# Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description

*Paul Z. Goldstein and Rob DeSalle**

**DNA barcodes, like traditional sources of taxonomic information, are potentially powerful heuristics in the identification of described species but require mindful analytical interpretation. The role of DNA barcoding in generating hypotheses of new taxa in need of formal taxonomic treatment is discussed, and it is emphasized that the recursive process of character evaluation is both necessary and best served by understanding the empirical mechanics of the discovery process. These undertakings carry enormous ramifications not only for the translation of DNA sequence data into taxonomic information but also for our comprehension of the magnitude of species diversity and its disappearance. This paper examines the potential strengths and pitfalls of integrating DNA sequence data, specifically in the form of DNA barcodes as they are currently generated and analyzed, with taxonomic practice.**

**Keywords:**
DNA barcoding; DNA taxonomy; species diagnosis; taxonomic impediment

DOI 10.1002/bies.201000036

Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA

***Corresponding author:**
Rob DeSalle
E-mail: desalle@amnh.org

**Abbreviations:**
**COI**; cytochrome oxidase subunit I; **M-OTU**; molecular operational taxonomic unit.

## Taxonomy in transition?

### Taxonomic impediments

In recent decades systematists have embraced revolutionary innovations on multiple technological fronts, ranging from the most basic tools for collecting and archiving specimens, to web-enhanced databasing endeavors, photo-microscopy, imaging, and analytical software and of course high-throughput DNA sequencing. Despite these advances, the Linnaean system and its concomitant nomenclatural codes remain fundamentally intact and successful as an informatics system summarizing observed natural hierarchy and maintaining a paper trail of primary literature devoted to description, classification, and revision.

Nevertheless, a bottleneck exists in the facility with which species new to science are described and specimens identified, and it has been tempting to attribute this bottleneck to some flaw in the science of taxonomy or to the strictures of the nomenclatural codes themselves rather than to the magnitude of the Earth's biodiversity relative to available expertise. The number of recognizable species on Earth remains controversial to within an order of magnitude (cf. [1]), and the obvious relevance of systematics to conservation efforts [2], highlighted by the recognition that the Earth's species are being extinguished more quickly than they can be discovered and described, has been underappreciated [3, 4]. Charged with the responsibilities of biological discovery and description, the dissemination of fundamental biological information, and the stewardship of collections that serve as the world's biological libraries, systematists have an overwhelming challenge, and are confronted with equally overwhelming requests to provide other biologists with the service of identifying organisms for them. Notwithstanding the broader aims of biological systematics, the rest of the world feels an immediate need for rapid, accurate taxonomic determinations, and the term ''taxonomic impediment'' [5] has been used variously to refer to the bottlenecks in time, manpower, training, availability, and publication constraining the rapid description of biological diversity and the ease with which biological determinations can be made by non-specialists [4, 6, 7].

### Enter DNA barcoding

Beginning in 2003, drawing on decades of advances in the field of molecular systematics, the initiative to ''barcode'' life on Earth was set out in a series of papers by Hebert et al. [8–11]. Exploiting the utility of molecular data to pinpoint cryptic species diversity (e.g. [12]), the DNA barcoding endeavor proposed to generate high-throughput DNA sequence data for multiple individuals from as many populations of as many species as possible with the stated goals of enhancing the availability of systematic output, relieving the burden of identification on taxonomists, pairing life stages of conspecific biological samples (e.g. [13, 14]), and ''provid[ing] a bio-literacy tool for the general public'' ([15] p. 1806). The idea – and the appeal – of DNA barcoding was to take a step towards a day when, in the words of Janzen et al. ([16] p. 24), ''there will be a hand-held barcorder. . .for everyone to use as their linkage between the wild world and what humanity knows about it''. DNA barcoding has been widely embraced for its hopeful efficacy in both the identification of biological specimens and the discovery of species; in short, for its potential to call attention to, if not ameliorate, the taxonomic impediment. Indeed, one of the primary selling points of DNA barcoding has been its promise as a taxonomic research tool as well as a basic identification tool. With growing comprehension of the magnitude, severity, and immediacy of the biodiversity crisis, and based on the enthusiastic responses from various segments of the scientific community (e.g. ecologists and conservation biologists) and from federal agencies (e.g. [17]), one would think, upon a superficial gaze at the recent literature, that DNA barcoding represented a novel, powerful, and empirically sound tool for the identification of specimens and the discovery of species.

Almost simultaneously came the suggestion of ''DNA taxonomy'' [18], a proposal – formally independent of the barcoding endeavor – to renovate the longstanding Linnaean system by establishing a taxonomic framework based on DNA sequence data. Both the practice of DNA barcoding and the notion of DNA taxonomy raised concerns among many systematists and evolutionary biologists for reasons of a practical (e.g. [19]), empirical (e.g. [20, 21]), philosophical (e.g. [8]), and economic (e.g. [22]) nature. The ensuing debate has been highly charged and at times characterized by selective scholarship and orthogonal argumentation. This debate does not need to be dissected in full here; suffice it to say that strong concerns have been raised surrounding the empirical approaches associated with DNA barcode data and their potential to impede rather than enhance the practice of taxonomy and the dissemination of reliable taxonomic information [23]. Although DNA barcoding has shown promise in many contexts, its practice has also raised concerns directed at its exclusive reliance on mitochondrial DNA [24–27]. Notwithstanding these issues, critiques of some of DNA barcoding's most obvious methodological shortcomings have either been redirected (e.g. [28]) or simply gone unanswered (e.g. [29]).

### DNA barcoding is not DNA taxonomy

Unfortunately, while some proponents of DNA barcoding have overlapped with those of DNA taxonomy in arguing against Linnaean taxonomy, the frequent linkage of the two groups in the literature has clouded a number of important issues. Fueled by increasing frustration at ''the'' taxonomic impediment, these concerns parallel criticisms of the standards and practices of systematics itself in the form of assertions that taxonomic practice is inherently too slow to manage the biodiversity crisis and that Linnaean taxonomy is outmoded, unstable, pedantic, counter-intellectual, or otherwise in retrograde.

Although some of the criticisms subsequently leveled against DNA taxonomy apply to certain barcoding applications, not all the shortcomings are shared, and although barcode data intersect potentially with current taxonomic approaches, it must be clarified that DNA barcoding is promoted as an investigative tool, whereas DNA taxonomy is intended as an overhaul of Linnaean classification and nomenclature. The justification of DNA barcoding relies on the consensus that species determination is too unwieldy to be undertaken by non-specialists and on the supposition that phenograms based on short sections of mitochondrial DNA provide a legitimate shortcut by which anyone can reliably identify a specimen without recourse to literature or scientific training. DNA taxonomy, in sharp contrast, represents a genuine paradigm shift, an up-ending of the Linnaean system and its replacement with a series of unique monikers linked to multiple gene sequences intended to form the ''scaffold'' of post-modern systematics [18, 30–32]. The proponents of DNA taxonomy, to their credit, do not promote the singular power of any gene, endorsing more exhaustive analyses of multiple genes and an as yet inchoately articulated process by which ''taxa'' are denoted by sequences themselves; in short a substitute process for species description that circumvents and replaces the standards of binominal nomenclature but does not establish strong criteria for testing taxonomic hypotheses or for maintaining nomenclatural stability or priority [33, 34].

In this paper we confine ourselves first to the empirical and operational aspects of DNA barcoding as it is most commonly practiced and address mechanical pitfalls shared by DNA barcoding, DNA taxonomy, and, potentially, certain biological nomenclatural codes. Our purpose is to clarify the concerns surrounding the use of DNA barcodes and suggest ways to ensure that some measure of their value may be rescued. Since DNA barcoding is not a taxonomic method per se, it is not our intention to paint a picture of competing, mutually exclusive paradigms between DNA barcoding and modern taxonomy, rather to elaborate the means by which results of DNA barcoding may dovetail successfully with taxonomic practice. It is our contention that DNA barcoding realizes its potential in the realm of what has come to be referred to as integrative taxonomy [35–37], and that the degree to which barcode data are taken to represent taxonomic determinations is dictated by the same rules of nomenclature that apply to all forms of data in systematics.

## Operational mechanics and the deceptive simplicity of DNA barcoding

The goals of DNA barcoding are deceptively simple: a DNA barcode – most commonly the ~650 base pair stretch of the

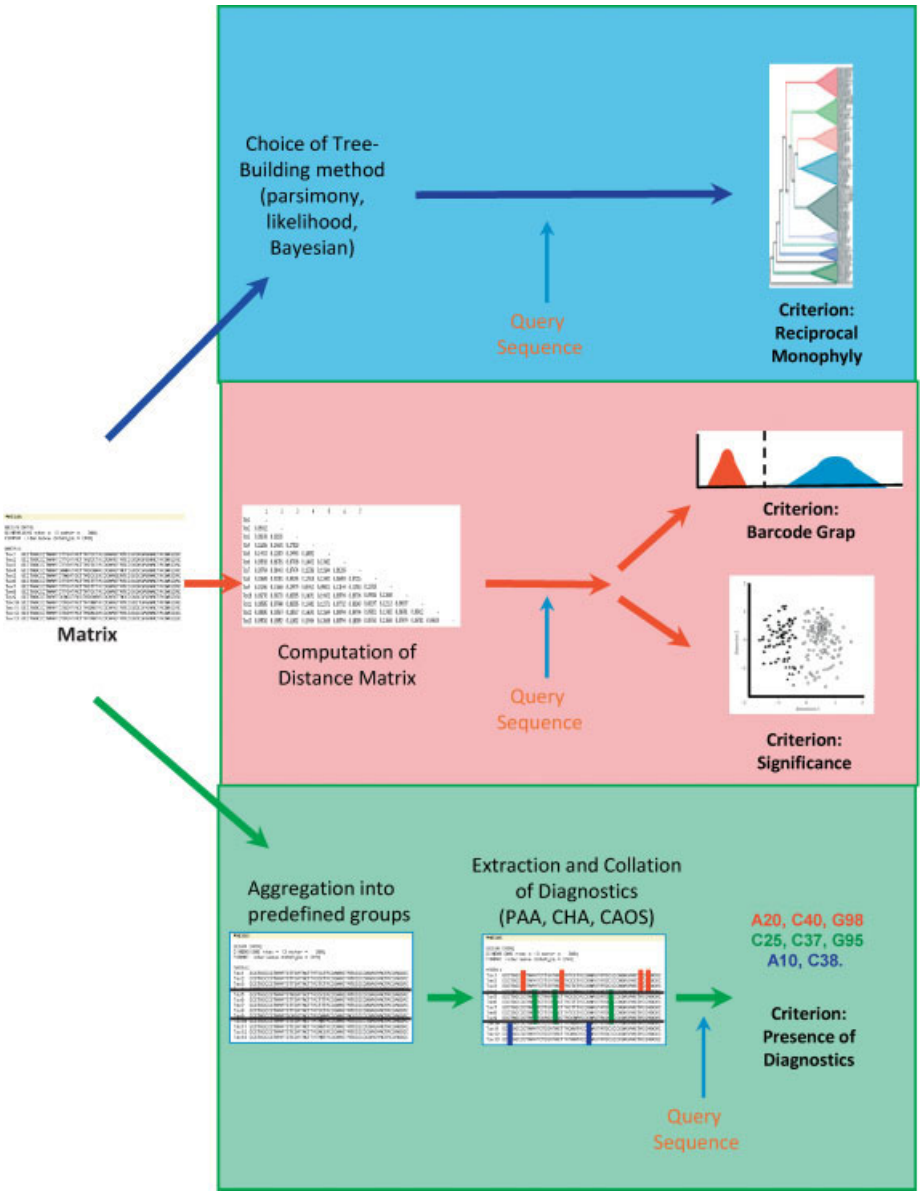


**Figure 1.** Operational pathways of barcoding analyses, proceeding from the compilation of a DNA sequence matrix to putative species determination (identification). In the bottom (green) pathway sequences are aggregated on the basis of shared characters, and diagnostics are extracted and collated. The red, green, and blue bars in the diagram indicate diagnostics for the first, second, and third species, respectively. Finally, a query sequence (in orange) from an unidentified sample is examined for diagnostics, which, if present, are used to affix a species determination. The red A20, C40, G98, blue C25, C37, G95, and green A10, C38 refer to positions in the sequence (the number) and the state of the position (A, G, C, or T) that are diagnostic for the three species in the matrix. In the top (blue) and middle (red) pathways, the diagnostic step is effectively skipped, and thereby the operations of delimitation and determination are conflated: in the top, tree-based pathway, a query sequence (in orange) is analyzed via a tree-generating method, and the topology of the resulting tree is used to ''locate'' the query sequence and affix a determination based on the graphically defined criterion of reciprocal monophyly (although simple adjacency is commonly used); the middle (red) pathway employs statistical methods (not necessarily accompanied by tree graphics) to identify species based either on the criterion of fit to a pre-designated barcode gap or on the criterion of statistical significance. Note that the top two approaches are not mutually exclusive.

mitochondrial cytochrome oxidase subunit I or COI for animals – is intended to serve as a provisional identifier of a biological specimen or sample (Fig. 1). To be sure, the identification of biological samples using DNA has been in practice for over two decades, starting with the forensic applications of DNA sequences pioneered by Jeffreys et al. [38] and the identification of bacterial strains developed by Woese [39]. What the neologism of DNA barcoding proffered was the standardization of gene region(s), a pedestrian analysis for purposes of identification, and (at least potentially) a centralized resource of frozen tissues and DNA samples.

Barcode sequences are commonly analyzed with a distance algorithm and individual samples are assigned taxonomically according to their proximity to other samples as measured by percent sequence divergence or visualized on a tree, most commonly generated by neighbor-joining algorithm [40]. The apparent appeal of neighbor-joining is that it is fast, computationally simple, and definitive, always retrieving a single tree regardless of ambiguity in the data. DNA barcode data are generally organized by submission to one or more databases such as GenBank or the Barcode of Life Database (BoLD; [41]), and specimens from these studies archived. In addition to being employed as nomenclatural arbiters outside formal taxonomic treatment, DNA barcodes can also be treated as sources of character information, i.e. as potential components of formal species descriptions within formal taxonomic treatment, and as potential evidence for higher taxonomic assignment or re-assignment (e.g. genus level or family level). Perhaps most intriguingly, barcode data are presented as an actual means of discovery of cryptic species by virtue of their revealing hidden variation congruent with other sources of data [42–44].

Methods for analyzing DNA sequences for purposes of taxonomic assignment may be examined according to how

## Box 1

### Identifying species using barcodes

Approaches to delimiting species using molecular data (cf. [99]) range from using diagnostic characters or combinations of characters to distance methods, coalescent approaches (e.g. Yule models of stochastic lineage growth of Pons et al. [100]) and likelihood ratio tests [101]. Statistical approaches for delimiting species are as varied and controversial as species concepts and criteria themselves [99, 102–104]. Although long abandoned by phylogenetic systematists for their statistical shortcomings, distance methods [105] have been resurrected in species delimitation methods used most commonly in DNA barcoding studies.

The first step in any DNA barcoding analysis is the construction of a matrix of DNA sequences, which may be used directly either as a series of character states or converted to a matrix of genetic distances. The outcome can then be subjected to a tree-building algorithm, in the first case to generate a cladogram summarizing shared character states among terminals, and in the second case to generate a phenogram summarizing distances among terminals. Alternatively, the matrix can be used without tree-generating graphical approaches to visualize clusters or discover structure.

***Tree-based approaches:*** Trees may be constructed, as in strict phylogenetic studies, using character-based approaches such as parsimony or maximum likelihood, but most studies featuring DNA barcoding generate a distance matrix and subject it to a tree building operation, commonly the neighbor-joining algorithm. If the tree-generating algorithm is phenetic (distance-based), the phenogram represents a summary of divergence percentages; if it is character-based, the cladogram represents a summary of character state distributions. If a tree-generating algorithm is used, the tree itself is taken to represent both a statement of proximity of a given specimen to other (named) biological entities and a graphical tool with which to delimit species boundaries. Implicit in all tree-based criteria for delimiting species is some notion of reciprocal monophyly (but see [106]) – a graphical or topological criterion that neither of two sister lineages be visually nested within the other – regardless of what method is used to generate the tree. Although the term is rarely used in the barcoding literature, reciprocal monophyly, as it was originally articulated in the context of phylogeography, remains the most widely recognized (if not practiced) topographical criterion formalized for the purpose of species delimitation.

Independent of tree-building but often presented alongside trees, many have invoked a distance cutoff as an indicator of differentiation [8, 30, 107]. This has been dubbed the ''DNA barcode gap'', below which two individuals are deemed conspecific and above which they are not [60, 61]. Essentially a section of unoccupied range computed from the primary matrix of pairwise distances, the barcode gap describes the idealized scenario in which the distribution of all conspecific pairwise distances and the distribution of all (congeneric) heterospecific pairwise distances do not overlap; in other words, the span between the largest conspecific pairwise distance and the smallest congeneric heterospecific pairwise distance, assuming it exists and is a positive number. As Meier et al. [60] point out, many published papers incorrectly define the ''barcode gap'' as the difference between the means of intraspecific versus interspecific distances, often going so far as to use all interspecific distances, and not just congeneric interspecific distances, which is even more misleading.

***Tree-less distance methods:*** Alternatively, sequences may be analyzed with statistical methods to evaluate distributions of character states. Some approaches utilize vector methods for visualizing data [108]. Studies focused on understanding the process of speciation often involve coalescent approaches in the generation of trees, but these are generally beyond the purview of barcoding for species identification and tend to utilize multiple loci. Analysis of molecular variance (AMOVA), multivariate analysis, and the molecular operational taxonomic unit (M-OTU) approach of Blaxter and coworkers [30, 109, 110] employ distance-based clustering methods to identify putative taxonomic entities based on DNA sequences and do not employ trees, although they are often accompanied by phenograms.

***Molecular operational taxonomic units:*** M-OTUs are aggregations of sequences united by some index of similarity (usually a predefined distance cutoff) intended to correspond to biological reality (''phylotypes'' or ''genospecies''; [30]). Originally developed for nematodes [107], a group with the same kinds of taxonomic overload problems such as fungi and bacteria, M-OTUs were developed to be unobtrusive with respect to existing nomenclature. The M-OTU approach places flagged entities into a taxonomic ''bin'' for the purposes of evaluating ecological diversity and census studies. Any entity in this system can ''graduate'' to the formal taxonomic system when it has been described formally.

***Diagnostics:*** A final methodological class of approaches eschews tree-based criteria for making taxonomic statements [46, 50, 53] taking into account the fact that the relationships among interbreeding organisms are not accurately presented as necessarily hierarchic, nor species as necessarily monophyletic. These include population aggregation analysis (PAA) of Davis and Nixon [111], which defines the minimal terminals of a cladogram as aggregates of individuals sharing fixed traits (characters), and cladistic haplotype analysis or CHA [112], which addresses homoplasy among suites of diagnostic characters, as well as the methods of Sarkar et al. [113], DeSalle et al. [50], and Little and Stevenson [46]. These methods are not encumbered by reliance on distances and are amenable to incorporation with other forms of data (e.g. morphological, behavioral) in phylogenetic analyses. An important, additional advantage of tree-less diagnostics is that they present a straightforward sense of the statistical reliability of all assessments of similarity, and do not return a positive result if no matching diagnostics are found (unlike routine barcode analyses or, for that matter, the BLAST algorithm, which always returns a ''closest'' match but requires the user to evaluate its reliability). This is an important reason why distance-reliant database querying can be misinterpreted.

the sequences are compared analytically, how those comparisons are visualized, and which criteria are used to differentiate among taxa (Box 1). Each step in the identification and discovery process – querying a sequence library or a sequence against a ''backbone'' tree, applying an a priori criterion for species delimitation, flagging novel sequences as potentially new species (discovery), and using barcode data as part of a formal taxonomic description – carries specific empirical requirements and epistemological ramifications. Unfortunately, while the empirical goals of DNA barcoding (identification and discovery) are explicit, its empirical rationale has not been so clearly articulated as it has been asserted.

The first and most immediate utility of DNA barcoding – the procedure of species identification (determination) portrayed in the vision of hand-held ''barcorders'' – is deceptively straightforward because the determination of an organism to species necessarily relies on existing species descriptions and enumerations of diagnostic criteria. As such, the process of identification is both comparative and empirical by nature: for a DNA barcode to function as an identifier, there must exist (i) an archived library of sequences reliably determined to species; (ii) a standardized means of comparing sequences; (iii) a justifiable and generalizable criterion for the delimitation of species; and (iv) species-specific diagnostic criteria, in other words criteria or thresholds for accepting or refuting a given taxonomic conjecture, and hence for the attribution of a specific nomenclatural epithet. Although specimen identification using DNA barcode data necessarily depends on retrievable species circumscriptions and diagnoses, the operations of species discovery and specimen identification are commonly muddled by the use of methods to simultaneously diagnose, determine, and potentially discover species.

There are at least two primary operational concerns surrounding the interpretation of DNA barcode data. The first is the use of the tree graphic itself as a criterion for species recognition (see Box 1). Certainly a branching diagram is visually satisfying: it conveys proximity, describing overall genetic similarity or summarizing character state distributions depending on the analysis used to generate it. But quantifying graphical proximity or evaluating species membership of a given ''branch'' based on its being adjacent to a clade of conspecific sequences is necessarily arbitrary [45, 46]. This is not a methodological argument per se, but a straightforward outcome of graph theory.

The most widely embraced graphical criterion for species delimitation is that of reciprocal monophyly (Fig. 1), which originates in the phylogeography literature but has not been widely taken up in the barcoding literature (barcoding proponents emphasize that barcoding is not intended for phylogenetic, historical, or evolutionary reconstruction even though they employ trees to infer species identity). Branching diagrams – however resolved – are also imperfect descriptors of organismal relationships below the species level, regardless of one's definition of species [47]. In part, the issue is one of shoe-horning organisms into graphical depictions of nested relationships where none may exist. Even if a given branch is nested within a group of conspecific branches, this is neither a necessary nor sufficient criterion for inclusion in that species since species need not be monophyletic. The fact that many species are demonstrably ''paraphyletic'' [48] is a necessary outcome of macroevolution, not merely an obstacle to the use of trees in the arbitration of species limits. Terms such as monophyly and paraphyly have implications beyond the graphical, and were not intended to apply below the species level.

Thus, the visualization of data as a tree may appear to suggest kinship but may actually compromise identification and circumscription at the species level [45], a methodological fracture potentially compounded by the second operational concern: reliance on distance metrics themselves as criteria for species delimitation. The use of distance data to generate trees results in the visual presentation of what amount to uninformative nodes that may belie relatedness. Although advocates of DNA barcoding have, to their credit, made it explicit that barcode analyses are not intended to be phylogenetic [49], the shortcomings of distance-reliant methods are not limited to phylogenetics. Tree-based methods and distance-based methods are neither mutually exclusive nor inextricably linked logically, but are commonly applied jointly in the barcoding literature.

Long recognized by phylogeneticists, the drawbacks of distance methods have been addressed, at least in a general way, in the context of DNA barcoding [45, 46, 50, 51], as they have in the context of species delimitation broadly [52]. Rather than reiterate these treatments, suffice it to say that distances do not enable the extraction of precise characters in the service of actual species identification; distance data can not by definition be diagnostic [52]. At best they indicate numerical proximity (whether or not visualized as a tree), always returning positive results (a nearest neighbor) even when no conspecific exist. Since DNA barcodes are intended to make precise taxonomic statements, it must be recognized that ''clusters at a 75% jackknife'', for example, is not such a statement, whereas ''diagnosed at positions COI-72(A), COI-198(G), COI-675(T)...etc.'' is. Moreover, neither pairwise nor patristic distances can resolve contradictory circumscriptions; they can not arbitrate conflict, serve as tests, or provide anything but ''arbitrary and capricious'' criteria for species delimitation [2, 46]. In short, relying on distance methods makes it difficult for the scientist to distinguish among identifying specimens of described species, flagging specimens of undescribed or cryptic species, and linking sequence data to taxonomically valid names via formal description (Figs. 1 and 2). When, on the other hand, characters – be they base pairs, behaviors, or morphological features – are explicitly tied to the delimitation of species, these operations are more readily parsed: groups of organisms can be surveyed for characters that differentiate them (''distinguishing marks and scars'') and queried with novel sequences [21, 46, 53] such that hypotheses of conspecificity are corroborated or refuted.

## Species discovery and taxonomic assignment are independent

The data generated by DNA barcoding and the degree to which they complement existing taxonomic information vary with the scale and purpose of the study (Box 2), but barcode data commonly reveal structure among clusters of sampled individuals and raise questions of whether such clusters represent

## Box 2

**Major kinds of DNA barcoding studies**

***Cottage Industry studies:*** In what can be thought of as cottage industry endeavors, the first cohesive barcoding studies to emerge focused on small groups or co-occurring groups of closely related organisms and, in addition to the general goal of facilitating the identification of specimens, were designed to tease out or corroborate cryptic species that were already suspected. A premier example of such a study is the oft-cited paper by Hebert et al. [10], which used DNA barcoding to support the recognition of ten cryptic species hiding under the nomenclatural epithet of *Astraptes fulgerator* (Lepidoptera: Hesperiidae). Hebert et al. suggested that ten species of skippers otherwise indistinguishable as adults were readily parsed with their barcodes. Although their analyses were challenged in part by Brower [96], who concluded that no more than seven species could be thus delineated based on the available molecular data, their results were consistent with observed differences in larval morphology and host plant associations, although it was suggested that the cryptic species were ''revealed'' by barcoding.

***Taxon-focused industrial scale studies:*** As the Consortium for the Barcode of Life (CBoL) was established and quickly gained momentum, initiatives grew in scope and more ambitious, industrial scale studies emerged that required the coordination offered by CBoL (http://www.barcoding.si.edu/). These were delimited taxonomically, concentrating on large groups of organisms such as fish (FishBOL [114]; http://www.fishbol.org/), birds (AllBirds; http://www.barcodingbirds.org/), and land plants [115]. The larger taxonomically comprehensive studies are, of course, capable of generating hypotheses of previously unknown species. Taxonomically focused sub-projects within the Census of Marine Life (CoML) that have an explicit barcoding goal include the International Census of Marine Microbes (ICoMM) and Census of Marine Zooplankton (CMarZ).

***Biogeographically focused industrial scale studies:*** Some taxonomically focused industrial scale studies have targeted one or more groups of organisms occupying a given area (e.g. Moorea [116]) or ecosystem [e.g. MarBOL (http://www.marinebarcoding.org/)] or both, such as North American birds [117], Australian fishes [118], and tropical Lepidoptera [94, 119]. One goal of such studies is to generate a resource for species identification; another is to gauge cryptic diversity of an understudied fauna. The Mosquito Barcoding Initiative (MBI) states (http://www.mooreabiocode.org/) that their objective is to construct ''a library of genetic markers and physical identifiers for every species of plant, animal and fungi on the island of Moorea''. CoML includes biogeographically and ecologically focused projects, such as Marine Ecology (MAR-ECO; http://www.mar-eco.no), Arctic Ocean Diversity (ArcOD; http://www.coml.org/projects/arctic-diversity-arcod), the Census of Coral Reef Ecosystems (Creefs; http://www.creefs.org), and the Census of Antarctic Marine Life (CAML; http://www.caml.aq).

***Program-focused industrial scale studies:*** Many DNA barcoding initiatives have arisen that are focused on a research goal external to taxonomy per se or directed by an initiative such as conservation (Barcoding for Endangered Species Conservation; BESC), disease reservoir research (Mosquito Barcoding Initiative; MBI) or agricultural pest research [International Network for Barcoding Invasive and Pest Species; INBIPS (http://barcoding.si.edu/INBIPS.htm) and Tephritid Barcode Initiative; TBI (http://connect.barcodeoflife.net/profiles/blogs/project-tephritid-barcode)].

discrete entities meriting formal description. Ambiguously assigned sequences or clusters of sequences are ''flagged'' as species potentially new to science, and DNA barcoding has suggested high levels of apparent crypsis in taxonomically varied organisms (e.g. [54–56]). If results are to be taken at face value, traditional estimates of species richness in many groups may represent significant underestimates.

The operations commonly undertaken in barcode studies to identify or ''flag'' entities for taxonomic attention are largely the same as those undertaken for purposes of species determination: failure of a given cluster of sequences to meet a topographical or distance criterion results in the hypothesis of a novel entity or would-be species (Fig. 2). In their roles as both taxonomic arbiters and means of putative discovery, DNA barcodes rely heavily on congruence with other ''genetically independent'' observations ([44] p. 77; see also [43] p. 151). As such a common path of inference proceeds simply by inspecting a tree topology to see whether or not groups of organisms hypothesized to be conspecific based on criteria such as habitat, morphology, life history do in fact ''tree'' together. As discussed earlier, simple inspection of trees generated by barcode data to evaluate species assignment is problematic for several reasons, the most obvious of which arises when species are incompletely sampled [45], resulting in an apparent but inaccurate grouping with the most similar ''available'' sequences. Authors have also suggested variously that DNA barcode gaps or other quantities suffice to distinguish species, and that a criterion of interspecific variability (for example, one ten times greater than intraspecific variability; [11]) is necessary and sufficient to draw conclusions of heterospecificity (see Box 1). These criteria have met with several related obstacles, the most compelling of which is that even if a barcode gap exists (in some cases it has been shown empirically not to [57, 58]), its magnitude is far from consistent. Speaking once again to the reliance on distance measures, high intraspecific variability demonstrably confounds the utility of barcoding in numerous groups [59–63] and over broad biological parameter space [64]. Burns et al. ([42] p. 138), attribute the practice of arbitrary distance-based cutoffs to taxonomists rather than barcoders, concluding that ''clearly

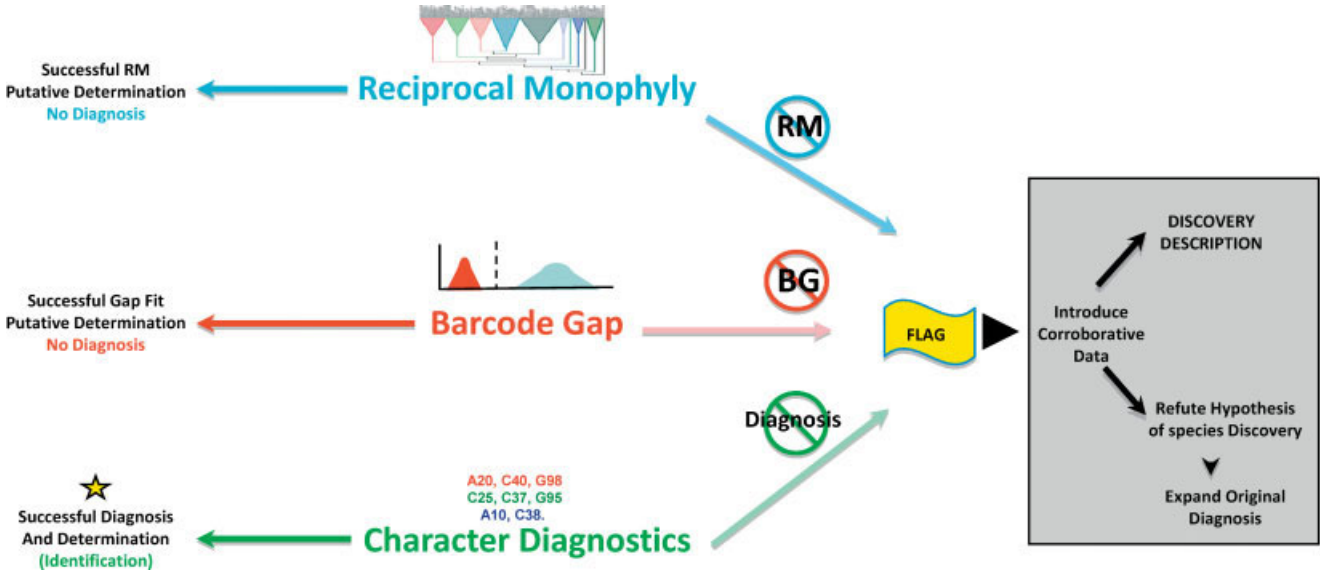


**Figure 2.** Possible results of approaches outlined in previous figure (colors correspond). Arrows pointing left indicate what constitutes positive identification under each framework: the finding of reciprocal monophyly (RM; blue) leads to a putative determination of a query sequence but no diagnosis. The presence of a barcode gap (BG; red) renders a positive result as well, but again no diagnosis. Successful match of character diagnostic(s) (Diagnosis; green) leads to successful determination of the query sequence. Failure to determine a given sequence under the RM or BG criteria requires that the sequence be flagged (yellow) as potentially new, and the procedure is reiterated upon the addition of molecular data. But even in the event of meeting one or both the RM and BG criteria no precise diagnoses emerge, and the species determination must be regarded as tentative – one made by virtue of graphical or mathematical proximity rather than observational hypothesis testing. If there is failure to retrieve diagnostics from the query sequence (light green line) then the specimen is flagged and either tested with addition of data (molecular, morphological, behavioral, etc.) or the original diagnostics are expanded to include the query sequence. The output determination from a successfully diagnosed sequence is accompanied by diagnostic characters amenable to species description.

the chronic practice of various taxonomists of setting arbitrary levels of differentiation for delimiting species is unrealistic'' and that ([42] pp. 146, 152) ''the designation of some percentage or degree of divergence as a point which individuals should be considered conspecific is unrealistic (even though many taxonomists have done so, in various contexts, for a great many years)''. These valid characterizations of distance-based criteria actually apply with greater strength to common barcoding practices, including those employed by Burns et al., than to proper taxonomic practice.

## Species description is a critical bottleneck

The practical nexus of DNA barcoding, DNA taxonomy, and systematics is that of formal description. Although the operations used by barcoders for delimiting described species and discovering new ones are essentially the same, the discovery of new species ultimately requires taxonomic treatment. DNA barcoding studies have informally addressed taxonomic problems (e.g. [65]; see also reference list at http://www.dnabarcodes.org/publications), and barcode data are increasingly used loosely to corroborate nomenclatural changes [66]. But most studies featuring barcodes have yet to realize that potential in formal taxonomic revision, suggesting instead that ''formal taxonomic work remains to be done'' ([67] p. 4775). Ignoring the nomenclatural problem by assigning informal names to undescribed species relegates them to taxonomic orphanhood: the most obvious problem with nameless species designations being that they tend to get lost in the shuffle. Unlike DNA taxonomy, DNA barcoding is promoted as intended to work within the Linnaean system. Taxonomic nomenclature is governed by several bodies (e.g. International Union of Microbiological Societies, 1992, http://www.iums.org/; International Commission on Zoological Nomenclature, 1999, http://www.iczn.org/iczn/index.jsp; International Botanical Congress, 2000, http://www.bgbm.fu-berlin.de/iapt/nomenclature/code/SaintLouis/0000St.Luistitle.htm), and acceptable criteria for species delimitation are both controversial and varied [68–70], as is the treatment of molecular data by taxonomic authors (Box 3). Bacterial taxonomy, for example, appears to rely on a combination of character-based data and distances: DNA-DNA hybridization temperature (the technique that most classically lends itself to distance methods) has been the ''gold standard'' for the description of new species of bacteria, but newer more genome-oriented approaches are evolving [71–73]. Molecular data are not explicitly treated by the zoological or botanical codes, but both contain provisions for such data to be used in species description. Not every ''new'' entity posited on the basis of barcode data need be formally described upon its first published recognition, but broad claims of vast troves of newly discovered species should be considered tentative until corroborated formally by additional data. Some degree of low-level variation, however structured, is best left outside the provenance of systematics proper. The important practical questions surround the curation of DNA type specimens by museums and herbaria and what to do with potentially new

## Box 3

**DNA barcode information in taxonomic description**

To date molecular data have been included in formal taxonomic publications in various and unstandardized ways. Although a growing number of taxonomic publications incorporate DNA sequences in species descriptions, only recently have there appeared barcoding papers that formally address taxonomic repercussions. Meanwhile, taxonomic authors vary in their treatment of molecular data, which may appear in the species description itself or in the diagnosis, depending on the author's preference. Below we provide examples from the entomological literature to demonstrate a range of approaches.

***Simple list of GenBank accessions***: Many taxonomic studies simply list the GenBank Accession number of the new species in the species description. No further discussion of the DNA sequence data is made in the description and the DNA sequence information is treated as an ''add on''.

***Inclusion of the DNA barcode sequence***: Burns et al. [42] include the entire barcode sequence in the description without a GenBank accession number:

> ''Type specimens: Holotype: male (Figs. 14, 26, 45, 46 [see arrow]), voucher code 06-SRNP-31674 (Janzen & Hallwachs 2006), Sendero Memos, Sector Pitilla, Area de Conservación Guanacaste, Costa Rica, 740 m, latitude10.98171, longitude −85.42785. Deposition: National Museum of Natural History, Smithsonian Institution (USNM). Labeled (yellow): LEGS AWAY/FOR DNA. DNA barcode (658 bp) of holotype (coded MHAHH575-06|06-SRNP-31674|Neoxeniades pluviasilva): AACTTTATAC. . .[fill in 638 bases]. . . ACATTTATTT.''

***Presentation of raw distance measures***: Mengual and Thompson [35, 120] allude to raw distance measures following the description of diagnostics and a morphological key:

> ''The uncorrected pairwise distance between the two species of *Palpada* for the COI gene was 2.80%, and the alignment by eye of the ITS2 sequences required 4 changes and 3 indels. On the other hand, the sequences of 28S gene for *P. prietorum* specimens XP117 and XP118 were identical, and they differed only 0.16% from the sequence of XP104, *P. ruficeps*. These results are in congruence with the expected mutational rate of the different gene fragments.''

***Qualitative description of DNA sequence information***: Some descriptions comment on DNA sequence variation without enumerating precise characters or diagnostics. Hastings et al. ([121] p. 36) state that morphological traits are used in species diagnosis and that the barcodes ''support'' these inferences:

> ''The new key to New World *Sphecius* (Holliday & Coelho 2006) uses morphological traits and geographic distributions to distinguish among five named species of *Sphecius*, including the four species, *S. convallis*, *S. grandis*, *S. spectabilis*, and *S. speciosus*, that were part of this study. The [DNA] barcodes support identification of *S. spectabilis* and *S. grandis* as species distinctly different [sic] from *S. speciosus* and *S. convallis.*''

In this case there is no attempt to extract diagnostics at the DNA sequence level for the description, which contradicts nomenclatural recommendations.

***Hybrid approaches***: Some studies verge on detailing the sequence diagnostics, but fall back on distance-based inferences. Fochetti and Tierno de Figueroa ([122] p. 45; http://www.sciencedaily.com/releases/2009/07/090715101507.htm) summarize diagnostic characters and present distance data in their description of several species of *Tyrrhenoleuctra*. The description of these species contains several ''phylogenetic'' representations and tables of genetic distances of the newly described taxa from one another:

> ''. . .the combined aligned dataset of all analyzed *Tyrrhenoleuctra* comprised 1,660 nucleotide positions (12S rRNA: 344; COI: 1316). Of these, 1,149 positions were constant, 282 (219, excluding gaps) variable positions were parsimony-uninformative and 229 (227, excluding gaps) variable positions were parsimony-informative. A pairwise distance calculation (MEGA 4.1; Kumar et al. 2008) using the COI sequence data (Table 1; sequence data not shown, available on request from first author) shows that *T. antoninoi* is on average more than 10% distant from the other species. Genetic variation of intraspecific populations ranged from 0.002 to 0.011 with populations of *T. zavattarii* and from 0.000 to 0.020 and 0.010 for populations of *T. tangerina* and *Tyrrhenoleuctra* sp. C, respectively.''

There is an interesting twist to this particular description in that the barcode sequences are touted as having utility in two areas:

> ''The results of this study provide evidence of the utility of COI barcoding. Despite the limitations of the method as a sole means of species identification and delineation, it has generated two new lines of inquiry: (i) the status of the two barcode groups of *S. grandis*, and (ii) the relationship between *S. convallis* and *S. speciosus*. Future work to resolve these two questions will follow several lines of investigation, including analysis of nuclear genes, closer examination of morphological variation, and field studies of the behavioral ecology of sympatric populations of these wasps.''

In other words the DNA barcodes are seen as useful in ''flagging'' specific questions for future research. The DNA barcodes generate hypotheses that need to be tested using other approaches and more information.

Schmidt [123], in an otherwise fine and thoughtful revision of the genus *Grammia* (Lepidoptera: Noctuidae) attributed the failure of barcodes to resolve species to ''extensive gene tree nonmonophyly'', but proceeded to present a

**Box 3 (*continued*)**

distance tree and included both barcode reference numbers and maximum percent differences in species descriptions.

***Presentation of a branching diagram***: Apart from presenting the sequences themselves, commonly only by their GenBank accession numbers, a popular way of presenting these data is in the form of a distance tree (e.g. [42]). Õunap and Viidalepp [124], while acknowledging the limitations of barcode data, use tree-based species delimitation and do not elucidate molecular characters even in description.

***Reporting diagnostic information***: In some very rare cases DNA barcoding information is analyzed diagnostically. Fisher and Smith [125] successfully integrated barcode data into a morphological taxonomic revision of the Malagasy species of *Anachetus* (Hymenoptera: Formicidae) in which two new species are described. They present ''diagnostic barcoding loci'' as follows:

''*A. boltoni:* ATCT-42-45 & RTTAR-66-70''

In noting diagnostic barcode data as part of taxonomic description and alluding to sequence data from multiple ribosomal genetic markers exclusive of the barcoded COI, Fisher and Smith provide a compelling example and a cogent discussion of integrating barcode data with other biological information in taxonomic revision. They also include raw distance data (percent divergence) in the species description and present neighbor-joining phylogenetic analysis trees, but it is unclear what these are intended to add to the taxonomic content of their paper.

taxa that can only be identified on the basis of DNA sequence information.

Given the rapid generation of molecular data and hypotheses of undescribed cryptic species, there is genuine concern that available taxonomic expertise may be insufficient to describe the world's biota in the foreseeable future, and that science awaits a backlog of unnamed species even more enormous than previously imagined. An important challenge is the archiving of molecular data for taxonomic treatment, while rendering it available for research. Following examples by various authors who invoked some form of interim taxonomy (e.g. [16, 74]), Schindel and Miller ([75] p. 111) have advocated a system, amenable to DNA barcode data, involving the use of ''standardized taxon labels as intermediate products in the process of producing new taxonomic names''. Fungal systematists have developed Emerencia (http://emerencia.math. chalmers.se), a bioinformatics tool to monitor the ''identity of fungal ITS sequences whose taxonomic annotations are poorly resolved'' ([76] p. 178). Since much of the recent work characterizing fungal diversity begins with environmental sampling using the ITS region, an effective source of DNA barcode information, researchers have adopted this approach to accommodate undescribed species.

Mechanisms along the lines advocated by Nilsson et al. [76] for fungal taxonomy are appealing for their capacity to accelerate the treatment of undetermined sequences, and should be developed further. La Salle et al. [77] suggest a need to automate the collection and analysis of morphological character information to implement species discovery and description, predicting that ''[d]escriptions of species boundaries will be web-based, and dynamically updated with digitization of new specimens which add to our understanding of phenotypic variability within a species. Taxonomic products will incorporate real-time accrual of changes and improvements, including new species as they are discovered and described'' ([77] p. 51). Although we do not promote the notion that taxonomy will or should one day be automated, automating species identification has gained momentum [78–81] in large measure from the barcoding endeavor. As barcode data are sure to be used in formal taxonomy with growing frequency, standards for doing so are desirable.

## DNA barcoding and the opportunity for integrative taxonomy

Ideally, taxonomic works incorporate all available data (morphological, molecular, behavioral, biogeographic, ecological) and certain of the established codes of nomenclature are intended to accommodate diverse sources of data, if not explicitly molecular. One of the first suggestions that taxonomic descriptive systems might incorporate macromolecular information was that of Doyen and Slobodchikoff [82]. Protein electrophoresis had already been introduced to the study of evolution [83, 84] and chromosomal data had been used in botanical systems for decades. Doyen and Slobodchikoff relied heavily on reproductive isolation (hence a biological species concept) as a criterion in species delimitation, but nevertheless recognized the importance of incorporating diverse information with taxonomy. The combination of molecular and morphological data in phylogenetic studies became a hotly debated topic in the 1990s, when phylogeneticists argued variously for the superiority of molecules to morphology, the superiority of morphology to molecules, the use of data to test one another in separate versus combined analyses, and so forth. Meanwhile the use of behavioral characters became a growing focus of the discussion, and studies began to appear that demonstrated the relationship between empirical rigor and the simultaneous analysis of combined data [85].

Since the introduction of DNA barcoding, authors have responded variously by emphasizing the importance of contextualizing barcode data within a broader scientific endeavor [3, 86, 87], stressing the intellectual content and critical nature of taxonomy [3, 19, 33, 88], and decrying ''one-dimensional systematics'' [89], and have collectively endorsed the practice of ''integrative taxonomy'' [35, 36, 46, 51, 61, 65, 87, 90–93], suggesting that DNA barcode data are most successfully used as a means of complementing other sources of information. The term ''integrative taxonomy'' was birthed as a result of

taxonomists' realizing not only the opportunity to incorporate all relevant information into a species description, but also that doing so strengthens both the empirical foundations of systematics and the Linnaean framework itself. Proponents of barcoding have recognized its potential contribution to systematics [16, 94] but appear to resist suggestions that established barcoding analyses be refined.

Packer et al. [43, 95] suggested that barcodes be "used as a first approximation to delimit taxa for which variation within species makes it difficult to discern subtle signal from morphological noise. In these cases, deep divergences indicate a lack of genetic cohesion among reproductively isolated taxa for which morphological differentiation has not yet arisen, or has not developed sufficiently for easy recognition using traditional methods." Probably without intending to, Packer et al. [17, 95] resurrected the "morphology versus molecules" debate by interpreting the early reaction of taxonomists to DNA barcoding as defensive, threatened, even panic-stricken. Packer et al. ([17] p. 1098) asserted that DNA barcoding is "generally far superior and clearly outperforms morphology in analysis of difficult species complexes". Their characterization of the "mediocrity of morphology" served to perpetuate the perception that DNA barcoding is fundamentally at odds with traditional taxonomic practice.

But barcoding has been most compellingly presented when used in concert with other data (e.g. [90, 92, 93]) as it was in perhaps the best-known early barcoding study, the *Astraptes* work of Hebert et al. [10], which suggested that ten cryptic species of skipper butterflies were hiding under one nomenclatural umbrella. Although reanalysis of their data suggested fewer species [96], and although all the putatively cryptic species were discernable by inspecting their conspicuously variable caterpillars, the larval morphology data behaved synergistically with host plant association and barcode data in arranging congruent clusters of species. Similar examples have since been presented for other hesperiids (e.g. [42–44]). From the outset, both barcoders and systematists appeared to agree that multiple sources of data are desirable as tests, at least in the very broad sense. What remains is to reconcile the precise mechanics of barcoding analysis with the empirical and philosophical rigor of systematics.

## Summary and future challenge

DNA barcoding is a potentially powerful heuristic in identifying specimens of described species and generating hypotheses of new ones, but its power is amplified when its limitations are recognized and its utility not overstated or over-sold in the name of novelty. At issue are neither preferred phylogenetic inference methods, species concepts, nor the superiority of one class of data over another, but rather how explicitly the analytical operations themselves are linked to the empirical steps of identifying, discovering, and describing species. Many, if not most users of taxonomic information may not know or even care about the language of nomenclatural bylaws, the mechanics of phylogenetic inference, or the nuances of species concepts and criteria. But for reasons both numerous and obvious, there is a certain immediacy to the need for taxonomic information, and in many circles the need for fast, definitive answers to taxonomic questions often trumps the reality of painstaking effort required to retrieve accurate ones reliably.

The deceptive simplicity of barcoding has perhaps enabled a somewhat shallow interpretation of the endeavors of systematics. The danger in such a view is a trade-off between the advertised ease of generating barcode data and the actual rigor of treating the data empirically: instead of fueling synergy, the barcode paradigm presents as philosophically and analytically chimeric. Although Hebert et al. ([8] p. 313) assure us that "taxonomic expertise is collapsing" and that they are "convinced that the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon 'barcodes'", Packer et al. [95] advise systematists not to "panic", and Hebert and Barrett [28] insist with seeming optimism that synergy exists between barcoders and practicing systematists, as do Gregory [97] and Janzen et al. [16, 98]. With well over 1,000 publications involving barcoding, it is not premature to ask how the barcoding endeavor may better dovetail with taxonomy. We suggest that the powers of barcoding can best be realized by recognizing the intellectual content (*sensu* Lipscomb et al.; [88]) of taxonomy and at least rendering barcoding analyses more amenable to taxonomic practice. Clearly barcoding endeavors are promising and (apparently) often successful, but greater success might be realized with some simple modifications, such as more careful, character-based analyses amenable to direct empirical testing from non-molecular data sources. The legacy of DNA barcoding has the potential to extend far beyond a database of short sequences, towards a bank of genomic DNA, and the analytical tools exist to accommodate data that will surely exceed short snippets of organellar sequence.

The growing number of DNA barcoding endeavors puts a fine point on the immediacy of the biodiversity crisis and the severity of the taxonomic impediment by suggesting the existence of hidden diversity of an order not widely anticipated. The enormity of cryptic species diversity is daunting, and its documentation demands creative approaches. Unsurprisingly, molecular data have begun to illuminate patterns and structure to that diversity, challenging the scientific community to absorb an accelerating torrent of information without eroding the ontological foundations of the system we use to understand life on Earth. Just as it remains a challenge in barcoding initiatives to substantiate findings with taxonomic ground truth, it is incumbent on the systematic community to accommodate molecular data and to decide on levels of variation worthy of formal taxonomic treatment. The empirical challenge is to streamline the formal taxonomic process without resorting to shortcuts that allow data – and species – to slip through the cracks.

## References


1. **Hamilton AJ**, **Basset Y**, **Benke KK**, **Grimbacher PS**, et al. 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. *Am Nat* **176**: 90–5.
2. **Mace GM.** 2004. The role of taxonomy in species conservation. *Philos Trans R Soc Lond B* **359**: 711–9.
3. **Wheeler QD.** 2005. Losing the plot: DNA "barcodes" and taxonomy. *Cladistics* **21**: 405–7.
4. **de Carvalho MR**, **Bockmann FA**, **Amorim DS**, **Devivo M**, et al. 2005. Revisiting the taxonomic impediment. *Science* **307**: 353–4.

5. **Taylor RW.** 1983. Descriptive taxonomy: past, present, and future. In Highley E, Taylor RW, eds; *Australian Systematic Entomology: A Bicentenary Perspective*. Melbourne: CSIRO. p 93–134.
6. **Giangrande A.** 2003. Biodiversity, conservation, and the 'Taxonomic impediment.' *Aquatic Conserv: Mar Freshw Ecosyst* **13**: 451–9.
7. **de Carvalho MR**, **Bockmann FA**, **Amorim DS**, **Brandão CRF**, et al. 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evol Biol* **34**: 140–3.
8. **Hebert PDN**, **Cywinska A**, **Ball SL**, **deWaard JR.** 2003. Biological identifications through DNA barcodes. *Proc R Soc Lond B* **270**: 313–21.
9. **Hebert PDN**, **Ratnasingham S**, **deWaard JR.** 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc B* **270**: 596–9.
10. **Hebert PDN**, **Penton EH**, **Burns J**, **Janzen DH**, et al. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly, *Astraptes fulgerator*. *Proc Natl Acad Sci USA* **101**: 14812–7.
11. **Hebert PDN**, **Stoeckle MY**, **Zemlak TS**, **Francis CM.** 2004. Identification of birds through DNA barcodes. *Plos Biol* **2**: 1657–63.
12. **Brower AVZ.** 1996. A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, as revealed by cladistic analysis of mitochondrial DNA sequences. *Zool J Linn Soc* **116**: 317–32.
13. **Miller KB**, **Alarie Y**, **Wolfe GW**, **Whiting MF.** 2005. Association of insect life stages using DNA sequences: the larvae of *Philodytes umbrinus* (Motschulsky) (Coleoptera: Dytiscidae). *Syst Entomol* **30**: 499–509.
14. **Webb KE**, **Barnes DKA**, **Clark MS**, **Bowden DA.** 2006. DNA barcoding: a molecular tool to identify Antarctic marine larvae. *Deep Sea Res Part II* **53**: 1053–60.
15. **Savolainen V**, **Cowan RS**, **Vogler AP**, **Roderick GK**, et al. 2005. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos Trans R Soc B* **360**: 1805–11.
16. **Janzen DH**, **Hallwachs W**, **Blandin P**, **Burns JM**, et al. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour* **9**: 1–26.
17. **Packer L**, **Grixti JC**, **Roughley RE**, **Hanner R.** 2009. The status of taxonomy in Canada and the impact of DNA barcoding. *Can J Zool* **87**: 1097–1110.
18. **Tautz D**, **Arctander P**, **Minelli A**, **Thomas RH**, et al. 2003. A plea for DNA taxonomy. *Trends Ecol Evol* **18**: 70–4.
19. **Wheeler QD.** 2004. Taxonomic triage and the poverty of phylogeny. *Philos Trans R Soc Lond B* **359**: 571–83.
20. **Moritz C**, **Cicero C.** 2004. DNA barcoding: promise and pitfalls. *Plos Biol* **2**: 1529–31.
21. **DeSalle R.** 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conserv Biol* **20**: 1545–7.
22. **Cameron S**, **Rubinoff D**, **Will K.** 2006. Who will actually use DNA barcoding and what will it cost? *Syst Biol* **55**: 844–7.
23. **Ebach MC**, **Holdrege C.** 2005. DNA barcoding is no substitute for taxonomy. *Nature* **434**: 697–8.
24. **Ballard JW**, **Whitlock MC.** 2004. The incomplete history of mitochondria. *Mol Ecol* **13**: 729–44.
25. **Rubinoff D**, **Holland BS.** 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst Biol* **54**: 952–61.
26. **Schmidt BC**, **Sperling FAH.** 2008. Widespread decoupling of mtDNA variation and species integrity in *Grammia* tiger moths (Lepoptera: Noctuidae). *Syst Entomol* **33**: 613–34.
27. **Hurst GD**, **Jiggins FM.** 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic, and phylogenetic studies: the effects of inherited symbionts. *Proc Biol Sci* **272**: 1525–34.
28. **Hebert PDN**, **Barrett RDH.** 2005. Reply to the comment by L. Prendini on ''Identifying spiders through DNA barcodes''. *Can J Zool* **83**: 505–6.
29. **Hebert PDN**, **Gregory TR.** 2005. The Promise of DNA Barcoding for Taxonomy. *Syst Biol* **54**: 852–9.
30. **Blaxter ML.** 2004. The promise of DNA taxonomy. *Philos Trans R Soc Lond B* **359**: 669–79.
31. **Vogler AP**, **Monaghan MT.** 2006. Recent advances in DNA taxonomy. *J Zool Syst Evol Res* **45**: 1–10.
32. **Köhler F.** 2007. From DNA taxonomy to barcoding – how a vague idea evolved into a biosystematic tool. *Mitteilungen aus dem Museum für Naturkunde Berlin. Zoologische Reihe* **83**: 44–51.
33. **Seberg O**, **Humphries CJ**, **Knapp S**, **Stephenson DW**, et al. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol Evol* **18**: 63–5.
34. **Kristiansen KA**, **Cileborg M**, **Bkova LD**, **Jørgensen T**, et al. 2009. DNA Taxonomy – the Riddle of *Oxychloë* (Juncaceae). *Syst Bot* **30**: 284–9.
35. **Will KW**, **Mishler BD**, **Wheeler QD.** 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* **54**: 844–51.
36. **Dayrat B.** 2005. Towards integrative taxonomy. *Biol J Linn Soc* **85**: 407–15.
37. **Padial JM**, **Miralles A**, **De la Riva I**, **Vences M.** 2010. The integrative future of taxonomy. *Front Zool* **7**: 16.
38. **Jeffreys AJ**, **Wilson V**, **Thein SL.** 1985. Hypervariable 'Minisatellite' regions in human DNA. *Nature* **314**: 67–73.
39. **Woese CR.** 1987. Bacterial evolution. *Microbiol Rev* **51**: 221–71.
40. **Saitou N**, **Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–25.
41. **Ratnasingham S**, **Hebert PDN.** 2007. BOLD: the barcode of life data system (http://www.barcodinglife.org) *Mol Ecol Notes* **7**: 355–64.
42. **Burns JM**, **Janzen DH**, **Hajibabaei M**, **Hallwachs W**, et al. 2007. DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *J Lepid Soc* **61**: 138–53.
43. **Burns JM**, **Janzen DH**, **Hallwachs W**, **Hajibabaei M**, et al. 2009. Genitalia, DNA barcodes, and life histories synonymize *Telles* with *Thracides* – a genus in which *Telles arcalaus* looks out of place (Hesperiidae: Hesperiinae). *J Lepid Soc* **63**: 141–53.
44. **Burns JM**, **Janzen DH**, **Hallwachs W**, **Hajibabaei M**, et al. 2010. Genitalia, DNA barcodes, larval facies, and foodplants place the mimetic species *Neoxeniades molion* in *Rhinthon* (Hesperiidae: Hesperiinae). *J Lepid Soc* **64**: 69–78.
45. **Will KW**, **Rubinoff D.** 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* **20**: 47–55.
46. **Little DP**, **Stevenson DW.** 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* **23**: 1–21.
47. **Goldstein PZ**, **DeSalle R.** 2000. Phylogenetic species, nested hierarchies, and character fixation. *Cladistics* **16**: 364–84.
48. **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.
49. **Hajibabaei M**, **Singer GAC**, **Hebert PDN**, **Hickey DA.** 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics, and population genetics. *Trends Genet* **23**: 167–72.
50. **DeSalle R**, **Egan MG**, **Siddall M.** 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos Trans R Soc B* **360**: 1905–16.
51. **Prendini L.** 2005. Comment on ''Identifying spiders through DNAbarcodes''. *Can J Zool* **83**: 498–504.
52. **Ferguson JWH.** 2002. On the use of genetic divergence for identifying species. *Biol J Linn Soc* **75**: 509–16.
53. **Bergmann T**, **Hadrys H**, **Breves G**, **Schierwater B.** 2009. Character-based DNA barcoding: a superior tool for species classification. *Bergmann Berl. 446 Münch. Tierärztl. Wochenschr.* **122**: 446–50.
54. **Witt JDS**, **Threloff DL**, **Hebert PDN.** 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Mol Ecol* **15**: 3073–82.
55. **Saunders GW.** 2008. A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. The foliose *Dilsea-Neodilsea* complex and *Weeksia*. *Botany* **86**: 773–89.
56. **Oliver PM**, **Adams M**, **Lee MSY**, **Hutchinson MN**, et al. 2009. Cryptic diversity in vertebrates: molecular data double estimates of species diversity in a radiation of Australian lizards (Diplodactylus, Gekkota). *Proc R Soc B* **276**: 2001–7.
57. **Wiemers M**, **Fiedler K.** 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool* **4**: 1–16.
58. **Yassin A**, **Markow TA**, **Narechania A**, **O'Grady PM**, et al. 2010. The Genus *Drosophila* as a model for testing tree and character-based methods of species identification using DNA barcoding. *Mol Phylogenet Evol* **56**: 1–10.
59. **Meyer CP**, **Paulay G.** 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol* **3**: e 422.
60. **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. *Syst Biol* **55**: 715–28.
61. **Whitworth TL**, **Dawson RD**, **Magalon H**, **Baudry E.** 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proc R Soc Biol Sci Ser B* **274**: 1731–9.
62. **Trewick SA.** 2008. DNA barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics* **24**: 240–54.

63. **Spooner DM.** 2009. DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *Am J Bot* **96**: 1177–89.
64. **Hickerson MJ**, **Meyer CP**, **Moritz C.** 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst Biol* **55**: 729–39.
65. **Verbruggen H**, **De Clerck O**, **Kooistra W**, **Coppejans E.** 2005. Molecular and morphometric data pinpoint species boundaries in *Halimeda* section *Rhipsalis* (Bryopsidales, Chlorophyta). *J Phycol* **41**: 606–21.
66. **Lafontaine JD**, **Schmidt BC.** 2010. Annotated check list of the Noctuoidea (Insecta: Lepidoptera) of North America north of Mexico. *Zookeys* **40**: 1–239.
67. **Miller SE.** 2007. DNA barcoding and the renaissance of taxonomy. *Proc Natl Acad Sci USA* **104**: 4775–6.
68. **Knapp S**, **Lamas G**, **Lughadha EN**, **Novarino G.** 2004. Stability or stasis in the names of organisms: the evolving codes of nomenclature. *Philos Trans R Soc Lond B* **359**: 611–22.
69. **Adl SM**, **Leander BS**, **Simpson AG**, **Archibald JM**, et al. 2007. Diversity, nomenclature and taxonomy of protists. *Syst Biol* **56**: 684–9.
70. **Fraser C**, **Alm EJ**, **Polz MF**, **Spratt BG**, et al. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741–2.
71. **Stackebrandt E**, **Frederiksen W**, **Garrity GM**, **Grimont PA**, et al. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**: 1043–7.
72. **Stackebrandt E**, **Ebers J.** 2006. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* **33**: 152–5.
73. **Richter M**, **Rossello-Mora R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–31.
74. **Erwin TL.** 1991. Establishing a tropical species co-occurrence database. Part I. A plan for developing consistent biotic inventories in temperate and tropical habitats. *Memoria del Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Peru*, Vol. 20, p. 1–16.
75. **Schindel DE**, **Miller SE.** 2010. The on-ramp to taxonomic names. In Polaszek A, ed; *Systema Naturae 250 – The Linnaean Ark*. London, UK: Natural History Museum. p. 109–15.
76. **Nilsson RH**, **Kristiansson E**, **Ryberg M**, **Larsson K-H.** 2005. Approaching the taxonomic affiliation of unidentified sequences in public databases – an example from the mycorrhizal fungi. *BMC Bioinformatics* **6**: 178–90.
77. **La Salle J**, **Wheeler Q**, **Jackway P**, **Winterton S**, et al. 2009. Accelerating taxonomic discovery through automated character extraction. *Zootaxa* **2217**: 43–55.
78. **Gaston KJ**, **O'Neil MA.** 2004. Automated species identification: why not? *Philos Trans R Soc Lond B* **359**: 655–67.
79. **Clark BR**, **Charles H**, **Godfray J**, **Kitching IJ**, et al. 2009. Taxonomy as an eScience. *Philos Trans R Soc A* **367**: 953–66.
80. **Yeates D.** 2009. Viva La Revolución! Designing the digital renaissance in zoological Taxonomy. *Aust J Entomol* **48**: 189–93.
81. **Hibbett DS**, **Nilsson RH**, **Snyder M**, **Fonseca M**, et al. 2005. Automated phylogenetic taxonomy: an example in the Homobasidiomycetes (mushroom-forming fungi). *Syst Biol* **54**: 660–8.
82. **Doyen JT**, **Slobodchikoff CN**, 1974. An operational approach to species classification. *Syst Zool* **23**: 239–47.
83. **Lewontin RC**, **Hubby JL.** 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**: 595–609.
84. **Harris H.** 1966. Enzyme polymorphisms in man. *Proc R Soc Lond B Biol Sci* **164**: 298–310.
85. **Nixon KC**, **Carpenter JM.** 1996. On simultaneous analysis. *Cladistics* **12**: 221–41.
86. **Page TJ**, **Choy SC**, **Hughes JM.** 2005. The taxonomic feedback loop: symbiosis of morphology and molecules. *Biol Lett* **1**: 139–42.
87. **Schlick-Steiner BC**, **Seifert B**, **Stauffer C**, **Christian E**, et al. 2007. Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends Ecol Evol* **22**: 391–2.
88. **Lipscomb D**, **Platnick N**, **Wheeler Q.** 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol Evol* **18**: 65–6.
89. **Crisci JV.** 2006. One-dimensional systematist: perils in a time of steady progress. *Syst Bot* **31**: 217–21.
90. **Roe AD**, **Sperling FAH.** 2007. Population structure and species boundary delimitation of cryptic *Dioryctria* moths using an integrative approach. *Mol Ecol* **16**: 3617–33.
91. **de Carvalho MR**, **Bockmann FA**, **Amorim DS**, **Brandão CRF.** 2008. Systematics must embrace comparative biology and evolution, not speed and automation. *Evol Biol* **35**: 150–7.
92. **Gibbs J.** 2009. Integrative taxonomy identifies new (and old) species in the *Lasioglossum* (*Dialictus*) *tegulare* (Robertson) (Hymenoptera, Halictidae) species group. *Zootaxa* **2032**: 1–38.
93. **Gibbs J.** 2009. New species in the *Lasioglossum petrellum* species group identified through an integrative taxonomic approach. *Can Entomol* **141**: 371–96.
94. **Hajibabaei M**, **Janzen DH**, **Burns JM**, **Hallwachs W**, et al. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci USA* **103**: 968–71.
95. **Packer L**, **Gibbs J**, **Sheffield C**, **Hanner R.** 2009. DNA barcoding and the mediocrity of morphology. *Mol Ecol Resour* **9**: 42–50.
96. **Brower AVZ.** 2006. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Syst Biodiv* **4**: 127–32.
97. **Gregory TR.** 2005. DNA barcoding does not compete with taxonomy. *Nature* **434**: 1067–8.
98. **Janzen DH**, **Hajibabaei M**, **Burns JM**, **Hallwachs W**, et al. 2005. Wedding biodiversity inventory of a large complex Lepidoptera fauna with DNA barcoding. *Philos Trans R Soc Lond B Biol Sci* **360**: 1835–45.
99. **Sites JW**, **Marshall JC.** 2004. Operational criteria for delimiting species. *Annu Rev Ecol Evol Syst* **35**: 199–227.
100. **Pons J**, **Barraclough TG**, **Gomez-Zurita J**, **Cardoso A**, et al. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* **55**: 595–609.
101. **Matz MV**, **Nielsen R.** 2005. A likelihood ratio test for species membership based on DNA sequence data. *Philos Trans R Soc B* **360**: 1969–74.
102. **Nielsen R**, **Matz M.** 2006. Statistical approaches for DNA barcoding. *Syst Biol* **55**: 162–9.
103. **Bond JE**, **Stockman AK.** 2008. An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian Trapdoor Spiders with extreme genetic divergence and geographic structuring. *Sys Biol* **57**: 628–46.
104. **Zhang AB**, **Sikes DS**, **Muster C**, **Li SQ.** 2008. Inferring species membership using DNA sequences with back-propagation neural networks. *Syst Biol* **57**: 202–15.
105. **Steinke D**, **Vences M**, **Salzburger W**, **Meyer A.** 2005. TaxI: a software tool for DNA barcoding using distance methods. *Philos Trans R Soc B* **360**: 1975–80.
106. **Knowles LL**, **Carstens BC.** 2007. Delimiting species without monophyletic gene trees. *Syst Biol* **56**: 887–95.
107. **Floyd R**, **Abebe E**, **Papert A**, **Blaxter M.** 2002. Molecular barcodes for soil nematode identification. *Mol Ecol* **11**: 839–50.
108. **Sirovich L**, **Stoeckle MY**, **Zhang Y.** 2009. A scalable method for analysis and display of DNA sequences. *PLoS ONE* **4**: e7051.
109. **Blaxter M**, **Floyd R.** 2003. Molecular taxonomics for biodiversity surveys: already a reality. *Trends Ecol Evol* **18**: 268–9.
110. **Blaxter MJ**, **Mann T**, **Chapman F**, **Thomas C**, et al. 2005. Defining operational taxonomic units using DNA barcode data. *Philos Trans R Soc B* **360**: 1935–43.
111. **Davis JI**, **Nixon KC.** 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Syst Biol* **41**: 421–35.
112. **Brower AVZ.** 1999. Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. *Syst Biol* **48**: 199–213.
113. **Sarkar IN**, **Planet PJ**, **Bael TE**, **Stanley SE**, et al. 2002. Characteristic attributes of cancer microarrays. *J Biomed Inform* **35**: 111–22.
114. **Ward RD**, **Hanner R**, **Hebert PDN.** 2009. The campaign to DNA barcode all fishes, FISH-BOL. *J Fish Biol* **74**: 329–56.
115. **Chase MW**, **Salamin N**, **Wilkinson M**, **Dunwell JM**, et al. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc B* **360**: 1889–95.
116. **Check E.** 2006. Treasure island: pinning down a model ecosystem. *Nature* **439**: 378–9.
117. **Kerr KCR**, **Stoeckle MY**, **Dove CJ**, **Weigt LA**, et al. 2007. Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes* **7**: 535–43.
118. **Ward RD**, **Zemlak TS**, **Innes BH**, **Last PR**, et al. 2005. DNA barcoding Australia's fish species. *Philos Trans R Soc B* **360**: 1847–57.
119. **Silva-Brandão K**, **Lyra M**, **Freitas AVL.** 2009. Barcoding lepidoptera: current situation and perspectives on the usefulness of a contentious technique. *Neotrop Entomol* **38**: 441–5.

120. **Mengual X**, **Thompson FC.** 2008. A taxonomic review of the *Palpada ruficeps* species group, with the description of a new flower fly from Colombia (Diptera: Syrphidae). *Zootoxa* **1741**: 31–6.
121. **Hastings JM**, **Schultheis PJ**, **Whitson M**, **Holliday CW**, et al. 2008. DNA barcoding of new world cicada killers (Hymenoptera: Crabronidae). *Zootaxa* **1713**: 27–38.
122. **Fochetti R**, **Tierno deFigueroa JM.** 2009. A new species of Leuctridae discovered by means of molecular and biochemical approaches: *Tyrrhenoleuctra antoninoi* n. sp. (Insecta: Plecoptera). *Zootaxa* **2112**: 41–6.
123. **Schmidt BC.** 2009. Taxonomic revision of the genus *Grammia* Rambur (Lepidoptera: Noctuidae: Arctiinae). *Zool J Linn Soc* **156**: 507–97.
124. **Õunap E**, **Viidalepp J.** 2009. Description of *Crypsiphona tasmanica* sp. nov. (Lepidoptera: Geometridae:Geometrinae), with notes on limitations in using DNA barcodes for delimiting species. *Aust J Entomol* **48**: 113–24.
125. **Fisher BL**, **Smith MA.** 2008. A Revision of Malagasy Species of *Anochetus* Mayr and *Odontomachus* Latreille (Hymenoptera: Formicidae). *PLoS ONE* **3**: e1787.