

## **BIODIVERSITY INFORMATICS: MANAGING KNOWLEDGE BEYOND HUMANS AND MODEL ORGANISMS**

INDRA NEIL SARKAR<sup>†</sup>

*Marine Biological Laboratory, 7 MBL Street  
Woods Hole, MA 02543, USA  
E-mail: sarkar@mbl.edu*

In the biomedical domain, researchers strive to organize and describe organisms within the context of health and epidemiology. In the biodiversity domain, researchers seek to understand how organisms relate to one another, either historically through evolution or spatially and geographically in the environment. Currently, there is limited cross-communication between these domains. As a result, valuable knowledge that could inform studies in either domain often goes unnoticed.

Biodiversity knowledge has long been a valuable source for many biomedical advances [1]. Before the creation of synthetic compounds, medicinal compounds originated from solely natural plant and animal extracts. Although there are upward estimates of between 10-100 million organisms on Earth [2], biomedical research primarily focuses on only a fraction of these as “model” organisms [3]. Furthermore, much knowledge may be lost from the biomedical community because many organisms are only described within biodiversity resources. These studies can form the basis for research on evolution, speciation, and distribution, and also provide an important baseline for studies of not only conservation but also the study of emerging diseases. The integration of biodiversity knowledge from museum collections, for example, has provided significant insights into the etiology and distribution of diseases such as hantavirus [4]. Understanding the etiology of diseases and their host epidemiology may also further the development of vaccinations and treatments that can help prevent epidemics or pandemics, such as the looming threat of the avian flu [5]. For emergent diseases (e.g., malaria), biomedical researchers traditionally focus on a limited number of species (e.g., four species of *Plasmodium* [6]). This represents very little in terms of the phylogenetic diversity of diseases that are known to infect numerous other organisms (e.g., malaria affects birds, lizards, and other mammals [7-12]). The incorporation of biodiversity knowledge in the context of biomedical advances may lead to breakthrough therapies for many of the diseases that still plague human society [13, 14].

The genomic revolution has resulted in a deluge of sequence data and derivative knowledge (e.g., protein structure prediction and gene expression

---

<sup>†</sup> INS is funded in part by NSF-IIS-0241229, NSF-BDI-0421604, and the D.A.B. Lindberg Research Fellowship from the Medical Library Association.

experiments), which are the predominant data types in biomedical research. Three of the papers in this session examine how these data can be integrated, annotated, and interpreted in light of greater taxonomic sampling. First, Cadag *et al.* propose a semi-automated framework that uses a federated approach to incorporate relevant knowledge from heterogeneous resources to assist with gene annotation. Next, Ng *et al.* demonstrate a prototype application to suggest annotations for genes that are involved with biological pathways across a range of organisms. Finally, Hampikian and Andersen examine the existence and utility of genes sequence regions that are *not* present in organisms across the tree of life.

The final two papers in this session consider how sequence and sequence-derived information can be complemented with biodiversity data (e.g., morphological, ecological, and temporal data). First, Maglia *et al.* propose a framework to incorporate existing ontologies and their structures towards the development of an ontology for Amphibian morphology. Sautter *et al.* then describe a system to semi-automatically organize knowledge that is embedded in literature resources.

There has been considerable discussion in both the scientific [15-18] and popular media [5, 19] with regards to the need for methods and tools to organize and integrate biodiversity and biomedical knowledge within the context of legacy, existing, and newly generated data. Significant infrastructural and methodological advancements are needed to incorporate knowledge from both biomedical and biodiversity resources. To this end, it is hoped that the papers that follow will spark synergistic activities that benefit both biomedical and biodiversity communities.

## References

1. W. E. Muller, R. Batel, H. C. Schroder, and I. M. Muller, "Traditional and Modern Biomedical Prospecting: Part I-the History: Sustainable Exploitation of Biodiversity (Sponges and Invertebrates) in the Adriatic Sea in Rovinj (Croatia)," *Evid Based Complement Alternat Med*, vol. 1, pp. 71-82, 2004.
2. E. O. Wilson, "The encyclopedia of life," *Trends in Ecology and Evolution*, vol. 18, pp. 77-80, 2003.
3. B. L. Umminger, "Unconventional organisms as models in biological research," *J Exp Zool Suppl*, vol. 4, pp. 2-5, 1990.
4. T. L. Yates, J. N. Mills, C. A. Parmenter, T. G. Ksiazek, R. R. Parmenter, J. R. Vande-Castle, C. H. Calisher, S. T. Nichol, K. D. Abbott, J. C. Young, M. L. Morrison, B. J. Beaty, J. L. Dunnum, R. J. Baker, J. Salazar-Bravo, and C. J. Peters, "The Ecology and Evolutionary History of an Emergent

- Disease: Hantavirus Pulmonary Syndrome," *BioScience*, vol. 52, pp. 989-998, 2002.
5. D. G. McNeil, "Hitting the Flu at Its Source, Before It Hits Us," in *New York Times*. New York, 2005.
  6. M. T. Makler, C. J. Palmer, and A. L. Ager, "A review of practical techniques for the diagnosis of malaria," *Ann Trop Med Parasitol*, vol. 92, pp. 419-33, 1998.
  7. C. T. Atkinson, K. L. Woods, R. J. Dusek, L. S. Sileo, and W. M. Iko, "Wildlife disease and conservation in Hawaii: pathogenicity of avian malaria (*Plasmodium relictum*) in experimentally infected iiwi (*Vestiaria coccinea*)," *Parasitology*, vol. 111 Suppl, pp. S59-69, 1995.
  8. P. C. Garnham, "Recent research on malaria in mammals excluding man," *Adv Parasitol*, vol. 11, pp. 603-30, 1973.
  9. B. Mons and R. E. Sinden, "Laboratory models for research in vivo and in vitro on malaria parasites of mammals: Current status," *Parasitol Today*, vol. 6, pp. 3-7, 1990.
  10. R. E. Ricklefs and S. M. Fallon, "Diversification and host switching in avian malaria parasites," *Proc Biol Sci*, vol. 269, pp. 885-92, 2002.
  11. J. J. Schall, "Virulence of lizard malaria: the evolutionary ecology of an ancient parasite-host association," *Parasitology*, vol. 100 Suppl, pp. S35-52, 1990.
  12. J. J. Schall, "Lizards infected with malaria: physiological and behavioral consequences," *Science*, vol. 217, pp. 1057-9, 1982.
  13. L. A. Basso, L. H. da Silva, A. G. Fett-Neto, W. F. de Azevedo, Jr., S. Moreira Ide, M. S. Palma, J. B. Calixto, S. Astolfi Filho, R. R. dos Santos, M. B. Soares, and D. S. Santos, "The use of biodiversity as source of new chemical entities against defined molecular targets for treatment of malaria, tuberculosis, and T-cell mediated diseases--a review," *Mem Inst Oswaldo Cruz*, vol. 100, pp. 475-506, 2005.
  14. F. Pelaez, "The historical delivery of antibiotics from microbial natural products-Can history repeat?," *Biochem Pharmacol*, 2005.
  15. D. Agosti, "Biodiversity data are out of local taxonomists' reach," *Nature*, vol. 439, pp. 392, 2006.
  16. J. Soberon and A. T. Peterson, "Biodiversity informatics: managing and applying primary biodiversity data," *Philos Trans R Soc Lond B Biol Sci*, vol. 359, pp. 689-98, 2004.
  17. S. Blackmore, "Environment. Biodiversity update--progress in taxonomy," *Science*, vol. 298, pp. 365, 2002.
  18. F. A. Bisby, "The quiet revolution: biodiversity informatics and the internet," *Science*, vol. 289, pp. 2309-12, 2000.
  19. "Today we naming of parts: a global registry of animal species could shake up taxonomy," in *Economist*, 2006.