GENOMES to LIFE
BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES
U.S. DEPARTMENT OF ENERGY
INNOVATIVE APPROACHES ALONG UNCONVENTIONAL PATHS

Genomes To Life

Biology for 21st Century

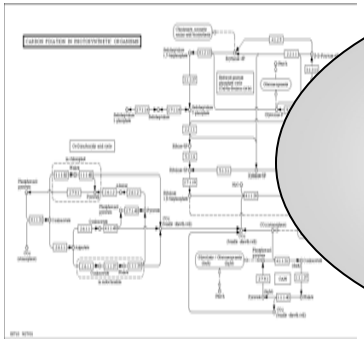
Al Geist
May 9, 2003
Computer Science and Mathematics Division

A Joint Initiative of the
Office of Advanced Scientific Computing Research
and Office of Biological and Environmental Research
http://DOEGenomesToLife.org

Office of Science
U.S. Department of Energy

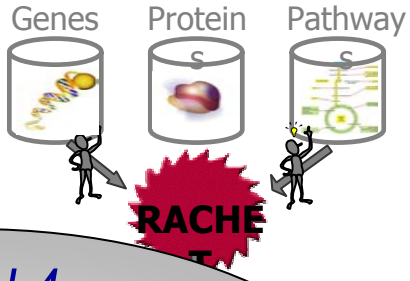
Four Goals of Genomes to Life

Goal 1
Identify and Characterize the Molecular Machines of Life

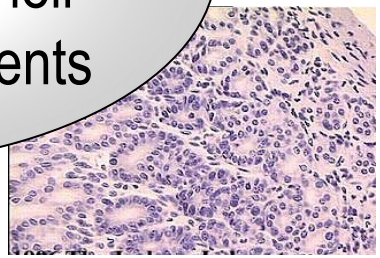


Goal 2
Characterize Gene Regulatory Networks

Understanding the essential processes of living systems



Goal 3
Characterize the Microbial Communities in their Natural Environments

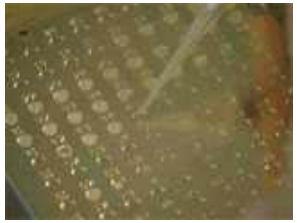


Goal 4
Develop the computational methods and capabilities to advance understanding of complex biological systems

www.genomes-to-life.org

DOE Genomes to Life Project

"The development of new effective and efficient methods and software tools to help both experimental and computational efforts characterize protein complexes, regulatory networks, and community behavior of microbes."



www.genomes-to-life.org

DOE Awards $103M to Five Large Projects

ORNL is a major winner

Goal 1 ***Genomes to life center for molecular and cellular systems***
led by ORNL and PNNL, w/ ANL, SNL, UNC, UU

Goal 4 ***Molecular Machines to Hierarchical modeling***
led by ORNL and SNL, w/ LBL, LANL, NCGR, UCSB, UTK, UM, UIUC

Goal 2 ***Rapid Deduction of Stress Pathways in Bacteria***
led by LBL, w/ ORNL, SNL, UCB, UMC, UW, Diversa

Goal 3 ***Microbial Ecology, Proteomics and Computational Optima***
led by Harvard Medical School w/ MIT, Brigham and Womens Hospital, Massachusetts General Hospital

Goal 3 ***Analysis of the Gene Expression of Microbial Communities***
led by U. Mass. W/ ANL, UT Memphis, and TIGR

Bioremediation

Carbon sequestration

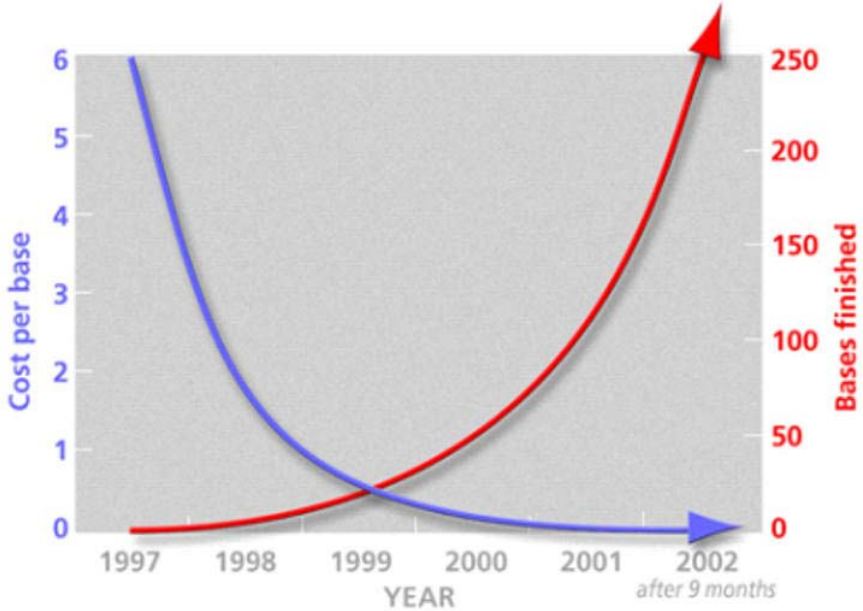
Energy production

www.genomes-to-life.org

Biology for the 21st Century

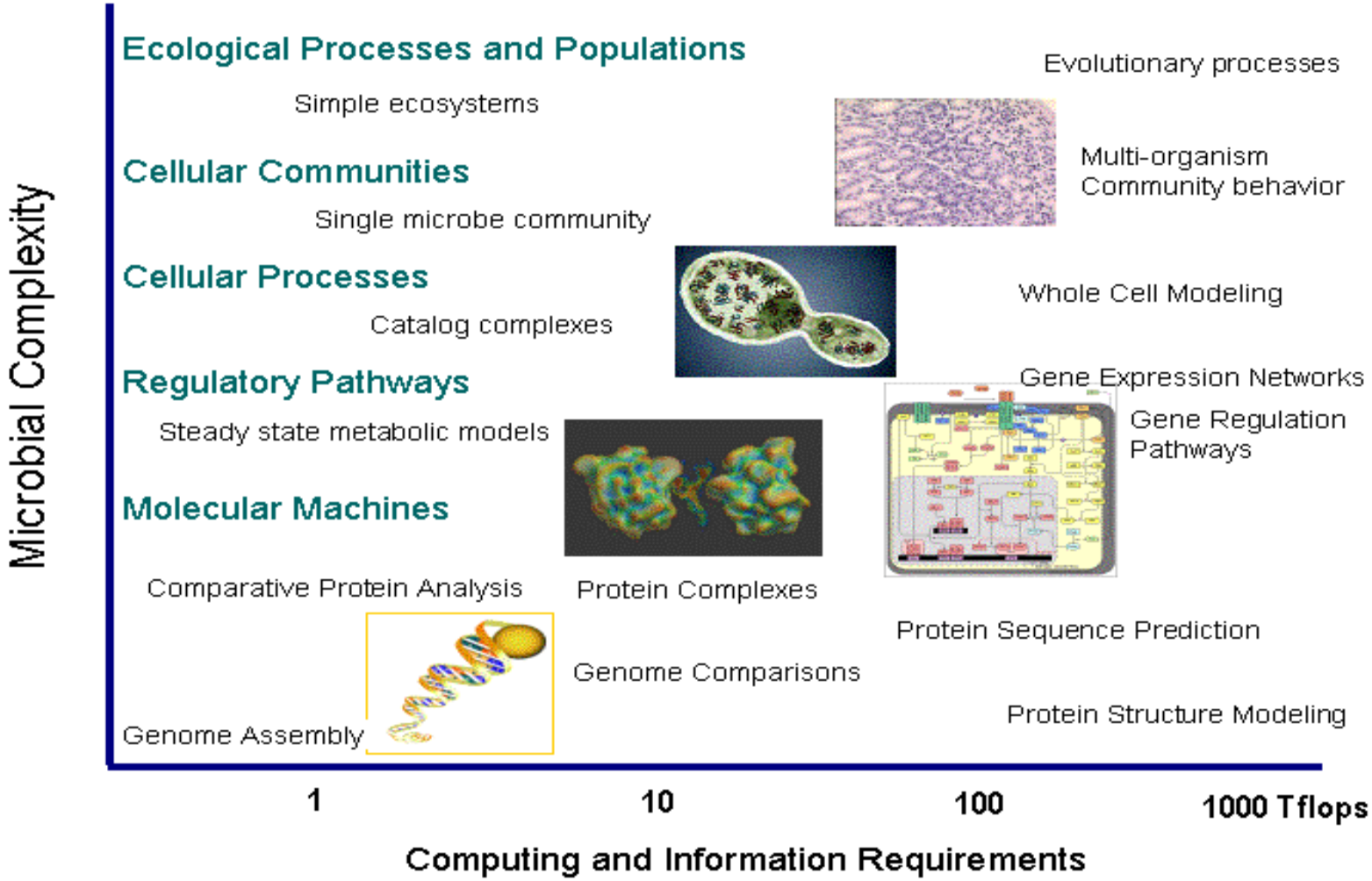
Growing reliance on on computing and data analysis

High-performance computing is essential to the high-throughput experimental approach to biology that has emerged in the last 10 years.



www.genomes-to-life.org

Broad Scale of Biological Complexity and Computational Requirements



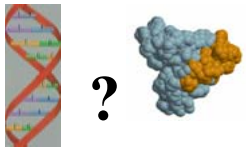
www.genomes-to-life.org

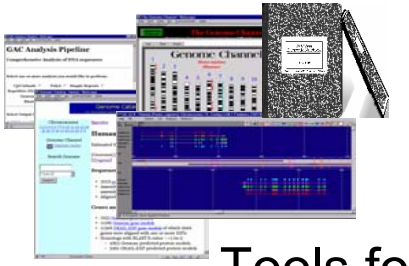
New Tools for New Challenges

- New high performance methods and software for characterizing protein complexes
- Efficient algorithms for determining regulatory pathways
- New approaches to cellular systems biology
- Improved methods for obtaining and evaluating microbial data
- Work environments & computational infrastructure for GTL
- Integrating all the pieces of GTL

Integrated Cyclic Process

High throughput Experiments

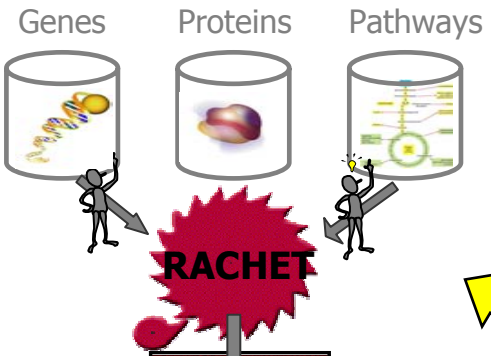


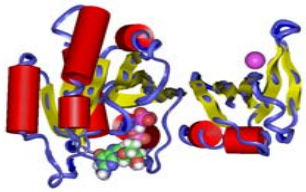


Tools for Ease of Use

Biology community

Distributed Databases





Simulation & Modeling

Advanced Computational Algorithms

Supercomputer Resources

Next Phase: GTL User Facilities

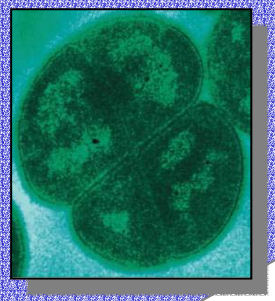
Open, Rapid User Access to Data

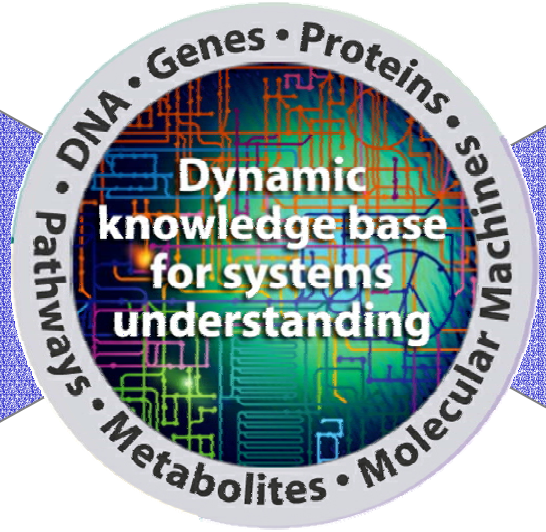
Technology Integration for Tomorrow's Biology

Advanced Technologies

Informatics and Databases

Microbial Genomics

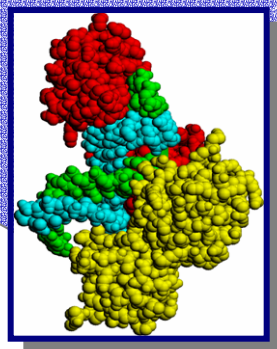




Biosystems Training

High-Throughput Analysis and Production

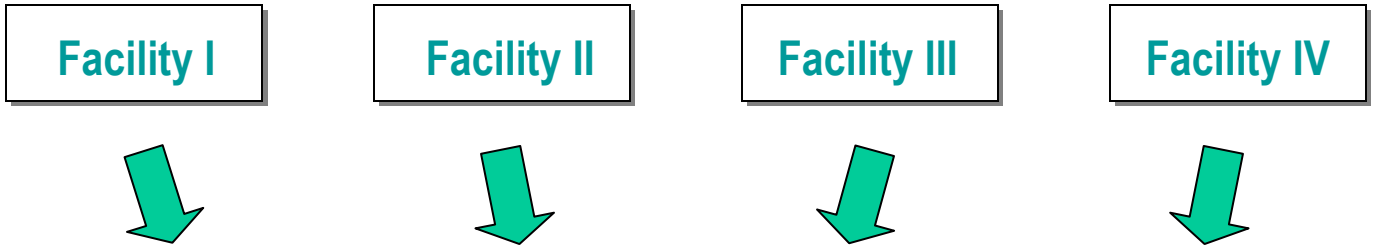
Computing and Simulation



Four GTL Facilities are Proposed

	Computational Needs
Facility I **Production and Characterization of Proteins**	Computer tools to make sense of the data, develop reliable rules to guide production, characterization, and for labelling sites
Facility II **Whole Proteome Analysis**	New computational tools for interpretation and modeling whole proteome data
Facility III **Characterization and Imaging Of Molecular Machines**	Modeling and simulation of molecular machines Protein interaction Mapping and Image analysis
Facility IV **Analysis and Modeling Of Cellular Systems**	New tools & infrastructure for analyses of cell state coupling experiment, analysis, and theory in a recursive fashion to reveal networks, function

Common Needs of GTL Facilities



- Databases and tools for interpreting, archiving, and disseminating data and models to the greater biological community.
- Integrated Large-Scale Experiment-Computing Cycles
- Integration of the large biology databases into a unified knowledge base for predictive modeling
- Both capability as well as capacity computation
- Simple user interfaces and transparent access to supercomputing resources

Office of Science
U.S. Department of Energy