Zootaxa 3754 (2): 195–200
www.mapress.com/zootaxa/

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

http://dx.doi.org/10.11646/zootaxa.3754.2.8
http://zoobank.org/urn:lsid:zoobank.org:pub:08FD1323-600E-457E-805D-4CB0BA16C1E7

Correcting the disconnect between phylogenetics and biodiversity informatics

JOSEPH T. MILLER & GARRY JOLLEY-ROGERS
Centre for Australian National Biodiversity Research, CSIRO Plant Industry, Canberra, ACT 2601 Australia.
E-mail: joe@acaciamulga.net, garry.jolleyrogers@gmail.com

Abstract

Rich collections of biodiversity data are now synthesized in publically available databases and phylogenetic knowledge now provides a sound understanding of the origin of organisms and their place in the tree of life. However, these knowledge bases are poorly linked, leading to underutilization or worse, an incorrect understanding of biodiversity because there is poor evolutionary context. We address this problem by integrating biodiversity information aggregated from many sources onto phylogenetic trees. PhyloJIVE connects biodiversity and phylogeny knowledge bases by providing an integrated evolutionary view of biodiversity data which in turn can improve biodiversity research and the conservation decision making process. Biodiversity science must assert the centrality of evolution to provide effective data to counteract global change and biodiversity loss.

Key words: Taxonomy, phylogeny, distribution, biodiversity, database

The biodiversity research community has divergent agendas. One side, the biological collection informatics community, is driven by large-scale collections-based projects, exemplified by the *Global Biodiversity Information Facility* (GBIF), *Encyclopedia of Life* (EOL), *Integrated Digitized Biocollections* (iDigBio), *Atlas of Living Australia* (ALA) (Beaman *et al.* 2007; Matsunaga *et al.* 2013) and online floras and faunas that describe biodiversity by species, taxonomies and localities. For example, a full description of the authoritative taxonomy, distribution and ecology, including images and literature references, is available at EOL for *Chelonoidis nigra* Abingdonii, the Galapagos giant tortoise, and tens of thousands of other species. The other side, the evolution and phylogenetics community, driven largely by the National Science Foundation and academia, is building the tree of life (Open Tree of Life 2013; APGIII 2009; NSF 2013), describing biodiversity through clades, diversifications and distributions. For example, the Angiosperm Phylogeny Group provides a synthesis of the evolutionary relationships (APGIII 2009) of the angiosperms and the NSF Assembling the Tree of Life (ATOL) initiative (Choumane *et al.* 2000) seeks to construct an evolutionary history for all major lineages of life.

However, biodiversity is neither a list of taxonomic entities with morphological and geographic attributes nor solely a phylogeny with divergence dates. Comparative biology is predicated on the expectation that closely related organisms share traits—such as morphology, ecology, biogeography, disease resistance, ecosystem services—that are not common in more distantly related organisms. The rich biodiversity data available through GBIF is underutilized because it is not integrated with the tree of life: only by applying the analytical power that comes from a phylogeny can we understand the relationships of biodiversity data (Mishler 2010).

While much work remains to accomplish the individual goals of each community, a synthesis is badly needed that will make integrated biodiversity data the foundation for scientifically based biodiversity-management decisions (van der Linde & Houle 2008; Varón *et al.* 2010; Parr *et al.* 2012; Jonathan *et al.* 2013). As Cracraft (2002) noted *"the ability to search multiple databases using the nodes of a phylogenetic tree may be the single most important contribution of systematics to conservation and sustainable use of biodiversity"*. Viewed through phylogeny, the evolutionary context of morphological, spatial and ecological data becomes clear and accessible to scientists and non-scientists alike.

The global biodiversity crisis, exacerbated by climate change, necessitates that science be better integrated so that decisions are made with the best available data provided by all communities. However, taxonomy and

Accepted by M. Ebach: 5 Dec. 2013; published: 14 Jan. 2014

References

APGIII (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, 161, 105–121.
http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x

Beaman, R., Macklin, J.A., Donoghue, M.J. & Hanken, J. (2007) Overcoming the digitization bottleneck in natural history collections. A summary report on a workshop held 7–9 September 2006 at Harvard University. Available from: http://www.etaxonomy.org/wiki/images/b/b3/Harvard_data_capture_wkshp_rpt_2006.pdf. (accessed 23 December 2013)

Chapman, A.D. (2005) Uses of Primary Species-Occurrence Data, version 1.0. Global Biodiversity Information Facility, Copenhagen.

Choumane, W., Winter, P., Weigand, F. & Kahl, G. (2000) Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer aerietinum* L.) within the genus *Cicer*. *Theoretical and Applied Genetics*, 101, 269–278.
http://dx.doi.org/10.1007/s001220051479

Cracraft, J. (2002) The Seven Great Questions of Systematic Biology: An Essential Foundation for Conservation and the Sustainable Use of Biodiversity. *Annals of the Missouri Botanical Garden*, 89, 157–144.
http://dx.doi.org/10.2307/3298558

Fukami, H., Budd, A.F., Paulay, G., Solé-Cava, A., Chen, C.A., Iwao, K. & Knowlton, N. (2004) Conventional taxonomy obscures deep divergence between Pacific and Atlantic corals. *Nature*, 427, 832–835.
http://dx.doi.org/10.1038/nature02339

GBIF (2013) http://www.gbif.org/occurrence/ (accessed 2 December 2013)

Grytnes, J.-A. & Romdal, T.S. (2008) Using Museum Collections to Estimate Diversity Patterns along Geographical Gradients. *Folia Geobotanica*, 43, 357–369.
http://dx.doi.org/10.1007/s12224-008-9017-6

Jolley-Rogers, G., Varghese, T., Harvey, P., dos Remedios, N. & Miller, J.T. (Under Revision) PhyloJIVE: Integrating biodiversity data with the Tree of Life. *Bioinformatics*.

Jonathan, B.L., Stevan, J.A., Gill, B., Brodie, E.D., Hibbett, D., Hoekstra, H.E., Mindell, D.P., Monteiro, A., Moritz, C., Orr, H.A., Petrov, D.A., Renner, S.S., Ricklefs, R.E., Soltis, P.S. & Turner, T.L. (2013) Evolutionary Biology for the 21st Century. *PLoS Biology*, 11, e1001466.
http://dx.doi.org/10.1371/journal.pbio.1001466

Mast, A.R. & Thiele, K. (2007) The transfer of *Dryandra* R.Br. to *Banksia* L.f. (Proteaceae). *Austalian Systematic Botany*, 20, 63–71.
http://dx.doi.org/10.1071/sb06016

Matsunaga, A., Thompson, A., Figueiredo, R.J., Gremain-Aubrey, C.C., Collins, M., Beaman, R.S., MacFadden, B.J., Riccardi, G., Soltis, P.S., Page, L.M. & Fortis, J.A.B. (2013) *A Computational and Storage-Cloud for Integration of Biodiversity. Proceedings of the 2013 IEEE 9th International Conference on e-Science*. Beijing, IEEE, China, pp 78–87.
http://dx.doi.org/10.1109/eScience.2013.48

Miller, J.T. & Bayer, R.J. (2001) Molecular phylogenetics of *Acacia* (Fabaceae : Mimosoideae) based on the chloroplast matK coding sequence and flanking trnK intron spacer regions. *American Journal of Botany*, 88, 697–705.
http://dx.doi.org/10.2307/2657071

Mishler, B.D. (2010) Species are not uniquely real biological entities. *In*: Ayala, F. & Arp, R. (Eds.), *Contemporary Debates in Philosophy of Biology*. Wiley-Blackwell, pp. 110–122.

NSF (2013) *Assembling the Tree of Life (ATOL)*. National Science Foundation, Arlington VA, 10 pp.

OECD (1999) *Final Report of the megascience forum working group on biological informatics*. OECD, Paris, 74 pp.

Open Tree of Life (2013) Available from: http://opentreeoflife.org/ (accessed 2 December 2013)

Page, R. (2006) Taxonomic Names, Metadata, and the Semantic Web. *Biodiversity Informatics*, 3, 1–15.

Page, R.D. (2008) Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics*, 9, 345–354.
http://dx.doi.org/10.1093/bib/bbn022

Page, R.D. (2012) Space, time, form: viewing the Tree of Life. *Trends in Ecology and Evolution,* 27, 113–120.
http://dx.doi.org/10.1016/j.tree.2011.12.002

Parr, C.S., Guralnick, R., Cellinese, N. & Page, R.D.M. (2012) Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology and Evolution*, 27, 94–103.
http://dx.doi.org/10.1016/j.tree.2011.11.001

Pinto, C.M., Baxter, B.D., Hanson, J.D., Mendez-Harclerode, F.M., Suchecki, J.R., Grijalva, M.J., Fulhorst, C.F. & Bradley, R.D. (2010) Using museum collections to detect pathogens. *Emerging Infectious Diseases*, 16, 356–7.
http://dx.doi.org/10.3201/eid1602.090998

SEWPAC (2013) Statement of Outstanding Universal Value of the Great Barrier Reef. Department of Sustainability, Environment, Water, Population and Communities, Canberra, Australia. Available from: http://www.environment.gov.au/heritage/places/world/great-barrier-reef/values.html (accessed 9 September 2013)

Shaffer, H.B., Fisher, R.N. & Davidson, C. (1998) The role of natural history collections in documenting species declines. *Trends in Ecology and Evolution*, 13, 27–30.
http://dx.doi.org/10.1016/s0169-5347(97)01177-4

Suarez, A.V. & Tsutsui, N.D. (2004) The Value of Museum Collections for Research and Society. *BioScience*, 54, 66–74.
http://dx.doi.org/10.1641/0006-3568(2004)054[0066:tvomcf]2.0.co;2

Van der Linde, K. & Houle, D. (2008) A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). *Insect Systematics and Evolution*, 39, 241–267.
http://dx.doi.org/10.1163/187631208788784237

Varón, A., Vinh, L.S. & Wheeler, W.C. (2010) POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*, 26, 72–85.
http://dx.doi.org/10.1111/j.1096-0031.2009.00282.x