

TAPROBANICA, ISSN 1800–427X. November, 2022. Vol. 11, No. 02: pp. 47–53.

© Research Center for Climate Change and Department of Biology, Faculty of Mathematics & Natural Sciences, University of Indonesia, Depok 16424, INDONESIA.

<http://www.taprobanica.org>

<https://doi.org/10.47605/tapro.v11i2.282>



The how and why of scientific naming: What's in a name?

Quite a lot, actually. Names organize and give meaning to our world. They give us identity and allow us to communicate clearly and efficiently. In the context of scientific names for living and fossil organisms, they link data from around the globe into a synthetic framework from which we can make powerful predictions and effective strategies: everything from governmental policies and next-generation medicines to the origins of diversity and its responses to climate change. These, and many other reasons, are the *why*, but *how* is equally important. Unlike folk taxonomies, which are rich in cultural tradition and applicable only at local scales, scientific nomenclature must abide specific rules to remain effective globally, unbounded by the vagaries of local customs. These names must also transcend time so that data from past and future generations can be meaningfully synergized for their greatest benefit. Not surprisingly, to achieve such lofty goals, an entire form of information science is needed, one complete with its own terminology and pragmatic rules. The *how* of scientific naming is as vital as the *why*, for to fail at the first means catastrophe for the fundamental reasons of the latter. Imagine bad practices for the scientific naming of drugs or diseases and the tsunami of disastrous impacts that would follow. Now imagine the same if we failed at communicating about disease-transmitting mosquitos or poisonous mushrooms. Scientific names have life-sustaining importance and we cannot be cavalier about taxonomic science and formalized nomenclature.

What is scientific nomenclature? A primer to “biodiversity semiotics”

Given that mathematics is the foundation of natural sciences, it follows that the axioms of the former apply to the latter (Fraenkel *et al.* 1973). Accordingly, the necessity of objects such as sets and natural kinds is fundamental to our knowledge of the world, be these infinite or empty sets in mathematics or a set of biological entities we classify as a species (*natural kinds*) (Maddy 1990, 1997). The realism of these sets and our need to discuss them is paramount (Maddy 1990, 1997, Kremer 1991, Linsky & Zalta 1995), and may be considered as the first-order of biological realism: *e.g.*, “there exists x such that x is *melanogaster* and x is a species.” The question about the structural properties and recognition of species is a philosophical challenge faced by virtually all biologists, seemingly with no agreed upon solution other than those that adopt a contextualized concept of species pluralism. The purpose here is not to further pace the halls of such debate as a discussion of nomenclature is not predicated upon one given theory of species over another.

To avoid misconceptions, formalized nomenclature (*i.e.*, scientific nomenclature) attempts to deploy pragmatic rules such that for each unique biological set (species, genus, family, and so forth) there is a unique name. These names serve as symbols for our communication about the objects and concepts to which they refer. Our biological theories and methodologies permit us to circumscribe these entities or sets (*i.e.*, the work of taxonomic science), and our systems of nomenclature provide a set of consistent rules by which we can assign names to these referents forming the language through which we discuss biology (*e.g.*, ICZN 1999).

One can view scientific nomenclature from many angles: as a data retrieval system aligned with information science, a suite of legalistic rules, and even from the perspective of semiotics. In fact, all of these are accurate and reflect different facets of nomenclature. Many have discussed nomenclature from the data retrieval and legal frameworks, so I shall refrain from strolling well-trodden paths. For current purposes, I choose to focus on a form of semiotics perspective and the importance of this for effective communication, which is one purpose of any system of language. This is what may be referred to as “biodiversity semiotics”, and is the philosophical underpinnings of our scientific nomenclature.

With all of the above in mind, let us return to a beginning. What is *Arabidopsis thaliana*? What is *Drosophila*? Seemingly simple questions, eh? And yet, there are considerable epistemic challenges embedded in such questions. What if in your mind *A. thaliana* is a flowering plant, and in mine it is a shellfish? How is it that we share a concept for *A. thaliana*? Is it important that we share a concept? The answer to the latter is yes. Scientific names are the very language of biological sciences, including all formal or informal disciplines that relate to any living or fossil organism—everything from medicine and ecology to governmental policies and international treaties. Like any language, its function is effective and precise communication, and consists of symbols (*i.e.*, names) that unite a *form* with a *meaning*, and these are then employed syntactically to build the language (Fig. 1). In a simplified semiotic context for scientific names, one can recognize examples of forms as *Drosophila*, *elegans*, *Arabidopsis*, *sapiens*, and so forth, while the meaning is the concept and object to which they are linked and meant to symbolize. Going beyond the name, a simple syntax to this language is, for example, that the generic name always proceeds the specific epithet: *Drosophila melanogaster*, rather than *Melanogaster drosophila*. The precision of this language comes from the degree to which the form and meaning are recognized by others in our global community. If *A. thaliana* means different things to different communities, then the precision is lost and errors ensue. Thus, unlike folk taxonomies, which are rich in cultural heritage and mythology and valuable in their own context, formal scientific nomenclature serves a different purpose and one that must maintain universal recognition.

How then do we maintain such precision for meaningful scientific communication? This is where the science of taxonomy and rules of nomenclature come into play. These operate, unfortunately often ignored and misunderstood by most scientific disciplines they gird, to maintain these meanings constantly and consistently, while remaining sufficiently flexible to preserve continuity as our biological theories and methods evolve. The fact that a name like *Drosophila melanogaster* has such meaning to science over the last 191 years, despite the intellectual revolutions of evolution, population genetics, and molecular developmental biology, attests to the success of our systems of nomenclature. These systems have evolved alongside the sciences they underpin, all so that we might enjoy a language from which to discuss biology. But at its core, the critical element is maintaining that form-meaning linkage, and to do this our hypotheses of species and their meaning must be exact, precise, and rooted in the best available data, ideally from as many sources as possible, and guided by best practices in taxonomic and nomenclatural sciences (Engel *et al.* 2021). Similarly, the way in which we establish the names for those meanings is just as precise, as is the syntax through which we combine these form-meaning symbols. Most practitioners of biology simply ignore nomenclature and take it for granted that when they speak of *Caenorhabditis elegans*, *Zea mays*, *Dictyostelium discoideum*, or *Saccharomyces cerevisiae*, that others will know what they mean. And yet this can only be taken for granted because of the work of taxonomists employing the best practices of nomenclatural science. This process is constantly in play not only for species newly discovered, but even for our seemingly most well-known organisms as our hypotheses and theories of what constitutes a species are debated and refined. Anyone who uses scientific names should be keenly concerned that nomenclatural science is robust, supported, and always executed with the greatest data and precision.

Accordingly, rules of nomenclature exist for the precise naming of molecules, chemical compounds, diseases, viruses, measures, species, and anything for which we need a formal, universal, scientific language. The rules, internationally agreed upon documents of legal form (note that many laws that relate to organisms actually cite the Codes of Nomenclature for establishing what is meant by a given species: *e.g.*, U.S. Code of Federal Regulations 21 CFR §101.4, among thousands of other national and international regulations), dictate what are the minimal criteria necessary to establish a formalized name for any given biological entity, be it a species, a genus, or a family, and how such concepts and data must be vouchered for arbitration during moments of ambiguity (*e.g.*, ICZN 1999). At their core they are tasked with establishing a precise and formal link between the *form* and *meaning* of our language symbols. If this link is weakened or broken, then our symbols are senseless. It is scientific nomenclature and taxonomic science that make it possible that my concept of *D. melanogaster* is the same as everyone else's. Thus, as in all disciplines, we require as much data as possible to ground our concepts, convey them to others, and imbue our names with exact meaning. Taxonomists therefore are continually testing hypotheses of species circumscription by data derived

from diverse sources, synthesizing genomics with variation in phenotypes, behaviors with biogeographic and phylogenetic history, and life histories with developmental stages. The application of names can have life or death consequences (e.g., bacteriological nomenclature even recognizes a class of names as *nomina periculosa*, or names that by their nature endanger health and security either through ambiguity or other difficulties), with ambiguity potentially being lethal. Data, as in all sciences, are key.

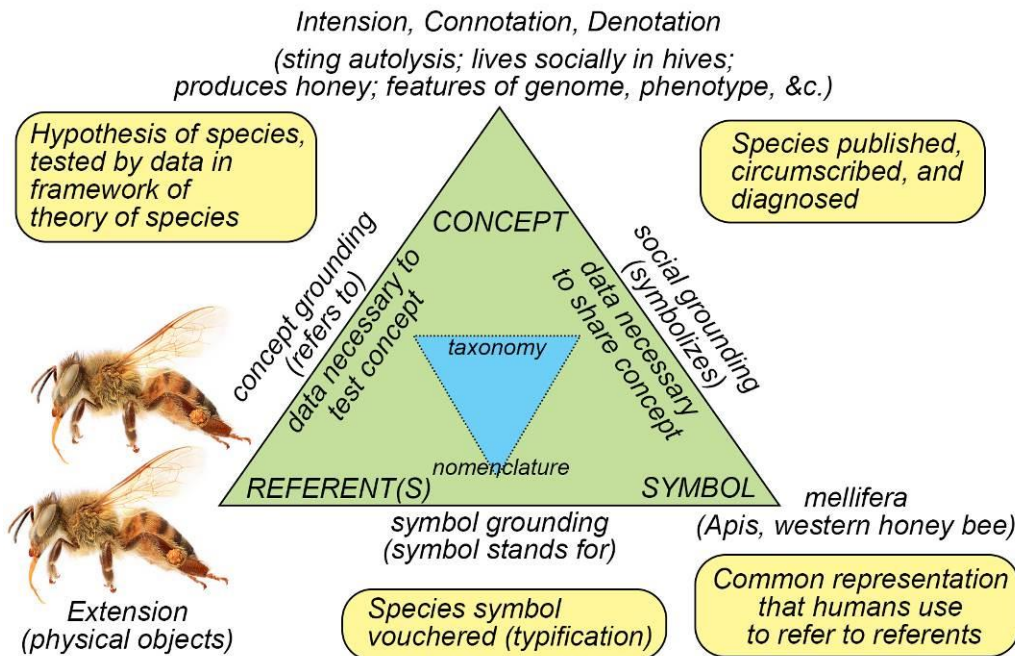


Figure 1. Biodiversity semiotics: an idealized semiotic perspective on scientific nomenclature, illustrated by the western honey bee (*Apis mellifera*). Data minimalism jeopardizes the concept grounding and social grounding of our concepts and referential symbols, leaving what is meant by our scientific names (symbols) ambiguous. Semiotic concepts presented around the triangle, biological activities relating to taxonomy and nomenclature enclosed in yellow bubbles.

What is data minimalism?

Our world is rich in life, and has been for nearly four billion years. Thus, there is much to discuss and share, with scientific names at the core. Indeed, although we have formally named in a scientific context about 2 million species, this is a fraction of the total diversity that exists or has ever existed. Accordingly, some scientists are eager to ‘finish the job’ of naming, believing that it is important to complete an inventory of all life as fast as possible, whatever it takes (e.g., Sharkey *et al.* 2021). Rather than carefully gather data and follow best practices, such proposals take a ‘data minimalist’ approach, leaving any potential collateral damage to future generations to clean up, all with the idea of having that ultimate list of names (Engel *et al.* 2021, Meier *et al.* 2022). The contention of such proponents is that taxonomy is failing to meet the challenge of biodiversity discovery in our current age of extinctions. Most species remain to be discovered and understood and researchers of this persuasion argue that these species shall be extinct before they are even named by science. This is certainly true. The argument goes further, stating that taxonomy is too slow and fails to capitalize on so-called innovative tools to speed the enumeration of species before biological diversity is consigned to oblivion. Accordingly, rather than ‘wasting’ time executing data-intensive taxonomic experiments, the data minimalists speed the process by gathering only the tiniest fraction of data to test a conception of species and then apply a name, claiming that this innovates discovery, and simultaneously eases species recognition and communication by the community. Their process in fact results not only in speedily conceiving of species identities, but also a rapid and minimal application of nomenclatural rules to establish the referent-symbol linkage and the social grounding of the concept-symbol ties (Engel *et al.* 2021).

Before addressing their solution to this challenge, it is worth addressing this criticism regarding the practice of taxonomy in biodiversity science. While on the surface this all seems legitimate, upon closer inspection one can see that these arguments are either outright red herrings meant to mislead or reveal a fundamental ignorance. In reality, taxonomy is one of the most synthetic disciplines, utilizing and uniting data from genomics and comparative morphology to physiology and ecology (Engel *et al.* 2021). Taxonomists regularly employ innovative tools ranging from the latest in nano-scale CT scans to transcriptomes and molecular developmental biology. The innovation is already there. Are there some taxonomists who rely heavily on morphology? Sure, and in the same manner that some researchers principally use histology for biomedical study. It remains a valid and important tool, and for some species it may be the only tool available. So, what about the lamented pace of the work? It is perhaps true that the speed of biodiversity discovery is hampered today. But this is not intrinsic to the science, and instead is the result of a steady decline in the number of trained taxonomists (Boero 2010, Britz *et al.* 2020, Engel *et al.* 2021). Nonetheless, while the number of taxonomists has declined, the rate of species discovery continues to climb, perhaps not as rapidly as one might wish, but in the face of fewer researchers the pace still increases, all while using the latest methods and tools. If the number of medical researchers were in decline, then we would lament the pace of medical advances, regardless of technical innovation. The tasks in science are great and we need armies of trained individuals to make progress (Wilson 2017). A lack of understanding that taxonomy and nomenclature are the foundation of all other biological disciplines has created a climate in which these sciences are dismissed and underfunded (Britz *et al.* 2020, Engel *et al.* 2021). If properly funded, then the numbers would grow and the so-called taxonomic impediment could be quickly dismissed. Another side to this, however, is the taxonomic gap, *i.e.*, the disconnect between the reality of biodiversity and our knowledge of it (Dubois 2010a,b, Raposo *et al.* 2020). Not only is it challenged by a diminishing workforce, factors preventing biodiversity exploration have kept those few remaining researchers from adequately engaging with the environments and organisms that are endangered. Furthermore, sampling of biodiversity, so critical for addressing the taxonomic gap, is impeded in most regions of the world to a level hitherto unimagined 50 years ago.

But let us set these matters aside for the moment. Even if we conceded the aforementioned false claims, would data minimalism be a wise response to the challenge at hand? The proponents of data minimalism advocate for the assignment of taxonomic names to COI barcode clusters (Cytochrome Oxidase I clusters, dubbed “BINs”, for Barcode Index Numbers) as computed by BOLD Systems (Barcode of Life Data Systems). A consensus barcode for any given cluster is assumed to “diagnose” the given species. In this way, hundreds of species could be wholesale sequenced for COI, clustered by BOLD, and then simply summarized by a consensus barcode sequence and given a name. Quick and easy. And yet, is it meaningful? One of the fundamental and critical roles of a taxonomist is to test species hypotheses. These rely on data, and ideally as much data as can be called upon to formulate a concept for any given species and what features, be they anatomical, behavioral, chemical, molecular, *etc.*, serve to circumscribe that biological unit from others in nature. Furthermore, those data are also used to identify and distinguish close relatives, ecological connections, life histories, and so forth. Ultimately, these data must allow other researchers to identify the species in question (*i.e.*, there must be meaning to the name), so that the same biological entity can be explored further. If they fail in this regard, then there was no purpose to the name to begin with. So, do BINs adequately test the very species they put forward and allow other researchers to unequivocally identify the species from other specimens, be they alive in the field or preserved in the lab? To the latter part of this question one includes: does this method erode the nomenclatural linkages foundational to effective communication? Empirical analyses of the same data used by the data minimalists indicate that the answer to the former of these is a resounding, no (Meier *et al.* 2022), and, I argue, to the latter an emphatic, yes. The use of BINs as advocated by the data minimalists cannot be applied consistently and COI alone fails to delimit species (Zamani *et al.* 2022). More importantly, the species established by the data-minimalist method adds to the taxonomic impediment considerably. Taxonomists, and other biologists, now hampered by not being able to determine what biological entities are represented by these names, must now spend considerable effort, funds, and time to properly test and characterize these species. Rather than doing the job properly once, the data-minimalist approach throws a dark veil over the meaning of the species they claim to discover, furthering us from our goal of elucidating biological diversity. In what scientific endeavor do you hear

researchers exclaim, “If only I had less data!”? The answer: none. There is a reason for this, and we should not accept the advice of those who advocate for such a stance and disguise it as innovation.

As we survey our world, is the goal really to know how many species there are on Earth? Not who these species *are*, but really only the most superficial interpretation of this question: that is, simply the actual number as enumerated by their names, nothing more, and achieved as quickly as possible, regardless of the consequences. I contend that this number, divorced from best practices in data and meaning, tells us little. What would it matter if the number was 10,000,001 or 10,000,000,001? All it really conveys is that there is a lot of different kinds of life out there. Any child could tell one that same message, saving billions in research investment. This *ultimate* number only has value through the data used to test and understand each of these species and the concepts that underlie them: those that are herbivores, those that are parasites, those that secrete toxins, those that pollinate, those that produce oxygen, those that fix nitrogen, those that live in total darkness under incredible pressures at the bottom of our seas, those that produce important biomedical compounds, those that could be the next model organism, and the diverse histories, occurrences, and interactions these have, both biotic and abiotic. Seems obvious? And yet, how do we know and communicate these things? It is through a systematic evaluation of as much pertinent data as we can muster, communicated through a properly formulated scientific language that is founded on best nomenclatural practices.

Such proposals for data minimalism are likely disastrous for us all. What good is a list of names poorly evaluated and based on the most minimal of data? Getting back to the *why* of scientific naming, these radical protocols erode the foundations for the sciences these names are meant to underpin. Rather than basing decisions on best practices, names are cavalierly established for species that others will have virtually no chance of recognizing. If this is the case, then what good was it to name them in the first place? Our scientific nomenclature will be as useless to predictive science as the names *basilisk* or *unicorn*. Speed should not come at the cost of good science and the goal is not the *list* of all names, but what these names must communicate to us all.

What are best practices in nomenclature?

One goal of science is to transcend our current abilities rather than strive for our limitations. The taxonomic gap is the concept of what we do not yet know about biological diversity, an admission of our missing and limited data. This challenge is only rectified by striving for voluminous data, not by constraining or binding ourselves to the least of our goals.

If we are to lose much diversity in the coming century, then it is imperative that what data we collect are good, maximized, and leveraged for the greatest benefit to our understanding, otherwise the result will have been wasted time and effort as species disappear. Given the empirical demonstration that data minimalism fails to delimit species or provide a means for understanding them by others (Meier *et al.* 2022), the endeavor can only lead to a waste of precious time. It will be more fruitful to have gained more good data from fewer species, than to have meaningless information for them all. The latter scenario will see us worse off than we began.

As in any science, there are recognized best practices. Unfortunately, like any legalistic work, the Codes of Nomenclature, of which there are several, are lengthy, complex, and have their own body of terminology needed to clearly express specific concepts. A full recitation on the Codes is well beyond the scope of the narrative provided here. Nonetheless, perhaps we can distill a suite of practices that could be considered foundational to any taxonomic and nomenclatural effort. Our scientific language must serve all disciplines exploring life, past and present. Accordingly, at minimum the meaning of the names must be clear and founded on more data, not less, and these data must be sufficient to allow for clear and unambiguous communication in ecology, developmental biology, conservation biology, medicine, environmental monitoring, and genetics, as well as all other domains of scientific inquiry and public discourse. The data must be maximized and sufficient to permit efficient and accurate (unequivocal) identification of a discrete evolutionary unit, ultimately allowing an understanding of that organism’s biology and interactions with other species and its environment and role (if any) with human culture, society, and welfare. These concepts of species must be vouchered for arbitration when problems arise and for retesting (repeating the ‘taxonomic experiment’, if you will) as new data are accumulated and new biological theories are advanced.

Every practitioner has a mandate to maximize their data and provide as robust a foundation as can be put forward for their hypotheses of species.

Why does any of this matter?

“Colorless green ideas sleep furiously”
–Noam Chomsky, 1957

As Wittgenstein famously wrote, “*The limits of my language, means the limits of my world.*” Admittedly, his context of linguistic philosophy and his intent differ greatly than that presented here, but there is a notion embedded in this epistemological word play that parallels scientific nomenclature. Indeed, it is a flipside analogue to the sentiment put forth by Linnaeus’ paraphrasing of Isidore of Seville, when he wrote, “*Nomina si nescis, perit et cognitio rerum*” [If you know not the names, then the knowledge of a thing is lost]. A failure to have a robust taxonomy and nomenclature, means that knowledge of our world is confused and lost. Language and the names within that language matter. Philosophers and linguists have known of and debated the importance of language for millennia. Why it is that biology is so cavalier with its own language?

The naming of a thing implies that we know something about it either directly, formally, or in an abstract sense, and that it is via this term that we share that knowledge. If this language is not precise, then the more confused and meaningless biology becomes. It is therefore imperative that we communicate with precision, and this precision requires as much data as we can obtain. In this way our scientific terms shape our universe and our understanding. Folk taxonomies or poorly executed scientific taxonomies lack this precision on a global scale. The former at least can be exacting within their region, but may overlap and lose that precision when applied outside of their cultural context. Thus, our formal scientific taxonomy is meant to provide universal applicability, regardless of cultural practices, time, or space. The term is an exact link to the vast body of data – and a constantly growing body of data – for the object. If data from different species becomes muddled under a single name, then our enterprise crumbles. In this sense, there can be no more fundamental and no more important discipline than taxonomy and its rules of nomenclature. Our solutions to today’s and tomorrow’s challenges rest upon proper concepts of species and the data derived from them. Each of us is living during a pandemic that is leaving millions dead. If we wish to understand and mitigate this scourge, we will need to properly understand what species served as the original source and what potential reservoirs there are for this virus and those like it. Confusion over those species’ identities has direct applicability to whether we meet this challenge, or foolishly suffer greater losses. We do live in an age of extinction, perhaps even our own. Given this, now is the time to strengthen our taxonomy with funding and best practices founded on the most robust data. The reverse shall only lead to miscommunicated babel and unimagined disasters.

Literature cited

- Boero, F. (2010). The study of species in the era of biodiversity: A tale of stupidity. *Diversity*, 2(1): 115–126.
- Britz, R., A. Hundsdörfer, and U. Fritz (2020). Funding, training, permits – the three big challenges of taxonomy. *Megataxa*, 1(1): 49–52.
- Dubois, A. (2010a). Zoological nomenclature in the century of extinctions: priority vs. usage. *Organisms, Diversity & Evolution*, 10(3): 259–274.
- Dubois, A. (2010b). Taxonomy in the century of extinctions: Taxonomic gap, taxonomic impediment, taxonomic urgency. *Taprobanica*, 2(1): 1–5.
- Engel, M.S., L.M.P. Ceriaco, G.M. Daniel *et al.* (2021). The taxonomic impediment: A shortage of taxonomists, not the lack of technical approaches. *Zoological Journal of the Linnean Society*, 193(2): 381–387.
- Fraenkel, A.A., Y. Bar-Hillel, and A. Lévy (1973). *Foundations of Set Theory*, 2nd Revised Edition. North Holland, Amsterdam.
- ICZN (1999). International Commission on Zoological Nomenclature. *International Code of Zoological Nomenclature*, 4th Edition. International Trust for Zoological Nomenclature, London.
- Kremer, M. (1991). Set-theoretic realism and arithmetic. *Philosophical Studies*, 64(3): 253–271.
- Maddy, P. (1990). *Realism in Mathematics*. Clarendon Press, Oxford.

- Maddy, P. (1997). *Naturalism in Mathematics*. Clarendon Press, Oxford.
- Linsky, B. and E.N. Zalta. (1995). Naturalized Platonism vs. Platonized Naturalism. *The Journal of Philosophy*, 92(10): 525–555.
- Meier, R., B.B. Blaimer, E. Buenaventura *et al.* (2022). A re-analysis of the data in Sharkey *et al.*'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. *Cladistics*, 38(2): 264–275.
- Raposo, M.A., G.M. Kirwan, A.C.C. Lourenço *et al.* (2020). On the notions of taxonomic 'impediment', 'gap', 'inflation' and 'anarchy', and their effects on the field of conservation. *Systematics & Biodiversity*, 19(3): 296–311.
- Sharkey, M.J., D.H. Janzen, W. Hallwachs *et al.* (2021). Minimalist revision and description of 403 new species in 11 subfamilies of Costa Rican braconid parasitoid wasps, including host records for 219 species. *ZooKeys*, 1013: 1–665.
- Wilson, E.O. 2017. Biodiversity research requires more boots on the ground. *Nature Ecology & Evolution*, 1(11): 1590–1591.
- Zamani, A., Z.F. Fric, H.F. Gante *et al.* (2022). DNA barcodes on their own are not enough to describe a species. *Systematic Entomology*, 47(3): 385–389.

Michael S. Engel

Natural History Museum and
Department of Ecology & Evolutionary Biology
University of Kansas
Lawrence, Kansas 66045
USA

Division of Invertebrate Zoology
American Museum of Natural History
New York, New York 10024
USA

Section Editor: *Taprobatica*, the Journal of Asian Biodiversity
November 22nd, 2022