10 Provisional Nomenclature
The On-Ramp to Taxonomic Names

David E. Schindel and Scott E. Miller

CONTENTS

THE OTHER TAXONOMIC IMPEDIMENT

Many authors have discussed the "taxonomic impediment"—the lack of trained taxonomists and the technical infrastructure needed to halt and reverse the loss of biodiversity (Lyal and Weitzman, 2003; Rodman and Cody, 2003). The Convention on Biological Diversity (CBD) entered into force at the end of 1993, and taxonomists were quick to point out the critical role that taxonomy would play in reaching the CBD's objectives. In 1995, the CBD Second Conference of the Parties (COP 2) called for a study of the lack of taxonomists needed to achieve the CBD's objectives*. In 1998, COP 4 endorsed the need to increase research capacity in taxonomy and a recommendation† to create the Global Taxonomy Initiative (http://www.cbd.int/gti/).

Evenhuis (2007) reviewed the use of the term taxonomic impediment as the limitations based on lack of external resources (e.g., funding, people, facilities). He went on to describe eight steps in the taxonomic process (presented in compressed form below). Evenhuis asserted that most taxonomists were enthusiastic about conducting the first four steps, but the last four were obstacles ("the other taxonomic impediment") that severely limited the progress of taxonomy. The steps he described were:

1. Venturing into nature
2. Collecting specimens
3. Sorting specimens into species
4. Discovering new or rare species among the sorted specimens
5. Confirming new discoveries through comparison with publications and types

* Decision II/8.7 of COP 2
† Decision IV/1 of COP 4 endorsed the Recommendation II/2 of the Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA)

6. Describing and illustrating new species
7. Submitting manuscripts for publication
8. Educating others about the new species

In this chapter, we will focus on steps 5, 6, and 7. We will argue that creating a system of provisional nomenclature can fundamentally alter and accelerate the taxonomic process.

STEPS AND MISSTEPS IN THE TAXONOMIC PROCESS

"Nomenclatural events"—including the publication of formal, Code-compliant Linnaean taxonomic names—are the critical datapoints in taxonomy. Each nomenclatural event places one or more formal names into the literature and the corpus of claims for nomenclatural priority. Each proposed name becomes subject to scrutiny by the research community that will test its formal validity. For example, was the name proposed in an allowed form, and was it accompanied by the required documentation (e.g., designation of a holotype, if required by a nomenclatural code, or deposition of the required type in a repository)? If a name was formally correct, does it represent a new taxon or is it based on a holotype that bears a previously proposed name, and may therefore be an objective synonym? Beyond these objective concerns lie the longer-term subjective considerations. Does the proposed name represent a taxon concept that has already been proposed, and may therefore be a subjective synonym? Does the name represent a taxon concept that is too broad and may contain distinct taxa that deserve to be formally recognized under different names?

In the traditional practice of taxonomy, researchers are expected to anticipate these and other challenges to the validity of the names they propose. Prior to publication, taxonomists prepare careful descriptions of new taxa and can devote years, sometimes decades, to examining the publications and type specimens associated with similar taxa to ensure that theirs are truly new and distinct. In principle, taxonomists are being asked to prove a negative assertion when they propose a new taxon, e.g., "This species has never been described before." In practice, taxonomists make the practical decision to formally propose and publish a new name when it seems very unlikely (but not certain) that they have overlooked a previous nomenclatural event that will invalidate their claim for priority.

This process bears many similarities to the system of patent protection for inventions. An inventor creates an innovation that he or she thinks is novel and significant. To avoid the cost of filing a patent claim that may later be judged invalid, an inventor normally conducts a patent search to determine whether the idea has already been proposed. The longer inventors wait to file a patent claim, the greater is the likelihood that someone will either steal their idea or devise it independently. Once a patent application is made, it enters the public domain where anyone can examine it and decide whether the new patent infringes on a previously filed patent claim. Lawsuits are the recourse open to an inventor who thinks a new patent is, in essence, a junior synonym of his or her earlier invention, just as taxonomic revisions are the recourse used by taxonomists to lump, split, or place names in synonymy.

The parallel between taxonomy and patenting ends here. Patenting can lead to licensing, product development, commercialization, and financial gain for the inventor. None of these consequences or incentives is associated with naming a new species. Except in rare circumstances, taxonomists do not profit from the naming of new taxa. Nevertheless, taxonomists often conduct their research on potential new species as solitary individuals behind closed doors, restricting pre-publication access to their findings as if they were trade secrets (see Figure 10.1). Conducted in this way, it can take a taxonomist years to decades to generate tangible products in the form of the nomenclatural events.

For an inventor, the goal of financial gain can be attained only if the asset (the innovation) is kept in private hands until it is ready for the marketplace. For a taxonomist, the goal is the professional recognition and personal satisfaction of discovering, describing, and recording new forms of biodiversity. For historical reasons, the taxonomic process has defined nomenclatural events as the goal-

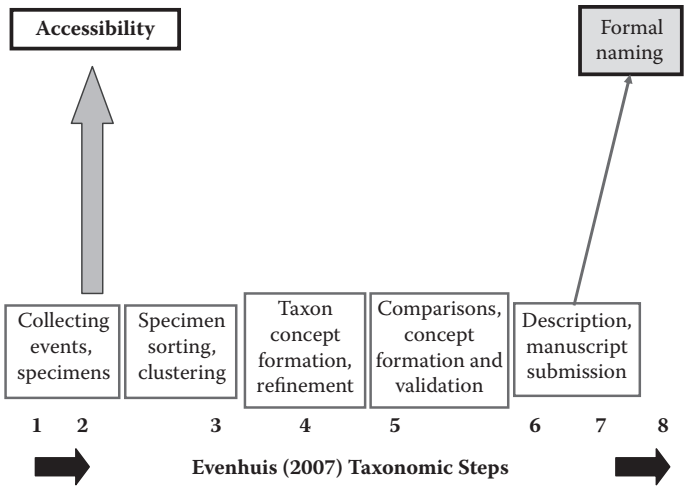

FIGURE 10.1 The results of each step in the taxonomic process, as traditionally practiced, are not shared until publication.

line over which taxonomists must cross if they are to receive formal recognition. Linnean names have become the only currency with which taxonomists can be rewarded. It is this system of incentives that is causing taxonomists to carry out their research as an individualistic, secretive pursuit.

In the following sections, we propose the introduction of standardized taxon labels as intermediate products in the process of producing new taxonomic names. As described, creating a standardized system of provisional nomenclature can create a new incentive and reward system for taxonomists and can greatly accelerate the documentation of biodiversity.

This is not a new problem, and others have suggested approaches in the past (e.g., the "interim taxonomy" of Erwin 1991, 1995), but a standardized system has never been widely adapted. Many taxonomists and ecologists have used various systems to refer in publication to species that lacked formal names. For example, Marks (1983: 534) explicitly formatted such references as "*Aedes* (*Finlaya*) 'Marks sp. no. 104' an undescribed species which has been studied in some detail by the author" versus named species and species for which identities were not resolved; and Holloway (1984) has made a practice of referring to undescribed lepidopteran species by the genitalia slide number of the critical voucher.

TAXON NAMES, CONCEPTS, AND LABELS

Taxon names provide the framework for information exchange and retrieval in taxonomy. The rules governing the formal naming of taxonomic units are overseen by several international commissions of nomenclature. Unlike patenting systems, taxonomy does not have centralized registries of patent applications or a regulated system for publishing patent applications. Polaszek et al. (2005) proposed the creation of ZooBank as a registry of taxonomic names for the International Commission of Zoological Nomenclature. Such a centralized registry would facilitate access to formal taxon names and their associated documentation.

Taxon concepts are biological hypotheses that are represented by taxon names. Concepts emerge in a taxonomist's thinking after examining specimens, sorting and separating them into groups of similar individuals, and seeking discontinuities in the variation among these groups. In most cases, these groups will correspond to previously described taxa, but some can emerge as new discoveries. The holotype designated for a new taxon is selected as a singular representation of the newly described taxon concept, and paratypes are often designated to represent the variation included by

the author under that name. Descriptions that form a critical part of a nomenclatural act are meant to transmit the author's concept of the taxon and its limits in various biological dimensions (e.g., morphology, development, geographical distribution, ecological preferences, DNA sequence variation).

The term *taxon label* has no generally accepted meaning, but we propose one here. Taxon labels with these characteristics can be used to communicate taxon concepts that are not yet ready for formal description and publication. In principle, they can be posted on public Websites, submitted to taxonomic databases, and even published without compromising the priority or clarity of taxonomic names.

> A taxon label is a unique, stable, text-phrase applied to an unpublished taxon concept. The text-phrase of a taxon label should link the concept to reference specimens and/or a geographic locality, and should be in a form that makes it clearly unusable as a formal taxon name.

This definition has the following important components:

Uniqueness. A taxon label must clearly separate a taxon concept from all other concepts. Labels such as *Genus sp. A* are inadequate for distinguishing concepts.

Stability. Once a taxon label has been applied and made public through Websites, databases, presentations, or publications, it should remain stable in its linkage to associated specimens and localities. Any change in the underlying concept should trigger a change in the label.

Unsuitability as taxon names. As illustrated below, the text-phrase of a taxon label can include multiple words, numerals, punctuation marks, and mixtures of upper- and lowercase letters. The use of these character types will make it impossible to misinterpret a taxon label as a taxon name. In this way, publication of a taxon label will not compromise the later publication of a taxon name based on the same concept.

Examples of Taxon Labels

Taxonomists have routinely used provisional names, but their practices have not been standardized or consistent. For example, records in GenBank carry taxon identifiers such as *Ocyptamus* sp. MZH S143_2004, *Argentinomyia* sp. CR-12; *Eunotia* sp. EUN392T, and *Xenopus* (*Silurana*) sp. new tetraploid. These are clearly not meant to be formal Linnean names, but neither are they interpretable to the user. Other than contacting the researcher who submitted the record to GenBank, there is no information available about the taxon concepts associated with these provisional names.

Janzen et al. (2009) proposed a system for interim nomenclature that would include provisional taxon labels in a variety of formats such as "Patelloa xanthuraDHJ02," "Belvosia Woodley06" and "Astraptes LOHAMP." Their proposed system does not fully meet the criteria we propose above. Specifically, appending numerals to a person's name does not provide clear uniqueness and stability, and capitalization is not a secure way to distinguish taxon labels from taxon names. Is "Belvosia Woodley06" a provisional species in the genus Belvosia or a provisional label for a new genus? Could "Astraptes LOHAMP" be transcribed as "Astraptes lohamp," which can be mistaken for a formal name?

In contrast, a few communities of practice in taxonomy have developed standardized systems of provisional nomenclature. Barker (2005) described a decision by the Council of the Heads of Australian Herbariums (CHAH) to regularize the formation of "informal names" (termed "taxon labels" here, to distinguish them from taxon names). The proposed standard CHAH format for a taxon label was:

- *Genus-name sp. Phrasename (Voucher-specimen identifier)* Source
- where:
- Genus-name is a previously published generic name.
- "sp." is a standard delimiter that indicates species rank.

- Phrasename refers to a locality.
- (Voucher-specimen identifier) is a two-part field consisting of a collector's name and the voucher specimen number attached to the exemplar of the taxon concept, or its herbarium sheet number.
- Source refers to the name of the concept's proposer.

The CHAH constructed this standard to address two problems. First, state-level floras or censuses had been conducted independently and in some cases different taxon names had been published for a single species. "Informal names" provided a system of communication among regions that can aid the process of constructing a consensus taxonomic list for the country as a whole. Second, non-taxonomists responsible for the conservation of rare and endangered plant species needed a way to cite taxa in regulations. The CHAH standard provided an objective reference system without being mistaken for taxonomic judgments.

The CHAH standard offers a starting place for discussion of the preferred format for taxon labels. We suggest that TDWG (Biodiversity Information Standards, formerly known as the Taxonomic Databases Working Group) is an appropriate organization for developing a global standard for taxon labels.

Taxon Labels in the Taxonomic Process

Taxonomic research requires the skill, experience, and judgment of a craftsman, and the steps in the taxonomic process (especially Evenhuis's steps 5, 6, and 7) require precision and scholarship. However, nothing in the process requires it to be carried out as the proprietary work of an individual. Quite the contrary, enlisting collaborators would make the work of examining the literature and type specimen collections faster and easier. The widespread availability of information technology and the Internet are reducing distance as an obstacle to taxonomic research (see Godfray et al., 2007; Zhang, 2008).

The proposed system of taxon labels would enable researchers to make their interim results accessible on public databases and Websites. Articles could be published with references to taxon labels linked to the databases and Websites that provide the information associated with the taxon labels. Ecologists and other non-taxonomists could publish results using taxon labels, thereby avoiding the delay often associated with waiting for taxonomists to put formal names on specimens.

Putting interim results into the public domain would have two significant impacts. First, taxonomists could cite these interim products as publications and other deliverables. Second, authors could engage other taxonomists as collaborators in making critical comparisons with type specimens. Initiatives such as the European Distributed Institute of Taxonomy (EDIT; see http://www.e-taxonomy.eu/) are developing tools for "cybertaxonomy" that will enable Web-based collaboration and consensus-based taxonomy research conducted by teams of researchers, not solitary individuals. Figure 10.2 illustrates how taxon labels can be used for information exchange, collaboration, and publication of interim results.

Taxon Labels and DNA Barcoding

Hebert et al. (2003) proposed the use of a short, standardized DNA sequence as a diagnostic marker for species identification—a "DNA barcode" analogous to the Universal Product Code used to link products in stores to inventory records. Since that proposal, the Barcode of Life Initiative has mushroomed and has gathered DNA barcode records from almost 700,000 specimens representing more than 60,000 species. There are at least three circumstances in which taxon labels could be used:

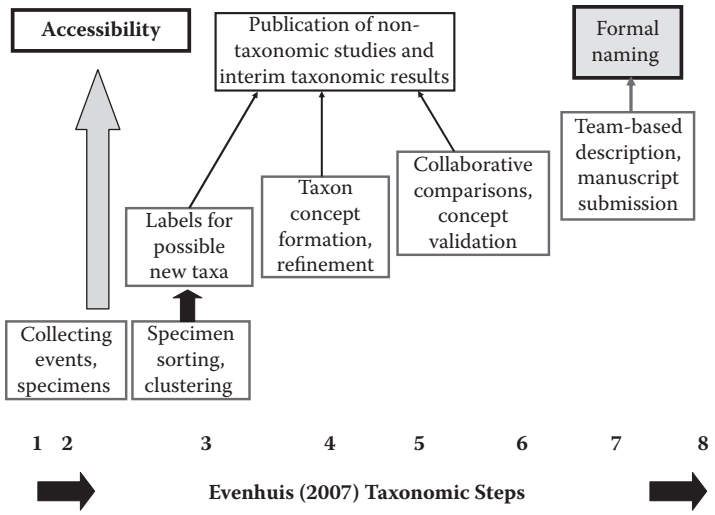

FIGURE 10.2 How taxon labels can be used for information exchange, collaboration, and publication of interim results.

1. Barcodes have uncovered hidden variation within species that may merit recognition as new species. These intraspecific barcode clusters need to be studied by taxonomists with experience in that group, and attaching a taxon label to each cluster would facilitate this research.
2. Field collections are also yielding specimens that do not appear to belong to any known species. Taxon labels attached to these specimens would facilitate communication between specialists as barcode data and digital images of specimens are exchanged for study.
3. Metagenomic analyses of environmental samples are producing enormous volumes of data on short gene fragments. The barcode gene sequences of the organisms in these environmental mixtures can be assembled to produce populations of barcodes. Clusters of these barcode records reflect discontinuities in genetic variation, but in the absence of discrete voucher specimens there is no way to associate taxonomic names to these clusters. Taxon labels provide placeholders for these barcode clusters until they can be associated with barcodes from identified specimens.

CONCLUSIONS

Formal taxonomic names that comply with nomenclatural codes are the mainstays of taxonomy and biodiversity research, but they are not the only medium for information exchange. Taxon labels can be very useful in conveying the results of non-taxonomic research and for accelerating the progress of taxonomic research. To achieve these ends, taxon labels will need to be unique, stable, and formatted to convey critical information without interfering with formal nomenclature.

ACKNOWLEDGMENTS

This work was supported by Alfred P. Sloan Foundation grant 2008-3-4 to the Smithsonian Institution for support of the Consortium for the Barcode of Life. The authors benefited from discussions with Dan Janzen, Arthur Chapman, James Croft and Terry Erwin.

REFERENCES

Barker, B. (2005). Standardising informal names in Australian publications, *Australian Systematic Botany Society Newsletter* 122:10–11.

Erwin, T.L. (1991). Establishing a Tropical Species Co-occurrence Database. Part 1. A plan for developing consistent biotic inventories in temperate and tropical habitats. Memoria del Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Peru 20:1–16.

Erwin, T.L. (1995). Measuring arthropod biodiversity in the tropical forest canopy. In M.D. Lowman, and N.M. Nadkarni, Eds. Forest Canopies. Academic Press. 109–127.

Evenhuis, N.L. (2007). Helping solve the "other" taxonomic impediment: Completing the eight steps to total enlightenment and taxonomic nirvana, *Zootaxa* 1407: 3–12.

Godfray, H.C.J., Clark, B.R., Kitching, I.J., Mayo, S.J., and Scoble, M.J. (2007). The Web and the structure of taxonomy, *Systematic Biology*, 56:6, 943—955

Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society London B* 270:313–321.

Holloway, J.D. 1984. The Moths of Borneo: Family *Notodontidae. Malayan Nature Journal* 37:1–107.

Janzen, D.H., W. Hallwachs, P. Blandin, J.M. Burns, J.-M. Cadiou, I. Chacon, et al. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources* 9 (supplement 1):1–26.

Lyal, C.H.C. and Weitzmann, A.L. (2004). Taxonomy: Exploring the impediment. *Science,* 305, 1106.

Marks, E.N. (1983). Mosquitoes of the Purari River lowlands. In T. Petr, Ed. The Purari: tropical environment of a high rainfall river basin. Dordrecht, Netherlands: Dr. W. Junk Publishers. 531–550.

Polaszek, A., Agosti, D., Alonso-Zarazaga, M., Beccaloni, G., de Place Bjørn, P., Bouchet, P. et al. (2005). A universal register for animal names. *Nature* 437: 477.

Rodman, J.E. and Cody, J.H. (2003). The taxonomic impediment overcome: NSF's Partnerships for Enhancing Expertise in Taxonomy (PEET) as a model. *Systematic Biology,* 52(3), 428–435.

Zhang, Z.Q. (2008). Zoological taxonomy at 250: Showcasing species descriptions in the cyber era, *Zootaxa* 1671: 1–2.