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

Rudolf Meier[1,2], Bonny Blaimer[2], Eliana Buenaventura[2], Emily Hartop[3,4], Thomas von Rintelen[2], Amrita Srivathsan[1], Darren Yeo[1]

[1] Department of Biological Sciences, National University of Singapore, Singapore

[2] Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Center for Integrative Biodiversity Discovery, Berlin, Germany

[3] Zoology Department, Stockholms Universitet, Stockholm, Sweden

[4] Station Linné, Öland, Sweden

Abstract

Halting biodiversity decline is one of the most critical challenges for humanity, but biodiversity assessment and monitoring are hampered by taxonomic impediments. We here distinguish between a "dark taxon impediment" caused by a large number of undescribed species and a "superficial description impediment" caused by species descriptions so imprecise that type specimens have to be consulted in order to resolve species identities. Recently, Sharkey et al. (2021) proposed to address the dark taxon impediment for Costa Rican braconid wasps by describing 403 species based on barcode clusters ("BINs") computed by BOLD Systems. The default assumption of the revision is that BIN=Species (e.g., BOLD:ACM9419 becomes *Bracon federicomatarritai* Sharkey, sp. nov.) and therefore the diagnoses of most species consist only of a consensus barcode. We here argue that this type of "minimalist revision" is unnecessary and undesirable. It is unnecessary because barcode clusters (e.g. BINs) already provide grouping statements that overcome many of the obstacles associated with dark taxon impediments. However, minimalist revisions are also undesirable and problematic because the diagnoses are only based on one character system – that in the case of Sharkey et al. was poorly analyzed. Furthermore, the revision relies on units that violate basic rules of reproducibility because the BINs were delimited by a proprietary algorithm (RESL) that is applied to a mixture of public and private data. Here, we illustrate that many of the BINs described as species are unstable when the available public data are reanalyzed, reiterate that *COI* mostly measures time of divergence, and that BOLD Systems violates key principles of open science. We conclude by urging authors, reviewers, editors, and grantors to only publish and fund projects that adhere to modern standards of reproducibility.

After being a mostly academic subject for many decades, biodiversity decline is now also a major topic for decision makers (e.g., economists, CEOs, political leaders) (World Economic Forum, 2020). This expanded interest in biodiversity research has highlighted once more that one of the most important jobs in science is unfinished: the discovery and description of earth's biodiversity (May, R.M. 2011). Any quantitative analysis of the problem reveals that most of the species-level diversity and biomass is concentrated in taxonomically poorly known microbial, fungal, and invertebrate clades. Within invertebrates are pockets of light (e.g., butterflies, dragonflies, bees), but also very large areas occupied by "dark taxa" with unknown numbers of species (Hausmann, A., Krogmann, L., et al. 2020). These dark taxa are avoided by taxonomists and other biologists alike because they are not only hyperdiverse but also includes hyper-abundant species. This avoidance contributes to what is known as the "taxonomic impediment" which arguably has two different sources. Firstly, there is the "dark taxon impediment". Most animal species are unknown and undescribed. The proportion of the unknown, or "dark", fauna in a clade increases with species richness and specimen abundance and decreases with body size. Secondly, there is the "superficial description impediment". Most species descriptions from the first 150 years of systematics are too superficial by today's standards to allow for the identification of the taxon without inspecting type specimens, collecting additional data, and re-describing the species. Addressing this superficial description impediment has been extremely time-consuming because it requires museum visits or loans, type digitization, lengthy searches for misplaced specimens, and re-descriptions of what should have been described in greater detail in the first place. Indeed, new species descriptions could be accelerated manifold if taxonomists did not have to spend so much time on resolving existing descriptions.

Both problems combined create the toxic mélange known as the taxonomic impediment, the composition of which roughly follows a latitudinal gradient. In the temperate regions, due to a long history of taxonomic work, much of the taxonomic impediment is caused by superficial descriptions. The species have been described – many multiple times – but few well enough that names can be resolved without consulting types (i.e., re-descriptions are needed). As one approaches the equator, the "superficial description impediment" is largely replaced by the "dark taxon impediment", but there are usually just enough poor-quality species descriptions that they still interfere significantly with biodiversity discovery and description. What the 21st century undoubtedly needs is faster biodiversity discovery, but what should be avoided at all costs is creating a new and large "superficial description impediment"; i.e., descriptions of large numbers of species that are so poorly documented and supported by evidence that future generations of taxonomists will have to regularly consult types and generate additional data.

Yet, this is what Sharkey et al. (2021) are proposing in their "Minimalist revision and description of 403 new species in 11 subfamilies of Costa Rican braconid parasitoid wasps, including host records for 219 species". The authors state: "… we view barcode-based descriptions as a first pass in an iterative approach to solve the taxonomic impediment of megadiverse and under-taxonomically resourced groups that standard technical and biopolitical approaches have not been able to tackle." This means that Sharkey et al. delegate the critical "iterative" work to future generations of taxonomists. They will have to start revisions by first revisiting the species descriptions, types, and specimens of Sharkey et al.'s species to resolve species boundaries based on data that should have been collected and analyzed at the time of description.

Sharkey et al.'s (2021) proposal is a shift from the status quo. Barcode clusters are widely used as "first pass" grouping statements that require further validation before description. Policy shifts like this should come with a thorough analysis of costs and benefits. Presumably, the main benefit of Sharkey et al.'s (2021) proposal is generating scientific names. Such names can be essential for announcing evidence for a new species and for filing and retrieving biological information associated with these species. Indeed, new taxa discovered based on morphological evidence only become known to the scientific community once they are described and a scientific name is assigned. However, this is not the case for taxa that are discovered based on molecular evidence. For them, the evidence is shared rapidly and efficiently via dynamically evolving sequence databases that usually also include associated biological information (e.g., locality and host information). This means that describing first-pass groupings based on DNA sequences must come with benefits that outweigh the costs of leaving them undescribed as barcode clusters such as Barcode Index Numbers (BINs). We here argue that this is only the case when the first-pass taxa are validated as species by collecting enough evidence for rigorous species delimitation ("integrative taxonomy"). If the validation involves morphological data, there are additional benefits. It facilities the comparison with described species, it allows for the inclusion of historical specimens lacking high-quality DNA, and avoids that biologists who have no access to molecular data can participate in the biodiversity discovery process. This is particularly important given that many biologists in biodiverse countries have limited access to sequencing data.

However, the vast majority of Sharkey et al.'s (2021) species descriptions lack any of these benefits that come with data integration. Instead, the "minimalist revision" creates a host of problems. The first is that the species diagnoses are static compilations of data from two dynamically evolving databases (Barcode of Life Data Systems and a database summarizing host information). How cumbersome this is becomes clear when a new braconid barcode from Costa Rica, obtained after the publication of Sharkey et al. (2021), needs to be identified. The identification should be based on the databases because they contain all available evidence, while the 665-page "minimalist revision" became outdated the moment a new braconid sequence was added after the publication date. Even if the new sequence has a good hit to a named BIN, assigning the scientific name to the new sequence is precarious because, as we will show, many of the described BINs are not stable. In this reply, we illustrate these problems by firstly re-analyzing the public data underlying the minimalist revision. Secondly, we reiterate the theoretical and empirical reasons why the cytochrome oxidase I gene (*COI*) should not be used as the only data source for species descriptions. Thirdly, BOLD Systems calculates BINs using an algorithm that is not publicly available and based on a large amount of "private" data that are inaccessible. This means that BINs violate basic requirements for reproducibility in science. We conclude with suggestions as to how several of these issues can be addressed without impeding or slowing biodiversity discovery.

(1) Many BINs are unstable.

To test the stability of the BINs used in Sharkey et al. (2021), we re-analyzed the available underlying data. Conventionally, such re-analyses of data from published studies are straightforward because the data are made available by the authors. However, the "molecular data" in the supplementary material of Sharkey et al. (2021) consist only of neighbor-joining (NJ) trees. We thus proceeded to obtain the sequences directly from BOLD by consulting the BIN numbers in the revision and NJ trees in the supplement. Note, however, that BINs are calculated based on public and private data with the ratio between the two apparently being approximately 3:1 (see below). In addition, the NJ trees in the supplement contained non-Costa Rican BINs from other Neotropical countries that were not covered by

the revision. Consequently, we ended up with three datasets of increasing size: (1) public barcodes for only those BINs described by Sharkey et al. (2021), (2) public barcodes for all BINs found on the NJ trees in the supplementary materials, and (3) all public barcodes in the BOLD database for the 11 braconid subfamilies covered by the revision of the Costa Rican fauna (see supplementary data).

All three datasets were analyzed with several species delimitation methods as recommended by Carstens et al. (2013). The overall analyses followed Yeo et al. (2020): objective clustering with uncorrected p-distance thresholds at 1-5% using TaxonDNA (Meier, R., Shiyang, K., et al. 2006), ASAP – a new implementation of ABGD – where we only kept the results obtained with the top five partitions (= lowest p-values) (Puillandre, N., Lambert, A., et al. 2012, Puillandre, N., Brouillet, S., et al. 2021), and Single-rate PTP (Zhang, J., Kapli, P., et al. 2013) based on maximum likelihood trees generated from the barcode data with RAxML v.8 (Stamatakis, A.J.B. 2014). These re-analyses were used to test whether the data analyzed with the three methods unambiguously supported the BINs that were described as species by Sharkey et al. (2021). Arguably, this is a minimum requirement for using BIN=Species as default because it is hard to justify describing barcode clusters as species if the former are not even stable despite sparse sampling.

We find that in dataset 1, 131 of the 401 BINs (32.7%) conflict with at least one of the species delimitation treatments. The instability of these BINs increases as more *COI* evidence is analyzed (Fig. 1), with 283 of the 615 BINs (46.0%) from dataset 2, and 2912 of the 3896 BINs (74.7%) from dataset 3 in conflict with at least one of the analyses. The corresponding numbers for only those 401 BINs described as species in Sharkey et al. (2021) are 138 (34.4%) and 276 (68.8%). This means that as braconid barcodes are sampled more densely across the Neotropics and the world, an increasing number of the BINs described as species in Sharkey et al. (2021) are stable.

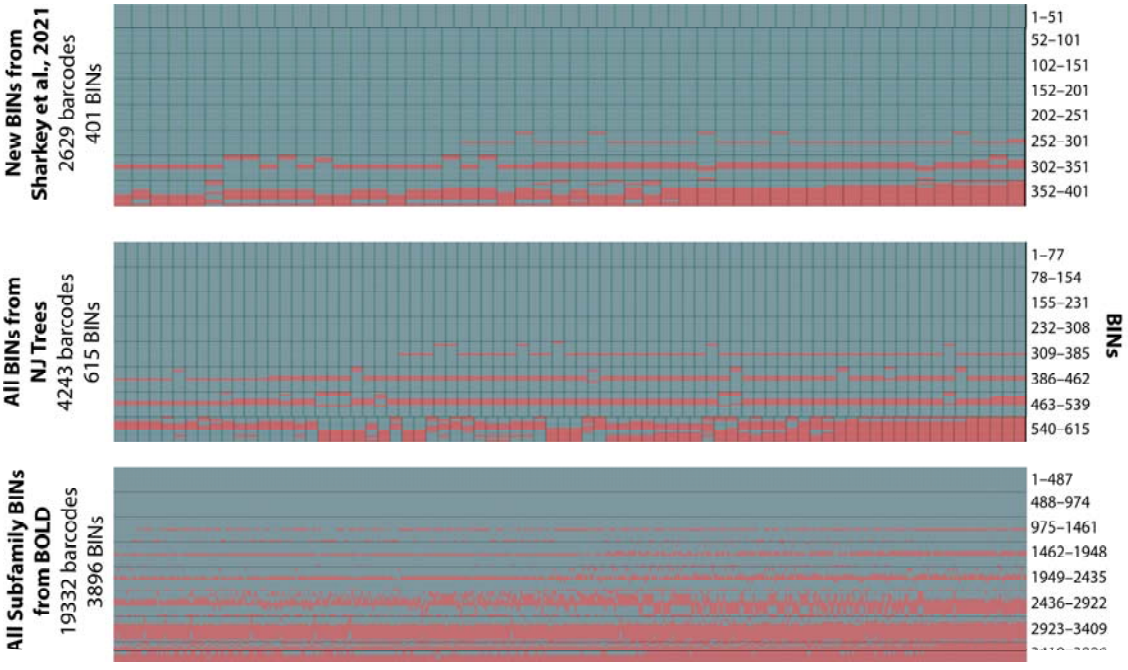

Figure 1. Instability heatmaps for three datasets of increasing size. In order to accommodate a large number of BINs, each heatmap is split into 8 blocks that are separated by dark horizontal lines (number of bins represented in each block is provided). Rows within each block represent species delimitation

algorithms (PTP, objective clustering: 1 – 5% p-distance, ASAP: top 5 priors) and BINs in columns (blue = congruent BINs; pink = incongruent BIN).

To obtain lower-bound estimates for instability, we identified the parameters for each species delimitation algorithm (objective clustering, ASAP) whose application maximized congruence with BINs. We find that 38 of the 401 BINs (9.48%) described in the paper are incongruent even after the methods were optimized to replicate BINs as well as possible (Fig. 2). Incongruence again increases as more of the available barcodes and BINs are included [dataset 2: 82 of the 615 BINs incongruent (13.33%), dataset 3: 1340 of the 3896 BINs (34.39%) incongruent]. These analyses illustrate that the available *COI* data only ambiguously support 10-30% of the BINs described as new species. This finding is congruent with virtually all studies that have looked into the stability of barcode clusters (Virgilio, M., Backeljau, T., et al. 2010, Kekkonen, M. and Hebert, P.D.N. 2014, Srivathsan, A., Hartop, E., et al. 2019, Yeo, D., Srivathsan, A., et al. 2020, Hartop, E., Srivathsan, A., et al. 2021). Furthermore, this 10-30% estimate is a low estimate because denser *COI* sampling closes barcoding gaps between species, and thus leads to more uncertainty with regard to the boundaries of barcode clusters (Bergsten, J., Bilton, D.T., et al. 2012). Even with the current extent of sampling, Sharkey et al. (2021) neglected to include non-Costa Rican barcodes in their minimalist revision: "Braconid specimens from the following New World countries appear to be relatively well-sampled in BOLD: Canada, USA, Belize, Argentina, French Guiana, and Mexico. There is a small number of cases where specimens from these countries fall in the same BIN as one of our Costa Rican species, but they were not studied. More sampling between these disparate localities, and more genomic and/or morphological and behavioral data will help resolve these species-level cases, which are beyond the scope of this paper." It is difficult to understand why any scientist would leave out available evidence that is relevant for a given study. If workload was the main concern, one could have revised a few of the 11 subfamilies based on all evidence instead of covering all 11.

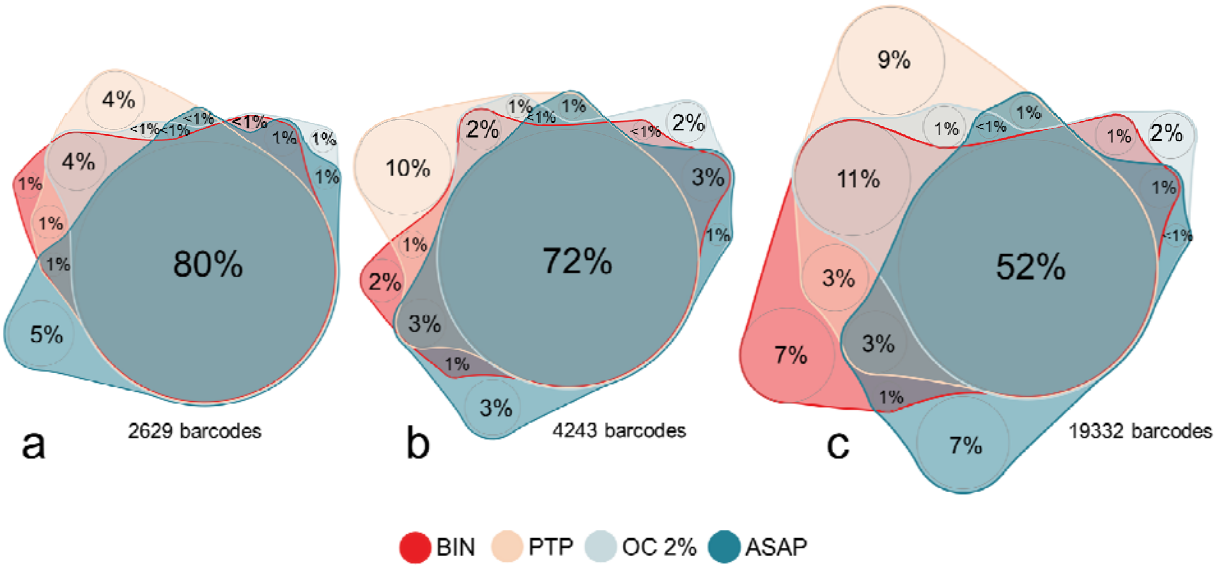

Figure 2. Barcode cluster congruence across methods using optimal thresholds for BIN, PTP, Objective Clustering (OC) (2%) and ASAP (d -3) for (a) dataset I, (b) dataset II, (c) dataset III.

The instability of BINs directly affects the value of the species diagnoses in Sharkey et al. (2021). The diagnoses for most species consist of a consensus barcode based on the publicly available barcode data. This raises three issues. The first is that some of the diagnoses were already incorrect at the time of species description because they lacked the information from the private barcodes that were used for delimiting the BINs. The second is that static diagnoses are inappropriate summaries of data obtained from dynamic databases. Not checking the BOLD database for new sequences would be tantamount to ignoring evidence. Thirdly, given that the application of other species delimitation algorithms would have yielded other barcode clusters, many of the described BINs were already unstable at the time the minimalist revision was compiled. This means that the species diagnoses in Sharkey et al. (2021) are for units with poor support. It is thus near certain that many of Sharkey et al.'s (2021) new species will have to be revisited in the near future in order to obtain sufficient data to establish their validity. "Minimalist revisions" are thus only minimalist at the time of description, while creating a major "superficial description impediment" in the future. Sharkey et al. (2021) acknowledge this problem when they consider their approach to be a "first pass in an iterative approach" but they never explain why they are willing to accept the disadvantages of BIN=Species when there are so few if any benefits.

We suspect that Sharkey et al. (2021) overlooked the instability of BINs because they relied too much on BINs supplied by BOLD Systems. As in all scientific studies, it is important that the authors analyze the underlying data thoroughly in order to present robust conclusions. For species descriptions based on molecular data, this means utilizing different species delimitation algorithms as is common practice (Virgilio, M., Backeljau, T., et al. 2010, Carstens, B.C., Pelletier, T.A., et al. 2013, Kekkonen, M. and Hebert, P.D.N. 2014, Yeo, D., Srivathsan, A., et al. 2020). It also means avoiding the results obtained with algorithms that are not publicly available (RESL) and operate on a mixture of public and private data.

Evidence for BIN instability is also readily available when one traces the "BIN history" of specimens. Here is the history for three BINs described as species in Sharkey et al.'s (2021) minimalist revision (see suppl. materials for details):

- BOLD:AAH8697 (*Heterogamus donstonei*) has 54 members of which 17 are public (but 18 are available on the BIN page). Of the 18 barcodes, 14 barcodes are stable. One unstable specimen (BCLDQ0377) was originally placed in a fly BIN (BOLD:AAG1770: Diptera: Muscidae: *Lispocephala varians*; 5 June 2010) before moving on to the current wasp BIN (19 June 2010). The three remaining unstable barcodes were all originally placed into the current BIN in September 2012, but subsequently shifted into two BINs without public members (in May 2013). All three then shifted back into the current BIN on 8 August 2015.
- BOLD:AAV3035 (*Pneumagathis erythrogastra*, new combination in Sharkey et al. (2021)) has 14 members of which 12 are barcode compliant. Only two barcodes are stable. The remaining 10 have similar histories. They were first placed in a stonefly BIN (BOLD:AAC5216: Plecoptera: Chloroperlidae: *Alloperla severa*; March 2010) before shifting to a BIN that is no longer available (January 2011). It is unclear when these specimens shifted to the current BIN because this is not revealed by the delta view tool in the BOLD Systems workbench. The two other barcodes were first placed in BIN(s) that are no longer available. One (specimen H1170) then shifted to the current BIN in September 2012 but later shifted to a BIN with a single private member in May 2013. This specimen returned to the current BIN in June 2018. The other specimen (specimen H7621) shifted from the original placement (BIN has ceased to exist) to the current BIN in June 2018.

- BOLD:ABY5286 (*Chelonus scottmilleri*). The history for the 15 public barcodes is detailed in Figure 3. It involves two BINs that are "no longer available" (BOLD:AAA2380 and BOLD:AAK1017) and two others that could not be accessed as they are "awaiting compliance with metadata requirements" (BOLD:AAD2009 and BOLD:ABX7466). Note that the two currently valid BINs (BOLD:AAA7014 and BOLD:ABX5499) contain species from different subfamilies of Braconidae (Microgastrinae: *Apanteles anamartinezae* and Cheloninae: *Chelonus scottshawi*, respectively).

Instability is not the only problem with BINs. For example, BOLD:ABX6701 was described by Sharkey et al. (2021) as *Plesiocoelus vanachterbergi*. The consensus barcode includes two single indels and was presumably obtained from the seven "barcode-compliant" sequences in BOLD Systems (as of April 18, 2021). Only one of the seven barcodes is translatable to amino acids with the remaining six having deletions. Sequences showing these attributes are often derived from pseudogenes, and it is conceivable that *P. vanachterbergi* was described based on paralogs. Note that this is likely due to a lapse in quality control because BOLD is supposed to check for shifts in reading frames (Ratnasingham, S. and Hebert, P.D.N. 2013).

All this highlights that relying heavily on BIN=Species for describing species is not advisable and thorough data analysis is important. Indeed, depending on the time of BIN description, two of Sharkey et al.'s (2021) wasp taxa would have been described as a fly or stonefly species, respectively. Similarly, biologists using BOLD as an identification tool should be aware that using BINs can lead to misleading conclusions. This makes it all the more important that BIN history can be traced easily. However, this is difficult in BOLD Systems. One can only reconstruct BIN evolution by tracing each individual specimen –one at a time – and clicking through pages of changes over time (this will still only reveal information on publicly available sequences). This is unfortunate because "taxon concepts" matter greatly in taxonomy (Franz, N.M. 2005, Meier, R. 2016, Packer, L., Monckton, S.K., et al. 2018) and it would be particularly straightforward to establish a "BIN-concept tracking" feature in BOLD Systems given that it is relatively a new database and only deals with one type of data.

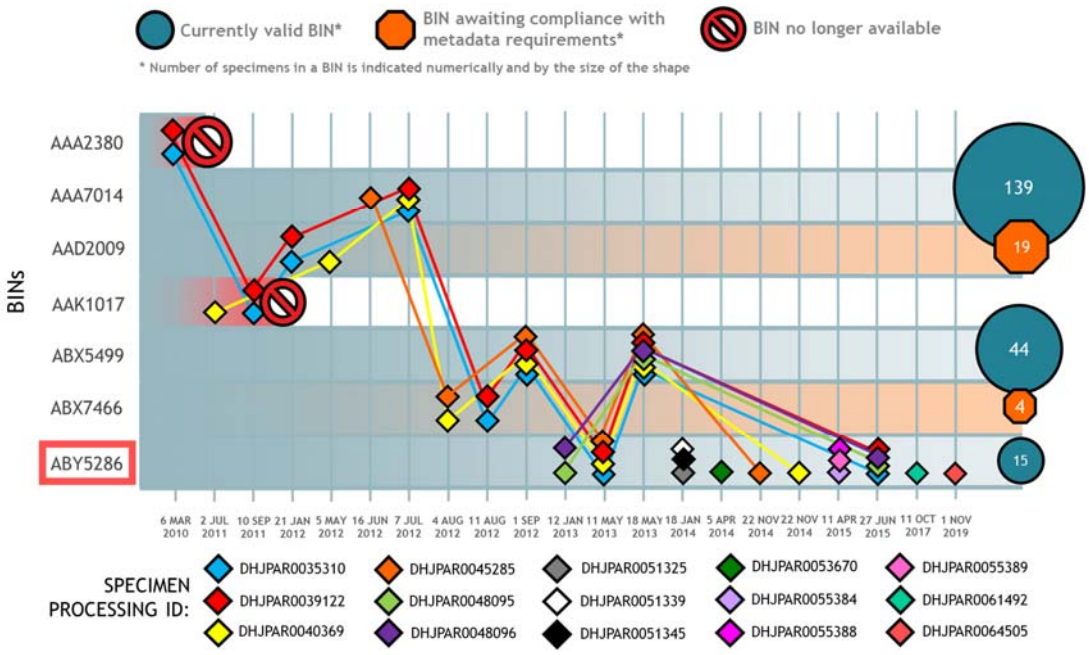

Figure 3. BIN membership through time for the 15 public barcodes in BOLD:ABY5286 (*Chelonus scottmilleri*). Two BINs are "no longer available"; two others cannot be accessed as they are "awaiting compliance with metadata requirements"; the two currently valid BINs contain species from different subfamilies (AAA7014 Microgastrinae: *Apanteles anamartinezae* and ABX5499 Cheloninae: *Chelonus scottshawi*). They contain 139 barcodes (all public) and 44 barcodes (although 45 listed as public), respectively.

The instability of species estimates obtained with barcodes is also evident when evaluating barcoding publications. Many of these are widely cited and published in high profile journals (e.g., PNAS), but the results reveal much instability over time. We initially focused on publications that were highlighted in Sharkey et al. (2021). We then added additional ones that allowed for a direct comparison of species numbers (=authors provided a list of GenBank accession or BOLD sample IDs). We obtained the sequences in February-April 2021 using these numbers by querying GenBank, the Sequence-ID tool of GBIF, and BOLD Systems. Table 1 compares the number of species and BINs for those species that had molecular data. We find that, for example, the barcodes for the 10 species in *Astraptes* belong to 5 BINs (Hebert et al. 2004), the 32 species of *Belvosia* reported in Smith et al. (2006) belong to 20 BINs, and the barcodes in Smith et al. (2007) are in 61 BINs instead of the 73 species reported in the paper. These are widely cited publications and yet it remains unclear which "species" numbers or boundaries should be trusted. Either BINs are not equal to species or the species reported in the publications are not species.

Table 1: **Comparison of published species and BIN numbers (April 2021)**.

Study	Taxa	Journal	BINs in BOLD	Species or BINs published	Deviation
Hebert et al. 2004*	*Astraptes* (Lepidoptera: Hesperiidae)	PNAS	5	10 species	100%
Smith et al. 2006*	*Belvosia* (Diptera: Tachinidae)	PNAS	20	32 species	60.0%
Raupach et al. 2016	Carabidae (Coleoptera)	ZooKeys	51	69 BINs	35.3%

Chacón et al. 2013*	*Dunama* (Lepidoptera: Notodontidae, Nystaleinae)	ZooKeys	6	8 species	33.3%
Gibbs 2018	*Lasioglossum* (Hymenoptera: Halictidae)	Genome	85	110 species	29.4%
Fernandez-Triana et al. 2014*	*Apanteles* (Hymenoptera: Braconidae)	ZooKeys	240	180 species	25.0%
Smith et al. 2007*	various Tachinidae (Diptera)	PNAS	61	73 species	19.7%
Janzen et al. 2011*	Eudaminae and Pyrginae *(Hesperiidae) skippers*	PLoS One	294	347[§] species	18.0%
Hajibabaei et al. 2006*	Hesperiidae, Sphingidae, Saturniidae (Lepidoptera)	PNAS	600	521 species	13.2%
Fleming et al. 2019	*Hyphantrophaga* (Diptera, Tachinidae)	Biodiversity Data Journal	25	22 species	12.0%
Bartolini et al. 2020	*Anastrepha* (Diptera, Tephritidae)	PLoS One	17	15 species	11.8%
Pohjoismäki et al. 2016	Tachinidae (Diptera)	PLoS One	331	366 BINs	10.6%
Smith et al. 2008*	Braconidae (Hymenoptera)	PNAS	287	313 species	9.1%
Fleming et al. 2014*	*Houghia* (Diptera: Tachinidae)	Zootaxa	34	35 species	2.9%
Ortiz et al. 2017	Erebidae (Lepidoptera)	Biodiversity Data Journal	157	160 species	1.9%
Burns et al. 2008*	*Perichares* (Lepidoptera: Hesperiidae)	PNAS	5	5 species	0%

* Study cited in Sharkey *et al.*

[§] number includes 44 putative barcode clusters.

(2) BINs and the origin and evolution biodiversity.

Many BINs are unstable, but it is equally important to ask the question whether there are theoretical reasons why *COI* barcodes should not be the sole data source for delimiting species. The answer in the literature is a fairly unanimous "yes", even by researchers who extensively use DNA sequences in their programs (Meier, R., Shiyang, K., et al. 2006, Puillandre, N., Lambert, A., et al. 2012, Puillandre, N., Modica, M.V., et al. 2012, Zhang, J., Kapli, P., et al. 2013). *COI* barcode clusters are unlikely to reflect species boundaries for those taxa where introgression, lineage sorting, and maternal inheritance obfuscated the species-level signal in the sequence (Will, K.W. and Rubinoff, D. 2004, DeSalle, R., Egan, M.G., et al. 2005). Additional reasons are rooted in what we know about *COI* protein evolution. In closely related species, the gene tends to be under strong stabilizing selection at the amino-acid sequence level (Roe, A.D. and Sperling, F.A.H. 2007, Kwong, S., Srivathsan, A., et al. 2012, Pentinsaari, M., Salmela, H., et al. 2016). This explains why most of the evolutionary change in the *COI* DNA sequence is synonymous and concentrated in 3rd positions. Nucleotide fixation in these positions is likely caused by genetic drift; i.e., *COI* distances between BINs and closely related species mostly measure time of divergence. This can be illustrated by the three datasets that we analyzed here. We find numerous BINs whose barcodes translate into identical amino acid sequences (Table S1: dataset 1: 11 cases involving 26 BINs; dataset 2: 21 cases and 49 BINs; dataset 3: 28 cases and 57 BINs – see supplementary materials for details) and the proportion across the literature can be as high as 80-90% (Kwong, S., Srivathsan, A., et al. 2012). These BINs have no known, functionally significant biological differences because the biology of these wasp "species" is not at all affected by which triplet codons are used to code for identical proteins. Given that most biologists associate speciation with the origin of biologically meaningful differences, describing such BINs as species rests on the hope that the correlation between time of

divergence and the origin of new species is strong enough that biologically meaningful differences will later be found. However, most biologists are rightfully skeptical of results based on correlations and we would argue that it would be healthy to adopt the same position here. Correlations warrant more exploration; i.e., barcode clusters should only be treated as first pass taxa, but not described.

The biggest concern is not that BINs can be "wrong" but that the error is non-random and will lead to misleading conclusions with regard to the evolution of biodiversity. Species originating during adaptive radiations will be systematically lumped and described as one species while old species will be systematically split. Evidence for this problem can be found throughout BOLD Systems. Anyone who has worked with the database will have found many cases where one BIN contains sequences that were identified as multiple species. Conversely, it is common to find species that have barcodes in multiple BINs. This could be due to misidentification or due to genuine lumping/splitting. Fortunately, there are taxa for which there is enough morphological, genetic, and behavioral data that species boundaries are well understood. For these, one can test whether the sequences of recently diverged species end up in one BIN and the sequences for old species are found in several BINs. Such data are available, for example, for many species of Sepsidae (Diptera) (Puniamoorthy, N., Ismail, M.R.B., et al. 2009, Puniamoorthy, N., Kotrba, M., et al. 2010, Tan, D.S., Ang, Y., et al. 2010, Ang, Y., Puniamoorthy, J., et al. 2013, Ang, Y.C., Wong, L.J., et al. 2013, Araujo, D., Tuan, M., et al. 2014, Rohner, P.T., Ang, Y., et al. 2014, Ang, Y., Rajaratnam, G., et al. 2017). Known pairs of closely related species that are distinct with regard to morphology and behavior are routinely found lumped into the same BIN (e.g., *Sepsis neocynipsea* and *S. cynipsea*: BOLD:AAC2855; *Sepsis orthocnemis* and *S. fulgens*: BOLD:AAJ7599; *Themira lucida and T. flavicoxa*: BOLD:AAD7140). *Sepsis punctum* is an example for the opposite. The populations in North America and Europe can interbreed but are split into two BINs (BOLD:AAG5639; BOLD:ACS2531). If taxonomists had applied BIN=Species as a default for these cases, several young species would have been overlooked and some old species would have been split. Some will argue that this degree of error is acceptable, but this error is non-random and thus systematically misinforms about speciation. After all, non-taxonomists approach species with the assumption that they are meaningful units that are predictive of biological properties.

One can only wonder how many of the 403 new species in the minimalist revision will impede with future research on how braconid species diversity in Costa Rica originated. Sharkey et al. (2021) show awareness of this problem and occasionally deviate from the BIN=Species default. For example, they recognized that one BIN contained seven morphologically very distinct species and described them as such. Highlighting these cases and describing the species is arguably the most important contribution of the minimalist revision. However, unfortunately it remains unclear how rigorously external morphology and host data were used in challenging BINs. Even though Sharkey et al. (2021) claimed that "COI data never lumped specimens that were markedly different morphologically in the same BIN; nor did it lump specimens with divergent host data", the assessment of their BINs under different species delimitation methods would most likely invalidate this assertion. Indeed, the role of morphology in braconid taxonomy is discussed in contradictory ways. On the one hand, Sharkey et al. (2021) argue that morphological evidence is not suitable for braconid taxonomy. On the other hand, they point out that subtle morphological differences often agree with barcode clusters, which would imply that the morphological evidence was misinterpreted. Yet, perceived conflict tends to be resolved in favor of BINs. This is sometimes justified by pointing to host associations (Sharkey, M.J., Janzen, D.H., et al. 2021). However, host data are only available for half the species and letting host data overwrite

morphology makes hidden assumptions about the host specificity of the wasp species. Overall, we must conclude that the methods used in the revision are too poorly described to fully understand how the authors gathered and treated evidence. Furthermore, even if morphology were to fail for braconid taxonomy, this cannot be used to justify BIN=Species. Instead, this should trigger the search for other character systems that are more suitable (Puillandre, N., Modica, M.V., et al. 2012). Until the data for these systems are available, one may as well let BINs be BINs. After all, BOLD Systems was designed as a workbench and not as a species-delimitation platform (Ratnasingham, S. and Hebert, P.D.N. 2013).

(3) BINs, BOLD Systems, Open Science, and Reproducibility:

The minimum standard for reproducibility in scientific publications is that the results can be replicated based on the described methods and data/tools in the manuscript. There are three reasons why the "minimalist revision" does not meet this standard. All three are related to open science issues within BOLD Systems. Firstly, the algorithm for calculating BINs (RESL: Refined Single Linkage) is not publicly available. Secondly, it is unclear how the underlying sequences are aligned. Thirdly, BINs are calculated based on public and private barcodes:

Open Science and Reproducibility: RESL. RESL computes BINs, but the widely cited reference paper only outlines the general strategy (Ratnasingham, S. and Hebert, P.D.N. 2013), while the code remains proprietary. This makes it impossible to vary parameters, such as the pairwise distance used for the initial clustering, threshold used for merging neighbors, and inflation parameters for cluster refinement. Furthermore, OTUs with very similar "Silhouette scores" are not available. This opacity and lack of user control gives the impression that BINs are stable, but in reality they are more akin to a mean that lacks a standard deviation, or a tree branch that lacks information on node support. Overall, we know very little about what the optimal settings for RESL would be because it was simultaneously proposed and implemented based on eight small datasets that together comprised only 18,843 barcodes compared to today's >8 million sequences stored in BOLD Systems (Ratnasingham, S. and Hebert, P.D.N. 2013). Three of these training datasets were for moths/butterflies, two for birds, and two for fish (last one: bees) (Ratnasingham, S. and Hebert, P.D.N. 2013). Despite the limited taxonomic range, the optimal clustering thresholds for these datasets varied from 0.7% to 1.8%. Eventually, "2.2% was adopted as it represents the upper 99% confidence limit for the optimal thresholds in the eight test datasets ($x\square$=1.26), SD =0.40)" (Ratnasingham, S. and Hebert, P.D.N. 2013). In the publication that optimized RESL, the same eight datasets were also analyzed with four other species delimitation algorithms (ABGD, CROP, GMYC, jMOTU). Overall, RESL and GMYC performed best with 89% congruence with morphology. Note, however, that this level of congruence was only achieved after RESL was applied to the same eight datasets for which it was trained. A proper test should have involved additional datasets for different taxa.

All remaining comparisons of RESL to other species delimitation algorithms are compromised because RESL can only be used within BOLD Systems and the other delimitation algorithms cannot be applied to BOLD System's private data. The only way to get BIN assignments from RESL is via submission of sequences to BOLD Systems. BINs are then calculated by BOLD Systems administrators during the next scheduled BIN calculation run based on aligned public and private data. This means that the researcher will not be able to obtain BIN assignments for only the submitted data, but it will be equally impossible to compare BINs to the units obtained with other algorithms, since the private data and alignments are not available for analysis with the latter. Effectively, this means that the comparative performance of

RESL remains unclear. Note, however, that it is unlikely that it is the best available method for clustering barcodes. As discussed by Ratnasingham, S. and Hebert, P.D.N. (2011), the algorithm was developed with speed and scalability in mind and the authors already anticipated that "[f]uture research will undoubtedly reveal analytical approaches that are better at recognizing species boundaries from sequence information than RESL." Such algorithms that were designed with precision in mind have since become available and are suitable for the kind of smaller datasets that underlie most published studies. This also pertains to the data used in Sharkey et al.'s (2021) revision. This means that the authors could and should have analyzed the relevant public data with these tools.

Open Science and Reproducibility: Alignment. It is unclear how BOLD Systems aligns barcode sequences. Again, only the general strategy is described: "Each sequence that passes all quality checks is translated to amino acids and aligned to a Hidden Markov Model (HMM) of the COI protein [43]. The aligned amino acids are back translated to nucleotides to produce a multiple sequence alignment", but the "parameters for BOLD's alignment of sequences are not publicly available" (Nugent, C.M., Elliott, T.A., et al. 2020). Evidence for alignment is ubiquitous in FASTA downloads of individual BINs that often include gaps in multiples of three not required for the BIN-only data. Given the number of sequences in BOLD Systems, alignment is nontrivial, and the details should be published so that the BINs on BOLD Systems are reproducible.

Open Science and Reproducibility: Private data. BINs are calculated based on public and private barcodes. The ratio of public and private barcodes is difficult to determine because the relevant numbers vary between what is listed on BOLD Systems under "Identification engine" and "Taxonomy" on the one hand, and what can be downloaded from BOLD Systems via the "Public Data Portal" on the other hand. We here estimate that approximately one quarter of all data are private. In order to obtain this estimate, we downloaded the data for each taxonomic category listed under Animals in BOLD Systems' "Taxonomy". Given the large size of the Arthropoda and Insecta datasets, sub-taxa were downloaded separately. Downloading was done either directly from the web browser or by using BOLD Systems' API between 27 March-11 April 2021. Overall, we obtained 6,165,928 COI barcodes >500 bp (6,508,300 overall) of the 8,480,276 (as of 18 April 2021) barcodes that are listed as "All Barcode Records on BOLD". We estimate that obtaining these data and setting up the bioinformatics pipeline cost >200 million C$, given that one project alone (Barcode500K: 2010-2015) had a budget of 125 million C$ (https://ibol.org/programs/barcode-500k/). It is very likely that some of these funds were also used for generating private data. This includes private BINs that have remained private for many years. Many databases in science include embargoed private data (e.g., NCBI), but the embargo can only be imposed for short time periods (e.g., 12 months) and all scientific work on these databases is restricted to the public data. This ensures that at any point in time, the proportion of private data is under control and results remain reproducible.

Scientific standards of barcoding studies. An overall dangerously relaxed attitude towards scientific standards also affects other aspects of Sharkey et al.'s (2021) minimalist revision and many other barcoding studies. Almost 10 years ago, Collins, R. and Cruickshank, R.J.M.e.r. (2013) described "The seven deadly sins of DNA barcoding". This included the use of Neighbor-Joining (NJ) trees. NJ trees are known to yield biased results when the input order of taxa is not randomized during tie-breaking when ties are encountered during tree construction (Takezaki, N. 1998). Yet, it remains unclear whether the NJ algorithm used by BOLD System breaks ties randomly. Secondly, BOLD Systems' NJ trees lack support values. This is problematic because NJ algorithms only generate one tree even when multiple trees have

a similarly good fit to the data. Thirdly, the NJ trees in BOLD Systems are based on K2P (Kimura-2-Parameter) distances, although all model testing indicates that this is an inappropriate model for *COI* (Collins, R.A., Boykin, L.M., et al. 2012, Srivathsan, A. and Meier, R. 2012), whereas RESL uses uncorrected distances. This means the NJ trees and BINs in BOLD Systems are based on different measures of barcode similarity. Many authors – including Sharkey et al. (2021) – argue that the use of K2P NJ trees is acceptable because they are "only" used for visualization. However, visualization tools also need justification and there is no study that shows that K2P NJ trees outperform other tools when it comes to illustrating which species are well-corroborated (if anything, there is evidence to the contrary: (Virgilio, M., Backeljau, T., et al. 2010)). Fourthly, it is unclear how the criteria for barcode compliance were developed. According to iBOL, "DNA barcodes" have to be >500 bp and retain <1% N. However, the length requirement has repeatedly been found to be unfounded (Karim, M. and Abid, M.R. 2020, Yeo, D., Srivathsan, A., et al. 2020, Piper, A.M., Cogan, N.O.I., et al. 2021), which is probably also why Sharkey et al. (2021) describe two species based on barcodes that are too short to be "DNA barcodes." All these issues only persist because there is too much leniency by reviewers, journals, and grantors when dealing with barcoding studies and projects. However, ultimately it will be beneficial for the field if the scientific standards were raised.

Change is needed. We highlight all these issues because iBOL is currently raising funds for a new barcoding campaign that is supposed to last 7 years and cost 180 million C$. DNA barcodes were proposed almost 10 years ago as a convenient identification tool. Barcodes subsequently evolved into a species discovery tool at a time when biodiversity decline was still mostly a largely academic concern. However, now it is seen as an immediate threat to the planet's survival. This means that strategies and structures that were designed 10 years ago may no longer be appropriate. One example is the relaxed attitude towards open science and scientific rigor. Proprietary algorithms and alignment parameters, large amounts of private data, K2P NJ trees without support values, etc. have no place in today's science. Arguably, the same applies to barcoding procedures that involve large-scale specimen movements across continents. The main barcoding facility is currently in Canada because efficient use of Sanger sequencing required automation given that each amplicon has to be processed separately. However, barcoding is now accomplished with 2nd and 3rd generation sequencing technologies. This means that it is now more efficient and cost effective to barcode in decentralized facilities all over the world (cf https://ccdb.ca/pricing/ with (Meier, R., Wong, W.H., et al. 2016, Wang, W.Y., Srivathsan, A., et al. 2018, Srivathsan, A., Lee, L., et al. 2021). Overall, organizations of a certain size tend to become static unless there is a healthy amount of scrutiny. In the case of barcoding, this scrutiny should come from contributors, journal editors, manuscript reviewers, and grantors. Journals rightfully insist on reproducible science, this means that they should only publish manuscripts that use barcode clusters and species hypotheses that can be replicated based on publicly available resources. As for grantors, one would hope that they become vigilant about adherence to open access policies. For example, Genome Canada is a major funder of DNA barcoding and has an explicit data sharing policy that contradicts how RESL and alignment parameters are handled: "Genome Canada is strongly committed to the principle of rapid sharing of the outputs of Genome Canada-funded research including open access to publications, release of data and sharing of unique resources to the scientific community" (https://www.genomecanada.ca/en/programs/guidelines-and-policies).

We would like to conclude our critique of Sharkey et al. (2021) with some suggestions. Firstly, one should not describe barcode clusters as species. Such clusters are first-pass taxa, but there is no benefit

to describing them as species without testing them with other character systems. This is because barcode evidence is more appropriately shared on dynamic databases and there are many disadvantages to describing BINs as species. Too many of them are unstable and BINs=Species misconstrues how speciation occurs. Secondly, it is time to publish RESL and provide access to the private data or exclude it from BIN calculation. Thirdly, BOLD Systems currently only provides one type of barcode cluster (BINs), but it could easily implement additional clustering algorithms for subsets of data. This would provide users with much needed information on which barcode clusters are stable. Fourthly, BOLD Systems currently lacks BIN tracking over time. As pointed out earlier, BIN compositions can change over time. A tracking too for "BIN concepts" is essential for validating the results of studies that used BINs computed in the past (e.g., see stonefly/fly examples). Lastly, the relationship between BINs and species in BOLD Systems should be revisited. The original publication describing BOLD Systems highlighted the workbench character of the database. BINs known to contain several verified species were supposed to be labeled with decimal numbers ("BOLD:AAB2314.1"). However, this option has been largely ignored. This means that the BIN matches of searches within BOLD Systems or GBIF's SequenceID lead to BINs, even if they are known to contradict species boundaries. BINs containing sequences for multiple species should be flagged.

Conclusions:

COI barcodes are extremely important for biodiversity discovery but using BIN=Species as a default for species descriptions without proper data analysis and confirmatory evidence is problematic and unnecessary. Barcode clusters are "first pass" grouping statements and should only be converted into species upon validation with other data sources. Otherwise, the next generation of taxonomists will be burdened by a "superficial description impediment". Indeed, in many ways Sharkey et al.'s (2021) BIN=Species default is reminiscent of the problems with species descriptions from the 19th century. The descriptions were too superficial because they relied too much on one kind of data that were not analyzed sufficiently (19th century: superficial external morphology; now: *COI* BINs). Furthermore, the sole character system that was used was poorly sampled (19th century: mostly temperate species; now: ca. 8 million barcodes for >10 million species). Sharkey et al. (2021) are pessimistic about new 21st century solutions to the dark-taxon impediment. We do not share this sentiment. Large throughput imaging and sequencing are already available and not vague promises for the future. Similarly, the data can now be analyzed with increasingly sophisticated algorithms that will provide taxonomists with a solid foundation for species descriptions that can be based on multiple sources of data (Hartop, E., Srivathsan, A., et al. 2021). These data will be particularly suitable for generating automatic species descriptions. Now is not the time to promote "minimalist revisions" that mostly replace alphanumerical identifiers for unstable BINs with species names. To prevent further damage to taxonomy, we urge journal editors, reviewers of taxonomic manuscripts, and grantors to insist that publications are based on thoroughly analyzed public data. Manuscripts relying on BINs and K2P NJ trees should trigger the request for proper data analysis during manuscript review. Regarding the open-science and governance issues, journal editors and grantors should treat everyone equally; algorithms and data must be public or else the results are not publishable and grant proposals cannot be funded.

Acknowledgements

We would like to thank Roderic D. M. Page and two additional colleagues for providing valuable comments on the manuscript. This work was supported by a Ministry of Education grant on biodiversity discovery (R-154-000-A22-112).

References

Ang Y, Puniamoorthy J, Pont AC, Bartak M, Blanckenhorn WU, Eberhard WG, Puniamoorthy N, Silva VC, Munari L, Meier R. 2013. A plea for digital reference collections and other science-based digitization initiatives in taxonomy: Sepsidnet as exemplar. Systematic Entomology, 38:637-644.

Ang Y, Rajaratnam G, Su KF, Meier R. 2017. Hidden in the urban parks of New York City: Themira lohmanus, a new species of Sepsidae described based on morphology, DNA sequences, mating behavior, and reproductive isolation (Sepsidae, Diptera). ZooKeys:95-111.

Ang YC, Wong LJ, Meier R. 2013. Using seemingly unnecessary illustrations to improve the diagnostic usefulness of descriptions in taxonomy-a case study on Perochaeta orientalis (Diptera, Sepsidae). Zookeys:9-27.

Araujo D, Tuan M, Yew J, Meier R. 2014. Analysing small insect glands with UV-LDI MS: high-resolution spatial analysis reveals the chemical composition and use of the osmeterium secretion in Themira superba (Sepsidae: Diptera). Journal of evolutionary biology, 27:1744-1750.

Bartolini I, Rivera J, Nolazco N, Olortegui A. 2020. Towards the implementation of a DNA barcode library for the identification of Peruvian species of Anastrepha (Diptera: Tephritidae). Plos One, 15.

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, Geijer J, Herrmann J, Foster GN, et al. 2012. The Effect of Geographical Scale of Sampling on DNA Barcoding. Systematic Biology, 61:851–869.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN. 2008. DNA barcodes and cryptic species of skipper butterflies in the genus Perichares in Area de Conservacion Guanacaste, Costa Rica. Proceedings of the National Academy of Sciences of the United States of America, 105:6350-6355.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. Molecular Ecology, 22:4369-4383.

Chacon IA, Janzen DH, Hallwachs W, Sullivan JB, Hajibabaei M. 2013. Cryptic species within cryptic moths: new species of Dunama Schaus (Notodontidae, Nystaleinae) in Costa Rica. Zookeys:11-45.

Collins R, Cruickshank RJMer. 2013. The seven deadly sins of DNA barcoding, 13:969-975.

Collins RA, Boykin LM, Cruickshank RH, Armstrong KFJMiE, Evolution. 2012. Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification, 3:457-465.

DeSalle R, Egan MG, Siddall M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B-Biological Sciences, 360:1905-1916.

Fernandez-Triana JL, Whitfield JB, Rodriguez JJ, Smith MA, Janzen DH, Hallwachs WD, Hajibabaei M, Burns JM, Solis MA, Brown J, et al. 2014. Review of Apanteles sensu stricto (Hymenoptera, Braconidae, Microgastrinae) from Area de Conservacion Guanacaste, northwestern Costa Rica, with keys to all described species from Mesoamerica. Zookeys:1-565.

Fleming AJ, Wood DM, Smith MA, Dapkeyl T, Hallwachs W, Janzen D. 2019. Twenty-two new species in the genus Hyphantrophaga Townsend (Diptera: Tachinidae) from Area de Conservacion Guanacaste, with a key to the species of Mesoamerica. Biodiversity Data Journal, 7.

Fleming AJ, Wood DM, Smith MA, Hallwachs W, Janzen DH. 2014. Revision of the New World species of Houghia Coquillett (Diptera, Tachinidae) reared from caterpillars in Area de Conservacion Guanacaste, Costa Rica. Zootaxa, 3858:1-90.

Franz NM. 2005. On the lack of good scientific reasons for the growing phylogeny/classification gap. Cladistics, 21:495-500.

Gibbs J. 2018. DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. Genome, 61:21-31.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences of the United States of America, 103:968-971.

Hartop E, Srivathsan A, Ronquist F, Meier R. 2021. Large-scale Integrative Taxonomy (LIT): resolving the data conundrum for dark taxa. bioRxiv

Hausmann A, Krogmann L, Peters RS, Rduch V, Schmidt S. 2020. GBOL III: Dark taxa. Barcode Bulletin, 10.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences of the United States of America, 101:14812-14817.

Janzen DH, Hallwachs W, Burns JM, Hajibabaei M, Bertrand C, Hebert PDN. 2011. Reading the Complex Skipper Butterfly Fauna of One Tropical Place. Plos One, 6.

Karim M, Abid MR. 2020. Accuracy responses in species identification varying DNA barcode lengths with a Naïve Bayes classifier: Efficacy of mini-barcode under a supervised machine learning approach. bioRxiv.

Kekkonen M, Hebert PDN. 2014. DNA barcode-based delineation of putative species: efficient start for taxonomic workflows, 14:706-715.

Kwong S, Srivathsan A, Vaidya G, Meier R. 2012. Is the COI barcoding gene involved in speciation through intergenomic conflict? Molecular Phylogenetics and Evolution, 62:1009-1012.

May RM. 2011. Why Worry about How Many Species and Their Loss? PLoS Biol, 9.

Meier R. 2016. Citation of taxonomic publications: the why, when, what and what not. Systematic Entomology:DOI: 10.1111/syen.12215.

Meier R, Shiyang K, Vaidya G, Ng PKL. 2006. DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. Systematic Biology, 55:715-728.

Meier R, Wong WH, Srivathsan A, Foo MS. 2016. $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. Cladistics, 32:100-110.

Nugent CM, Elliott TA, Ratnasingham S, Adamowicz SJ. 2020. coil: an R package for cytochrome c oxidase I (COI) DNA barcode data cleaning, translation, and error evaluation, 63:291-305.

Ortiz AS, Rubio RM, Guerrero JJ, Garre MJ, Serrano J, Hebert PDN, Hausmann A. 2017. Close congruence between Barcode Index Numbers (bins) and species boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. Biodiversity Data Journal, 5.

Packer L, Monckton SK, Onuferko TM, Ferrari RR. 2018. Validating taxonomic identifications in entomological research. Insect Conservation and Diversity, 11:1-12.

Pentinsaari M, Salmela H, Mutanen M, Roslin T. 2016. Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. Scientific Reports, 6:35275.

Piper AM, Cogan NOI, Cunningham JP, Blacket MJ. 2021. Computational Evaluation of DNA Metabarcoding for Universal Diagnostics of Invasive Insect Pests. bioRxiv

Pohjoismaki JLO, Kahanpaa J, Mutanen M. 2016. DNA Barcodes for the Northern European Tachinid Flies (Diptera: Tachinidae). Plos One, 11.

Puillandre N, Brouillet S, Achaz GJMER. 2021. ASAP: assemble species by automatic partitioning, 21:609-620.

Puillandre N, Lambert A, Brouillet S, ACHAZ GJMe. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation, 21:1864-1877.

Puillandre N, Modica MV, Zhang Y, Sirovich L, Boisselier M-c, Cruaud C, Holford M, Samadi S. 2012. Large-scale species delimitation method for hyperdiverse groups. Molecular Ecology, 21:2671-2691.

Puniamoorthy N, Ismail MRB, Tan DSH, Meier R. 2009. From kissing to belly stridulation: comparative analysis reveals surprising diversity, rapid evolution, and much homoplasy in the mating behaviour of 27 species of sepsid flies (Diptera: Sepsidae). Journal of Evolutionary Biology, 22:2146-2156.

Puniamoorthy N, Kotrba M, Meier R. 2010. Unlocking the "Black box": internal female genitalia in Sepsidae (Diptera) evolve fast and are species-specific. BMC Evolutionary Biology, 10.

Ratnasingham S, Hebert PDN. 2011. BOLD's role in barcode data management and analysis: a response. Molecular Ecology Resources, 11:941-942.

Ratnasingham S, Hebert PDN. 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. Plos One, 8:e66213.

Raupach MJ, Hannig K, Moriniere J, Hendrich L. 2016. A DNA barcode library for ground beetles (Insecta, Coleoptera, Carabidae) of Germany: The genus Bembidion Latreille, 1802 and allied taxa. Zookeys:121-141.

Roe AD, Sperling FAH. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. Molecular Phylogenetics and Evolution, 44:325-345.

Rohner PT, Ang Y, Lei Z, Puniamoorthy N, Blanckenhorn WU, Meier R. 2014. Genetic data confirm the species status of Sepsis nigripes Meigen (Diptera: Sepsidae) and adds one species to the Alpine fauna while questioning the synonymy of Sepsis helvetica Munari. Invertebrate Systematics, 28:555-563.

Sharkey MJ, Janzen DH, Hallwachs W, Chapman EG, Smith MA, Dapkey T, Brown A, Ratnasingham S, Naik S, Manjunath R, 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 4.

Smith MA, Rodriguez JJ, Whitfield JB, Deans AR, Janzen DH, Hallwachs W, Hebert PDN. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences of the United States of America, 105:12359-12364.

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN. 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proceedings of the National Academy of Sciences of the United States of America, 104:4967-4972.

Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera : Tachinidae). Proceedings of the National Academy of Sciences of the United States of America, 103:3657-3662.

Srivathsan A, Hartop E, Puniamoorthy J, Lee WT, Kutty SN, Kurina O, Meier R. 2019. Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. BMC Biology, 17:96.

Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier RJb. 2021. MinION barcodes: biodiversity discovery and identification by everyone, for everyone. BioRxiv.

Srivathsan A, Meier R. 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics, 28:190-194.

Stamatakis AJB. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, 30:1312-1313.

Takezaki N. 1998. Tie trees generated by distance methods of phylogenetic reconstruction. Molecular Biology and Evolution, 15:727-737.

Tan DS, Ang Y, Lim GS, Ismail MRB, Meier R. 2010. From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of Sepsis pyrrhosoma (Sepsidae: Diptera). Zoologica Scripta, 39:51-61.

Virgilio M, Backeljau T, Nevado B, De Meyer MJBb. 2010. Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics, 11:1-10.

Wang WY, Srivathsan A, Foo M, Yamane SK, Meier R. 2018. Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. Molecular Ecology Resources, 18:490-501.

Will KW, Rubinoff D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics, 20:47–55.

World Economic Forum, 2020. The Global Risks Report: https://www.weforum.org/reports/the-global-risks-report-2020 .

Yeo D, Srivathsan A, Meier R. 2020. Longer is Not Always Better: Optimizing Barcode Length for Large-Scale Species Discovery and Identification. Systematic Biology, 69:999-1015.

Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics, 29:2869-2876.

Supplementary Information

Materials & Methods

Data retrieval. Barcode data were extracted from BOLD for the three datasets via the following approaches BIN numbers were added to the sequence headers of each barcode for all datasets:

- Dataset 1: New BINs from Sharkey et al. (2021): A list of BINs were compiled from the Sharkey et al. (2021) paper. The BINs were then retrieved from BOLD via the following url: http://www.BOLD Systems.org/index.php/API_Public/sequence?bin=BOLD:**<BIN>**
- Dataset 2: All BINs from NJ Trees: A list of BINs were compiled from the NJ trees in the supplementary of Sharkey et al. (2021). The BINs were then retrieved from BOLD via the following url: http://www.BOLD Systems.org/index.php/API_Public/sequence?bin=BOLD:**<BIN>**
- Dataset 3: All subfamily BINs from BOLD: The subfamily names included in the paper were used as search terms in BOLD System's public data portal. The barcodes were downloaded in FASTA format and compiled.

Species delimitation. The sequences in all three datasets were first aligned with MAFFT v7 (Katoh and Standley 2013) using a gap opening penalty of 5.0. Dataset 3 initially produced an untranslatable alignment with multiple indels. Sequences with a large number of internal gaps were separated from the other barcodes, with the remaining processed using a gap opening penalty of 5.0, while the --add function was used to add the problematic barcodes to the alignment using the default gap opening penalty. All alignments and scripts are available from https://figshare.com/articles/dataset/BOLD_braconid_datasets/14452992. The aligned barcodes were then processed using three species delimitation algorithms that represent both distance and tree-based approaches. Objective clustering under TaxonDNA (Meier et al. 2006) was performed using a range of distance thresholds (1-5% p-distance) common in the *COI* barcoding literature for Metazoa (cluster algorithm in TaxonDNA: Meier et al. 2006, 2016, Ratnasingham & Hebert 2013). ASAP (Puillandre et al. 2021) was performed using simple distances (d -3) and the top five partitions (lowest p-values) were used in this study. A maximum likelihood tree was generated with RAxML v.8 (Stamatakis 2014) via its rapid hill-climbing algorithm (-f d) under the GTRCAT model on CIPRES (Miller et al. 2011). The resulting best tree was processed with single-rate PTP (--single) (Zhang et al. 2013) using the mPTP scripts (Kapli et al. 2017).

Python scripts were used to identify which clusters from the above species delimitation algorithms and parameters were congruent with the BINs in a pairwise manner. The most congruent parameters for each delimitation algorithm (PTP, ASAP, objective clustering under TaxonDNA) were identified and cluster congruence visualized with nVenn diagrams (Pérez-Silva et al. 2018).

Translation and amino acid distances. The barcodes for the three datasets were aligned with TranslatorX (Abascal et al., 2010) using the invertebrate mitochondrial genetic code with the MAFFT method. The aligned amino acid sequences were processed using the UCLUST algorithm (Edgar 2010) with the cluster_fast command at -id 1.0 to obtain clusters of identical amino acid sequences. A Python script was then used to identify and quantify which of these identical clusters contained multiple BINs.

Table S1. List of BINs with identical amino acid sequences

Dataset 1:

Case	BINs	Case	BINs
1	AAD7334, ABU7454	7	AAA5374, AAB8975, ACK7801
2	AAK1016, ABX5499, ABY5286, ACJ3551	8	AAB1652, ABA9319, ACX5348
3	AAB1658, ADB1219	9	AAE0329, ADJ3083
4	AAA5372, AAJ4092	10	AAJ3968, ABZ7672
5	ACJ4722, ACL5572	11	ABZ9478, ACJ4626
6	ADF4538, ADA8821		

Dataset 2:

Case	BINs	Case	BINs
1	AAB8975, ACK7801	12	ACL5572, ACJ4722
2	AAM5951, ACG8400, ACJ2495, ACO7975	13	ADF4538, ADA8821
3	AAK1016, ABX5499, ABY5286, ACJ3551	14	ACJ5020, ADC9608
4	ACN9714, ADC9450	15	AAA5374, AAB8975, ACK7801, AEB4766
5	AAB1658, ADB1219	16	AAB1652, ABA9319, ACX5348
6	AAD7334, ABU7454	17	AAE0329, ADJ3083
7	ACA4726, ACA4727, ADC9733, AEB4152	18	AAJ3968, ABZ7672
8	ACB1292, ACY8111	19	ABZ9478, ACJ4626
9	AAA5372, AAJ4092	20	ACJ5330, ADC9429, ADC9430
10	ABA9319, ABY3599	21	ACK7467, AEF5717
11	ACJ2111, ACN0949		

Dataset 3:

Case	BINs	Case	BINs
1	AAU8750, ACF4955, ADJ2714	15	AAH8628, AAV7498
2	AAG1434, ADA8867	16	ACF4946, AAH8726
3	AAD7425, AAU9842	17	ACE1136, AAK5490
4	AAH0622, ACQ8886	18	ACD2708, AAJ1227
5	ACE4571, AEA6054	19	ACI8848, AAV7450
6	AAG1395, ADE1516	20	ACT1567, ACV2385
7	ABY5568, AAD9345	21	AEG0007, ACM0677
8	AAZ3033, ACM0741, ACU4378	22	ADI0213, AAH8773
9	AAG1350, ACO1868	23	ACJ5330, ADC9430
10	AAU8556, ADN5159	24	ACI7729, ACV3169
11	AAG1277, ACA6737	25	ADO9563, AAZ3033
12	ABU5601, AAH8651	26	ADA0629, ACP6154
13	ABY9715, ACE4773	27	AAH8868, ACQ9364
14	ADH5015, ACE6023	28	AAB6029, ABY9714

Survey of BOLD Systems. A download of all COI-5P from BOLD Systems cannot be done directly as the query is too large. Data was instead downloaded for each taxonomic category listed under Animals in

BOLD Systems Taxonomy http://boldsystems.org/index.php/TaxBrowser_Home. Given the large size of Arthropoda dataset, the sub-taxa were downloaded separately and the same approach was adopted to Insecta dataset. The downloads were done either directly from the web browser or using BOLD Systems API, and "Combined data" was obtained for each. A separate download was also conducted for all data available under Data Releases section. As it is stated that this release only contains preliminary information of taxonomy, the updated metadata was obtained from the full download. Summary statistics available online in BOLD Systems were recorded and search terms were optimized based on the full data downloaded and these summary statistics. If the numbers were identical/very similar, those search terms were applied for 2015 dataset.

Table Sxxx. Metazoan data available in BOLD Systems Public database

		# Seqs (2021 only)	# species (2021 only)	#COI-5P (2021/2015)	# BINs (2021/2015)
Total	**11,722,663**	**7,471,895**	**302,179**	**6,508,300/2,859,684**	**570,500/320,473**
Arthropoda	10,321,635	6,572,565	230093	5,678,581/2,747,509	489,784/300,166
Chordata	842,287	518,022	35724	493,448/75,717	40,529/13,054
Mollusca	247,044	187,019	17764	179,263/10,096	20,919/2340
Annelida	104,937	61,922	4465	58,439/14,579	9272/3176
Echinodermata	57,578	29,768	1866	26,069/7684	2220/874
Cnidaria	29,900	20,353	3206	18,256/2004	1379/323
Nematoda	35,218	22,852	2609	14,490/657	1241/112
Rotifera	13,268	9307	509	9175/439	988/108
Platyhelminthes	38,963	29,271	2714	11,435/430	971/91
Porifera	7889	5099	1431	4537/71	730/44
Nemertea	5683	3858	414	3800/129	569/40
Bryozoa	4137	1842	245	1652/159	377/96
Tardigrada	3006	1769	232	1578/105	336/15
Onychophora	1393	1363	183	1353/5	246/2
Gastrotricha	1351	712	206	372/0	192/0
Acanthocephala	2306	2145	108	2032/2	182/2
Chaetognatha	1743	1344	102	1292/30	181/8
Sipuncula	1318	670	74	652/29	128/8
Others (<100 BINs)	3007	2014	234	1876/39	256/14