

Challenges for taxonomy

The discipline will have to reinvent itself if it is to survive and flourish.

H. Charles J. Godfray

Taxonomy, the classification of living things, has its origins in ancient Greece and in its modern form dates back nearly 250 years, to when Linnaeus introduced the binomial classification still used today. Linnaeus, of course, hugely underestimated the number of plants and animals on Earth. As subsequent workers began to describe more and more species, often in ignorance of each others' work, the resulting confusion and chaos threatened to destroy the whole enterprise while still in its infancy. In today's jargon, we might call this the first bioinformatics crisis. Using the tools then available, nineteenth-century taxonomists solved this crisis in a brilliant way that has served the subject well since then. They invented a complex set of rules that determine how a species should be named and associated with a type specimen; how generic and higher taxonomic categories should be handled; and how conflicts over the application of names should be resolved. All these rules revolved around publications in books and scientific journals, and their descendants form the current codes of zoological and biological nomenclature.

But today much of taxonomy is perceived to be facing a new crisis — a lack of prestige and resources that is crippling the continuing cataloguing of biodiversity. In the United Kingdom, a Parliamentary Select Committee is currently conducting an enquiry into the health of the subject for the second time in 10 years, and similar concerns are being expressed around the world. In this article I shall first explore why descriptive taxonomy is in such straits (in contrast, its sister subject, phylogenetic taxonomy, is flourishing). Then, after this essentially negative exercise, I will argue that taxonomy can prosper again, but only if it reinvents itself as a twenty-first-century information science. It needs to adopt some of the solutions that molecular biologists have developed to cope with the second bioinformatics crisis: the huge explosion of sequence, genomic, proteomic and other molecular data.

The problem

Why can't descriptive taxonomy attract large-scale funds in the same way as other big programmes like the Human Genome Project or the Sloan Digital Sky Survey? All three projects are enabling science: not in themselves generating new ideas or testing hypotheses, but allowing many new areas of research to be opened up.

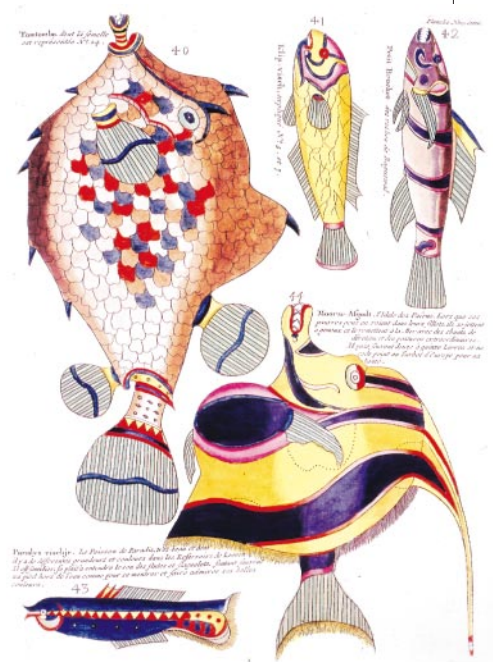
One reason is that taxonomists lack clearly achievable goals that are both realistic and relevant. Of course it would be great to describe every species of organism on Earth, but we are still monumentally uncertain as to how many species there are (probably somewhere between 4 million and 10 million); this goal is just not realistic at present. There are various projects aimed at listing, for example, all the valid described species of animal in Europe, or butterflies on Earth (see Box 1, overleaf). These aims are eminently achievable and very worthwhile, but the results are like raw, unannotated DNA sequences: unexciting and of relatively little value in themselves to non-specialists. Taxonomists need to agree on deliverable projects that will receive wide support across the biological and environmental sciences, and attract public interest.

A second problem is part of the legacy of more than 200 years of systematics. Many taxonomists spend most of their career trying to interpret the work of nineteenth-century systematists: deconstructing their often inadequate published descriptions, or scouring the world's museums for type material that is often in very poor condition. A depressing fraction of published systematic research concerns these issues. In some taxonomic groups the past acts as a dead weight on the subject, the complex synonymy and scattered type material deterring anyone from attempting a modern revision. As Frank-Thorsten Krell pointed out in Correspondence (*Nature* 415, 957; 2002), "original descriptions have to be referred to for ever, independent of the paper's quality".

The problems do not always lie in the past. Even today, many species are being described poorly in isolated publications, with no attempt to relate a new taxon to existing species and classifications. Many of these 'new' species will have been described before, so sorting out the mess will be the headache of the next generation of taxonomists. It is not surprising if funding bodies view much of what taxonomists do as poor value for money.

One of the astonishing things about

This discipline is made for the web: it is information-rich and often requires copious illustrations.



From paper to screen: is it time for taxonomy to break with tradition and unify on the Internet?

being a scientist at this particular time in history is the vast amount of information that is available, essentially free, via one's desktop computer. I can download the sequences of millions of genes, the positions of countless stars. Yet, with a few wonderful exceptions, the quantity of taxonomic information available on the web is pitiful, and what is present (typically simple lists) is of little use to non-taxonomists. But surely taxonomy is made for the web: it is an information-rich subject, often requiring copious illustrations. At present, the output of much taxonomy is expensive printed monographs, or papers in low-circulation journals available only in specialized libraries. These are not attractive 'deliverables' for major research funders.

Two models of taxonomy

The taxonomy of a group of organisms does not reside in a single publication or a single institution, but instead is an ill-defined integral of the accumulated literature on that group. The literature is bound together and cross-references itself using the venerable rules of taxonomy encapsulated in the codes. But this is not the only way to organize a taxonomy. The taxonomy of a particular group could reside in one place and be administered by a single organization. It could be self-contained and require reference to no other sources.

My main argument is that to address the problems outlined above, and for taxonomy to flourish now and in the future, it has to move from the first to the second model: from having a distributed to a unitary organization. Such a massive task could only be accomplished group by group, as resources became available. I believe a number of things would then follow. First, the only logical way to organize a unitary taxonomy and to make it widely available is on the web. The web is currently used, if used at all, as an adjunct to the distributed, printed taxonomy, but I think it should replace it. Second, the core of taxonomy is a description of each species and a means of distinguishing among them; to this core has been added the exercise of resolving their evolutionary relationships. I believe that taxonomy needs to expand to include other aspects of the species' biology, to become an information science that curates our accumulated knowledge of that species in the way a gene annotation in a genome database organizes our knowledge of a particular protein. Third, I think it is essential that the unitary taxonomy of different groups evolves from the present taxonomy. We must preserve the achievements of 250 years of distributed taxonomy, dispensing with the bad legacy of the past but retaining the good.

To illustrate how this could be done I

Box 1: Taxonomy on the web

The current codes of zoological and botanical nomenclature do not allow original descriptions to be made purely on the web, but nevertheless there is a substantial amount of taxonomy on the Internet. The Natural History Portal of the Natural History Museum in London (www.nhm.ac.uk/portal/index.html) provides an excellent entry into these resources, which include such sites as the International Plant Name Index (www.ipni.org) that covers all higher plants; the ant database (www.antbase.org) featured recently in *Nature's* News section (416, 115; 2002); and the Tree of Life project (tolweb.org/tree), a database of phylogenies.

The most common data available are catalogues of species names and lists of museum specimens, although some identification keys and other information-rich sites are becoming available.

An ambitious project led by Species 2000 (www.sp2000.org) and the Integrated Taxonomic Information System (www.itis.usda.gov) aims to catalogue the world's biota, and these sites themselves also link to the Global Biodiversity Information Facility (www.gbif.org), intended to be a general clearing house for biodiversity information.

Finally, the All Species Foundation (www.all-species.org) has set itself the goal of making an inventory of all species on Earth in the next 25 years.

shall sketch one possible way a unitary taxonomy might be achieved. I am not a professional taxonomist and am under no illusion that what follows will be the best or even a viable model, but I hope it will bring out the issues involved.

A unitary taxonomy

Introduce as a formal taxonomic procedure the 'first web revision'. This would be a revision of a major group of organisms to a standard decided on by the International Commission on Zoological Nomenclature, or the International Botanical Congress, or equivalent body (let's just call it the international committee). The revision would include a traditional description of each taxon and the location of type material. It might also include material not currently required in a formal description, for example keys and, for many groups, photographs or other illustrations. For some organisms a gene sequence might be required. It would also include a treatment of existing known synonyms to preserve contact with the older literature. This draft first web revision would be placed on the web for comments from the community, then after changes have been made in response, it would become the unitary taxonomy of the group.

What would this mean? First, from this time onwards all future work on the group need refer only to the set of species in the first web revision and then later to those in the '*n*th (that is, current) web revision'. The taxonomy of the group is thus at a stroke liberated from nineteenth-century descriptions and potentially undiscovered synonyms. If I think I have discovered a new species I need only to check that it is not already in the web revision. So what happens if I describe a new species and then someone discovers that Linnaeus or someone had already described it in an overlooked work? Well, that interesting nugget of historical information can be added to the species' web page, but the name doesn't change. What happens if I want to lump, split or add species, or revise their higher classification? Then I submit a revision that is mounted on the web for refereeing and comment. If, as a result, it is accepted, it becomes incorporated into the current (*n*+1th) web revision. At any one time there is just a single current web revision to which people refer, linked to all previous revisions (which are maintained on the web, so that in future I can easily see what was understood by species *x* in year *y*).

A major difference between this way of doing taxonomy and the *status quo* is that a unitary taxonomy needs administration: both the physical implementation on servers and networks, and the intellectual administration of the current web revision. One virtue of the present system is that if no one is interested in a group's taxonomy it can quietly slumber in the library. But the collections and

type material that underpin distributed taxonomies do require administration, which is currently undertaken by our great museums and herbaria. Nearly all these organizations are enthusiastically embracing modern web technologies. Hosting web revisions is something I see as a logical extension of their moves towards becoming, in part, modern information storehouses. It is absolutely clear, however, that they need more money in order to do this. They might also undertake the intellectual administration of the web revision — the refereeing and editing — although they would probably devolve this to committees drawn from a wider constituency (the equivalent of a journal's editorial board).

However it worked, standards would need to be set and monitored by the international committee, who would also determine which institute houses which taxonomy, and would prevent duplication of effort.

Advantages

I believe that what I have described is evolutionary rather than revolutionary in that it preserves the hard-won successes of current taxonomy while dispensing with the historical baggage. It is also evolutionary in that groups would move to the new unitary taxonomy as resources became available. It would set a series of achievable targets that could be used to spur major funding initiatives, for example the first web revision of mosquitoes, reptiles or plants (and I hope *Nature* or *Science* might celebrate these milestones as they do completed genome sequences).

I believe that major government and private research funders would consider construction and maintenance of a unitary taxonomy — universally accessible, and the foundation of all future work on the group — much more attractive to support than taxonomy as presently practised. It might also attract new sources of funding. It surely isn't impossible that a major company might sponsor the web revision of, say, the Lepidoptera (butterflies and moths); and if it wants to put its logo on the site, then why not?

The web revision would become an information hub, both through its contents and through its links to other sites. Links to molecular databases will facilitate the increasing usefulness of molecular techniques in species identification. There are already exciting web-based phylogenetic projects (see Box 1) that aim ultimately to build a phylogeny of all living organisms; clearly, one would build in reciprocal links to these sites. Today, a reference to a species in a scientific article usually gives just the scientific name and possibly the authority, but seldom refers (or gives credit) to the taxonomic revision upon which the identification is based. As increasing numbers of journals go electronic, the mention of a species can more and more easily be linked to its position in the current web revision. Were the status of the

species to change, the link would take you to the contemporary web revision and then forward to the current conception of the taxon. These links could also be used to produce a much-needed, fair 'citation count' for taxonomists. Finally, as an increasing amount of the scientific literature becomes available online through projects such as JSTOR (www.jstor.org), one can imagine links between a species description and important early papers on its taxonomy and biology, again maintaining links with the good legacy of distributed taxonomy.

Many taxonomic works are very hard for non-specialists to use, sometimes because of real difficulties in telling many species apart, but more often because of the telegraphic jargon and lack of illustration imposed on taxonomists by the expense of publication in print. The web has far fewer constraints, and provides the space needed for taxonomists to be understood. Taxonomy often pays insufficient attention to its 'end users', the ecologists, conservationists, pest managers and amateur naturalists who need or want to identify animals and plants. I hope that, overlaid on the current web revision, there would be higher-level information, the equivalent of the regional field guides and floras used by field workers. For many, this 'entry level' would be all that is required, but where needed the user could burrow deeper, right through to the primary taxonomic sources. Today, few people would seriously think about taking a computer into the field as a substitute for a field guide, but that will undoubtedly change and taxonomists should be ready.

Finally, the taxonomy should be available free (without access charges) to anyone who can log onto the Internet. This will raise the profile of taxonomy and increase the number of people who actually use the fruits of taxonomic research. Longer-term positive benefits will be for a new, young generation of naturalists, stalking their prey using digital cameras, downloading their captures into PCs, then identifying them over the web — exposing them to taxonomy as an active discipline, at the heart of modern biology.

Disadvantages

One disadvantage of a unitary taxonomy is the requirement for more administration, with its attendant costs. My assertion is that the advantages of a unitary taxonomy will prime sufficient new funds to counter-balance this, but if I'm wrong the project fails. There are also considerable technological challenges in developing the web software to support the taxonomies.

A possible criticism is that the proposal is top-down, at variance with the individualistic tradition of taxonomy. Would one clique be able to impose its view of how a group is classified? The international committee would be empowered to set standards, but rejected contributions to a group's taxono-



Harnessing the power of the web would allow all contributions to taxonomy to be collated.

my should also be stored on the web. Even if they are not incorporated in the current web revision they can at least influence future scholarship and research.

An important issue is the degree to which a treatment should be 'complete' before it is a candidate for a first web revision. Could a series of intractable species complexes requiring detailed research delay completion of a revision? The ideal solution would be to commission new taxonomic research to sort out these problems, but if this is not possible I would favour a category of 'provisional taxon', where the need for further study is clearly highlighted. After all, the heterochromatin-rich gaps in the human genome sequence did not delay the announcement of its 'completion'.

Is a web-based taxonomy as permanent as a paper-based one, and are people without computers disenfranchised, especially those in less wealthy countries? I believe the first is a non-issue; there is not (as far as I know) a paper back-up to the human genome database, and the international committee would set rigid standards for archiving and backup. Access is a much more important matter, but very many more people are at present disenfranchised by their inability to get to a specialist library, or to order a reprint, or even by being unaware that certain literature exists. The web-based taxonomy must be completely downloadable so that even continuous access to the Internet is not essential, and, if all else fails, a paper copy could be printed. It might spread the geographical distribution of taxonomic activity if some sites were hosted by developing countries with strengths in computing, such as India.

Conclusions

I find that the commonest reaction of taxonomists to these ideas is the worry that it is an attempted technological fix that distracts attention from what they (and I) perceive to be the overwhelmingly critical issue — the lack of people and resources devoted to descriptive taxonomy. The counter-argument is that the technological fix is not an end in itself; it is the means of making grassroots taxonomy more accessible and useful,

and thus attracting people and funds into the field. But is such a root-and-branch change in the culture of taxonomy really needed? Although there is near-universal agreement about the current depressed state of descriptive taxonomy, wouldn't more funding alone solve the problem?

I think not: indeed, descriptive taxonomy might disappear completely for 'difficult' groups such as many insects and nematodes. Just as Moore's law says that microprocessor power doubles every 18 months, there must be a parallel law that says DNA sequencing power increases geometrically. In 10 or 20 years' time it will be simpler to take an individual organism and get enough sequence data to assign it to a 'sequence cluster' (equivalent to species) than to key it down using traditional methods, let alone describe it as new. Just as bacterial taxonomy is now nearly all sequence-based, a new way of classifying insects, nematodes and perhaps even many plants and fish might evolve that is totally divorced from current taxonomy — a point also made forcibly by Robert May, president of Britain's Royal Society.

Would the death of large swathes of present-day systematics matter? Yes it would, because we would be throwing away so much of what we have learned in the past 250 years about the planet's biota, a lot of which we would then have to relearn. But unless taxonomy is unitary, web-based and able to accommodate these radical new ways of doing biology, I fear it will be sidelined.

The rigidity built into the current rules and codes of taxonomy — which include prohibition of purely electronic description — is part of their success, and changes should not be made lightly. But I suspect these rules are now a brake on progress, imprisoning the subject in outdated methodologies, and rendering it difficult or impossible to attract the major funds needed to reverse its slow decline. Surely it is time to experiment — time for the international taxonomic community to come together and countenance a unitary web revision of one or a few major groups of organisms (and to work out exactly how a unitary taxonomy should operate). This venture must be sanctioned and supported by the existing international committees, or no serious taxonomist will waste his or her time on it; no institution will administer it; and no agency will fund it. If successful, it will change how taxonomy is done for ever; if it fails it would not be difficult to revert to the *status quo ante*. There is everything to gain and little to lose. ■

H. Charles J. Godfray is at the NERC Centre for Population Biology, Department of Biological Sciences, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.

Acknowledgements

I am grateful to the many taxonomists and other biologists who have debated these issues with me.