

Biodiversity databases should gain support from journals

Xiaolei Huang¹ and Gexia Qiao^{1,2}

¹Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China

²National Zoological Museum of China, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China

Good biodiversity data are fundamental to biodiversity research, natural resource management and conservation policy making. However, owing to challenges such as funding and scientist involvement, gathering data into public biodiversity databases is not easy [1]. Recently, a group of key journals in evolution and ecology began to implement formally a joint data-archiving policy with the aim to preserve paper-related data [2]. This is a good example of how the scientific community promotes efforts in data sharing. We suggest here that biodiversity databases should seek cooperation and coordination with biodiversity and conservation journals. Adoption of a data-archiving policy by such journals would sustainably increase the data quantity and quality of biodiversity databases.

Although biodiversity databases, such as the Global Biodiversity Information Facility (GBIF: <http://www.gbif.org>), have made substantial progress during the past few years [3], they are likely to lag behind the expanding amount of biodiversity data. GBIF represents a concerted global effort to digitize biodiversity occurrence data from natural-history collections, which archive life on Earth over the past few centuries [4]. Although a large set of data has been gathered, many blank spots (countries or regions) still exist on the map [1,5]. For example, among the more than 276 million occurrence records in the GBIF network, only 727 206 (0.26%) and 533 790 (0.19%) records are from China and Russia, respectively (<http://data.gbif.org>, accessed May 16, 2011). These numbers do not demonstrate the size and digitization degree of these two countries. The current operational mode of GBIF is likely to be responsible for this imbalance. GBIF requires each participant (country or international organization) to sign a Memorandum of Understanding; and the participants have a responsibility to organize and share their own biodiversity data through GBIF. Such a 'top-down' mode has limitations. First, a country can choose not to be a GBIF participant and, sometimes, a participant does not make expected progress because of factors such as bureaucratic inertia. Second, under this government-oriented mode, it is common that some individuals provide only some of their data. Therefore, converting scientists from data hoarders to data sharers is still a problem [1]. Third, an excessive focus on data from archived specimens in natural-history collections can impede the gathering and

incorporation of data on recent changes of species distributions under rapid global climate change.

A workable way to resolve this situation for biodiversity databases is to gain support from journals. GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) could serve as a model for biodiversity databases. A development that was crucial to the success of GenBank was convincing journal editors to make electronic submission of sequence data a condition for publication [6]. A routine data-archiving policy would benefit both the journals and biodiversity databases. If raw data were archived, journal editors and reviewers would have an idea of the completeness and quality of data sets used in papers, and could therefore enhance and assure the reliability of those papers. Most importantly, by cooperating with journals, biodiversity databases could devise a more sustainable methodology for data collection. We believe that implementation of a data-archiving policy by biodiversity and conservation journals would drive a large amount of species occurrence data across broad geographic and taxonomic ranges into biodiversity databases. Our preliminary survey of publications in the *Journal of Biogeography* during 2010 revealed that, in the 104 papers in which numbers of species and distributions had been explicitly provided or could be figured out (approximately half of all the papers), more than 367 000 occurrence records for 12 782 species were covered. Given that quite a few papers without explicit numbers of records were also derived from sizeable biodiversity data sets, an impressive amount of data could be provided by this journal alone in just one year. To match the requirements of a data-archiving policy, biodiversity databases should do more to provide standard data formats and more friendly submission tools for journal authors and data submitters.

Acknowledgments

We thank Ruiling Zhang for her help in journal data survey. XLH is funded by NSFC-30900133 and O529YX5105 from the Key Laboratory of Zoological Systematics and Evolution, Chinese Academy of Sciences; and GXQ by NSFC-31025024, 31061160186 and 30830017. We also thank Paul Craze for valuable suggestions.

References

- 1 Thomas, C. (2009) Biodiversity databases spread, prompting unification call. *Science* 324, 1632–1633
- 2 Whitlock, M.C. (2011) Data archiving in ecology and evolution: best practices. *Trends Ecol. Evol.* 26, 61–65
- 3 King, N. and Penman, D. (2009) A step toward unification. *Science* 325, 1205–1206

- 4 Edwards, J.L. *et al.* (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289, 2312–2314
- 5 Yesson, C. *et al.* (2007) How global is the global biodiversity information facility? *PLoS ONE* 2, e1124

- 6 Strasser, B.J. (2008) GenBank: natural history in the 21st century? *Science* 322, 537–538

0169-5347/\$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tree.2011.05.006 Trends in Ecology and Evolution, August 2011, Vol. 26, No. 8

Letters

The need for jumpstarting amphibian genome projects

Federico C.F. Calboli¹, Matthew C. Fisher², Trenton W.J. Garner³ and Robert Jehle⁴

¹ Department of Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London W2 1PG, UK

² Department of Infectious Disease Epidemiology, Imperial College London, Norfolk Place, London W2 1PG, UK

³ Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

⁴ Center for Environmental Systems and Wildlife Research, University of Salford, Salford M5 4WT, UK

Amphibian genomic diversity is rapidly disappearing because this class of vertebrates is facing an unprecedented decline [1]. Amphibians are the most ancient class of land-dwelling vertebrates, emerging 360 million years ago to colonise all continents except Antarctica. Their highly diversified genomes therefore hold important keys to understanding, among other traits, vertebrate terrestrialisation. Despite presenting significant hurdles to modern sequencing technology – due to an average size of 9.36 Gb for anurans to 35.90 Gb for salamanders [2] – amphibian genomes hold great promise for furthering our understanding of both vertebrate evolution and development. Here we stress that amphibian genomics is also urgently needed for amphibian conservation.

Although 41% of amphibian species are categorised as at least ‘threatened’ by the IUCN Red List (by far the highest proportion of all vertebrate groups), amphibians have benefited least from conservation efforts [3]. Vulnerable to a host of human activities, amphibians are also threatened by novel pathogens. The recent worldwide emergence of the panzootic fungus *Batrachochytridium dendrobatis* (*Bd*) and *Ranavirus* have caused catastrophic population collapses and extinctions in a number of species [4]. The 10K genomes project, a project dedicated to the collection of the genome sequences of 10 000 vertebrate species [5], lists only 13 amphibian species in its ‘top 101’ species list, and only two (*Bufo garagarizan* and *Andrias davidianus*) have funds pledged for the sequencing of their genomes. This state of play raises the concern that genome sequencing projects will be too late to achieve their full potential for future conservation of the Amphibia.

Without direct and aggressive conservation action a large number of amphibian species will be lost in the near future, either through epizootics or through anthropogenic habitat modification. Population genetic studies based on a modest number of variable loci have revealed important insights such as the spatio-temporal structure of endangered amphibian populations, and approaches based on large-scale sequence information now have the potential to accelerate amphibian conservation manyfold. For instance, genomics would help detect which loci are under selection and give a measure of the adaptive potential within populations by

revealing which genes and alleles are involved in responses to rapid natural selection, such as that posed by climate change or emerging infection [6]. Because amphibian genomes are huge, a first functional step towards their genotyping would be to initially concentrate on the transcriptome, building Expressed Sequence Tags (EST) libraries. These libraries can be used to create linkage maps [7] and to build databases of Single Nucleotide Polymorphisms (SNPs) [8], both vital tools for discovering the genes underlying adaptation, and for characterising how much variation is present at these loci in wild populations. The size of amphibian genomes also suggests a high level of gene duplication; because genome structural variations play a role in the determination of phenotypes [9], genomics will give the opportunity to identify the full extent of variation underlying adaptation. This information is necessary for scientifically managing amphibian captive breeding projects with the aim of releasing genetically-fit progeny into nature and increasing their fitness against specific threats.

We believe amphibian genomics projects will rapidly yield important insights. Many amphibian traits important for coping with environmental unpredictability, including infectious diseases, seem to be dictated by relatively simple genetic mechanisms [10] and evolutionarily ancient stress responses may be governed by fundamental molecular mechanisms with broad pleiotropic effects [11]. Recent efforts to understand the *Bd*–amphibian interaction using the recently published genome of *Silurana tropicalis* has shown that an intact innate antimicrobial-peptide immune response is critical to controlling infection, and that adaptive immunity appears to be downregulated [12]. Extending such studies to non-model species requires a focused attack on acquiring new genome sequences from affected amphibian species. Additionally, because the *Bd* panzootic is so recent, amphibian genomics will provide ‘real time’ data on the evolution and development of host–pathogen interaction, making it a perfect (multi-species) model system to understand both the evolutionary genetics and epidemiology of infectious diseases.

The wealth of published literature on amphibian biology, risk factor analyses of amphibian traits associated with decline, and the IUCN Red List, would lay the groundwork for selecting the initial panel of amphibian species. Selecting target species will be moot without the involvement of

Corresponding author: Calboli, F.C.F. (f.calboli@imperial.ac.uk).