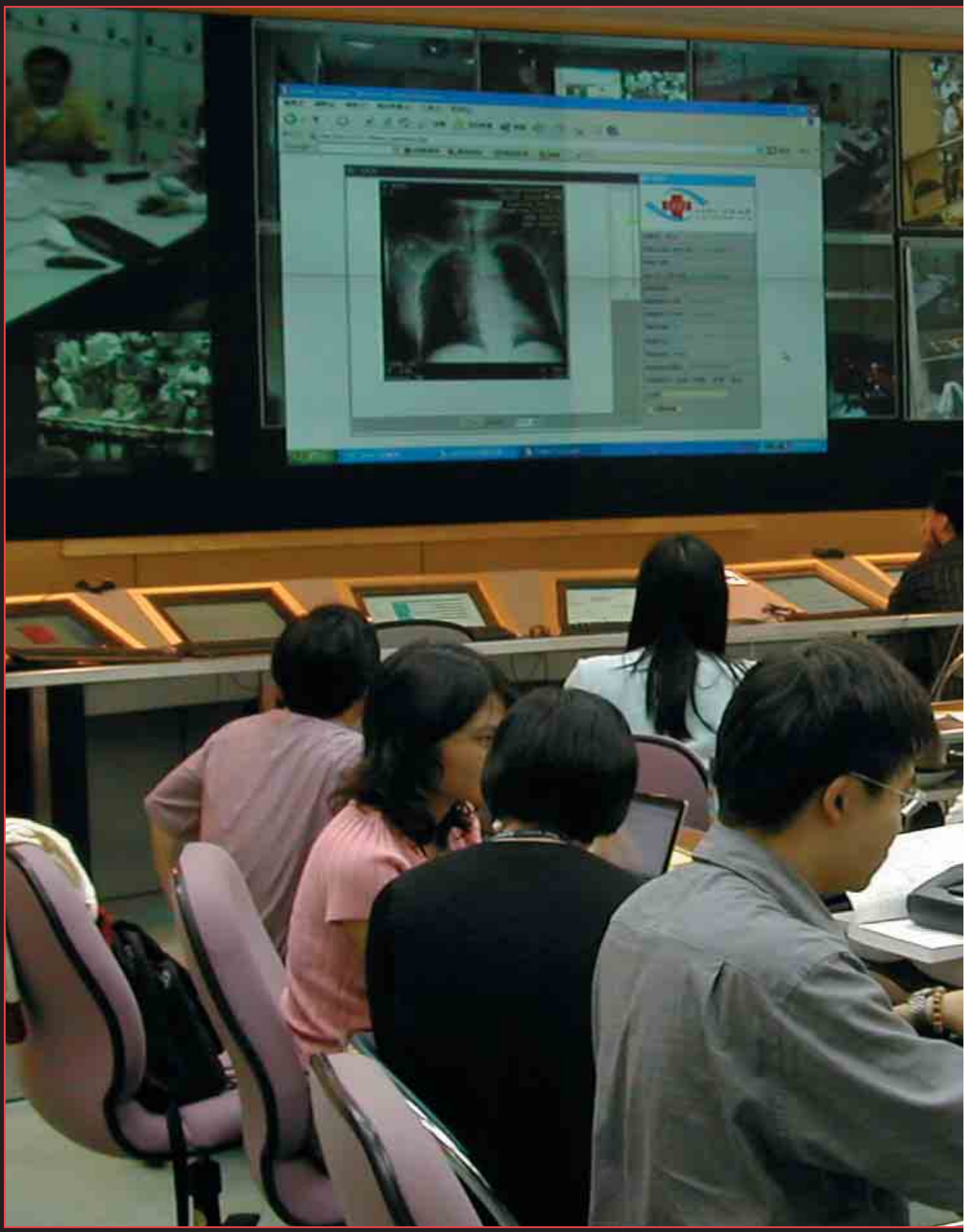
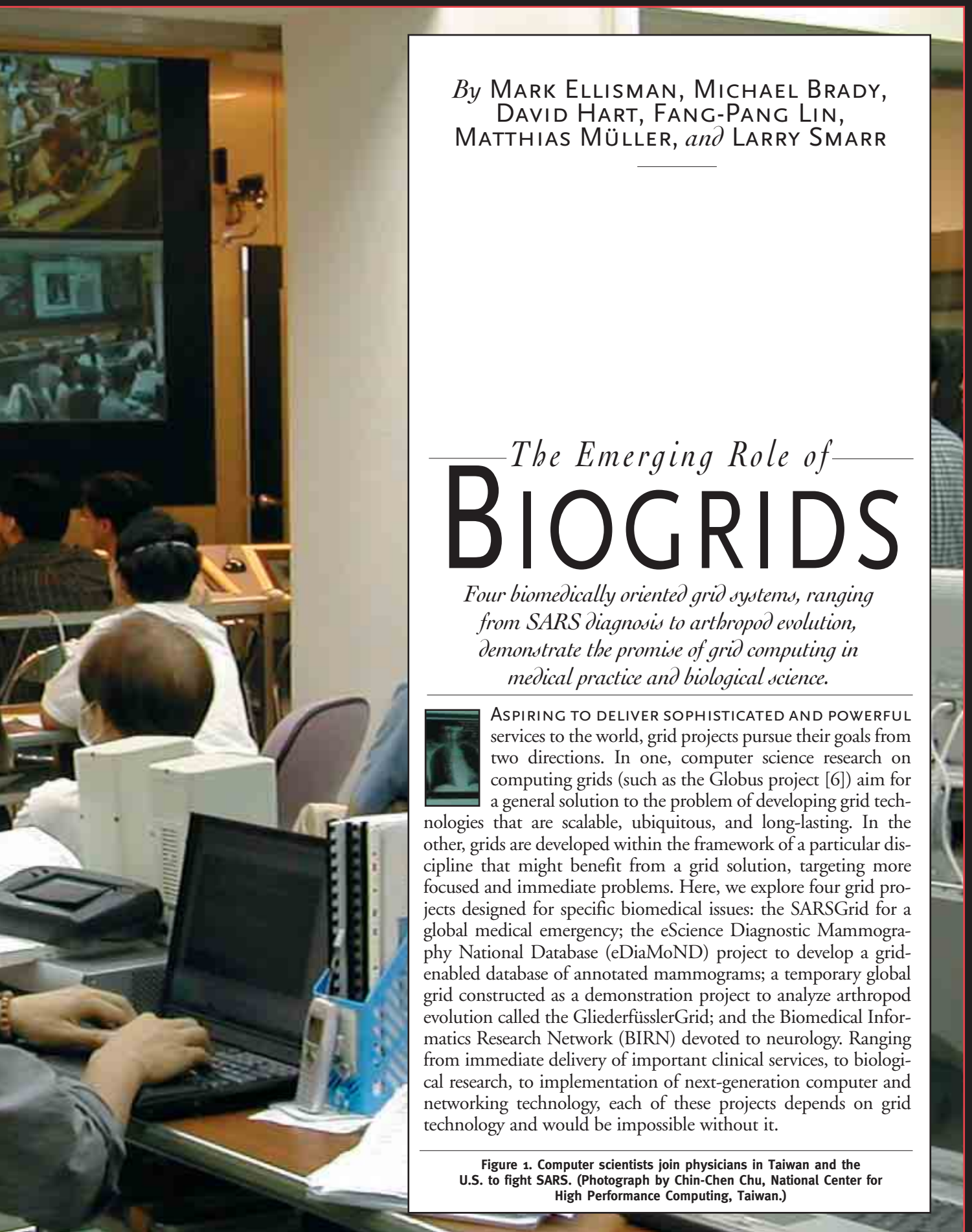




By Mark Ellisman, Michael Brady, David Hart, Fang-Pang Lin, Matthias Müller, *and* Larry Smarr

The Emerging Role of

BIOGRIDS

Four biomedically oriented grid systems, ranging from SARS diagnosis to arthropod evolution, demonstrate the promise of grid computing in medical practice and biological science.

Aspiring to deliver sophisticated and powerful services to the world, grid projects pursue their goals from two directions. In one, computer science research on computing grids (such as the Globus project [6]) aim for a general solution to the problem of developing grid technologies that are scalable, ubiquitous, and long-lasting. In the other, grids are developed within the framework of a particular discipline that might benefit from a grid solution, targeting more focused and immediate problems. Here, we explore four grid projects designed for specific biomedical issues: the SARSGrid for a global medical emergency; the eScience Diagnostic Mammography National Database (eDiaMoND) project to develop a grid-enabled database of annotated mammograms; a temporary global grid constructed as a demonstration project to analyze arthropod evolution called the GliederfüsslerGrid; and the Biomedical Informatics Research Network (BIRN) devoted to neurology. Ranging from immediate delivery of important clinical services, to biological research, to implementation of next-generation computer and networking technology, each of these projects depends on grid technology and would be impossible without it.

Figure 1. Computer scientists join physicians in Taiwan and the U.S. to fight SARS. (Photograph by Chin-Chen Chu, National Center for High Performance Computing, Taiwan.)

Grid applications and human networking to combat SARS. The disease known as severe acute respiratory syndrome, or SARS, was first reported in Asia in February 2003. It spread quickly and resulted in a global outbreak over the following months before finally being contained in July 2003. A major outbreak in Taiwan occurred in April 2003, confined mostly to hospitals. However, when a hospital was affected, many medical personnel had to be quarantined, thus reducing the numbers available to treat the sick.

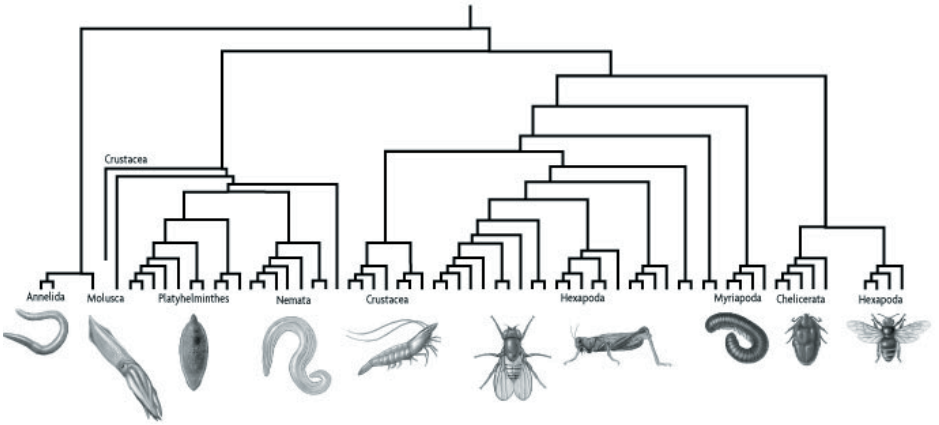


Figure 2. Evolutionary tree with some of the species studied. In this run the smallest subtree, including the files, locusts, and bees, also contains millipedes and spiders, suggesting hexapods might not form a complete single evolutionary unit. (Jennifer Fairman, Fairman Studios, www.fairmanstudios.com)

Remote medical diagnosis, supporting the sharing of expertise, is a significant benefit in such circumstances. The National Center for High-Performance Computing (NCHC) in Hsinchu, Taiwan, was called in to leverage its existing grid-based collaborative projects with hospitals and manufacturers for online asthma monitoring to help address the emergency. NCHC adopted several tactics, including use of the Access Grid (which supports group-to-group interactions across the grid) for communication among hospitals and the U.S. Centers for Disease Control and Prevention in Atlanta, retrieving information via existing grid projects, and quickly deploying a dedicated network. This work was a race against the spread of the disease and helped save the lives of at least two infected patients.

The first collective system was developed in two weeks. Addressing the practical problem of medical staff shortages in particular hospitals, it allowed doctors both in and outside Taiwan to participate, in real time, in the analysis and diagnosis of individual cases, without risk of contracting the virus (see Figure 1). The collaborative technology provided remote accessibility that was indeed better than being there. Development of this grid-based infrastructure continues in preparation for the possibility that SARS (and other diseases) could emerge again. The effort clearly demonstrates that grid collaboration technology can save lives.

eDiaMonND. Mammograms are an important tool for maintaining women's health by detecting breast cancer early when the prognosis for treatment is optimal. As a result, the screening age is being extended to 70 years. At the same time, there is a movement among breast cancer specialists to two-view screening in order to improve the accuracy of the diagnosis. But these initiatives are also expected to increase the typical radiologist's workload by 50%. Consequently, there is great interest in using information technology to cope with this emerging demand.

Managing mammogram data involves further technological challenges. First, a single mammogram of a patient is often less helpful than the difference between a current mammogram and one from the same patient one or two years earlier. Secondly, mammograms of two different patients taken on the same instrument may look more like each other than two mammograms from the same patient taken on two different instruments. Moreover, people move from place to place over time, so their records inevitably become scattered and sometimes even lost.

These factors create a critical need for a shared, common distributed database of mammograms that can be used in clinical work, providing secure access to a patient's medical history and earlier mammograms; standard formats to enable comparison; and a library of examples for training and diagnosis. The eDiaMoND Project (www.ediamond.ox.ac.uk) at Oxford University is building such a database, initially from four of the U.K.'s 92 breast-screening centers. In order to address the consistency issue, a normalized form of each image is generated using the Standard Mammographic Form [2] software.

eDiaMoND aims to help develop grid middleware, as well as demonstrate the value of a virtual mammography image store constructed from the image stores at each of the collaborating screening centers. The image stores on which eDiaMoND is based remain at the originating center, while the virtual store is constructed by "federating," or integrating, those image stores to create the illusion of a single database. eDiaMoND expects to deliver a data-mining service

and a remote reading service, permitting mammograms created by Hospital X and stored at Hospital Y to be read by an authorized user at Hospital Z. The project is expected to demonstrate how grid-based systems can support epidemiological studies and contribute to quality control at the breast-care units, while improving patient care.

Arthropod evolution. The annual ACM/IEEE Supercomputing (SC) conference is widely regarded as the premier gathering for the high-performance computing and communications community. It features a friendly competition called the HPC Challenge (www.sc-conference.org/sc2003/tech_hpc.php), an opportunity for various research teams to pull together resources to demonstrate proofs of concept. These challenges often help accelerate development of new technologies; perhaps the most dramatic example was the 1995 I-way project [3], which paved the way for the creation of the very high-speed Backbone Network Service and then the Abilene network.

A biologically motivated project was among the winners in 2003, when Indiana University and the High Performance Computing Center of Stuttgart (HLRS) led an international collaboration to create a temporary global grid performing a massive analysis of arthropod evolution. Arthropods are invertebrates with hard exoskeletons and jointed legs (insects). For many years it was believed that all six-legged arthropods (hexapods, such as ants and flies) were a single evolutionary group; recently, however, it has been suggested this may not be so.

Indiana University has long maintained and managed development of the parallel version of the well-known phylogenetic code fastDNAml [5, 8]. This code performs a heuristic search for the most likely phylogenetic tree, based on aligned genetic sequences. The order in which taxa are added to the tree, as well as the subsampling of the genetic loci considered in the analysis, are randomized. Many individual analyses are performed to increase the chance of finding the best tree. The computational requirements for phylogenetic analyses are formidable (see Bader's article in this section). A typical journal or proceedings article in evolutionary biology might analyze 100 randomizations of the data. Because hexapod evolution is the subject of significant debate, the researchers hoped to analyze 300 randomizations, even though such a task requires thousands of hours of CPU time.

HLRS has developed a grid-aware Message Passing Interface (MPI) library called the Parallel Computer eXtension, or PACX-MPI [7], and a collaborative, virtual reality software application called the COllaborative VIsualization and Simulation Environment (COVISE) [9]. Both were used to build a grid execution environment enabling the use of MPI-based parallel code across different computing systems. In this assembly HLRS software engineers employed a two-level approach. The PACX-MPI library was responsible for combining several computers into one large virtual resource. A special module within COVISE distributed the work among these virtual supercomputers while ensuring fault-tolerance. The interaction with users not only allowed for analyzing the final data, it also steered the ongoing simulation.

Computer centers around the world were invited to participate in the project, called Global Analysis of Arthropod Evolution. GliederfüsslerGrid, a computing grid of 641 CPUs and eight types of computing systems, was assembled, spanning every continent except Antarctica (see Figure 2). For approximately 10 days during the 2003 SC conference, one of the largest computing grids ever assembled was focused on answering an intriguing question in evolutionary biology; the biological results are still being analyzed.

Biomedical Informatics Research Network. BIRN [4] is a consortium of 14 universities (22 research groups) participating in one or more of the following three testbed projects involving brain imaging, human neurological disorders, and associated animal models of neurological disease:

Function BIRN. For studying regional brain dysfunctions related to the progression and treatment of schizophrenia;
Morphometry BIRN. For examining unipolar depression, mild Alzheimer's disease, and mild cognitive impairment; and
Mouse BIRN. For studying animal models of diseases and disorders (such as multiple sclerosis, Alzheimer's, schizophrenia, Parkinson's disease, Attention Deficit Hyperactivity Disorder, Tourette's syndrome, autism, and brain cancer).

These testbeds, along with the overall infrastructure of the BIRN and the growth and extension of these technologies to additional biomedical research and clinical care communities are supported by the BIRN Coordinating Center at the University of California, San Diego. BIRN is exploring the use of a virtual data grid (VDG) to support multiscale brain mapping. The BIRN VDG was brought online in less than a year, providing standardized storage, computation, and networking equipment to each of the of the 14 BIRN sites. Grid middleware includes the Storage Resource Broker (SRB) [1], which provides a uniform interface to heterogeneous data resources over a network, and Globus [6], which allows resource discovery, monitoring, utilization, and management over a network. All

files within the SRB environment are part of a single grid file system, where a file's logical location is independent of its physical location. Data is replicated across sites under SRB control. Once imaging data is stored within the BIRN VDG, users from any collaborating site are able to interact with that data through the BIRN portal, which is a workflow and application integration environment.

Single volumetric data sets acquired by BIRN researchers through such techniques as electron microscopic tomography are common at 2048 × 2048 × 512 pixels. New energy-filtered electron microscopes now being developed will allow for correction of chromatic aberration and consequently the use of much thicker specimens. With these technologies, individual data sets will commonly exceed 4K × 4K × 2K pixels within two years, and 12K × 12K × 2K pixels within five years. While the process of reconstructing the volumes from raw projection data can be addressed with the current generation of supercomputers, it is not yet possible to interactively explore, segment, and analyze the resultant volumes.

For this reason, BIRN provides an application driver for the National Science Foundation-funded OptIPuter [10], which seeks to upgrade the VDG to a LambdaGrid by replacing the Internet2 (in use today) with a dedicated optical network. End users can still come in over Internet2 to access the system, but once a query is inside the LambdaGrid, response times will be improved by an order of magnitude. The next-generation BIRN will provide to a distributed audience of biomedical researchers a set of storage, computation, and visualization systems that accomplish a set of tasks no single supercomputer alone can accomplish today.

Conclusion

The four applications of grid computing to biomedical problems explored here show that the grid can play an important role in the delivery of health care, as well as a deeper understanding of evolutionary processes. Biogrids are increasingly important in the development of new computing applications for the life sciences and in providing immediate medical benefits to individual patients and even to those only at risk of getting sick. The development of targeted grids providing practical solutions to individual problems is also important in the overall development of grid technology, providing immediate, visible rewards from the continuing investment. C

References

1. Baru, C., Moore, R., Rajasekar, A., and Wan, M. The SDSC Storage Resource Broker. In *Proceedings of the IBM Center for Advanced Studies (CASCON'98) Conference* (Toronto, Canada, Nov. 30–Dec. 3). IBM Press, Toronto, 1998.
2. Birchard, K. Online system to allow easy access to mammograms. *Medical Post 38,* 40 (Nov. 5, 2002); www.medicalpost.com/mpcontent/article.jsp?content=/content/EXTRACT/RAWART/3840/51B.html.
3. DeFanti, T., Foster, I., Papka, M., Stevens, R., and Kuhfuss, T. Overview of the I-WAY: Wide-area visual supercomputing. *Int. J. Supercomput. Applic. 10,* 2 (Summer/Fall 1996), 123–130.
4. Ellisman, M. and Peltier, S. Medical data federation: The biomedical informatics research network. In *The Grid: Blueprint for a New Computing Infrastructure, 2nd Ed.,* I. Foster and C. Kesselman, Eds. Morgan Kaufmann, San Francisco, 2004.
5. Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol. 17* (1981), 368–376.
6. Foster, I., Kesselman, C., and Tuecke, S. The anatomy of the grid: Enabling scalable virtual organizations. *Int. J. Supercomput. Applic. 15,* 3 (2001), 200–222.
7. Keller, R., Krammer, B., Müller, M., Resch, M., and Gabriel, E. Towards efficient execution of MPI applications on the grid: Porting and optimization issues. *J. Grid Comput. 1,* 2 (2003), 133–149.
8. Olsen, G., Matsuda, H., Hagstrom, R., and Overbeek, R. fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci. 10* (Feb. 1994), 41–48.
9. Rantzau, D., Frank, K., Lang, U., Rainer, D., and Wossner, U. COVISE in the CUBE: An environment for analyzing large and complex simulation data. In *Proceedings of the Second Workshop on Immersive Projection Technology* (Ames, IA, May 11–12). Iowa Center for Emerging Manufacturing Technology, Ames, IA, 1998.
10. Smarr, L., Chien, A., DeFanti, T., Leigh, J., and Papadopoulos, P. The OptIPuter. *Commun. ACM 46,* 11 (Nov. 2003), 58–66.

Mark Ellisman (mark@ncmir.ucsd.edu) is director of the National Center for Microscopy and Imaging Research at the University of California, San Diego.
Michael Brady (MBrady9258@aol.com) is BP Professor of Information Engineering at Oxford University in the U.K.
David Hart (dhart@indiana.edu) is manager of the High Performance Computing Support Group at Indiana University.
Fang-Pang Lin (c00fpl00@nchc.org.tw) is head of the Grid Computing Division in the National Center for High-Performance Computing in Hsinchu, Taiwan.
Matthias Müller (mueller@hlrs.de) is deputy director and head of the Technical and Scientific Computing Group at the High Performance Computing Center Stuttgart in Germany.
Larry Smarr (lsmarr@ucsd.edu) is director of the California Institute for Telecommunications and Information Technology and Harry E. Gruber Professor in the Department of Computer Science and Engineering at the University of California, San Diego.

The National Center for Microscopy and Imaging Research at the University of California, San Diego is supported by grants from the National Institutes of Health (grants P41RR04050 and R01NS14718), the National Science Foundation (grants ASC975249 and MCB9728338), the Branfman Foundation, the Michael J. Fox Foundation, and the W.M. Keck Foundation.

The BP Professor of Information Engineering Chair is endowed by BP. The eDiaMoND project is supported by Oxford University, IBM, and the U.K. government.

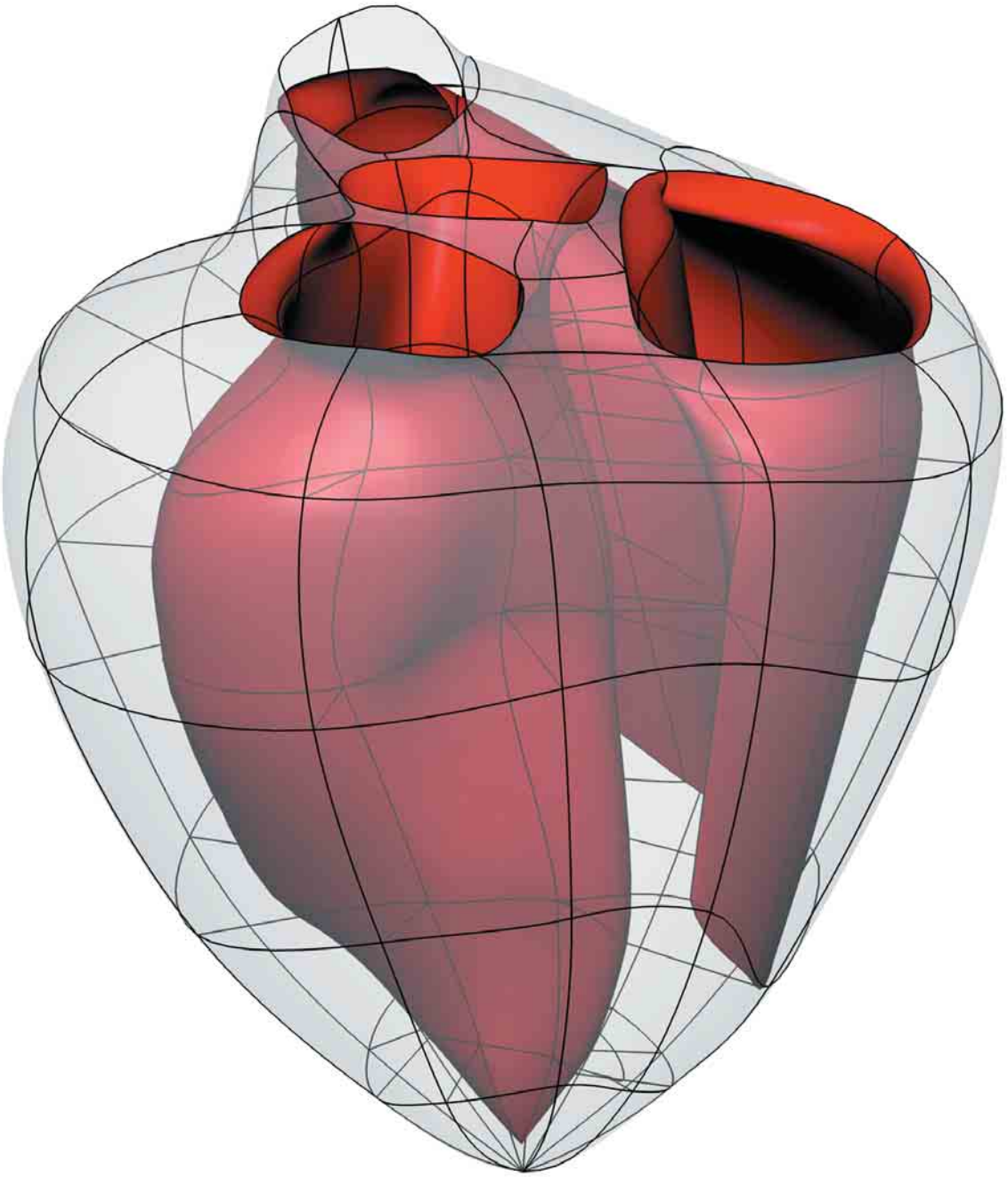
Indiana University's research in life science applications in information technology is supported in part by the National Science Foundation (grant 0116050) and by Shared University Research grants from IBM Inc. and by the Lilly Endowment, Inc. through support of the Indiana Genomics Initiative.

Grid computing research at the High Performance Computing Center Stuttgart is supported in part by DAMIEN (grant IST-2000-25406) and CROSSGRID (grant IST-2001-32243).

The Biomedical Informatics Research Network is funded by the National Center for Research Resources at the National Institutes of Health (grants RR04050, RR08605, and DC03192) and the National Science Foundation (grants ASC975249 and MCB9728338).

The OptIPuter project is funded by the National Science Foundation Information Technology Research program (grant SCI0225642).

© 2004 ACM 0001-0782/04/1100 $5.00



FINITE ELEMENT MODEL FOR STUDYING THE COUPLED ELECTROMECHANICS OF THE HUMAN HEART.
Peter Hunter, The Bioengineering Institute at the University of Auckland, New Zealand.