CSIRO PUBLISHING

www.publish.csiro.au/journals/is *Invertebrate Systematics*, 2004, **18**, 1–6

Review

The molecularisation of taxonomy

Michael S. Y. Lee

Department of Environmental Biology, University of Adelaide, SA 5005, Australia & Earth Sciences Section, South Australian Museum, North Terrace, Adelaide SA 5000, Australia. Email: michael.s.lee@adelaide.edu.au

Abstract. The recent proposal for a new system of biological taxonomy based primarily on DNA sequences from one or a few chosen ('standard') genes sequenced across all taxa appears inadvisable for both practical and theoretical reasons. While nucleotide sequences are more objective than traditional (e.g. morphological) data in some respects (character choice, character delineation, character state identity), in other respects both are inherently subjective (homology/alignment, divergence metrics). Sequence divergence in standard gene(s) is an extremely crude method for determining species limits; more appropriate markers (potentially directly linked to species criteria such as reproductive isolation) should be and often are used. It is thus worth persisting with the plurality of genetic, anatomical and ethological criteria currently used to hypothesise ('identify') and test species boundaries. However, once species boundaries have been thus discerned, use of sequences from standard genes to diagnose those boundaries (and place individuals with respect to those boundaries) is highly feasible, though subject to error like any single type of marker. In many cases this approach might have advantages over morphological diagnoses. However, unless an appropriate taxonomic framework constructed using all appropriate biological information is already in place, such molecular diagnoses will be premature.

Additional keywords: biodiversity, molecular systematics, species.

Introduction

In recent years, biological systematics has seen the rapid growth of molecular phylogenetics and the increasing ease of sequencing DNA, along with the attrition of morphological approaches as experts with irreplaceable organismal expertise retire without replacement. These trends have led to recent proposals to place taxonomy firmly within the molecular age, by using sequences in a chosen ('standard') gene as the universal reference for naming species (Tautz *et al*. 2003) and for identifying biological specimens (Hebert *et al*. 2003). Tautz *et al*. (2003) have suggested that existing rules of nomenclature, where taxonomic debates are resolved with reference to type specimens, be replaced with a new system where such debates are resolved with reference to 'type' gene sequences from voucher specimens. Molecular data, of course, is regularly used in exactly this fashion to resolve the validity and boundaries of species (especially in morphologically problematic groups), but Tautz *et al*.'s proposal is radical in making molecular approaches mandatory, and furthermore in specifying the use of standard gene sequences (such as nuclear rRNA and/or mitochondrial Cytochrome *b*). Along similar lines, Hebert *et al*. (2003) demonstrated the feasibility of broad-scale use of sequences of a standard gene for specimen identification, through establishment of a database containing mitochondrial *CO1* sequences across a broad sample of metazoans. An unknown taxon is identified at coarse ('phylum' or 'order') levels by comparison of its amino acid sequence with existing organisms in the database, and assigned to finer levels (e.g. 'genus' and 'species') by analysis of its nucleotide sequence.

These proposals have elicited a flurry of responses, some broadly in support (Blaxter and Floyd 2003; Baker *et al*. 2003) but many against (Seberg *et al*. 2003; Lipscomb *et al*. 2003; Dunn 2003). These exchanges, however, have been very brief due to space constraints, and sometimes (Proudlove and Wood 2003) conflated the issue of molecular approaches for diagnosing species with the issue of a formalised molecular taxonomy using standard genes. A more detailed evaluation of the situation is thus attempted here. Because Linnaean higher taxa are arbitrary constructs (e.g. de Queiroz and Gauthier 1992; Ereshefsky 2001), this discussion is restricted to α or species-level taxonomy. The arguments for and against a molecular taxonomic framework are evaluated, and it is concluded that a formal taxonomic framework for naming and delimiting species based on one or even several 'standard' gene sequences is inadvisable on both practical and theoretical grounds. Most

 10.1071/IS03021 1445-5226/04/010001

importantly, genetic divergence is too crude a proxy for reproductive isolation and thus species boundaries (Ballard *et al*. 2002; Ferguson 2002). It is thus worth persisting with the plurality of genetic, anatomical and ethological criteria currently used to identify and test species boundaries. However, once species boundaries have been thus identified, using a standard gene sequence to place individuals with respect to those boundaries can be highly feasible (though not infallible), and in many cases might have advantages over traditional morphological approaches.

Improved objectivity

The digital nature of DNA sequences, and thus the presumed objectivity of sequence divergence, has been argued to improve nomenclatural stability (Tautz *et al*. 2003). However, others have pointed out that this reasoning would be valid only if DNA sequences were identical within species and differed between species; like any other organismal trait, DNA sequences vary within and between species, and there would continue to be arguments over species limits (splitting and lumping), and thus the identity of particular individuals (Lipscomb *et al*. 2003). In this section, only the problem of objectively inferring degree of divergence is addressed; the issue of using such divergence, whether molecular or morphological, to determine species limits and affinity, is addressed in a following section.

Deoxyribonucleic acid sequences are clearly more objective than phenotypic data in several aspects. Choosing and evaluating the same set of characters across a diverse range of organisms ('character standardisation') will be more straightforward for molecular data. Although the choice of gene region(s) to use as the standard reference will be unavoidably subjective, once that choice is made, biologists can routinely sequence every single nucleotide for that region for each organism. In contrast, it will be much more difficult to devise an analogous 'standard' list of taxonomically-useful phenotypic characters that are applicable across any large group, and biologists will always need to select particular phenotypic characters to include in their analyses (Hillis and Wiens 2000). Also, once the problematic issue of alignment (see below) has been dealt with, character delineation is much less of an issue: each aligned nucleotide position is a clearly definable unit character. In contrast, the number of ways of atomising an organism into unit phenotypic characters is much more problematic (Wagner 2001). Finally, molecular characters are more similar to each other than are morphological characters. Each character (aligned position) has the same five possible states (four bases and gap): each state is discrete and easily identifiable (sequencing error notwithstanding). In contrast, most morphological traits vary continuously, and identification of the number and limits of states within each character is often largely arbitrary (e.g. Goldman 1988; Thiele 1993). Thus, it might be argued that in evaluating divergence, each (observed or inferred) molecular change can be treated as a standard unit and assigned equal weight, whereas this assumption is more difficult to justify with morphological data. However, this view might be countered by emphasising that different substitutions can have different effects (e.g. transitions *v.* transversions, silent *v.* replacement substitutions, changes in active *v.* non-active sites). If so, simply counting all observed or inferred molecular substitutions for the purposes of calculating divergence might be as coarse an approximation as counting the number of changes in heterogeneous morphological characters.

In many other aspects, nucleotide data is just as inherently subjective as morphological data. Even if the task is merely quantification of similarity, there are major subjectivities involved. For instance, even calculating the simple percentage divergence entails a choice of alignment methods and gap costs. Alignment (identifying positional homologies in sequences of differing lengths from two organisms) entails problems similar to those involved when determining the homology of phenotypic traits in different organisms (e.g. Wheeler 2001). As in morphology, great errors can occur where fundamentally different structures are assumed equivalent: for instance, if a mitochondrial sequence from one organism is aligned (homologised) with a nuclear paralogue from another. Detection of such errors is not necessarily more straightforward for molecular data (Olson and Yoder 2002). Even if the homology of regions to be aligned is beyond doubt, the precise positional homologies for each site will vary according to what alignment algorithm is used, and what costs are assigned to gaps, stems/loops, and transitions/transversions (e.g. Gatesy *et al*. 1993). There is at present no widely-accepted, objective method for choosing a single algorithm and set of transformation costs for aligning sequences. The different alignments generated by alternative procedures will, of course, lead to different inferred levels of divergence. Even for a single alignment (however obtained), there is no single accepted way to calculate *the* degree of molecular similarity or divergence. Rather, there is again a plethora of methods, from simple raw distances which tally the number of differing sites, to complex models for correcting for multiple substitutions (e.g. Page and Holmes 1998, Nei and Kumar 2000). For many of these methods, there are a range of possible parameter values which will also affect calculated divergences / branch lengths. This is no different to the problem of picking one of the numerous possible metrics for calculating phenetic distances based on morphological characters, a problem which undermined the purported objectiveness of numerical taxonomy (Hull 1988). In molecular and morphological cases, both absolute and relative (rank order) distances between organisms is highly dependent on the distance metric employed. Finally, even if a single objective method to calculate sequence divergence was somehow devised and universally adopted, lineage-specific

rates of molecular evolution would still make it difficult to equate a certain percentage divergence with 'species-ness' (see below).

For the purposes of calculating similarity, molecular data are more objective than morphological traits in some respects (*fide* Tautz *et al*. 2003): character standardisation, character delineation, character-state identity, and (arguably) character equivalency. However, in other respects—homology/alignment, similarity metrics, models for inferring branch lengths on phenograms or phylogenies—determination of similarity based on DNA characters is just as subjective as for morphological traits (*fide* Lipscomb *et al*. 2003). However, the question of whether molecular data are easier and more cost-effective to gather (Tautz *et al*. 2003) is a separate issue and will be discussed later.

Species boundaries

The current codes of nomenclature provide rules for naming species using a type organism and a Linnaean rank (e.g. Ereshefsky 2001). *Homo sapiens,* for instance, consists of the type organism (Linnaeus: Spamer 1999) and all organisms belonging to the same species as the type. The focal point of a species is marked by the type specimen, and boundaries (and thus, contents) of that species determined by employing some species concept. However, the codes are totally silent on the issue of species concepts, i.e. they do not 'define' what sort of biological entity constitutes a species. The boundaries of *Homo sapiens* could be very broad or very narrow depending on the species concept adopted, with consequent implications for assigning (especially fossil) organisms to this species or to another. Until there is universal agreement on a particular species concept, there is no objective way to draw species boundaries and resolve such disputes. The statement that 'organism X belongs to species Y' is not testable, since the boundaries of species Y can be broadened or reduced at will by invoking alternative species concepts, even biologically unreasonable ones (Lee 2003). No amount of character information—morphological or molecular—can create taxonomic stability if the boundaries of species are highly labile and subjective.

The arguments below adopt the widespread view that considers species as reproductively isolated lineages of organisms (e.g. Mayr 1963; de Queiroz 1998); however, these arguments will likely apply if other species concepts are adopted, provided these concepts provide rigorous criteria for defining species limits. If species are reproductively-isolated lineages, the boundaries of species can be drawn with at least some precision (partial interbreeding notwithstanding: see Lee 2003: fig. 1), turning the statement 'organism X belongs to species Y' into an hypothesis (organism X and the type of species Y belong to the same interbreeding lineage) that can be corroborated or refuted by observations. However, using genetic divergences in a standard gene to evaluate this hypothesis and thus determine species identity (the consequence of the system proposed by Tautz *et al*. 2003) is an overly crude approach (Ballard *et al*. 2002; Ferguson 2002). Divergence in any widely sequenced, candidate 'standard' gene is not causally related to reproductive isolation. Rather, in most speciation events, genetic divergence and reproductive isolation are only very loosely associated, both increasing in an irregular fashion with time in allopatry. Numerous examples exist documenting the loose correlation between genetic divergence and reproductive isolation (see overviews in Johns and Avise 1998; Ferguson 2002). This means that it will never be possible to equate a particular level of genetic divergence with reproductive isolation (species boundaries). For groups where reproductive isolation evolves very slowly, even highly genetically divergent organisms would be capable of interbreeding (and thus conspecific under reproductive species concepts), while in groups where reproductive isolation evolves very rapidly, the converse would hold. A taxonomic system where 'speciesness' is determined principally by molecular divergence in a standard gene (or any other trait, for that matter) would incorrectly split species in the former case, and lump species in the latter. This issue was briefly acknowledged by Tautz *et al*. (2003: box 1), who, as experienced systematists, were undoubtedly aware of the problem. However, no straightforward solution was presented.

If one accepts that species boundaries should be equated with the limits to interbreeding, then the most appropriate markers for diagnosing species boundaries will be those causally related to reproductive compatibility. No widely-sequenced genes fit this criterion. However, some molecular markers might be directly implicated in the origin of reproductive isolation (e.g. Wu 2001), and other commonly used traits such as allozymes (e.g. Fitzpatrick 2002), genital morphology (e.g. Eberhard 1985), mating vocalisations and courtship behaviour (e.g. Ryan 2001) are potentially causally associated with cladogenesis. Divergence in any these markers would be more accurate proxies of species boundaries than sequences in 'housekeeping' genes (and, it should be acknowledged, gross morphology). The boundaries of reproductive isolation can of course be rigorously inferred using genetic (including molecular) methods, but these require more sophisticated techniques than divergence comparisons with single 'type' sequence. Phylogenetic analysis of multiple individuals and often multiple loci is required. The identification of reciprocally monophyletic subgroups in a putative species, each diagnosed by alleles at multiple unlinked loci would suggest reproductive isolation between these subgroups (e.g. Avise 1999). However, such inferences need not be made according to molecular sequences—similar patterns in allozymes or morphological traits would be sufficient. As acknowledged by Tautz *et al*. (2003), the problems of lateral gene transfer, lineage sorting and homoplasy mean that the converse does not hold: if there *is* incongruence between gene trees, it cannot be concluded

that all organisms form a single lineage. Also, a single fixed difference between sympatric groups of a diploid sexually reproducing species is indicative of absence of interbreeding and thus sufficient to both recognise and characterise two species (Richardson *et al*. 1986). Although such differences may be apparent in patterns of morphological, physiological, or ecological variation, they have typically been revealed in allozyme electrophoretic studies, the results of which can usually be interpreted directly in terms of Mendelian genotypes at particular loci (Richardson *et al*. 1986; Avise 1994).

Species boundaries thus cannot usually be determined by evaluation of sequence divergence in a standard gene or genes; rather, multiple lines of evidence need to be considered. Genetic divergence is, like morphological similarity, at best a crude approximation for reproductive isolation, subject to refinement based on other sources (Ferguson 2002). If so, there is no reason to grant either measure central importance in species definitions. Once the boundaries of species have been determined using all appropriate biological information, however, a profile for some standard gene for each species can be constructed, allowing identification of organisms using sequences from this gene (Hebert *et al*. 2003). This would be analogous to identifying *post hoc* the best morphological characters separating identified species (e.g. through discriminant function analysis), and measuring these variables to provisionally identify new specimens. Whether molecular or morphological approaches to identification are more economical might of course vary according to group (see later).

Fossils

None of the commentaries on molecular taxonomy have discussed the issue of fossil organisms, for which DNA sequences will usually not be available (V. Weisbecker, personal communication). A molecular taxonomy would result in two different nomenclatural systems, a traditional one for fossil organisms, emphasising type specimens and hard anatomy, and a molecular one for recent forms emphasising standard gene sequences. This might be justified in the sense that species boundaries can be more accurately inferred in recent organisms where multiple sources of information are available (osteology, soft anatomy, behaviour, genetics). It is generally accepted that palaeospecies are only crude approximations of biological species: a single palaeospecies might be a complex of cryptic reproductively isolated lineages, while two palaeospecies might represent a highly polymorphic lineage. Under this view, different taxonomic systems might be appropriate, reflecting vastly different degrees of scientific precision.

However, an alternative view is that there is no fundamental difference between concepts of fossil and living species. In both cases, if one accepts that species are lineages, the boundaries of species are defined according to reproductive isolation. However, these boundaries are rarely directly observed (even in recent species) but rather, are usually inferred through more readily observable traits, such as morphology, behaviour and molecular sequences. In some well studied extant species, these boundaries have been very precisely inferred using a battery of different data sources, but in most extant species, only morphological similarity has been used to infer reproductive isolation. Highly morphologically divergent organisms are considered different species under the assumption that such disparity is indicative of reproductive isolation, an hypothesis that can undergo further testing with additional data. The ontological status of such morphologically inferred recent species is thus no different from that of fossil species. Under this view, it might seem better to have a taxonomic system that can be applied to all life, rather than to only the portion of life that existed recently enough to retain amplifiable DNA.

Molecular diagnoses

The above discussion suggests that any attempt to base taxonomy on a standard gene region (or regions) is unworkable; while Tautz *et al*. (2003) briefly acknowledge (in a box set aside from the main text) the utility of other lines of evidence in constructing taxonomies, the title of their paper and the bulk of their discussion advocates making standard DNA sequences 'central' to taxonomy. While sequences from genes with appropriate evolutionary dynamics are highly informative tools for inferring species boundaries and relationships, the appropriate genes will vary according to the problem investigated. Use of divergences in a standard gene region for all such studies is too blunt an instrument; no single gene (or for that matter, any other trait) is likely to be a panacea for all taxonomic problems. Some recent discussion has conflated the two issues, by stating the undoubted utility of molecular methods and using this observation to support a formalised molecular taxonomy based on divergences in one or a few standard genes (Proudlove and Wood 2003)—an unwarranted extrapolation. However, once species have been delimited using a plurality of methods, assigning organisms to these species using molecular methods is a promising prospect. Hebert *et al*. (2003) have demonstrated that once a standard gene (in their example, mitochondrial *CO1*) has been sequenced from a broad cross-section of organisms, an unknown organism may be identified from its sequence for this gene. Fast neighbour-joining cluster analysis will link the unknown sequence with some species(s) in the database, usually its closest relatives. In their example, unknown taxa were identified at broad levels ('phylum' and 'order') using amino acid sequences, and at finer levels ('genus' and 'species') using nucleotide sequences. It should be stressed that species limits are not defined using *CO1* sequences alone, but rather inferred using a plurality of methods to accurately gauge the boundaries of reproductive isolation.

Once these species boundaries are known, however, *CO1* sequences can be used as a tool (like any other trait) to help place a specimen with respect to these boundaries. The question of interest is whether a standard gene sequence is more reliable and economical than other markers used, especially morphology.

There is a large and growing literature documenting the ability of molecular markers for inferring relationships around the species level where traditional morphological approaches have failed (e.g. Leaché and Reeder 2002; Morando *et al*. 2003). For certain conspicuous, well known groups (e.g. butterflies, terrestrial vertebrates, higher plants) there is enough morphological distinctness and existing expertise to make visual morphological identification usually more efficient than molecular identification. In such groups, morphological traits might function at least as well as any single molecular marker (e.g. Dunn 2003). For other groups, molecular techniques could be the only feasible method. The broad-scale use of molecular species identification has numerous undoubted advantages. Such techniques might be more tractable than morphological methods in group where morphology is either simple or difficult to examine macroscopically: perhaps viruses, bacteria, protists and nematodes (see Hebert *et al*. 2003; Blaxter and Floyd 2003; Proudlove and Wood 2003). Molecular methods might also be useful in megadiverse groups where many taxa are not yet described and in identification keys: molecular approaches would be able to quickly relate specimens to known species that are most genetically most similar. Such methods can readily identify different life-cycle stages, or processed tissue (e.g. cetacean products), which are often difficult to identify using morphological methods (Baker *et al*. 2003). Also, the technique uses very broad generic skills—molecular sequencing—so that a trained molecular taxonomist can work across all of life. In contrast, a morphological taxonomist has expertise that is only applicable to particular organisms. Even though he or she will have generic skills, such as ability to read identification keys and understand systematic concepts, different sets of traits and techniques would need to be learned for different group of organisms. Finally, the cost-effectiveness and accessibility of molecular identification will improve as sequencing technology becomes faster, cheaper and easier to use.

However, the broad-scale use of molecular species identification carries some undesirable consequences. It is largely designed to circumvent, rather than rectify, the increasing lack of morphological taxonomic expertise. Instead of training biologists who can identify, observe and study organisms in the field, it could instead train technicians who can only identify organisms after grinding them up and feeding them into a machine. In particular, to be of use to most biologists and the general public, descriptions and diagnoses of most species will need to describe salient morphological features, rather than molecular sequences. Indeed, even in apparently well known groups such as mammals, there are now severe problems generated by molecular systematists without enough whole organismal expertise to properly identify the specimens sequenced (Ruedas *et al*. 2000). There will be little hope of conceptual advances unless biologists can relate gene sequences to the anatomy and biology of the relevant organisms, using species descriptions compiled by trained whole-organism biologists. Rather, as stressed by Lipscomb *et al*. (2003: 65), much of biology will be reduced to the situation characterising unculturable prokaryotes, where systematists 'collect sequence data from the environment, compile data bases of the results, and construct 'classifications' that reflect only the degree of similarity displayed by those sequences'. In particular, as stressed above, a taxonomic framework based on multiple sources of information needs to be already in place before molecular taxonomic identification can be used as a convenient (though not infallible) tool to assign individual organisms to species (Baker *et al*. 2003). Such a sound taxonomic framework, where species (lineages) are comprehensively characterised based on morphology, ecology and genetics, is not yet in place for most of life. A premature shift towards taxonomic identification based on a standard gene might delay its construction.

Finally, there are implications for the conservation and appreciation of biodiversity. One cost-effective, efficient and socially beneficial method of addressing the impending shortfall in taxonomists has been to train parataxonomists: people (often locals from developing countries) with enough taxonomic skills to visually assign organisms to morphospecies (or morphogenera) with high accuracy. These informally trained biologists with field and technical skills, appreciation of organisms, and commitment to conservation often become local ambassadors and activists for biodiversity and environmental sustainability (e.g. Basset *et al*. 2000). Also, the approach of employing local parataxonomists brings educational and economic benefits to remote regions and engenders mutually beneficial interaction between biologists and indigenous people. If molecular taxonomy is over-emphasised, such outreach will be stymied, with the study of biodiversity, and its funding, shifting away from local people and field facilities towards molecular biologists and laboratories centralised in large cities.

Acknowledgments

I thank Adam Skinner, Vera Weisbecker and other members of the Evolutionary Biology Unit discussion group at the South Australian Museum and three anonymous reviewers for valuable comments and the Australian Research Council for funding.

References

Avise, J. C. (1994). 'Molecular Markers, Natural History, and Evolution.' (Chapman and Hall: New York, USA.)

Avise, J. C. (1999). 'Phylogeography: The History and Formation of Species.' (Harvard University Press: Cambridge, MA, USA.)

Baker, C. S., Dalebout, M. L., Lavery, S., and Ross, H. A. (2003). www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. *Trends in Ecology & Evolution* **18**, 271–272. doi:10.1016/S0169-5347(03)00101-0

Ballard, J. W. O., Chernoff, B., and James, A. C. (2002). Divergence of mitochondrial DNA is not corroborated by nuclear DNA, morphology, or behaviour in *Drosophila simulans*. *Evolution* **56**, 527–545.

Basset, Y., Novotny, V., Miller, S. E., and Pyle, R. (2000). Quantifying biodiversity: experience with parataxonomists and digital photography in Papua New Guinea and Guyana. *Bioscience* **50**, 899–908.

Blaxter, M., and Floyd, R. (2003). Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology & Evolution* **18**, 268–269. doi:10.1016/S0169-5347(03)00102-2

de Queiroz, K. 1998, The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In 'Endless Forms: Species and Speciation'. (Eds D. J. Howard and S. H. Berlocher.) pp. 57–75. (Oxford University Press: New York, USA.)

de Queiroz, K., and Gauthier, J. (1992). Phylogenetic taxonomy. *Annual Review of Ecology and Systematics* **23**, 449–480. doi:10.1146/ANNUREV.ECOLSYS.23.1.449

Dunn, C. P. (2003). Keeping taxonomy based in morphology. *Trends in Ecology & Evolution* **18**, 270–271. doi:10.1016/S0169-5347(03)00094-6

Eberhard, W. G. 1985. 'Sexual Selection and Animal Genitalia.' (Harvard University Press: Cambridge, MA, USA.)

Ereshefsky, M. 2001. 'The Poverty of the Linnaean Hierarchy: A Philosophical study of Biological Taxonomy.' (Cambridge University Press: Cambridge, UK.)

Ferguson, J. W. H. (2002). On the use of genetic divergence for identifying species. *Biological Journal of the Linnaean Society* **75**, 509–516. doi:10.1046/J.1095-8312.2002.00042.X

Fitzpatrick, B. M. (2002). Molecular correlates of reproductive isolation. *Evolution* **56**, 191–198.

Gatesy, J., DeSalle, R., and Wheeler, W. C. (1993). Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Molecular Phylogenetics and Evolution* **2**, 152–157.

Goldman, N. (1988). Methods for discrete coding of variable morphological features for numerical analysis. *Cladistics* **4**, 59–71.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **270**, 313–321. doi:10.1098/RSPB.2002.2218

Hillis, D. M., and Wiens, J. J. (2000). Molecules versus morphology in systematics. In 'Phylogenetic Analysis of Morphological Data'. (Ed. J. J. Wiens.) pp. 1–18. (Smithsonian Institution Press: Washington, DC, USA.)

Hull, D. L. 1988. 'Science as a Process.' (The University of Chicago Press: Chicago, IL, USA.)

Johns, G. C., and Avise, J. C. (1998). A comparative study of genetic distance in the vertebrates from the mitochondrial cytochrome *b* gene. *Molecular Biology and Evolution* **15**, 1481–1490.

Leaché, A. D., and Reeder, T. W. (2002). Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood and Bayesian approaches. *Systematic Biology* **51**, 44–68. doi:10.1080/106351502753475871

Lee, M. S. Y. (2003). Species concepts and species reality: salvaging a Linnaean rank. *Journal of Evolutionary Biology* **16**, 179–188. doi:10.1046/J.1420-9101.2003.00520.X

Lipscomb, D., Platnick, N., and Wheeler, Q. (2003). The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology & Evolution* **18**, 65–66. doi:10.1016/S0169-5347(02)00060-5

Mayr, E. 1963. 'Animal Species and Evolution.' (Harvard University Press: Cambridge, MA, USA.)

Morando, M., Avila, L. J., and Sites, J. W., Jr (2003). Sampling strategies for delimiting species: genes, individuals, and populations in the *Liolaemus elongatus-kreigi* complex (Squamata: Liolaemidae) in the Andean-Patagonian South America. *Systematic Biology* **52**, 159–185.

Nei, M., and Kumar, S. 2000. 'Molecular Evolution and Phylogenetics.' (Oxford University Press: Oxford, UK.).

Olson, L. E., and Yoder, A. D. (2002). Using secondary structure to identify ribsomal numts: cautionary tales from the human genome. *Molecular Biology and Evolution* **19**, 93–100.

Page, R. D. M., and Holmes, E. C. (1998). 'Molecular Evolution: A Phylogenetic Approach.' (Blackwell Science: Oxford, UK.)

Proudlove, G., and Wood, P. J. (2003). The blind leading the blind: cryptic subterranean species and DNA taxonomy. *Trends in Ecology & Evolution* **18**, 272–273. doi:10.1016/S0169-5347(03)00095-8

Richardson, B. J., Baverstock, P. R., and Adams, M. (1986). 'Allozyme Electrophoresis: A Handbook for Animal Systematics and Population Structure.' (Academic Press: Sydney, Australia.)

Ruedas, L. A., Salazar-Bravo, J. S., Dragoo, J. W., and Yates, T. L. (2000). The importance of being earnest: what, if anything, constitutes a "specimen examined?" *Molecular Phylogenetics and Evolution* **17**, 129–132. doi:10.1006/MPEV.2000.0737

Ryan, M. J. (2001). 'Anuran Communication.' (Smithsonian Institution Press: Washington, DC, USA.)

Seberg, O., Humphries, C. J., Knapp, S., Stevenson, D. W., Petersen, G., Scharff, N., and Andersen, N. M. (2003). Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution* **18**, 63–65. doi:10.1016/S0169-5347(02)00059-9

Spamer, E. E. (1999). Know thyself: responsible science and the lectotype of *Homo sapiens* Linnaeus 1758. *Proceedings. Academy of Natural Sciences of Philadelphia* **149**, 109–114.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., and Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution* **18**, 70–74. doi:10.1016/S0169-5347(02)00041-1

Thiele, K. (1993). The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* **9**, 275–304. doi:10.1006/CLAD.1993.1020

Wagner, G. P. ed. (2001). 'The Character Concept in Evolutionary Biology.' (Academic Press: San Diego, CA, USA.)

Wheeler, W. C. (2001). 'Homology and DNA sequence data. In 'The Character Concept in Evolutionary Biology'. (Ed. G. Wagner.) pp. 303–317. (Academic Press: San Diego, CA, USA.)

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851–865. doi:10.1046/J.1420-9101.2001.00335.X

Manuscript received 16 June 2003; revised and accepted 22 September 2003.