



HAL
open science

DNA Barcoding, species delineation and taxonomy: a historical perspective

Nicolas Hubert, Robert Hanner

► **To cite this version:**

Nicolas Hubert, Robert Hanner. DNA Barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes, 2015, 3 (1), 10.1515/dna-2015-0006 . hal-01958691

HAL Id: hal-01958691

<https://hal.science/hal-01958691>

Submitted on 3 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Review Article

Open Access

Nicolas Hubert, Robert Hanner

DNA Barcoding, species delineation and taxonomy: a historical perspective

DOI 10.1515/dna-2015-0006

Received December 11, 2014; accepted March 17, 2015

Abstract: DNA barcoding is a system designed to provide species identification by using standardized gene regions as internal species tag. Foreseen since its earlier development as a solution to speed up the pace of species discovery, DNA barcoding has established as a mature field of biodiversity sciences filling the conceptual gap between traditional taxonomy and different fields of molecular systematics. Initially proposed as a tool for species identification, DNA barcoding has also been applied in taxonomy routines for automated species delineation. Species identification and species delineation, however, should be considered as distinct activities relying on different theoretical and methodological backgrounds. The aim of the present review is to provide an overview of the use of DNA sequences in taxonomy, since the earliest development of molecular taxonomy until the development of DNA barcoding. We further present the differences between procedures of species identification and species delineation and highlight how DNA barcoding proposed a new paradigm that helps promote more sustainable practices in taxonomy.

Keywords: DNA barcodes; species delineation; specimens identification; coalescent theory; divergence threshold; integrative taxonomy

1 Introduction

DNA barcoding is a system designed to provide accurate, fast and automatable species identification by using short and standardized gene regions as internal species tag [1]. It has been initially proposed to circumvent the lack of taxonomists and available of tools for species identification (*i.e.* taxonomic impediment) by enhancing our capacity to accurately document biodiversity and as such, DNA barcoding was foreseen, since its earliest development, as a solution to speed up the pace of species discovery and open new perspectives in conservation [2-4]. While entering an era of mass extinction, it becomes evident that describing biological diversity with traditional approaches happens at a much slower rate than the rate of species loss [2,5], hence jeopardizing the conservation of earth biotas.

Since the initial proposal by Hebert and colleagues [6], DNA barcoding has established as a mature field of biodiversity sciences filling the conceptual gap between traditional taxonomy and different fields of molecular systematics [6-8]. The rationale behind DNA barcoding originates from the accumulation of evidence during the last two decades, as a consequence of an increasing access to genomic data, that the characterization and documentation of biodiversity using phenotypes is currently bridled by several limits inherent to morphological characters [9-13]. First, intraspecific phenotypic variation often overlaps that of sister taxa in nature, which can lead to incorrect identifications or species delineations [9,11]. Second, DNA barcodes are effective whatever the life stages under scrutiny [14,15] or available biological materials for identification [16,17]. Third, spectacular levels of cryptic diversity have been frequently reported using DNA barcoding [11-13,18-20]. Altogether these advantages have opened new perspectives in fields as diversified as functional ecology [13], taxonomy [1,8,12], biogeography [18,21,22], conservation [23], wildlife forensics [16,17,24-26] and biodiversity socio-economics [27,28].

It has become manifest during the initial building of the DNA barcodes reference libraries that DNA barcoding

*Corresponding author: Nicolas Hubert, Institut de Recherche pour le Développement (IRD), UMR226 ISE-M, Bât. 22 - CC065, Place Eugène Bataillon, 34095 Montpellier cedex 5, France, E-mail: nicolas.hubert@ird.fr

Robert Hanner: Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

would have to face the undescribed diversity of earth biotas [e.g. 12,13,18] and that DNA barcodes might help to speed up the pace of species discovery through automated delineation of mitochondrial lineages. While species delineation was not initially intended as a primary goal of DNA barcoding [6], the integration of DNA barcoding as a routine in taxonomy has called considerable attention during the last decade and led to several conceptual and methodological advances that rejuvenated the practice of taxonomy [1].

Most of the controversy on DNA barcoding has been initially revolving around its use for species delineation but the same criticisms have been applied later to both specimen identification or species delineation. The use of DNA sequences for either identification or species delineation is embedded in the conceptual framework of molecular taxonomy [1-4,29,30] and the coalescent theory [31-33]. The present paper aims at reviewing those foundations and outlines the historical development that led to the establishment of DNA barcoding from conceptual and methodological perspectives. First, we will briefly sketch the constraints faced by the worldwide community of taxonomists in the 90's that led to the development of molecular taxonomy and DNA barcoding. Second, we will describe how molecular taxonomy helped establish criteria for a global system of molecular identification to ensure its stability. Third, we will list the connections that have established between DNA barcoding and taxonomy. Fourth, we will review the applications of DNA barcoding in taxonomy and its limits in light of the coalescent theory and properties of gene genealogies. Finally, we discuss the future directions pointed out recently in molecular taxonomy.

2 From the taxonomic impediment to DNA barcoding

During the second conference of the parties of the Convention on Biological Diversity (CBD) held in Jakarta in 1995, the participant countries have explicitly formulated through the concept of taxonomic impediment the major concern raised by the worldwide community of taxonomists since the 90's about the increasing disinterest from governments and funding agencies for taxonomy. Unfortunately, several global initiatives such as the Global Taxonomic Initiative (GTI) launched in the context of the CBD early in 2002 failed to embrace a massive adhesion and failed to help reach the CBD goal to slow-down the pace of species loss by 2010. Several challenges prevented the emergence of a global project including the settlement

of a universal information system in taxonomy and the digitizing of the collections in national museums, both calling for a more massive investment in taxonomy as a research priority by the nations [30,34]. Another challenge is caused by the lack of consensus on the morphological characters to be used by the community of taxonomists, a limit that was to be overcome by the use of DNA sequences due to the universality of the genetic code [1-4,29,30]. Moreover, the ease of access to sequencing facilities was expected by a large community to counterbalance the impact of the taxonomic impediment in conservation and basic biodiversity sciences [28].

A potential solution to the gap in global information systems in taxonomy was proposed by the international Barcode of Life project (iBOL) through the creation of a database system enabling the repository of sequences but also offering a workbench for a collective assembly of the DNA barcode libraries [29,35]. This project led to the launch in 2005 of the Barcode of Life datasytem (www.boldsystems.org) as a solution to an interactive online database enabling collective assembly and curation of the libraries. The development of such a global system had to face several challenges: (i) identification based on molecular data should be reproducible to enable a universal use, (ii) owing to the state of the art of the world inventory of living beings, the system should ensure the storage of collateral data enabling further taxonomic studies, (iii) open access to the data should be guaranteed in agreement with the Access and Benefit Sharing (ABS) principle established by the CBD.

3 DNA barcoding: toward the establishment of a global information system

The first step in the development of a global system of molecular identification is the constitution of the DNA barcodes reference libraries for known species based either on large scale field sampling campaigns [18,36-38] or sequencing collections in natural history museums whenever possible (e.g. birds and insects [39,40]). In order to ensure the reproducibility of molecular identifications based on DNA barcodes reference libraries, however, the BOLD database currently hosts specimen records for which essentially, seven data elements are listed (Fig. 1):

1. Species name
2. Voucher data
3. Collection record
4. Identifier of the specimen

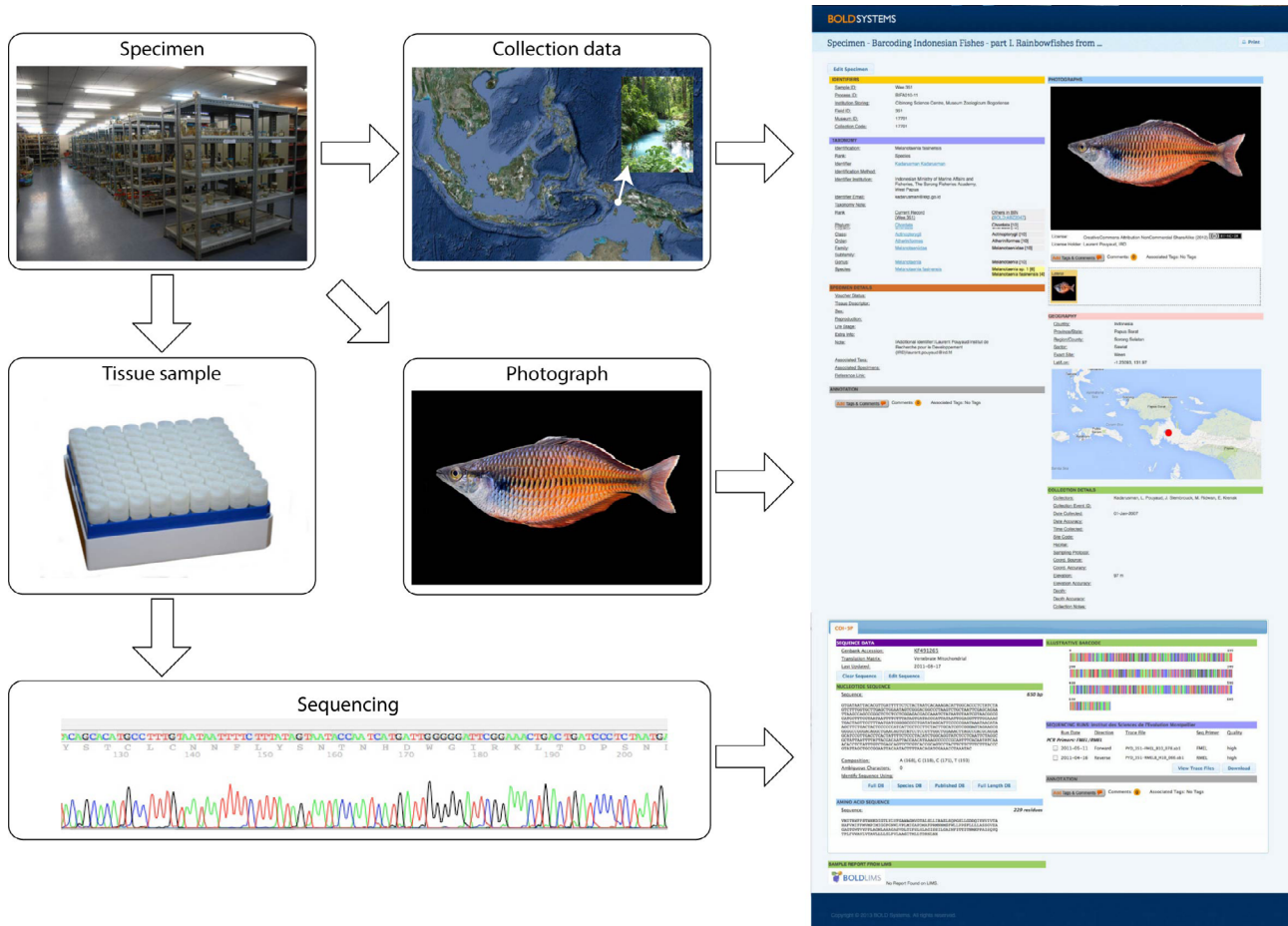


Figure 1. Structure of a specimen record in BOLD. The BARCODE keyword in genbank is reserved for the records compliant with the following scheme including a voucher specimen in a biological collection, a tissue sample in a bio-repository, collection data, a specimen photograph and a DNA barcode including primary data (e.g. trace files).

5. COI sequence of at least 500 bp
6. PCR primers used to generate the amplicon
7. Trace files

Altogether, these data allow users to access raw data at any step during the production of DNA barcodes in order to: (i) ensure the reproducibility of the PCR and sequencing protocols, (ii) allow the validation and detection of potential discrepancies in the initial identification of specimens by the community of users, (iii) ensure traceability by providing contacts to the key peoples involved in generating data, (iv) allow further taxonomic studies when discrepancies are detected between molecules and phenotypes as a consequence of cryptic diversity or the detection of new evolutionary lineages.

To further improve reliability and reproducibility of DNA barcoding, the Consortium for the Barcode of Life (cBOL), in cooperation with GenBank and the other members of the International Nucleotide Sequence Database Collaboration (INSDC), have created and implemented the BARCODE data standard. “BARCODE” is

a reserved keyword for those records in an INSDC database that meet a higher quality standard and are compliant with the following requirements:

1. Bi-directional sequences of at least 500 base-pairs from the approved barcode region of COI, containing no ambiguous sites
2. Links to electropherogram trace files available in the NCBI Trace Archive
3. Sequences for the forward and reverse PCR amplification primers
4. Species names that refer to documented names in a taxonomic publication or other documentation of the species concept used
5. Links to voucher specimens using the approved format of institutional acronym:collection code:catalog ID number.

Altogether, these data allow connecting vouchers specimens, further available for screening diagnostic morphological characters any time undescribed diversity

is detected, and DNA barcodes. This way, the essential link between genomes and phenotypes is ensured and vouchers specimens play the similar role for genomes as type specimens for species names in taxonomy guaranteeing nomenclatural stability by linking species name to specimens instead of concepts (*i.e.* the delineation of a species by a given author at a given time) necessarily varying through time as taxonomic knowledge accumulates [41].

4 DNA barcoding: how it complements taxonomy

The initial goal of iBOL is the settlement of a universal system based on DNA barcode reference libraries upon which molecular identifications rely (Fig. 2). Yet, DNA barcoding has proved able to capture a large majority of the diversity in the case of well-know faunas such as the North-American fishes and birds with only very

few discrepancies between molecules and phenotypes [36,39]. It proved also, however, to constitute a powerful approach when dealing with hyperdiverse tropical fauna by facilitating the delineation of new evolutionary lineages representing instances of new species sometimes at unexpected rates as recently emphasized in arthropods and crustaceans [11-13,20,42,43] or fishes [18,19,37,44]. Worth mentioning, DNA barcoding also revealed some overlooked taxa in well-studied temperate faunas [36,39,40].

Despite that the usefulness of DNA sequences for taxonomy is not disputed, DNA barcoding has been controversial in some scientific circles based on the rationale that providing speeding up the inventory of living beings means simplifying procedures, so an integrative approach of taxonomy was needed rather than DNA barcoding [45-47]. Since its earliest development, DNA barcoding faced the undescribed component of biodiversity, sometimes in an order of magnitude higher than expected, and its ease of access highlighted the

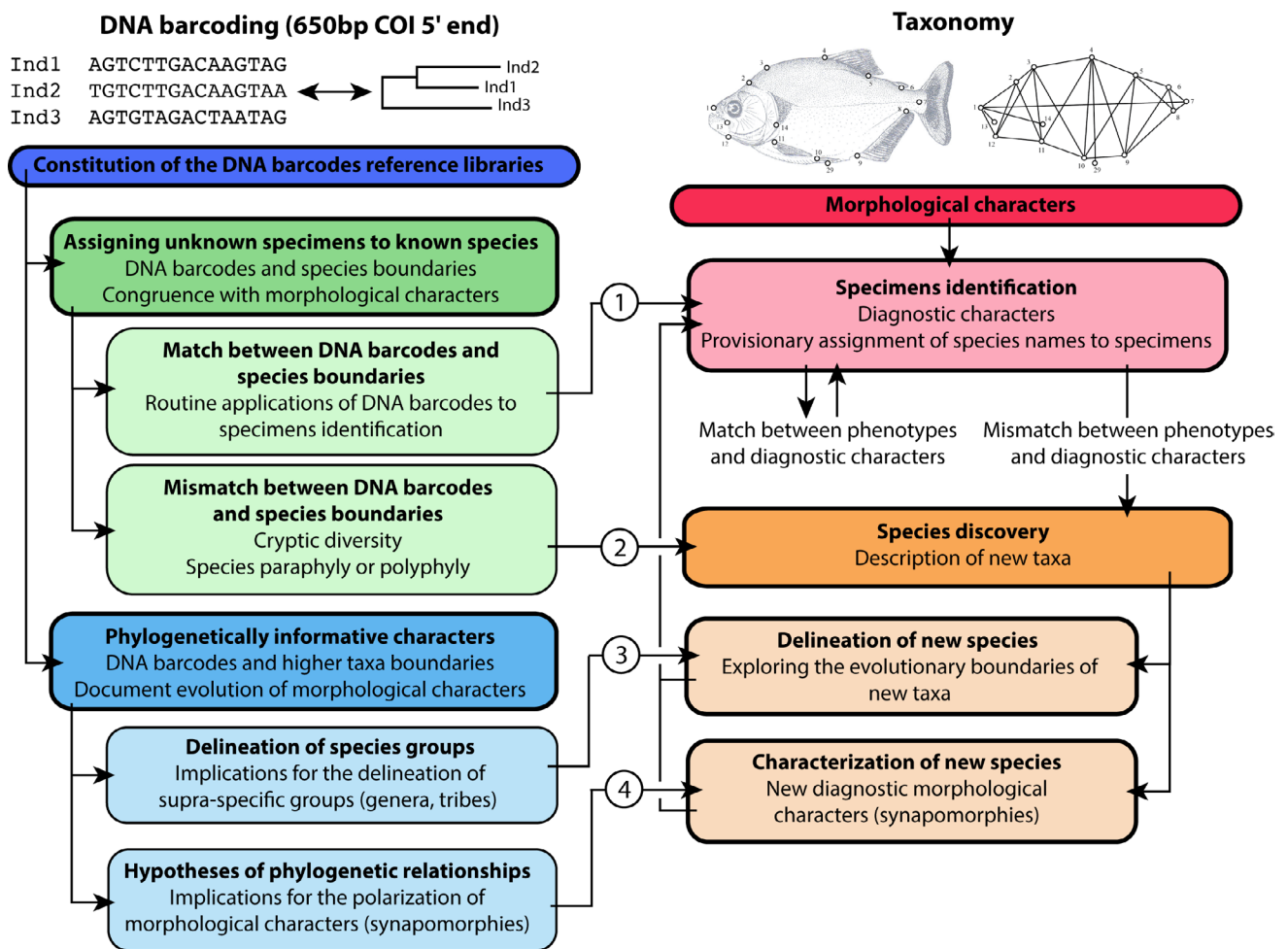


Figure 2. Conceptual links between DNA barcoding and taxonomy.

benefit of using DNA barcoding as a first step during species inventories (Fig. 2). Actually, iterative procedures including DNA barcoding, taxonomy and natural history have been successfully applied for the delineation and description of new species in megadiverse, yet poorly described, biotas [12,40,48]. Integrative approaches have recently demonstrated to be highly useful in speeding the pace of species discovery and description and the term turbo-taxonomy has been even applied recently by Butcher and colleagues [49] to the procedures including DNA barcoding as a first step toward the fast description of species based on the combination of COI sequences, concise morphological descriptions by an expert taxonomist, and high-resolution digital imaging to streamline the description of larger number of species [49,50]. Thus, instead of proposing a replacement to an integrative approach for the description of living beings, DNA barcoding is more frequently integrated as a routine in large-scale biodiversity inventories and this integrative approach is also well exemplified by recent large scale DNA barcoding projects focusing on collections in national museums and as such allowing the integration of the legacy of more than a century of natural history into the constitution of DNA barcode libraries [40].

Implication of DNA barcoding for species delineation in an integrative framework should be considered separately from the routine of specimen identification for known species because they rely on distinct practices of taxonomy. Most of the controversy on DNA barcoding has been revolving around the threat that developing automated molecular identification will expedite the decline of taxonomy by monopolizing funds that would be devoted otherwise to taxonomy. The recent literature on turbo-taxonomy, however, evidenced that instead of expediting the decline of taxonomy, DNA barcoding has triggered the return of funding in alpha taxonomy and reinvigorated the field by linking several fields of systematics (e.g. taxonomy vs. phylogeny) that were evolving independently from each others since more than a decade [1,8]. In fact, DNA barcoding opened new perspective in species delineation through automated and standardized protocols for DNA sequencing and data labeling and exemplified by the large-scale campaign above mentioned.

The taxonomic impediment has been dramatically limiting the expertise worldwide for species identification. Species identification, however, is not the primary goal of taxonomy, by contrast with species delineation, while being of high societal importance [1]. The use of DNA barcoding *per se* for species delineation would have been a step back in taxonomy from a conceptual perspective;

however, its application for species identification provided a universal answer to the taxonomic impediment by enabling fast and automated identifications. Taxonomy is still an active field of research, even after more than two centuries of inventory of living beings, and hypotheses of species delineation are constantly being re-examined and revised. In this context, incorporating up to date taxonomic knowledge is a concern [12,20,51] and depositing voucher specimens in national repositories has been explicitly defined as a mandatory step to ensure accurate and up-to-date identifications by enabling taxonomists to validate or perform new identifications any times discrepancies between DNA barcodes and the interpretation of morphological characters are observed. In turn, DNA barcoding has been reinvigorating national collections worldwide through the development of new reference collections linking genotypes and phenotypes.

5 Applications of DNA barcoding

5.1 Coalescent theory and limits of the one gene approach based on mitochondrial genomes

The rise of the coalescent theory in the early 80's largely opened new perspectives in the understanding of gene genealogies in populations and species [31-33]. The coalescent theory is a sampling theory based upon individual-based models that has proved to open new perspectives compared to population-based models to analyze sequence polymorphism in natural populations [52]. The coalescent theory describes the sampling of genes in populations that happen at each generation. Considering a diploid population of effective size Ne that sexually reproduce and $2Ne$ the number of copies of a gene in the population, the probability that two copies of a gene at generation t come from the same ancestor at the previous generation t_1 is $1/2Ne$ and the probability that they don't coalesce is $1-1/2Ne$. Thus, the probability that two sequences coalescence at time x is given by equation (1):

$$(1) P(t = x) = \frac{1}{2Ne} \left(1 - \frac{1}{2Ne}\right)^x$$

Given that $2Ne$ is large, this become in a continuous formulation

$$(2) P(t = x) = \frac{1}{2Ne} e^{\left(-\frac{x}{2Ne}\right)}$$

The probability that a coalescence occur at time x in a population of i sequences is given by eqn 3

$$(3) P(t = x) = \frac{C_i^2}{2Ne} e^{-\frac{x C_i^2}{2Ne}}$$

The expected age of the genealogy is given in eqn 4:

$$(4) TMRCA = \frac{4Ne}{1-\frac{1}{i}}$$

Thus, for large numbers of genes i , the expected age of a genealogy for a diploid population of effective size Ne is $4Ne$. For genomes inherited from a single parent such as mitochondrial DNA, the expected age of a genealogy becomes Ne . This result explains in part the choice of mitochondrial genes for DNA barcoding since the pace of genetic drift is expected to be four time faster here compared to nuclear genes and as such, mitochondrial genes are expected to become diagnostic of isolated lineages faster compared to nuclear DNA [53].

The coalescent theory established a simple relationship between population size and coalescence dynamic that enabled some straightforward predictions regarding the distribution of molecular polymorphism after the divergence of two populations (Fig. 3). From the equation 4, it appears that if the isolation of two populations is younger on average than $4Ne$ for nuclear DNA or Ne for mitochondrial DNA, genes in a population may coalesce with genes from other populations at a time prior to the isolation of the populations. This phenomenon has been formalized as the retention of ancestral polymorphism and constitutes a well know limit of the single gene approach using genes with an uniparental inheritance (Fig. 4). Since mitochondrial DNA is maternally inherited, it is impossible to disentangle ancestral polymorphism from recent gene flow in the origin of shared polymorphism since only part of the genetic material from the parents is left at the next generation [54,55]. Thus, species polyphyly and paraphyly are expected for recently diverged species [56]. The percent of failure of DNA barcoding as a consequence of this shortcomings has been demonstrated to be low for vertebrates, however, as less than 10% of the species show mixed genealogies in the recent studies based on comprehensive continental sampling [36,39]. Nevertheless, specimens identification to the species level may be done a posteriori by integrating geographic information (Fig. 5).

5.2 Cryptic diversity and the discovery of new species

Much of the controversy related to the integration of DNA barcoding into the taxonomic workflow has been related to the use of DNA sequences in delineating species [45-47]. The global campaign to DNA barcode animal and plant species gave rise to a large array of data release papers

describing the effectiveness of DNA barcoding in capturing species boundaries for large data set [18,20,37,57], sometimes from a comprehensive continental perspective [36,39,40]. All the large-scale studies conducted in tropical ecosystems resulted in high levels of cryptic diversity as revealed by DNA barcodes. This situation led to two distinct consequences for the practice of taxonomy in tropical ecosystems and the collective curating of the DNA barcodes reference libraries.

From a practical perspective, mismatch between DNA barcodes clusters and nominal species are frequently observed in the tropics due to the higher occurrence of species polyphyly and paraphyly compared to temperate faunas due to: (i) a higher impact of the taxonomic impediment in the tropics, (ii) a more intricate practice of taxonomy due to the higher diversity [58]. In addition, a growing body of evidence supports that species turn-over through time is slower in the tropics and tropical species are in average older than species in the temperate biomes and host older polymorphisms [59,60], sometimes leading to unexpectedly deep coalescent in tropical species.

In order to facilitate the assembly of DNA barcodes reference libraries, Barcode Index Number (BIN) have been recently created in order to facilitate the checking of mismatch between nominal species and DNA barcodes clusters and ease their discrimination through standard procedures [61]. The attribution of BIN numbers is also designed to facilitate the taxonomic workflow by indexing cryptic lineages readily flagged by DNA barcodes. Along the same line, this index enables to speed up the application for automated species identification anytime the state of the art in taxonomy do not enable the use of species names to label DNA barcodes, as generally observed in tropical faunas. More importantly, however, BIN ease the taxonomic workflow including DNA barcodes as a preliminary step for diversity sorting in mega-diverse ecosystems and enable the implementation of iterative procedures to document and describe biodiversity (Fig. 6).

Riedel and colleagues [50], who described 101 new species of weevil beetle in their study have listed the benefits of such a fast-track iterative procedure for taxonomy: (i) using DNA barcoding to produce a phylogenetic backbone eases the selection of specimens with close DNA sequence affinities to screen morphological characters, (ii) giving up the creation of a traditional identification key that is time-consuming for large faunas using morphological characters, (iii) reduction of species description to the essential diagnostic characters, (iv) reduction of the description of intraspecific variations of limited utility for interspecific comparisons, (v) reduction of the number of illustrations by using highly resolved images, (vi) diagnosis can be

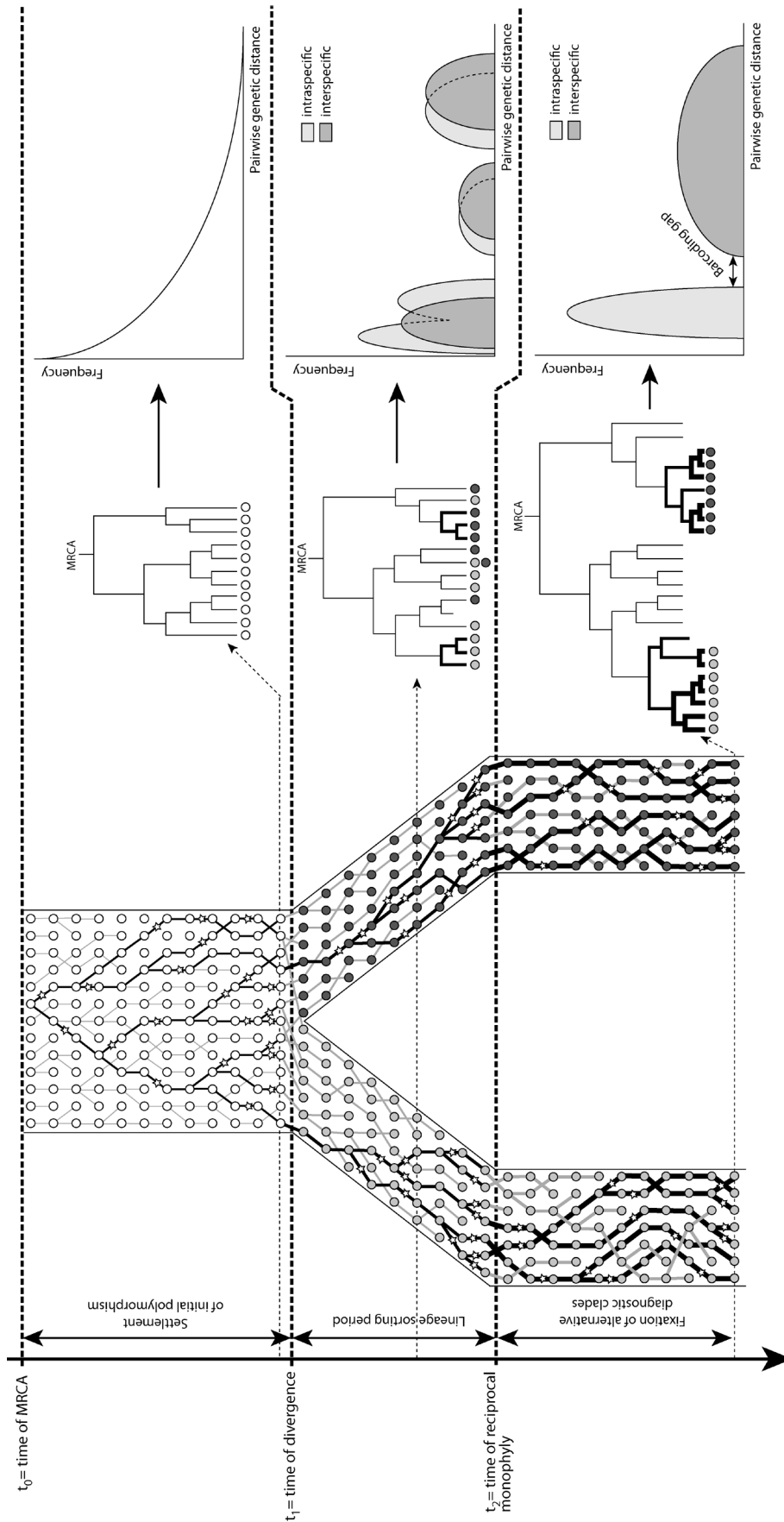


Figure 3. Gene genealogy and coalescent theory. Relationship between Barcoding gap and line of descent of mitochondrial genomes during the settlement of two lineages. Stars represent mutation events leading to new haplotypes, light and dark grey for the ancestral population, light and dark grey for each diverging lineage).

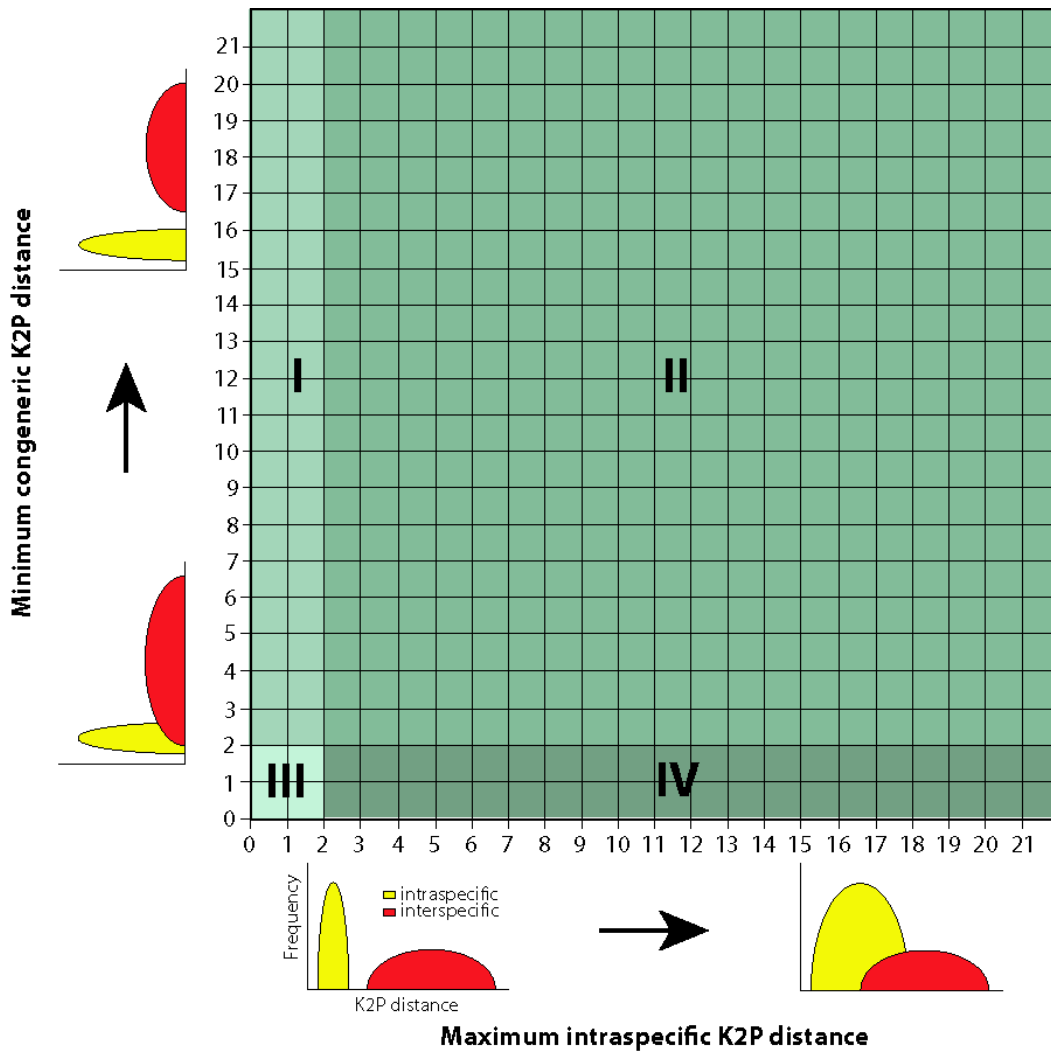


Figure 4. Compared distribution of intraspecific and interspecific genetic distances and associated patterns following a 2% threshold. See text for details.

focused among species with close phylogenetic similarities, (vii) digging out information from type specimens can be shortened and eased through digitization. Altogether, these arguments point out that the use of DNA sequences as the ‘key elements’ in integrative taxonomic studies jointly with the use of digitized information system will pave the way of more sustainable practices in taxonomy with a pace finally compatible with the ultimate goal foreseen by Linnaeus to tackle the inventory of earth living beings [49,50].

5.3 Assigning unknown specimen to know species through DNA barcodes

From an analytical perspective, the initial proposal by Hebert et al [6] was based on the observation that the vast majority of the species analyzed exhibited genetic distances of more than 2 percents while sequence

divergence within species were largely smaller averaging around 0.1 percent [20, 36, 39, 62]. Following these observations, it has been suggested that the distributions of intra- and inter-specific distance do not overlap [63] and exhibit a barcoding gap (Fig. 3). Later, Meyer and Paulay explored the error rates of identifications using a varying threshold approach by estimating the relative frequency of false positive (*i.e.* a conspecific diverging by more than the threshold to the nearest species is attributed to a distinct species) and false-negative (*i.e.* a heterospecific sequence diverging by less than the threshold from the nearest species is attributed to the same species) [64]. The authors [64] demonstrated that the cumulative percent of false- positive and negative might be optimized at 33% for a 0.02 threshold of divergence or 18% when accounting for undescribed evolutionary lineages, but a threshold approach cannot eliminate error rates.

Four different cases have been identified based on a threshold for species divergence (Fig. 4), the two percent threshold being the most commonly used:

- (1) Case I: intraspecific distance is smaller than 2% and interspecific distance is higher than 2%, the species had achieved reciprocal monophyly and results are concordant with current taxonomy.
- (2) Case II: both intraspecific and interspecific distances are higher than 2%, the species is composite and encompasses several lineages.
- (3) Case III: both intraspecific and interspecific distances are lower than 2%, species has recently diverged from its sister-species and either ancestral polymorphism or introgressive hybridization occurs. Synonymy may also apply in this case.
- (4) Case IV: intraspecific distance are greater than 2% while interspecific distance are smaller than 2%, specimens have been probably misidentified and a proper reassessment is needed.

The relative occurrence of the first three cases, given that case IV is an artefact due to misidentification, determine the effectiveness of DNA barcoding for the assignment of sequences from unknown specimens to known species. Most of the case II found during the recent large scale DNA barcoding surveys published so far turn out to fall into case I once cryptic diversity was properly accounted in the analyses [11-13,18,20,36,44,63-65]. When based on case I, DNA barcoding has been applied successfully for purposes as diverse as the identification to the species level of introduced and invasive species [26,66], market substitution [16,67], conservation of endangered wildlife [68,69], identification of early ontogenetic stages [11,14,15,70-72] and assignation of sexual morphotypes to species [73] among the most straightforward applications.

6 What is next for DNA barcoding in taxonomy?

6.1 Objective tools for species delineation

During the last decade, decision models have been developed for either species identification or species delineation based on the rich theoretical background of the coalescent and phylogenetic theories [61,74,75]. These new methods of species delineation based on DNA sequences may be sorted into three categories depending on the algorithms implemented and the processes of

sequence sorting involved during classification decisions:

- (1) phylogenetic methods based on models defining branching patterns along the trees but not modeling coalescent dynamics inside species [74]. These methods are essentially focused on optimizing the parameters of phylogenetic models to identify an objective divergence threshold for clustering individuals as implemented in Spider [76], ABGD [77,78], BIN [61], Bayesian inferences [79,80]. These methods are fundamentally clustering-based decision models, however, variable thresholds may be used through iterative procedures as in ABGD or BIN.
- (2) coalescent methods based on Kingman's model of gene sorting within species [31] but not modeling branching patterns among species [81,82]. These methods are similar in essence with the phylogenetic methods as they produce a classification of sequences after optimizing the parameters of the model. They differ, however, in defining this classification based on coalescent instead of phylogenetic models.
- (3) phylogenetic-coalescent methods based on mixed models including phylogenetic and coalescent components [74]. The most popular model is the General Mixed Yule-Coalescent model [GMYC,83] that takes advantage of the Kingman's coalescent model [31] and Yule's diversification model [84] to optimize the likelihood of the transition between species diversification (*i.e.* speciation rate) and coalescent dynamic (*i.e.* mutation and sorting of genes). In its initial formulation, lineages are delineated when they exceed the threshold value [83], however, this model has been further extended for multiple thresholds [48].

These methods have been developed to deal with the assignment uncertainties inherent to using universal threshold with case II and III or when DNA barcode reference libraries are still partial. These methods, however, require either a representative sampling of intraspecific genealogies (*i.e.* coalescent-based and mixed phylogenetic-coalescent methods), as methods may be sensitive to departure from initial assumptions regarding population size, nucleotidic diversity or populations structure [74,82,85] or a reliable knowledge of phylogenetic relationships when transition threshold (*i.e.* speciation vs. mutation) are estimated based on phylogenetic trees (*e.g.* GMYC). Sampling of gene genealogies and/or phylogenetic trees, however, are rarely optimal, except in limited cases of well know faunas but these methods constitute a major improvement in the objective delineation of putative species and they have proven to speed up the pace of

species inventories, particularly in the case of diverse, yet poorly known, fauna [e.g. 48].

6.2 Integrating DNA barcoding in iterative procedures for species delineation

The use of a single gene approach presents some important limits, particularly if based on mitochondrial genes. Owing to its maternal inheritance, the relative contribution of gene flow and ancestral polymorphism to the origin of shared mitochondrial polymorphism among lineages may be disentangled only in particular cases (*i.e.* geographic isolation; Fig. 5). It should be distinguished, however, between specimens identification and species delineation. Coalescent-based methods specifically designed to disentangle the relative contribution of both phenomena might be able to shed light on the origin of shared polymorphism [e.g. 55]. By contrast, classification methods developed for species delineation are based on thresholds and as such, classify clusters of closely related sequences (*i.e.* monophyletic units).

So far, large-scale studies have demonstrated that

shared polymorphism was bridling the accuracy of species identifications in nearly 10 percents of the cases for vertebrates [e.g. 36,39]. This estimate, however, is based on temperate and well-known faunas and was obtained during the building of DNA barcode reference libraries for species with detailed a priori knowledge on species boundaries. Species richness, however, has been shown to enhance the impact of spatial scale on the effectiveness of DNA barcoding [e.g. 86] because more closely related species are likely to be sampled while increasing spatial scale [85,86]. Increasing spatial scales, however, not only affects the accuracy of species identification through the decrease of inter-specific genetic distances but also increases the opportunity of species range overlap and shared polymorphism (*i.e.* hybridization), a phenomenon compromising the use of DNA barcoding for species delineation (e.g. Fig. 5).

Delineating species with no a priori knowledge about their boundaries appear to be more dramatically affected by shared ancestral polymorphism when based on a single gene approach as no external information helps to distinguish false positive and false negative

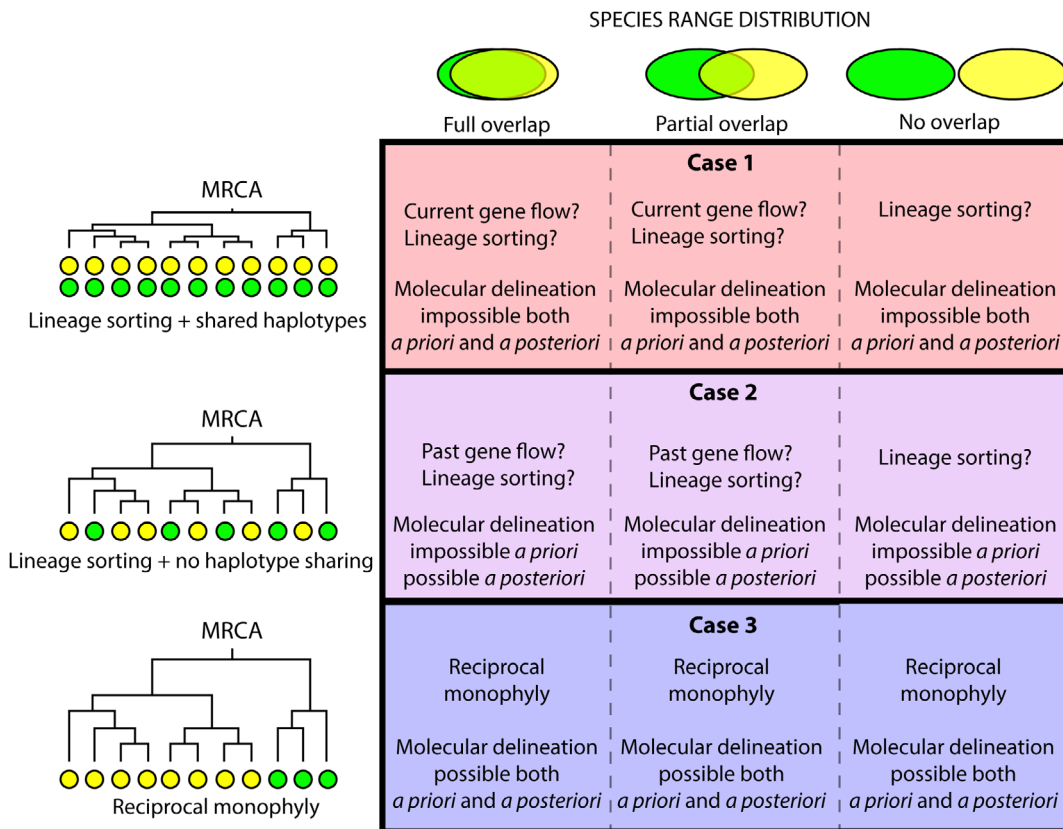


Figure 5. Accuracy of species delineation based on DNA barcoding. Accuracy of species delineation based on mitochondrial DNA is depending on the spatial context and state of the line of descent. A priori species delineation do not account for geographic information or morphological characters (e.g. use of a threshold). A posteriori species delineation account for geographic information, morphological characters and the occurrence of haplotype sharing.

[64,87]. The development of decision-models based either on coalescent or phylogenetic theory has been a major improvement in the objective classification of molecular lineages, although, these models cannot circumvent the inherent limits of a single gene approach [45-47]. The automated and objective classification of mitochondrial lineages enabled by the use of DNA barcode reference libraries, however, has been foreseen by several authors as a potential solution to the time consuming sorting of specimens during inventories of unknown fauna, particularly in tropical areas exhibiting high levels of species richness [49,50]. This procedure enables to speed up the taxonomic workflow through a preliminary sorting of specimens according to their DNA barcodes

follow by an iterative procedure involving natural history, morphology and DNA barcoding (e.g. Fig. 6, [12]). This procedure is making the most of DNA barcoding since fast automated delineations can be performed, instead of the time consuming sorting of specimens based on their morphological attributes, but mitochondrial lineages are further validated or invalidated by using other sources of evidence including additional molecular markers (e.g. nuclear markers with biparental inheritance), life history traits (e.g. host plants for phytophagous beetles) and morphology [1-4,12,49,50]. Recent studies have demonstrated that integrating upstream a preliminary step of DNA barcoding was actually speeding up the taxonomic workflow up to 20 times [50]. In addition, the

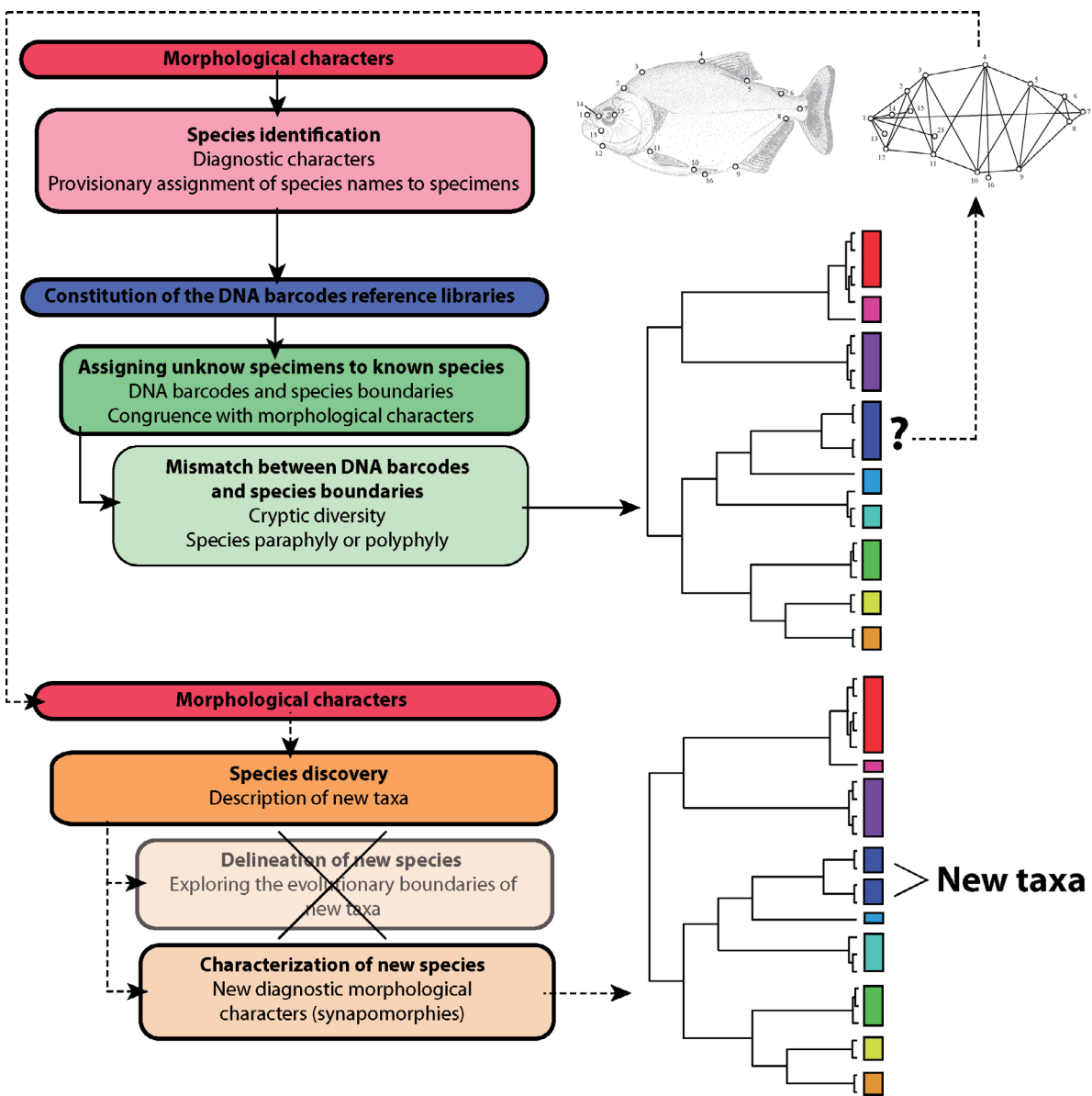


Figure 6. Typical iterative procedure involving DNA barcoding during a taxonomic workflow.

development of multi-locus approach of DNA barcoding, as exemplified in plants [88], and the multiple links developed by DNA barcoding with the taxonomic workflow (e.g. Fig. 2) have demonstrated that DNA barcoding was moving toward an integrative approach of species delineation [11-13,20,22,36,39,51,57].

6.3 Toward more integration with online biodiversity information systems

One of the most impacting limits of the taxonomic impediment during the last decade has been probably the lack of global biodiversity information systems, a gap already pointed out during the 1995's CDB conference of the parties. BOLD has been a major step forward in that direction by applying the same concepts at work for traditional biological collections and nomenclature to the curation of DNA sequence libraries [35]. In doing so, DNA barcoding and BOLD opened new perspectives in providing a data repository associated to an online workbench enabling the collective management of DNA barcode reference libraries from their initiation to their application for routine identification and linking DNA sequence data with voucher specimens repositories. In the context of an integrative taxonomy workflow, BOLD makes the most of the DNA barcode data standards by introducing an access to all primary data at any steps and promoting more sustainable practices during the production of DNA sequences for taxonomic purposes.

Along the same line, several other initiatives have been launched to summarize and ease access to primary data and biological repositories for taxonomic purposes. The Global Registry of Biorepositories (GRBio), for instance, aims at offering a consolidated clearinghouse of information about biological collections and repositories worldwide in order to facilitate electronic linkages of this information (<http://grbio.org>). The accessibility to species description has been recently considered from a data storage perspective recently by Species-ID (<http://species-id.net>). This portal hosts several links to available online tools for morphological identification and hosts species page containing the basic information of a species regarding its type locality, diagnosis and location of type specimens. Riedel and colleagues [50] have recently detailed the benefit of considering species description as primary data to be stored in public portal. The same approach has been developed for phenotypes through initiative such as phenomicDB (<http://www.phenomicDB.de>) and Phenoscape (<http://kb.phenoscape.org>), which aims at storing phenotypic data for comparative purposes.

The examples above-mentioned illustrate the

benefit of global biodiversity information systems on the development of more sustainable practices in taxonomy by guaranteeing access to primary data but also to biological collections and repositories. These new tools demonstrate that taxonomic expertise and new technology are actually compatible and help making the taxonomic workflow more transparent and sustainable [50]. The increasing integration of DNA barcoding in the taxonomic workflow is expected to reinforce this trend and help trigger more connections between those large-scale initiatives of data repositories.

7 Conclusions

DNA barcoding has been subject to major development during the last decade on analytical procedures, data analyses and its application expended into a large array of biodiversity sciences and taxonomy is no exception. Aside of linking the DNA world with a traditional approach of taxonomy based on morphological characters, DNA barcoding settled new standards of data quality, accessibility and reproducibility making the use of DNA sequences in others field of biology more sustainable. After almost a decade, DNA barcoding introduced automated, fast and objective methods of biodiversity screening based on DNA sequences than opened unprecedented perspectives for the global inventory of earth living beings. DNA barcoding also challenge taxonomy and questioned several of its oldest practices regarding the description of morphological characters and the production of species keys that proved to be frequently irrelevant for the community in species rich areas such as tropical ecosystems. The use of DNA sequences not only provided objective methods for species delineation and new tools for species identification but more importantly, challenged the way we collect, keep and make biodiversity knowledge publicly available and paved the way for more sustainable practices in taxonomy.

Acknowledgements: N.H. and R.H. have been funded by the IRD and University of Guelph, respectively. We thank the anonymous reviewers for their thoughtful comments. This publication has ISEM number 2015-085.

Conflict of interest: Authors declare nothing to disclose.

References

- [1] Hebert P.D.N., Gregory T.R., The promise of DNA barcoding for taxonomy, *Syst. Biol.*, 2005, 54, 852-859

- [2] Blaxter M., Molecular systematics: Counting angels with DNA, *Nature*, 2003, 421, 122-124
- [3] Tautz D., Arctander P., Minelli A., Thomas R.H., Vogler A.P., DNA points the way ahead in taxonomy, *Nature*, 2002, 418, 479
- [4] Tautz D., Arctander P., Minelli A., Thomas R.H., Vogler A.P., A plea for DNA taxonomy, *Trends Ecol. Evol.*, 2003, 18, 70-74
- [5] May R.E., How many species are there on earth?, *Science*, 1988, 241, 1441-1449
- [6] Hebert P.D.N., Cywinska A., Ball S.L., de Waard J.R., Biological identifications through DNA barcodes, *Proc. R. Soc. London B*, 2003, 270, 313-321
- [7] Hajibabaei M., Singer G.A.C., Hebert P.D.N., Hickey D.A., DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics, *Trends Ecol. Evol.*, 2007, 23, 167-172
- [8] Miller S.E., DNA barcoding and the renaissance of taxonomy, *Proc. Nat. Acad. Sci. USA*, 2007, 104, 4775-4776
- [9] Packer L., Gibbs J., Sheffield C., Hanner R., DNA barcoding and the mediocrity of morphology, *Mol. Ecol. Resour.*, 2009, 9, 42-50
- [10] Durand J.D., Shen K.-N., Chen W.-J., Jamandre B.W., Blel H., Diop K., *et al.*, Systematics of the grey mullets (teleostei: Mugiliformes: Mugilidae): Molecular phylogenetic evidence challenges two centuries of morphology-based taxonomy, *Mol. Phylogen. Evol.*, 2012, 64, 73-92
- [11] Hebert P.D.N., Penton E.H., Burns J.M., Janzen D.H., Hallwachs W., Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *astraptes fulgerator*, *Proc. Nat. Acad. Sci. USA*, 2004, 101, 14812-14817
- [12] Smith A.M., Rodriguez J.J., Whitfield J.B., Deans A.R., Janzen D.H., Hallwachs W., *et al.*, Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections, *Proc. Nat. Acad. Sci. USA*, 2008, 105, 12359-12364
- [13] Smith M.A., Wood D.M., Janzen D.H., Hallwachs W., Hebert P.D.N., DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (diptera, tachinidae) are not all generalists, *Proc. Nat. Acad. Sci. USA*, 2007, 104, 4967-4972
- [14] Hubert N., Delrieu-Trottin E., Irisson J.O., Meyer C., Planes S., Identifying early stages of coral reef fishes through DNA barcoding : A test case with the families acanthuridae and holocentridae, *Mol. Phylogen. Evol.*, 2010, 55, 1195-1203
- [15] Ko H.-L., Wang Y.-T., Chiu T.-S., Lee M.-A., Leu M.-Y., Chang K.-Z., *et al.*, Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding, *PLoS One*, 2013, 8, e53451
- [16] Wong E.H.-K., Hanner R.H., DNA barcoding detects market substitution in north american seafood, *Food Resour. Int.*, 2008, 41, 828-837
- [17] Holmes B.H., Steinke D., Ward R.D., Identification of shark and ray fins using DNA barcoding, *Fisheries Resour.*, 2009, 95, 280-288
- [18] Hubert N., Meyer C., Bruggemann J.H., Guérin F., Komeno R.J.L., Espiau B., *et al.*, Cryptic diversity in indo-pacific coral reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis, *PLoS One*, 2012, 7, e28987
- [19] Jaafar T.N.A.M., Taylor M.I., Mohd Nor S.A., De Bruyn M., Carvalho G.R., DNA barcoding reveals cryptic diversity within commercially exploited indo-malay carangidae (teleostei: Perciformes), *PLoS One*, 2012, 7, e49623
- [20] Janzen D.H., Hajibabaei M., Burns J.M., Hallwachs W., Remigio E., Hebert P.D.N., Wedding biodiversity inventory of a large and complex lepidoptera fauna with DNA barcoding, *Phil. Trans. R. Soc. B*, 2005, 360, 1835-1845
- [21] Fouquet A., Gilles A., Vences M., Marty C., Blanc M., Gemmill N.J., Underestimation of species richness in neotropical frogs revealed by mtDNA analyses, *PLoS One*, 2007, 10, e1109
- [22] Kerr K.C.R., Lijtmaer D.A., Barreira A.S., Hebert P.D.N., Tubaro P.L., Probing evolutionary patterns in neotropical birds through DNA barcodes, *PLoS One*, 2009, 4, e4379
- [23] Forest F., Grenyer R., Rouget M., Davies J., Cowling R.M., Faith D.P., *et al.*, Preserving the evolutionary potential of floras in biodiversity hotspots, *Nature*, 2007, 445, 757-760
- [24] Ardura A., Linde A.R., Moreira J.C., Garcia-Vazquez E., DNA barcoding for conservation and management of amazonian commercial fish, *Biol. Conserv.*, 2010, 143, 1438-1443
- [25] Armstrong K.F., Ball S.L., DNA barcodes for biosecurity: Invasive species identification, *Phil. Trans. R. Soc. B*, 2005, 360, 1813-1823
- [26] Floyd R., Lima J., deWaard J.R., Humble L., Hanner R., Common goals: Policy implications of DNA barcoding as a protocol for identification of arthropod pests, *Biol. Invasions*, 2010, 12, 2947-2954
- [27] Stribling J.B., Environmental protection using DNA barcodes or taxa?, *Bioscience*, 2006, 56, 878-879
- [28] Vernooij R., Haribabu E., Muller M.R., Vogel J.H., Hebert P.D.N., Schindel D.E., *et al.*, Barcoding life to conserve biological diversity: Beyond the taxonomic imperative, *PLoS Biol.*, 2010, 8, e1000417
- [29] Godfray H.C.J., To boldly sequence, *Trends Ecol. Evol.*, 2006, 21, 503-504
- [30] Godfray H.C.J., Linnaeus in the information age, *Nature*, 2007, 417, 17-19
- [31] Kingman J.F.C., The coalescent, *Stochastic Process App.*, 1982, 13, 245-248
- [32] Hudson R., Testing the constant rate neutral allele model with protein sequence data, *Evolution*, 1982, 37, 203-217
- [33] Tajima F., Evolutionary relationships of DNA sequences in finite populations, *Genetics*, 1983, 105, 437-460
- [34] Agnarsson I., Kuntner M., Taxonomy in a changing world: Seeking solutions for a science in crisis, *Syst. Biol.*, 2007, 56, 531-539
- [35] Ratnasingham S., Hebert P.D.N., Bold: The barcode of life data system (<http://www.Barcodinglife.Org>), *Mol. Ecol. Notes*, 2007, 7, 355-364
- [36] April J., Mayden R., L., Hanner R.H., Bernatchez L., Genetic calibration of species diversity among north america's freshwater fishes, *Proc. Nat. Acad. Sci. USA*, 2011, 108, 10602-10607
- [37] Pereira L.H.G., Hanner R., Foresti F., Oliveira C., Can DNA barcoding accurately discriminate megadiverse neotropical freshwater fish fauna ?, *BMC Genetics*, 2013, 14, 20
- [38] Weigt L.A., Baldwin C.C., Driskell A., Smith D.G., Ormos A., Reyier E.A., Using DNA barcoding to assess caribbean reef fish biodiversity: Expanding taxonomic and geographic coverage, *PLoS One*, 2012, 7, e41059
- [39] Kerr K.C., Stoeckle M.Y., Dove C.J., Weigt L.A., Francis C.M., Hebert P.D.N., Comprehensive DNA barcode coverage of north american birds, *Mol. Ecol. Notes*, 2007, 7, 535-543

- [40] Hebert P.D.N., deWaard J.R., Zakharov E., Prosser S.W.J., Sones J.E., McKeown J.T.A., *et al.*, A DNA 'barcode blitz': Rapid digitization and sequencing of a natural history collection, *PLoS One*, 2013, 8, e68535
- [41] Ride W.D.L., Cogger H.G., Dupuis C., Kraus O., Minelli A., Thompson F.C., *et al.*, International code of zoological nomenclature, ed Nomenclature T.I.T.f.Z., National History Museum, London, 2012
- [42] Witt J.D.S., Threlloff D.L., Hebert P.D.N., DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: Implications for desert spring conservation, *Mol. Ecol.*, 2006, 15, 3073-3082
- [43] Smith A.M., Fisher B.L., Hebert P.D.N., DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of madagascar, *Phil. Trans. R. Soc. B*, 2005, 360, 1825-1834
- [44] Kadarusman, Hubert N., Hadiaty R.K., Sudarto, Paradis E., Pouyaud L., Cryptic diversity in indo-australian rainbowfishes revealed by DNA barcoding: Implications for conservation in a biodiversity hotspot candidate., *PLoS One*, 2012, 7, e40627
- [45] Will K., W., Rubinoff, D., Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification, *Cladistics*, 2004, 20, 47-55
- [46] Ebach M.C., Holdredge C., DNA barcoding is no substitute for taxonomy, *Nature*, 2005, 434, 697
- [47] Will K.W., Mishler B.D., Wheeler Q.D., The perils of DNA barcoding and the need for integrative taxonomy, *Syst. Biol.*, 2005, 54, 844-851
- [48] Monaghan M.T., Wild R., Elliot M., Fujisawa T., Balke M., Inward D., *et al.*, Accelerated species inventories on madagascar using coalescent-based models of species delineation, *Syst. Biol.*, 2009, 58, 298-311
- [49] Butcher B.A., Smith M.A., Sharkey M.J., Quicke D.L.J., A turbo-taxonomic study of thai aleiodes (aleiodes) and aleiodes (arcaleiodes) (hymenoptera: Braconidae: Rogadinae) based largely on coi barcoded specimens, with rapid descriptions of 179 new species, *Zootaxa*, 2012, 3457, 1-232
- [50] Riedel A., Sagata K., Suhardjono Y.R., Tänzler R., Balke M., Integrative taxonomy on the fast track - towards more sustainability in biodiversity research, *Frontiers Ecol.*, 2013, 10, 15
- [51] Hubert N., Hanner R.H., Holm E., Mandrak N.E., Taylor E.B., Burrige M., *et al.*, Identifying canadian freshwater fishes through DNA barcodes, *PLoS One*, 2008, 3, e2490
- [52] Emerson B., C., Paradis, E., Thébaud, C., Revealing the demographic histories of species using DNA sequences, *Trends Ecol. Evol.*, 2001, 16, 707-716
- [53] Birky C.W.J., Fuerst, P., Maruyama, T., Organelle gene diversity under migration, and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes, *Genetics*, 1989, 121, 613-627
- [54] Pamlilo P., Nei, M., Relationships between gene trees and species trees, *Mol. Biol. Evol.*, 1988, 5, 568-581
- [55] Nielsen R., Wakeley, J., Distinguishing migration from isolation: A markov chain approach, *Genetics*, 2001, 158, 885-896
- [56] Funk D.J., Omland K.E., Species-level paraphyly and polyphyly: Frequency, causes and consequences, with insights from animal mitochondrial DNA, *Ann. Rev. Ecol. Syst.*, 2003, 34, 397-423
- [57] Hajibabaei M., Janzen D.H., Burns J.M., Hallwachs W., Hebert P.D.N., DNA barcodes distinguish species of tropical lepidoptera, *Proc. Nat. Acad. Sci. USA*, 2006, 103, 968-971
- [58] Lamoureux J.F., Morrison J.C., Ricketts T.H., Olson D.M., Dinerstein E., McKnight M., *et al.*, Global tests of biodiversity concordance and the importance of endemism, *Nature*, 2006, 440, 212-214
- [59] Weir J.T., Schluter D., The latitudinal gradient in recent speciation and extinction rates of birds and mammals, *Science*, 2007, 315, 1574-1576
- [60] Wiens J.J., Donoghue M.J., Historical biogeography, ecology and species richness, *Trends Ecol. Evol.*, 2004, 19, 639-644
- [61] Ratnasingham S., Hebert P.D.N., A DNA-based registry for all animal species: The barcode index number (bin) system, *PLoS One*, 2013, 8, e66213
- [62] Ward R.D., Hanner R.H., Hebert P.D.N., The campaign to DNA barcode all fishes, fish-bol, *J. Fish Biol.*, 2009, 74, 329-356
- [63] Hebert P.D.N., Stoeckle M.Y., Zemlak T.S., Francis C.M., Identification of birds through DNA barcodes, *PLoS Biol.*, 2004, 2, e312
- [64] Meyer C., Paulay G., DNA barcoding: Error rates based on comprehensive sampling, *PLoS*, 2005, 3, 2229-2238
- [65] Ward R.D., Zemlak T.S., Innes B.H., Last P.R., Hebert P.D.N., DNA barcoding australia's fish species, *Phil. Trans. R. Soc. B*, 2005, 360, 1847-1857
- [66] Collins R.A., Armstrong K.F., Meier R., Yi Y., Brown S.D.J., Cruickshank R.H., *et al.*, Barcoding and border biosecurity: Identifying cyprinid fishes in the aquarium trade, *PLoS One*, 2012, 7, e28381
- [67] Hanner R., Becker S., Ivanova N.V., Steinke D., Fish-bol and seafood identification: Geographically dispersed case studies reveal systemic market substitution across canada, *Mit. DNA*, 2011, 22, 106-122
- [68] Francis C.M., Borisenko A.V., Ivanova N.V., Eger J.L., Lim B.K., Guillion-Servent A., *et al.*, The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia, *PLoS One*, 2010, 5, e0012575
- [69] Rasmussen R.S., Morrissey M.T., Hebert P.D.N., DNA barcoding of commercially important salmon and trout species (onchorhynchus and salmo) from north america, *J. Agri. Food Chem.*, 2009, 57, 8379-8385
- [70] Puillandre N., Strong E.E., Bouchet P., Boisselier M.-C., Couloux A., Samadi S., Identifying gastropod spawn from DNA barcodes: Possible but not yet practicable, *Mol. Ecol. Resour.*, 2009, 9, 1311-1321
- [71] Richard B., Decaens T., Rougerie R., James S.W., Porco D., Hebert P.D., Re-integrating earthworm juveniles into soil biodiversity studies: Species identification through DNA barcoding, *Mol. Ecol. Resour.*, 2010, 10, 606-614
- [72] Stur E., Ekrem T., Exploring unknown life stages of arctic tanytarsini (diptera: Chironomidae) with DNA barcoding, *Zootaxa*, 2011, 2743, 27-39
- [73] Rougerie R., Laguerre M., A striking case of sexual dimorphism revealed by the analysis of DNA barcodes in the arctiid moth *Senecauxia coraliae* de toulgoët, 1990 from French Guiana (Lepidoptera: Arctiidae), *Annales de la Sociétés Entomologique de France*, 2010, 46, 477-480
- [74] Austerlitz F., David O., Schaeffer B., Bleakley K., Olteanu M., Leblais R., *et al.*, DNA barcode analysis: A comparison

- of phylogenetic and statistical classification methods, *BMC Bioinf.*, 2009, 10, S10
- [75] Kekkonen M., Hebert P.D.N., DNA barcode-based delineation of putative species: Efficient start for taxonomic workflows, *Mol. Ecol. Resour.*, 2014, 14, 706-715
- [76] Brown S.D.J., Collins R.A., Boyer S., Lefort C., Malumbres-Olarte J., Vink C.J., *et al.*, Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding, *Mol. Ecol. Resour.*, 2012, 12, 562-565
- [77] Puillandre N., Modica M.V., Zhang Y., Sirovitch L., Boisselier M.-C., Cruaud C., *et al.*, Large scale species delimitation method for hyperdiverse groups, *Mol. Ecol.*, 2012, 21, 2671-2691
- [78] Puillandre N., Lambert A., Brouillet S., Achaz G., Abgd, automatic barcode gap discovery for primary species delimitation, *Mol. Ecol.*, 2012, 21, 1864-1877
- [79] Munch K., Boomsma W., Huelsenbeck J., P., Willerslev E., Nielsen R., Statistical assignment of DNA sequences using Bayesian phylogenetics, *Syst. Biol.*, 2008, 57, 750-757
- [80] Munch K., Boomsma W., Willerslev E., Nielsen E.G., Fast phylogenetic DNA barcoding, *Proc. R. Soc. London B*, 2008, 363, 3997-4002
- [81] Abdo Z., Golding G.B., A step toward barcoding life: A model-based, decision-theoretic method to assign genes to preexisting species groups, *Syst. Biol.*, 2007, 56, 44-56
- [82] David O., Larédo C., Leblois R., Schaeffer B., Vergne N., Coalescent-based DNA barcoding: Multilocus analysis and robustness, *J. Comput. Biol.*, 2012, 19, 271-278
- [83] Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., *et al.*, Sequence-based species delimitation for the DNA taxonomy of undescribed insects, *Syst. Biol.*, 2006, 55, 595-606
- [84] Yule G.U., A mathematical theory of evolution based on the conclusions of Dr J.C. Willis, *Phil. Trans. R. Soc. B-Biol. Sci.*, 1924, 213, 21-87
- [85] Bergsten J., Bilton D.T., Fujisawa T., Elliott M., Monaghan M.T., Balke M., *et al.*, The effect of geographical scale of sampling on DNA barcoding, *Syst. Biol.*, 2012, 61, 851-869
- [86] Geiger M.F., Herder F., Monaghan M.T., Almada V., Barbieri R., Bariche M., *et al.*, Spatial heterogeneity in the Mediterranean biodiversity hotspot affects barcoding accuracy of its freshwater fishes, *Mol. Ecol. Resour.*, 2014, 14, 1210-1221
- [87] Hickerson M.J., Meyer C., Moritz C., DNA barcoding will often fail to discover new animal species over broad parameter space, *Syst. Biol.*, 2006, 55, 729-739
- [88] Hollingsworth P.M., Graham C.H., Little D.P., Choosing and using a plant DNA barcode, *PLoS One*, 2011, 6, e19254