Accepted on 29 July 2006
J Zool Syst Evol Res doi: 10.1111/j.1439-0469.2006.00384.x

[1]*Department of Entomology, Natural History Museum, London, UK;* [2]*Department of Biology, Imperial College London, Ascot, UK*

Recent advances in DNA taxonomy

A. P. Vogler[1, 2] and M. T. Monaghan[1, 2]

Abstract

Large-scale DNA sequencing of living species holds great promise in taxonomy, but has been controversial. In this article, we review the recent advances that follow the dramatic increase in data generation. We distinguish DNA taxonomy from DNA barcoding, where the former directly concerns the circumscription and delineation of species using evolutionary species concepts and the latter is a means of identifying *a priori* entities by sequence similarity. A key finding from recent studies in animals is that variation in mitochondrial DNA (mtDNA) is partitioned as tight clusters of closely related genotypes, which group specimens largely according to traditionally recognized species limits, and which are congruent with nuclear markers. This finding provides confidence to use sequence variation as the primary information for species delimitation in poorly known groups. A number of recent, large-scale studies support the power of mtDNA in species recognition, and previous application of molecular techniques to taxonomically complicated cases has likely led to an overestimate of the proportion of species with polyphyletic mtDNA haplotypes. The continued development of DNA taxonomy will lead to more refined sampling strategies and data analyses than those that are presently used. Sophisticated statistical methods of grouping have already been developed based on sequence similarity; yet, the units defined in this way have largely unknown evolutionary relevance. In future, a standard DNA taxonomic analysis will include broad sampling of the target taxa across their geographic range, followed by large-scale sequencing of representative samples for a DNA profile of the group, and algorithmic procedures for delineating species limits. The taxonomic system will be derived from the data rather than expert opinion, and hypothesized species entities can be tested against morphology, biogeography and other data, providing an evolutionary justification of the procedures used for species delimitation. Discrepancies between DNA and other data are used to refine species delimitations via a feedback loop that incorporates new data. We argue, however, that the use of DNA methodology in taxonomy (including DNA barcoding) will remain controversial until it is better founded in existing theory of evolutionary biology and phylogenetics.

Key words: Species concepts – molecular taxonomy – DNA barcoding – molecular operational taxonomic units

Introduction

Bold new approaches are needed to modernize taxonomy (Godfray 2002; Wilson 2003; Janzen 2004), and DNA-based methods have the potential to provide the much-needed quantum leap in the speed and precision of taxonomic procedures. The use of DNA sequences in taxonomy dates back 30 years to when ribosomal RNA probes were developed for the identification and phylogenetics of eubacteria and archaebacteria (Fox et al. 1980). Molecular tools have been widely used for species separation and identification throughout the past two decades (e.g. Baker and Palumbi 1994; Sperling et al. 1994; DeSalle and Birstein 1996) and one of the earliest uses of the term 'molecular taxonomy' appeared in this journal (Scherer and Sontag 1986). A drastic increase of activity in the field is underway at present, following recent proposals for the creation of sequence databases that represent all or most living species on Earth (Hebert et al. 2003a; Tautz et al. 2003). This large-scale application of molecular data is clearly bound to revolutionize taxonomy (Savolainen et al. 2005), but the validity and practicalities of molecular approaches to taxonomy have been subject to a variety of criticisms (e.g. Lipscomb et al. 2003; Moritz and Cicero 2004; Will and Rubinoff 2004; Wheeler 2005; Crisci 2006).

In this article, we review the recent advances in this fast-developing field of applying DNA sequence data in taxonomic practice. We distinguish three general approaches to data collection and analysis and suggest that much confusion and criticism regarding DNA-based taxonomy has arisen from imprecise use of terminology and the diverse aims of various approaches. This is aggravated by the omission of rigorous evolutionary theory, particularly with regard to the underlying species concepts (e.g. Moritz and Cicero 2004; Wheeler 2004; Will et al. 2005). As it will be shown, species delimitation is of central importance wherever DNA approaches are to fulfil the role of morphology-based taxon concepts of the current Linnean system. Rigorous application of evolutionary theory to species circumscription will, in turn, shed light on unresolved secondary issues in DNA taxonomy, such as sampling strategy of specimens and the choice of suitable gene regions. It is concluded that recent results from a number of large-scale DNA sequencing studies promise a great advance for taxonomic practice as a whole. The field is now at a stage where much of the early criticism can be rejected and procedures need to be formalized, which link DNA-based information to the existing taxonomic system.

Recent applications of molecular sequence data in taxonomy

Most recent studies can be grouped into three general approaches that are referred to as DNA taxonomy, DNA barcoding and molecular operational taxonomic units (MOTU) delineation in this article. The terms themselves sometimes lack a clear definition in the literature, and some confusion has arisen from their inconsistent application. A major distinction should be made between species identification, generally associated with the idea of 'molecular barcodes', and species circumscription and delineation, broadly referred to as 'DNA taxonomy'. Defining of each approach and illustration of the different procedures and aims inherent to each are discussed in the following sections. An important distinction is to be made in the treatment of the individual organisms as the basic items in these analyses, and the taxonomic entities into which these individuals are

grouped. We refer to the individuals as 'collection objects' (or 'catalogue objects'), as they represent the objects in a physical collection or the specimen entries in a database, unlike the 'taxonomy objects', which correspond to the Linnean binomials under which the individuals are subsumed (Fig. 1). The impact of DNA sequences will also propagate upward in the classification hierarchy, based on phylogenetic interpretation of DNA taxonomy data (e.g. Franz 2005; Sereno 2005). Yet, as the recent activities have focussed largely on species-level aspects of taxonomy, we limit our discussion to this lower level of classification only. We first introduce three main approaches to the interpretation of sequence data in species delimitation and identification.

Molecular operational taxonomic units

MOTUs were proposed by Floyd et al. (2002) in an effort to use sequence similarity for grouping morphologically cryptic meiofauna (e.g. nematodes from soil samples) into genetically defined entities. A MOTU is defined as a group of sequences that differed from one another by a maximum number of base pairs, e.g. two or three nucleotides in a 500-bp region of 18S rRNA (Floyd et al. 2002). Molecular surveys of this type have revealed a large amount of DNA diversity in locally collected samples (e.g. Bensch et al. 2004). Yet, it is unclear how well MOTU diversity corresponds to species or ecological diversity. To date, MOTUs have not been put in the context of established species concepts, and it remains to be tested if

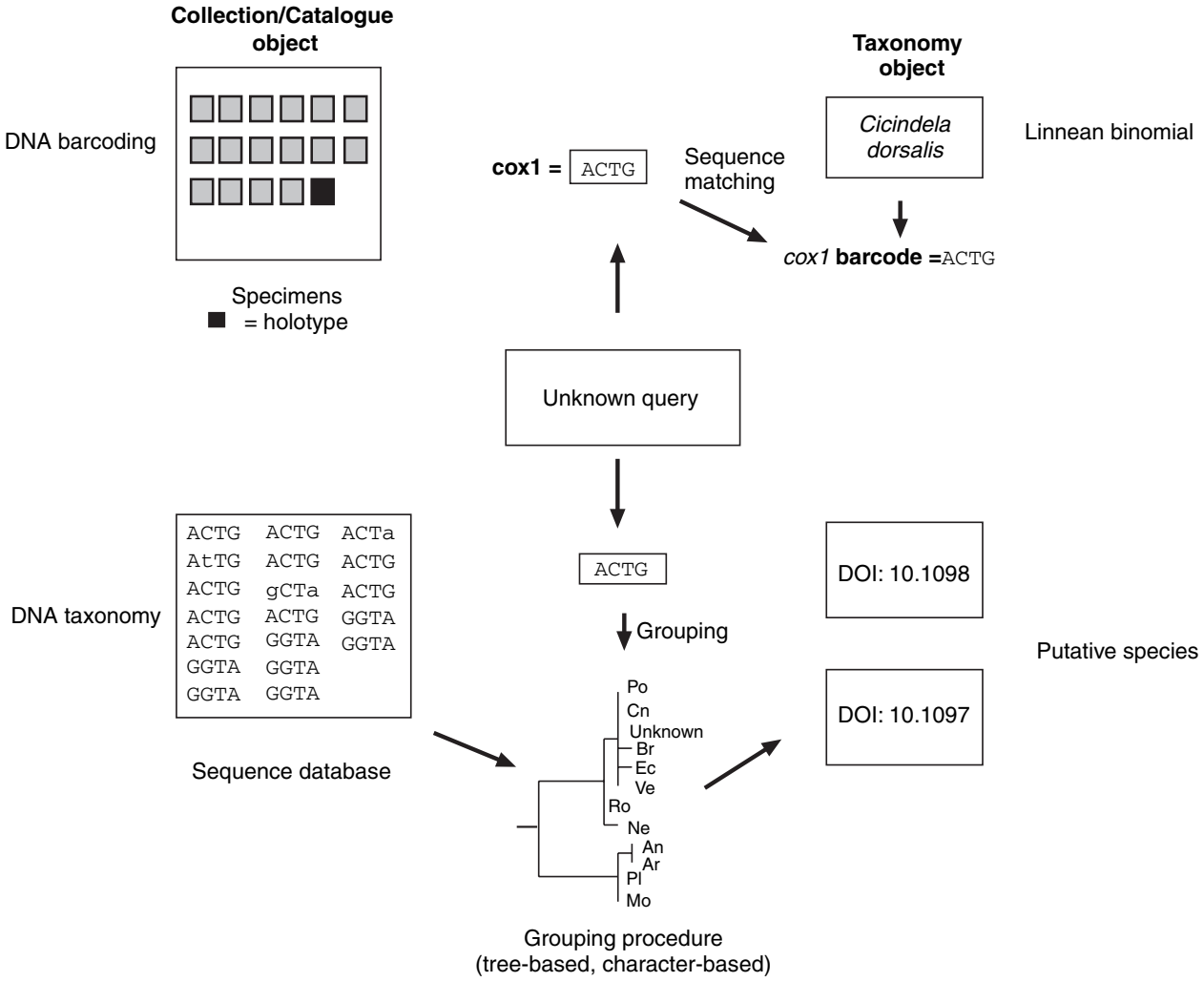

Fig. 1. Schematic illustration of the differences between DNA barcoding and DNA taxonomy. We make a distinction between collection/catalogue objects (specimen entries in an inventory, corresponding to specimens in a physical collection) and taxonomy objects (the names in a taxonomy). (1) In traditional taxonomy, the collection/catalogue objects include the holotype and are grouped based on morphological traits and assigned a species name; the taxonomy objects are the Linnean binomials (and other terms of the higher classification). Specimen(s) identified as members of a given species are then sequenced to provide the 'barcode'. The identity of a query sequence obtained from an unknown individual (in italics, underlined) is determined by a match to the database of barcodes. (2) DNA taxonomy uses the sequences as the primary catalogue/collection objects, which are grouped to represent the taxonomy objects. As these sequences might differ slightly (small letters in the sequence), a range of grouping procedures can be used to identify the species-level entities from the sequence information (see main text). To illustrate this, we show a tree with terminals labelled with arbitrary locality codes. Identification of unknowns is against this set of sequences, i.e. they are included in the same kind of grouping procedure that established these groups in the first place. The group so defined can be assigned any type of name (including a digital object identifier, DOI). Many sets of sequences will correspond to existing Linnean names. Note that the MOTU approach is similar to the latter procedure (the example given uses a cut-off of a 3-bp to delimit MOTUs if the short branches correspond to 1 bp), with the main distinction that the MOTU is not understood to constitute an evolutionarily justified species

they are equivalent to the species-level groups identified in traditional taxonomy. More recently, 'reverse taxonomy' (Markmann and Tautz 2005) employs similar methodologies in what is essentially a first step in taxonomy – identifying the level of diversity present and provide a starting point for further biological analysis (Blaxter 2004).

DNA barcoding

DNA barcoding was proposed by Hebert et al. (2003a) as a method for identifying unknown specimens. Short mitochondrial DNA (mtDNA) sequences (usually the 5′ half of the *cox1* gene) are used to group unknown individuals with *a priori*-defined taxonomic entities based on sequence similarity, deriving a species identification from DNA rather than morphological characters. In the terminology introduced earlier (Fig. 1), the collection objects are subsumed under a Linnean name (the taxonomy object) by means of traditional procedures, and the sequence is fitted to the taxonomy object retrospectively. Inexact matches are either grouped with taxa present in the database or identified as new to the database based on whether they fall within a threshold of sequence similarity. This is justified by the observation that variation among species is normally lower than interspecies variation (Hebert et al. 2003b; Hebert and Gregory 2005).

While intuitively appealing, the notion of low intraspecies versus high interspecies divergences is not borne out in many groups, as intrapopulation variation may exceed divergences between species (Avise 2000), in particular where the evolution of reproductive barriers is sudden because of sexual selection, ecological shifts, chromosomal rearrangements and other factors (Coyne and Orr 1998). As such, DNA barcoding is not predictive, i.e. it fails when an identical sequence is not available and a limit for admissible divergence has not been established. Hence, DNA barcoding is limited in its potential, as it requires a near complete database of vouchers against which individuals can be placed (Moritz and Cicero 2004; Will and Rubinoff 2004). Lineage-specific cut-off values of intraspecific sequence divergence have been proposed as a rule-of-thumb to remedy this problem, but calibrations of this 'barcoding gap' are problematic (Meyer and Paulay 2005) and the magnitude of sequence divergence will vary among lineages (Hebert and Gregory 2005). Thus, the approach can function only as an identification tool – a by-product of a classification system established in a traditional way – but not as a taxonomic system itself.

DNA taxonomy

In contrast, DNA taxonomy in the strict sense would refer to the notion that the DNA sequences themselves serve as the taxonomic reference system. Following this idea, the DNA sequences constitute the catalogue objects, from which the taxonomy objects have to be derived (Fig. 1). The latter are groups of sequences which take on a role equivalent to Linnean binomials in the traditional taxonomy (i.e. they serve as the term to which biological information is being associated; see Thiele and Yeates 2002). Given variation between individuals, the grouping procedures are a critical step. Rather than accepting any arbitrary groups, e.g. MOTUs of a particular cut-off, the aim in DNA taxonomy is to identify groups that correspond to entities of reproductively coherent individuals (the species), i.e. to determine a hierarchical level roughly equivalent to the binomials of the traditional system (which are generally considered to represent the true species in nature). While recourse to the Linnean nomenclature provides important evidence for the correct level of the taxonomic hierarchy, it does not follow that the DNA groups are only valid if they correspond precisely to the existing species names; this has been frequently misunderstood in the recent literature (Wheeler 2004; Meyer and Paulay 2005).

For some taxa, a growing DNA taxonomy is now available (e.g. Powers 2004; Verbruggen et al. 2005), where sequences serve as the principal means of linking vouchers and collateral information (and hence already represent the primary portrayal of the group for communication). Once in place, the DNA taxonomy also provides a framework for routine identification, and can then serve as the primary database for standard DNA barcoding. Yet, in contrast to DNA barcoding in the strict sense (species identification against a database of mitochondrial *coxI* sequences), a DNA taxonomy for a particular group of organisms may be based on one or more regions of mtDNA or nuclear DNA, and can be derived from phylogenetic and clustering methods using any gene region (Pons et al. 2006). In addition, the sequences define hierarchical groupings of living organisms into which all species can be included at some level (Tautz et al. 2003; Savolainen et al. 2005). Therefore, the DNA taxonomy can also be useful for fitting unknown or unrepresented species into the database using phylogenetic information. Because confidence in many nodes of trees derived from a short sequence fragments may be low, the sequencing of multiple genes is desirable for this purpose.

A review of progress to date

As the activities to generate DNA data for entire taxonomic groups are increasing rapidly, major questions about the feasibility and consequences of these approaches can now be answered. Large-scale taxonomic sequencing has been performed on various groups of metazoan animals, including nematodes (Floyd et al. 2002), cowries (Meyer and Paulay 2005), tardigrades (Blaxter et al. 2004), chelicerates (Barrett and Hebert 2005), Lepidoptera (Hebert et al. 2003a; Hajibabaei et al. 2006), birds (Hebert et al. 2004b) and fishes (Ward et al. 2005), as well as fungi, red algae (Saunders 2005) and preliminary studies on flowering plants (Kress et al. 2005). Many of these studies are based on local samples (single populations) of target groups and hence are taxonomically incomplete, i.e. they only include a subset of species from a more widely distributed phylogenetic lineage. Local samples also rarely encompass the genetic variation found among populations. As a result, they are a useful starting point for more complete DNA databases in the future, but their contribution to the test of specific taxonomic questions may be limited (Moritz and Cicero 2004; Prendini 2005). In addition to these broad-based sequencing studies that did not pursue specific taxonomic questions, recent DNA analyses have also addressed narrow problems on particular lineages, e.g. the taxonomy of a group based on larval and adult individuals (Paquin and Hedin 2004; Scheffer et al. 2006); the correspondence of ecomorphological types with mtDNA groups (Hebert et al. 2004a; Smith et al. 2006); the geographical patterns of diversity in taxonomically poorly studied insects (Smith et al. 2005); the match of morphologically recognized entities and mtDNA in various species complexes

(Cardoso and Vogler 2005; Gompert et al. 2006); and the DNA 'profiling' of unknown faunas (Markmann and Tautz 2005; Monaghan et al. 2006; Pons et al. 2006). These studies pertain to different streams of inquiry and together with the broader surveys of sequencing studies described earlier are useful to evaluate the prospects for the proposed DNA-based approaches to taxonomy.

The key finding of studies to date is that the vast majority of sequence variation in nature is partitioned into clearly defined clusters that are easily visualized using a variety of tree building methods. There are now several well-documented cases where morphologically or cytologically defined taxa align closely with mtDNA clusters (Hugall et al. 2002; Hebert et al. 2003b, 2004b; Dalebout et al. 2004; Hogg and Hebert 2004; Paquin and Hedin 2004; Sharley et al. 2004; Barrett and Hebert 2005; Hebert and Gregory 2005; Meyer and Paulay 2005; Monaghan et al. 2005; Page et al. 2005; Ball and Armstrong 2006), in agreement with well-established findings from earlier phylogeographic analyses (Avise and Walker 1999). A figure of 97.8% has been cited for the proportion of species distinguished by prior taxonomic work that can be separated based on unique mtDNA sequences in tropical moths (Hajibabaei et al. 2006). Other studies have demonstrated a close match of mtDNA with nuclear markers (Bensch et al. 2004; Gaines et al. 2005; Monaghan et al. 2005; Smith et al. 2006), providing further evidence for the utility of mtDNA as a proxy for species circumscription. Where sequences attributed to a single Linnean species name were found to be highly divergent, subsequent revision of the morphological evidence in some cases suggested the existence of hitherto unrecognized species-level forms that had been overlooked (Hebert et al. 2004a,b).

Incongruence in DNA-based taxonomy

Coincident to many of the studies cited earlier, a number of concerns have been raised as to the validity of equating mtDNA clusters with species boundaries. First, the clusters may simply represent an artefact because of insufficient taxonomical and geographical sampling; clusters may collapse, once closely related species are included and geographic variation is fully quantified (Moritz and Cicero 2004; Sperling 2004; Prendini 2005). Secondly, the mtDNA clusters may represent an incorrect image of species boundaries because of the stochastic processes affecting single genetic loci, their female-limited mode of inheritance, and the apparently greater propensity of mtDNA for gene flow across otherwise separated gene pools (Mallet and Willmott 2003; Will and Rubinoff 2004).

Focussing first on the argument of incomplete sampling, several studies have now provided a wider geographical and taxonomical representation of focal groups, whereby the strong clustering tends to be preserved when species are sampled across their ranges. Moths (*Acronicta* spp.) from distant localities in North America still retained a tight clustering in phylogenetic analysis, unequivocally associating individuals as members of these clusters (Hebert and Gregory 2005). In a densely sampled radiation of tiger beetles across interior Australia covering most known populations for some 50 species, mtDNA clusters remained clearly recognizable, despite the narrow geographical and taxonomical sampling, and clusters showed restricted geographical distribution indicating their internal coherence while being geographically and phylogenetically separated from other such clusters (Pons et al. 2006). Similarly, mtDNA clusters corresponded closely to ecological and behavioural traits pertaining to host specificity and habitat parameters in moths and parasitic flies, also supporting the validity of mtDNA clusters to reveal biologically meaningful groupings (Hebert et al. 2004a; Smith et al. 2006). Finally, mtDNA sequencing for a nearly complete fauna of day-flying butterflies (Papilionoidea and Hesperoidea) of Madagascar (D. Lees and A. P. Vogler, unpublished data), including three large radiations of up to 60 species each and additional sampling of geographically and morphologically divergent populations, did not reveal a single case where mtDNA would have resulted in groupings that were inconsistent with morphologically defined species (although in several cases species exhibited indistinguishable genotypes). Based on these examples, it can be expected that denser geographical and taxonomical sampling may result in the discovery of new clusters, and perhaps reduce their divergence from each other, but they are unlikely to erode the clustering altogether.

The high degree of congruence of mtDNA groups and traditionally defined taxa appears to contradict the reported mismatch of established species boundaries and mtDNA distributions attributable to stochastic lineage sorting and hybridization. According to Funk and Omland's (2003) widely cited review based on broad surveys of the animal literature, a staggering 23% of cases showed mtDNA to be polyphyletic with respect to the named species entities. A similar figure was obtained by Meyer and Paulay (2005) for cowries. Indeed, an 'error rate' (Meyer and Paulay 2005) of this magnitude would be entirely unacceptable for a taxonomic system of general practical utility. Yet, the notion of incongruence of gene tree and species tree assumes the existence of at least two data sets with features amenable to formal comparison. Where traditional classifications are based on non-explicit or intuitive data interpretations, incongruence of a gene tree with the 'true' clade history may be concluded without actually testing conflict (Brower et al. 1996). A taxon circumscription that has a limited empirical basis is easily refuted by any hypothesis that contradicts it, and in these cases of 'incongruence' the gene tree should be provisionally accepted as the best available hypothesis of taxonomic grouping (Brower et al. 1996). Tests of congruence require, by definition, an external reference and are thus entirely dependent on the morphological entity being delimited (i.e. accuracy of the taxonomy) and accuracy of the identifications. Differences in species concepts and their implementation also play a large role in species delimitation. Literature surveys revealed a 48.7% higher count of species with the application of a phylogenetic species concept compared with studies applying the biological species concept on the same organisms (Agapow et al. 2004). Therefore, redefinition of taxa may frequently be necessary and may greatly alter the results of 'tests' of DNA barcoding. Where the nomenclatural framework is relaxed and molecular evidence has been taken into account, e.g. by establishing more realistic groupings referred to as 'evolutionarily significant units' (Meyer and Paulay 2005), the number of polyphyletic species is reduced, indicating that even well-studied groups may be in need of taxonomic revision before accurate tests of incongruence can be conducted.

This should not ignore the many established cases of reticulate evolution and lack of discrete entities which result in ambiguity of species delineation. Yet, this phenomenon appears mostly limited to a small set of 'taxonomically

complex groups' (Ennos et al. 2005), which share properties resulting from peculiar genetic systems, hybridogenetic origin or partial gene flow in zones of contact. The issue is also usually confined to lineages where the divergences between species are small, as is the case in species radiations on islands (Gillespie and Roderick 2002; Monaghan et al. 2006), in species complexes consisting of multiple subgroups (e.g. Cardoso and Vogler 2005; Gompert et al. 2006), or in morphologically conservative groups (Hebert et al. 2004a). These cases are likely to be a minority (Besansky et al. 2003; Hajibabaei et al. 2006), but the prevalence of applying molecular techniques in these complicated cases will mean that literature surveys overestimate the proportion of polyphyletic species.

The validity of short sequence fragments as species markers

A final argument raised against DNA-based taxonomy is the claim that a single short piece of sequence is insufficient to represent the complexities of species-level differences, while morphology as an amalgam of many evolutionary differences is greatly more informative (Lipscomb et al. 2003; Mallet and Willmott 2003; Wheeler 2004; Will and Rubinoff 2004; Brower 2006). In addition, such morphologically defined species have a greater chance of being reproductively isolated, because of the complexity of characters that may also be involved directly in reproductive isolation, preventing the anastomosis of separated groups in the future (Q. D. Wheeler, pers. comm.).

However, assumptions about the characters used to define the species go beyond what most taxonomic research is usually able to achieve. Species delineation, unless studied directly by some measure of reproductive compatibility (Mayr 1942), is based on the pattern of character variation to recognize ('diagnose') geographically or otherwise defined hypothetical groups (Cracraft 1983; Sites and Marshall 2003). The evidence is indirect: it is inferred that groups are separated, because consistent character differences only exist if gene flow and recombination is suppressed for extended periods. The observed character variation is a proxy for the recognition of historically separated groups. Even in the rare cases where diagnostic characters can be implicated in particular biological functions or reproductive incompatibility (e.g. genitalic differences), taxonomy uses this information only to inform on past processes indicating population separation. DNA sequences reflect these patterns of evolutionary separation in the same way as morphological characters.

The fact that these sequence fragments are comparatively short is equally irrelevant: even a very short sequence down to a single nucleotide change (Goldstein and Desalle 2003), or a combination of sites providing a unique diagnostic sequence signature (DeSalle et al. 2005), may fulfil the requirement for diagnostic changes, and any part of the genome (mtDNA and any nuclear marker) can potentially serve in this function. The sequence information is used for grouping according to historical separation of populations, rather than based on an understanding of biological functions or 'speciation genes' separating the groups. Importantly, diagnosable groups do not represent some arbitrary level of the evolutionary hierarchy; because of the population-based procedures for species delimitation (below) these groups are likely to correspond to the species (i.e. a unique level of biological organization that results from evolutionary processes acting on populations and producing discontinuities; Brower et al. 1996, Coyne and Orr 1998). Specifically, in the absence of gene flow among populations, haplotypes diverge among (geographically or reproductively separated) species, and thus estimates of the coalescent within a species should be accelerated in mtDNA compared with any nuclear encoded marker because of the smaller effective population size (N_e) of the former. Frequent selective sweeps acting on mtDNA appear to increase the depth of subdivision and ease of recognition further (Bazin et al. 2006). But whatever the exact biological processes, taxonomy assesses the distribution of character variation, whereas the kind of variation (or the magnitude, such as a percentage of sequence divergence) is of secondary importance.

Interpreting molecular evidence: the practice of DNA-based taxonomy

Having established the evolutionary framework of DNA taxonomy, what remains is to formalize how it can be implemented in everyday taxonomic practice. The strong clustering of sequence variation is the key observation of existing research, and provides the basis for species recognition and delimitation. Yet, the literature to date is vague about how species boundaries should be interpreted. A prerequisite is that sequence variation is sufficiently high to encounter genetic polymorphisms separating the species. This is not always the case, in particularly in plants (Kress et al. 2005) or where slowly evolving genes have been used in animals (Monaghan et al. 2005). Secondly, the identification of presumed species-level entities is complicated because of intraspecific variation. Analytical procedures developed to date have focussed on capturing this divergence using statistical methods including multidimensional scaling (Hebert et al. 2003a), pairwise similarity (Steinke et al. 2005), likelihood (Matz and Nielsen 2005) and Bayesian (Nielsen and Matz 2006) methods, the latter two rooted in coalescent theory.

These methods have met with varying levels of success in assigning a query sequence to a particular group of database sequences (Hebert et al. 2003b; Paquin and Hedin 2004; Barber and Boyce 2006; Scheffer et al. 2006), and while simulations suggest that assignments are fairly robust to variation of population sizes and mutation rates (Nielsen and Matz 2006), any similarity-based approaches in the practice of DNA taxonomy are inherently limited. Because the methods are general tests of group membership based on sequence similarity, the resulting group assignments can be at any hierarchical level and may have little relevance with regard to matching a query at the species level (Nielsen and Matz 2006). This has not always been appreciated in the recent literature (Janzen et al. 2005; Meyer and Paulay 2005) where the position in the tree is used to decide species limits without a clear criterion (Brower 2006). Instead, well-established quantitative methods for species delineation (Cracraft 1983; Sites and Marshall 2003) draw inferences based on *a priori*-hypothesized putative groups whose existence is subsequently tested by analysis of the new DNA or morphological evidence (Davis and Nixon 1992; Brower 2006). Species delimitation may be 'character based' or 'tree based'. Under the former, species limits are defined by the presence of 'diagnostic' traits in a set of populations, i.e. character states uniformly present in all members of the set, but nowhere else. This has been implemented in population aggregation analysis (Davis and Nixon 1992), which can be modified to allow for homoplasy of

diagnostic characters (Brower 1999). Tree-based methods define these sets on the basis of the tree topology, testing for 'exclusivity', i.e. clades that are exclusive to a set of populations but absent from other populations (Wiens and Penkrot 2002).

While readily applicable in taxonomic studies based on DNA sequences (Sites and Marshall 2003), the use of these procedures in species delimitation requires evidence of populational coherence prior to the analysis. This knowledge, however, may be difficult to obtain as populations frequently lack easily discernible spatial or genetic boundaries (Schaefer 2006), in particular where closely related species occur in sympatry. This constitutes a great difficulty with the application of the underlying phylogenetic species concept. Yet, this problem is by no means specific to DNA-based data, and the fact that quantitative methods have rarely been applied in morphology-based species delineation (e.g. Cracraft 1992) further suggests the difficulty of obtaining appropriate characters in traditional taxonomy and the subjectivity of many current species designations.

The problems resulting from the populational approach to species delimitation could be overcome if it was possible to establish species boundaries from the sequences themselves, without the requirement of prior population definitions before they are subjected to tests of aggregation. In a recent study, Pons et al. (2006) (and also D. Fontaneto, E Herniou, C Boschetti, M Caprioli, G Melone, C Ricci and TG Barraclough (in prep.) developed a method whereby the clustering observed in DNA taxonomy data is analysed with respect to the branching rate in a clock-constrained phylogram. These trees generally show a very steep transition from low branching rates throughout most of the tree, to very fast branching rates near the tips (Fig. 2). This shift can be interpreted as a change from phylogenetic branching (macroevolution) to population-level processes (microevolution), and can be described, respectively, with simple lineage birth models used in studies of clade evolution (Yule 1924; Nee 2001) and equations from coalescence theory (Hudson 1991; Wakeley 2006). These new quantitative procedures, therefore, can infer the elusive species boundary directly from the transition in branching rate and constitute an exciting possibility to define species from sequence variation (rather than from predefined populations), solving problems of recognizing separated populations in sympatry or across different developmental stages. Pons et al.'s (2006) analysis of branch length also takes into account uncertainty of species limits by permitting confidence intervals when allocating species defining nodes, a desirable property where species limits are weakly developed (Hey et al. 2003).

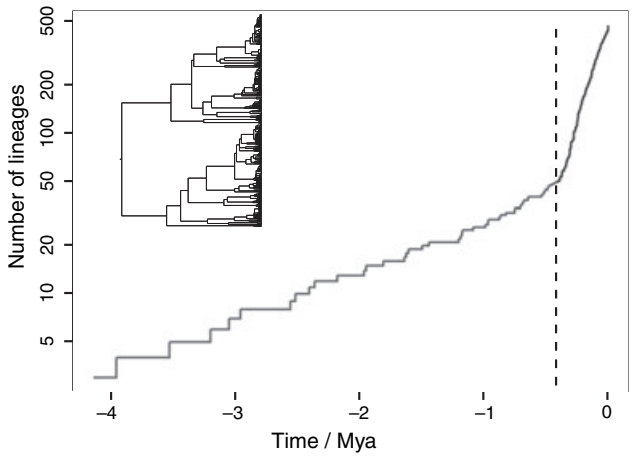

Fig. 2. A semi-logarithmic lineage-through-time plot derived from a phylogram for a sample of sequences from a 'DNA profiling' study (Pons et al. 2006). The rapid increase in diversification rates that presumably demarcates the species boundary (vertical line) is clearly evident

Implementing a DNA taxonomic system: a growing database

With this kind of analytical tool in hand, and ever faster protocols for sequencing, it is predicted that taxonomy will increasingly be based on DNA information. A standard DNA taxonomic analysis (see Fig. 3) could include broad sampling of the target taxa across their geographic range, followed by large-scale sequencing of representative samples for a 'DNA profile' of the group, and a subsequent quantitative analysis to test species limits (Hebert et al. 2003a; Smith et al. 2005; Monaghan et al. 2006; Pons et al. 2006). Sampling across a geographic region would yield putative species by clustering local samples, providing a system that is easily linked to other studies using similar protocols. The DNA taxonomic system inevitably emerges from these individual studies. Given the algorithmic procedures for analysis, the system will be derived from the data, rather than an author's decision. Hypothesized species entities can be further scrutinized against other data, including morphology, biogeography and others (DeSalle et al. 2005), and if corroborated, these tests also provide an evolutionary justification of the procedures used for species delimitation. Where discrepancies occur, further data may be generated and integrated into the growing database, for a new iteration of the process (Fig. 3).

It is predicted that DNA databases will increasingly be the point of reference for taxonomic information in the future, taking on the role of physical specimen collections and taxonomic literature of today. Once a sequence database entry is available for a good proportion of the existing species, each of them recognizable as a distinct cluster of sequences, a variety of existing fast search algorithms can give the correct identification from among millions of species in seconds. These DNA data are accessible to everyone in the research community, applicable to all developmental stages, and not affected by researcher bias (Tautz et al. 2003). There is little doubt that, once these databases attain good coverage across the spectrum of a group, they will be consulted increasingly by the 'end-user', while the physical collections will decrease in importance. Increasingly, DNA profiles of entire communities or ecosystems will be obtained, from which conclusions about macroecological and macroevolutionary patterns can be derived (Pons et al. 2006). It is conceivable that much of the poorly known biodiversity of the planet will be investigated largely on the DNA level to the study of species diversity, spatial turnover, range sizes and lineage history, diversification rates, and to link these factors to questions about paleoclimatic and geological history, environmental factors and disturbance and others. This may ultimately overcome the 'taxonomic impediment' that prevents such studies today for most of the non-vertebrate biodiversity.

This leaves the question of how well DNA-based taxonomies reflect the distribution and extent of species diversity on Earth, and to what degree DNA taxonomies match the

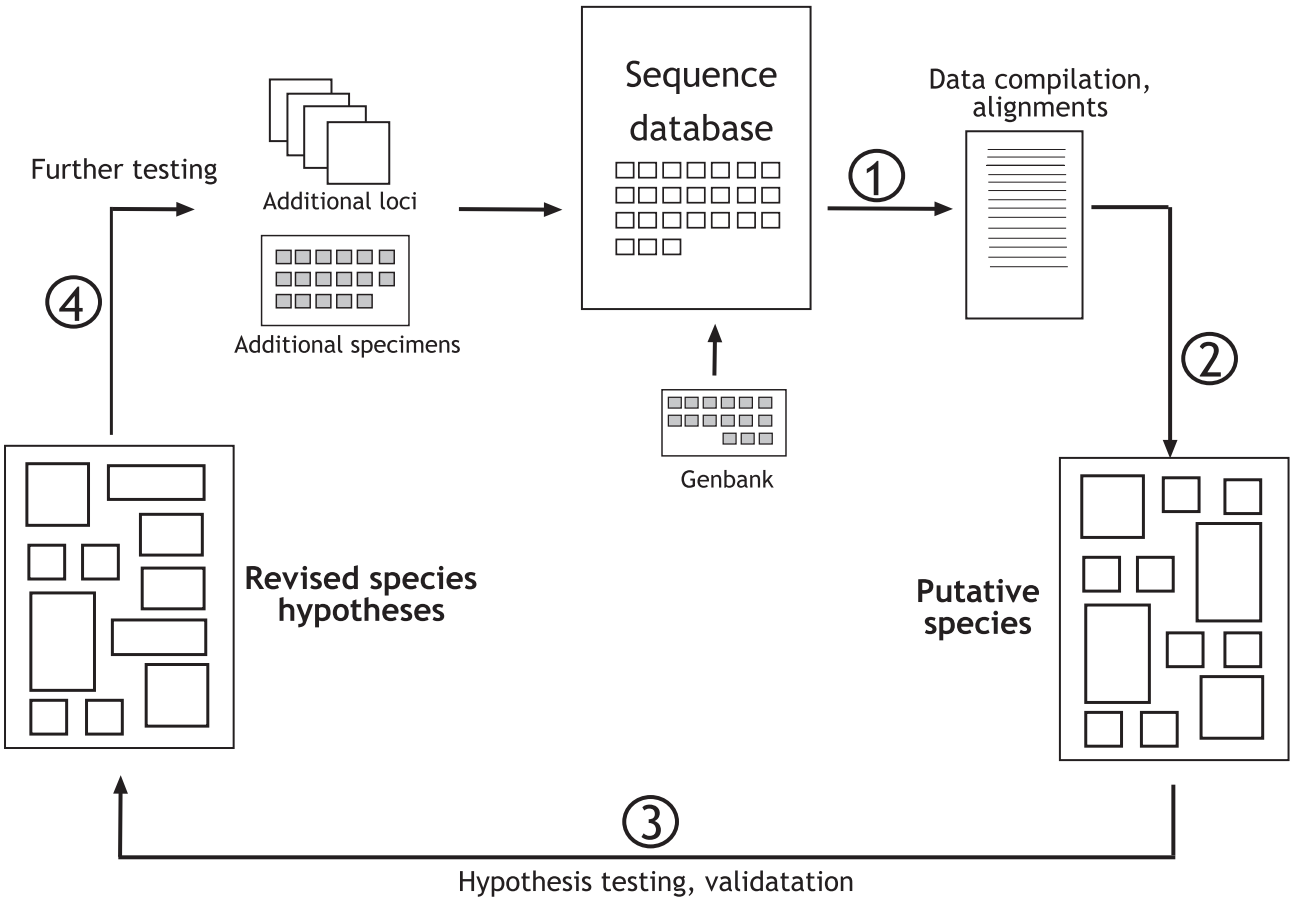

Fig. 3. A conceptual flow chart of procedures in DNA taxonomy. Sequences are compiled from a primary database prior to quantitative analysis of species delimitation (step 1). Grouping and species delimitation are achieved through various procedures (main text) to provide primary hypotheses of species limits and phylogenetic relationships (step 2). Hypothesized species entities are further scrutinized against other data (morphology, biogeography, distributional ranges, etc.) for corroboration of primary hypotheses (see DeSalle et al. 2005), and hypotheses are revised if necessary (Step 3). Additional populations or additional genetic loci may be selected to test the original hypotheses or resolving discrepancies (Step 4). These new data are included in the growing database, which is also fuelled from other sources such as GenBank, for a new iteration of the process

existing system. Several issues about the utility of DNA-based approaches remain largely unanswered. For example, how complete sampling of a clade of species is needed and at what geographic scale before the conclusions of the iterative analyses (Fig. 3) stabilize? Similarly, despite claims by some that the *cox1* gene is sufficient as a universal barcoding marker (Marshall 2005), it remains unclear what are the most appropriate markers to reflect species boundaries, and what is the minimum number of loci (mtDNA and nuclear) that would provide a stable system? To what degree would species limits be captured by different markers, and what is the level of incongruence when using morphological and molecular data? Finally, what is the phylogenetic information content of these sequences, and will these data also produce a solid image of the Tree of Life? It is apparent that these conclusions can only be drawn against a background of established knowledge about species limits and phylogeny that has been accumulated by diverse disciplines including classical taxonomy, ecology and behavioural biology, biogeography, palaeontology and others, which collectively have produced our evolutionary understanding of the living world. Clearly, this information is the necessary background for justifying the utility of DNA-based approaches (step 4 in Fig. 3), and the success of this in the few cases where it has been properly assessed, testifies to the great strength of the approach. Detailed questions about the specifics of sampling strategy, utility of gene markers, length of gene fragments needed and others, will also have to be assessed in this framework.

Conclusions

While DNA barcoding has focussed mainly on practical conveniences, these approaches have frequently ignored the existing, well-developed evolutionary theory of species delimitation. As such, the generation of 'barcode' data will be useful for the identification of the 10–20% of Earth's biodiversity that has been formally described, but will fail to advance the study of all other taxa and may not truly improve the ability to catalogue and understand the evolution of biodiversity. Measures of sequence similarity have been widely used to examine recent large-scale sequence data, and sophisticated statistical methods of grouping based on sequence similarity have been developed. Yet, the evolutionary relevance of units defined in this way remains unclear. Instead, future efforts should be directed towards improving the algorithms for the test of theoretically sound species concepts that consider the biological nature of species as isolated sets of populations lacking recombination. Unfortunately, the evolutionary-hypo-

thesis based nature of taxonomy (including systematics) continues to be ignored in the DNA-barcoding literature, as has been pointed out repeatedly (e.g. Wheeler 2004; Will and Rubinoff 2004). This has resulted in the outright dismissal of DNA-based methods by many classical taxonomists, and led to an unnecessary alienation of researchers along methodological lines. The DNA barcoding community has responded by proposing a conciliatory approach that uses the sequence information only to aid the study of curated voucher specimens and the continued application of the Linnean names (Gregory 2005). Although this may placate taxonomists concerned about funding streams, it does not solve the fundamental problems of determining exactly how DNA will be integrated into the taxonomic system and establishing the theoretical relevance of these data.

In fact, it would be insufficient if DNA taxonomy had no more than an auxiliary role while the use of traditional methods as the basis for species delimitation and classification is continued. This simply does not take full advantage of the DNA information, and will perpetuate the dependency of taxonomy on the slow search for diagnostic morphological characters and the need for expert training to recognize them. Since DNA taxonomy was first suggested just a few years ago, its power has exceeded all expectations. Yet, as critics rightly pointed out, sequence data in themselves are of little value, unless independent evidence can provide the evolutionary justification for the DNA approach. The major role of non-DNA data, therefore, is to provide the background knowledge corroborating the evolutionary interpretation of the DNA data (DeSalle et al. 2005; Pons et al. 2006). This also requires to link the existing taxonomic system to these DNA-based entities, so that the current biological information associated to the existing names will be not lost. Hence, there will be a continued need for carefully curated DNA databases from specimens correctly identified by specialist taxonomists, and a unified effort of taxonomists, bioinformaticians and molecular systematists.

Acknowledgements

We are grateful to Tim Barraclough, Anabela Cardoso, Ryan Gregory, David Lees and Joan Pons for discussions about DNA taxonomy and related topics. We thank Darrel Siebert for his highly relevant contribution at the DNA and Museum Collections workshop held at the Natural History Museum in February 2006. We also thank Dr W. Westheide for suggesting the topic and providing editorial help, and thank the referees for useful comments. Our research on DNA taxonomy is funded by the BBSRC (grant BBS/B/04358) and the Leverhulme Trust (F/00696/I).

Zusammenfassung

Fortschritte in der DNA-Taxonomie

Die großangelegte DNA-Sequenzierung von lebendenden Organismen enthält große Versprechungen für die Taxonomie, ist aber umstritten. Hier beschreiben wir Fortschritte auf diesem Gebiet, die aus der drastischen Zunahme der Datenerzeugung resultieren. Wir unterscheiden DNA-Taxonomie von 'DNA barcoding', wobei der erste Begriff direkt die Umgrenzung der Arten auf der Basis von evolutionsbiologischen Artenkonzepten betrifft, während DNA barcoding nur die Identifizierung von *a priori* definierten Arten auf Grund von Ähnlichkeitskriterien beschreibt. Ein wichtiger Befund bisheriger Studien an Tieren ist, dass sich die Variation der mitochondrialen DNA in Gruppen von nah verwandten Genotypen gruppiert, deren Ausmaß sowohl den traditionsgemäß anerkannten Artgrenzen entspricht als auch mit der Variation in Kerngenen übereinstimmt. Diese Befunde zeigen, dass der Gebrauch von DNA-Sequenzen auch als Primärinformation für Artenbegrenzung in taxonomisch weniger bekannten Gruppen sinnvoll ist. Molekulare Techniken sind in der Vergangenheit meist an taxonomisch komplizierten Fällen angewandt worden, was wahrscheinlich zu einer Überschätzung der Arten mit polyphyletischen mtDNA-Haplotypen geführt hat. Die anhaltende Weiterentwicklung der DNA-Taxonomie führt jetzt zu verbesserten Strategien für die Auswahl von Individuen und zu verbesserten Datenanalysen. Statistische Methoden der Gruppenbilding auf Grund des Kriteriums der Ähnlichkeit sind schon vorhanden; jedoch haben die Einheiten (Arten), die in dieser Weise definiert werden, keinen klaren Bezug zur Evolution der Gruppe. In Zukunft wird eine DNA-taxonomische Standardanalyse die weiträumige Aufsammlung der untersuchten Taxa über ihr geographisches Areal einschliessen, gefolgt von der weitreichenden Sequenzierung von representativen Arten für ein 'DNA-Profil' der Gruppe sowie algorithmische Verfahren für die Artbegrenzung. Das taxonomische System wird von den Daten und nicht von der Expertenmeinung abgeleitet. Art-Hypothesen können gegen Morphologie, Biogeographie und andere Daten geprüft werden und bringen eine evolutionsbiologische Rechtfertigung der Methoden, die für Artbegrenzung verwendet wurden. Diskrepanzen zwischen DNA und anderen Daten erlauben eine taxonomische Rückkopplungsschleife, die nach Einschluß von neuen Daten zu verfeinerten Artkonzepten führt. Wir argumentieren jedoch, dass der Gebrauch von DNA-Methoden in der Taxonomie (einschließlich DNA barcoding) umstritten bleibt, bis diese besser mit vorhandenen Theorien der Evolutionsbiologie und der Phylogenetik zu begründen sind.

References

Agapow PM, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, Purvis A (2004) The impact of species concept on biodiversity studies. Quart Rev Biol **79:**161–179.

Avise JC (2000) Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA.

Avise JC, Walker DE (1999) Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. Proc Natl Acad Sci **96:**992–995.

Baker CS, Palumbi SR (1994) Which whales are hunted – a molecular genetic approach to monitoring whaling. Science **265:**1538–1539.

Ball SL, Armstrong KF (2006) DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can J For Res **36:**337–350.

Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proc R Soc B **273:**2053–2061.

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. Can J Zool **83:**481–491.

Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. Science **312:**570–572.

Bensch S, Perez-Tris J, Waldenstrom J, Hellgren O (2004) Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? Evolution **58:**1617–1621.

Besansky NJ, Severson DW, Ferdig MT (2003) DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol **19:**545–546.

Blaxter ML (2004) The promise of a DNA taxonomy. Phil Trans R Soc B **359:**669–679.

Blaxter M, Elsworth B, Daub J (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. Proc R Soc B **271:**S189–S192.

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.

Brower AVZ (2006) Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst Biodivers **4:**127–132.

Brower AVZ, DeSalle R, Vogler A (1996) Gene trees, species trees, and systematics. Annu Rev Ecol Syst **27:**423–450.

Cardoso A, Vogler AP (2005) DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida* species group (Coleoptera: Cicindelidae). Mol Ecol **14:**3531–3546.
Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. Phil Trans R Soc B **353:**287–305.
Cracraft J (1983) Species concept and speciation analysis. Curr Ornithol **1:**159–187.
Cracraft J (1992) The species of the birds-of-paradise (Paradisaeidae): applying the phylogenetic species concept to complex pattern of diversification. Cladistics **8:**1–43.
Crisci JV (2006) One-dimensional systematist: perils in a time of steady progress. Syst Bot **31:**217–221.
Dalebout ML, Baker CS, Mead JG, Cockcroft VG, Yamada TK (2004) A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. J Hered **95:**459–473.
Davis JI, Nixon KC (1992) Populations, genetic variation, and the delimitation of phylogenetic species. Syst Biol **41:**421–435.
DeSalle R, Birstein VJ (1996) PCR identification of black caviar. Nature **381:**197–198.
DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Phil Trans R Soc B **360:**1905–1916.
Ennos AE, French GC, Hollingsworth PM (2005) Conserving taxonomic complexity. Trends Ecol Evol **20:**164–168.
Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. Mol Ecol **11:**839–850.
Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA, Wolfe RS, Balch WE, Tanner RS, Magrum LJ, Zablen LB, Blakemore R, Gupta R, Bonen L, Lewis BJ, Stahl DA, Luehrsen KR, Chen KN, Woese CR (1980) The phylogeny of prokaryotes. Science **209:**457–463.
Franz NM (2005) On the lack of good scientific reasons for the growing phylogeny/classification gap. Cladistics **21:**495–500.
Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Syst **34:**397–423.
Gaines CA, Hare MP, Beck SE, Rosenbaum HC (2005) Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. Proc R Soc B **272:**533–542.
Gillespie RG, Roderick GK (2002) Arthopods on islands: colonisation, speciation, and conservation. Annu Rev Entomol **47:**595–632.
Godfray HCJ (2002) Challenges for taxonomy – the discipline will have to reinvent itself if it is to survive and flourish. Nature **417:**17–19.
Goldstein PZ, Desalle R (2003) Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. Mol Ecol **12:**1993–1998.
Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM (2006) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. Mol Ecol **15:**1759–1768.
Gregory TR (2005) DNA barcoding does not compete with taxonomy. Nature **434:**1067–1067.
Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci **103:**968–971.
Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol **54:**852–859.
Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. Proc R Soc B **270:**313–321.
Hebert PDN, Ratnasingham S, DeWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc B **270** (Suppl.)**:** S96–99.
Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci **101:**14812–14817.
Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. PLoS Biol **2:**1657–1663.
Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. Trends Ecol Evol **18:**597–603.
Hogg ID, Hebert PDN (2004) Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Can J Zool **82:**749–754.
Hudson RR (1991) Gene genealogies and the coalescent process. Oxf Surv Evol Biol **7:**1–44.
Hugall A, Moritz C, Moussalli A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail Gnarosophia bellendenkerensis (Brazier 1875). Proc Natl Acad Sci **99:**6112–6117.
Janzen DH (2004) Now is the time. Phil Trans R Soc B **359:**731–732.
Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Phil Trans R Soc B **360:**1835–1845.
Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci **102:**8369–8374.
Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. Trends Ecol Evol **18:**65–68.
Mallet J, Willmott K (2003) Taxonomy: renaissance or Tower of Babel? Trends Ecol Evol **18:**57–59.
Markmann M, Tautz D (2005) Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. Phil Trans R Soc B **360:**1917–1924.
Marshall E (2005) Taxonomy – will DNA bar codes breathe life into classification? Science **307:**1037–1037.
Matz MV, Nielsen R (2005) A likelihood ratio test for species membership based on DNA sequence data. Phil Trans R Soc B **360:**1969–1974.
Mayr E (1942) Systematics and the Origin of Species. Columbia University Press, New York.
Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol **3:**2229–2238.
Monaghan MT, Balke M, Gregory TR, Vogler AP (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Phil Trans R Soc B **360:**1925–1933.
Monaghan MT, Balke M, Pons J, Vogler AP (2006) Beyond barcodes: complex DNA taxonomy of a south pacific island radiation. Proc R Soc B **273:**887–893.
Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. PLoS Biol **2:**1529–1531.
Nee S (2001) Inferring speciation rates from phylogenies. Evolution **55:**661–668.
Nielsen R, Matz M (2006) Statistical approaches for DNA barcoding. Syst Biol **55:**162–169.
Page TJ, Choy SC, Hughes JM (2005) The taxonomic feedback loop: symbiosis of morphology and molecules. Biol Lett **1:**139–142.
Paquin P, Hedin M (2004) The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Cicurina* (Araneae: Dictynidae) from Texas caves. Mol Ecol **13:**3239–3255.
Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol **55:**595–609.
Powers T (2004) Nematode molecular diagnostics: from bands to barcodes. Ann Rev Phytopathol **42:**367–383.
Prendini L (2005) Comment on "identifying spiders through DNA barcodes". Can J Zool **83:**498–504.
Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Phil Trans R Soc B **360:**1879–1888.
Savolainen V, Cowan RS, Vogler AP, Roderick G, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Phil Trans R Soc B **360:**1805–1811.
Schaefer JA (2006) Towards maturation of the population concept. Oikos **112:**236–240.
Scheffer SJ, Lewis ML, Joshi RC (2006) DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines. Ann Entomol Soc Am **99:**204–210.

Scherer S, Sontag C (1986) Molecular taxonomy and evolution of the Anatidae. Z Zool Syst Evolutionsforsch **24:**1–19.
Sereno PC (2005) The logical basis of phylogenetic taxonomy. Syst Biol **54:**595–619.
Sharley DJ, Pettigrove V, Parsons YM (2004) Molecular identification of *Chironomus* spp. (Diptera) for biomonitoring of aquatic ecosystems. Aust J Entomol **43:**359–365.
Sites JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. Trends Ecol Evol **18:**462–470.
Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Phil Trans R Soc B **360:**1825–1834.
Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). Proc Natl Acad Sci **103:**3657–3662.
Sperling FAH (2004) DNA barcoding: deus ex machina. Newsl Biol Surv Can **22:**50–53.
Sperling FAH, Anderson GS, Hickey DA (1994) A DNA-based approach to the identification of insect species used for postmortem interval estimation. J Forensic Sci **39:**418–427.
Steinke D, Vences M, Salzburger W, Meyer A (2005) TaxI: a software tool for DNA barcoding using distance methods. Phil Trans R Soc B **360:**1975–1980.
Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends Ecol Evol **18:**70–74.
Thiele K, Yeates D (2002) Tension arises from duality at the heart of taxonomy – names must both represent a volatile hypothesis and provide a key to lasting information. Nature **419:**337–337.
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:**606–621.
Wakeley J (2006) Coalescent Theory: An Introduction. Roberts & Co., Greenwood Village, CO.
Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Phil Trans R Soc B **360:**1847–1857.
Wheeler QD (2004) Taxonomic triage and the poverty of phylogeny. Phil Trans R Soc B **359:**571–583.
Wheeler QD (2005) Losing the plot: DNA "barcodes" and taxonomy. Cladistics **21:**405–407.
Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst Biol **51:**69–91.
Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics **20:**47–55.
Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Syst Biol **54:**844–851.
Wilson EO (2003) The encyclopedia of life. Trends Ecol Evol **18:**77–80.
Yule GU (1924) A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, FRS. Phil Trans R Soc B **213:**21–87.

Authors' addresses: Alfried P. Vogler (for correspondence) and Michael T. Monaghan, Department of Entomology, Natural History Museum, Cromwell Road, London SW7 5BD, UK. E-mail: apv@nhm.ac.uk, m.monaghan@uhm.ac.uk