i. Species diagnosis and DNA taxonomy

Dirk Ahrens

Museum A. Koenig Bonn, Leibniz Institute for the Analysis of Biodiversity Change; Germany. Email: ahrens.dirk_col@gmx.de, https://orcid.org/0000-0003-3524-7153

ii. Summary/Abstract

The use of DNA has helped to improve and speed up species identification and delimitation. However, it also provides new challenges to taxonomists. Incongruence of outcome from various markers and delimitation methods, bias from sampling and skewed species distribution, implemented models, and the choice of methods/ priors may mislead results and also may, in conclusion, increase elements of subjectivity in species taxonomy. Lack of direct diagnostic outcome from most contemporary molecular delimitation approaches and need of reference to existing and best sampled trait reference systems, reveals the need of refining criteria of species diagnosis and diagnosability in the current framework of nomenclature codes and good practices to avoid nomenclatorial instability, parallel taxonomies, and consequently more and new taxonomic impediment.

iii. Key Words Integrative taxonomy, nomenclature, DNA barcoding, species delimitation, morphology, good practice, taxonomic impediment

1. Introduction

DNA taxonomy[1] - to characterize species based on DNA sequences - have become a well-feasible and highly effective in the past two decades ***(1, 2, 3, 4)***. Meanwhile, the array and number of markers (Table 1) as well as approaches for species identification and delimitation have considerably diversified ***(5, 6, 7)***. Among these, DNA barcoding ***(1)*** is one of the most popular ones ***(4, 8)***. DNA taxonomy has increased the quality and reproducibility of species' characterization and enabled rapid assessments of biodiversity ***(9)***. A major advantage of DNA Barcoding is the ability to standardize and automatize species recognition by using a single,

[1] I do not refer to the older but stricter definition in which DNA taxonomy was meant as a framework to replace the Linnean name-based taxonomy (see ***(4, 11)***). Since Tautz's et al. proposal ***(11)*** was not accepted, the term is being used in the above, wider sense in most of the literature ***(3)***.

standardized gene fragment (e.g., ***(1, 5, 7, 10)***) which already includes (by definition) all its potential characters. DNA-based species delimitation made way for the direct inference of species boundaries from unknown samples ***(12)***, allowed association of different life stages with each other ***(13)***, and the study of environmental DNA and bulk samples ***(9)***.

2.Species, barcodes, and taxonomy

2.1 From species to barcodes

The approach of matching unknown samples with DNA sequences of known (identified and named) specimens is the basic principle behind the DNA-based species identification approach. The known specimens are recruited from (morphospecies-informed) reference libraries which are being built in the various DNA barcoding initiatives which are ultimately mirrored in NCBI, EMBL or BOLD ***(14).*** However, these currently cover not more than ca 2-3% of World's known species diversity (Fig. 1). The success of identification is strictly linked to the completeness of these libraries (in terms of species and coverage of their variation). Libraries are often geographically quite restricted.

Successful identification results are supposed to capture unambiguously the "match" between reference and query sequences. This is ultimately linked to the question, how to define "match"? This definition is scientifically strictly linked to the recognition of species boundaries since libraries are not supposed to reflect the entire spectrum of genetic variation of each species.

Another issue for the success and long life of the reference libraries is the continuously curated taxonomic consistency ***(15)***. Integration of large-scale data and a consideration of an evolving taxonomy (new synonymies, newly discovered species etc.), not only in invertebrates, makes a strictly organized and regulated data processing and taxon entry/referencing system indispensable. Alternative nomenclatures (by different specimen providers or by updated taxonomy in time) may effectively mess up the reference libraries and make quality checks of and identification with barcode data very tedious or impossible. The same problem is known from decades of faunistic literature in diverse arthropod groups (e.g., beetles)

having a continuously evolving taxonomy. Here, auxiliary numerical species identifiers have been in successful use for almost 36 years in a limited geographical extent ***(16, 17)***.

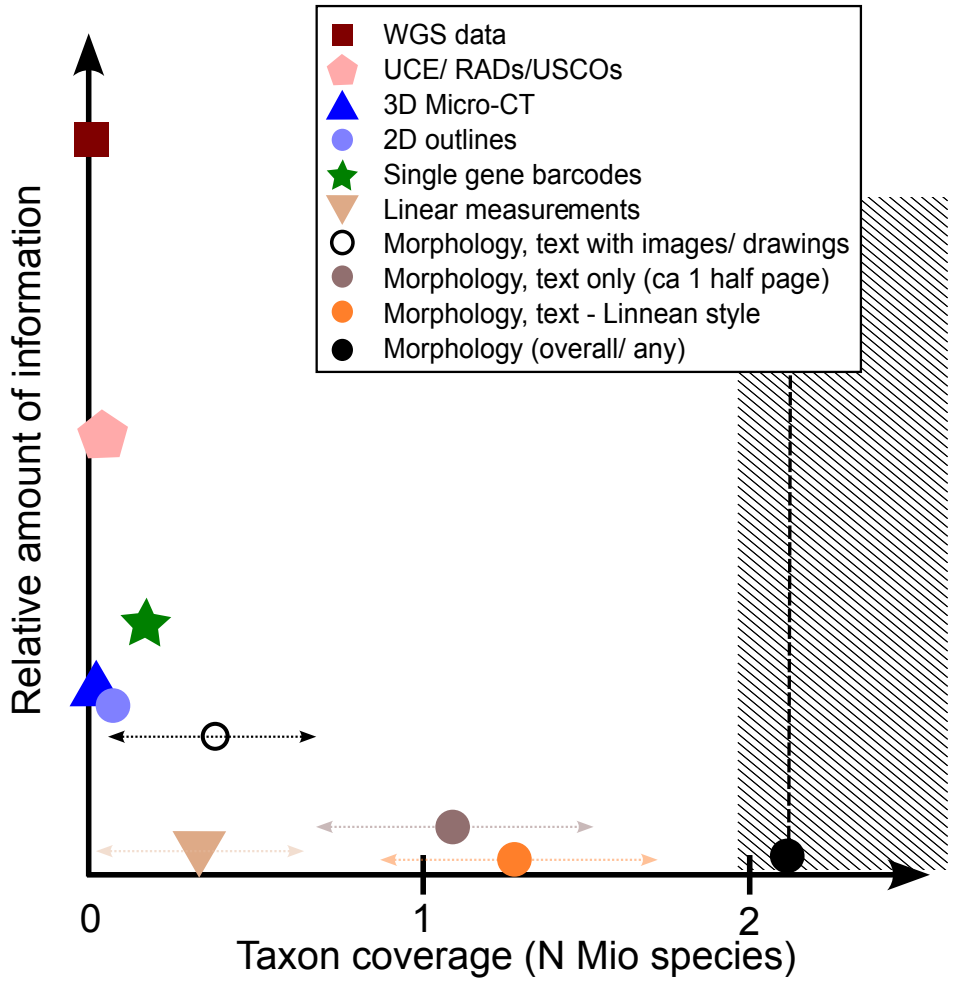

Fig. 1. Approximate schematic estimate of trait information content and species coverage of different character traits used for diagnosing species in the light of validly described taxa (grey vertical line). (Categories of taxonomic investigation (used as follows: Linnean style descriptions (ca 2 lines, morphology), 19th century's description (one page text, no images, morphology), 20th century's description (one page text, with images, genitalia figures, morphology), Morphometric data/ Artificial intelligence (on 2D images), DNA barcodes, Genomic data (AFLP, RADseqs), 3D CT scans, 3D surface images, full genomes). Descriptions based on words (black), images (white) and molecular characters (grey).

BOLD is a good example for clear and sustainable reference data requirements. The same we cannot say about current taxonomic science and how it is organized (see ***(18)***). ZOOBANK ***(19)*** was a good beginning, in terms of standards and rigor (e.g., for the availability of new species names from electronic publications in the era of digital information). To meet the needs for a modern taxonomic science and biodiversity

data mining/management as well as today's standards of digital data sharing and integration policies, more rigor and standardization might be needed (particularly in the publication step), which however, will not come without costs for the underfunded science taxonomy.

2.2 From DNA barcodes to species

Since the beginning, there have been a vivid controversy about how to use DNA markers and in particular single gene barcodes in the context of species taxonomy (e.g., ***(4, 11, 20, 21, 22, 23, 24, 25)*** and many more). The discussion intensified again after some authors have been starting to formally describe and name species exclusively using barcodes (e.g., ***(26, 27, 28)***[2]; see also ***(18, 29, 30, 31, 32, 33)***). In this context it became evident ***(18)*** that criteria of a species diagnosis, which is essential foundation for the availability of a newly established taxon name, are ambiguous in the Code for Zoological Nomenclature ***(34)***.

2.3 Taxonomy and nomenclatural stability

In past, taxonomic and nomenclatural stability was mainly challenged by two issues: 1) misused taxon names due to poor species descriptions and lacking revision of the type specimens (also due to lack of clear rules and guidelines in shape of a code of nomenclature until 19th century; ***(35)***), and by 2) competing species hypotheses derived from different species concepts (e.g., ***(36)***) or different authority 'opinions' (e.g., ***(37)***; see also ***(38)***).

So far, problems of competing species concepts have affected mainly the vertebrate taxonomist's community ***(36)***. Invertebrate taxonomists instead mainly have been focusing on morphology, for a variety of reasons but probably most of all because they mainly have been experiencing most of their examined specimens as dead vouchers in collections. A major trigger behind this tendency has been apparently to observe and recognize their species alive and the number of "taxonomists" *dealing*

[2] In some cases, diagnoses were made specific by highlighting diagnostic nucleotide positions, and in others, such as ***(27)***, delimitation was based on threshold clustering (2%) as implemented in the Barcode of Life Index Number (BIN; ***(39)***).

with the same group of organisms in a respective geographic area (including also the large rapidly interfering user community, e.g., conservation managers and ecologists).

Particularly in vertebrates (and more recently not only), probably due to the many "authorities" but also due to the fact that many vertebrate species are more considered by conservation and wildlife observations, there is a tendency of a wish of democratically or autocratically (by the scientific community and key users) decided taxonomies (expressed as lists of accepted taxa, e.g., ***(38, 40, 41, 42, 43)***). Such ideas are opposed by calls for an only science-based taxonomy on the other hand ***(44)***.

While cataloging life has been used since the very beginning of taxonomic research and has been very important to the user of taxonomy, taxon lists (peer-reviewed or not) have now in the era of cybertaxonomy an enormous impact. They are being replicated on the internet among the users and are generally also used as taxon backbone for large biodiversity information databases such as GBIF, NCBI, BOLD, etc. However, sometimes these lists lose track of their original reference, include errors or different lists are competing against each other ***(42)***.

A related, but more recent problem is that traditional, previously mainly morphology-based taxonomies (in invertebrates) are being challenged by minimalistic approaches that seek to define species not relating to existing diagnostic reference systems (i.e., morphology) ***(28)*** partly without proper diagnoses. Despite we are used to controversy, what is new here is the dimension of number of taxa affected and the sudden shift to an imaginary, non-existing species concept (BINs; ***(39)***), and the potential that such rather harmful approaches are copied by other scientists. A similar threat and a potential factor favoring a taxonomic impediment, many scientists see in the establishment of new taxa without using physical specimens (i.e., type specimens) at hand of photos or images ***(45, 46, 47)***, or environmental DNA ***(18, 48, 49)***.

Beyond the scope of this discussion, but to be mentioned to complete the overview of issues hampering taxonomic and nomenclatural stability, are by far more extreme cases which are known under the term "taxonomic vandalism". Many cases were clearly addressed as such (e.g., ***(50, 51, 52, 53)***), while other needed centuries to resurface (e.g., ***(54)***; see ***(55)***).

With the rise of DNA-based species delimitation in the last two decades, the above-mentioned challenges have been additionally complicated by a major weight of DNA-based approaches in the taxonomic treatment of species, particularly revealing often common patterns of cryptic biodiversity often revolutionizing the definition of widely accepted species. However, what matters here are not only the incongruent results with morphology-based species diagnoses but also the inconsistencies among DNA-based delimitations itself (Fig. 2) deriving from a multitude of molecular markers (Table 1), sampling issues, or different methodological approaches of species delimitation (not necessarily based on different species concepts) (Table 2) and/or their prior choices using the same data and species concepts (e.g., ***(33, 56, 57, 58, 59)***).

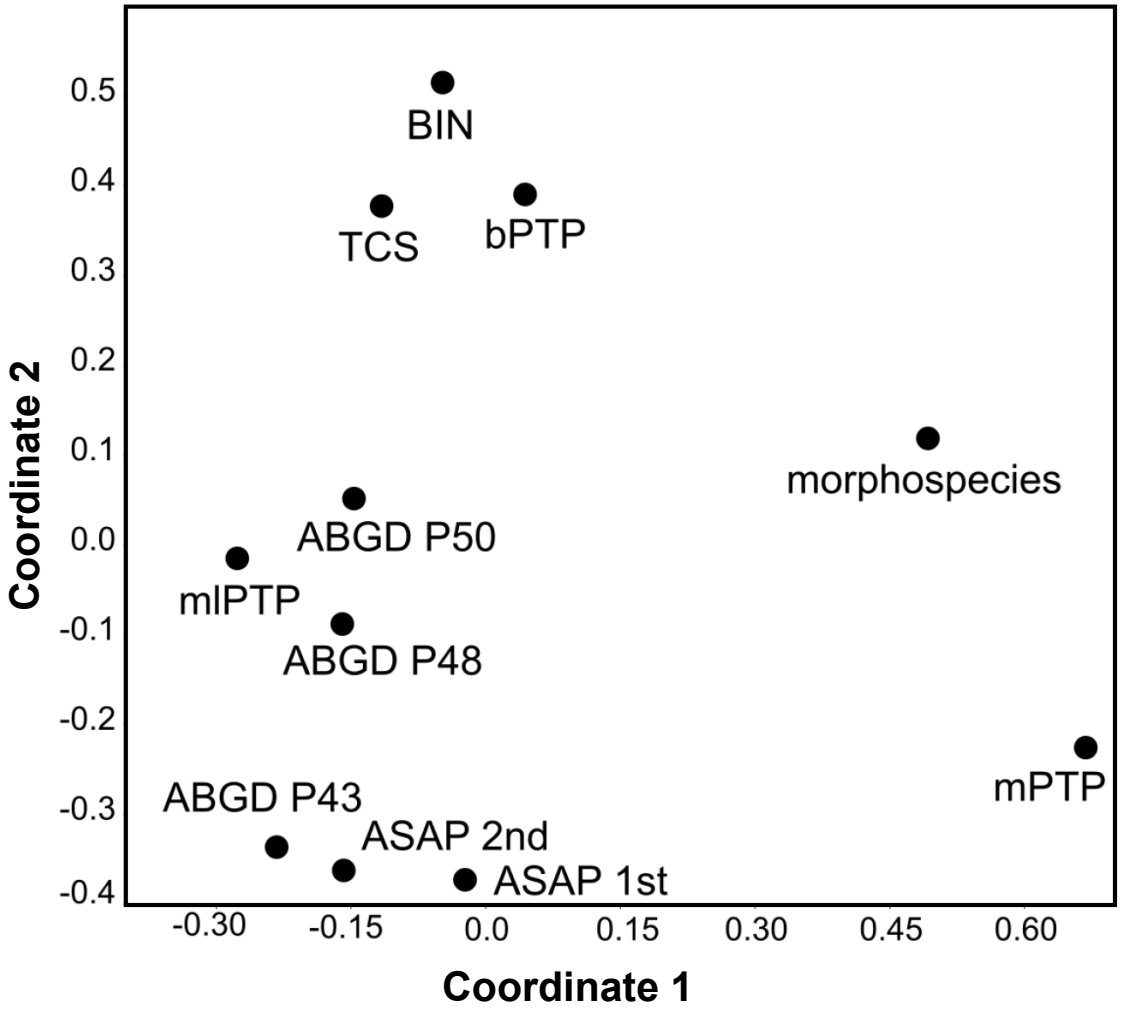

Fig. 2. Incongruence of results of various species delimitation approaches as graphically expressed by principal coordinate analysis (PCoA) of species delimitation results on pairwise match ratios ***(56)*** based on a study case of scarab chafer beetles ***(59)***.

3. Diagnosability

3.1. Bias for the objective diagnosability of species and the problems of the nature of species

With the integration of different, possibly competing species concepts ***(82)*** it became clear that in the process of speciation many of these concepts can be accommodated. An ideal delimitation method is thus not linked to any particular mechanism of reproductive isolation but based on detecting the ultimate outcome of speciation – i.e., the genetic isolation on an evolutionary timescale ***(83)***, or also morphological divergence as an expression of the former. Traditionally, species have been identified and described using morphological traits which is why in one or another form morphological characteristics for almost all species exist (Fig. 1). However, morphological characters may be subject to stasis, convergent evolution, or polymorphisms as they can be under similar selective pressure.
Morphology has other shortcomings:

1) when used in quantitative traits analysis of morphometric data, its multidimensional variation makes it difficult to objectively define distinctive entities (e.g., clustering) ***(57)***. Traits that differ between some species can be noise to diagnose others which is why taxonomists traditionally seek hierarchical comparisons (of only two or few taxa) in which the variation is distributed over a lesser number of dimensions (i.e., axes). Therefore, statistical testing for species differences needs an a priori assignment of species membership to be tested which may easily lead into circularity with a priori bias ***(58)***.
2) Morphological characterizations of taxa (diagnoses) are generally based on verbal descriptions or verbal coding of discrete character states. The more taxa are known, the more characters and character states become evident. Diagnoses of taxa then need to be updated to consider newly enclosed, new taxa, which is why morphology-based taxonomy needs updates by taxonomic revisions (the situation may be alleviated by more thorough digital imaging of specimens and the use of highly informative genital organs, for example in insects).

Both phenomena underlie probably also the phenomenon of the so-called “cryptic species”, as many species once distinguished by molecular characters can subsequently also be differentiated easily by morphology.

Instead, molecular genetic data are supposed to provide additional information about many factors related to species identification, including population identities, levels of recent or ancient gene flow, degree of hybridization, and phylogenetic relationships among prospective species ***(83)***. Empirical studies of DNA- and integrative taxonomy in the almost past two decades have demonstrated that a DNA-based species delimitation is very complex, too ***(5)***.

The oversimplified assumption that the “barcode gap” in a single mitochondrial marker will distinguish species ***(84, 85)***, has been recognized as flawed and methodologically outdated ***(8)***. Despite its apparent simplicity and increasingly low cost, mtDNA (e.g., COI) is not perfect for inferring species boundaries due to many issues: species delimitation and identification based on information from a single mitochondrial gene is prone to errors due to extrachromosomal inheritance and accordingly reduced rate of gene flow, little recombination, incomplete lineage sorting, sex-biased dispersal, asymmetrical introgression, and/or *Wolbachia*-mediated genetic sweeps ***(86, 87, 88, 89)***. Coalescence times are three to four times faster than in nuclear markers ***(90, 91)***. Its inaccuracy is also result of widely occurring mitochondrial paraphyly of species, which increases with geographical upscaling and more extensive sampling of the analyzed data ***(92)***, common sex-biased dispersal and thus biased patterns of COI divergence as well as ancestral polymorphism leading to false “cryptic” diversity (e.g., ***(58, 93)***). The species number inferred with COI data may in such cases exceed the true species number up to ten times ***(58)***, while outcome with different delimitation methods based on even the same marker and sampling can be consistently incongruent (Fig. 2) ***(33, 59)***, even if samples are not subject to geographical sampling bias ***(94)***. Consequently, divergence in mtDNA is not always stringently a result of speciation, particularly if it is not correlated with evidence from nuclear genomic DNA and/or morphology.

The outcome of species delimitation is affected by the choice of the method of species delimitation ***(33, 56, 65, 95)***, but also by factors which could be summarized as the nature of species and their assemblages such as (unbalanced) sampling (see ***(33)***), effective population size (of each species), fluctuation of effective population

size (within a group of taxa whose species limits are to infer), and depth of phylogenetic sampling ***(33, 56, 95)***. Singletons (i.e., species occurring with a single specimen only) do not impact the delimitation if they are not too abundant ***(56)***. Even highly sophisticated multispecies-coalescent approaches using multi-gene nuclear or genomic DNA data have been found to be highly sensitive to subjective initial prior choices for the analysis parameters (e.g., the population mutation rate (theta) is generally unknown; ***(57, 58)***) and result easily in over-splitting of true species ***(58, 96, 97, 98, 99)***. Subsequent correction with an empirics-based genealogical divergence index (gdi; ***(80)***) may work for some cases but is to some extent also biased (as thresholds for species boundary are not inferred directly from the data; ***(10)***) (showing a certain analogy to the barcode gap approach).
All in all, as highlighted already in ***(18)***, results of a DNA taxonomy (and not only that of DNA barcodes) need to be critically discussed to derive final conclusions. Depending on the case, the outcome may be yet subjective (i.e., if someone refrains to diagnose and name all over-split entities).

3.2 From species delimitation to species diagnosis

Independently from the molecular markers used, DNA-based species delimitation can be subdivided into validation methods (predefined entities are tested) and de-novo inference methods ***(68)*** (without a priori defined entities). Alternatively, we may distinguish character-based, distance-based, and tree-based methods ***(4, 60)***. Some need a calibration with known morphospecies, and are consequently to consider quite circular (e.g., AGBD, ASAP). With the time of doing empirical research using DNA taxonomy, it became quite evident that it is not difficult to diagnose divergent nucleotides in molecular data (e.g., between populations) but rather hard to infer the coherence of the divergent populations. This might have been the reason while population aggregation analysis and cladistic haplotype analysis, as character-based species inference, were almost completely abandoned in literature. What today methods of species delimitations offer as output are statistical likelihoods or probabilities of group membership of a number of certain specimens to a taxon, in which also these probabilities are subject to empirical thresholds ***(100)***. Consequently, contemporary methods used provide any substantial evidence or

basis for a species diagnosis which is not derived. Since Article 13.1.1 ***(34)*** states that a criterion of a name's availability is the presence of a "description or definition that states in words *characters* that are purported to differentiate the taxon", additional input/ information from the taxonomist would be required. If a simple group assignment based on such delimitation approaches' output shall fulfill the requirements of the Code (example: ***(28)***), the Code would need to be revised to implement such a thing. The potential ambiguity of considering simply mentioned base pairs as a species diagnosis was indicated by ***(18***), which is where lies the dilemma of the current Code ***(34)***. In consequence, either quite a number of names of the aforementioned example ***(28)*** (I stopped counting after some dozens) might result validly available, or even not.

Either way of interpretation, giving a diagnosis in this way is a nonsense (and fortunately it was only a short COI fragment and not part of a genomic USCO marker set composed of several thousands of nucleotide base pairs) as the statement of diagnosis is not logically comprehensible in terms of human language (this is what I suppose is the intended demand of Article 13.1.1; ***(34)***).[3]

Consequently, even more sophisticated cases that employ DNA-based species delimitation shall implement a meaningful character-based diagnosis for establishing a new animal species. And it becomes evident and logical, that meaningful can mean ***only*** a diagnosis based also on characters *and* morphology.

In contrast to that, according to ***(101)***, the International Code of Nomenclature for algae, fungi, and plants ***(102)*** allows descriptions[4] of new species exclusively based on DNA sequences, although the code interestingly does not explicitly specify that either (Art. 38.1, ***(102)***[5]).

3.3 A continued plea for integration[6]

[3] Based on that, I am afraid that the names in ***(28)*** lacking other diagnosis than base pair abbreviations would be invalid.

[4] According to ***(101)***, "A description gives the physical properties of the taxon, e.g. morphology, colour, odour, flowering time, chromosomes, chemical properties, or DNA sequence data"; but see also see ***(109)***.

[5] The term description is not defined in the glossary, in contrast to the term diagnosis ("a statement of that which in the opinion of its author distinguishes a taxon from other taxa (Art. 38.2); a diagnosis (or a description) is required for valid publication of a name of a new taxon (Art. 38.1(a)").

[6] - if not, there is a risk of more taxonomic impediment

From the above said follows that a meaningful implementation of DNA-based species delimitation (i.e., DNA taxonomy) implies an integrative taxonomy approach (***(103, 104)***; see also the majority of examples summarized in ***(29)***). Integrative taxonomy was introduced as a comprehensive framework to delimit and describe taxa by integrating information from different types of data and methodologies ***(22, 105, 106)***. Although there have been different concepts on how and to which degree to integrate data (e.g., ***(104, 107, 108)***), it was found that in simultaneous analyses always prevail splitting data ***(58)***.

However, computational power is currently the biggest problem to analyze large data sets of many taxa, specimens, and genes using for example the multispecies coalescent ***(74)***. This hampers some more robust testing for integrating different types of data or taxon sampling, but also imposes a huge time frame until data are ready for evaluation ***(10, Dietz et al, unpublished data)***. Analyses running for weeks is not really what we need to face biodiversity crisis.

Thus, take home message in this context: morphology is THE essential and likely most meaningful proxy for species diagnoses, even if species are circumscribed based on genomic evidence, no matter which other information is included in data analysis.

Continuously new approaches are suggested with more informative and better resolving data (e.g., ***(7, 10, 110)***). However, only, if these data cover sufficiently World's biodiversity (Fig. 1), we can properly employ them to discover the unknown. None of these methods can be exclusively taken not into account for so far accumulated data in regard to discovering new species due to their lack of connection with the other known species not investigated for this new trait.

In this context, although single marker DNA barcoding has been found to be imperfect in capturing true species boundaries, it would be unwise to abandon this well working approach for species identification and environmental monitoring, especially now that reference libraries appear to become more and more complete for at least some defined geographical areas.

The same applies for morphology, the only trait system for which data exists for (almost) all known species. Filling up species knowledge with new, more informative data types (e.g., WGS data, 3D images, MicroCTs, etc.) is expected to be complementary for so far rather than exhaustive (vision of many major natural history museums; ***(111)***).

Learning from the debate of past 30 years about the taxonomic impediment ***(112, 113, 114***), we may have understood its major causes and, hopefully, also how to avoid a further aggravation ***(18)***. A liberal interpretation and extension of the code(s) in the sense of ***(101)*** would allow the growth of incompatible trait sets that need continuous (taxonomic) "revision" to complete taxon sampling for one or the other trait. This would hardly allow us with the limited resources ***(113)*** to complete our knowledge about world's biodiversity. Just imagine, taxa barcoded only for the COI 3-prime end, would never be captured (i.e., match) with Hebert's COI barcode (i.e., the 5-prime end of COI; ***(1)***), only until the moment both fragments are captured. Mammals and frogs, which traditionally have been sequenced for other mtDNA markers (***(115, 116, 117),*** e.g., CytB and 16S, respectively, are partly less well covered in COI databases and risk to be captured less confidently if their marker is not especially addressed. Even worse, the use of AFLP or RAD sequences, whose recovery is often highly (i.e., study case) specific, would hardly allow reproducible data that are comparable with previously obtained data ***(7)***. A scenario of n-different parallel taxonomies (where n is the number of different traits being exclusively used for species diagnosis) was already predicted by ***(18)*** with reference to morphology and any potential DNA marker-based taxonomic system.
The only solution and chance to not renounce on modern high-throughput species recognition or highly informative genomic DNA approaches including all possible evidence (and traits) that allows to recognize new species, is to work integratively, *and* just to keep sticking with morphology. Agreement (see paragraph below) on at least one(some) reference system(s) will prevent us from wasting future resources.

4. Future perspectives

4.1 Good practices, codes, nomenclatural stability

The major goal of nomenclatural regulations ***(34, 102)*** is stability of each taxon's nomenclature. Given the experience from the past (see above), it is to question whether the approach of codes to not infer with issues of species delimitation and diagnosis is wise and yet timely since taxonomic uncertainty and taxonomic impediment are two sides of the same medal (see also: ***(40, 118)***. The

(mis)conception that taxonomy (i.e., species diagnoses and delimitation) and nomenclature/ nomenclatural stability are unlinked ("…types are name-bearers, not 'standards for species delimitation'…" ***(119)***) is perceived as an abnormality to everyone working with taxonomy, since taxonomic entities are generally marked with names. No nomenclatural stability without taxonomic stability. Neglecting these minimal common foundations and the complexity of taxonomy as a science (by non-emending the codes) is simply spoken comparable with a situation in which traffic rules from Linnean times or of 100 years ago (almost non-existing) are applied to today's traffic chaos. They do not entirely reflect what we know and what taxonomy needs.

This is particularly crucial, since taxonomy is experiencing a renaissance as hypothesis-based science, with DNA barcoding as its flagship (apart from its methodological pros and cons), after decades of neglect and shortage of funding. Its label of being not a science but a subjective, opinion driven field of descriptive biology is what drew off funding and made it unattractive for students to start with and to follow as a professional career. Instead, the success of DNA barcoding catalyzed the development and empirical exploration of new data systems (genomic data, 3D morphological data, and mass spectrometry, e.g., ***(7, 110, 120)***), computing power, and a manifold of methodological approaches for species identification and delimitation which has increased the capabilities to infer species diagnoses in many ways. All this fertilized ideas of boosting up the discovery of species and to integrate results from different lines of evidence ***(103, 105, 121)***, also with the aim to quantify competing species hypotheses and to simultaneously analyze different lines of evidence ***(77, 104)***.

The past 30 years of good practice agreements in biological sciences have shown that scientific rigor can be implemented in agreed (even nomenclatorial) rules, and in the long run, from which the entire scientific community benefits. This starts with proper citation of the published records (digital object identifier (DOI) system), species registration (Zoobank), the deposition of voucher specimens on which scientific results are based, the submission and sharing of data (BOLD, GBIF, NCBI, Dryad, etc.), peer-review, minimal data standards (BOLD). If we would have more such good practice agreements implemented with the Code(s), museums would not need to invest huge resources for the digitization of their collections (if done automatically upon publication of every taxonomic revision, in a similar way as

sequences are uploaded to NCBI for any molecular study published since the early 1990ies; ***(122)***), distribution data could flow into global data bases, and images could be collected in shared databases for immediate access and artificial intelligence applications.

Agreed standard trait sets would help to use more efficiently resources and most of all would avoid parallel taxonomies like those have existed in the old times when the taxonomists did not know the species of contemporary colleagues.

With more such agreements many resources could be used more efficiently. But of course, there must be acceptance in the different scientific communities, which also implies more interdisciplinarity, i.e., molecular biologists need to be involved more in taxonomy and species discovery, while taxonomists (including scientists being commissioners in the nomenclatural commissions) need to study and get training in modern methods of taxonomy. In other words, the entire community needs to move to a new level of understanding.

The current codes already include several such recommendations for a good scientific practice, so why not update them when our knowledge has improved to a level that we do understand that we need them. Defining species without a reference specimen (type) based on images or DNA sequences alone, should be possible but should remain an exception when justified. Such a reference specimen allows further data exploration and hypothesis verification, and thus allows for continuous improvement of scientific rigor. Codes do not need to justify poor scientific practice from the past (i.e., describing species based on drawings or images) keeping such rules alive (a cutoff date, as for the electronic publications or type specimen concept would do it very well).

And new technologies are at the doorstep for which we need to be ready, either to use them for our goals (rapid species discovery) using image analysis (e.g., ***(123, 124, 125, 126, 127)***), but also to keep species identification (particularly approaches that do not retain a reference specimen) clearly separate from species taxonomy and discovery. In this context, natural history museums need also to realize that virtual (or extended specimen) collections may at maximum enhance the use of the physical specimens but alone ***(111)*** they are of only limited use for uncovering the global biodiversity compared to the resources they consume since major taxonomic research and species diagnosis continue to be sustainable and scientifically sound only if specimen based.

4.2 Solutions and conclusions

Nomenclatural codes urgently need updates to be more explicit about species diagnoses and should impose for data compatibility and consequently for more sustainability to avoid waste of resources, parallel taxonomies, and new taxonomic impediment (standard trait reference systems, e.g., morphology, COI barcodes etc.) (see also ***(29, 128)***). Integration of taxonomic publications and (meta)data digitization/ databasing (collection records, images) would be crucial to implement too but it goes partly beyond the principal topic here and requires most of all generally more funding (storage, data management). Solutions should also consider the issues of exclusive access to funding and barcoding facilities which is seen by some researchers as a continuation of a not-inclusive policy ***(31, 33)*** and which could enforce more integrative than exclusive solutions. Therefore, morphology SHALL become a mandatory diagnostic trait in the ICZN and should obligatorily accompany any other diagnosis of a species being recognized by an alternative species concept. Different species concepts are reconcilable ***(82)***, and a hypothesis-based taxonomy should seek to perform this process of reconciliation through integrative approaches ***(30, 103, 121)***.

5. References

1. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–21
2. DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc Lond B Biol Sci 360(1462):1905-16
3. Vogler AP, Monaghan MT (2006) Recent advances in DNA taxonomy. J Zool Syst Evol Res 45:1–10
4. Goldstein PZ, DeSalle R (2010) Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. Bioessays 33:135–147
5. Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. Mol Ecol 22:4369–4383
6. Fontaneto D, Flot J-F, Tang CQ (2015) Guidelines for DNA taxonomy, with a focus on the meiofauna. Mar Biodiv 45:433–451

7. Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B (2020) A plea for implementing a standardized set of nuclear markers in DNA taxonomy. Trends Ecol Evol 35 (4):336–345
8. DeSalle R, Goldstein P (2019) Review and interpretation of trends in DNA barcoding. Front Ecol Evol 7(302):1-11
9. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21:2045–2050
10. Dietz L, Eberle J, Kukowka S, Mayer C, Bohacz C, Baur, H, Espeland M, Huber B, Hutter, C, Mengual X, Peters RS, Vences M, Wesener T, Willmott K, Misof B, Niehuis O, Ahrens D (2023) Standardized nuclear markers improve and homogenize species delimitation in Metazoa. Methods Ecol Evol 14:543–555
11. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends Ecol Evol 18:70–74
12. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol 55:595–609
13. Ahrens D, Monaghan MT & Vogler AP (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). Mol Phylog Evol 44:436–449
14. Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364
15. Rulik B, Eberle J, Jung M, Köhler F, Apfel W, Weigel A, Kopetz A, Köhler J, Hadulla K, Hartmann M, Fritzlar F, Schmidt J, Astrin J, Wägele W, Geiger M, Ahrens D (2017) Using taxonomic consistency with semi-automatized data preprocessing for high quality Barcode data. Methods Ecol Evol 8:1878–1887
16. Lucht WH (1987) Die Käfer Mitteleuropas. Katalog Goecke & Evers, Krefeld 12
17. Bleich O, Gütlich S, Köhler F (2023) Verzeichnis und Verbreitungsatlas der Käfer Deutschlands. http://www.colkat.de/en (last access Feb 12, 2023)
18. Ahrens D, Ahyong ST, Ballerio A, Barclay MVL, Eberle J, Espeland M, Huber BA, Mengual X, Pacheco TL, Peters RS, Rulik B, Vaz-de-Mello F, Wesener T, Krell F-T (2021) Is it time to describe new species without diagnoses? – A comment on Sharkey et al (2021). Zootaxa 5027(2):151–159
19. Polaszek A, et al (2005) A universal register for animal names. Nature 437:477
20. 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*. Proc Natl Acad Sci USA 101 (41):14812–14817
21. Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47-55
22. Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA Barcoding and the need for integrative taxonomy. Syst Biol 54(5):844–851
23. DeSalle R (2006) Species discovery versus species identification in DNA Barcoding efforts: response to Rubinoff. Conserv Biol 20(5):1545–1547

24. Packer L, Gibbs J, Sheffield C, Hanner R (2009) DNA barcoding and the mediocrity of morphology. Mol Ecol Res 9:42-50
25. Hołyński RB (2010) Taxonomy and the mediocrity of DNA barcoding – some remarks on Packer et al. 2009: DNA barcoding and the mediocrity of morphology. Arthropod Syst Phylogeny 68 (1):143–150
26. Brower AVZ (2010) Alleviating the taxonomic impediment of DNA barcoding and setting a bad precedent: names for ten species of '*Astraptes fulgerator*' (Lepidoptera: Hesperiidae: Eudaminae). with DNA-based diagnoses Syst Biodivers 8:485–491
27. Meierotto S, Sharkey MJ, Janzen DH, Hallwachs W, Hebert PDN, Chapman EG, Smith MA (2019) A revolutionary protocol to describe understudied hyperdiverse taxa and overcome the taxonomic impediment. Dtsch Entomol Z 66 (2):119–145
28. Sharkey MJ, Janzen DH, Hallwachs W, Chapman EG, Smith MA, Dapkey T, Brown A, Ratnasingham S, Naik S, Manjunath R, Perez K, Milton M, Hebert P, Shaw SR, Kittel RN, Solis MA, Metz MA, Goldstein PZ, Brown JW, Quicke DLJ, van Achterberg C, Brown BV, Burns JM (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
29. Renner, SS (2016) A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. Syst Biol 65:1085–1095
30. Zamani A, Dal Pos D, Fric ZF, Orfinger AB, Scherz MD, Sucháčková Bartoňová A, Gante HF (2022b) The future of zoological taxonomy is integrative, not minimalist. Syst Biodivers 20(1):1-14
31. Zamani A, Fric ZF, Gante HF, Hopkins T, Orfinger AB, Scherz MD et al (2022a) DNA barcodes on their own are not enough to describe a species. Syst Entomol 47(3):385– 389
32. Zamani A, Vahtera V, Sääksjärvi IE, Scherz MD (2021) The omission of critical data in the pursuit of 'revolutionary' methods to accelerate the description of species. Syst Entomol 46:1–4
33. Meier R, Blaimer BB, Buenaventura E, Hartop E, von Rintelen T, Srivathsan A, Yeo D (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 38:264-275
34. ICZN (1999) International Code of Zoological Nomenclature. 4th Edition. International Trust for Zoological Nomenclature, London, XXIX + 306 pp
35. Blanchard R, Maehrenthal F von, Stiles C W (1905) Règles internationales de la nomenclature zoologique adoptées par les Congrès Internationaux de Zoologie. International Rules of Zoological Nomenclature. Internationale Regeln der Zoologischen Nomenklatur – Paris (Rudeval)
36. Minelli A (2020) Taxonomy needs pluralism, but a controlled and manageable one. Megataxa 001 (1):9–18
37. Aleixo A (2023) On species concepts, species delimitation criteria, taxonomy committees, and biases: a response to Lima (2022a). Ornithol Res 31:62–70
38. Garnett ST, Christidis L (2017) Taxonomy anarchy hampers conservation. Nature 546(7656):25–27

39. Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PloS One 8:e66213
40. Garnett ST, Christidis L (2018) Science- based taxonomy still needs better governance: Response to Thomson et al. PloS Biol 16(3):e2005249
41. Raposo MA, Stopiglia R, Brito GRR, Bockmann FA, Kirwan GM, Gayon J, Dubois A (2017) What really hampers taxonomy and conservation? A riposte to Garnett and Christidis (2017). Zootaxa 4317 (1):179–184
42. McClure CJW et al (2020) Towards reconciliation of the four world bird lists: hotspots of disagreement in taxonomy of raptors. Proc R Soc B 287:20200683
43. Garnett ST, Christidis L, Conix S, Costello MJ, Zachos FE, Bánki OS, et al (2020) Principles for creating a single authoritative list of the world's species. PloS Biol 18(7):e3000736
44. Thomson SA, Pyle RL, Ahyong ST, Alonso-Zarazaga M, Ammirati J, Araya JF, et al (2018) Taxonomy based on science is necessary for global conservation. PLoS Biol 16(3):e2005075
45. Ceríaco LMP, Gutiérrez EE, Dubois A et al (2016) Photography-based taxonomy is inadequate, unnecessary, and potentially harmful for biological sciences. Zootaxa 4196 (3):435–445
46. Amorim DS et al (2016) Timeless standards for species delimitation. Zootaxa 4137 (1):121–128
47. Sluys R (2021) Attaching names to biological species: the use and value of type specimens in systematic zoology and natural history collections. Biol Theory 16:49–61
48. Hongsanan S, Jeewon R, Purahong W, Xie N, Liu J-K, Jayawardena RS, Ekanayaka AH, Dissanayake A, Raspé O, Hyde KD, Stadler M, Peršoh D (2018) Can we use environmental DNA as holotypes? Fungal Divers 92:1–30
49. Wu B, Hussain M, Zhang W, Stadler M, Liu X, Xiang M (2019) Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology 10 (3):127–140
50. Anonymous (2007) Vandalism in taxonomy. Koleopt Rundsch 77:38+60+88+146
51. Kaiser H, Crother BI, Kelly CMR, Luiselli L, O'Shea M, Ota H, Passos P, Schleip WD, Wüster W (2013) Best practices: In the 21st century, taxonomic decisions in Herpetology are acceptable only when supported by a body of evidence and published via peer-review. Herpetol Rev 44(1):8–23
52. Páll-Gergely B, Hunyadi A, Auffenberg K (2020) Taxonomic vandalism in malacology: comments on molluscan taxa recently described by N. N. Thach and colleagues (2014–2019). Folia Malacol 28(1):35-76
53. Wüster W, Thomson SA, O'Shea M, Kaiser H (2021) Review: Confronting taxonomic vandalism in biology: conscientious community self-organization can preserve nomenclatural stability. Biol J Linn Soc 133:645–670
54. Gistel JNFX (1857) Achthundert und zwanzig neue oder unbeschriebene wirbellose Thiere. p 513–606. In: Vacuna oder die Geheimnisse aus der organischen und leblosen Welt Unterdruckte Originalien-Sammlung von grösstenteils noch lebenden und verstorbenen Gelehrten aus dem Gebiete sämmtlicher Naturwissenschaften, der Medizin, Literaturgeschichte, des Forst- und Jagdwesen, der Oekonomie, Geschichte, Biographie, und der freien schönen

Künste, herausgegeben von Professor Dr Johannes Gistel Zweiter Band – Straubing: Schorner, 1031pp

55. Dellacasa G (1983) Sistematica e nomenclatura degli Aphodiini italiani (Coleoptera Scarabaeidae: Aphodiinae). Monogr Mus Reg Sci Nat 1:1- 464
56. Ahrens D, Fujisawa T, Krammer H-J, Eberle J, Fabrizi S, Vogler AP (2016) Rarity and incomplete sampling in DNA-based species delimitation. Syst Biol 65 (3):478-494
57. Eberle J, Warnock RCM, Ahrens D (2016) Bayesian species delimitation in *Pleophylla* chafers (Coleoptera)—the importance of prior choice and morphology. BMC Evol Biol 16:94
58. Eberle J, Fabrizi S, Bazzato E, Rossi M, Stella Columba M, Cillo D, Uliana M, Sparacio I, Sabatinelli G, Warnock RCM, Carpaneto GM, Ahrens D (2019) Sex-biased dispersal obscures species boundaries in integrative species delimitation approaches. Syst Biol 68:441–459
59. Ranasinghe UGSL, Eberle J, Thormann J, Bohacz C, Benjamin SP, Ahrens D (2022) Multiple species delimitation approaches with COI barcodes poorly fit each other and with morphospecies – an integrative taxonomy case of Sri Lankan Sericini chafers (Coleoptera: Scarabaeidae). Ecol Evol 12(5):e8942
60. Sites JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. Trends Ecol Evol 18:462–470
61. Davis JI, Nixon KC (1992) Populations, genetic variation, and the delimitation of phylogenetic species. Syst Biol 41:421–435
62. Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data. 3 Cladogram estimation. Genetics 132:619–633
63. Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. Mol Ecol 10:779–791
64. Brower AV (1999) Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. Syst Biol 48(1):199-213
65. Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. Syst Biol 62:707–724
66. Brown SD, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12(3):562-5
67. Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21:1864–1877
68. Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements, Bioinformatics 29: 2869–76
69. Kapli T, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T (2017) Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics 33(11):1630–1638
70. Puillandre N, Brouillet S, Achaz G (2021) ASAP: assemble species by automatic partitioning. Mol Ecol Res 21:609–620

71. Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
72. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19(9):1655-64
73. Yang ZH, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci USA 107(20):9264–9
74. Flouri T, Jiao X, Rannala B, Yang Z (2018) Species tree inference with BPP using genomic sequences and the multispecies coalescent. Mol Biol Evol 35:2585–2593
75. O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. Syst Biol 59:59–73
76. Grummer JA, Bryson RW, Reeder TW (2014) Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Syst Biol 63:119–133
77. Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492–507
78. Jones G, Aydin Z, Oxelman B (2015) DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31(7):991–998
79. Fujisawa T, Aswad A, Barraclough TG (2016) A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. Syst Biol 65:759-71
80. Jackson N, Carstens B, Morales A, O'Meara BC (2017) Species delimitation with gene flow. Syst Biol 66:799–812
81. Rabiee M, Mirarab S (2020) SODA: Multi-locus species delimitation using quartet frequencies. Bioinformatics 36:5623–5631
82. De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56 (6):879–886
83. Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. Mol Biol Evol 31(12):3125–3135
84. Meyer CP, Paulay G (2005) DNA Barcoding: error rates based on comprehensive sampling. PloS Biology 3 (12):e422
85. Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist?— a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers Zool 4:8
86. Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst 34:397–423
87. Ballard JW, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13:729–744
88. Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24:386–393
89. Klopfstein S, Kropf C, Baur H (2016) *Wolbachia* endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). Zool J Linn Soc 177:541–557
90. Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121:613–627

91. Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55 (5):859–868
92. Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, Geijer J, Herrmann J, Foster GN, Ribera I, Nilsson AN, Barraclough TG, Vogler AP (2012) The effect of geographical scale of sampling on DNA barcoding. Syst Biol 61:851–869
93. Ahrens D, Fabrizi S, Sipek P, Lago P (2013) Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. Mol Phylog Evol 69:83–94
94. Lukic D, Eberle J, Thormann J, Holzschuh C, Ahrens D (2021) Excluding spatial sampling bias does not eliminate over-splitting in species delimitation analyses. Ecol Evol 11(15):10327-10337
95. Esselstyn JA, Evans BJ, Sedlock JL, Anwarali Khan F a, Heaney LR (2012) Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. Proc R Soc B Biol Sci 279:3678–3686
96. Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. Proc Natl Acad Sci USA 114:1607–1612
97. Barley A J, Brown J M, Thomson RC (2018) Impact of model violations on the inference of species boundaries under the multispecies coalescent. Syst Biol 67:269–284
98. Chambers EA, Hillis DM (2020) The multispecies coalescent over-splits species in the case of geographically widespread taxa. Syst Biol 69:184–193
99. Leaché AD, Zhu T, Rannala B, Yang Z (2019) The spectre of too many species. Syst Biol 68:168–181
100. Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc R Soc B, Biol Sci 277:3071–3077
101. Turland N (2019) The Code decoded. Advanced Books. https://doiorg/103897/abe38075
102. Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Kusber W-H, Li D-Z, Marhold K, May TW, McNeill J, Monro AM, Prado J, Price MJ, Smith GF (eds) (2018) International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159 Glashütten: Koeltz Botanical Books DOI https://doiorg/1012705/Code2018
103. Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. Annu Rev Entomol 55:421–438
104. Yeates DK, Seago A, Nelson A, Cameron SL, Joseph L, Trueman JWH (2011) Integrative taxonomy, or iterative taxonomy? Syst Entomol 36:209–217
105. Dayrat B (2005) Towards integrative taxonomy. Biol J Linn Soc 85:407–415
106. Pante E, Schoelinck C, Puillandre N (2015) From integrative taxonomy to species description: one step beyond. Syst Biol 64(1):152–160
107. Miralles A, Vences M (2013) New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. PloS One 8(7):e68242

108. Edwards DL, Knowles LL (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? P Roy Soc B-Biol Sci 281 (1777):20132765
109. Hassemer G, Prado J, Baldini RM (2020) Diagnoses and descriptions in plant taxonomy: are we making proper use of them? Taxon 69:1-4
110. Raupach, MJ, Amann R, Wheeler Q, et al (2016) The application of "-omics" technologies for the classification and identification of animals. Org Divers Evol 16:1–12
111. Johnson KR, Owens IFP, Global Collection Group (2023) A global approach for natural history museum collections. Science 379(6638):1192-1194
112. Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: impediment or expedient? Science 303 (5656):285
113. Engel MS, et al (2021) The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. Zool J Linn Soc 193(2):381–387
114. Löbl I, Klausnitzer B, Hartmann M (2022) Das stille Aussterben von Arten und Taxonomen – ein Appell an Wissenschaftspolitik und Legislative. Entomol Nachr Ber 66(3):217-226
115. Vences M, Thomas M, Bonett RM, Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos Trans R Soc Lond B Biol Sci 360:1859–1868
116. Smith MA, Poyarkov NA Jr, Hebert PD (2008) DNA Barcoding: CO1 DNA barcoding amphibians: take the chance, meet the challenge. Mol Ecol Res 8(2):235-46
117. Ivanova NV, Clare EL, Borisenko AV (2012) DNA barcoding in mammals. Methods Mol Biol 858:153-82
118. Dubois, A (2011) The International Code of Zoological Nomenclature must be drastically improved before it is too late. Bionomina 2:1–104
119. Pape T, et al (2016) Species can be named from photos. Nature 537:307
120. Murugaiyan J, Roesler U (2017) MALDI-TOF MS Profiling-Advances in Species Identification of Pests, Parasites, and Vectors. Front Cell Infect Microbiol 15 (7):184
121. Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. Front Zool 7:16
122. Mizrachi I (2007) GenBank: The Nucleotide Sequence Database. 2002 Oct 9 [Updated 2007 Aug 22]. In: McEntyre J, Ostell J (editors) The NCBI Handbook [Internet] Bethesda (MD): National Center for Biotechnology Information (US); 2002- Chapter 1. Available from: https://wwwncbinlmnihgov/books/NBK21105/
123. Buschbacher K, Ahrens D, Espeland M, Steinhage V (2019) Image-based species identification of wild bees using convolutional neural networks. Ecol Inform 55:101017
124. Klasen M, Ahrens D, Eberle J, Steinhage V (2022) Image-based automated species identification: Can virtual data augmentation overcome problems of insufficient sampling? Syst Biol 71 (2):320–333
125. Borowiec ML, Dikow RB, Frandsen PB, McKeeken A, Valentini G, White AE (2022) Deep learning as a tool for ecology and evolution. Methods Ecol Evol 13:1640– 1660
126. van Klink R, August T, Bas Y, Bodesheim P, Bonn A, Fossøy F, Høye TT, Jongejans E, Menz MHM, Miraldo A, Roslin T, Roy HE, Ruczyński I, Schigel D, Schäffler L, Sheard JK,

Svenningsen C, Tschan GF, Wäldchen J, Zizka VMA, Åström J, Bowler DE (2022) Emerging technologies revolutionise insect ecology and monitoring. Trends Ecol Evol 37 (10):872-885

127. Fujisawa T, Noguerales V, Meramveliotakis E, Papadopoulou A, Vogler AP (2023) Image-based taxonomic classification of bulk insect biodiversity samples using deep learning and domain adaptation. Syst Entomol: 1– 15. https://doiorg/101111/syen12583
128. Orr MC, Feijó A, Chesters D, Vogler AP, Bossert S, Ferrari RR, Costello MJ, Hughes AC, Krogmann L, Ascher JS, Zhou X, Li D-Z, Bai M, Chen J, Ge D, Luo A, Qiao G, Williams PH, Zhang A, Ma K, Zhang F, Zhu C-D (2022) Six steps for building a technological knowledge base for future taxonomic work. Natl Sci Rev 9 (12):nwac284

Table 1. Common markers in modern DNA taxonomy (from ***(7)***, modified) (wg = whole genome).

	Origin of marker	Taxonomic presence/ number of markers	Copy number	Variability	Orthology assessment	Alignment	Data generation
COI/ Cyt b	mitochondrial	universally present; normally one variant	many	high, but often insufficient to discern closely related taxa	easy to homologize; NUMTs can pose problems	easy	targeted PCR; genome skimming, wg sequencing
ribosomal DNA	nuclear (18S, 28S) or mitochondrial (16S)	universally present; few markers	many	In many groups resolution insufficient in closely related species; ITS more variable	homologizable	difficult, numerous insertions/deletions or ITS hamper comparison across broad taxon sampling	targeted PCR/ DNA enrichment, genome skimming; wg sequencing
RADs	nuclear	not universal; many markers	unknown	unpredictable, but typically high	difficult or impossible across distantly related species	difficult or impossible across distantly related species	highly protocol specific
UCEs	nuclear	wide universal presence	almost all single-copy	nearly none; the flanking regions are analyzed, whose degree of variability is unpredictable	flanking regions bordered normally by only one homologizable ultra-conserved element (orthology of flanking regions questionable)	flanking regions can be difficult	target DNA enrichment; genome skimming; wg sequencing
metazoan level USCOs	nuclear	near universal presence of 978 genes in Metazoa	single-copy in > 90 % of species	variability of coding sequence sections sufficiently high	Ortho DB (Introns usually excluded)	easy	target DNA enrichment; genome skimming; wg sequencing

Table 2. Popular methods and approaches of modern DNA taxonomy (for details see also ***(6, 60)***).

Method	Main references	Species inference	Inference basis	Implied monophyly
single locus				
Population aggregation analysis	***(61)***	validation	character	yes
Statistical parsimony analysis (TCS)	***(62, 63)***	de novo	character	no
Cladistic haplotype analysis	***(64)***	validation	tree	yes
Generalized mixed yule coalescent (GMYC)*	***(12, 65)***	de novo	tree	yes
Distance-based clustering (e.g., SPIDER)	***(66)***	de novo	distance	no
Automatic barcode gap discovery (AGBD)	***(67)***	de novo	distance	no
Poisson tree process modeling (PTP)*	***(68, 69)***	de novo	tree	yes
Barcode index number (BIN) system	***(39)***	de novo	distance	no
Assemble species by automatic partitioning (ASAP)	***(70)***	de novo	distance	no
multi locus				
Population structure analysis (Structure)	***(71)***	validation	character	no
Population structure analysis (Admixture)	***(72)***	validation	character	no
Bayesian phylogenetics & phylogeography method (BPP)	***(73, 74)***	validation	tree	yes
Brownie	***(75)***		tree	yes
Bayes factor delimitation (BFD)	***(76)***	validation	tree	yes
Integrated BPP (iBPP)	***(77)***	validation	tree	yes
DISSECT	***(78)***	de novo	tree	yes
Trinomial distribution of triplets model (tr2)	***(79)***	de novo	tree	yes
BPP combined with genealogical divergence index (gdi)	***(80)***	validation	tree + distance	yes
Multi-locus species delimitation using quartet frequencies (SODA)	***(81)***	de novo	tree	yes