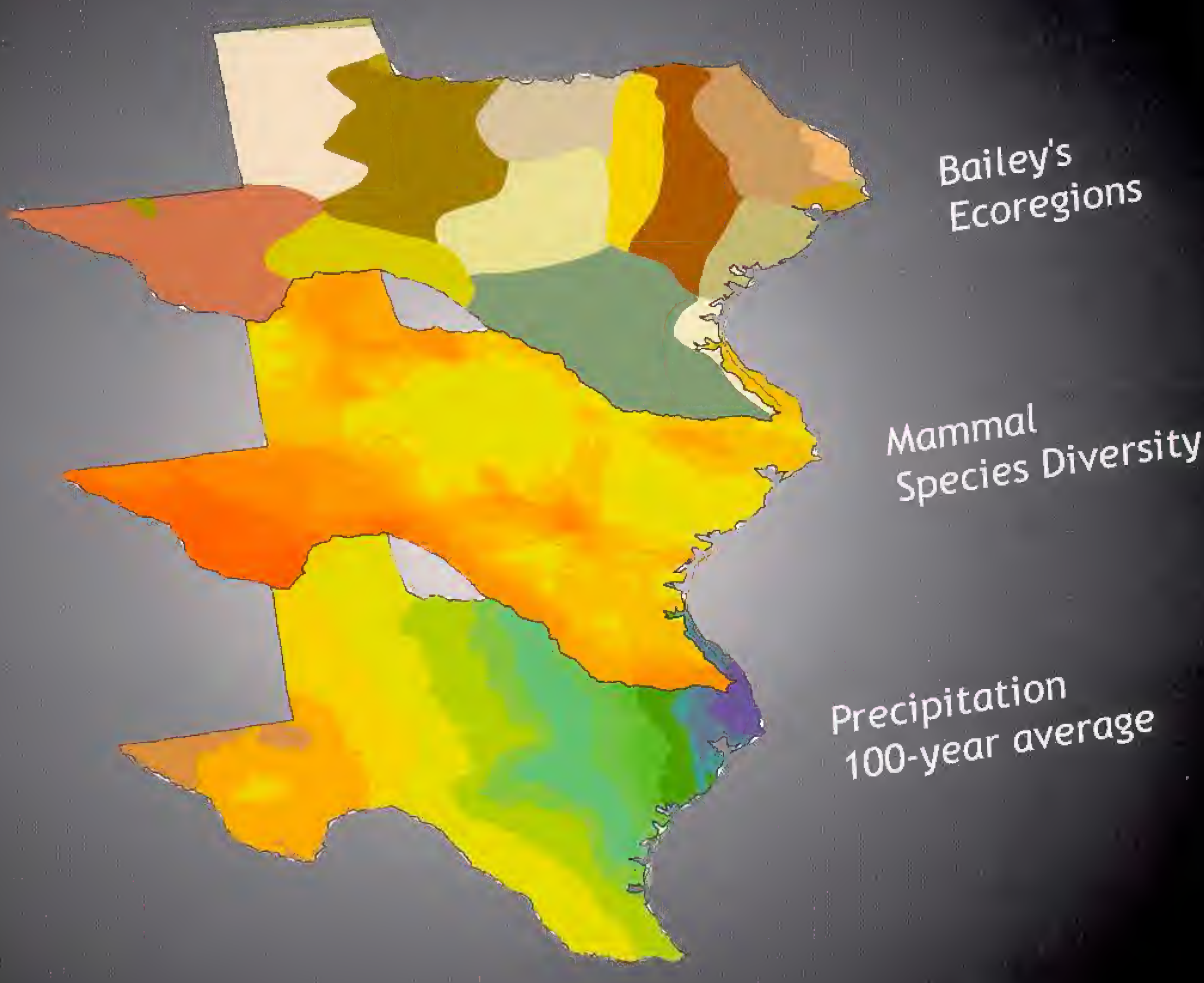
Occasional Papers

Museum of Texas Tech University

Number 187 *21 December 1998*



Bioinformatics, Museums and Society:

Integrating Biological Data for Knowledge-Based Decisions

Texas Tech University's Role in Bioinformatics

"As we envision it, our role at Texas Tech is fourfold. One aspect is to develop a specimen and geographic information systems niche in the field of bioinformatics to the point where it may be globally exploited. A second aspect is a role unique to universities, namely, to develop the philosophical underpinnings of what will become a new paradigm that blends and synthesizes historical and contemporary information about living things. A third aspect is to interface the collections of museums into a bioinformatics-compatible format in order that our archival collections become a valuable resource for information on biocomplexity issues. The final aspect is to prepare and educate the specialists necessary to serve the broad needs of society."

Front cover: Example of bioinformatic analysis from Texas Tech University employing the mammal collection and the Geographic Information Systems (GIS) databases in comparison to the ecological regions described by Bailey et al. (1994). The computer-generated successive Texas maps show mammal species diversity (lowest in yellow, highest in red) and 100 year precipitation data (lowest in brown, highest in blue).

Bioinformatics, Museums and Society: Integrating Biological Data for Knowledge-Based Decisions

Robert J. Baker, Carleton J. Phillips, Robert D. Bradley, John M. Burns, Daniel Cooke, Gary F. Edson, Donald R. Haragan, Clyde Jones, R. Richard Monk, John T. Montford, David J. Schmidly, and Nick C. Parker

Bioinformatics is an emerging field of science that integrates biological with geographical, climatological, health related, and economic databases to provide a framework for understanding biocomplexity within the context of significant societal issues (Grace 1997, Fischman, 1996). Bioinformatics is a systems approach that allows for retrospective analysis of past events, real-time analysis of complex contemporary data sets, and, ideally, positioning for an unpredictable future.

Until recently, 'biology' was thought of as a scientific field that produced basic information ranging from fundamental 'facts of life' to arcane details of natural history. In reality, the biological sciences comprise critical underpinnings of human health, environmental quality and law, biodiversity, planetary exploration, economic and commercial development, and most importantly, far-reaching policy decisions at all levels of government (Brenner, 1998). Given this dramatic scope and truly remarkable educational and economic potential, Texas Tech University proposes to make bioinformatics a centerpiece of multidisciplinary education and research.

As bioinformatics has become more widely used, its definition has broadened. One current formal definition from the world wide web (<http://www.bioplanet.com/whatis.html>) is "the systematic development and application of computing systems and computational solution techniques analyzing data obtained by experiments, modeling, database search, and instrumentation regarding biological aspects." At Texas Tech University, we have extended this definition to include "delivery of data and historical and real-time synthesis to all potential users" (Parker et. al, 1998). Because of our educational mission, we think of users as students and the general public along with scientists, economic developers and professional decision-makers. "Users" are central to the whole purpose and limited only by human ingenuity.

Decisions clearly will be influenced and limited by information processing and the accuracy, reliability, and detail of information. Implicit is the conviction that benefits of bioinformatics to society range from economic competitive position and sustainable development to local health service and quality of life. Scientific research is an expensive enterprise, but in our judgement, bioinformatics and associated data mining is the most efficient means of controlling costs. If the United States shirks this opportunity, other nations undoubtedly will rise to the occasion, giving them a great competitive advantage in use of information. The result could be erosion of our preeminent position in scientific research and its applications.

The very nature of "knowledge," "information," and "data" is changing. Some of the most important elements are: (a) the existence of significant historical data sets; (b) an exponential increase in volume and diversity of new knowledge and raw data; (c) a global need for rapid acquisition and processing of information central to decision-making, strategic planning and education; and (d) mechanisms by which non-specialists and non-experts in science (*e.g.*, economists, business and political leaders, planners and students) can gain access to reliable and useful syntheses of available information (Marshall, 1996).

The birth and development of bioinformatics can be thought of as a response to the growing challenge that has become more pressing with each passing year. The major elements of this challenge are as follows:

- Far more scientific data (in terms of volume, complexity and diversity) exist than can be used by classical methods.
- Most scientists (data collectors) are trained in more and more specialized areas, studying finer and finer details of a limited system.
- The current educational process produces highly trained but narrowly focused scientists with a limited field of intellectual vision.

Baker et al., 1998. Bioinformatics, Museums and Society: Integrating Biological Data for Knowledge-Based Decisions
Occasional Papers, Museum of Texas Tech University 187:i+1-4

- The technology that can synthesize megavolumes of detailed knowledge, old and new information, and data sets into a global 'scientific' worldview is available and expanding rapidly.
- For the most part, biologists are not computer scientists and relational database experts, and the problem of data management is beyond the current training of most biological scientists.

The Role of Museums and Voucher Specimens

Natural History museums, such as the one at Texas Tech, represent a wealth of biological information that has been carefully documented in a time and space format. Such collections might represent the most valuable biological resources in the world. Museum collections contain millions of specimens of plants and animals, many of which were preserved in such a way as to render them scientific treasure troves. Nearly all scientific names are referenced to a voucher specimen. Over time, voucher specimens increase in value through additional study and scientific publication. The huge numbers of voucher specimens are the foundation for thousands of scientific articles and books. Museums have enhanced the relevance of collections by modernizing: e.g., creating archives of preserved tissues, listing of collections and biological studies online and making data sets available for use in relational studies and other computer analyses (Baker, 1994; Baker and Yates, 1998).

Modern methods of preservation of tissues (liver, kidney, heart, lung, spleen, reproductive organs) either chemically, or by cryopreservation, greatly enhance the value of voucher specimens (Dessauer and Hafner, 1984; Phillips, 1985; and Longmire et al, 1997). Such materials are sources of DNA, RNA, proteins, other complex molecules, tissues and microorganisms that provide libraries of information on subjects including the history of life, biocomplexity (Mervis, 1998), life history strategies, disease and pollution. Genes recovered from such materials can be used in molecular biology procedures to develop transgenic plants, animals and microorganisms that can help solve world problems such as food production, insect and drought resistance, and production of pharmaceuticals (Wilson, 1989).

There is an historical component to voucher specimens. Biological collections in museums provide a means of assessing change over time. A collection that spans one hundred years can be used to trace shifting ranges of plants and animals, changing genetic characteristics, trends in species richness and patterns of abundance. This data set then can be matched to data about temperature and precipitation, landscape and landscape uses, and management practices or policies. The value of historical or retrospective information scarcely can be over-estimated. One need only think of the controversy about 'global warming' to appreciate the need for historical reference points.

Texas Tech has assumed an active leadership role in bioinformatics (Baker et. al, 1997) by becoming a leading repository for biological specimens. We have increased our ability to archive and computerize specimen-based information for input into the large relational databases by implementing more efficient procedures such as bar coding all specimens (Monk, 1998). In less than three decades, we have amassed one of the largest collections of mammals in the United States and one of the largest collections of chemically and cryopreserved tissues from wild mammals in the world.

The Vision and Opportunity

Our vision at Texas Tech University is to develop bioinformatics as a centerpiece of our research and educational mission. In this process we will integrate strengths from Biological Sciences and our museum science program, biological research collections, geographic information systems (GIS) and landscape analysis programs, Institute of Environmental and Human Health (TIEHH), Imagine Analysis Laboratory, research programs in plant and animal genomics, and computer science programs. To achieve our ambitious goals, it will be necessary to implement the following.

- Develop a multidisciplinary research team consisting of relational database managers, computer scientists and biologists from a broad array of disciplines.
- Create a research structure that provides a work environment where the team interacts and works to-

gether on a daily basis.

- Develop a system of regular team meetings, seminars, and discussion groups that integrates and focuses on problems and research.
- Construct a curriculum of bioinformatics-based courses that fill the needs of educational growth in bioinformatics and that address undergraduate, graduate and postdoctoral education.
- Provide hardware (super computer capabilities) and related equipment with the appropriate software to support the bioinformatics agenda.
- Develop computer language(s) suitable for interface with the GIS world and for data mining on distributed databases.
- Create databases such as rain fall, soil types, land use and distribution of chemicals (fertilizers, pesticides, herbicides, air pollution).
- Texas Tech will continue to grow as a repository for biological specimens. By wisely choosing specimens to be preserved we can address numerous concerns including health and agricultural issues along with issues critical to understanding biocomplexity.
- Acquire collections from other institutions no longer engaged in specimen-based research. In this unprecedented era of tight budgets and priority setting, many institutions have abandoned natural history and specimen-based research. Collections housed at such institutions continue to occupy space and require resources for maintenance. Even more important, they are not being used to their maximum benefit for society and science. We will provide institutions with an alternative for caring for these orphaned collections by placing them in an environment where they will be properly archived and available for use.
- Establish a web site that provides users access to the databases including the museum collections.

As we envision it, our role at Texas Tech is fourfold. One aspect is to develop a specimen and geographic information systems niche in the field of bioinformatics to the point where it can be globally exploited. A second aspect is a role unique to universities, namely, to develop the philosophical underpinnings of what will become a new paradigm that blends and synthesizes historical and contemporary information about living things. A third aspect is to interface the collections of museums into a bioinformatics-compatible format in order that our archival collections become a valuable resource for information on biocomplexity issues. The final aspect is to prepare and educate experts capable of serving broad societal needs.

In summary, Texas Tech University has recognized the fundamental significance of bioinformatics as a logical approach to enhance educational and economic growth. By developing a bioinformatics program as a centerpiece, the University is building on its existing strengths. In our opinion, implementation of a bioinformatics program will be the catalyst for true multidisciplinary education, research, and public service.

Acknowledgments

We thank H. H. Genoways and Laura Wiggins for critically reviewing the manuscript. Supported by National Science Foundation grant number DBI-9808928, R. D. Bradley, PI.

Literature Cited

Bailey, R.G., P.E. Avers, T. King, and W.H. McNab, eds. 1994. Ecoregions and subregions of the United States. USDA Forest Service, Washington, DC.

Baker, R.J. and T.L. Yates. 1998. Net-wielding Anachronisms. Letter to the Editor, Science, 282:1048-1049.

Baker, R.J., R.A. Albin, R.D. Bradley, J.J. Bull, J. Burns, K.A. Clark, G. Edson, R.E. Estrada, E. Farley, C.B. Fedler, B.M. Gharaibeh, R.L. Hammer, C. Jones, R.R. Monk, J.T. Montford, G. Moore, N.C. Parker, J. Rawlings, A. Sansom, D.J. Schmidly, R.W. Sims, H. Wichman, and F.D. Yancey. 1997. Natural Science Database: Resource management and public health. Pages 10-20 *in* M. Shaughnessy (ed.) Collaboration: The 'Key' to Success. Proceedings of the 4th Annual Conference, Organization of Fish and Wildlife Information Managers. Key Largo, FL. 96 pp.

Baker, R. J. 1994. Some thoughts on conservation, biodiversity, museums, molecular characters, systematics and basic research. Journal of Mammalogy, 75:277-287.

Brenner, S. 1998. The impact of society on science. Science 282:1411-1412.

PUBLICATIONS OF THE MUSEUM OF TEXAS TECH UNIVERSITY

It was through the efforts of Horn Professor J Knox Jones, as director of Academic Publications, that Texas Tech University initiated several publications series including the Occasional Papers of the Museum. This and future editions in the series are a memorial to his dedication to excellence in academic publications. Professor Jones enjoyed editing scientific publications and served the scientific community as an editor for the Journal of Mammalogy, Evolution, The Texas Journal of Science, Occasional Papers of the Museum, and Special Publications of the Museum. It is with special fondness that we remember Dr. J Knox Jones.

Institutional subscriptions are available through the Museum of Texas Tech University, attn: NSRL Publications Secretary, Box 43191, Lubbock, TX 79409-3191. Individuals may also purchase separate numbers of the Occasional Papers directly from the Museum of Texas Tech University.

Layout and Design: R. Richard Monk
Cover Design: Eric Schultz and Eric Holt

This book was set in Times New Roman and printed on acid-free paper that meets the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources.

Printed: 21 December 1998

Library of Congress Cataloging-in-Publication Data

Occasional Papers, Number 187
Series Editor: Robert J. Baker

BIOINFORMATICS, MUSEUMS AND SOCIETY:
INTEGRATING BIOLOGICAL DATA FOR KNOWLEDGE-BASED DECISIONS

By: Robert J. Baker, Carleton J. Phillips, Robert D. Bradley, John M. Burns, Daniel Cooke, Gary F. Edson, Donald R. Haragan, Clyde Jones, R. Richard Monk, John T. Montford, David J. Schmidly, and Nick C. Parker

ISSN 0149-175X

Museum of Texas Tech University
Lubbock, TX 79409-3191 USA
(806)742-2442

SEAL OF TEXAS TECH UNIVERSITY ™