LINKING BIG: THE CONTINUING PROMISE OF EVOLUTIONARY SYNTHESIS

Brian Sidlauskas,[1,2,3,4] **Ganeshkumar Ganapathy,**[1] **Einat Hazkani-Covo,**[1,5] **Kristin P. Jenkins,**[1] **Hilmar Lapp,**[1] **Lauren W. McCall,**[1] **Samantha Price,**[1,6] **Ryan Scherle,**[1] **Paula A. Spaeth,**[1,7] **and David M. Kidd**[1,4,8]

[1]*National Evolutionary Synthesis Center, Suite A200, 2024 West Main Street, Durham, North Carolina 27705*

  [2]*E-mail: brian.sidlauskas@oregonstate.edu*

[3]*Department of Fisheries and Wildlife, Oregon State University, 104 Nash Hall, Corvallis, Oregon 97331*

[5]*Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710*

[6]*Section of Ecology and Evolution, University of California Davis, 1 Shields Avenue, Davis, California 95616*

[7]*Department of Biology and Natural Resources, Northland College, 1411 Ellis Avenue, Ashland, Wisconsin 54806*

[8]*NERC Centre for Population Biology, Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire, SL5 7PY, United Kingdom*

Synthetic science promises an unparalleled ability to find new meaning in old data, extant results, or previously unconnected methods and concepts, but pursuing synthesis can be a difficult and risky endeavor. Our experience as biologists, informaticians, and educators at the National Evolutionary Synthesis Center has affirmed that synthesis can yield major insights, but also revealed that technological hurdles, prevailing academic culture, and general confusion about the nature of synthesis can hamper its progress. By presenting our view of what synthesis is, why it will continue to drive progress in evolutionary biology, and how to remove barriers to its progress, we provide a map to a future in which all scientists can engage productively in synthetic research.

KEY WORDS: Data sharing, education, informatics, interdisciplinary, ontology, scientific culture, visualization.

Why Synthesize?

Synthesis has become both more necessary and more difficult as evolutionary studies have become more diffuse and more specialized. Knowing more and more about less and less may mean that the relationships are lost and the grand pattern and great processes of life are overlooked.

G.G. Simpson (1944)

As articulated eloquently by Simpson more than 60-years ago, any attempt to elucidate the history and nature of life from a grand perspective must link clues from different scientific fields and disparate geographic, temporal, and physical scales. For exam-

ple, Darwin's (1859) fundamental insights came from the union of Malthusian population dynamics, careful observation of artificial selection and natural variation, Smith's economic principles, and geological evidence of the immense age of the Earth. Similarly, the 20th century transformation of biology and the unification of microevolution and macroevolution during the Modern Synthesis resulted from discourse and collaboration between researchers working on heritability, paleontology, development, speciation, Mendelian genetics, natural selection, and systematics (Fisher 1930; Haldane 1932; Wright 1932; Dobzhansky 1937; Mayr 1942; Simpson 1944; Mayr and Provine 1980). In either case, breakthroughs sprang from new linkages among scales of analysis ranging from the detailed perspective provided by single generations of flies or pigeons to the vast view of deep time provided by the fossil record.

[4]These authors led this project; other authors contributed equally and are in alphabetical order.

Synthesis continues to drive progress in evolutionary biology. For example, the current projects involved in NSF's "Assembling the Tree of Life" initiative require the aggregation of huge quantities of data and the combined expertise of systematists, geneticists, morphologists, theoreticians, and computer scientists, whereas the recent integration of evolutionary biology and developmental biology has led to the emergence of evo–devo as an entirely new field (Raff 2000). A synthetic approach also best addresses many important biological and social issues. Effective strategies to conserve biodiversity draw from evolution, ecology, biogeography, economics, and politics; epidemiology relies upon evolution, public health, and medicine; and climate change research links biogeography, climatology, biogeochemistry, landscape use, evolution, and ecology.

Evolutionary theory has also been synthesized productively with fields well outside biology (Derry 2009), including philosophy (Hull 1988; Mayr 2001), economics (Galor and Moav 2002), linguistics (Cavalli-Sforza 2000), and computer science (Mitchell and Taylor 1999; Lemmon and Milinkovitch 2002). Although each of the above syntheses represents an important conceptual advance in its own right, the crossover of evolutionary thought into the humanities and social sciences also provides valuable examples that help demonstrate evolution's increasing relevance to issues that students and the public find familiar and meaningful. Because syntheses that link evolution to other fields often focus on large scales and provide overarching answers to big questions, they tend to be highly accessible to the general public and can be invaluable in communicating scientific ideas.

As postdocs, informaticians, and educators affiliated with the National Evolutionary Synthesis Center (NESCent), the continuing ability of the synthetic approach to address complex questions at the largest scales and generate major advances in evolutionary biology inspires and motivates our work. At the same time, our experience has highlighted several ongoing technological and cultural barriers to pursuing this mode of inquiry. By calling attention to the importance of synthesis and suggesting ways to overcome hurdles to its progress, we hope to help promote a scientific environment that encourages researchers to link as big as they dare.

What Is Synthesis?

To foster synthesis, one must first recognize it. There is no single synthetic approach to science, and indeed, the opinion expressed herein grew out of a lively series of discussions on what is and is not synthetic. Although all syntheses "combine separate elements or components" (Ritchey 1991) to make a "coherent" (Ritchey 1991) or "complex" ("synthesis 6a" The Oxford English Dictionary. 2nd ed. 1989) whole, in a scientific context it is the extraction of otherwise unobtainable insight from a combination of disparate elements that best diagnoses synthesis. Novel insight can accrue from the combination of data, methods, results, or overarching concepts, with linkages between each of these types of elements creating four basic synthetic modes (Fig. 1). Because not all
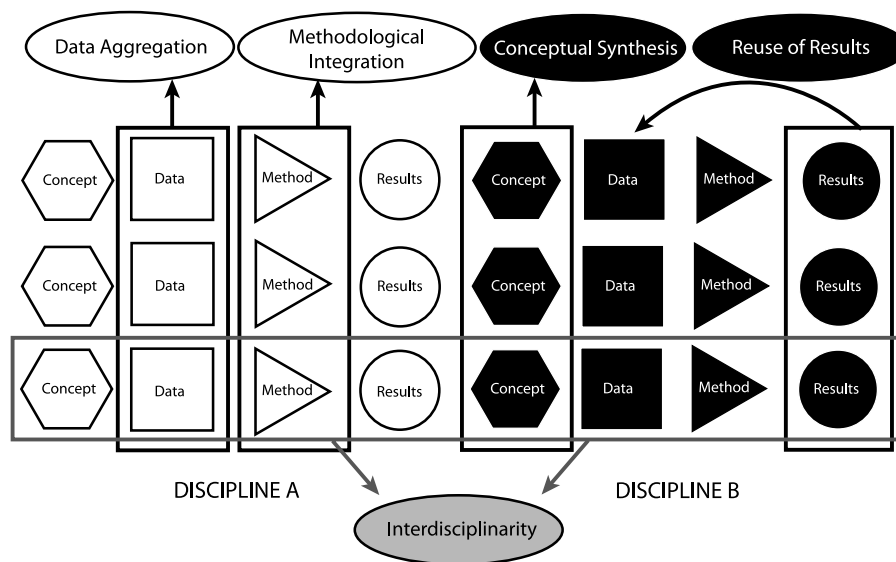

Figure 1. Schematic illustrating four modes of synthesis among elements from varied studies. Individual studies (horizontal sets) within either of two disciplines (white and black) begin with underlying concepts (hexagons), then collect data (squares), and perform analyses (triangles) to yield results (circles). Syntheses between like elements (vertical boxes) define the four basic modes of synthesis, with reuse of results being an unusual mode in that it converts one element (results) into another (data). Simple, but still effective syntheses may include only similar elements within single disciplines, whereas more complex syntheses may incorporate heterogeneous elements or span multiple disciplines (horizontal box).

syntheses cross disciplines (Sepkoski 1982, 2002; Bininda-Emonds et al. 2007) and because some actually erode disciplinary boundaries or create new fields of study like evo–devo (Raff 2000) and phylogenomics (Eisen and Fraser 2003), we recognize interdisciplinarity as a separate aspect that may enhance any of the four synthetic modes rather than as a necessary component.

The first mode of synthesis, data aggregation, reinterprets the raw data underlying prior investigations to answer questions at new and typically larger scales, as in supermatrix-based phylogenetic inference (de Queiroz and Gatesy 2007; Smith and Donoghue 2008). The second, reuse of results, differs from data aggregation by using extant results as data in a new context, such as in meta-analyses (Reed and Frankham 2001; Parmesan and Yohe 2003) or the construction of supertrees (Bininda-Emonds et al. 2007). The third mode, methodological integration, links two or more methods to create a new analytical pathway, as in phylogenomics (Eisen 1998b; Eisen and Fraser 2003) or the recent combination of ecological niche models with ancestral state reconstruction that permits inference of ancestral ecologies and species ranges (Evans et al. 2009). The fourth, conceptual synthesis, bridges the theories or paradigms of thoughts that underlie and motivate prior studies rather than extracting the elements of the studies themselves, as in the modern synthesis (Mayr and Provine 1980) or the growing links between evolution and epidemiology (Galvani 2003) or community ecology (Webb et al. 2002; Johnson and Stinchcombe 2007). As illustrated by the examples in Box 1, each of these four modes of synthesis has strong potential to provide previously unattainable insights across the full spectrum of biology. Combinations of these modes can be more powerful still (Darwin 1859; Simpson 1944).

Box 1: The Modes of Synthesis in Action

EXAMPLE 1: REVEALING THE HISTORY OF LIFE THROUGH THE MARINE PALEOBIOLOGY DATABASE

Primary mode of synthesis: Data aggregation

Over more than 20 years of research, J. John Sepkoski, Jr. compiled a database of first and last appearances of marine fossil taxa (Sepkoski 1982, 2002) that exemplifies the power of aggregate data. The more than 37,000 entries in the database cut across taxa, time, and geography to reveal emergent patterns over more than 500 million years of life that could not be extracted from the component data in isolation. Macroevolutionary patterns directly inferred from the Sepkoski database included the increase in biological diversity across the grand scale of the Phanerozoic (Sepkoski et al.

1981), the number and nature of mass extinctions (Raup and Sepkoski 1982), and the bias towards the evolution of ecological novelty in near-shore communities (Jablonski et al. 1983). The Sepkoski database catalyzed the rise of the entire subfield of synthetic paleobiology, which teases new patterns from old data to test current macroevolutionary questions, such as whether taxa originating in the wake of mass extinctions enjoy increased longevity (yes, Miller and Foote 2003), whether extinctions and originations within a time interval tends to occur in short bursts or longer sustained intervals (bursts, Foote 2005), and whether major geological changes correlate with and might explain major changes in biodiversity across the history of life (yes, Peters 2005). The information contained in the Sepkoski database's even more ambitious direct descendant, the Paleobiology database (Alroy et al. 2000), has supported 101 distinct publications since its inception in 1998. Clearly, these databases revolutionized paleobiology and continue to spark major analyses of large-scale evolutionary phenomena.

EXAMPLE 2: INFERRING THE IMPACT OF CLIMATE CHANGE ON NATURAL SYSTEMS

Primary mode of synthesis: Reuse of results

Although countering the effects of climate change on natural systems is currently recognized as one of the most pressing challenges facing humankind (IPCC 2007), not until two landmark meta-analytical approaches (Parmesan and Yohe 2003; Root et al. 2003) synthesized the results of hundreds of previous studies could the fingerprint of climate change on entire biota be distinguished beyond reasonable doubt. In isolation, few of the component studies could conclusively attribute shifts in species' distribution, abundance, phenology, or morphology to global climate change as opposed to local nonclimatic phenomena such as habitat loss (Hughes 2000; Parmesan and Yohe 2003; Thuiller 2007). By combining the larger body of results and applying meta-analytic statistics and probability models, the synthetic studies were able to determine with very high confidence that climate change must drive biotic change in a large proportion of cases. These two papers were hailed as being among the top 10 "breakthroughs" of 2003 (The News Editorial Staffs 2003), received more than 2300 total citations in just 5 years (Google Scholar, accessed April 26, 2009), form a large part of the foundation for current governmental policy on climate change (IPCC 2007; Rosenzweig et al. 2007), and illustrate that even conceptually and computationally simple syntheses within a single discipline possess remarkable power to tease out statistically significant effects that are inaccessible from the original elements in isolation.

EXAMPLE 3: PHYLOGENOMICS: USING EVOLUTIONARY TOOLS TO STUDY GENOMES
Primary mode of synthesis: Methodological integration

Phylogenomics integrates phylogenetic comparative methods with the analysis of genomic data and thus enables "scientists to design better experiments and generate new insights" (Eisen 1998b) concerning the origin and structure of the genome. For example, the prediction of gene function can be improved by incorporating the evolutionary history of the genes themselves and reconstructing their historical sequence and function using a phylogenetic framework (Eisen 1998a; Thornton 2004; Fisher 2008). The phylogenomic approach can identify gene duplications (orthologs vs. paralogs), infer evolutionary rate variation among taxa, and separate sequence convergences from shared origins, all of which confounded earlier methods that relied on raw sequence similarity (Eisen 1998b). By combining methods for assembling enormous datasets on the genomic architecture of species with methods for inferring the evolutionary history of those genomes, the new discipline of phylogenomics stands poised to reveal how the instruction book for building organisms changed and diversified across the branches of the tree of life (Eisen and Fraser 2003).

EXAMPLE 4: INTEGRATING EVOLUTION WITH MEDICINE TO IMPROVE HUMAN HEALTH
Primary modes of synthesis: Conceptual synthesis and methodological integration

By providing important new explanations for the origin and causation of human diseases and by promising to yield novel treatments, the emerging discipline of evolutionary medicine represents one of the most fruitful examples of conceptual synthesis between evolution and an entirely distinct field of study (Colwell 2004; Stearns and Koella 2007). The incontrovertible value of viewing a modern person as the result of past evolutionary pressures, tied intimately to population genetics and historical environments, can assist physicians in understanding puzzling medical conditions (Antolin 2009). For example, a rise in asthma rates has been traced to a hyperactive immune response in a subset of the population that may have originally evolved as an adaptation to parasites (Lau and Mtricardi 2006). This synthesis is also methodological, as the union of medicine with genomic information and evolutionary analysis has resulted in new techniques such as genome wide scans (The International HapMap Consortium 2005; Wellcome Trust Case Control Consortium 2007), new chemotherapy approaches (Nowak et al. 2004), the study of complex mental illness through synthesis of quantitative trait loci with public health and historical records (Susser et al. 1998), and the ability to mine large databases like the Fram-ingham Heart study (Levy and Brink 2005) to yield results far beyond the original focus of the study. In all of the above cases, application of evolutionary concepts and methods to a previously intractable medical issue yielded novel insight. In recognition of the explanatory power of this conceptual synthesis, medical schools are increasingly adding evolutionary theory to their curricula (Nesse and Schiffman 2003; Nesse et al. 2009) and the once strong barriers between these disciplines are eroding rapidly.

Towards a New Culture of Synthetic Science

Although the rise of reductionism and disciplinary thinking as the dominant paradigm in biology has stymied efforts to train students, fund research, publish results, and secure tenure (Rhoten and Parker 2004; Hurtado and Sharkness 2008) in more synthetic and interdisciplinary pursuits (Metzger and Zare 1999; Pfirman et al. 2005; McPeek 2006), a new cultural shift is leading synthetic science to receive more support than ever before (Collins 2002; Colwell 2004; McPeek 2006). As researchers fortunate enough to have received support from one of NSF's three biological synthesis centers, we have experienced firsthand the benefit of working in a culture that promotes synthesis. At the same time, we have encountered several cultural obstacles to conducting and communicating synthetic science outside NESCent. By calling attention to these difficulties at each stage of an academic career and drawing on our experience to suggest ways to overcome them, we hope to help foster increased support for synthesis within the field of evolutionary biology.

TRAINING

In order for synthesis to become mainstream, it must become easier for scientists to develop skills in aggregating and reusing information at every career stage, beginning with the student years (Carpenter et al. 2009). Established investigators can help students build skills in synthesis by engaging them in the assembly and analysis of datasets or offering workshops on database construction and integrative methods, such as the annual summer course in analytical paleobiology offered by the Paleobiology Database (Alroy et al. 2000). Graduate advisors can also support ambitious theses that span a diversity of modes of analysis or encourage students in distinct but complementary fields to pursue collaborative side-projects. No matter the form, the most effective training in synthesis will address information literacy (the ability to locate relevant information and assemble a knowledge base), statistical literacy (understanding how statistical manipulation affects data and inference), and data literacy (the skills required to manipulate and present data; Shields 2004). The erection of more interdisciplinary degree and research programs (Collins 2002; Frost et al.

2004) would also help train students in synthesis because such environments tend to encourage interaction between people with different skill sets and can counteract the prevailing push to specialize very quickly. Recent increases in funding for interdisciplinary undergraduate and graduate research experiences such as NSF's Research Experiences for Undergraduates and Integrative Graduate Education and Research Traineeship programs (Collins 2002), promote such training and will hopefully help more students engage in synthetic science.

JOB SEARCH AND TENURE EVALUATION

During the post-degree job search, synthetic researchers can face unique hurdles. The need to find a permanent home in a highly disciplinary academic environment makes following a purely synthetic research path risky for a young researcher. Job advertisements frequently seek highly specialized individuals skilled at analysis within a particular evolutionary field; rarely does a department search specifically for a synthetic or interdisciplinary scientist. In response, we have often found ourselves trying to repackage our research programs to fit narrow job descriptions, or emphasizing the familiar methodological and conceptual elements of our research rather than novel combinations that are frequently our primary thrusts. Solutions here are not easy, but as science moves to a more synthetic mindset and as more interdisciplinary units form, more job announcements will likely seek synthetic scientists.

At the level of entering faculty, more departments need to adopt published guidelines in developing and evaluating interdisciplinary or synthetic positions (National Academy of Sciences 2005; Pfirman et al. 2005; Pfirman et al. 2007). Above all else, such guidelines agree that review boards addressing tenure, publication, and funding decisions must be broad enough to provide a balanced review of the applicant's work. Otherwise, interdisciplinary faculty will continue to face a major "sweat inequity" (Frost et al. 2004), essentially needing to earn tenure in more than one academic department, while synthetic faculty may lack advocates during the critical tenure review process.

GRANT REVIEW

For better or worse, success in grantsmanship plays a major role in defining the success of a scientific career, and here synthetic scientists may encounter another hurdle. Synthetic research proposals tend to fare poorly in competition against disciplinary proposals because typical panel members feel ill-equipped to evaluate all the components of the research, because novel methods may lack the rigor of established methods, and because innovative research often carries a high risk of failure (Metzger and Zare 1999). Although a pair of recent government reports (National Institutes of Health 2004; National Research Council 2008) addressed these issues and encouraged funding agencies to develop programs for

synthetic research with appropriate peer review and an awareness of the inherently risky nature of these proposals, the funding barrier is a still a major obstacle. The creation of new granting opportunities like NSF's Opportunities for Promoting Understanding through Synthesis (OPUS) or additional synthesis centers such as the just-funded National Institute for Mathematical and Biological Synthesis (NIMBioS) or Germany's BiK-F (Biodiversity and Climate Research Center) will undoubtedly help (see also Carpenter et al. 2009), but a simpler solution may be to attempt to correct the funding bias in traditional grant programs by including more synthetically minded scientists on review panels and as program officers. The onus for that correction rests squarely within the community of synthesizers and upon our willingness to volunteer.

PUBLICATION

Publication issues present enormous barriers throughout a synthetic researcher's career, and it is not trivial to find an appropriate venue for publication of synthetic or interdisciplinary work. Many mainstream journals emphasize primary data collection over synthesis, and reviewers who are unfamiliar with all areas and methods spanned by a synthetic research question may in turn criticize the science as unoriginal or badly conceived (Lee 2006; Alberts et al. 2008). Synthetic datasets themselves can be extremely valuable and require a high level of skill to assemble, but are also difficult to publish. Even when synthetic papers or datasets are published, the large lists of authors that tend to appear on highly collaborative papers and the rigidity of citation indices can obscure measurement of a synthetic researcher's contribution and impact (Collins 2002). In this case, solutions are reasonably straightforward. Journal editors should encourage the publication of synthetic datasets and analyses and should make an effort to invite objective reviews from scientists who are also engaged in synthesis, authors should suggest such reviewers, and post-tenure, synthetic scientists should consider serving in editorial capacities.

LANGUAGE BARRIERS

Collaboration can catalyze synthesis powerfully by facilitating an exchange of information, methods, and insights that can address topics broader than any individual can tackle alone, but only if the collaborators can communicate effectively. The lack of a common frame of reference can present a serious barrier to communication and effective collaboration, particularly when collaborators cross disciplines. Overcoming that obstacle often requires the development of new frameworks based upon a common language or ontology. For example, the recent Phenoscape initiative (Mabee et al. 2007) built bridges between evolutionary morphologists and developmental biologists by linking the terminology of zebrafish development and genomics to the language of

morphological systematics with a formal ontology (Dahdul et al. in press). Because identical terms had very different meanings in the evolutionary and developmental communities, significant effort from several full time programmers and scientists was needed to reconcile the different semantic and conceptual histories. Despite the awkwardness of the initial communication, the resulting ontology-based framework allowed the creation of a massive database linking morphological shifts to potential developmental and genetic causes that will likely drive new synthetic discoveries for years. Notably, that resource was only achievable through the intentional breakdown of a cultural barrier to synthesis.

Overcoming Technological Impediments to Synthesis

Synthetic linkages drive progress in evolutionary biology, and the field's entry into the age of big data, fast computation, and instant communication theoretically sets the stage for bigger syntheses and greater insights than have ever before been possible. Examples like Phenoscape (Mabee et al. 2007) illustrate the potential of the expanding technological toolbox to handle an increasing flood of information (The Editorial Staff 2008) and facilitate synthesis, but technology can also hinder synthesis if a proliferation of methods, data standards, languages, and protocols hampers communication and interoperability. Without careful attention to how we record, store, link, and disseminate information, we run the risk of knowing so much, in so many different ways, that it becomes impossible to connect the various bits of knowledge practically.

New analysis of aggregate data underlies many synthetic studies, but the vast majority of data supporting previous studies are unavailable, often because the data are lost or preserved in inaccessible forms (notebooks, floppy disks). Although several repositories aim to preserve biological data in the long term, these repositories typically focus on a single type of data such as molecular sequences (GenBank), protein structures (PDB), or phylogenetic trees (TreeBASE) or on a single organism (e.g., ZFIN, FlyBase). Data that fall outside these few data types or model organisms is often lost, because individual researchers lack the technology, funding, and expertise to preserve this data on their own. Data aggregation can also be impeded when a culture of data sharing does not prevail. For example, in a survey of 1240 geneticists (Campbell et al. 2002), 47% had been denied at least one request for data or materials in the preceding 3 years, and 28% reported that they had been unable to confirm published research because of their inability to obtain data described in a publication. Awareness of these issues is slowly increasing, as evidenced by a recent special section on data sharing in *Nature*. One of that section's highlights is a report from the mouse research community (Schofield et al. 2009) that describes and advocates principles that call for all data to be shared at the time of publication. These principles include a call for journals and funding agencies to mandate archiving of data and materials, provide funds directed towards archiving, and enforce archiving mandates.

In a similar spirit, the Dryad data repository (National Evolutionary Synthesis Center and UNC Metadata Research Center 2009), a project spearheaded by NESCent, is working to alleviate the problem of data availability by providing an open-access home for ecological and evolutionary data that does not fit into more specialized repositories. Dryad actively works with a coalition of journals and scientific societies to make deposition of all data a normal part of the research workflow. As more journals require data deposition as part of the manuscript publication process, the opportunities for potential syntheses linking such data will increase substantially.

Powerful as aggregate data may be, a collection of data files is of little value without adequate metadata–information about relationships between files, the ways data are formatted within each file, and details about how the data were collected and manipulated. Data repositories and the organizations that house them can avoid this pitfall by standardizing metadata formats, developing expressive vocabularies, and establishing a consensus of what information is essential for a dataset to be reusable. To be most effective, these efforts need to be supported by the majority of researchers within the community, whether they are personally involved in synthesis or not.

Methods as encoded in software have become an integral part of the scientific process, and effective synthesis often requires interoperability and integration between such programs. Similar issues of availability and adequate metadata description affect the area of software tools (Veretnik et al. 2008; Wren 2008). Aside from concerns about one's ability to recreate published results, the loss of software tools can be a significant obstacle in synthetic work and may prevent the replication of analyses on new, broader datasets. Source-code repositories such as SourceForge, Bioinformatics.org, or Google Code can increase the life span and accessibility of code, but commonly accepted best practices for software preservation repositories do not yet exist. Development of such should take high priority.

Combining data from multiple sources, processing this data with a variety of software packages, or visualizing emergent patterns (see Box 2) is easiest when the data are in standardized formats and placed within a common frame of reference (Borer et al. 2009). Comprehensive data storage formats like NeXML (www.nexml.org), which standardizes the bewildering array of custom extensions to the popular but limited NEXUS format (Maddison et al. 1997), represent one good way to improve the ability to link data of different types within or among studies. Widespread use of ontologies and other common frames of references like the Tree-of-Life Web Project (Maddison and Schulz
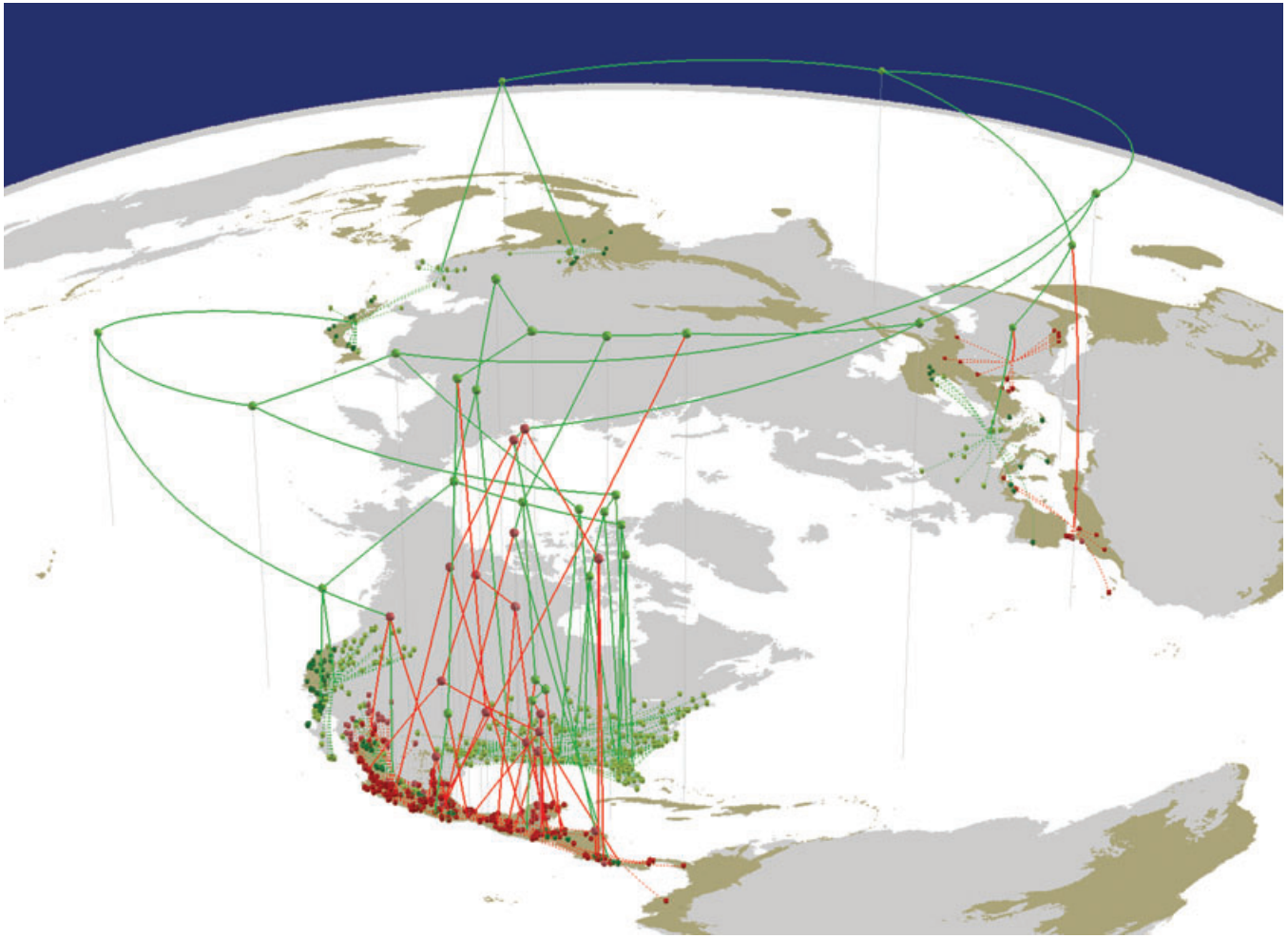
Figure 2. Visualizing phylogenetic diversity, biogeography, and conservation status. A geophylogeny (Kidd and Liu 2008; Kidd in review) reveals clades of hylid frogs that could be endangered or threatened if Conservation International's global biodiversity hot spots are not protected. Red branches indicate that >75% of descendant species live within hot spots, which appear in brown.

2007) or anthropology's Ethnographic Atlas (Gray 1998), will also vastly improve our ability to combine dissimilar data or to visualize patterns that emerge as a consequence of the linkage. One good example appears in Figure 2, which shows how the union of biogeographic, phylogenetic, and biopolitical information in a common geographic frame of reference reveals the concentration of potentially endangered species of frogs in certain clades.

The good news threaded through these examples is that the technological tools necessary to tear down barriers to synthesis exist or are being developed actively. However, the speed and efficiency with which we take advantage of these new capacities will largely be set by the willingness of the evolutionary biology community to embrace a culture in which sharing is normative, methods exist to be combined, and the potential longevity and utility of data exceeds the life span of the scientists that create it. Thus, informatic barriers to synthesis are also largely cultural. As such, the responsibility for fostering a future in which uncovering the "grand pattern and great processes of life" (Simpson 1944)

takes center stage rests with each scientist's willingness to participate in science as a collective, rather than individual enterprise.

Box 2: Visualization in The Service of Synthesis

As the amount of data available for synthesis continues to grow, so does the need for new methods to display and explore these increasingly complex data spaces (Frankel and Reid 2008). Successful data visualization promotes scientific discovery and communication by harnessing the power of human cognition to detect patterns within complex datasets.

The visualization of phylogenetic trees is integral to evolutionary biology (Hewitt 2001), but displaying the increasingly large trees at the cutting edge in ways that reveal both their large- and small-scale structure is very difficult. When the difficulties of projecting large multidimensional data into

limited space are overcome, visualizations of large phylogenies yield major new insights, such as the discovery that the extinction of nonavian dinosaurs did not alter the species diversification rate of present-day mammals (Bininda-Emonds et al. 2007) or that generation time predicts the rate of molecular evolution at a grand scale across the tree of life for plants (Smith and Donoghue 2008). Combined visualizations that synthesize phylogenies with other types of data, such as geography (Janies et al. 2007; Kidd in review), conservation priority (see Fig. 2), or morphology (Sidlauskas 2008) can be even more insightful.

Collaboration between scientists and specialists in the fields of informatics, technology, cognition, and graphic design can provide the tools needed to visualize complex synthetic datasets (Frankel and Reid 2008). New media, such as Virtual Earths, 3D graphics, animation, large-format video walls, and immersive virtual reality systems are increasingly being employed to display and interact with phylogenetic trees (Janies et al. 2007; Kidd in review). Related software enhancements maximize drawing clarity (Sanderson 2006) or judiciously distort images, as in Dendroscope's (Huson et al. 2007) use of fish-eye perspectives. Regardless of medium, input from cognitive science and careful graphic design (Tufte 2001) can ensure that visualizations exploit our innate human perceptive abilities to maximal effect (Ware 2004). For example, simple rearrangement of tips on a phylogeny can make it easier to read a tree correctly as a set of hierarchically nested groups rather than incorrectly as a linear progression from ancestral to more modern species (Baum et al. 2005) and help reinforce the key concepts of common ancestry and relationships among species in a classroom setting (Baum et al. 2005). Such advances in visualization suggest a future in which a lack of capacity to display complex, graphical datasets will not limit our ability to learn from them.

Conclusion

Synthesis offers unparalleled potential to address questions at the largest scales and to generate major scientific advances, which has led us to pursue synthetic science despite a suite of cultural and informatic obstacles. We predict that the accelerating ability to link data, concepts, methods, and results from a variety of fields will place synthesis at the heart of many future advances in evolutionary biology and in applied sciences to which it connects. Thus, we advocate an increased recognition of the key role of synthesis in evolutionary biology and continued effort to promote synthetic approaches across disciplines and at all levels of the academic hierarchy. As more and more obstacles are overcome successfully, synthesis will offer an even better ability to see further and deeper by allowing us to stand on the shoulders of more giants than has ever before been possible.

LITERATURE CITED

"synthesis 6a" The Oxford English Dictionary. 2nd ed. 1989. OED Online. Oxford Univ. Press, Oxford, UK.

Alberts, B., B. Hanson, and K. L. Kelner. 2008. Reviewing peer review. Science 321:15.

Alroy, J., C. Marshall, and A. Miller. 2000. The paleobiology database. Available at http://www.paleodb.org. Accessed April 19, 2009.

Antolin, M. F. 2009. Evolutionary biology of disease and Darwinian medicine. Pp. 281–298 in M. Ruse, and J. Travis, eds. Evolution: the first four billion years. The Belknap Press of Harvard Univ. Press, Cambridge, MA.

Baum, D., S. Smith, and S. Donovan. 2005. The tree-thinking challenge. Science 310:979–980.

Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature 446:507–512.

Borer, E. T., E. W. Seabloom, M. B. Jones, and M. Schildhauer. 2009. Some simple guidelines for effective data management. Bull. Ecol. Soc. Am. 90:205–214.

Campbell, E. G., B. R. Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N. A. Holtzman, and D. Blumenthal. 2002. Data withholding in academic genetics: evidence from a national survey. J. Am. Med. Assoc. 287:473–480.

Carpenter, S. R., E. V. Armbrust, P. W. Arzberger, F. S. Chapin, J. J. Elser, E. J. Hackett, A. R. Ives, P. M. Kareiva, M. A. Leibold, P. Lundberg, et al. 2009. Accelerate synthesis in ecology and environmental sciences. BioScience 59:699–701.

Cavalli-Sforza, L. L. 2000. Genes, peoples, and languages. Univ. of California Press, CA.

Collins, J. 2002. May you live in interesting times: using multidisciplinary and interdisciplinary programs to cope with change in the life sciences. BioScience 52:75–83.

Colwell, R. 2004. Biocomplexity and a new public health domain. EcoHealth 1:6–7.

Dahdul, W. M., J. G. Lundberg, P. E. Midford, J. P. Balhoff, H. Lapp, T. J. Vision, M. A. Haendel, M. Westerfield, and P. M. Mabee. in press. The teleost anatomy ontology: anatomical representation for the genomics age. Syst. Biol. In press.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

de Queiroz, A., and J. Gatesy. 2007. The supermatrix approach to systematics. Trends Ecol. Evol. 22:34–41.

Derry, J. F. 2009. Darwin in disguise. Trends Ecol. Evol. 24:73–79.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia Univ. Press, New York, NY.

Eisen, J. A. 1998a. A phylogenomic study of the Muts family of proteins. Nucleic Acids Res. 26:4291–4300.

———. 1998b. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res. 8:163–167.

Eisen, J. A., and C. M. Fraser. 2003. Phylogenomics: intersection of evolution and genomics. Science 300:1706–1707.

Evans, M. E. K., S. A. Smith, R. S. Flynn, and M. J. Donoghue. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses (Oenothera, sections Anogra and Kleinia). Am. Nat. 173:225–240.

Fisher, K. M. 2008. Bayesian reconstruction of ancestral expression of the LEA gene families reveals propagule-derived desiccation tolerance in resurrection plants. Am. J. Bot. 95:506–515.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon Press, Oxford, UK.

Foote, M. 2005. Pulsed origination and extinction in the marine realm. Paleobiology 31:6–20.

Frankel, F., and R. Reid. 2008. Big data: distilling meaning from data. Nature 455:30–30.

Frost, S. H., P. M. Jean, D. Teodorescu, and A. B. Brown. 2004. Research at the crossroads: how intellectual initiatives across disciplines evolve. Rev. High. Educ. 27:461–479.

Galor, O., and O. Moav. 2002. Natural selection and the origin of economic growth. Q. J. Econ. 117:1133–1191.

Galvani, A. 2003. Epidemiology meets evolutionary ecology. Trends Ecol. Evol. 18:132–139.

Gray, P. J. 1998. Ethnographic atlas codebook. World Cult. 10:86–136.

Haldane, J. B. S. 1932. The causes of evolution. Longman, Green and Co., London, UK.

Hewitt, G. M. 2001. Speciation, hybrid zones and phylogeography – or seeing genes in space and time. Mol. Ecol. 3:537–549.

Hughes, L. 2000. Biological consequences of global warming: is the signal already apparent? Trends Ecol. Evol. 15:56–61.

Hull, D. L. 1988. Science as a process: an evolutionary account of the social and conceptual development of science. Univ. of Chicago Press, Chicago, IL.

Hurtado, S., and J. Sharkness. 2008. Scholarship is changing, and so must tenure review. Academe 94.

Huson, D., D. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460.

IPCC, ed. 2007. Climate change 2007: impacts, adaptation and vulnerability. Contribution of working group II to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press, Cambridge, U.K.

Jablonski, D., J. J. Sepkoski Jr., D. J. Bottjer, and P. M. Sheehan. 1983. Onshore-offshore patterns in the evolution of phanerozoic shelf communities. Science 222:1123–1125.

Janies, D., A. W. Hill, R. Guralnick, F. Habib, E. Waltari, and W. C. Wheeler. 2007. Genomic analysis and geographic visualization of the spread of avian influenza (H5N1). Syst. Biol. 56:321–329.

Johnson, M. T. J., and J. R. Stinchcombe. 2007. An emerging synthesis between community ecology and evolutionary biology. Trends Ecol. Evol. 22:250–257.

Kidd, D. M. in press. Geophylogenies and the map of life. Syst. Biol. *In press*.

Kidd, D. M., and X. Liu. 2008. GEOPHYLOBUILDER 1.0: an ArcGIS extension for creating 'geophylogenies'. Mol. Ecol. Resour. 8:88–91.

Lau, K., and P. M. Mtricardi. 2006. Worms, asthma and the hygiene hypothesis. Lancet 367:1556–1558.

Lee, C. 2006. Perspective: peer review of interdisciplinary scientific papers. Nature: web forum. Available at http://www.nature.com/nature/peerreview/debate/nature05034.html; doi:10.1038/nature05034. Accessed April 9, 2009.

Lemmon, A. R., and M. C. Milinkovitch. 2002. The metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. Proc. Natl. Acad. Sci. USA 99:10516–10521.

Levy, D., and S. Brink. 2005. A change of heart: how the people of Framingham, Massachusetts, helped unravel the mysteries of cardiovascular disease. Knopf, New York, NY.

Mabee, P. M., M. Ashburner, Q. Cronk, G. V. Gkoutos, M. Haendel, E. Segerdell, C. Mungall, and M. Westerfield. 2007. Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol. Evol. 22:345–350.

Maddison, D. R., and K.-S. Schulz. 2007. The tree of life web project. Available at http://tolweb.org. Accessed April 9, 2009.

Maddison, D. R., D. L. Swofford, and W. P. Maddison. 1997. NEXUS: an extensible file format for systematic information. Syst. Biol. 46:590–621.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York, NY.

———. 2001. The philosophical foundations of Darwinism. Proc. Am. Philos. Soc. 145:488–495.

Mayr, E., and W. B. Provine, eds. 1980. The evolutionary synthesis: perspectives on the unification of biology. Harvard Univ. Press, Cambridge, MA.

McPeek, M. A. 2006. What hypotheses are you willing to entertain? Am. Nat. 168:S1–S3.

Metzger, N., and R. N. Zare. 1999. Interdisciplinary research: from belief to reality. Science 642–643.

Miller, A. I., and M. Foote. 2003. Increased longevities of post-paleozoic marine genera after mass extinctions. Science 302:1030–1032.

Mitchell, M., and C. E. Taylor. 1999. Evolutionary computation: an overview. Annu. Rev. Ecol. Syst. 30:593–616.

National Academy of Sciences. 2005. Facilitating interdisciplinary research. The National Academic Press, Washington, DC.

National Evolutionary Synthesis Center and UNC Metadata Research Center. 2009. Dryad. Available at http://www.datadryad.org. Accessed September 24, 2009. National Evolutionary Synthesis Center, Durham, NC.

National Institutes of Health. 2004. NIH roadmap for medical research. Available at http://nihroadmap.nih.gov. Accessed May 20, 2009.

National Research Council. 2008. The role of theory in advancing 21st-century biology: catalyzing transformative research. The National Academies Press, Washington, D.C.

Nesse, R. M., and J. D. Schiffman. 2003. Evolutionary biology in the medical school curriculum. BioScience 53:585–587.

Nowak, M., F. Michor, and Y. Iwasa. 2004. Evolutionary dynamics of tumor suppressor gene inactivation. Proc. Natl. Acad. Sci. USA 101:10635–10638.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42.

Peters, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. Proc. Natl. Acad. Sci. USA 102:12326–12331.

Pfirman, S. L., J. P. Collins, S. Lowes, and A. F. Michaels. 2005. Collaborative efforts: promoting interdisciplinary scholars. Chron. High. Educ. 51:B15–B16.

Pfirman, S., P. Martin, L. Berry, M. Fletcher, M. Hempel, R. Southard, D. Hornback, and B. Morehouse. 2007. Interdisciplinary hiring, tenure and promotion: guidance for individuals and institutions. Available at http://www.ncseonline.org/CEDD/cms.cfm?id=2042. Accessed April

24, 2009. Council of Environmental Deans and Directors Interdisciplinary Tenure and Career Development Committee.

Raff, R. A. 2000. Evo-Devo: the evolution of a new discipline. Nat. Rev. Genet. 1:74–79.

Raup, D. M., and J. J. Sepkoski, Jr. 1982. Mass extinctions in the marine fossil record. Science 215:1501–1503.

Reed, D. H., and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095–1103.

Rhoten, D., and A. Parker. 2004. Risks and rewards of an interdisciplinary research path. Science 306:2046.

Ritchey, T. 1991. Analysis and synthesis: on scientific method – based on a study by Bernhard Riemann. Sys. Res. 8:22–41.

Root, T. L., J. T. Price, K. R. Hall, S. H. Schneider, C. Rosenzweig, and J. A. Pounds. 2003. Fingerprints of global warming on wild animals and plants. Nature 421:57–60.

Rosenzweig, C., G. Casassa, D. J. Karoly, A. Imeson, C. Liu, A. Menzel, S. Rawlins, T. L. Root, B. Seguin, and P. Tryjanowski. 2007. Assessment of observed changes and responses in natural and managed systems. Pp. 79–131 in M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, eds. Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press, Cambridge, MA.

Sanderson, M. J. 2006. Paloverde: an OpenGL 3D phylogeny browser. Bioinformatics 22:1004–1006.

Schofield, P. N., T. Bubela, T. Weaver, L. Portilla, S. D. Brown, J. M. Hancock, D. Einhorn, G. Tocchini-Valentini, M. Hrabe de Angelis, and N. Rosenthal. 2009. Post-publication sharing of data and tools. Nature 461:171–173.

Sepkoski, J. J., Jr. 1982. A compendium of fossil marine families. Milwaukee Public Museum Contrib. Biol. Geol. 51.

———. 2002. A compendium of fossil marine animal genera. Paleontological Research Institution, Ithaca, NY.

Sepkoski, J. J., Jr., R. K. Bambach, D. M. Raup, and J. W. Valentine. 1981. Phanerozoic marine diversity and the fossil record. Nature 293:435–437.

Shields, M. 2004. Information literacy, statistical literacy, data literacy. IASSIST Q. 28:6–11.

Sidlauskas, B. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution 62:3135–3156.

Simpson, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York, NY.

Smith, S. A., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Stearns, S. C., and J. C. Koella. 2007. Evolution in health and disease. Oxford Univ. Press, Oxford, U.K.

Susser, E., H. W. Hoek, and A. Brown. 1998. Neurodevelopmental disorders after prenatal famine: the story of the Dutch Famine Study. Am. J. Epidemiol. 147:213–216.

The Editorial Staff. 2008. Community cleverness required. Nature 455:1–1.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437:1299.

The News Editorial Staffs. 2003. Breakthrough of the year: the runners-up. Science 302:2039–2045.

Thornton, J. W. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. 5:366–375.

Thuiller, W. 2007. Biodiversity: climate change and the ecologist. Nature 448:550–552.

Tufte, E. R. 2001. The visual display of quantitative information (2nd ed.). Graphics Press, Cheshire, CT.

Veretnik, S., J. L. Fink, and P. E. Bourne. 2008. Computational biology resources lack persistence and usability. PLoS Comp. Biol. 4:e1000136.

Ware, C. 2004. Information visualization: perception for design. Morgan Kaufmann, San Francisco, CA.

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33:475–505.

Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common disease and 3000 shared controls. Nature 447:661–678.

Wren, J. D. 2008. URL decay in MEDLINE—a 4-year follow-up study. Bioinformatics 24:1381–1385.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. Sixth Int. Congr. Genet. 1:356–366.

Associate Editor: M. Rausher