Journal of Vertebrate Paleontology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ujvp20

From card catalogs to computers: databases in vertebrate paleontology

Mark D. Uhen [a], Anthony D. Barnosky [b], Brian Bills [c], Jessica Blois [d], Matthew T. Carrano [e], Marc A. Carrasco [b], Gregory M. Erickson [f], Jussi T. Eronen [g h], Mikael Fortelius [g], Russell W. Graham [i], Eric C. Grimm [j], Maureen A. O'Leary [k], Austin Mast [f l], William H. Piel [m], P. David Polly [n] & Laura K. Säilä [g]

[a] Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, Virginia, 22030, U.S.A.

[b] Department of Integrative Biology and Museum of Paleontology, University of California, Berkeley, California, 94720, U.S.A.

[c] Center for Environmental Informatics, Pennsylvania State University, University Park, Pennsylvania, 16802, U.S.A.

[d] Center for Climate Research, University of Wisconsin, Madison, Wisconsin, 53706, U.S.A.

[e] Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, D.C., 20013, U.S.A.

[f] Department of Biological Science, The Florida State University, Tallahassee, Florida, 32306, U.S.A.

[g] Department of Geosciences and Geography, HY-00014, University of Helsinki, Finland

[h] Senckenberg Research Institute und Nature Museum, Biodiversity and Climate Research Centre LOEWE BiK-F, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany

[i] Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania, 16802, U.S.A.

[j] Research and Collections Center, Illinois State Museum, Springfield, Illinois, 62703, U.S.A.

[k] Department of Anatomical Sciences, School of Medicine, Stony Brook University, Stony Brook, New York, 11794, U.S.A.

[l] Robert K. Godfrey Herbarium, Florida State University, Tallahassee, Florida, 32306, U.S.A.

[m] Yale-NUS College, Singapore, 138614, Singapore

[n] Department of Geological Sciences, Indiana University, Bloomington, Indiana, 47405, U.S.A.

Published online: 08 Jan 2013.

To cite this article: Mark D. Uhen, Anthony D. Barnosky, Brian Bills, Jessica Blois, Matthew T. Carrano, Marc A. Carrasco, Gregory M. Erickson, Jussi T. Eronen, Mikael Fortelius, Russell W. Graham, Eric C. Grimm, Maureen A. O'Leary, Austin Mast, William H. Piel, P. David Polly & Laura K. Säilä (2013) From card catalogs to computers: databases in vertebrate paleontology, Journal of Vertebrate Paleontology, 33:1, 13-28, DOI: 10.1080/02724634.2012.716114

To link to this article: http://dx.doi.org/10.1080/02724634.2012.716114

Journal of Vertebrate Paleontology 33(1):13–28, January 2013

INVITED ARTICLE

FROM CARD CATALOGS TO COMPUTERS: DATABASES IN VERTEBRATE PALEONTOLOGY

MARK D. UHEN,*,[1] ANTHONY D. BARNOSKY,[2] BRIAN BILLS,[3] JESSICA BLOIS,[4]
MATTHEW T. CARRANO,[5] MARC A. CARRASCO,[2] GREGORY M. ERICKSON,[6] JUSSI T. ERONEN,[7,8]
MIKAEL FORTELIUS,[7] RUSSELL W. GRAHAM,[9] ERIC C. GRIMM,[10] MAUREEN A. O'LEARY,[11] AUSTIN MAST,[6,12]
WILLIAM H. PIEL,[13] P. DAVID POLLY,[14] and LAURA K. SÄILÄ[7]
[1]Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, Virginia 22030, U.S.A., muhen@gmu.edu;
[2]Department of Integrative Biology and Museum of Paleontology, University of California, Berkeley, California 94720, U.S.A., barnosky@berkeley.edu; carrasco@berkeley.edu;
[3]Center for Environmental Informatics, Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A., bbills@essc.psu.edu;
[4]Center for Climate Research, University of Wisconsin, Madison, Wisconsin 53706, U.S.A., blois@wisc.edu;
[5]Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20013, U.S.A., carranom@si.edu;
[6]Department of Biological Science, The Florida State University, Tallahassee, Florida 32306, U.S.A., gerickson@bio.fsu.edu; amast@bio.fsu.edu;
[7]Department of Geosciences and Geography, HY-00014, University of Helsinki, Finland, mikael.fortelius@helsinki.fi; laura.saila@helsinki.fi;
[8]Senckenberg Research Institute und Nature Museum, Biodiversity and Climate Research Centre LOEWE BiK-F, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany, jussi.t.eronen@helsinki.fi;
[9]Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A., rgraham@ems.psu.edu;
[10]Research and Collections Center, Illinois State Museum, Springfield, Illinois 62703, U.S.A., grimm@museum.state.il.us;
[11]Department of Anatomical Sciences, School of Medicine, Stony Brook University, Stony Brook, New York 11794, U.S.A., moleary@notes.cc.sunysb.edu;
[12]Robert K. Godfrey Herbarium, Florida State University, Tallahassee, Florida 32306, U.S.A., amast@bio.fsu.edu;
[13]Yale-NUS College, Singapore 138614, Singapore, william.piel@yale-nus.edu.sg;
[14]Department of Geological Sciences, Indiana University, Bloomington, Indiana 47405, U.S.A., pdpolly@indiana.edu

ABSTRACT—Data, whether images, measurements, counts, occurrences, or character codings, are a cornerstone of vertebrate paleontology. Every published paper, master's thesis, and doctoral dissertation relies on these data to document patterns and processes in evolution, ecology, taphonomy, geography, geologic time, and functional morphology, to name just a few. In turn, the vertebrate paleontology community relies on published data in order to reproduce and verify others' work, as well as to expand upon published analyses in new ways without having to reconstitute data sets that have been used by earlier authors and to accurately preserve data for future generations of researchers. Here, we review several databases that are of interest to vertebrate paleontologists and strongly advocate for more deposition of basic research data in publicly accessible databases by vertebrate paleontologists.

INTRODUCTION

The field of vertebrate paleontology is celebrating a birthday of sorts. Certainly it is difficult to define the exact starting point for any scientific endeavor, but a key moment in our history was the publication of Cuvier's Recherches sur les Ossemens Fossiles de Quadrupèdes in 1812 (Cuvier, 1812). Since then, vertebrate paleontologists have accumulated vast quantities of both geological and biological data at a dramatically accelerating rate during the last decades of the 20th century (see Fig. 1). These data form the foundations for a rich scientific literature that remains in vigorous use.

Yet, the basic data behind the paleontological literature are often not readily available, even in the Information Age. Many quantitative studies published in the Journal of Vertebrate Paleontology (and other journals) do not include the supporting data for important graphs or statistical summaries. Even if the data are initially available, eventual data loss is common. 'Data available upon request from the authors' often disappear as workers leave academia, retire, or die, whereas their institutions frequently lack archiving policies altogether. Web pages for institutions, individuals, and even supplementary data posted by journals are notoriously ephemeral. In fact, a recent study of the biomedical literature found that between 8 and 29 percent of supplementary information eventually becomes unavailable (Anderson et al., 2006). This review focuses on several persistent, publicly accessible data repositories of interest to vertebrate paleontologists, with the goal of encouraging such scientists to make use of them for long-term data preservation.

The Nature of Paleontological Databases

Before reviewing several data repositories, we need to clarify our use of the word 'databases' and summarize their history in paleontology. Databases are collections of data and the

*Corresponding author.

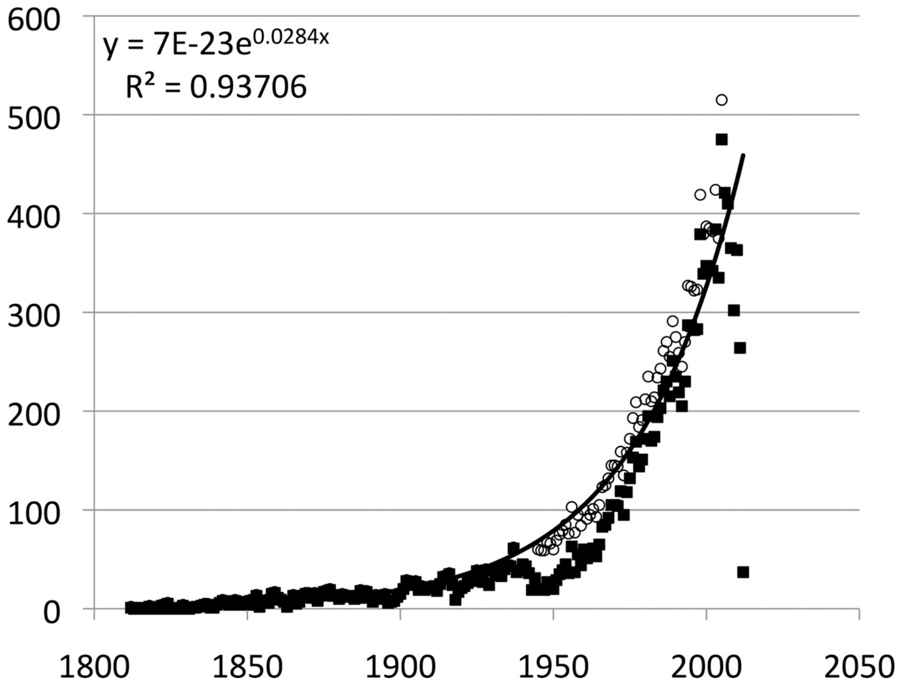

FIGURE 1. Growth in vertebrate paleontology publication output. Data plotted in black squares represents the number of references in the Paleobiology Database (PaleoDB) from every year from 1812 to 2012. Notice that there is a significant downward offset during World War II (1939–1945) and that the peak in number of references occurs in 2005. The drop in the number of references from 2005 to 2012 represents the lag time from publication to entry into the PaleoDB. Although this is not a random sample of the literature, it is almost certainly random with respect to the year of publication. The open circles represent the same data, excluding points from 1939 to 1945, and has the post-1945 data points increased by 40, to adjust for the offset noted during World War II, and excludes the 2005–2012 lag. If the World War II data points and those from 2006 to 2012 are excluded, the publication output in vertebrate paleontology is well modeled by the exponential growth curve shown.

organizational structure in which those data are stored and organized along with the means to usefully retrieve subsets of information (after MacLeod and Guralnick, 2000). There are two basic types of paleontological databases: collections databases and research databases.

Collections databases are structured around lists of specimens and the associated information cataloged in museum collections. They come in three basic varieties. First, many institutions have created custom databases built on commercially available database platforms. Second, many institutions use broadly available freeware such as Specify, which was developed with funds from the U.S. National Science Foundation (NSF) at the University of Kansas (specifysoftware.org). Third, many institutions use commercially available products such as KE EMu (www.kesoftware.com) or PastPerfect (www.museumsoftware.com). No matter what variety of collections database is deployed, all of these classes are customizable to be as complex or as simple as is necessary.

A new resource for paleontology collections is the Advancing Digitization of Biological Collections (ADBC) program, which is a collaboration between the NSF's Directorates for Biological Sciences and Geosciences. The program supports two previously unfunded digitization activities: digitization efforts of Thematic Collections Networks (TCNs) centered on a research question and a national hub to facilitate the digitization activities and online deployment of the data. ADBC's national hub is iDigBio (www.idigbio.org; the Integrated Digitized Biocollections project), whose mission is to develop a national infrastructure to support the vision of ADBC by overseeing implementation of standards and best practices for digitization; to build and deploy a customized cloud computing environment for collections; to recruit and train personnel, including underserved groups; to engage the research community, collections community, citizen scientists, and the public through education and outreach activities; and to plan for long-term sustainability of the national digitization effort.

Although collections databases are an essential and valued part of the management of vertebrate paleontology resources and contain vast quantities of data available for the clever data miner, the main focus of our review is on research databases.

Research databases are usually created to address a particular set of scientific questions, and therefore collect data tailored to a subset of paleontological inquiry. These data may be extracted from published literature (e.g., taxonomic lists and associated information) or derive from unpublished primary information. Often, unpublished data are deposited in a research database as it is assembled for study and eventual publication. These databases can be archival (i.e., an objective archiving of available information) or research-filtered (data pass through an analytical filter prior to incorporation in a given field).

Research databases often have the advantage of high-quality information that has passed through a peer-review filter; their disadvantage is that they typically exclude information about the vast number of unpublished specimens. In contrast, collections databases offer a more complete listing of all the specimens that have been placed in museums; their downside is that they have less uniformity of information quality than research databases, because relatively few of the specimens have passed through the peer-review filter. Thus, collections databases and research databases offer complementary information, but have remained difficult to effectively merge.

Today, we tend to think of databases as electronic entities, but they have existed for some time as printed compilations of data in books, tables, ledgers, photographs, and card files. Card files are probably the closest analogue to modern digital databases because they share the same dynamic quality. Books and tables in journals become static the moment they are printed, but card files can be continually added to and updated, like digital databases. It is also worth mentioning that virtually anything that can be accomplished with a digital database could, in theory, also be done with an analog database, albeit with significantly greater difficulty (Uhen, 2000).

History

Benton (1999) gives an excellent overview of the early history of research databases and their use in paleontology, and we discuss only the highlights most relevant to the development of paleontological (especially vertebrate) research databases. Benton's review begins with a mention of a book by Phillips (1860) who was the first to publish a diversity curve, which was based in turn on a compilation of data on British fossils by Morris (1854). (Both of these volumes are available as instant downloads from Google Books, which underscores how accessible information can be through the Internet.)

The growth of paleontology through the 19th and early 20th centuries created a demand for authoritative compilations of data about fossil taxa. Miller's (1889) North American Geology and Paleontology and Shimmer and Shrock's (1944) Index Fossils of North America were among the first and last examples, respectively, of pre-digital efforts to catalog past life comprehensively across time and taxa. In subsequent decades, it became necessary to break such compilations into more manageable taxonomic subsets, with Moore's Treatise on Invertebrate Paleontology, the first edition of which was completed in the 1940s, being one of the most monumental examples. The Bibliography of Fossil Vertebrates, an indexed compilation of publications on vertebrate paleontology, first appeared in this era (Hay, 1902). The Modern Synthesis invigorated the compilation of paleontological data for analytical purposes related to the study of evolution and Earth history. Of particular note is Newell's (1952) paper on invertebrate evolution in which he compiled massive amounts of data from the then-published record to document changes in the diversity of groups ranging from foraminiferans to echinoderms.

The beginning of the modern era of computerized paleontological databases can be marked by the publication of several works in the late 1960s, including The Fossil Record (Harland et al., 1967), which is the first published database specifically designed for the study of the history of life (Benton, 1999). Many others expanded on this work in the 1970s, 1980s, 1990s, including Sepkoski's compendia of marine fossils, first at the family level (Sepkoski, 1982), and then at the genus level (Sepkoski, 2002), which facilitated numerous studies in the marine realm. Benton's own The Fossil Record 2 (1993; www.fossilrecord.net/fossilrecord/index.html) added terrestrial animals.

With the advent of relational database software, collection management databases became increasingly common in the 1970s and 1980s, first running on mainframe computers, and then on workstations and personal computers. Graham et al. (1987) published a compendium of late Quaternary fossil mammal sites from the Great Plains and prairies of the United States and Canada. This was the first systematically organized database to provide faunal lists along with extensive metadata that could be used to evaluate the data themselves. This print publication was a precursor to the FAUNMAP database (discussed below).

The next leap forward happened with the advent of the Internet and the World Wide Web (WWW), which was used to distribute paleontological information as early as 1991. By this time, much paleontological data had been translated to digital form, either as museum collections databases or as research databases such as Sepkoski's and Benton's. Prior to the 1990s and the spread of easy-to-use graphical-interface Web browsers, however, these databases were still distributed on paper or another form of physical media (in those days, backup tapes or the now obsolete floppy disks) for computers. Although distribution via physical computer media saved a great deal of time and eliminated one layer of possible transcription errors, it was just as static as publication on paper. The University of California Museum of Paleontology (UCMP) was the first to provide interactive, integrated data via the WWW in 1993 when graduate students R. P. Guralnick and P. D. Polly interfaced the UCMP's existing collection database with interpretive material, with multiple portals aimed at scientists and the general public (Schwarzer, 2012). The posting of databases on the Internet, which became preferred practice beginning in the late 1990s and early 2000s (e.g., the Paleobiology Database, Alroy's first iteration of the North American Mammal Database, FAUNMAP, MIOMAP, NOW, etc.), initially allowed community-wide access to information and provided a means for database developers to continually update data as more became available and as users identified mistakes. The first online journal of paleontology, Palaeontologia Electronica, was launched in 1996 and contains early examples of integrated data and scientific publication.

A more recent innovation in paleontological databases is the advent of modified wiki approaches to data addition (e.g., Paleobiology Database). In the past, most databases were constructed by individuals (such as Sepkoski or Benton) or small consortia of people who may (or may not) have been experts on the data. Presently it is becoming more common for databases to open data contribution to anyone with relevant expertise. This community approach, facilitated by the open access of the WWW and the dramatic reduction in the cost and increase in power of consumer-grade computers, has the great advantage of distributing the work of data collection and entry among a large group of knowledgeable individuals.

Data Portals

A growing trend in bioinformatics is to build data portals that link freestanding databases through a common search engine, thereby providing the ability to choose data fields from separate databases in order to assemble sets of information that otherwise cannot be easily obtained. The greatest advances in this regard so far have been in linking individual, freestanding museum collections databases, such that a single search retrieves records from all of the participating collections. A key example for vertebrates is VertNet (vertnet.org/index.php), which as of this writing links 72 different museum collections through a portal that is mirrored at five different servers. This approach has also proven fruitful in paleontology, as exemplified by the Paleontology Portal (www.paleoportal.org/portal/), which links collections data from 14 different museums.

In principle, this approach would also prove useful in linking several existing freestanding research databases, each of which has particular strengths, but none of which completely overlaps the others in terms of fields and data. Linking research databases with collections databases also would represent a major breakthrough in facilitating local-, regional-, and global-scale analyses—for example, answering questions that depend on integrating morphological and diversity information, or on abundance data (which can often be extracted from collections databases, but are seldom contained in research databases).

In practice, paleontological data portals already are beginning to develop, as exemplified by NEOMAP (www.ucmp.berkeley.edu/neomap/), which links FAUNMAP and MIOMAP. Whereas the latter both remain freestanding databases as described above, NEOMAP provides a way of easily and uniformly extracting and analyzing point-occurrence data for all published late Oligocene through Holocene mammals in the United States, and for many Quaternary localities in Canada, using the same tools described for MIOMAP. This joint search of separate data sets facilitates analyses that were previously much more time consuming, if they were even practical, prior to the ability to search them through a common portal (Carrasco et al., 2009; Barnosky et al., 2011a, 2011b; Stegner and Holmes, in press). Funded by the NSF Sedimentary Geology and Paleobiology and Ecology Programs and served through the UCMP, NEOMAP was launched online in 2010.

NEOMAP is an example of a portal that was built to facilitate specific analyses, in that case to detect how anthropogenic impacts are affecting ecosystem dynamics on Earth through analysis of species-area relationships through time. Larger-scale data portals for linking paleontology research databases and collections databases also are underway. For instance, the VertNet project is linking with FAUNMAP and MIOMAP to facilitate analyses that require simultaneous treatment of both paleontological and neontological vertebrate data. At the same time, the Paleontology Portal is experimenting with linking museum collections databases with research databases; it now links MorphoBank with 14 collections databases, and in the past, has also linked with MIOMAP. As such linkages become more common, the scope of research problems that can be addressed should continue to increase dramatically.

Paleoinformatics

In 2000, MacLeod et al. coined a new term, paleoinformatics, for issues relating to the electronic creation, management, and retrieval of paleontological data, formed in parallel to the term bioinformatics (MacLeod et al., 2000; MacLeod and Guralnick, 2000). The term has recently started to appear in the scientific literature (at least three hits as a key word in GeoRef) (Dolven and Skjerpen, 2011), as the recognition of the value of paleontological data has begun to grow, especially in the areas of understanding biodiversity trends and maintenance (Fortelius et al., 2002; Alroy, 2008; Barnosky et al., 2011b; Johnson et al., 2011). In many ways, like the field of paleontology itself in relation to biology and geology, paleoinformatics combines many of the aspects of both bioinformatics and geoinformatics. Data structures regarding taxonomy, morphology, and ecology are virtually identical to those from bioinformatics, whereas data structures regarding geologic time, rock units, and sedimentology are similar to those in geoinformatics.

Copyright Concerns

Constructing a database of previously published information may present legal challenges, and additional legal issues regarding use of the data contained therein may also need to be considered, even if the data have not been previously published. Dolven and Skjerpen (2011) expressed concern that data deposited in a database that had been gleaned from copyrighted publications created potential copyright issues. This concern seems unwarranted, because (at least under U.S. copyright law) facts or data cannot be copyrighted even if the works derived from such facts are (Brown and Denicola, 1998).

On the other side of this issue, compilations of previously published data do enjoy copyright protection from wholesale copying and reproduction without permission, provided that the compilation involves the collection of preexisting data; those data are selected, coordinated, and arranged; and an original work of authorship is created by virtue of the selection, coordination, or arrangement (Brown and Denicola, 1998). Thus, even if data do not enjoy copyright protection, the unique attributes and presentation of data that are provided by databases do enjoy copyright protection.

The situation is much more complicated when it comes to images that are included in a database. Under U.S. copyright law, the simple act of taking a photograph or producing a graph or image of any kind is considered an original work of authorship whose copyright is held by the person who created it (Brown and Denicola, 1998). Thus, any inclusion of images in a publicly accessible database can create copyright concerns that can and should be addressed. Most print journals require permission to use figures on a case-by-case basis. Some online journals use Creative Commons licenses to govern the use of visual material. Most of the databases discussed here make use of Creative Commons licenses that grant use of the copyrighted images for noncommercial purposes. (See creativecommons.org for more information on the use of Creative Commons licenses.)

Current Status of Paleoinformatics in Vertebrate Paleontology

Several freestanding, Web-based databases are now available to the community at large. Most of these databases were originally developed to address specific research questions, and have since grown to become useful in many ancillary applications. For the most part, each one is a 'stand-alone' database, meaning it has a more-or-less unique data structure, resides on its own server or servers, is maintained by a certain working group, has analysis tools that are specific to the database, and deals with metadata (documenting how the data in the database were actually obtained) in its own way. Some of the databases allow immediate download of all of their data; others require various permissions or joining the database group.

Future breakthroughs may well be possible through building distributed database networks, which would vastly enhance the abilities to pick and choose data fields from separate databases in order to assemble sets of information that presently cannot be easily obtained.

DATABASE REVIEWS

Several databases have been identified for inclusion in this review based on the following criteria: (1) including vertebrate fossil data or associated data of some sort; (2) the database is active and dynamic, that is, data are continually being added and updated; and (3) being research (rather than collection) databases. Reviews are presented in alphabetical order. These databases are summarized in Table 1.

Ancient Human Occupation of Britain (AHOB)
(www.ahobproject.org)

Basic Statistics—Number of records/data sets: 510 sites, 1858 faunal occurrence records, 300 archaeological assessments, 133

TABLE 1. List of databases reviewed herein, with their respective URLs and a summary of major data types included in each database.

		Included data types*					
Database	URL	Matrices	Occurrences	Stratigraphy	Taxonomy	Images	References
AHOB	www.ahobproject.org/database/		X	X		X	X
FAUNMAP	www.neotomadb.org www.ucmp.berkeley.edu/neomap/		X	X	X		X
MIOMAP	www.ucmp.berkeley.edu/miomap/		X	X	X		X
Morphbank	www.morphbank.net/			X		X	X
MorphoBank	www.morphobank.org	X			X	X	X
Neotoma	www.neotomadb.org		X	X	X		X
NOW	www.helsinki.fi/science/now/		X	X	X		X
PaleoDB	paleodb.org/		X	X	X	X	X
Polyglot Palaeontologist	paleoglot.org/						X
TreeBASE	www.treebase.org	X	X				X

*Many of the databases include additional data not covered under one of the broad categories listed.

stable isotope readings, 141 photos. Number of contributors: ca. 20

Goals and Mission—The goals of AHOB are to document data relating to British and European Quaternary sites investigated by the Ancient Human Occupation of Britain project.

History—The Ancient Human Occupation of Britain (AHOB) project is a multi-institutional program to better understand the geographic, climatic, and environmental contexts of the earliest human occupation of northwestern Europe (Stringer, 2006, 2011). The AHOB database was an integral part of the project from its start in 2001. The goal of the database is to document the data on which project findings are based, thus it includes both published and new data and does not aim to be comprehensive (Polly and Stringer, 2011).

Governance and Funding—The AHOB database is governed by the Ancient Human Occupation of Britain project; C. B. Stringer and N. Ashton, directors; P. D. Polly, coordinator. The AHOB project has been funded by three program grants from The Leverhulme Trust (2001–2013).

Participation in Ancient Human Occupation of Britain—Regular and Associate members of the AHOB project contribute data. Data are currently available only to project members, and will be made publicly available at the end of the project in 2013.

Included Data Types—AHOB includes data on Quaternary sites from Britain and Europe. Data are organized by site, and include site coordinates, faunal lists, indication of archaeological industry, site descriptions, photos, dates, publications, stable isotope readings, and other site-specific information. Morphometric and other kinds of specimen-based information are included for a limited number of taxa and sites.

Data Sources—Data are derived from publications and original research by project members.

Data Protection—Access is restricted to project members during the life of the project, after which access will be publicly granted.

Analysis—AHOB facilitates (bio)geographic, stratigraphic, climatic analyses of Britain.

Links to Other Databases—AHOB is integrated with GoogleEarth, and will have integration with NOW, PaleoDB, NESPOS, and PaleoAnth portal.

Publications—More than 250 publications have results from the AHOB project, all of which are recorded in the database. Associated data for many of these papers are stored in the database as well. Polly and Stringer (2011) published a description of the database and its scope. Polly and Eronen (2011) published an analysis of Quaternary climate at British sites, with an assessment of the long-term stability of the relationship between species and climate, using data from the AHOB database.

FAUNMAP (www.neotomadb.org and www.ucmp.berkeley.edu/neomap/)

Basic Statistics—Two versions of FAUNMAP are now available online and will continue to be online for the foreseeable future. One is the version served through the NEOMAP portal, and here termed FAUNMAP III. FAUNMAP III combines FAUNMAP I and FAUNMAP II with updates from literature through 2003. A second version is served through the Neotoma Paleoecology Database (FAUNMAP IV; see below). As of April 2012, the FAUNMAP III database contains information on all published localities (5015) and taxon occurrences (61,640) of Pliocene through Holocene mammals that were reported from the 48 contiguous states of the United States, Canada, and Alaska in peer-reviewed and gray literature through 2003.

Goals and Mission—FAUNMAP was originally produced as a research database, with the principal goal of examining changes in mammalian community composition over the past 40,000 years. Three fundamental ecological questions were addressed: (1) Do mammalian communities respond to environmental changes as tightly linked, highly coevolved assemblages of species or do they respond as individual species in accordance with their own tolerance limits? (2) How has provinciality of mammalian distributions changed? (3) Have environments become more or less patchy?

History—Graham et al. (1987) published a compendium of data and its synthesis on the late Quaternary mammal faunas of the Great Plains and Prairies, which was a precursor to FAUNMAP but established its basic structure. The FAUNMAP database was initially completed in two parts. FAUNMAP I spans the time period from 40,000 years ago to 500 years ago (i.e., the effective range of radiocarbon dating and prehistory in North America) and it was geographically restricted to the contiguous 48 states of the United States. FAUNMAP I was a community project led by R. W. Graham and E. L. Lundelius Jr., with the original data assembled in Paradox and served through the Illinois State Museum. Because the Web was not universally available at that time, FAUNMAP I was also distributed as a cross-referenced hardcopy with a diskette (Graham et al., 1994). Primary data entry began on FAUNMAP I in 1990 and concluded in 1994 at the Illinois State Museum. Over the past 15 years, periodic updates have been made, the last one in 2003. FAUNMAP II was a project undertaken by Graham and Lundelius to extend the FAUNMAP I–style data back through the entire Pleistocene and Pliocene and incorporate data from Canada and Alaska. Data entry for FAUNMAP II occurred between 1999 and 2003 at the Denver Museum of Natural History (now known as the Denver Museum of Nature & Science) and assembled in Access. FAUNMAP III, primarily constructed for NEOMAP by M. Carrasco, is a combination of FAUNMAP I and II.

Governance and Funding—Both FAUNMAP I and FAUNMAP II were funded by the NSF Earth Sciences Division (NSF EAR 900514 and ESH 9807499, respectively). The Illinois State Museum and Denver Museum of Nature & Science provided institutional support. Initially, FAUNMAP I was assembled by a consortium of Quaternary vertebrate paleontologists, each with regional expertise (Regional Collaborators), who met at a series of workshops over three years, under the direction of Graham and Lundelius. The structure, data types, data sources, and research databases (Graham et al., 1994) were approved by the Regional Collaborators at these meetings. FAUNMAP II followed the same guidelines.

Participation in FAUNMAP—FAUNMAP was not designed for scientists to contribute data to the database directly. Instead, publications could be sent to Graham and/or Lundelius who then had these data uploaded into the database. There were not any restrictions on who could use the online database. In the new Neotoma Paleoecology Database (see below), vertebrate data can now be directly submitted for incorporation into the database.

Included Data Types—The primary data are published occurrences of mammals of Pliocene through Holocene age from the 48 contiguous states of the United States, Canada, and Alaska. The basic data in FAUNMAP are species occurrences and counts (number of identified specimens [NISP] and/or minimum number of individuals [MNI]) if available. Metadata include geographic location, geochronologic ages (both relative and 'absolute'), site types, depositional environments, and cultural associations, all of which are tied to bibliographic data. Analysis units, which may be identified by depth and/or stratigraphy, chronology, and, sometimes, cultural affiliation, provide for the subdivision of sites into temporal increments. Most Holocene sites have faunas derived from archaeological excavations, whereas most pre-Holocene sites are paleontological.

Data Sources—FAUNMAP data are derived from the literature and require that specimens are in public repositories and have geographic and chronological data associated with them.

Data Protection—Because FAUNMAP is served by both NEOMAP and Neotoma, data protection is determined by the protocols of those two entities.

Analysis—Initially, FAUNMAP was served by the Illinois State Museum, from which one could download the database and work with it on a personal computer. In addition, FAUNMAP was linked with ARC INFO on the ISM server and investigators could create maps of the occurrences of individual species for different time periods. Maps of modern distributions could be overlain on these dispersal maps as well as the location of the Laurentide Ice Sheet at different time intervals. Printouts of these maps would include lists of the sites and their basic data. Faunal lists could be obtained for the sites by mousing over the individual sites on the ISM server. Many of these services have been transferred to NEOMAP, Neotoma, or both.

Links to Other Databases—FAUNMAP is now linked to NEOMAP and Neotoma.

Publications—The primary publication derived from FAUNMAP I was Graham et al. (1996). However, the database has been used by many other scientists to provide a broad spectrum of publications that are available on the Neotoma Web Page under Publications (www.neotomadb.org/index.php/references). The power of these public databases is shown by the increase in the number of published papers that have resulted from them through time.

Miocene Mammal Mapping Project (MIOMAP)
(www.ucmp.berkeley.edu/miomap/)

Basic Statistics—As of April, 2012, the MIOMAP database contains information on all published localities (3977) and taxon occurrences (18,231) of late Oligocene (Arikareean NALMA) through Miocene mammals that were reported from the United States in peer-reviewed literature through mid-2011. It also provides a bibliography of references (16,094 citations out of 1398 references) from which data were mined, which can be downloaded in Word or EndNote files. Data were assembled largely by M. Carrasco, A. D. Barnosky, E. B. Davis, and B. Kraatz.

Goals and Mission—The database was produced with the principal goal of conducting research on how major disruptions to the physical environment affected species richness, evolutionary patterns, and biogeographic patterns in mammals from approximately 30 million to 5 million years ago (Arikareean through Hemphillian Land Mammal Ages). Environmental disruptions of interest initially included middle Miocene tectonism in the northern Rocky Mountains and Great Basin, and climatic warming events of the late Oligocene and mid-Miocene. Subsequently the database has been used in a variety of research applications, ranging from developing methods to correct for sampling biases to understanding biodiversity baselines relevant to conservation biology.

History—See Carrasco et al. (2007) for details. Assembly of the database began in the year 2000 to address the research questions noted above, and MIOMAP was launched online for the general community in 2005. During the initial five years, data assembly concentrated on states in the western half of the United States. M. A. Carrasco extracted the data from the literature and entered it for most of the states. For the states of Idaho, Montana, and Wyoming, data were originally compiled by A. D. Barnosky; checked, added to, and entered by S. Minter and B. P. Kraatz; and updated by A. D. Barnosky and M. A. Carrasco. The Nevada portion of the database was developed by E. B. Davis. From 2005 to 2009, Carrasco completed data entry from the remaining (eastern and southern) states, Web delivery and analysis tools were improved, and linkages with the FAUNMAP database through the NEOMAP portal were established in order to address questions related to anthropogenic effects on the long-term biodiversity baseline for mammals. The database was largely completed by 2010; since then, maintenance has been ongoing and new data are entered as they become available.

Governance and Funding—The development of and research using MIOMAP was largely funded by the NSF from 2000 to 2010. The University of California Museum of Paleontology has been a partner from the outset and provides the long-term storage and delivery of the data via its servers and Web staff. The database has been directed from its beginning by A. D. Barnosky and M. A. Carrasco, with involvement of graduate and undergraduate students who have helped with extracting data from the literature and in programming tasks. UCMP provides ongoing technical support for the Web site and the database, and the Berkeley Natural History Museums have provided and supported the Berkeley Mapper application, which was developed by J. Deck and extended and customized for MIOMAP by E. Davis.

Participation in MIOMAP—There is no formal membership criterion. Those interested in contributing data should contact A. D. Barnosky. The database is fully downloadable and all parts of the Web site are fully functional for anyone with a Web connection.

Included Data Types—The primary data are published occurrences of mammals of late Oligocene (Arikareean) through Miocene (Hemphillian) from the United States. For each species occurrence, the following data were entered: taxonomic name (as of 2011), latitude and longitude, relative age range, geological occurrence, and taphonomic information. Details are explained on the MIOMAP Web site at www.ucmp.berkeley.edu/miomap/about/data.html. When possible, counts for MNI and NISP per taxon per locality were included. Ancillary data include high-quality digital images linked to original type descriptions for many Miocene type specimens. Full metadata following Federal Geographic Data Committee format are posted on the Web site.

Data Sources—The vast majority of data are information extracted from peer-reviewed published literature; in two cases (the Cabbage Patch Beds and Anceney, Montana), information from doctoral dissertations (by D. Rasmussen and J. Sutton, respectively) was also utilized. Unpublished specimens were included for a few key areas for which little published information existed (Hepburn's Mesa, Montana; Railroad Canyon, Idaho; the state of Nevada). In most cases these specimens were from areas in which Barnosky had worked extensively and for which better identifications than had been published were available (Hepburn's Mesa, Railroad Canyon) or for which primary field notes and examination of key specimens was possible (Nevada). Unpublished specimen counts were also included for a few localities by surveying museum online databases. See the Web site for details (www.ucmp.berkeley.edu/miomap/about/data-acquisition.html).

Data Protection—Copies of the database reside on servers in the UCMP that are backed up on a regular basis both off site and on site. All of the servers are secure servers.

Analysis—Besides the standard query interfaces that allow searches by taxon, geographic area, temporal bin, etc., MIOMAP also provides tools to facilitate biodiversity analyses. A key feature is the ability to select a polygon on a map, get a list of the taxa within the polygon and the geographic area encompassed by the polygon, and export the information into a table that can be directly imported into the EstimateS diversity analysis program.

Links to Other Databases—MIOMAP is linked to FAUNMAP through the NEOMAP portal.

Publications—As of 2012, the MIOMAP database has been utilized in at least 28 peer-reviewed papers by MIOMAP researchers, including publications in Science, Proceedings of the National Academy of Sciences of the United States of America (PNAS), PLoS Biology, Journal of Vertebrate Paleontology (JVP), and two books. Recent examples include

Barnosky et al., 2007, 2011a, 2011b; Barnosky and Kraatz, 2007; Carrasco et al., 2007, 2009; and Barnosky, 2008, 2009. Other researchers have used the MIOMAP data as a key part of at least five additional peer-reviewed publications. There have also been at least 50 professional-meeting presentations and abstracts that have resulted. Most of the MIOMAP-based publications are listed online at www.ucmp.berkeley.edu/miomap/results/index.html.

Morphbank (www.morphbank.net/)

Basic Statistics—There are 350,254 public and 42,679 nonpublic images in Morphbank as of April 2012; nonpublic images have yet to be released by the image contributor. As of that same date, there are 591 user accounts representing 46 countries.

Goals and Mission—Morphbank is a repository of biological and paleontological images and associated data about the images and their subjects, as well as a collaborative environment for research and education in those areas.

History—Work began on Morphbank in 1998 by a Swedish-Spanish-American group of entomologists for the purpose of their collaborative morphological phylogenetic research involving images, and in 2000, they launched a Morphbank Web site. In 2002, the first paper based on Morphbank images was published (Fontal-Cazalla et al., 2002). In 2005, the scope of Morphbank's functionality expanded to accommodate other areas of biology and paleontology with a grant to a group of principal investigators in Florida State University (FSU)'s School of Computational Science and Department of Biological Science from the NSF's Division of Biological Infrastructure. In 2009, the same NSF Division awarded principal investigators (PIs) at FSU's College of Communication and Information and Department of Biological Science a collaborative grant to integrate Morphbank more seamlessly with Specify (specifysoftware.org/; a natural history collection database based at the University of Kansas's Biodiversity Institute) and Morphster (www.morphster.org/; a morphological ontology management system at the University of Texas at Austin's Departments of Computer Science and Geological Science).

Governance and Funding—Morphbank is currently managed by FSU's Institute for Digital Information and Scientific Communication (iDigInfo; www.idiginfo.org). Past sponsors of Morphbank software development and outreach activities include the NSF, FSU, The National Evolutionary Synthesis Center, the Swedish Research Research Council, and Uppsala University. Current sponsors of Morphbank include the NSF and FSU's iDigInfo.

Participation in Morphbank—New Morphbank account requests are submitted via an online form. Account requests are vetted by Morphbank staff to confirm that the user is a member of the biological or paleontological research community prior to account creation. Accounts are free. An account holder can submit images and data and edit and delete images and data that have not yet been made public on the Web site. The user submitting content is presented as the 'Submitter'; the Submitter can choose to credit another user as the 'Contributor' of the images and data, for example, if the PI (the Contributor) has hired a technician (the Submitter) to submit the content.

Users can create one or more Groups of users for collaboration. The role assigned each user within a Group determines the user's ability to manage the Group's content. A 'Scientist' may add content for the Group, but that user may only modify or delete that content that s/he added; a 'Lead Scientist' has the ability to add content but can also modify or delete any content held by the Group regardless of their status as Submitter; the single 'Coordinator' for the Group has the privileges of a Lead Scientist but can also add and remove Group members and change member roles. A user can have different roles in different Groups (e.g., be a Scientist in one and a Coordinator in another).

Included Data Types—Morphbank can store and display data about the images (image metadata), the subject of the images (standard view descriptions, specimen records, localities, taxonomic names, and annotations), and the research publications that are the sources of data. A mapping of Morphbank fields to those in Darwin Core (rs.tdwg.org/dwc/) and the Access to Biological Collection Data schema (wiki.tdwg.org/ABCD/) is available at www.morphbank.net/About/Manual/dwcabcdmb.php. For paleontologists, Morphbank is able to store information about the lithostratigraphic group, formation, membership, and bed in which the subject of an image was found. Users can submit institutional or project logos to display on the pages for objects.

Morphbank also allows a user to create one or more Collections of images that are each displayed in a slide sorter interface. In that interface, image tiles can be dragged and dropped into a new serial order, which is stored in the system.

Data Sources—The two principle sources of images, image metadata, and information about the subject(s) of the images in Morphbank are biodiversity researchers engaged in new data collection (e.g., for taxonomy, phylogenetics, comparative morphology, anatomy, and histology) and natural history collections engaged in digitization of their specimens. Data can be submitted via a Web interface, an Excel workbook, XML, or Specify; for a more complete description of these options, see www.morphbank.net/About/Manual/submit.php.

The taxonomic name data in Morphbank was bolstered with an import of names from the Integrated Taxonomic Information Service (www.itis.gov) in 2004, but new names are regularly added by users.

Data Protection—The Submitter provides a date to publish an image and specifies the Group managing the image upon submission. Only the Submitter and any user in the chosen Group can see the image prior to its release. Visitors to Morphbank that are not logged in (Guests) can only search and view images and data that have reached their date-to-publish.

The Morphbank project strives to facilitate sharing of images and data. Morphbank requires that all images that are uploaded be associated with a Creative Commons License 3.0 (BY-NC-SA; creativecommons.org/licenses/by-nc-sa/3.0/us/) or a less restrictive license. The CCL 3.0 license allows another user to share and adapt the image, but attribution must be given to the Contributor of the image, the image cannot be used for commercial purposes, and the image (even after it is altered in some way) must be shared with the same or similar license. Morphbank is a collaborator with Encyclopedia of Life (EOL; eol.org) and gives Submitters the option to share with EOL upon submission of images and data.

Morphbank has Web services that permit queries of the database with output returned in a variety of formats, including XML, RDF, RSS, and images. For a more complete description of Morphbank's Web services, visit www.morphbank.net/About/Manual/services.php or see services.morphbank.net/mb3/.

Analysis—Morphbank allows users to share (in Groups), organize (in the slide sorter interface), and comment upon (with annotations) images early in the collaborative research workflow by providing researchers an environment in which they can work privately prior to release of their content. With publication of their research, researchers can publish their images on Morphbank, making the step from private analysis to public release an easy one.

One example of a paleontology project that has used Morphbank is the study of Alberto Prieto-Márquez and Greg Erickson on hadrosaurid dinosaurs (Prieto-Márquez et al., 2007). That work generated the first vertebrate fossil uploads to Morphbank to document museum holdings (e.g., www.morphbank.net/?id=

475738; www.morphbank.net/?id=143314), skeletal/histological materials, and graphics from peer-reviewed publications (e.g., from Leidy, 1858; Erickson et al., 2004; Prieto-Márquez et al., 2006; Erickson et al., 2007). During his Ph.D. dissertation research, Prieto-Márquez (2008) conducted a global survey of iguanodontian museum holdings, examining and photographing nearly every known specimen. Approximately 1700 of his images have been uploaded to Morphbank. Prieto-Marquez's Morphbank uploads are notable in that they also illustrate character-state codings using extensive annotations in Morphbank (e.g., those seen in the illustrative collection of his annotations at www.morphbank.net/myCollection/?id=796367). Ultimately, these efforts have made it possible for any researcher interested in the osteology of these animals to remotely access most of the world's holdings.

Links to Other Databases—URLs can be added to an image, specimen, view, locality, publication, and taxonomic name record; these external links are displayed in the page for each of these objects. For example, a user might choose to create links to sequences in GenBank (www.ncbi.nlm.nih.gov/genbank) from images of the source of the DNA data.

Links into Morphbank are simple and persistent: each image, specimen, view, locality, publication, and taxonomic name has a unique identifier in the system, and one constructs the Morphbank link as (morphbank.net/) + the unique identifier for the specific object.

MorphoBank (www.morphobank.org)

Basic Statistics—MorphoBank (O'Leary and Kauffman, 2011) currently has almost 1,000 registered scientists and students worldwide, compiling phenomic (e.g., morphological) or combined matrices or publishing annotated phenomic data that supplement published material. These content providers have work in over 600 separate Projects, one fourth of which have been released to the public. Several large, collaborative phenomic projects sponsored by the NSF, including the Assembling the Tree of Life projects for Mammalia and for bivalves, have their data in MorphoBank to be published in the near future. The site contains over 90,000 media elements, that is, images (two- and three-dimension [2D and 3D]) and video and over 500 phenomic matrices that are being released by scientists as they publish their work.

MorphoBank tools are used by an average of 200 active researchers a week to develop and analyze their password-protected data sets, many of which become the basis of published papers. Over 1500 visitors a month explore published data on MorphoBank.org.

Goals and Mission—The primary goals of MorphoBank are (1) to make it easier to build phylogenetic matrices from phenomic data; (2) to make this process more repeatable by storing and displaying such matrices via the Web; and (3) to enhance this practice with tools to associate images with matrix cells. These efforts are primarily designed to aid scientists who collect phenomic data to build evolutionary trees. MorphoBank facilitates and encourages the documentation of homology statements with images because morphological homology is often more clearly explained with labeled pictures than with words alone.

To accomplish this goal, MorphoBank provides a collaborative workspace for individuals or teams building phenomic matrices, and a place to store matrices long term so that scientists and the public may access the historic legacy of comparative morphological phylogenetics. Increasingly, scientists form research teams to build phylogenetic matrices, and working on separate copies of the data on separate desktops quickly becomes problematic. MorphoBank provides these scientists a private virtual environment to work together dynamically on the same matrix, in real time. The MorphoBank database is fronted by a user-friendly Web application so that scientists and students can access their data at any time from anywhere in the world. MorphoBank's major innovation is a robust, Web-based matrix editor that permits (but does not require) users to load images (and other media, including video) to document their homology statements. For example, if a scientist has a character 'Behavior: aquatic (0); arboreal (1)' those states could be illustrated in a matrix cell with a video or a picture. The program also has tools to save time and minimize scoring error. These include tools to allow authors to build rules dictating that certain characters are to be scored inapplicable if another character is scored to a particular state, and tools to automate and check character scoring for large blocks of inapplicable cells.

History—The idea for MorphoBank emerged in the late 1990s during meetings sponsored by the NSF that led to the creation of the 'Assembling the Tree of Life' program. Subsequently another NSF-sponsored meeting of morphological systematists at the American Museum of Natural History in 2001 was arranged by M. O'Leary and J. Caira for the community to contribute to the early growth of the site. Ekdale et al. (2011) is a particularly important milestone in the growth of the site because it is a research article that puts the site's tools to extensive use.

Governance and Funding—MorphoBank is overseen by an Executive Committee of comparative biologists and paleontologists. This committee is chaired by N. Simmons and includes five other senior scientists with diverse taxonomic expertise including vertebrate paleontology (P. Crane, G. Giribet, M. Novacek, D. Stevenson, and M. Wake) and a student representative (J. Wolfe). The day-to-day operations of the project are overseen by M. O'Leary, who supervises a software development team of four including engineers/computer scientists (S. Kaufman and K. Alphonse), Web developers and designers (M. Passarotti and A. Waller), and a curator from library science (N. Milbrodt). To promote transparency in project management, all work directed by O'Leary is performed in a project management database that is visible to the team.

Feedback on the software reaches the team daily from users worldwide. All feedback is entered into the project management system and queued for development. It is a priority to listen very closely to user requests and the site has been improved enormously with these formative evaluations.

MorphoBank has been supported by multiple grants from the NSF's Division of Earth Sciences and Division of Environmental Biology, as well as grants from NESCent (National Evolutionary Synthesis Center) and support from the American Museum of Natural History.

Participation in MorphoBank—Published MorphoBank projects and their matrices and images are freely available to anyone in the world without registration or other restriction (although media reuse varies from image to image and is specified by the copyright holder, not by MorphoBank). Registration is required to submit content to MorphoBank; permission is granted broadly to the scientific community. Student members are asked to register their advisors and to include them on their projects so that the student work is supervised.

Included Data Types—The research data MorphoBank contains are matrices (in Nexus or TNT format; it is also a place to convert between formats) and media (2D and 3D images, video). These are linked to a range of metadata (e.g., taxa, specimen, repository, anatomical views, copyright, extinct/extant), which are detailed in the schema posted on the site. Media may be used to illustrate not only the scoring of matrix cells, but also characters and taxa. A visual annotation tool allows researchers to mark up images with comments and references to specific anatomical features.

Data Sources—MorphoBank has not absorbed mass input from other databases; its population of data are entirely newly entered by scientists. MorphoBank has grown organically to contain over 500 separate projects in use by the community of 1,000 registered scientists and students. The work the site contains is a reflection of day-to-day data collection by comparative biologists.

Data Protection—An advantage of a Web-based system like MorphoBank is that maintenance tasks such as data backup can be centralized. MorphoBank runs on servers hosted by the State University of New York at Stony Brook. All data on the main MorphoBank server at Stony Brook is backed up to high-capacity tapes, as well as to off-site mirror servers at the American Museum of Natural History on a nightly basis. By distributing backups across servers and locations, we are able to minimize risk of data loss and recover quickly in the event of an outage.

As an added level of security, users are encouraged to maintain their own backups of project data. To this end, MorphoBank generates a project-based download feature that provides, on demand, a snapshot of project data that includes all original high-resolution media, matrices and associated taxonomic, and specimen data in a zipped archive.

Analysis—MorphoBank facilitates phenomic and combined data (morphology [phenomic] + molecular data) phylogenetic systematics research by hosting and displaying the core data (Nexus, TNT files) needed to repeat, and build on, existing phylogenetic studies. The sophisticated matrix editor is designed to support morphological data; a full combined matrix can be stored in MorphoBank's 'Documents' folder. MorphoBank also serves as a repository for supplementary images associated with a publication but not affiliated with a matrix (Project 67; Liu and Yang, 2006) and a standard permalink exists for all projects making them easy to find. The site is not a tool for tree analysis or tree visualization.

Links to Other Databases—MorphoBank is linked to the Paleobiology Database (PaleoDB) and uBio, allowing users to search these sites for opinions on higher taxonomy. Published MorphoBank data are also searchable within the PaleoPortal.org collections portal, which incorporates specimen and locality data from over a dozen North American paleontology collections.

Publications—Many contributions to MorphoBank integrate information on phylogenetic matrices that include vertebrate fossil data, and we mention a few of those here. Ekdale et al. (2011) supplied almost 600 images to support homology statements in their large combined analysis on cetaceans (Fig. 1). Others, including Gunnell and Simmons (2005) on fossil bats, Hill (2005) on fossil archosaurs, Claeson et al. (2010) on fossil batoids, and Nesbitt et al. (2011) on dinosaurs, have used the site to deposit their phylogenetic work, making it directly available for others to expand upon.

Neotoma Paleoecology Database (www.neotomadb.org)

Basic Statistics—As of April, 2012, the Neotoma Paleoecology Database contains information for Pliocene through Holocene mammals and fossil-pollen data. The faunal component of the database contains 44,508 taxon occurrences in 4987 analysis units from 2929 sites (FAUNMAP I). FAUNMAP II is slated for incorporation into Neotoma in the near future, and thus will add 20,750 taxon occurrences in 4487 analysis units from 2053 sites. Thus, the mammal component of Neotoma (FAUNMAP IV) will soon have a total of 65,258 taxon occurrences in 9474 analysis units from 4982 sites. These sites include Pliocene through Holocene mammals that were reported from the 48 contiguous states of the United States, Canada, and Alaska in peer-reviewed and gray literature through the present. In addition, there are 1836 stratigraphic pollen data sets from 1658 sites plus 2441 pollen surface samples. These data will be supplemented by plant macrofossils, other vertebrates, ostracodes, diatoms, and many other data types in the near future.

Goals and Mission—Neotoma is an open-access, community-implemented database in Microsoft SQL Server, a highly scalable client-server relational database management system suitable for Web servicing. Inasmuch as Neotoma makes data available via Web services, it is not necessary to download the database to have full access, although that option is available. Neotoma is developing a variety of tools for use with the data. Neotoma Web services, Application Programming Interfaces (APIs), and a software development kit (SDK) will also allow researchers to develop their own desktop or Web-based software applications, which will have real-time access to the current database. These tools can be made available from other Web sites (e.g., MIOMAP, PaleoDB, NOAA Paleoclimate Database) and thus serve as alternative clients to Neotoma or portions of Neotoma, or they may be distributed from the Neotoma Web site for others to use.

The Neotoma Paleoecology Database is a relational database comprising a number of constituent 'databases' (e.g., vertebrates [FAUNMAP IV], pollen, plant macrofossils, beetles) that share similar structure, which facilitates data entry from diverse disciplines as well as easy comparisons of different data types. The temporal coverage is the Pliocene, Pleistocene, and Holocene, the time during which modern species, including humans, and modern ecosystems appeared. The versatile structure of the Neotoma database makes it amenable to the incorporation of other types of data. The philosophy behind Neotoma is (1) open and easy data access, (2) easy cross-taxa data retrieval, (3) high-quality data (contributed and maintained with quality control by disciplinary communities), (4) easy interfaces with other databases, (5) stimulation of new and innovative research, and (6) cost-effective data management. The power of Neotoma is that vertebrate data, as well as any other data types, are easily compared with other proxies (e.g., vegetation, isoscapes, ostracodes).

History—The Neotoma Paleoecology Database was initiated by E. C. Grimm and R. W. Graham in 2006 in order to merge FAUNMAP I and II and the North American Pollen Database (NAPD) that was originally served by the National Oceanic and Atmospheric Administration (NOAA) Paleoclimate Database. Both FAUNMAP and NAPD were developed independently in the early 1990s. Graham and Grimm quickly realized the power of a comprehensive database such as Neotoma and enlisted experts in other taxonomic groups (e.g., plant macrofossils, beetles, ostracodes, diatoms) as collaborators in the effort. The vertebrate data in Neotoma will be referred to as FAUNMAP IV, which will include FAUNMAP I and II as well as new and future data uploaded into Neotoma.

Governance and Funding—Each constituent database or working group within Neotoma, including FAUNMAP IV, is responsible for its own governance and appointment of data stewards. The governing body, currently just for North American data, determines the type and format of the data and metadata as well as taxonomic, chronological, and other disciplinary standards for thorough data documentation. Likewise, if other geographic areas wish to contribute data, they would establish their own governing body and data stewards. This structure allows the appropriate proxy community (e.g., mammals, pollen, isotopes, taphonomy) and geographic regions (e.g., North America, South America, Australia, etc.) to set and maintain the quality standards for their own disciplinary data.

NSF EAR (0622349, 0622289) supported the first efforts to develop the Neotoma structure and merge the faunal and pollen databases. Neotoma now has a five-year grant from NSF EAR Geoinformatics (0948652, 0947459) to expand the database with other proxies, to make it a Web services database, and to develop various tools to facilitate use of the database.

Participation in Neotoma—Membership is not required to participate in Neotoma and any one can contribute data as described on the Neotoma Web Page (www.neotomadb.org/index.

php/data/category/contribution). However, only appointed data stewards can directly enter data into the Neotoma database (see Data Sources below).

Included Data Types—FAUNMAP IV is the vertebrate constituent of Neotoma and is broadly illustrative of the kinds of data stored in the database, although each constituent database has data and organization appropriate for the data type. Neotoma vertebrate data will not be restricted to mammals but will include all of the other vertebrate classes (fish, amphibians, reptiles, and birds). Isotopic and taphonomic data derived from fossil vertebrates from the Plio-Pleistocene and Holocene will be hosted in Neotoma and linked to the FAUNMAP IV data. The taphonomy database within Neotoma will host a variety of types of taphonomic data (actualistic studies, landscape surveys, live-dead comparisons, site-specific data for fossil sites, etc.) for plants, invertebrates, and vertebrates.

Data Sources—Data are derived from a host of sources. Most of the data in FAUNMAP IV are derived from the literature but individuals can now contribute their data directly to Neotoma. In addition, Neotoma will allow for specimen specific data that will facilitate uploading the primary research data used in the original site analysis.

Data Protection—A software interface is currently under development that will enable data stewards to remotely upload and manage data in Neotoma. The client end of this interface is built on the Tilia software originally designed for managing microfossil data and associated metadata from sediment cores. Tilia ensures consistency in data entry through a standard structure with a series of pull-down pick lists. Web services will enable Tilia to check data against lookup tables in the main database, to add new lookup items (such as new taxa names or depositional environments) if necessary, and finally to upload the data once they have been checked for completeness and consistency. Data contributors can enter data in an end-user version of Tilia, or they can submit data to Neotoma in other electronic formats (e.g., Excel, text files) or even as paper hardcopy. The Neotoma Web page provides instructions on how Tilia can be acquired (www.neotomadb.org/index.php/data/category/tilia). Those wishing to develop a data consortium (data type or geographic regions) should contact one of the data managers (www.neotomadb.org/index.php/contacts/investigators).

Before data are uploaded into the database, they are reviewed by a data steward for completeness and consistency. This review, which is automated by the software interface, ensures that site names and locations are provided, names of investigators are included, bibliographic data are entered, taxonomic names are spelled correctly, and so on. One important purpose of this review is to ensure that contributed data sets are credited to the appropriate people and publications. Data awaiting review are stored in an online 'holding tank' as flat files where they are available to anyone to use but do not have a guarantee of standardization as do data extracted from the relational database. Following review, files are uploaded by the data steward to the database.

As discussed above, anyone can contribute data but only appointed data stewards can enter data directly into the database or make changes in the database. Because many investigators routinely manage microfossil data in Tilia for their own use, this software has proved very useful for database submission, because these data can then be quickly uploaded to the database when they are ready for public release. The expansion of Tilia to better accommodate faunal data may similarly expedite the flow of these data to Neotoma. Data are automatically backed up on servers at the Pennsylvania State University that are located in separate facilities as a contingency against disasters.

Analysis—Currently, the primary data-exploration tool is Neotoma Explorer, which is an interactive Web application for querying and visualizing information from the database. The basic concepts of Neotoma Explorer are (1) discover: find information quickly and efficiently with easy-to-use tools that filter the database by spatial, temporal, and metadata criteria; (2) explore: interactively present data and metadata from discovery so user may decide if data set meets needs; and (3) share: get data and information in a variety of useful formats (e.g., downloads, reports, graphics). Neotoma Explorer currently delivers data in five different formats: site summaries, data summaries, stratigraphic diagrams (similar to pollen diagrams), raw data, and maps. A special feature of Neotoma, and unlike many other databases, is that all of the data can be downloaded.

Investigators can either use the basic mapping features of Neotoma Explorer or specialized analytical mapping applications such as Neotoma TaxaMapper. TaxaMapper enables the mapping of taxa occurrences from the database at different time periods. Users can create multiple layers of taxon versus time, combine them into new layers with AND/OR operations, and customize layer symbology. These maps can be used to investigate, or document, species dispersals and interactions. Modern species distributions can be overlain to compare past occurrences to their modern ranges. Maps can be saved for publication or placed in other software and modified.

In the original FAUNMAP, chronology was handled in two different ways. Predefined bins for certain time periods (e.g., Full Glacial, Late Glacial, Early Holocene) were created in a Research Database (Graham et al., 1994). These binned chronological data are available in Neotoma. It was also possible to download original chronological data (e.g., radiocarbon ages, K-Ar ages) and create new or alternative chronologies, for example, ones that are based on a more recent radiocarbon calibration curve. Neotoma retains this function but enhances it with new chronological tools that construct age models for stratified sites.

Links to Other Databases—FAUNMAP IV data will link to GenBank. The mammal component of FAUNMAP IV will also be available through NEOMAP, which is a portal at the University of California at Berkeley that serves both the MIOMAP and FAUNMAP III databases. Neotoma is working with the NOAA Paleoclimate Program to establish Web services that enable Neotoma metadata to be discovered and data sets linked to NOAA Web search tools.

Publications—Neotoma maintains a bibliography of publications that have resulted from the database (www.neotomadb.org/index.php/references) as well as publications used in assembling the database.

New and Old Worlds: Database of Fossil Mammals (NOW; Formerly Known As Neogene of the Old World)

(www.helsinki.fi/science/now/)

Basic Statistics—The NOW database contains 3063 references, 13,361 species/taxa, 8353 localities, and 79,210 locality/species combinations. There are 95 NOW contributors (55 of whom are active) from over 17 institutions from Australia, China, Finland, France, Germany, Greece, Iran, Netherlands, Russia, Spain, Sweden, Turkey, United Kingdom, and the U.S.A.

Goals and Mission—The NOW fossil mammal database contains information mainly about Eurasian Miocene to Pleistocene land mammal taxa and localities. The emphasis of the database has been on the European Miocene and Pliocene but African localities are currently being added and updated, and North American localities will also be visible to the public in 2014. The temporal scale is also currently becoming wider, with both Pleistocene and Paleogene localities being added into the database. The NOW database is maintained and coordinated at the University of Helsinki by M. Fortelius in collaboration with an international advisory board.

History—A workshop organized by R. Bernor, V. Fahlbusch, and S. Rietschel in 1992 at Schloss Reisensburg in Bavaria, Germany, was the starting point in the development of NOW.

Participants in the workshop had been requested to compile revised lists of taxa and localities to produce a database for later distribution within the group. The purpose of the workshop was to review the evidence for provinciality and diachrony of change between Central Europe and the eastern Mediterranean realm during the Middle and Late Miocene.

Based on the Schloss Reisensberg workshop, the database was released to the public in December 1996 on the day that "The Evolution of Eurasian Neogene Mammal Faunas" (Bernor et al., 1996) was published. The first Web-based user interface was launched in 1999 as a clone of the one developed by the Evolution of Terrestrial Ecosystems Consortium for the Evolution of Terrestrial Ecosystems (ETE) database. A main source of coordinated data input and revision was also the European Science Foundation Network on Hominoid Evolution and Environmental Change in the Neogene of Europe (1995–1998). For most of its existence, the acronym 'NOW' has been derived from the words 'Neogene of the Old World' but because NOW has since expanded to cover most of the Eurasian continent for the post-Oligocene (ca. 25–0.01 million years ago), and is still expanding in both temporal and geographic scales, the words behind the acronym were officially changed to 'New and Old Worlds' in early 2012.

In 2009, there was an agreement with C. Janis that the North American fossil mammal data, based on the two volumes of Tertiary Mammals of North America (Janis et al., 1998, 2008), would become available through NOW. These data are in the process of being uploaded to the database and will be released to the public during 2014. In 2010, North American coordinators also joined the advisory board. Meanwhile, the African locality and taxon occurrences of NOW are currently being updated based on the Cenozoic Mammals of Africa book (Werdelin and Sanders, 2010) and other recent literature.

Governance and Funding—NOW is governed by an Advisory Board, led by M. Fortelius (Coordinator). The advisory board consists of a large number of professional paleontologists, acting as coordinators for the geographic, stratigraphic, taphonomic, paleoecological, taxonomical, and ecomorphological information stored in the database (www.helsinki.fi/science/now/board.html). All the members of the board are volunteers. The day-to-day business of NOW is run from the virtual office at University of Helsinki, Finland. In addition to the main coordinator, the NOW office includes the Associate NOW coordinators J. Eronen and L. Säilä, and office manager S. Sova. The database itself is also housed at the University of Helsinki. Between 2004 and 2011, the database was maintained as part of the Institute of Biotechnology's information technology (IT) structure, and the database will permanently move to the Natural History Museum's servers at the University of Helsinki in the near future. The NOW database has always been funded by research grants, without any stable source of funding.

Participation in NOW—The NOW database is fully public. The data are freely available through a Web-based graphic user interface. There is no charge for membership or any other service provided by the NOW database, nor does NOW database pay for the data. A username membership is open to all researchers who agree to abide by policy and want to contribute. Anyone who would like to become a member of NOW should contact the NOW office in Helsinki. Users have different editing capabilities, depending on their expertise and needs. Username is not necessary for data download, however.

Included Data Types—The nucleus of the database is a locality table and a species table, relationally linked by means of a locality-species correlation table (in essence, a table of localities and their faunal lists). Additions and updates are tied to a table of references. The aim of the database is that each species (including higher taxa such as 'Machairodontini indet. large sp.' or 'Rhinocerotidae indet. indet.') is given certain attributes describing its anatomy and inferred diet, locomotion, and other properties. Similarly, the geographic location, age, stratigraphy, lithology, taphonomy, and environmental interpretation are described for each locality to the extent that these are known. Users can perform searches based on locality name, country/geographic area, and age, as well as taxonomic names and groups. Lists of localities and/or taxon occurrences defined by the search criteria can be exported. In addition, they can be viewed on a simple border map or a Google Map, and the resulting images can be saved.

Data Sources—Sources for the data in the NOW database are mostly published scientific literature but unpublished data generated by members, data from previously published data compendia, data from individual research projects, and personal communications from advisory board members are also included.

Data Protection—The vast majority of the data are freely available for browsing and downloading. However, locality/collections data and their associated taxonomic lists can be restricted to other database members, database working group members, or themselves for a limited time. Also large data sets that are in the process of being uploaded are private until they have been thoroughly reviewed. The restriction of data is discouraged, because this goes against the philosophy of a public access database. Nevertheless, restrictions allow members and working groups to enter and utilize their own data sets in private for limited time periods and assures that the data become available to the public once the study has been published. Not all data sets that have been added to NOW are of equal quality, and users should be aware of this fact. Revision of the data by NOW advisory board members and all users is an ongoing process and an integral part of the NOW project.

Analysis—The analyses facilitated by the NOW database are varied. Most studies utilizing the data have focused on the relationships between fossil mammals, climate, and environment. Also, studies on community composition, trophic levels, trait evolution, taxonomy, paleoecology, temporal ranges, occupancy, and biogeography have been common.

Links to Other Databases—The MorphoBrowser database and interface is also operated at University of Helsinki (Institute of Biotechnology, J. Jernvall Evo-Devo research group) and utilizes the same taxonomic tables as the NOW database. MorphoBrowser is a 3D visualization and searching tool for mammalian teeth, accessible over the Web (morphobrowser.biocenter.helsinki.fi). MorphoBrowser allows users to view a diverse range of tooth morphologies found in both extinct and extant mammals, with digitized 3D models of teeth freely rotatable on the screen. In addition, data such as crown typing (Jernvall et al., 1996), shape descriptors generated directly from the 3D data (Pljusnin et al., 2008), and orientation patch count (OPC; a measure of tooth crown complexity correlated with dietary category; Evans et al., 2007) are displayed alongside the 3D tooth images. These parameters are also searchable, allowing comparisons between nearest neighbor teeth for each categorization to be made. In the near future, many more taxa present in the NOW database will have 3D information from their dentitions stored in MorphoBrowser, thus providing an additional layer of ecomophological information available, and MorphoBrowser user access will be made comparable to NOW.

Publications—There are a number of publications in top journals (Nature, PNAS, etc.) using data from NOW database, as well as large number of publications in more discipline-specific journals (e.g., American Naturalist, Paleobiology, Global and Planetary Change, Palaeoclimatology, Palaeogeography, Palaeoecology, PLoS, Computational Biology, Journal of Human Evolution, Evolutionary Ecology Research), ranging from data mining to climatology to paleobiology in topics. For a representative sample of studies, see, for example: Agustí et al. (2001); Fortelius et al. (2002); Jernvall and Fortelius (2002, 2004); Ukkonen et al. (2005); Puolamäki et al. (2006); Liow et al. (2008);

Eronen et al. (2009, 2010); Raia et al. (2011, 2012); and Tang et al. (2011).

Paleobiology Database (PaleoDB) (paleodb.org/)

Basic Statistics—As of April 2012, the PaleoDB contains 41,109 references, 23,2062 taxa (of all levels of taxonomy), and 1,017,422 occurrences in 126,724 separate collections. These data were entered by 296 contributors from 120 institutions in 22 countries.

Goals and Mission—The Paleobiology Database (PaleoDB) is an open resource for the global scientific community managed and run by a diverse group of paleobiologists. Its purpose is to provide global, collection-based occurrence and taxonomic data that are both temporally and bibliographically referenced for organisms of all geological ages, as well as software for analysis of the data. The long-term goal is to encourage collaborative efforts to answer large-scale paleobiological questions by developing a useful infrastructure and development of large, publicly accessible data sets (Alroy, 2000).

History—The PaleoDB originated in the National Center for Ecological Analysis and Synthesis (NCEAS)-funded Phanerozoic Marine Paleofaunal Database initiative, which operated from August 1998 through August 2000. The scope was expanded in 2000 to include terrestrial and, thus, all fossils. As the database has grown, it has incorporated data from several legacy databases. In 2002, the ETE database was incorporated into the PaleoDB. In addition, all of the taxonomic and age range data from the Sepkoski Compendium was uploaded into the PaleoDB in 2003 (Sepkoski, 2002). The PaleoDB also incorporates data from Alroy's North American Fossil Mammal Systematics Database, first uploaded to the PaleoDB in 2003 (see, e.g., Alroy, 1996).

Governance and Funding—The Database Management Committee makes decisions regarding data access, data storage, assignment of credit, and membership. The PaleoDB has a main office at Macquarie University that houses the main database server and staff. Mirror servers are maintained at the Museum für Naturkunde in Berlin and the University of Wisconsin–Madison.

The PaleoDB's core facility is funded by the Australian Research Council and its mirror servers are operated independently. The PaleoDB's analytical paleobiology training workshop has been funded by several collaborating organizations and has guaranteed support through 2013. Data entry projects are funded by grants to individual members. The PaleoDB was originally funded from 2000 to 2006 by a grant from the NSF's Biocomplexity program, and also has received funding from the NSF's Sedimentary Geology and Paleobiology program. Further development of the database is being funded by an NSF geoinformatics grant.

Participation in PaleoDB—Database membership is open to all researchers who agree to abide by PaleoDB policy and contribute a significant amount of data. There is no charge for membership or any other service provided by the PaleoDB and PaleoDB does not make payments to receive existing data sets.

There are three classes of PaleoDB members. Authorizers, who are professional paleontologists, are granted full database privileges. Most Authorizers have doctoral degrees, but serious researchers without graduate-level training also may be considered for membership. To become an Authorizer, one needs only to contact the Database Coordinator with credentials and a proposal for contribution to the database. Generally, students working under a supervisor towards a graduate degree are classified as the second class of member, the Enterer. Enterers work under the supervision of a particular Authorizer. Contact an Authorizer directly to discuss becoming an Enterer. Undergraduate students who are using the database during a class or for undergraduate research constitute the third class of members. Student members have limited access to enter data and do so under the supervision of an Authorizer. Student membership status may only be requested by an Authorizer by contacting the Database Coordinator.

Included Data Types—The PaleoDB currently includes several classes of data, including bibliographic references, taxonomic names, taxonomic synonymies and classifications, primary collection data, taxonomic occurrences, reidentifications of occurrences, and geologic time scales. Limited information on ecological and taphonomic attributes of higher taxa and species, measurements of specimens, and data about the digital images of specimens may also be included. The data are stored in various tables relationally linked with record ID numbers.

Data Sources—Sources for the data in the PaleoDB are mostly published scientific literature. Some data are unpublished data generated by members, and other data are from previously published data compendia (e.g., Carroll, 1988; Sepkoski, 2002) or from legacy databases that have been incorporated into the PaleoDB.

Data Protection—Bibliographic and taxonomic data entered into the PaleoDB are released to the public immediately. Collections data and their associated taxonomic lists can be released to the public immediately or restricted to other database members, database working group members, or themselves for up to two years for previously published data and up to five years for unpublished data. This allows members and working groups to enter vast data sets over many years for their own use, and also guarantees that these data will ultimately be available to the public once the research has been published.

Links to Other Databases—PaleoDB is linked directly to the Neptune database (a database of microfossil occurrences reported in Deep Sea Drilling Project and Ocean Drilling Program samples), and users can search and download data from both Neptune and PaleoDB simultaneously. PaleoDB and Neptune are both part of the CHRONOS (www.chronos.org) database consortium. PaleoDB also serves data to the Global Biodiversity Information Facility (GBIF; www.gbif.org).

Analysis—Many different types of paleobiological analyses are facilitated by the PaleoDB, including analysis of diversity, biogeography, taxonomy, taxonomic history, temporal range, paleoecology, etc.

Publications—Database members can mark publications as official publications of the PaleoDB, but this is not required. Many publications have resulted from both database members, and non-database members who have downloaded data sets using the public interface. A Google Scholar search for the term 'Paleobiology Database' yields about 680 hits, most of which are scientific papers that have used PaleoDB data in some fashion, with a few duplicates, and a few that aren't scientific publications. A similar search of GeoRef yields 134 hits. Good examples of official database publications that show the breadth of contributions resulting from PaleoDB data include: Alroy (2001, 2002); Fara (2004); Marx and Uhen (2010); Butler et al. (2011). A list of official PaleoDB publications can be found at paleodb.science.mq.edu.au/cgi-bin/bridge.pl?a=publications.

Polyglot Paleontologist (paleoglot.org)

Basic Statistics—As of February 29, 2012, the Polyglot Paleontologist hosts English translations of nearly 450 different publications, provided by 95 contributors working from 11 different original languages.

Goals and Mission—The Polyglot Paleontologist is a free database managed by M. Carrano at the Smithsonian's National Museum of Natural History. It serves as a centralized repository for English translations of non-English paleontological and related scientific literature. By providing a stable access point for translations, the Polyglot Paleontologist allows reliable free access and the opportunity to cite (or reference) translations along with the original references in published works.

History—The Polyglot Paleontologist was created at Stony Brook University in 1999 and hosted there until 2003, when it was migrated to a server at the Smithsonian Institution.

Governance and Funding—As a small database, the Polyglot Paleontologist is managed by a single individual and decisions about its governance are ad hoc. It requires no regular funding but receives infrastructural support from the Smithsonian's National Museum of Natural History. The site has also been supported by a Smithsonian Seidell Grant (for the dissemination of previously published scientific works).

Participation in Polyglot Paleontologist—There is no formal membership but site users are encouraged to sign up for periodic e-mail updates, which typically announce the availability of new translations. Anyone may submit additional translations to the database by sending an electronic file (preferably in PDF format) to M. Carrano along with appropriate bibliographic and translation metadata.

Included Data Types—The database hosts both PDF and DOC files, the latter mostly a legacy of the early stages of Web site development and associated limits on file sizes.

Data Sources—The translations are primarily based on formal scientific publications, but a small number also derive from unpublished works and literature for the general public.

Publications—Users are encouraged to cite both the translations and the database in the acknowledgements of any published paper that utilizes them. This allows readers to know of, and access, these same translations but also creates a 'chain of evidence' for the information contained therein.

TreeBASE (www.treebase.org)

Basic Statistics—As of April 1, 2012 TreeBASE includes data from 3056 studies; with 5965 character matrices, 8912 phylogenies, and 495,612 operational taxonomic units (OTUs) that map to 84,717 distinct taxa. These data sets were produced by 6370 distinct authors with 1438 new contributors since 2010.

Goals and Mission—TreeBASE is a research tool and digital library of phylogenetic data governed by the Phyloinformatics Research Foundation, Inc. (PRF; www.phylofoundation.org), and presently hosted by NESCent (www.nescent.org). Its purpose is to archive and disseminate phylogenetic trees, and the phylogenetic data used to generate them, that were used in peer-reviewed journals, books, and other publications. A primary goal is to attenuate the loss of scientific data that happens when large, complex, structured digital objects used in phylogenetic research are degraded into analog forms (e.g., figures in a printed publication), or digital forms that prevent data reuse (e.g., PDF files), or stored on resources that fail to retain critical information, such as the dealignment of sequences and the exclusion of insertion-deletion coding when data are submitted to GenBank (e.g., see Sanderson et al., 1993). Another goal is to promote the sharing and reuse of phylogenetic knowledge and to provide a resource for discovery of phylogenetic relationships.

History—M. Donoghue, T. Eriksson, W. Piel, and M. Sanderson initially developed TreeBASE's prototype in the mid-1990s. In 1998, it began accepting submissions from the scientific community and disseminating data through a Web interface built and curated by Piel. A complete redevelopment of the software was funded through the CIPRES project (www.phylo.org) in 2006, led by Val Tannen and Piel, with programming support from M. Miller's staff at the San Diego Supercomputer Center (SDSC), including L. Chan, M. Gujral, and J. Ruan. Later the team expanded with M. J. Dominus joining in 2007 and R. A. Vos in 2008 with support from the pPOD project (phylodata.seas.upenn.edu) (Piel et al., 2009). In 2009, Y. Guo and H. Shyket joined the project with support from Dryad (www.datadryad.org). In the Spring of 2010, the new version was deployed and launched at NESCent with help from H. Lapp, J. Auman, and V. Gapeyev, and continues to be hosted there under agreement between the PRF and NESCent.

Governance and Funding—The Board of Directors of the PRF, a Connecticut Non-Stock Corporation, provides oversight for decisions regarding database policies, support, and research. TreeBASE has received support as a component of several NSF-funded projects and it has received support from hosting institutions such as the Harvard University Herbarium, Leiden University, the University at Buffalo, the Yale Peabody Museum, and, presently, the National Evolutionary Synthesis Center. The PRF welcomes efforts by the scientific community to improve TreeBASE and welcomes efforts to seek funding for this purpose. All source code can be downloaded from the TreeBASE subversion repository at SourceForge under a BSD license (sourceforge.net/projects/treebase).

Participation in TreeBASE—TreeBASE welcomes submissions of phylogenetic data that are formatted in NEXUS (Maddison et al., 1997) and that were used to generate results published in any peer-reviewed scientific publication, including journals, books, and theses. Typically one of the authors on the publication submits the data, but submissions by third-party persons are also possible. Scientists who are willing to serve as data curators are welcome to request administrative access to TreeBASE, subject to approval by the PRF. Administrative access provides them with the ability to improve or augment the quality of data and metadata as well as the ability to approve new submissions by the public.

Included Data Types—TreeBASE accepts phylogenetic data in the form of taxon-by-character matrices, such as continuous characters (i.e., floating-point numbers) and discrete characters (i.e., morphological states, nucleotide sequences, and protein sequences), and phylogenies, with or without branch lengths or clade support values. Submitters of morphological data are encouraged to supply character name and character-state descriptions using the NEXUS 'CHARSTATELABELS' command. Submitters of molecular data are encouraged to define the location of different genes or loci using 'CHARSET' commands. Taxon labels supplied in the 'TAXA BLOCK' of the NEXUS file are mapped to uBio NamebankID identifiers and NCBI TaxID identifiers. GenBank accession numbers and basic Darwin Core (rs.tdwg.org/dwc) metadata, such as collection codes in compliance with the Registry of Biological Repositories (www.biorepositories.org) and collecting locality and georeferencing metadata, are accepted in a tab-separated text file. Submitters manually enter metadata about the analyses performed on the character matrices to produce the phylogenies.

Data Protection—Data released by TreeBASE are freely available to the public and are viewed by the PRF as electronically encoded ideas, as opposed to art, and therefore not subject to copyright restrictions. Submissions to TreeBASE proceed through three stages. First, submissions of state 'in progress' are only viewable and editable by the submitter and TreeBASE staff; the manuscript can be in preparation, submitted, in press, or published. When the manuscript is accepted for publication and the submitter wishes the data to be made public, the submitter can toggle the status to 'ready' state. Second, submissions of state 'ready' are viewable but no longer editable by the submitter except for citation metadata; the manuscript must be accepted for publication, with or without minor revision. These submissions are flagged for TreeBASE staff for review, and if they meet minimum standards they are pushed to 'published' state. Third, submissions of state 'published' are viewable by the public and are no longer editable by the submitter except for citation metadata.

TreeBASE provides submitters with a special URL that can be used by the journal editors or reviewers to view data for submissions with states 'in progress' or 'ready.' This feature provides reviewers advanced access to the data so that inspection of the data can be part of the peer-review process. Normally, submissions for manuscripts that are published must also be made public; however, embargo periods are allowed subject to the policies of the journal.

Analysis—Consumers of TreeBASE data can download digital serializations of phylogenetic data in NEXUS or NeXML (Vos et al., 2012) formats, and then reanalyze, augment, or otherwise repurpose the data on their own local computers. Programmatic access to data is available using the PhyloWS API standard (evoinfo.nescent.org/PhyloWS), and interaction with this API can be found in Bio:Phylo libraries for PERL (Vos et al., 2011) and libraries for R (e.g., www.carlboettiger.info/archives/3019). Interfacing with the database allows users to perform meta-analyses regarding broad-scale patterns in published phylogenetic results (e.g., Piel et al., 2003) or, for example, for performing auto-assembly of supertrees. Instruction in the practical use of this API is provided by the summer course, 'Computational Phyloinformatics' (academy.nescent.org/wiki/Computational_phyloinformatics).

Publications—Although TreeBASE welcomes phylogenetic data for all taxa, extant species are dominant because of the availability of DNA sequences. Examples of studies in TreeBASE that are pertinent to vertebrate paleontology include those relevant to the controversies regarding the placement of cetaceans relative to bovines (e.g., studies by J. Gatesy, J. Geisler, J. Gittleman, R. W. Meredith, M. O'Leary, M. Springer, J. Thewissen, etc.). It is hoped that more paleontologists will adopt TreeBASE as a resource for sharing morphological data sets and phylogenies involving extinct taxa.

SUMMARY AND CONCLUSIONS

Science, and society as a whole, has now permanently settled the new frontier of the 'Information Age,' where integration of vast amounts of data routinely leads to new discoveries. Correspondingly, paleoinformatics is developing as a vibrant subfield of paleontology. Among key insights that have already emerged through research using the databases mentioned above are how biodiversity and taxonomic patterns vary through time (Alroy, 2002, 2008), both in the marine and terrestrial realms (Alroy et al., 2001); how to establish ecological baselines in the context of defining conservation targets (Carrasco et al., 2009; Hadly and Barnosky, 2009; Stegner and Holmes, in press); calibrating extinction trajectories and their implications for conservation (Barnosky et al., 2011a, 2011b); and understanding links between climate change and community composition (Graham et al., 1996) or ecomorphology (Fortelius et al., 2002; Polly and Eronen, 2011).

In addition to research contributions, the individual databases on which paleoinformatics relies are now serving as data repositories for information acquired through public funding. For example, the U.S. NSF recently implemented requirements that their funded research projects have a Data Management Plan to allow preservation and access to key data. For fossil-related data, the logical repositories are paleontological databases, particularly those originally funded by the NSF. Currently, the NSF Earth Sciences (EAR) policy encourages, but does not require, deposition of data generated from grants in a public repository. The United Kingdom research councils and public funding bodies in other countries also have similar policies.

Online archiving of data offers numerous benefits toward fulfilling data archiving mandates but also expands research opportunities by (1) making data easily available for other researchers to inspect or add into additional analyses; and (2) making data accessible (potentially) in perpetuity, such that future generations can rely and build upon an ever-growing foundation of knowledge. Of course, these same points could be made about traditional paper archiving, but the great improvement offered by databases comes through combining ease of access with the breadth of available data, which in turn makes practical new types of analyses that can address ever-more complex questions.

The National Science Board (NSB) extensively documented the status of NSF policy regarding long-lived data collections and highlighted several important issues regarding how NSF funds and cares for these collections in the long-term (National Science Board, 2005). In the same year, H. Richard Lane of the NSF discussed many of the practical issues being tackled by the researchers creating large databases in paleontology (Lane, 2005). The NSB highlighted that (1) the NSF had many separate data-archiving policies, but no consistent, agency-wide policy; and (2) the NSF was initiating many long-lived databases via principle-investigator-led grant projects, without the policies or infrastructure to ensure that these databases would remain available, much less interactive (National Science Board, 2005). Lane (2005) listed several projects that were being funded or considered by the NSF in 2005, including several that are still active (PaleoDB, MIOMAP, Paleontology Portal) and many that got off to a good start, but are now seemingly inactive (Paleostrat, SESAR, EARTHTIME). These latter are, at least in part, victims of the lack of a long-term strategy for data preservation by the geoscience community. The demonstrated feasibility of linking freestanding data sets through common portals may offer increasingly efficient mechanisms to keep existing databases widely available through distributed networks, while simultaneously providing the ability to control content and quality locally.

Current work also demonstrates the feasibility of creating a clearinghouse or portal that encompasses a much wider range of paleontological data than is currently the case, such as information about all fossil specimens collected with public funds (akin to GBIF). Linking with a public repository (akin to PubMed Central) for research papers that report on analyses of paleontological databases (and the specimen information contained within them) also is technologically feasible, and is an especially relevant need with respect to public-funded data and research.

ACKNOWLEDGMENTS

We would first like to thank all of the vertebrate paleontologists who have painstakingly unearthed, prepared, studied, and described fossils. Their work have always been, and always will be, the foundation on which our science is based. Without them, we would have no taxa in our character matrices, no occurrence data for our diversity curves, nor any paleoecological data to interpret past climates. Second, we would like to thank all of the people who have already begun to deposit their current data in online databases. Your actions in this regard are both fruitful to your own research program, but also benefit the community greatly by providing easy access to your data for further analysis and research. Finally, thanks to all of those who keystroke legacy data into databases, giving us all new access to old data that may have been temporarily lost from field through neglect. The PIs of Morphbank gratefully acknowledge funding from the US National Science Foundation for its development (awards 0446224 to A.M., G.E., and colleagues and award 0851313 to A.M. and colleagues).

LITERATURE CITED

Agustí, J., L. Cabrera, M. Garcés, W. Krijgsman, O. Oms, and J. M. Parés. 2001. A calibrated mammal scale for the Neogene of Western Europe. State of the art. Earth Science Reviews 52:247–260.

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. Palaeogeography, Palaeoclimatology, Palaeoecology 127:285–311.

Alroy, J. 2000. Frequently Asked Questions. Paleobiology Database. Available at http://paleodb.science.mq.edu.au/cgi-bin/bridge.pl?a=displayPage&page=paleodbFAQ. Accessed February 1, 2012.

Alroy, J. 2001. A multispecies overkill simulation of the End-Pleistocene megafaunal mass extinction. Science 292:1893–1896.

Alroy, J. 2002. How many named species are valid? Proceedings of the National Academy of Sciences of the United States of America 99:3706–3711.

Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. Proceedings of the National Academy of Sciences of the United States of America 105:11536–11542.

Alroy, J., C. R. Marshall, R. K. Bambach, K. Bezosko, M. Foote, F. T. Fürsich, T. A. Hansen, S. M. Holland, L. C. Ivany, D. Jablonski, D. K. Jacobs, D. C. Jones, M. A. Kosnik, S. Lidgard, S. Low, A. I. Miller, P. M. Novack-Gottshall, T. D. Olszewski, M. E. Patzkowsky, D. M. Raup, K. Roy, J. J. Sepkoski Jr., M. G. Sommers, P. J. Wagner, and A. Webber. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. Proceedings of the National Academy of Sciences of the United States of America 98:6261–6266.

Anderson, N. R., P. Tarczy-Hornoch, and R. E. Bumgarner. 2006. On the persistence of supplementary resources in biomedical publications. BMC Bioinformatics 7:1–7.

Barnosky, A. D. 2008. Megafauna biomass tradeoff as a driver of Quaternary and future extinctions. Proceedings of the National Academy of Sciences of the United States of America 105:11543–11548.

Barnosky, A. D. 2009. Heatstroke: Nature in the Age of Global Warming. Island Press, Washington, D.C., 269 pp.

Barnosky, A. D., and B. P. Kraatz. 2007. The role of climatic change in the evolution of mammals. BioScience 57:523–532.

Barnosky, A. D., M. A. Carrasco, and R. W. Graham. 2011. Collateral mammal diversity loss associated with late Quaternary megafaunal extinctions and implications for the future; pp. 179–189 in A. McGowan, and A. B. Smith (eds.), Comparing the Geological and Fossil Records: Implications for Biodiversity Studies.

Barnosky, A. D., F. Bibi, S. S. B. Hopkins, and R. Nichols. 2007. Biostratigraphy and magnetostratigraphy of the mid-Miocene Railroad Canyon sequence, Montana and Idaho, and age of the mid-Tertiary unconformity west of the continental divide. Journal of Vertebrate Paleontology 27:204–224.

Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. R. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, and E. A. Ferrer. 2011b. Has the Earth's sixth mass extinction already arrived? Nature 417:51–57.

Benton, M. J. 1993. The Fossil Record 2. Chapman & Hall, London, 845 pp.

Benton, M. J. 1999. The history of life: large databases in palaeontology; pp. 249–283 in D. A. T. Harper (ed.), Numerical Palaeobiology. John Wiley and Sons, New York.

Bernor, R. L., V. Fahlbusch, P. Andrews, H. Bruijn, M. de. Fortelius, F. Rögl, F. F. Steininger, and L. Werdelin. 1996. The evolution of western Eurasian Neogene mammal faunas: a chronologic, systematic, biogeographic, and paleoenvironmental synthesis; pp. 449–470 in R. L. Bernor, V. Fahlbusch, and W. Mittmann (eds.), The Evolution of Western Eurasian Neogene Mammal Faunas. Columbia University Press, New York.

Brown, R. S., and R. C. Denicola. 1998. Cases on Copyright: Unfair Competition, and Related Topics Bearing on the Protection of Literary, Musical, and Artistic Works, seventh edition. Foundation Press, New York, 828 pp.

Butler, R. J., R. B. J. Benson, M. T. Carrano, and P. D. Mannion. 2011. Sea-level, dinosaur diversity, and sampling biases: investigating the 'common cause' hypothesis in the terrestrial realm. Proceedings of the Royal Society B 278:1165–1170.

Carrasco, M. A., A. D. Barnosky, and R. W. Graham. 2009. Quantifying the extent of North American mammal extinction relative to the pre-anthropogenic baseline. PLoS ONE 4:e8331.

Carrasco, M. A., A. D. Barnosky, B. P. Kraatz, and E. B. Davis. 2007. The Miocene Mammal Mapping Project (MIOMAP); an online database of Arikareean through Hemphillian fossil mammals. Bulletin of Carnegie Museum of Natural History 39:183–188.

Carroll, R. L. 1988. Vertebrate Paleontology and Evolution. W. H. Freeman and Company, New York, 698 pp.

Claeson, K. M., M. A. O'Leary, E. M. Roberts, F. Sissoko, M. L. Bouaré, L. Tapanila, D. Goodwin, and M. D. Gottfried. 2010. First Mesozoic record of the stingray *Myliobatis wurnoensis* from the Late Cretaceous of Mali and a phylogenetic analysis of Myliobatidae (Batoidea) incorporating dental characters. Acta Palaeontological Polonica 55:655–674.

Cuvier, G. 1812. Recherches sur les Ossements Fossiles de Quadrupèdes. Chez Deterville, Paris.

Dolven, J. K., and H. Skjerpen. 2011. Paleoinformatics: past, present and future perspectives; pp. 45–52 in A. M. T. Elewa (ed.), Computational Paleontology. Springer-Verlag, Berlin.

Ekdale, E. G., A. Berta, and T. A. Deméré. 2011. The comparative osteology of the petrotympanic complex (ear region) of extant baleen whales (Cetacea: Mysticeti). PLoS ONE 6:e21311.

Erickson, G. M., K. Curry Rogers, D. D. Varricchio, M. A. Norell, and X. Xu. 2007. Growth patterns in brooding dinosaurs reveals the timing of sexual maturity in non-avian dinosaurs and genesis of the avian condition. Biology Letters 3:558–561.

Erickson, G. M., P. J. Makovicky, P. J. Currie, M. A. Norell, S. A. Yerby, and C. A. Brochu. 2004. Gigantism and comparative life-history parameters of tyrannosaurid dinosaurus. Nature 430:772–775.

Eronen, J. T., M. Mirzaie Ataabadi, A. Micheels, A. Karme, R. L. Bernor, and M. Fortelius. 2009. Distribution history and climatic controls of the Late Miocene Pikermian chronofauna. Proceedings of the National Academy of Sciences of the United States of America 106:11867–11871.

Eronen, J. T., K. Puolamäki, L. Liu, K. Lintulaakso, J. Damuth, C. Janis, and M. Fortelius. 2010. Precipitation and large herbivorous mammals, part II: application to fossil data. Evolutionary Ecology Research 12:235–248.

Evans, A. R., G. P. Wilson, M. Fortelius, and J. Jernvall. 2007. High-level similarity of dentitions in carnivorans and rodents. Nature 445:78–81.

Fara, E. 2004. Estimating minimum global species diversity for groups with a poor fossil record: a case study of Late Jurassic-Eocene lissamphibians. Palaeogeography Palaeoclimatology Palaeoecology 207:59–82.

Fontal-Cazalla, F. M., M. L. Buffington, G. Nordlander, J. Liljeblad, P. Ros-Farré, J. L. Nieves-Aldrey, J. Pujade-Villar, and F. Ronquist. 2002. Phylogeny of the Eucoilinae (Hymenoptera: Cynipoidea: Figitidae). Cladistics 18:154–199.

Fortelius, M., J. T. Eronen, J. Jernvall, L. Liu, D. Pushkina, J. Rinne, A. Tesakov, I. Vislobokova, Z. Zhang, and L. Zhou. 2002. Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. Evolutionary Ecology Research 4:1005–1016.

Graham, R. W., H. A. Semken Jr., and M. A. Graham (eds.). 1987. Late Quaternary Mammalian Biogeography and Environments of the Great Plains and Prairies, Volume 22. Illinois State Museum, Springfield, Illinois, 491 pp.

Graham, R. W., E. L. Lundelius Jr., E. Anderson, A. D. Barnosky, J. A. Burns, C. S. Churcher, D. K. Grayson, R. D. Guthrie, C. R. Harington, G. T. Jefferson, L. D. Martin, H. G. McDonald, R. E. Morlan, H. A. Semken, S. D. Webb, and M. C. Wilson. 1994. FAUNMAP: a database documenting late Quaternary distributions of mammal species in the United States. Illinois State Museum Scientific Papers 25:1–690.

Graham, R. W., E. L. Lundelius Jr., M. A. Graham, E. K. Schroeder, R. S. Toomey, E. Anderson, A. D. Barnosky, J. A. Burns, C. S. Churcher, D. K. Grayson, R. D. Guthrie, C. R. Harington, G. T. Jefferson, L. D. Martin, H. G. McDonald, R. E. Morlan, H. A. Semken, S. D. Webb, L. Werdelin, and M. C. Wilson. 1996. Spatial response of mammals to late Quaternary environmental fluctuations. Science 272:1601–1606.

Gunnell, G. F., and N. B. Simmons. 2005. Fossil evidence and the origin of bats. Journal of Mammalian Evolution 12:209–246.

Hadly, E. A., and A. D. Barnosky. 2009. Vertebrate fossils and the future of conservation biology. Conservation Paleobiology: Using the Past to Manage for the Future. Dietl, G. P., and Flessa, K. W. (eds.). Paleontological Society Papers 15:39–59.

Harland, W. B., C. H. Holland, M. R. House, A. B. Reynolds, M. J. S. Rudwick, G. E. Satterthwaite, L. B. H. Tarlo, and E. C. Willey (eds.). 1967. The Fossil Record: A Symposium with Documentation. Geological Society of London, London, 827 pp.

Hay, O. P. 1902. Bibliography and Catalogue of the Fossil Vertebrata of North America. Government Printing Office, Washington D.C., 868 pp.

Hill, R. V. 2005. Integration of morphological data sets for phylogenetic analysis of Amniota: the importance of integumentary characters and increased taxonomic sampling. Systematic Biology 54:530–547.

Janis, C. M., G. F. Gunnell, and M. D. Uhen (eds.). 2008. Evolution of Tertiary Mammals of North America, Volume II. Cambridge University Press, Cambridge, U.K., 795 pp.

Janis, C. M., K. M. Scott, and L. L. Jacobs (eds.). 1998. Evolution of Tertiary Mammals of North America, Volume I. Cambridge University Press, Cambridge, U.K., 691 pp.

Jernvall, J., and M. Fortelius. 2002. Common mammals drive the evolutionary increase of hypsodonty in the Neogene. Nature 417:538–540.

Jernvall, J., and M. Fortelius. 2004. Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. American Naturalist 164:614–624.

Jernvall, J., J. P. Hunter, and M. Fortelius. 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiations. Science 274:1489–1492.

Johnson, K. G., S. J. Brooks, P. B. Fenberg, A. G. Glover, K. E. James, A. M. Lister, E. Michel, M. Spencer, J. A. Todd, E. Valsami-Jones, J. Young, and J. R. Stewart. 2011. Climate change and biosphere response: unlocking the collections vault. BioScience 61:147–153.

Lane, H. R. 2005. Paleoinformatics at the National Science Foundation: IT or not IT? American Paleontologist 13:22–26.

Leidy, J. 1858. [*Hadrosaurus foulkii*, a new saurian from the Cretaceous of New Jersey, related to *Iguanodon*]. Proceedings of the Academy of Natural Sciences of Philadelphia 10:213–218.

Liow, L. H., M. Fortelius, E. Bingham, K. Lintulaakso, H. Manilla, L. Flynn, and N. C. Stenseth. 2008. Higher origination and extinction rates in larger mammals. Proceedings of the National Academy of Sciences of the United States of America 105:6097–6102.

Liu, X. Y., and D. Yang. 2006. Revision of the fishfly genus *Ctenochauliodes* van der Weele (Megaloptera, Corydalidae). Zoologica Scripta 35:473–490.

MacLeod, N., and R. Guralnick. 2000. Paleoinformatics; pp. 31–36 in R. H. Lane, F. F. Steininger, R. L. Kaesler, W. Ziegler, and J. H. Lipps (eds.), Fossils and the Future: Paleontology in the 21st Century. E. Schweizerbart, Frankfurt.

MacLeod, N., P. Diver, R. Guralnick, D. B. Lazarus, and B. Malmgren. 2000. Computers, quantification, and databases; pp. 191–201 in R. H. Lane, F. F. Steininger, R. L. Kaesler, W. Ziegler, and J. H. Lipps (eds.), Fossils and the Future: Paleontology in the 21st Century. E. Schweizerbart, Frankfurt.

Maddison, D. R., D. L. Swofford, and W. P. Maddison. 1997. NEXUS: an extensible file format for scientific information. Systematic Biology 46:590–621.

Marx, F. G., and M. D. Uhen. 2010. Climate, critters, and cetaceans: Cenozoic drivers of the evolution of modern whales. Science 327:993–996.

Miller, S. A. 1889. North American Geology and Paleontology. Western Methodist Book Concern, Cincinnati, Ohio, 718 pp.

Morris, J. 1854. A Catalogue of British Fossils: Comprising the Genera and Species Hitherto Described, second edition. Published by the author, London, 372 pp.

National Science Board. 2005. Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. National Science Board Report 05-40, 87 pp.

Nesbitt, S. J., D. T. Ksepka, and J. A. Clarke. 2011. Podargiform affinities of the enigmatic *Fluvioviridavis platyrhamphus* and the early diversification of Strisores (“Caprimulgiformes” + Apodiformes). PLoS ONE 6:e7062.

Newell, N. D. 1952. Periodicity in invertebrate evolution. Journal of Paleontology 26:371–385.

O’Leary, M. A., and S. A. Kauffman. 2011. MorphoBank: phylophenomics in the “cloud.” Cladistics 27:1–9.

Phillips, J. 1860. Life on the Earth Its Origin and Succession. MacMilland and Co., London, 224 pp.

Piel, W. H., M. J. Sanderson, and M. J. Donoghue. 2003. The small-world dynamics of tree networks and data mining in phyloinformatics. Bioinformatics 19:1162–1168.

Piel, W. H., L. Chan, M. Dominus, J. Ruan, R. A. Vos, and V. Tannen. 2009. TreeBASE v. 2: a database of phylogenetic knowledge. e-BioSphere 09:216–217.

Pljusnin, I., A. R. Evans, A. Karme, A. Gionis, and J. Jernvall. 2008. Automated 3D phenotype analysis using data mining. PLoS ONE 3:e1742.

Polly, P. D., and J. T. Eronen. 2011. Mammal associations in the Pleistocene of Britain: implications of ecological nich modeling and a method for reconstructing palaeoclimate; pp. 279–304 in N. Ashton, S. G. Lewis, and C. B. Stringer (eds.), The Ancient Human Occupation of Britain. Elsevier, Amsterdam.

Polly, P. D., and C. B. Stringer. 2011. The Ancient Human Occupation of Britain (AHOB) Database. Pleistocene Databases: Acquisition, Storing, Sharing, R., Macchiarelli, and G.-C., Weniger (eds.). Wissenschaftliche Schriften des Neanderthal Museums 4:51–60.

Prieto-Márquez, A. 2008. Phylogeny and historical biogeography of hadrosaurid dinosaurs. Ph.D. dissertation, Florida State University, Tallahassee, Florida, 934 pp.

Prieto-Márquez, A., D. B. Weishampel, and J. R. Horner. 2006. The dinosaur *Hadrosaurus foulkii*, from the Campanian of the East Coast of North America, with a reevaluation of the genus. Acta Palaeontologica Polonica 51:77–98.

Prieto-Márquez, A., G. M. Erickson, K. Seltmann, F. Ronquist, G. A. Riccardi, C. Maneva-Jakimoska, N. Jammigumpula, A. Mast, S. Winner, W. Blanco, A. Deans, D. Gaitros, D. Paul, and C. Gaitros. 2007. Morphbank, an avenue to document and disseminate anatomical data: phylogenetic and paleohistological test cases. Journal of Morphology 268:1120.

Puolamäki, K., M. Fortelius, and H. Mannila. 2006. Seriation in paleontological data using Markov Chain Monte Carlo methods. PLoS Computational Biology 2:e6.

Raia, P., F. Carotenuto, J. T. Eronen, and M. Fortelius. 2011. Longer in tooth, shorter in the record? The evolutionary correlates of hypsodonty in Neogene ruminants. Proceedings of the Royal Society B 278:3474–3481.

Raia, P., F. Carotenuto, D. Passaro, D. Fulgione, and M. Fortelius. 2012. Ecological specialization in fossil mammals explains Cope’s rule. American Naturalist 179:328–337.

Sanderson, M. J., B. G. Baldwin, G. Bharathan, C. S. Campbell, D. Ferguson, J. M. Porter, C. Von Dohlen, M. F. Wojciechowski, and M. J. Donoghue. 1993. The growth of phylogenetic information and the need for a phylogenetic database. Systematic Biology 42:562–568.

Schwarzer, M. 2012. Riches, Rivals, and Radicals: 100 Years of Museums in America. American Association of Musuems, Washington, D.C., 263 pp.

Sepkoski, J. J., Jr. 1982. A compendium of fossil marine families. Milwaukee Public Museum, Contributions in Biology and Geology 51:1–125.

Sepkoski, J. J., Jr. 2002. A Compendium of Fossil Marine Animal Genera. Paleontology Research Institution, Ithaca, New York, 500 pp.

Shimer, H. W., and R. R. Shrock. 1944. Index Fossils of North America. J. Wiley & Sons, New York, 837 pp.

Stegner, M. A., and M. Holmes. In press. Using paleontological data to assess mammalian community structure: potential aid in conservation planning. Palaeogeography Palaeoclimatology Palaeoecology.

Stringer, C. B. 2006. Homo britannicus. Penguin, London, 242 pp.

Stringer, C. B. 2011. The changing landscapes of the earliest human occupation of Britain and Europe; pp. 1–10 in N. M. Ashton, S. G. Lewis, and C. B. Stringer (eds.), The Ancient Human Occupation of Britain. Elsevier, Amsterdam.

Tang, H., A. Micheels, J. T. Eronen, and M. Fortelius. 2011. A regional climate model experiment to investigate the Asian monsoon in the Late Miocene. Climate of the Past 7:847–868.

Uhen, M. D. 2000. The paleontological legacy of Eckert and Mauchly. Paleobiology 26:310–315.

Ukkonen, A., M. Fortelius, and H. Mannila. 2005. Finding partial orders from unordered 0–1 data; pp. 285–293 in R. Grossman, R. Bayardo, and K. P. Bennett (eds.), Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, 21–24 August 2005.

Vos, R. A., J. A. Caravas, K. Hartmann, M. A. Jensen, and C. Miller. 2011. Bio::Phylo-Phyloinformatic analysis using PERL. BMC Bioinformatics 12:63.

Vos, R. A., J. P. Balhoff, J. A. Caravas, M. T. Holder, H. Lapp, P. E. Midford, A. Piriyam, J. Sukumaran, X. Xia, and A. Stoltzfus. 2012. NeXML: rich, extensible, and verifiable representation of comparative data and metadata. Systematic Biology.

Werdelin, L., and W. S. Sanders (eds.). 2010. Cenozoic African Mammals. University of California Press, Berkeley, California, 986 pp.

Submitted May 7, 2012; revisions received July 18, 2012;
accepted July 19, 2012.
Handling editor: Blaire Van Valkenburgh.