but a set of core metrics for a minimum standards of review for taxa before they are added to the global list should be developed and implemented. Minimum criteria should be whether names meet with the requirements of the relevant Code of Nomenclature, whether the records and associated metadata (including references to sources) are complete and free from editorial errors, and whether reviewers have confidence that the content is accurate according to current knowledge.

We acknowledge that, as with peer review of any publication, reviewers and editors cannot check every observation and datum and, therefore, must trust that the authors honestly reported their results. Where errors, discrepancies or questionable information is detected, the species list could be subject to closer scrutiny and its editors asked to respond to reviewers' comments. This may result in revisions or rejection of list entries that are considered inadequate.

Arbitrating Taxonomic Disagreements

Inevitably, disputes in how some taxa should be represented on the global list will emerge. Such disputes are an inherent and not necessarily undesirable aspect of taxonomic practice (Thiele et. al. In Prep). Mechanisms for the arbitration of such disputes depends on which fundamental model is adopted for content creation and curation.

If standards and procedures are applied to the taxonomic community that provides and maintains a partial list, the onus for achieving consensus and arbitrating disputes would fall on each respective taxonomic community. If different factions within a community produce and maintain alternative competing lists, standards would need to be in place to either endorse one faction or to ensure that each faction is represented in producing a consensus partial list. If, however, standards are applied at the level of individual entries, then a more granular approach (however complex) could be implemented.

Governance of a global list is a complex process and can be achieved in many ways (Lien et al. In Prep.). For a global list to be widely adopted and maintained, solutions for arbitration among competing or alternative interpretations must be carefully designed and implemented by a wide array of stakeholders and other constituencies. In this paper we do not propose specific solutions to such complex issues but, rather, highlight technical considerations when designing a functional dispute resolution process.

An important consideration for establishing any system of arbitration is to separate facts from opinions. For the purposes of this discussion, facts are the properties and information that are

(mostly) objective. For example, whether or not a name is correctly published and available under the relevant Code of Nomenclature, or whether it has the correct Code-compliant spelling is, in most cases, a matter that can be determined objectively. By contrast, taxonomy is necessarily rife with opinion, much of which is legitimately subjective (Thiele et. al. In Prep). Most arbitration will involve selecting from among two or more alternative opinions about circumscription and classification. A key consideration, which could prove vital in the arbitration process, is grounding opinions in specific TNUs, rather than simply through the editors of partial lists. While it may be an entirely subjective opinion whether two populations are best treated as one or two species, or whether a species is best placed in one genus or another, it can usually be objectively declared that one publication treated a taxon one way, and another publication treated it in another way. These are facts about taxonomic opinion, best captured as TNUs, and represented in the global list through the TreatmentAuthors property (the citation following the "sec." part of a taxonomic treatment label). For example, it's a fact that Author "X" treated species "Y" as a valid taxon, and within a particular genus. Whether or not this treatment represents the "accepted" taxonomic arrangement is an opinion.

While this does not obviate the need for arbitration, it does clarify the functional steps of arbitration. Instead of the arbitration process directly determining whether a particular species name should be regarded as representing a valid species or as a synonym, or determining which genus a particular species should be placed within, the question can be framed in the form of 'which published treatment of a name best captures the accepted status?' While this distinction may seem trivial and academic, it has important implications for the arbitration process. Rather than amassing a collection of opinion and rationale for supporting a particular decision, the arbiters need only determine which taxonomic treatment, among competing alternatives, makes the most compelling case. While this final step certainly represents an opinion, it anchors that opinion to objective facts. This doesn't mean the opinion is any more or less likely to be "correct", but it does allow much more robust documentation for the basis of the opinion. In summary, the arbitration of disputes of fact will follow different protocols than the arbitration of disputes of opinion. And the latter will be greatly facilitated by abstracting the collections of opinions in the form of discernable facts, such as individual TNUs and their associated properties.

The local and the global

While, as noted previously, partial lists must be global in scope, there is an important consideration for regional influence. One of the ten principles for the governance of a global taxonomic list enumerated by Garnett et. al. (2010) is:

> 10. *A global listing process needs both to encompass global diversity and to accommodate local knowledge of that diversity*.

This can be important when dealing with regionalised taxonomic lists, some of which may be of very high quality. For example, the Australian botanical community maintains the Australian Plant Census (APC; Council of Heads of Australasian Herbaria 2010) using practices that conform closely with best-practice governance as envisaged for the global list: curation of the APC is broad-based and national in scope, with clear mechanisms for peer review, consensus decision-

making, and dispute arbitration. For these reasons the APC is widely accepted in Australia as the normative list of Australian plants. While some contributors to a global list of plants may have an important global perspective of particular plant groups, Australian contributors to the APC have detailed knowledge of Australian species and their circumscriptions.

A global list that accommodates regional knowledge is likely to be significantly more useful and accurate, both regionally and globally, than one that fails to do so, particularly for taxa that are endemic to the region. However, incorporating content from such regional lists should fall primarily within the purview of the communities who maintain the partial lists, rather than directly through the management of the global list itself.

Accessing and citing the Global List

All content represented on the list should be freely accessible and downloadable, and available under one of the standard Creative Commons (CC) licenses. The most liberal CC license is "No Rights Reserved" (CC0), which is effectively "public domain". The value and utility of the global list and its associated content is greatly enhanced if it is freely accessible and distributable. Any restrictions placed on the use of the list or its associated content will have a negative impact on its wide adoption. Thus, a strong argument can be made that the content of the global list should be designated as CC0.

At the same time, an enormous amount of work goes into formulating partial lists for incorporation into a global list, and attribution of such effort is a compelling and important motivator for those who would contribute to this effort. For this reason, another CC option is the "Attribution" license (CC-BY), which "lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation". While certainly appealing, one problem with this license as applied to a global list of names (and the entries thereon) is: 'who should receive the attribution?' Many different individuals play a role in the process of establishing each entry on the list, from the person or people who first discovered and collected the original specimens, to the taxonomist(s) who first described it, to the taxonomist(s) responsible for asserting the current status, to the compilers and editors of the partial list, and to the managers of the global list itself. In many cases, the naturalists, explorers, and collectors provide the most important and costly (both financially and personally) contributions to the process (Conniff 2011, 2021), yet are seldom featured in the attribution chain of any taxonomic name (often relegated to scant mention in the "collectors" property of a specimen record). The next greatest contribution to the process is usually in the form of the original description of the taxon, and the subsequent taxonomic revision that established its current status. Attribution for both of these roles is embedded within the core properties of the list entries themselves, in the form of the NomenclaturalAuthor and TreatmentAuthor properties. Although nowhere near as significant on a taxon-by-taxon basis as the contributions by the discoverers, collectors, nomenclatural authors and taxonomic authorities, the role played by the compilers of the partial lists is nevertheless significant and important. This attribution is also included among the properties of each list entry, in the form of the Source property. Thus, the only role in the process not already included within the global list content is the role of whatever overarching entity administers and manages the global list itself. Compared with the other roles already included by

attribution within the data, this role hardly seems significant enough to warrant a CC-BY license designation for the global list.

The CC licenses include other provisions in addition to Attribution, including "Share Alike" (SA), "No Derivis" (No Derivatives; ND), "Non-Commercial" (NC), and various combinations of these extensions. Restricting the creation of derivatives of the global list would be counterproductive to its fundamental intentions. Restricting commercial re-use would apply to situations such as a consulting firm providing environmental impact assessments, or an NGO website that offers merchandise to cover its costs. While there are situations for which researchers and their institutions are likely to want to restrict commercial usage, the CC-BY-NC license is too blunt an instrument for achieving the desired goal. The requirement for sharing derivatives would make sense in the context of the global list, but it only exists as an extension of CC-BY (i.e., there is no recognized "CC-SA" license). Therefore, unless a decision is made to enforce the CC-BY license on the global list, these other extensions are not relevant.

Perhaps the most sensible approach to managing the licensing for the global list is to declare it as CC0, but strongly encourage users (through clear recommendations and social and academic norms) to treat it as if it were covered by the CC-BY license. An important step to encourage this would be for the underlying data infrastructure to give users the tools that make it simple for them to give an appropriate citation. At the very least, a strong recommendation for preserving the various inherent attribution components (nomenclatural and treatment authorships, as well as the partial-list source) on subsequent distribution could be asserted. Additionally, encouragement for referencing the source of the global list as a whole could mirror normal practice within scientific and related information sources to cite sources. Whatever licensing is applied to the global list, it should be applied consistently for all records in the list equally. There is no rationale for asserting different license provisions for different parts of the global list, and application of such a heterogenous approach would be intractable and impractical. Also, because the list will change over time, any reference to it in part or as a whole needs to indicate a time-stamp for when it was accessed.

Rally around the Catalogue of Life

Most of the issues and questions articulated in this discussion have already been confronted and, to a greater or lesser degree, addressed by the Catalogue of Life (COL). In many respects, the goals sought by the IUBS Working Group on the Governance of Taxonomic Lists are aligned almost perfectly with those of COL. This is not to say that all the features of the envisioned global list have already been fulfilled by COL, nor that COL by itself represents the entirety of a widely accepted global list. However, more than any other initiative in taxonomy, the COL team has broader and deeper experience in addressing and solving the devilish details described in this paper.

A more detailed description of COL, its current implementation, and its future directions, is provided by Hobern et al. (In Prep.). The efforts and ambitions described in this series of papers should not, in any way, be construed as an effort to replace COL with something new. Indeed, many of the authors of these papers have decades of experience with COL, and are among the

world's experts in solving the many complex issues that an effort to establish a global list of accepted species will face. Instead, we seek harmonization between the ongoing efforts of COL with the broader constituency of content consumers represented by the IUBS initiative that has precipitated this series of papers (and others). Whereas COL represents the unambiguous foundation for implementing a global list of accepted names of taxa, the dialog emerging from this series of papers provides a broader scope and purpose of need than what COL has previously prioritized. In many ways, the convergence of the IUBS Working Group with COL establishes an opportunity to fulfill the original vision of COL in an expedited way. An overarching message from this paper is that future efforts should fully leverage the work already accomplished by COL, and provide robust support for the refinement of its goals and products to more effectively fulfill the needs of broad communities of biologists, conservationists, resource managers, and many others.

Acknowledgements

References

Alvarez, M., & Luebert, F. (2018). The taxlist package: managing plant taxonomic lists in R. *Biodiversity Data Journal*, 6, e23635. https://dx.doi.org/10.3897/BDJ.6.e23635

Anonymous (2002). VegBank Taxonomic Data Models. Ecological Society of America. Retrieved May 31, 2021, from http://www.vegbank.org/vegdocs/design/planttaxaoverview.html

Ardelean, A., Kervin, K., Kansakar, S., & Fautin, D. G. (2009). Syngraph: An application for graphic display and interactive use of synonym lists. *Zootaxa*, 2283, 29–39. https://doi.org/10.11646/zootaxa.2283.1.3

Balakrishnan, R. (2005). Species concepts, species boundaries and species identification: a view from the tropics. *Systematic Biology*, 54(4), 689–693.

Berendsohn, W. G. (1995). The concept of "potential taxa" in databases. *Taxon*, 44, 207–212.

Berendsohn, W. G. (1997) A taxonomic information model for botanical databases: the IOPI Model. *Taxon*, 46, 283–309.

Berendsohn, W. G., Anagnostopoulos, A., Hagedorn, G., Jakupovic, J., Nimis, P. L., Valdes, B., Giintsch, A., Pankhurst, R. J., & White, R. J. (1999). A comprehensive reference model for biological collections and surveys. *Taxon*, 48, 511–562.

Blum, S., Barker, K., Baskauf, S. J., Berendsohn, W. G., Buttigieg, P. L., Döring, M., Droege, G., Fichtmueller, D., Glöckler, F., Güntsch, A., & Guralnick, R. (2019). Integrating ABCD and DarwinCore: Toward a better foundation for biodiversity information standards. *Biodiversity Information Science and Standards*, 3, e37491. https://doi.org/10.3897/biss.3.37491

Brickell, C. D., Alexander, C., David, J. C., Hetterscheid, W. L. A., Leslie, A. C., Malecot, V., Xiaobai, J., & Cubey, J. J. (2009). *International code of nomenclature for cultivated plants, Eighth Edition*. International Society for Horticultural Science. Leuven

Callomon, P. (2016) *The nature of names: Japanese vernacular nomenclature in natural science*. Drexel University.

Catapano, T. (2010). *TaxPub: An Extension of the NLM/NCBI Journal Publishing DTD for Taxonomic Descriptions*. Proceedings of the Journal Article Tag Suite Conference 2010. http://www.ncbi.nlm.nih.gov/books/NBK47081

Chen, M., Yu, S., Franz, N., & Bowers, S. (2014). Euler/X: a toolkit for logic-based taxonomy integration. *arXiv preprint*, 1402.1992.

Chesser, R. T., Burns, K. J., Cicero, C., Dunn, J. L., Kratter, A. W., Lovette, I. J., Rasmussen, P. C., Remsen, J. V., Rising, J. D., Stotz, D. F., & Winker, K. (2017). Fifty-eighth supplement to the American Ornithological Society's check-list of North American birds. *The Auk*, 134(3), 751–773. https://doi.org/10.1642/AUK-17-72.1

Conix, S. (2018). Values, regulation, and species delimitation. *Zootaxa*, 4415(2), 390–392. https://doi.org/10.11646/zootaxa.4415.2.9

Conix et al. In Prep. Towards a global list of accepted species III: Independence and stakeholder inclusion.

Conniff, R. (2011). *The Species Seekers: Heroes, Fools, and the Mad Pursuit of Life on Earth*. W. W. Norton & Company.

Conniff, R. (2021, January 22). *The Wall of the Dead: A Memorial to Fallen Naturalists*. Retrieved February 7, 2021, from https://strangebehaviors.wordpress.com/2011/01/14/the-wall-of-the-dead/

Corliss, J. O. (1995). The ambiregnal protists and the codes of nomenclature: a brief review of the problem and of proposed solutions. *Bulletin of Zoological Nomenclature*. 52, 11–17.

Correia, R. A., Jarić, I., Jepson, P., Malhado, A. C., Alves, J. A., Ladle, R. J. (2018), Nomenclature instability in species culturomic assessments: why synonyms matter. *Ecological Indicators*, 90, 74–78. https://doi.org/10.1016/j.ecolind.2018.02.059

Costello, M. J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D. P., Hoeksema, B. W., Poore, G. C. B., van Soest, R. W. M., Stöhr, S., Walter, T. C., Vanhoorne, B., Decock, W., &

Appeltans, W. (2013a). Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS ONE* 8(1), e51629.

Costello, M. J., May, R. M., & Stork, N. E. (2013b). Can we name Earth's species before they go extinct? *Science*, 339, 413–416.

Costello, M. J., & Wieczorek, J. (2014). Best practice for biodiversity data management and publication. *Biological Conservation*, 173, 68–73.

Costello, M. J., Appeltans, W., Bailly, N., Berendsohn, W. G., de Jong, Y., Edwards, M., Froese, R., Huettmann, F., Los, W., Mees, J., Segers, H., & Bisby, F. A. (2014). Strategies for the sustainability of online open-access biodiversity databases. *Biological Conservation*, 173, 155–165.

Costello, M. J., Horton, T., & Kroh, A. (2018). Sustainable biodiversity databasing: international, collaborative, dynamic, centralised. *Trends in Ecology and Evolution*, 33(11), 803–805.

Cotterill, F. P., Groves, C. P., & Taylor, P. J. (2017). Taxonomy: refine rather than stabilize. *Nature*, 547(7662), 162–162. https://doi.org/10.1038/547162d

Council of Heads of Australasian Herbaria (2010). Australian Plant Census https://biodiversity.org.au/nsl/services/apc

Döring, M., Ower, G. D. (2019). The Catalogue of Life Data Package - A new format for exchanging nomenclatural and taxonomic information. *Biodiversity Information Science and Standards*, 3, e38711. https://doi.org/10.3897/biss.3.38771

Fichtmueller, D., Berendsohn, W. G., Droege, G., Glöckler, F., Güntsch, A., Hoffmann, J., Holetschek, J., Petersen, M., & Reimeier, F. (2019). ABCD 3.0 Ready to use. *Biodiversity Information Science and Standards*, 3, e37214. https://doi.org/10.3897/biss.3.37214

Franz, N. M., Peet, R. K., & Weakley, A. S. (2008). On the use of taxonomic concepts in support of biodiversity research and taxonomy. In The New taxonomy (Wheeler QD, ed.). *Systematics Association Special Volume, Taylor & Francis*, 76, 63–84.

Franz, N. M., & Cardona-Duque, J. (2013). Description of two new species and phylogenetic reassessment of Perelleschus Wibmer & O'Brien, 1986 (Coleoptera: Curculionidae), with a complete taxonomic concept history of Perelleschus sec. Franz & Cardona-Duque, 2013. *Systematics and Biodiversity*, 11(2), 209–236. https://doi.org/10.1080/14772000.2013.806371

Franz, N. M., & Thau, D. (2010). Biological taxonomy and ontology development: Scope and limitations. *Biodiversity Informatics* 7(1), 45–66. https://journals.ku.edu/index.php/jbi/article/view/3927

Franz, N., Gilbert, E., Ludäscher, B., & Weakley, A. (2016). Controlling the taxonomic variable: Taxonomic concept resolution for a southeastern United States herbarium portal. *Research Ideas and Outcomes*, 2, e10610. http://dx.doi.org/10.3897/rio.2.e10610

Funk, V. A., Herendeen, P., & Knapp, S. (2017) Taxonomy: naming algae, fungi, plants. *Nature*, 546(7660), 599–599. https://doi.org/10.1038/546599c

Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. *Nature*, 546(7656), 25–27. https://doi.org/10.1038/546025a

Garnett, S. T., & Christidis, L. (2018a). Science-based taxonomy still needs better governance: Response to Thomson et al. *PLoS Biology*, 16(3), e2005249. https://doi.org/10.1371/journal.pbio.2005249

Garnett, S. T., & Christidis, L. (2018b). Better governance will lead to better taxonomy: response to Raposo et al.(2017). *Zootaxa*, 4418(3): 296–298. https://doi.org/10.11646/zootaxa.4418.3.7

Garnett, S. T., Christidis, L., Pylem R. L., & Thomson, S. A. (2020a). *How a scientific spat over how to name species turned into a big plus for nature. The Conversation*, https://theconversation.com/how-a-scientific-spat-over-how-to-name-species-turned-into-a-big-plus-for-nature-138887 26 July 2020.

Garnett, S. T., Christidis, L., Conix, S., Costello, M. J., Zachos, F. E., Bánki, O. S., Bao, Y., Barik, S. K., Buckeridge, J. S., Hobern, D., Lien, A., Montgomery, N., Nikolaeva, S., Pyle, R. L., Thomson, S. A., van Dijk, P. P., Whalen, A., Zhang, Z.-Q., & Thiele, K. R. (2020b). Principles for creating a single authoritative list of the world's species. *PLoS Biology*, 18(7), e3000736. https://doi.org/10.1371/journal.pbio.3000736

Geoffroy, M., & Berendsohn, W. G. (2003). The concept problem in taxonomy: importance, components, approaches. In: MoReTax: Handling factual information linked to taxonomic concepts in biology (W.G. Berendsohn WG, ed.). *Schriftenreihe für Vegetationskunde, Federal Agency for Nature Conservation, Bonn, Germany*. 39, 5–14.

Gewirtz, P. (1996). On 'I Know It When I See It'. *Yale Law Journal*, 105, 1023–1047.

Gradstein, S. R., Sauer, M., Braun, M., Koperski, M., Ludiwig, G. (2001) TaxLink, a program for computer-assisted documentation of different circumscriptions of biological taxa. *Taxon*, 50, 1075–1084.

Hayward, B.W., Le Coze, F., Vachard, D.,& Gross, O. (2021). *World Foraminifera Database*. Retrieved January 25, 2021, from http://www.marinespecies.org/foraminifera https://doi.org/10.14284/305

Hobern et al. In Prep. Towards a global list of accepted species VI: The Catalogue of Life Checklist.

Holetschek, J., Dröge, G., Güntsch, A., & Berendsohn, W. G. (2012). The ABCD of primary biodiversity data access. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, 146(4), 771–779. https://doi.org/10.1080/11263504.2012.740085

Hollingsworth, P. M. (2017). Taxonomy: avoid extra bureaucracy. *Nature*, 546(7660), 600–600 https://doi.org/10.1038/546600a

Holstein, N., Luebert, F. (2017). Taxonomy: stable taxon boundaries. *Nature*, 548(7666), 158–158. https://doi.org/10.1038/548158d

International Commission on Zoological Nomenclature (ICZN) (1999). *International Code of Zoological Nomenclature*. International Trust for Zoological Nomenclature.

Jackson, M.D., Scherz, M. D., & Zona, S. (2017). Taxonomy is not beholden to its dependencies: a rebuttal to Garnett and Christidis (2017) *PeerJ*, 5, e3060v1. https://doi.org/10.7287/peerj.preprints.3060v1

Kennedy, J., Hyam, R., Kukla, R., & Paterson, T. (2006). Standard data model representation for taxonomic information. *OMICS: A Journal of Integrative Biology* 10, 220–230. https://10.1089/omi.2006.10.220

King, A. M., Lefkowitz, E., Adams, M. J., & Carstens, E. B. (eds.) (2011) *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses (Vol. 9)*. Elsevier.

Koperski, M., Sauer, M., Braun, W., & Gradstein, S. R. (2000). Referenzliste der Moose Deutschlands. *Schriftenreihe Vegetationsk*. 34, 1–519.

Krell, F. T., & Pyle, R. L. (2010). ZooBank progress report. *Bulletin of Zoological Nomenclature*, 67(4), 271.

Kroh A., Costello MJ, Horton T (2019) Biodiversity databases in the future: reply to Cene Fišer. Trends in Ecology and Evolution 34(3): 185–186.

Lambertz, M. (2017). Taxonomy: retain scientific autonomy. *Nature*, 546(7660), 600–600. https://doi.org/10.1038/546600b

Lepage D., Vaidya, G., & Guralnick, R. (2014). Avibase–a database system for managing and organizing taxonomic concepts. *ZooKeys*, 420, 117. https://doi.org/10.3897/zookeys.420.7089

Le Renard, J. (2000). TAXIS, a taxonomic information system for managing large biological collections. in: *Abstracts. TDWG 2000: Digitizing Biological Collections. Taxonomic Databases Working Group* (p. 18), 16th Annual Meeting, Frankfurt, November 10-12, 2000.

Lien et al. In Prep. Towards a global list of accepted species IV: Overcoming fragmentation in the governance of taxonomic lists.

Linnaeus, C. (1753). *Species plantarum, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas.* Impensis Laurentii Salvii

Linnaeus, C. (1758). *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata.* Impensis Laurentii Salvii

McClure, C. J. W., Lepage, D., Dunn, L., Anderson, D. L., Schulwitz, S. E., Camacho, L., Robinson, B. W., Christidis, L., Schulenberg, T. S., Iliff, M. J., Rasmussen, P. C., Johnson, J. (2020). Towards reconciliation of the four world bird lists: hotspots of disagreement in taxonomy of raptors. *Proceedings of the Royal Society B: Biological Sciences*, 287(1929), 20200683. https://doi.org/10.1098/rspb.2020.0683

Midford, P. E., Dececchi, T. A., Balhoff, J. P., Dahdul, W. M., Ibrahim, N., Lapp, H., Lundberg, J. G., Mabee, P. M., Sereno, P. C., Westerfield, M., & Vision, T. J. (2013). The vertebrate taxonomy ontology: a framework for reasoning across model organism and species phenotypes. *Journal of Biomedical Semantics*, 4(1), 1–6. https://doi.org/10.1186/2041-1480-4-34

Murray, A. E., Freudenstein, J., Reysenbach, A.-L. (2020). Roadmap for naming uncultivated Archaea and Bacteria. *Nature Microbiology*, 5, 987–994.

Nakada, T. (2010). Nomenclatural notes on some ambiregnal generic names (comments to Özdikmen, 2009). *Munis Entomology & Zoology*, 5(1), 204–208.

Parker, C. T., Tindall, B. J., & Garrity, G. M. (eds.) (2019). International code of nomenclature of prokaryotes: prokaryotic code (2008 revision). *International Journal of Systematic and Evolutionary Microbiology*, 69(1A), S1–S111. https://doi.org/10.1099/ijsem.0.000778

Page, L. M., Espinosa-Pérez, H., Findley, L. T., Gilbert, C. R., Lea, R. N., Mandrak, N. E., & Mayden, R. L. (2013). New Seventh Edition of Common and Scientific Names of Fishes: Changes include capitalization of common names. *Fisheries*. 38(4), 188–189.

Patterson, C. (1981). Significance of Fossils in Determining Evolutionary Relationships. *Annual Review of Ecology and Systematics*, 12, 195–223.

Patterson, D. J. (1986). Some problems of ambiregnal taxonomy and a possible solution. *Symposia Biologica Hungarica*, 33, 87–93.

Patterson, D. J., & Larsen, J. (1992). A perspective on protistan nomenclature. *The Journal of Protozoology*, 39(1), 125–131.

Patterson, D. J., Cooper, J., Kirk, P. M., Pyle, R. L., & Remsen, D. P. (2010), Names are key to the big new biology. *Trends in Ecology & Evolution*, 25(12), 686–691. https://doi.org/10.1016/j.tree.2010.09.004

Pullan, M. R., Watson, M. F., Kennedy, J. B., Raguenaud, C., & Hyam, R. (2000). The Prometheus Taxonomic Model: a practical approach to representing multiple classifications. *Taxon*, 49, 55–75.

Pyle, R. L. (2004). Taxonomer: a relational data model for managing information relevant to taxonomic research. *Phyloinformatics*, 1, 1–54. https://doi.org/10.5281/zenodo.59790

Pyle, R. L. (2016). Towards a Global Names Architecture: The future of indexing scientific names. *ZooKeys*, 550, 261. https://dx.doi.org/10.3897/zookeys.550.10009

Pyle, R. L., & Michel, E. (2008). ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. *Zootaxa*, 1950(1), 39–50. https://doi.org/10.11646/zootaxa.1950.1.6

Raposo, M. A., Stopiglia, R., Brito, G. R. R., Bockmann, F. A., Kirwan, G. M., Gayon, J., & Dubois, A. (2017). What really hampers taxonomy and conservation? A riposte to Garnett and Christidis (2017). *Zootaxa*, 4317(1), 179–184. https://doi.org/10.11646/zootaxa.4317.1.10

Raguenaud, C. (2002), *Managing complex taxonomic data in an object-oriented database*. [Unpublished manuscript representing partial PhD thesis, Napier University.] http://www.soc.napier.ac.uk/publication/op/getpublication/publicationid/1845313

Ruggiero, M. A., Gordon, D. P., Orrell, T. M., Bailly, N., Bourgoin, T., Brusca, R. C., Cavalier-Smith, T., Guiry, M. D., & Kirk, P. M. (2015). A Higher Level Classification of All Living Organisms. *PLoS ONE*, 10(4), e0119248. https://doi.org/10.1371/journal.pone.0119248

Senderov, V., Simov, K., Franz N., Stoev, P., Catapano, T., Agosti, D., Sautter, G., Morris, R. A., & Penev, L. (2018). OpenBiodiv-O: ontology of the OpenBiodiv knowledge management system. *Journal of Biomedical Semantics*, 9(5), 1–15. https://doi.org/10.1186/s13326-017-0174-5

Spencer, R. D., & Cross, R. G. (2007). The international code of botanical nomenclature (ICBN), the international code of nomenclature for cultivated plants (ICNCP), and the cultigen. *Taxon*, 56(3), 938–940.

Stearn, W. T. (ed.) (1953). *International code of nomenclature for cultivated plants*. London: Royal Horticultural Society.

Thiele et al. In Prep. Towards a global list of accepted species I: Taxonomy as map-making: the foundation for lists of accepted species.

Thomson, S. A., Pyle, R. L., Ahyong, S. T., Alonso-Zarazaga, M., Ammirati, J., Araya J. F., et al. (2018). Taxonomy based on science is necessary for global conservation. *PLoS Biology*, 16, e2005075. https://doi.org/10.1371/journal.pbio.2005075

Thomson et al. In Prep. Towards a global list of accepted species II: Consequences of inadequate taxonomic list governance

Trehane, P. (2002). 50 years of the International Code of Nomenclature for Cultivated Plants: Future prospects for the Code. In *XXVI International Horticultural Congress: IV International Symposium on Taxonomy of Cultivated Plants* (pp. 17–27).

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

Wheeler, Q. D., & Meier, R. (eds.) (2000). *Species concepts and phylogenetic theory: a debate*. Columbia University Press.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: an evolving community-developed biodiversity data standard. *PloS ONE*, 7(1), e29715. https://doi.org/10.1371/journal.pone.0029715

Ytow, N., Morse, D. R., Roberts, & D. M. (2001). Nomencurator: a nomenclatural history model to handle multiple taxonomic views. *Biological Journal of the Linnean Society*, 73(1), 81–98. https://doi.org/10.1111/j.1095-8312.2001.tb01348.x

Zachos, F. E. (2018). (New) Species concepts, species delimitation and the inherent limitations of taxonomy. *Journal of Genetics*, 97(4), 811–815. https://doi.org/10.1007/s12041-018-0965-1

Zhong, Y., Jung, S., Pramanik, S., & Beaman, J. H. (1996). Data model and comparison and query methods for interacting classifications in a taxonomic database. *Taxon*, 45, 223-241.