

Progressing towards a biological names register

How taxonomy could harness the indexing and organizational powers of the Internet.

Sir— There is much current debate about the decline in the number of taxonomists at a time of increasing need to monitor and manage our biodiversity resources. Various people and organizations have called for a reinvention of taxonomy and for new strategies in the delivery of knowledge, with an emphasis on bioinformatics solutions using the Internet (see, for example, H. C. J. Godfray *Nature* **417**, 17–19; D. Agosti and N. F. Johnson *Nature* **417**, 222; 2002).

Specialist aggregators (such as the Australian Biodiversity Information Facility, AlgaeBase, Antbase and FishBase) are forming unified global biodiversity catalogues for use by general aggregators such as Species 2000 and the Integrated Taxonomic Information System, or by more general web-based biodiversity projects such as the Tree of Life. Some Internet resources integrate information from distributed sources (for example, the Species Analyst, Fungalweb, micro*scope, Australia's Virtual Herbarium, and Global Searcher). What is missing is a unifying device to draw all these together on the Internet, and the tools to use the indexing and organizational power of taxonomy.

Virtually every biological database uses names, so these have the potential to act as a unifying device for biodiversity bioinformatics. This has not happened because — despite the best efforts of the codes of nomenclature — organisms may have more than one name, and names change in response to taxonomic insights. The situation can be managed by taxonomic name servers that map alternative names against each other.

Name servers can act as a biodiversity thesaurus, reconciling alternative names for the same taxon. Reconciliation functions can be embedded in Internet search engines or website search functions to allow novices or machines to find information under all of the relevant names, not just the one used to initiate a search. At a grander level, name servers may 'read' the name fields of biological databases, match entries against a master names list, and replace or add a standard entry drawn from a universal names register. Once populated with names, name servers offer realistic prospects for low-cost, accessible warehousing of distributed biological data, and of machine-to-machine dialogue about biology.

Conceptually, biological name servers are an extension of database thesauri or traditional synonymy lists, which catalogue alternative names. Software systems capable

of managing names are not common. Platypus, a database package for taxonomists developed through the Australian Biological Resources Study, is a stand-alone computer-based software, but systems capable of working through the Internet, such as that used by the Biodiversity Conservation Information System (BCIS), have considerably more appeal because they can index and organize distributed information. The Taxonomic Name Server (TNS) — part of uBio, the Universal Biological Indexer and Organizer project (www.ubio.org) — and a new concept called Octopus are emerging as broad, Internet-based biological name servers. TNS is the most advanced, with a development version in use by micro*scope (www.mbl.edu/microscope).

There are probably about 10 million names for about 1.7 million species. Name servers must deal with synonyms (alternative names for the same entity) and homonyms (identical names for different entities). Name servers will acquire their content from aggregators, the public domain, taxonomists and the literature. They will share content with each other, and return value-added information to the public domain. At this time, aggregators probably hold only about 10% of required names, and the rate of addition is slow — comprehensive lists are said to be 10–25 years away. This creates a challenge to accelerate and coordinate the compilation.

Currently, nomenclatural information about species is assembled by alpha taxonomists and passes through taxonomic revisions en route to expert aggregators such as FishBase or AlgaeBase, and then to general aggregators such as Species 2000. The rate-limiting elements are availability of taxonomic experts and the intellectual

effort of compiling definitive information.

A different approach is required if we intend to make rapid progress in the near future. Compiling names of organisms within an Internet-based name server will evade the rate-limiting step, allowing definitive taxonomic information to be acquired from authoritative remote sites as it becomes available. Using this approach, the uBio project is developing a universal indexing system for biology and expects to deliver a comprehensive and unified list of genera early next year. Once this list is in place within a homonym-aware environment, taxonomic service providers can read large lists of species names from, for example, the National Center for Biotechnology Information or BIOSIS, and assign each species to its correct location very quickly.

A major agency is needed to coordinate such a development, and to put in place an expert review panel (of at least 200 taxonomists) to oversee continuing taxonomic input. The Global Biodiversity Information Facility (GBIF) is the most obvious contender. Such an agency can coordinate the process, standardize products and processes, establish the credentials of the peer-review process, seek financial support for the peer-review panel, represent the interests of the compilation to governments and intergovernmental bodies, and ensure its integration with other initiatives. Most significantly, the involvement of GBIF will help to ensure that the names register and associated services remain in the public domain, and that all individuals who contribute are given appropriate recognition.

David J. Patterson

School of Biological Sciences, Heydon Laurence Building, University of Sydney, Sydney, New South Wales 2006, Australia

Peer review and the rewards of open access

Sir— Eugene Koonin's suggestion in Correspondence (*Nature* **422**, 374; 2003) that a journal should reward swift peer-reviewers by equally swiftly processing any articles that they themselves submit is an interesting one. Perhaps the converse should apply too: submissions from those who persistently underperform as reviewers should have their own work reviewed on a slower timescale.

An alternative reward system for good reviewers can be used in the open-access model of publishing, in which authors of

accepted articles pay a processing charge so that there are no subscription or access charges to the journal. In this model, swift reviewers can be rewarded by a reduction in this charge. A twist is applied by some journals published by Berkeley Electronic Press (www.bepress.com), in which submitting authors do not have to pay the processing charge if they contract to provide a timely review of an agreed number of papers. But they do have to provide credit-card details and are subsequently charged if they fail to deliver their reviews on time.

Peter Newmark

Editorial Director, BioMed Central, 34–42 Cleveland Street, London W1T 4LB, UK