



ONLINE ARTICLE

Missing “Links” in Bioinformatics Education:

Expanding Students’ Conceptions of Bioinformatics
Using a Biodiversity Database of Living & Fossil Reef Corals

ROSS H. NEHM ANN F. BUDD

Bioinformatics is a new and integrative branch of science bounded by the intersection of biology, computer science, and information technology (NCBI, 2004). Two stars shine brightly in the bioinformation constellation: the Human Genome Project (HGP) and Genbank (NCBI, 2004). The HGP, which has now completed a working draft of the genome of our species, strives to understand the basic genetic instructions that help construct the human phenotype (Buxeda & Moore-Russo, 2003). The Genbank database is designed to provide and encourage public access to the most up-to-date and comprehensive DNA sequence information from the HGP and other genetic research projects (Palladino, 2002). With the scientific and educational spotlights so frequently focused on the HGP and Genbank (e.g., Smith & Emmeluth, 2002), biology teachers may not have noticed the bioinformation revolution in other areas of biology (e.g., Bisby, 2000; Edwards et al., 2000; Budd et al., 2001). Here we introduce the expansive realm of bioinformatics beyond genetics and provide a series of lessons that employ a bioinformatics database to test hypotheses about biodiversity change over millions of years.

Ross H. NEHM, Ph.D., is an Assistant Professor in the Departments of Secondary Education and Biology at The City College, City University of New York, New York, NY, 10031; email: rnehm@ccny.cuny.edu. He studies the evolution and development of marine mollusks and student learning of evolution. ANN F. BUDD, Ph.D., is Professor and Chair of the Department of Geoscience at the University of Iowa, Iowa City, IA; email: ann-budd@uiowa.edu. She is an expert on coral evolution and Co-Editor of the Journal of Paleontology.

Missing “Links” in Bioinformatics

As several *ABT* articles about bioinformatics have discussed (e.g., Jegalian, 2000; Bloom, 2001; Smith & Emmeluth, 2002), the magnitude of biological information has exploded over the past decade as a result of technological advances (e.g., PCR, automated sequencing) that allow scientists to gather data more efficiently and effectively. These and many other articles imply that the staggering volume of biological data is restricted to genes and proteins, but this is only part of the story. As we know, biological information extends far beyond genetics, and new tools and technologies are revolutionizing how other forms of biological information are captured, stored, and studied. Technological advances in biological imaging, morphometrics (the quantitative analysis of form), biodiversity surveys using GPS and remote sensing, and digital microscopy have also contributed to the flood of new information, changing our understanding of biology from the level of cells to ecosystems. However, bioinformatics has often been defined as the “acquisition, storage, analysis, modeling and distribution of information embedded in DNA and protein sequence data” (Rowen et al., 1997). Should the “bio” in bioinformatics stand for biology, or genetics alone? We answer this question below.

Bioinformatics & Biodiversity

A new and pertinent area of bioinformatics is the development of databases that help identify, monitor, and study spatial and temporal biodiversity patterns (Sugden & Pennisi, 2000; Pennisi, 2000; Bisby, 2000; Edwards et

Table 1. Biodiversity databases on the Internet.

DATABASE	URL	TOPIC
Species 2000	http://www.sp2000.org/	Global species database
Global Biodiversity Information Facility	http://www.gbif.org/	Data portal to biodiversity
Tree of Life Project	http://tolweb.org/tree/phylogeny.html	Phylogeny of life
Integrated Taxonomic Information System	http://www.itis.usda.gov/	North American species
USDA Plants database	http://plants.usda.gov/	Plant diversity
National Biodiversity Institute	http://www.inbio.ac.cr/en/default.html	Costa Rica Biodiversity
The Diptera Site	http://www.sel.barc.usda.gov/diptera/biosys.htm	USDA fly systematics
Fishbase	http://www.fishbase.org/search.cfm	Fish database
Cephbase	http://www.cephbase.utmb.edu/	Cephalopod mollusk database
Census of Marine Life	http://www.coml.org/coml.htm	International database of marine life
Biogeoinformatics of Hexacorals	http://www.kgs.ku.edu/Hexacoral/	Coral database

al., 2000). Museum card boxes and paper lists of species, habitats, and distributions have been replaced by internationally accessible interactive digital databases (e.g., Species 2000; see Table 1). Biodiversity databases typically focus on specific taxonomic groups (e.g., the USDA Plants database), geographic regions (e.g., the Costa Rican Biodiversity Institute) or methods for studying biodiversity data (e.g., the WorldMap database). These databases are facilitating more rapid identification of new species as well as range distribution changes in previously identified species. Biodiversity bioinformatics is central to global conservation efforts. One overall goal of digital diversity projects is to establish an online record of all living species on Earth (Wilson, 2000). This goal remains elusive, however, because millions of species have yet to be discovered. Compounding this problem has been a gradual decrease in funding for basic taxonomic and systematic research over the past few decades (Wilson, 2000).

Bioinformatics & Biology Education

Concordant with the explosion of biological data, improvements in Internet access and analytical tools now provide easy access to biological information for teachers and students alike worldwide. Several curricula have been developed for teaching students about large-scale scientific endeavors and databases such as the HGP and Genbank (e.g., BSCS, 2004). Although teachers, students, and the general public should learn how to access, use and interpret genetic data, we must also provide students with a more comprehensive and balanced picture of advances in biological information tools and resources. Students must be brought up to speed on how computers, the Internet, and global information databases have revolutionized many fields of biology including ecology, oceanography, and evolutionary biology, *in addition to* fields commonly associated with the more restricted concept of “bioinformatics” (e.g., molecular medicine, forensics, and pharmacology). We argue for a more expanded definition of bioinformatics and introduce

a source of bioinformation that is a central component of biodiversity and evolution education: the fossil record.

Paleontology Has Evolved: Paleobiological Data in the Digital Age

Paleontology and its evolutionarily focused subfield, paleobiology, have long provided a useful pedagogical “entry point” for biology teachers when introducing standard curriculum topics (e.g., evolution, extinction, biodiversity, fossils) because the subject is captivating and accessible to most students. Often students are engaged by paleontologists’ field expeditions to remote lands that include collecting and mapping never-before collected—and often bizarre—fossils of extinct organisms. Although fieldwork is still central to the work of many paleobiologists, technological and computer-based approaches are revolutionizing paleobiological and evolutionary research (Kaiser, 2000; Budd et al., 2001). As in genetics and many other fields of biology, paleobiological data have grown dramatically over the past decade as a result of technological advances. Likewise, Internet databases that combine precise morphologic, microstructural, geographic, and phylogenetic information while performing real-time analyses are becoming standard research tools in paleobiology. Many paleobiological databases exist and include data ranging from deep-sea microorganisms to land mammals. A list of several of these databases is shown in Table 2.

Unfortunately, these new databases and analytical tools have yet to be used in the vast majority of secondary science classrooms, where paleontology is sometimes ignored altogether or characterized as an antiquated and qualitative discipline riddled with incomplete data and uncertain results. Paleobiology has become data-rich and high tech; online databases can provide students with a more complete and accurate picture of this discipline. The revolution in bioinformation extends to investigations of ancient life and the Earth’s history.

Table 2. Paleobiological databases on the Internet.

DATABASE	URL	TOPIC
Alaska Paleontological Database	http://www.alaskafossil.org/	Alaskan paleobiology
Evolution of Terrestrial Ecosystems	http://www.nmnh.si.edu/ete/	Land ecosystem evolutionary data
Faunmap	http://museum.state.il.us/research/faunmap/	Mapping database of land animals
Fossil Record 2	http://palaeo.gly.bris.ac.uk/frwhole/FR2.html	General fossil record data
Global Pollen Database	http://www.ngdc.noaa.gov/paleo/gpd.html	Data on fossil pollen through time
Miomap	http://www.ucmp.berkeley.edu/miomap/miomap_home_page.htm	Miocene age vertebrates
Neogene of the Old World (NOW)	http://www.helsinki.fi/science/now/	Neogene age fossils
Ocean Drilling Program Data and Samples	http://www.odsn.de/	Distributions of fossils from ocean cores
Orbitolinid Foraminifera	http://www.paleotax.de/orbitos/index.htm	Foraminifera fossils
Paleobank	http://paleo.ku.edu/paleobank.html	General fossil record data
Paleobiology Database	http://paleodb.org/cgi-bin/bridge.pl	Data on many fossil groups through time
WDC for Paleoclimatology	http://www.ngdc.noaa.gov/paleo/data.html	Data on past climates

Bioinformatics & Evolution Education

Evolution is a central, unifying, and integrative subject that makes use of a broad array of scientific data from fields ranging from geochemistry to molecular biology. Bioinformatics is one of several appropriate arenas for introducing the fossil record in its contemporary state to biology students because it covers many of the *National Science Education Standards* (National Academy of Sciences, 1996), including evolution, ecology, extinction, geological change, and technology. Paleobiological data can be compared and contrasted with genetic data in order to provide students with a broader introduction to the new field of bioinformatics.

One of the authors teaches a six-week module on bioinformatics to high school biology students who use both Genbank and a biodiversity bioinformatics database known as NMITA (Neogene Marine bIota of Tropical America: <http://nmita.geology.uiowa.edu/>) to test biological hypotheses and explore biodiversity change through geological time. Here we introduce this biodiversity bioinformatics database as an educational tool and provide a series of lessons that can be used in high school or college undergraduate classrooms to complement Genbank lessons (e.g., BSCS, 2004). The goal of these lessons is to provide students with a more comprehensive view of how databases are revolutionizing biology. NMITA and Genbank are very different databases; nevertheless, there are similarities in their structures and uses that provide common themes for introducing students to bioinformatics. In order to place NMITA within the conceptual array of biological databases, we provide a chart that compares and contrasts NMITA's paleobiological data relating to biodiversity to Genbank's data relating to genetics (Table 3). This has helped our high school biology students achieve a better understanding of the range of biological information that can be stored and studied using online databases.

NMITA: An Introduction

The Neogene Marine bIota of Tropical America (NMITA) Web site and database was established to provide global access to tropical American marine biodiversity data from the past 25 million years (Budd et al., 2001). These data are central to understanding current extinctions of Caribbean coral reefs and other invertebrates (see the cover photo of the journal *Science*, 9/15, 2003) as well as evolution in marine ecosystems (Nehm, 2001). NMITA contains over 4000 scientific images of more than 1000 taxa, as well as data and analysis tools. The majority of data were collected as part of two large-scale and interdisciplinary fossil sampling programs: the Panama Paleontology Project (PPP) coordinated by the Smithsonian Tropical Research Institute in Panama (Collins, 2005) and the Dominican Republic Project, formerly coordinated by the Natural History Museum in Basel, Switzerland (see Nehm, 2004). Fieldwork has thus far produced more than a million fossil specimens that have been used in many evolutionary studies (e.g., Jackson, Budd & Coates, 1996; Nehm, 2001). NMITA is designed to provide access for researchers, teachers, and students who are interested in systematics, stratigraphy, biodiversity, and evolutionary biology.

NMITA contains high-resolution digital images of hundreds of species of invertebrate and vertebrate animals from different periods of time and different localities in tropical America. These invertebrates include bryozoans, corals (zooxanthellate and azooxanthellate), mollusks (gastropods and bivalves), and arthropods (ostracodes). There are two main avenues for accessing information in NMITA: searching by biological group, or searching by locality/time. Users may click on individual species to receive standard taxonomic information (phylum, class, species, etc), synonyms (equivalent biological names for species), morphology (size, shape, distinguishing characteristics), and spatial and temporal distributions (localities and samples that the species occurs in through geological time). Alternatively, users may look for

Table 3. A comparison of NCBI (Genbank) and NMITA.

CATEGORY	NMITA	GENBANK
Biological scale	ORGANISMS Microscopic (tissues) and Macroscopic (animals)	PARTS OF ORGANISMS Microscopic (e.g., proteins, genes, chromosomes)
Research scale	Small (~30 scientists)	Large (thousands of scientists)
Units of study	Bags of sediment with fossils or living animals	Tissue from living organisms
Specimens	Tissues, Organs, Organisms, Species, Communities of fossil and living animals	Proteins, Genes, Chromosomes from living species
Methods for obtaining data	Field work, systematics, dating, thin sectioning, measurement, X-ray, etc.	Field work, systematics, tissue extraction, DNA/protein extraction, PCR, sequencing, etc.
Data	Taxon, geography, time, morphology, chemical structure, geochemistry	Taxon, chromosomes, DNA, proteins
Data type	Geographic, stratigraphic, taxonomic, morphological, systematic	Biochemical
Data form	Measurements, images of species	Molecular and biochemical structure
Taxonomic keys	Present	Absent
Spatial information	Localities on Earth	Locations on chromosomes or in cells
Temporal coverage	20 Million years (Miocene-Today)	Today
Data for evolutionary studies?	YES	YES
Data for biodiversity studies?	YES	LIMITED

information by clicking on maps and stratigraphic columns (rock layers) to gather data on the spatial and temporal distributions of species and their ecological associations.

The fossil collections digitally housed by NMITA are physically stored at the U.S. National Museum of Natural History (NMNH) of the Smithsonian Institution in Washington, DC and the Natural History Museum in Basel, Switzerland (NMB). Thus, NMITA stores specimen data in a similar way to how Genbank stores genetic data.

NMITA Computer Laboratory Lessons

We developed four inquiry-based biodiversity bioinformatics lessons for high school students and college non-majors using NMITA. These lessons, as well as the data in NMITA, are freely available to other teachers and researchers at <http://www.dominicanrepublicproject.org/Multimedia/Lessons/1.htm>. The overall purpose of these lessons is to provide students with working knowledge of how biodiversity bioinformatics databases are used by biologists to store data and retrieve data to test biological hypotheses. A spreadsheet of the *National Science Education Standards* that the lessons target may be downloaded at <http://www.dominicanrepublicproject.org/Multimedia/Lessons/1/nmita-national.pdf>.

Table 4. NMITA lessons and hypotheses that can be tested.

LESSON	POSSIBLE HYPOTHESES TO TEST
2	Modern reef coral diversity is greater than in all fossil samples. Reef coral biodiversity has decreased over the past 6 million years. Reef coral biodiversity has stayed the same over the past 6 million years.
3	The percentage of branching coral morphotypes is the same today as in the fossil record. Branching morphotypes are the most common reef corals in the Pliocene. All samples have similar percentages of branching morphotypes.
4	Most modern reef coral species first originated in the Pleistocene. Most modern reef coral species first originated in the Miocene. Patterns of generic and species-level diversity over time are equivalent.

[republicproject.org/Multimedia/Lessons/1/nmita-national.pdf](http://www.dominicanrepublicproject.org/Multimedia/Lessons/1/nmita-national.pdf). The first lesson is designed to help students become familiar with the overall structure of the NMITA database and learn about the types and formats of the data that are contained therein. This lesson takes students on a virtual journey to the Dominican Republic, a country located on the island of Hispaniola in the Caribbean Sea. On the island, students explore the marine reef coral fossils that occur in the sedimentary rock layers formed when parts of the island were under the Caribbean Sea. By the end of the first lesson, students should be familiar enough with the database to begin using it to gather larger amounts of data from different regions and answer specific questions about evolution and biodiversity.

Lesson 2 is designed to help students learn how to use NMITA to gather data and test hypotheses about coral reef biodiversity change through time. In Lesson 2 students collect coral reef biodiversity data from six different time periods in order to determine if coral reef biodiversity has increased, decreased, or stayed the same over the past eight million years in the Dominican Republic. Students collect, graph, and interpret data from three samples from each of the six time periods (Table 4).

Lesson 3 builds upon Lessons 1 and 2 and explores *ecological* and *morphological* patterns of biodiversity change in the same coral reef samples that students studied previously. In this lesson students use NMITA to gather data on the morphology, or shape, of the living and fossil coral species and plot these changes through time. Students determine whether reef coral morphological patterns have changed over the past eight million years.

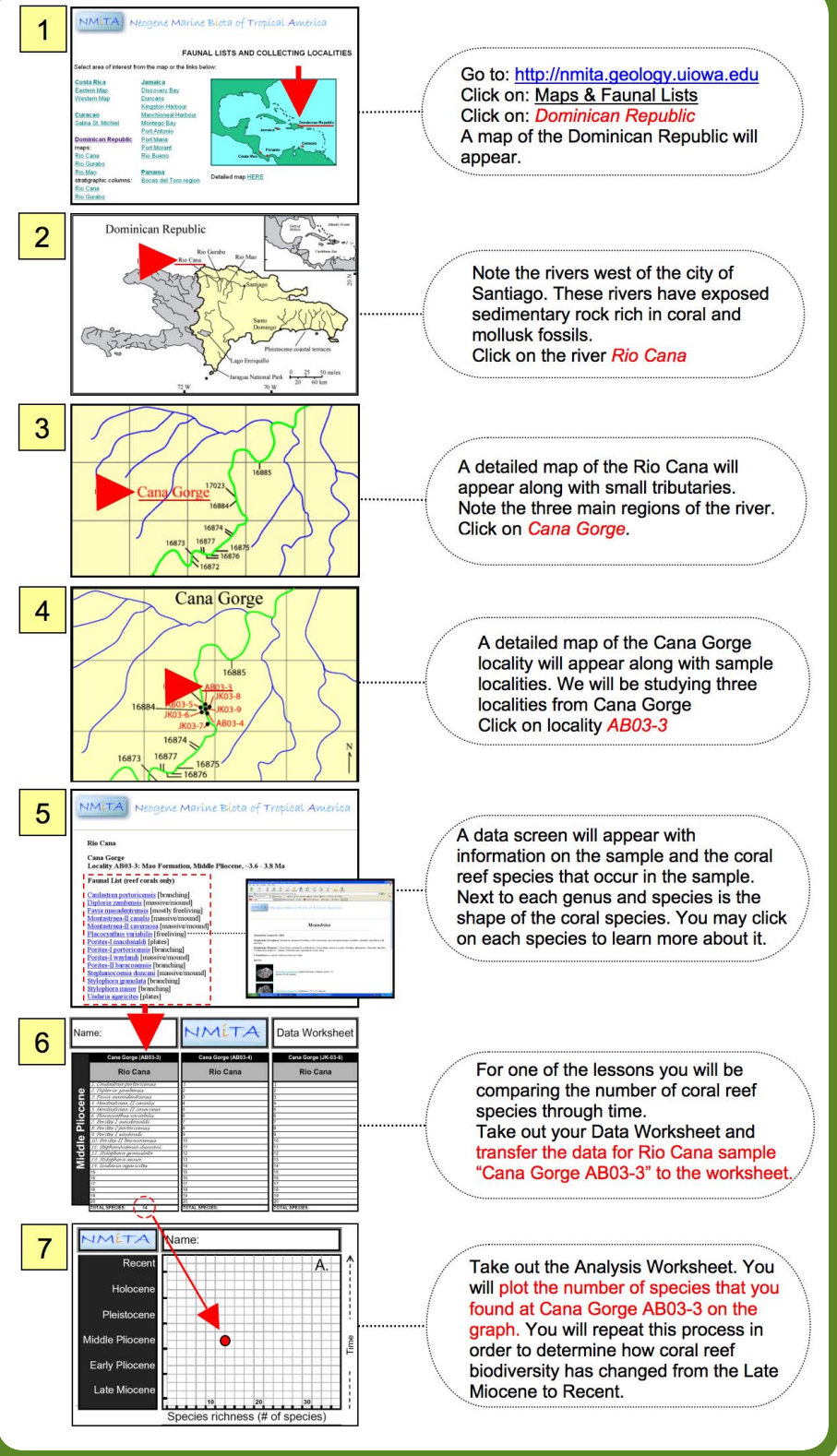
Lesson 4 has students use NMITA to determine *when* modern reef corals from the Dominican Republic originated. Did all the living genera and species of corals appear recently, or many millions of years ago? Did the modern genera and species appear together or randomly through time? NMITA's data on fossil reef corals, combined with data from living reefs, are used to answer these questions. Students collect data from three samples from each of the six time periods.

Assessment

We have introduced biodiversity bioinformatics (using the NMITA lessons) to a diverse group of 9th grade biology students at an urban high school in New York City as part of a six-week enrichment module over the past three years. The module included an overview of bioinformatics and an introduction to both genetic bioinformatics databases (e.g., NCBI's Genbank) and biodiversity bioinformatics databases (e.g., NMITA). The biodiversity bioinformatics component of this module included the four computer laboratory lessons using NMITA discussed previously; they are free and available at www.dominicanrepublicproject.org.

The enrichment module had six learning objectives

Figure 1. NMITA screen demonstrates how to gather biodiversity data (from Lesson 1).



(Table 5). In the third year of the course we assessed our success in achieving these objectives by collecting pre- and post module data on student performance. Student data were collected in two ways: (1) written pencil and paper tests; and (2) written laboratory reports that required the retrieval, application, and interpretation of data from bioinformatics

Table 5. Assessment data. Twenty students participated in the six-week enrichment module. Numbers indicate how many students were able to successfully complete the listed task before and after the module.

NATIONAL SCIENCE EDUCATION STANDARD	ASSESSMENT TOPIC	PARTICIPANTS	PRE-POST DATA (n)	P VALUE
Science and Technology Standard E: Science often advances with the introduction of new technologies. Solving technological problems often results in new scientific knowledge. New technologies often extend the current levels of scientific understanding and introduce new areas of research.	1. Awareness of new field of bioinformatics	20	Pre = 1; Post = 20	**P. <0.01
	2. Ability to define bioinformatics	20	Pre = 0; Post = 18	** P. <0.01
	3. Ability to explain why the new discipline of bioinformatics exists	20	Pre = 0; Post = 18	** P. <0.01
	4. Ability to name and locate at least two bioinformatics databases on the Internet	20	Pre = 0; Post = 20	** P. <0.01
	5. Ability to use at least one genetic and one biodiversity bioinformatics database to retrieve data	20	Pre = 0; Post = 20	** P. <0.01
	6. Ability to analyze and interpret data from a biodiversity bioinformatics database	20	Pre = 0; Post = 15	** P. <0.01

databases (e.g., NMITA Lesson 3). To analyze the data, a mastery approach to student achievement was taken. Using a percentage gain of post-test versus pre-test approach would have been less meaningful than a mastery approach because statistically significant percentage gains on the post-test were inevitable given that most of the students had no prior knowledge of bioinformatics. Before beginning the module all but one of the students were unaware of the meaning of bioinformatics and none of the students had used a bioinformatics database. Using a mastery approach, a predetermined high level of student achievement in bioinformatics content and procedural knowledge was set as acceptable (based on the *National Science Education Standards* and age-appropriateness). We assessed whether or not students had successfully mastered the learning objectives of the module (Table 5). Student scores were tabulated as categorical binary data (e.g., 0 = unsuccessful completion of task; 1 = successful completion of task). Incomplete and/or partially accurate responses, while indicative of student gains, were categorized as unsuccessful task achievement. The nonparametric McNemar test (SPSS, 1998) was used to test whether the binary mastery distributions were significantly different before and after the module. This test is appropriate for detecting changes in responses due to experimental intervention (SPSS, 1998). The modules were deemed successful in this small sample by the statistically significant change in the distribution of students who achieved a mastery of the content and procedural objectives in bioinformatics (Table 5). Larger samples will be required to determine if such success can be expected for high school students in general.

Conclusion

NMITA is a biodiversity database that may be used to introduce students to the expansive realm of bioinformatics beyond genetics. We have developed a series of lessons that have students use this database, thereby accessing real data that can be used to test hypotheses about biodiversity and evolution using fossils, while targeting the *National Science Education Standards*.

References

- Bisby, F.A. (2000). The quiet revolution: Biodiversity informatics and the internet. *Science*, 289(5488), 2309-2312.
- Bloom, M. (2001). Biology in silico: The bioinformatics revolution. The Human Genome Project is forging an alliance between biology and computers that promises to revolutionize our understanding of genetics. *The American Biology Teacher*, 63(6), 397-401.
- Biological Sciences Curriculum Study. (2004). Bioinformatics and the Human Genome Project. Available online at: <http://www.bsos.org/>.
- Budd, A.F., Foster, C.T., Dawson, J.P. & Johnson, K.G. (2001). The Neogene Marine Biota of Tropical America (NMITA) database: accounting for biodiversity in paleontology. *Journal of Paleontology*, 75(3), 743-751.
- Buxeda, R.J. & Moore-Russo, D.A. (2003). Enhancing biology instruction with the Human Genome Project. *The American Biology Teacher*, 65(9), 664-668.
- Collins, L. (2005). The Panama Paleontology Project. Available online at: <http://www.fiu.edu/~collinsl/pppimagemapnew.htm>.
- Edwards, J.L., Lane, M.A. & Nielsen, E.S. (2000). Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, 289(5488), 2312-2314.
- Jackson, J.B.C., Budd, A.F. & Coates, A. (Editors). (1996). *Evolution and Environment in Tropical America*. Chicago: University of Chicago Press.
- Jegalian, K. (2000). The Human Genome Project—its implications for educators. *The American Biology Teacher*, 62(9), 626-627.
- Kaiser, J. (2000). Fossil databases move to the web. *Science*, 289(5488), 2307.
- National Academy of Sciences. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- National Center for Biotechnology Information (NCBI). (2004). Available online at: <http://www.ncbi.nlm.nih.gov/>.
- Nehm, R.H. (2001). Calibrating spatial and temporal diversity patterns in tropical American *Prunum* and *Volvarina* (Gastropoda: Marginellidae). *Journal of Paleontology*, 75(3), 680-696.

- Nehm, R. H. (2004). The Dominican Republic Project. Available online at: <http://www.DominicanRepublicProject.org>.
- Palladino, M. A. (2002). Learning about the Human Genome Project via the Web: Internet resources for biology students. *The American Biology Teacher*, 64(2), 110–116.
- Pennisi, E. (2000). Taxonomic revival. *Science*, 289(5488), 2306–2308.
- Rowen, L., Mahairas, G. & Hood, L. (1997). Sequencing the human genome. *Science*, 278(5338), 605–607.
- Science*. (September 15, 2003). Cover photo.
- Smith, T. M. & Emmeluth, D.S. (2002). Introducing bioinformatics into the biology curriculum: Exploring the National Center for Biotechnology Information. *The American Biology Teacher*, 64(2), 93–99.
- Sugden, A. & Pennisi, E. (2000). Diversity digitized. *Science*, 289(5488), 2305.
- SPSS Version 9.0 (1998). Chicago, IL: SPSS, Inc.
- Wilson, E. O. (2000). A global biodiversity map. *Science*, 289(5488), 2279.