LSU’s Center for Computation & Technology (CCT)

Joel E. Tohline, Director
An LSU research center whose mission, in part, is to infuse and enable computation – especially at the high end – into the forefront research and creative activities of all disciplines.

- **Faculty lines** – currently, 26 (avg. 50/50 split appointments) across 11 departments and 6 colleges/schools; tenure resides in home department
- **Cyber-Infrastructure** – guide LSU’s (and state’s via LONI) cyber-infrastructure design to support research ➔ high-performance computing (HPC), networking, data storage/management, & to some extent, visualization; also associated HPC support staff
- **Enablement staff** – currently 12 senior research scientists (non-tenured; ideally, on soft money support) with HPC expertise who support a broad range of compute-intensive research projects
- **Economic development** – to date, most significant interactions have been with Louisiana’s burgeoning digital media industry (e.g., video game design; visual effects)
- **Education** – Influence design and content of interdisciplinary curricula; for example: (1) computational sciences, (2) visualization, and (3) digital media
Brief Historical Perspective

PART 1
New Faculty Positions in Multiple Areas

This year, through Governor M. J. "Mike" Foster Jr.'s leadership, the Louisiana legislature approved $32 million in new, recurring funding for a statewide Information Technology (IT) initiative to build research capacity and promote economic development and diversification throughout the state. Governor Foster proposed this initiative to complement Baton Rouge, the state's capital city, for economic development, which calls for advancement in six key technology areas: computer science, chemical engineering, chemistry, electrical and computer engineering, mechanical engineering, and physics.

LSU is currently offering 42 new positions in the areas of computer science, chemistry, chemical engineering, and mechanical engineering. Applications are due by January 18, 2013. Visit the Office of Research & Graduate Studies, 3110 W.qus, Baton Rouge, Louisiana 70803, for more information on these positions.

Year: 2001
This year, through Governor M. J. “Mike” Foster Jr.'s leadership, the Louisiana legislature approved $22 million in new, recurring funding for a statewide Information Technology (I.T.) initiative to build research capacity and promote economic development and diversification throughout the state. Governor Foster proposed this initiative to complement Vision 20/20, the state’s master plan for economic development, which calls for advancement in six key technology areas (www.lded.state.la.us/new/vision2020/contents.htm). Higher education has been identified as the driver for economic development in the I.T. arena and, as the state’s flagship institution, Louisiana State University (LSU) has been charged with a leadership role, receiving $7 million in recurring funds in the current budget year, to rise to more than $9 million on July 1, 2002. LSU is working closely with four sister institutions—Louisiana Tech, the University of Louisiana at Lafayette, the University of New Orleans, and Southern University—that are also targeted for investment under the governor’s initiative.
Governor Kathleen Blanco announces that the State is committing $40 million to the Louisiana Optical Network Initiative (LONI) over the next 10 years.
Faculty-Driven Research Activities

PART 2
CCT Focus Areas

- Material World
- Coast-to-Cosmos
- Cultural Computing
- AVATAR
- System Sci. & Engineering
- Core Computation
CCT Focus Areas

- Material World
- System Sci. & Engineering
- Cultural Computing
- AVATAR
- Coast-to-Cosmos

Office of Research & Economic Development
Relevance to Biological Sciences
(current & near term)

• Faculty lines:
  – Michal Brylinski: 50/50 joint appointment w/ CCT; active involvement in “Material World” focus area; priority queue on SuperMike II
  – CCT has committed to help with startup funds in connection with a “computational biology / microbial metagenomics” search that is underway in Biological Science (Brylinski is on search committee)
HPC in Louisiana Higher Education

2002 : SuperMike : ~ $3M from LSU (CCT & ITS)
1024 cores; 3.7 Tflops

11th in Top500

2006 : Tezpur : ~ $$ from LSU (CCT & ITS)
1440 cores; 15.3 Tflops

2007 : Queen Bee : ~ $5M thru BoR/LONI (Gov. Blanco)
5440 cores; 50.7 Tflops;
Became NSF-funded node on TeraGrid

2007 : Queen Bee : ~ $5M thru BoR/LONI (Gov. Blanco)
5440 cores; 50.7 Tflops;
Became NSF-funded node on TeraGrid

2012 : SuperMike II : $2.65M from LSU (CCT & ITS)
7040 cores; 112 + 37.5 Tflops
SuperMike-II
Relevance to Biological Sciences
*(current & near term)*

- **Cyber-Infrastructure:**
  - Tezpur (LSU) and Queen Bee (LONI) available, free of charge to LSU researchers
  - SuperMike II recently installed at LSU
    - 440 compute nodes: at 16 cores per node  ➔ 7040 cores
    - 50 nodes contain attached pair of GPUs to accelerate suitable codes
    - 8 nodes are tied together via ScaleMP  ➔ even serial codes can see 2 TBytes of RAM
    - In principle, able to execute Windows OS applications
  - **Network infrastructure**
    - Working closely with LSU’s ITS and LONI to build more steerable and higher bandwidth network connectivity across the campus and state that is smoothly integrated with national research networks
  - **Data storage and management**
    - Working closely with LSU’s ITS and LONI to provide more adequate data storage and data management/ curation
Enablement Activities

PART 4
Relevance to Biological Sciences
(current & near term)

• Enablement research activities
  – Honggao Liu, CCT Deputy Director
  – James Lupo, assistant director: Takes the lead in answering any computational research questions that arise in connection with the use of LSU/LONI’s high-performance computing infrastructure
  – Jinghua Ge – visualization expertise; has supported campus visualization lab and has helped develop an Honors course heavily utilizing visualization tools across the sciences. Example, interaction with Professor Homberger’s research on anatomical kinematics of, e.g., birds and cats
  – Computational Biology & Bioinformatics Team: Currently 2 senior research scientists (Joohyun Kim and Nayong Kim) focused on assisting bioinformatics and broader computational biology efforts, especially in connection with LBRN = Louisiana Biomedical Research Network
  – CCT search underway to hire a “Senior Bioinformatics Computational Scientist”
CCT Computational Biology & Bioinformatics Team

Joohyun Kim and Nayong Kim
Genome Sequencing -> RNA-Seq -> ChIP-Seq

Base calling/Quality Filtering

Alignment, Assembly

Duplicate removal, quality score recalibration, realignment

Genome Variation
SNP, Small InDel, CNV, and Structure Variation

Transcriptome Profiling

Genome-wide DNA-protein Interaction & Methylation

Downstream Analyses
Software tools

R/Bioconductor/Biopython

Protein Gene Prediction: Glimmer, GenMark.Hmm-p
ncRNA Gene Finding: Infernal, CMFinder, RNAz, Evofold
Homology Sequence Match: exonerate, BLAST
DNA motif Finding: MEME
Comparative genomics: CGView, DAVID
Functional genomics: GSEA, pathway analyses
Microarray analysis: R/Bioconductor modules
SNP: diBayes (Bioscope), BFAST, SAMTools, SOAPsnp
CNV: (Bioscope) and others
Small InDel: (Bioscope), SAMTools and others
Mapping: SSAHA2, BFAST, BWA, SHRiMP2, Novoalign, Bowtie, MAQ, Stampy, SOAP2
De Novo Assembly: EDENA, NGS Cell, ABySS, Velvet
Misc (NGS Seq. Analysis): samtools, ARTEMIS, BamView
Misc (others): blast2GO, DAVID
RNA-Seq: TopHat/TopHat-fusion, Cufflinks, Scripture, OASES, Trinity, and others
ChIP-Seq: MACS, and many
Phylogeny: MrBayes and others
Molecular Dynamics: NAMD, CHARMM, Gromacs, LAMMPS, TINKER
Visualization tools: VMD, IGV, BamView, Gbrowse

• Genome Analysis Framework: Bioscope, GATK

* DARE-NGS - DARE (Dynamic Application Runtime Environment)-based Science Gateway
Next-Generation Seq. Data Bioinformatics Infrastructure

DNA Seq. Center

Ion Torrent

Server

System B : Visualization

System A : Computation
- Located at Frey Building

NAS

IT Storage
(1 TB : $1K/yr)

Remote Users

Cloud computing

Bioport

DARE-NGS

LONI
(project space)

* Four-tier Infrastructure
* Modular architecture
* Integrated service (compute/data)
* Scalable & Extensible by DARE-NGS
DARE provides abstractions to developers of science gateways. These abstractions allow developers and scientists to focus on the unique requirements of their scientific applications and relevant workflows as opposed to focus on the “plumbing” of how to submit ensembles of simulations to several supercomputers concurrently and archive their results. DARE is the natural evolution of science gateway middleware. As resource platforms, network capabilities and data repositories grow in size, number and vary in interface, the emergence of a unifying framework was inevitable. Many of the critical features of the DARE framework are provided by SAGA and the Pilot-Job capability: SAGA-BigJob SAGA demonstrated the capability (and usefulness) of overcoming utilization issues associated with distributed compute and data resources, complex multi-level workflows and run-time decision making. Building a science gateway framework on top of SAGA was the next logical step. The DARE framework’s distinguishing features include support for HPDC infrastructure and application/application workflow agnosticism.
**Available Services** – three different types

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Description</td>
<td>Standalone Single Tool</td>
<td>Pipeline Tool</td>
<td>Dynamic Workflow-based Tool</td>
</tr>
<tr>
<td>Example Target Application</td>
<td>Mapping</td>
<td>ChIP-Seq, RNA-Seq</td>
<td>ChIP-Seq, RNA-Seq</td>
</tr>
<tr>
<td>Example of Existing Tools</td>
<td>Bfast, BWA, Bowtie, ABySS</td>
<td>Mapping+MACS, TopHat-Fusion, Trans-ABySS, Hydra, GATK</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Upcoming Services**
- RNA-Seq pipelines
- Structural Bioinformatics : eThread
DARE-NGS

Scale out performance for DNA sequence mapping using BFAST on HPC

<table>
<thead>
<tr>
<th>Compute System</th>
<th># of cores</th>
<th># of tasks</th>
<th>Task Concurrency</th>
<th>Bfast (mapping)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workstation</td>
<td>4</td>
<td>8</td>
<td>No</td>
<td>≈18 h (or 72 h)</td>
</tr>
<tr>
<td>Ranger (HPC)</td>
<td>64</td>
<td>4</td>
<td>Yes</td>
<td>6.5 h</td>
</tr>
<tr>
<td>Ranger (HPC)</td>
<td>128</td>
<td>8</td>
<td>Yes</td>
<td>3.4 h</td>
</tr>
<tr>
<td>Ranger (HPC)</td>
<td>256</td>
<td>16</td>
<td>Yes</td>
<td>1.95 h</td>
</tr>
</tbody>
</table>

Scale out performance for DNA sequence mapping using BWA with Map-Reduce

Varying number of workers, input size = 10 GB, reduces = 8, Number of reads/ chunk=625000
Relevance to Biological Sciences
(current & near term)

- Enablement research activities
  - Honggao Liu, CCT Deputy Director
  - James Lupo, assistant director: Takes the lead in answering any computational research questions that arise in connection with the use of LSU/LONI’s high-performance computing infrastructure
  - Jinghua Ge – visualization expertise; has supported campus visualization lab and has helped develop an Honors course heavily utilizing visualization tools across the sciences. Example, interaction with Professor Homberger’s research on anatomical kinematics of, e.g., birds and cats
  - Computational Biology & Bioinformatics Team: Currently 2 senior research scientists (Joohyun Kim and Nayong Kim) focused on assisting bioinformatics and broader computational biology efforts, especially in connection with LBRN = Louisiana Biomedical Research Network
  - CCT search underway to hire a “Senior Bioinformatics Computational Scientist”
Relevance to Biological Sciences
(current & near term)

- Enablement research activities
  - Honggao Liu, CCT Deputy Director
  - James Lupo, assistant director: Takes the lead in answering any computational research questions that arise in connection with the use of LSU/LONI’s high-performance computing infrastructure
  - Jinghua Ge – visualization expertise; has supported campus visualization lab and has helped develop an Honors course heavily utilizing visualization tools across the sciences. Example, interaction with Professor Homberger’s research on anatomical kinematics of, e.g., birds and cats
  - Computational Biology & Bioinformatics Team: Currently 2 senior research scientists (Joohyun Kim and Nayong Kim) focused on assisting bioinformatics and broader computational biology efforts, especially in connection with LBRN = Louisiana Biomedical Research Network
  - CCT search underway to hire a “Senior Bioinformatics Computational Scientist”
Visiting Panelists
(February 2012)

Boore, Jeffrey
CEO, Genome Project Solutions, Inc.
Adjunct Professor, Department of Integrative Biology, UC-Berkeley
http://genomeprojectsolution.com/Personal_home_pages/Jeffrey_Boore.html

Cherbas, Peter
Director, Center for Genomics and Bioinformatics
Professor, Department of Biology, Indiana University
http://www.bio.indiana.edu/faculty/directory/profile.php?person=cherbas

Collins, Jack
Director, Advanced Biomedical Computing Center
National Cancer Institute, Frederick, Maryland
http://isp.ncifcrf.gov/abcc/abcc-staff/abcc-staff-bio/?id=12

Jongeneel, Victor
Director of Bioinformatics
Institute for Genomic Biology, University of Illinois
http://www.ncsa.illinois.edu/News/10/0301Jongeneeljoins.html

*Kissing, Jessica
Director, Institute of Bioinformatics
Associate Professor, Department of Genetics, University of Georgia
http://www.genetics.uga.edu/people_bio_kissing.html

Pfrender, Michael
Associate Professor, Department of Biological Sciences
Evolutionary & Ecological Genomics Lab, University of Notre Dame
http://biology.nd.edu/people/faculty/pfrender/

Quackenbush, John
Professor, Department of Biostatistics, Harvard School of Public Health
http://134.174.190.199/faculty/john-quackenbush/

Thomas, W. Kelley
Director, Hubbard Center for Genome Studies, University of New Hampshire
http://www.unh.edu/news/cj_nr/2011/feb/bp03genome.cfm

Wang, Yue (Joseph)
Director, Computational Bioinformatics and Bio-imaging Laboratory
Bradley Department of Electrical & Computer Engineering, Virginia Tech
http://www.ece.vt.edu/faculty/ywang.php
Strengthening Bioinformatics Research at PBRC and LSU
Expert Panel Recommendations
15-17 February 2012

2. Recruit and hire a senior scientist with experience in coordinating biologists and bioinformatics in a research center environment
   a. This hire should be placed in the CCT and charged with coordinating the more centralized bioinformatics model described in recommendation #1, especially to support the analysis of genome sequences
   b. This should be a joint appointment across Biological Sciences, CCT and PBRC
   c. This hire should coordinate the research and service activities of “professional” hires identified in recommendation #3b
   d. In stages, this hire should also develop and coordinate an outreach component across the state and outside the university system – see related recommendation #6
3. Hire a mix of researchers whose collective expertise spans a variety of areas of emphasis
   a. Make a few key tenure-track faculty hires that are “genome enabled” in order to increase usage of the genomics and bioinformatics cores and increase interactions between biologists and research scientists with computational expertise
   b. Develop “professional” positions that are not tenure-track but that are intended to be strongly collaborative and service oriented
      i. These positions should probably largely reside within the CCT
      ii. Service efforts should be coordinated by the senior scientist described in recommendation #2
      iii. To offset cost of additional staff positions, recognize the involvement of the CCT on grant proposals and direct additional F&A to the CCT when these are successful applications
   c. Consider making “research faculty” hires as well
   d. Lay out a thoughtful strategy regarding where new positions will reside and how appropriate joint appointments might stimulate interdisciplinary collaborations
Bioinformatics Hire Search Committee

- Brown, Jeremy (Biological Sciences)
- Canavier, Carmen (LSUHSC Biology & Anatomy)
- Kim, Joo (Biological Sciences)
- Macaluso, Kevin (SVM’s Pathobiological Sciences)
- Monroe, Todd (Biological & Agricultural Engineering) – committee chair
- Mores, Chris (SVM’s Pathobiological Sciences)
- Salbaum, Michael (Pennington Biomedical Research Center)
- Ullmer, Brygg (CCT and Computer Science)
Senior Bioinformatics Computational Scientist  
(draft advertisement)

- The Center for Computation & Technology (CCT) at Louisiana State University invites applications for a senior research scientist position in Computational Bioinformatics, broadly defined. The successful candidate will recruit and lead an Interdisciplinary Research Support Group (IRSG) that will support and integrate data-intensive and computationally demanding research activities across various academic units on LSU’s main campus, at the LSU School of Veterinary Medicine, the Pennington Biomedical Research Center, and LSU’s Health Sciences Centers. The IRSG will support research in genomics, bioinformatics, biostatistics, biomolecular structure/function, systems biology modeling, computational neuroscience, and other areas.

- The new leader of the IRSG will be charged with mobilizing this infrastructure to support the cutting-edge, interdisciplinary research activities described above. S/he will participate in and lead the development of extramural grant proposals. Equally important, s/he will develop programs to assist faculty and scientists in their use of bioinformatics and computational resources -- by individual mentoring and by workshops and tutorials. The IRSG leader will be encouraged to develop collaborative ties with industrial scientists across Louisiana.

- Required Qualifications: Ph.D. in biology, computational science, or a related area with emphasis on bioinformatics data analysis; five years of experience.

- Additional Qualifications Desired: Experience leading bioinformatics and biostatistics projects, teams and software use and development. Experience with common software development languages and tools, software design, and architecture and with the scripting tools commonly used by bioinformaticists: PERL, GALAXY, R/Bioconductor, etc. Experience with large dataset management specific to next-generation sequencing. Experience in the development of web interfaces to bioinformatics tools. Experience with high-performance computing, parallel programming and/or programming frameworks. Experience using virtual collaborative environments.

- Appointment and salary will be commensurate with experience and qualifications. This is a non-tenure track research position.
THANK YOU