

Taxonomy: renaissance or Tower of Babel?

James Mallet^{1,2} and Keith Willmott²

¹Galton Laboratory, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK

²Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Taxonomy, the science of naming and classifying organisms, is the original bioinformatics and a fundamental basis for all biology. Yet over the past few decades, teaching and funding of taxonomy has declined. Last year, taxonomy suddenly became fashionable again, and revolutionary approaches to taxonomy using DNA and Internet technology are now being contemplated. For examples, see the article by Tautz *et al.* in this issue of *TREE*, and a separate paper by Hebert *et al.* in *Proc. R. Soc. Lond. Ser. B*. The new excitement about taxonomy is driven partly by advances in technology, and partly by newly perceived needs given the biodiversity crisis. To reform and build on what taxonomists have already accomplished, the biology community must now begin to seek consensus, and avoid fragmenting into vociferous subdisciplines with multiple, competing aims.

DNA sequences are much used in phylogenetic analysis because of the many potential combinations in only a few hundred base pairs. For the same reason, DNA sequences can also be useful in identification, and Hebert *et al.* [1] (see also [2]) now propose a central or mandatory role for DNA in defining and identifying species. The DNA sample and its sequence readout would act simultaneously as a key part of the type specimen [2] and as a kind of label for the taxon to which the specimen is deemed to belong [1,2]. If the original species name were superseded (Box 1), it is proposed that the DNA sequence would serve as the primary key for information retrieval, similar to a supermarket barcode [1].

In principle, a single DNA sequence is no different from a single unique name for every taxon – DNA taxonomy would still suffer the problem of name changes because the group of organisms circumscribed by a name expands and contracts according to differences of opinion [3,4] (Box 1). Most name changes that annoy biologists today are due not to confusion over names applied to type specimens, but to changed concepts of the taxa that include those types (Box 1).

One must also decide which sequence to use. Any particular gene can become uninformative where species are closely related [2]. This difficulty arises for two reasons. First, ancestral polymorphisms may persist for

millions of generations after speciation. Second, genes may introgress between closely related species long after intraspecific coalescence would otherwise have fixed divergent alleles. There are now excellent examples of identical or near-identical sequences in related species, even in well studied higher animals and flowering plants [5–8]. Just as we would not want to use a single morphological criterion to define or identify species, it is best not to rely on a single sequence as proposed by Hebert *et al.* [1]. We will almost always need several genomic regions to distinguish closely related species [2,9], a task where identification becomes tricky under any system.

We might be only one tenth of the way through describing the world's species [10], so one must also question whether it is sensible to add an extra requirement to the already slow process of describing new taxa, even if funds became available for DNA taxonomy. Therefore, we doubt DNA taxonomy will catch on as a mandatory step for species description in all organisms, and we believe that most biologists will prefer to see DNA sequence information as a supplement rather than a replacement for morphological data.

Nonetheless, today's Botanical and Zoological Codes do not specify particular characters for diagnosing new taxa [11,12], so DNA taxonomy is already valid (even though descriptions of visible features would be of more immediate use and are certainly more interesting than sequences to read). Microorganism species are routinely delimited using molecular methods [13] and, to elucidate the universal tree of life, it would clearly be useful to sequence the same genes in many taxa [1]. A 'horizontal genome project' and DNA archiving system to achieve this should command widespread support regardless of whether DNA becomes a requirement in species description.

Changing the rules for web taxonomy

Many Internet taxonomy initiatives exist, perhaps too many: All-Species (<http://www.allspecies.org/>), GBIF (<http://www.gbif.org/>), Species 2000 (<http://www.sp2000.org/>) and Tree of Life (<http://www.tolweb.org/tree/phylogeny.html>) are among the grandest projects, but >50 other projects exist worldwide. Charles Godfray [14] is only one of the most recent to propose integration of all taxonomy within a single peer-reviewed web portal, to replace our current system of species descriptions scattered across hundreds of hard-to-find print

Corresponding author: James Mallet (j.mallet@ucl.ac.uk).

Box 1. Why do names keep changing?

Names of taxa can change when earlier descriptions of the same taxon are rediscovered, or when the original description is found to refer to a different taxon. By the rules of nomenclature, the name must either change, or an application for conservation or suppression is required. However, these problems are largely solved in well studied groups, and there are mechanisms in the Codes for suppression of poorly defined names in more difficult groups. Once this basic taxonomy has been done, few further name changes should result from these issues.

Even after the correct associations between names and type specimens have been established, names can continue to change. Taxonomists differ in how they circumscribe species [a], largely because of conflicting opinions [b] rather than because of new information. For example, in Fig. 1, we show a distribution map of the imaginary taxon *Symbolus*, whose taxa are all described. The

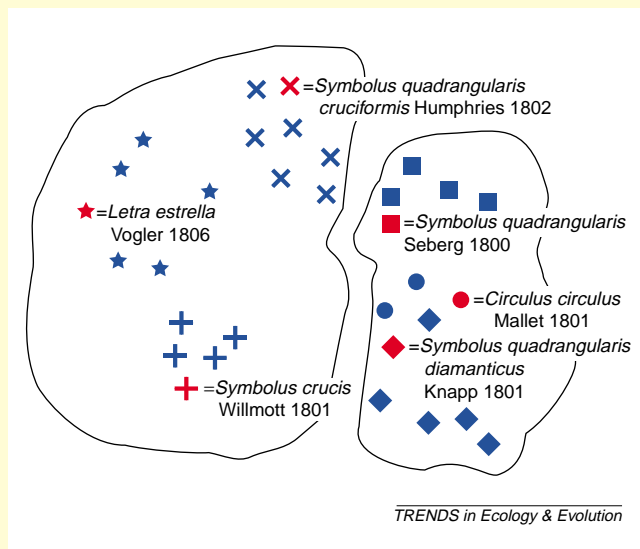


Fig. 1.

enclosing lines represent the species-level taxonomy applied by the imaginary taxonomist Mayne in a recent revision. Type specimens are shown in red with their names, authors and dates; other specimens are shown in blue.

Mayne recognizes two species in the genus, which, by rules of priority, must be called *Symbolus quadrangularis* and *S. crucis*. Mayne separates the two species because they overlap without intergrading, and because each forms a monophyletic unit in his estimated phylogeny. He suggests that all the other names apply to valid subspecies, except *circulus*, which was named by Mallet and placed in its own genus *Circulus*. Mayne shows convincingly that this name represents a hybrid from an intersubspecific hybrid zone between *S. q. quadrangularis* and *S. q. diamanticus*, and is therefore not a valid taxon.

Previous taxonomists might have had different ideas. Another taxonomist, call him Cracraft, could have argued in 1989 that all the names apply to good species on the grounds that each has its own diagnostic morphological characters, and that there is 'a parental pattern of ancestry and descent' within each taxon. An earlier circumscription by Rensch, in his review of the genus in 1930, could have included all of the taxa within the 'superspecies' *Symbolus quadrangularis*.

Thus, the names of the species and genera to which *estrella* belonged would have changed from *Letra estrella* in 1806 to *Symbolus quadrangularis* in 1930 to *Symbolus estrella* in 1989 to *Symbolus crucis* today. This instability would have been entirely due to changes in taxonomic philosophy since 1806.

Note: The taxonomic decisions in this box are for illustration only, and do not necessarily represent those that actual biologists would have taken. However, differences of opinion, such as those between Mayne and Cracraft in this hypothetical case, are extremely common today, leading to great potential instability.

References

- a Seberg, O. *et al.* (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18 DOI: 10.1016/S0169-5347(02)00059-9
- b Vane-Wright, R.I. (2003) Indifferent philosophy versus Almighty Authority: on consistency, consensus and unitary taxonomy. *Syst. Biodiv.* in press

journals. In addition, he proposes a 'first web revision' as the new starting point for nomenclature in every group; synonyms from earlier publications would be invalidated from then on. Alessandro Minelli, President of the International Commission for Zoological Nomenclature (and co-author on the DNA taxonomy proposal [2]), also argues in this issue of *TREE* for a mandatory central registry of names [15], although web taxonomy is not part of his proposal.

The bacterial code already has a unitary taxonomy: all valid bacterial names are now published in a single journal, with a new starting date for nomenclature of 1980 (<http://www.dsmz.de/bactnom/bactname.htm>). Similar reforms to macro-organism codes [11,12] might seem sensible, but zoologists and botanists have yet to ratify such proposals. The possibility that a few specialists might monopolize nomenclature was among the worries that led botanists to reject calls for a central registry of names at the International Botanical Congress in 1999 [12]; similar proposals were also rejected in the 1999 Zoological Code [11]. In addition, because many name changes are due to

differing species concepts rather than to confusion about name priority or identity (Box 1), stability of names would be unlikely to result even if a single registry became mandatory.

One is reminded of the furore about 'rank-free taxonomy' and the PhyloCode (<http://www.phylocode.org>) a few years ago. Supporters hoped to replace Linnaean nomenclature based on ranks (families, genera, species and subspecies; see Box 1) with a uninominal system for naming clades. Some versions would have done away with species altogether. The PhyloCode has not surfaced in the current unitary taxonomy debate, and the new proposals wisely sidestep the tricky issue of what constitutes a nameable taxon, as do existing codes of nomenclature.

Unitary taxonomy, DNA taxonomy and the PhyloCode proposals all argue that existing rules of nomenclature are inadequate. These codes derive from centuries of debate. For all their weaknesses, they are impressive achievements that can be adapted to reflect new needs. Reforms to nomenclature are continually needed, but the success and universality of

our current system requires reformers to act with sensitivity and only with broad consensus, to avoid fragmenting existing knowledge into multiple, incompatible systems.

The taxome project?

Taxonomic information could become much more unitary even under existing codes. GenBank and EMBL did not become primary sources of DNA sequence information by decree. They simply provided the best information, which is why we all now use them. Similarly, universal web taxonomy [10,14] does not depend on changing taxonomic rules. It requires establishment of a stable, well funded, and user-friendly information source that is open to all. Once a widely used taxonomic system exists, its architects will be able to sway the opinions of the minority of dissenters, and rules of nomenclature would change to suit the new medium. There is now real hope that the 'Encyclopaedia of Life' will obtain financial backing and achieve some of these aims [10]. The program will be especially effective if a drive for web delivery is coupled with a massive push to fund new descriptive taxonomy [10,14].

How will this all be funded? Internationally the picture looks mixed. In the USA, talk about taxonomy has led to hard cash – hardly a year goes by without a major competitive grants initiative in systematics, such as the US\$10 million programme for 2003 entitled 'Planetary Biodiversity Inventories: Mission to an (Almost) Unknown Planet' (<http://www.nsf.gov/pubs/2002/nsf02186/nsf02186.htm>). In stark contrast, a House of Lords report [16] concluded that taxonomy remains critically underfunded in the UK. Unfortunately, the Lords' recommendations were few and vague. They failed to argue for the competitive funds that would have made taxonomy attractive to UK universities under the Research Assessment Exercise (RAE) ranking system. Even if the Lords' spending recommendations were taken up (which currently seems unlikely), taxonomy will continue as uncompetitive science in the UK for the foreseeable future. In the European Community, the situation looks equally bleak: there is no taxonomy work-package in the current Framework VI call for proposals. Meanwhile, developing countries lack resources to implement major programmes similar to those in the USA, but are using improvements in their economic situation, as well as help from abroad to put together impressive national programmes. Latin American countries, for instance, are rapidly filling vacancies left behind by gaps in European taxonomy.

Conclusion

Biodiversity is in crisis, and taxonomy is now in vogue again. This newfound enthusiasm may well lead to upheavals in the nature of taxonomy itself. So far, there has been a lot of waste and very little progress.

Competing proposals, organizations and websites abound. Important people jet frequently to international biodiversity conferences in expensive locales, while few improvements in taxonomy are yet evident. We must now start to fund and carry out the basic and difficult descriptive work required to map the diversity of life. The good news is that the process seems to be gathering momentum.

Reform to taxonomy could be beneficial, but would-be reformers must now work towards consensus on the changes they advocate. Widely circulated articles, such as those in this issue of *TREE*, are only a first step. The alternatives are bleak. Minority initiatives might be established that risk fragmenting international taxonomic communication, or, more likely, all the excitement will simply be forgotten in a few years, and we will lose the opportunity provided by today's impetus for global taxonomy.

Acknowledgements

We are grateful for comments by A. Brower, V. Bull, C. Humphries, S. Knapp, J. Longino, J. Nicklin, F. Sperling, D. Vane-Wright, A. Whinnett and M. Zimmermann.

References

- 1 Hebert, P.D.N. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. Ser. B* in press
- 2 Tautz, D. *et al.* (2003) A plea for DNA taxonomy. *Trends Ecol. Evol.* 18 DOI: 10.1016/S0169-5347(02)00041-1
- 3 Seberg, O. *et al.* (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18 DOI: 10.1016/S0169-5347(02)00059-9
- 4 Lipscomb, D. *et al.* (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol.* 18 DOI: 10.1016/S0169-5347(02)00060-5
- 5 Kliman, R.M. *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156, 1913–1931
- 6 Machado, C.A. *et al.* (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and its close relatives. *Mol. Biol. Evol.* 19, 472–488
- 7 Beltrán, M.S. *et al.* (2002) Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol. Biol. Evol.* 19, 2176–2190
- 8 Chase, M.W. *et al.* (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Ann. Bot.* in press
- 9 Mallet, J. (2001) Species, concepts of. In *Encyclopedia of Biodiversity* (Vol. 5) (Levin, S.A., ed.), pp. 427–440, Academic Press
- 10 Wilson, E.O. (2003) The encyclopaedia of life. *Trends Ecol. Evol.* 18 DOI: 10.1016/S0169-5347(02)00040-X
- 11 International Commission on Zoological Nomenclature, (1999) *International Code of Zoological Nomenclature*, International Trust for Zoological Nomenclature
- 12 Greuter, W. *et al.* (2000) *International Code of Botanical Nomenclature (St Louis Code)*, Gantner/Koeltz
- 13 Claridge, M.F. *et al.* (eds) (1997) In *Species: The Units of Biodiversity* Chapman & Hall
- 14 Godfray, H.C.J. (2002) Challenges for taxonomy. *Nature* 417, 17–19
- 15 Minelli, A. (2003) The status of taxonomic literature. *Trends Ecol. Evol.* 18 DOI: 10.1016/S0169-5347(02)00051-4
- 16 House of Lords, (2002) *What on Earth? The Threat to the Science Underpinning Conservation*, H.M. Stationery Office Ltd