The Next Revolution in Bioscience for Environment and Agriculture?

Dr Tan Tin Wee
Board Director, International Society for Computational Biology (ISCB.org)
Secretariat, S* Life Science Informatics Alliance (s-star.org)
Founding Secretariat, Asia Pacific Bioinformatics Network (APBioNet.org)
ExCo/Past President, Assn for Medical and BioInformatics Singapore (AMBIS.org.sg)
Assoc Professor, Dept of Biochemistry, National University of Singapore (NUS.edu.sg)
Is there a What’s Next?

- The Green Revolution (pre-Molecular Biology days)
- Biotech/recombinant DNA technology Revolution (1980s-1999)
- Genomic era (mid1990s-now)
- “Post-Genomic era” (now)
What’s underlying the next revolution?

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics
- Phenomics
- ‘omics

The New Language (or “Calculus”) of Biology
- Computational Biology, Bioinformatics
- Scientific Computing Information Technology (IT)
- Internet Networking (Cyberinfrastructure)
NSF’s 21st Century Biology

- Integrative
- Synthetic
- Predictive

To build a CIBIO CyberInfrastructure for 21st Century Biology

From NSF’s report:
“Building a Cyberinfrastructure for Biological Sciences” 2005 and Beyond: A roadmap for consolidation and exponentiation
Genbank data is doubling every 12-18 months
Faster than Moore’s Law of the IT world
After the Post-Genome deluge of data, we now need to analyse data at the Terabyte level

But the New Biology is not just about the volume of data
it is as much about the inherent complexity of biological information

“And that’s why we need a computer.”
BioImaging – Ever Increasing Complexity of Biological Data

- linear DNA sequences through to complex movie clips of living cells labelled with quantum dots etc captured in confocal/fluorescence/electron microscopy and MRI etc:
  - 1D sequence data,
  - 2D image of pixels, microarray data
  - 3D Xray structures and molecular models
  - 4D NMR spectroscopy
  - 6D space-time-spectral records of every voxel in a tomographic slice.
Progression of Life Sciences in the Future

Sequence | Structure | Function | Interactions | Pathways | Regulatory Networks/Circuits | “E-Cell” simulation | Tissue/Organ Physiology

Computational
Informational

Experimental
Observational

Genomics
Transcriptomics
Proteomics
Metabolomics
Other ‘omics
ExPERI MENTAL
OBSERVATIONAL

Bioimaging
From John Wooley et al 2005

Symposium 1: Bioscience for Environment and Agriculture
The 5th Science Council of Asia (SCA) Conference, Ha Noi, Viet Nam
Richard Sinnott  
Director  
UK e-Science Centre  
At the 2nd Int'l Life Science Grid Computing Workshop  
LSGrid 2005, Biopolis, Singapore  

eScience, Systems Biology plant/crops... Environmental... All interconnected

Systems Biology?

+ links to plant/crops, environmental, health, ... information sources
1991

“Towards a Paradigm Shift in Biology”

- The new paradigm, now emerging, is... that the starting point of a biological investigation will be theoretical. An individual scientist will begin with a theoretical conjecture, only then turning to experiment to follow or test that hypothesis. *Walter Gilbert Nature (1991)*
What do some leading scientists think?

- Biology is in the middle of a major paradigm shift driven by computing.
  
  Eric Lander

- Computing has changed biology forever; most biologists just don’t know it yet.
  
  M Levitt

- Computational Biology will be as essential for the next quarter century of biology as molecular biology was for the past quarter century.
  
  W McGinnis
The world of scientific computing and advanced IT reached the level of being fully applicable to a wide range of deep biological research themes

*John Wooley, 2005*
Ecological and Environmental research

- Environmental Sensors for measuring physical, chemical, biological, meteorological, spatial, ecological parameters
- 24by7 monitoring of environmental and population events
- New technologies, methodologies and infrastructure for the environmental sciences

- Genes responsible for domestication of crops
- Mechanisms of polyploidization vs genome reduction
- Molecular mechanisms of symbiosis

From NSF Report CIBIO 2005

Symposium 1: Bioscience for Environment and Agriculture
The 5th Science Council of Asia (SCA) Conference, Ha Noi, Viet Nam
Workflow Integration for High Throughput research and Pipeline data handling architecture.

How they do it at the Joint Center for Structural Genomics (JCSG)
Workflow Integration in Laboratory Automation

GenBank
Genomic Data Source

swissprot
Proteomic Data Source

PCL CS4
Data Mining Tool

Data Visualization

User: BioProcesssing Technology Institute, Singapore

User: National Cancer Center, Singapore

The 5th Science Council of Asia (SCA) Conference, Ha Noi, Viet Nam
Cycle of the biological study with *in silico*

Wet Experiments → A Real Cell → Qualitative Modelling → Quantitative Modelling → Cell Programming → Run → Analyses → Interpretation

E-cell
- Model builder
- Algorithm modules
- Simulation visualiser
- Stanford BioX (US$ 150M)
- MIT CSBi (US$10M/yr)
- Princeton Sigler Inst for Integrative Genomics ICAHN Lab (US$40M)
- Duke Institute for Genome Sciences and Craig Venter’s TCAG (US$250M)
- UMichigan LSI (US$380M)
- QB3 UCaliforniaSF/Scruz/Berkeley (US$200M)
- Cornell LSI (US$140)
- UCSD JCSG
- NSF new CIBIO programme
Biology is Big Science these days

After the Genomes projects, industrial scale generation of data is no big deal.

Sophisticated bioinstrumentation from automated sequencers to microarray systems 24by7 churn out ever increasingly large scales of output, throughput and data generation.

Grid Computing tools are now available for supercomputing scale computing accessible from your desktop
APBioGrid 2002:
Asia Pacific’s role in Grid Computing for Life Science

in this issue

APBioNet Enlists Lion, Cray, KOOPrime
For Grid-Based Interoperability Project

THE ASIA Pacific Bioinformatics Network is bringing together a number of industry partners to build a grid-based system to share bioinformatics applications and workflows throughout the Asia Pacific region. The testbed project, a brainchild of Tan Tin Wee, secretariat of APBioNet and an associate profes-
Taverna-FreeFluo-myGrid workflow integration from UK eScience

Taverna Release: 1.3, 24th January 2005

Mailing Lists

Downloads

User Manual & Misc. Docs

Services & Components

Architecture and myGrid

Funding and Related Work

Follow-on Projects and Consumers

Papers, Presentations and other Outreach

Motivation & Background

Example Scenarios

Places, People & Collaborations

Taverna Scufl Workbench: http://taverna.res.net
Research Process Integration for Life Sciences

Data Management

Legacy Systems (e.g. HR, CRM, Analytical)

Application Management

Knowledge Management

Equipment

Symposium 1: Bioscience for Environment and Agriculture
The 5th Science Council of Asia (SCA) Conference, Ha Noi, Viet Nam

2005/6/22

www.koopprime.com
Laboratory Integration

- Allows users to select vendor plates for processing
- Generate in-house plates from vendor plates
- Print barcodes for each selected plate
- Start up legacy dispenser software
- Auto-import output files of dispenser into database
- Email user if there is any error in processing
Pipeline from R&D to Bio-Manufacturing of Diagnostics for Bird Flu and other emerging pathogens

Unknown Infectious Agent → Automated DNA Sequencing → Complete Genome/Sufficient Data → Semi-Automated Oligonucleotide Design

Automated Microarray Fabrication → Nucleotide Probes → Automated Oligonucleotide/Probe Synthesis → Lead Primer Targets

Prototype DNA Chips → Semi-Automated Diagnostic Assays

Yes → Large Scale Manufacture of DNA Chips

No → Re-Design of Probes

Deployable Diagnostic DNA Chips
A Typical Bio-Manufacturing Workflow

Recloning/amp
Chromo Walking
Gap Filling
Sample Processing
DNA Sequencing
Taverna
KOOPlatform

Primer Design

Homology Search w GRIDBLAST

Sequence Annotation
Probe Design
Generate Experiment
Microarray Data Mgt

Microarray Discovery
Microarray Visualization

Clinical Sample Prep
Probe List Prep

Hits List Generation
Diag Kits Manufacturing
21st Century BIO-Cyberinfrastructure

Changing How Science is Done
Providing the Tools to Swim in the Rapid Current of Data

From John Wooley et al 2005

Symposium 1: Bioscience for Environment and Agriculture
The 5th Science Council of Asia (SCA) Conference, Ha Noi, Viet Nam
The Language of Life Sciences

- “Calculus, in managing the infinitely small but large scale of events filled with redundancy, has been the language of the physical sciences.”

- “Biology has high information content, along with individuality, historicity and contingency… the biological sciences as a research discipline are said to be an information science. As such, information technology is the language of the life sciences, managing the discrete, non-symmetric, largely non-reducible, unique nature of biological systems and observations.”

John Wooley 2005
Biotech and InfoComm Technologies: Parallel Growth

Biotechnology

Genome Project
Microbial Genomes
Worm Genome
Dolly & DNA chips
Human Genome
Systems Biology

InfoCommunication Technology

Wais Gopher
WWW boom
ISP
Java
Dotcom boom
And crash
Grid Computing
Lambda Networking
Internet 2
Framework of Bioinformatics Development in Asia Pacific from 1991-2005

- **Policy**
- **Research**
- **Education & Manpower Training**
- **Compute Infrastructure**
- **Data Infrastructure**
- **Network Infrastructure**

Coordination

Planning
Training a new generation of Biologists for the Environment and Agriculture

- We must hook our individual computers in the worldwide network that gives us access to daily changes in the database and also makes immediate our communication with each other. The programs that display and analyze the material for us must be improved – and we must learn to use them more effectively. *Walter Gilbert Nature 1991*

- We must equip our students with the skills to carry tomorrow’s research today!
APBioBox APBioKnoppix collaboration to deliver free Computational Tools

- A/P Tan Tin Wee, National University of Singapore
- Adjunct Professor Shoba Ranganathan, NUS and Chair Professor, Macquarie University, Sydney
- Ong Guan Sin, Consultant programmer, Singapore Computer Systems Pte Ltd
- Funded by International Development Research Centre of Canada, under their PAN Pan Asia Networking ICT grants
CGIAR’s New Generation Challenge

- Programme to bring about a change in agricultural research and development
- Equip scientists with enabling skills
- Capacity building
- Crop Genomics and Informatics training programme
S* Life Science Informatics Alliance

- Stanford*Karolinska*Uppsala*
  SouthAfricanBioinformaticsInstitute (SANBI)*UCSD*NUS*USydney collaboration
- Online bioinformatics education since 2001 free of charge
- More than 1000 students have taken part
- Collaboration with AVIST
Conclusion

- Life Sciences, including environmental and agricultural sciences, have gone beyond simple sequence analysis of the 1980s and genome analysis of the 1990s.
- Computational aspects of biology will continue to increase.
- Database building, knowledge structuring and knowledge organisation and integration will continue.
- Imaging, Modeling and Simulation will emerge.
- Workflow integration and Pipelining will continue to accelerate high throughput, high efficiency research.

- In addition to observational/experimental, all biologists, including environmental and agricultural researchers, must be informationally and computationally competent in the 2000s.

“We must act now for the future”
Kiyoshi Kurokawa, 2005
Thank you SCA!

Thank you my hosts, Ha Noi, Viet Nam