ABSTRACT

Computational Biology Service Unit of the Cornell University Life Sciences Core Laboratories Center (CLC) provides an array of shared research resources and services to the university community and to outside investigators. The goal of the facility is to meet the increasing need of Cornell investigators for broad support in computational biology and bioinformatics.

INTRODUCTION

CBSU was founded in 2001 as a computational resource for the Tri-institutional Collaboration of Cornell University / Weill Cornell Medical College, Rockefeller University, and Memorial Sloan-Kettering Cancer Center, and later became part of the Cornell University Life Sciences Core Laboratories Center (CLC). In 2006, the CBSU was designated one of ten Microsoft High-Performance Computing Institutes worldwide.

Who can use our resources? The shared resources and services of the facility are open to all investigators at Cornell University. The facility supports research at all the university campuses, including the Ithaca campus, the Weill Cornell Medical College in New York City, and the Weill Cornell Medical College in Qatar. CBSU also provides resources and services to investigators at other academic institutions and to researchers at commercial enterprises. Some of the computational resources of the CBSU are freely available to researchers worldwide through web interface.

RESOURCES

The infrastructure of the facility includes the following resources: 2 dedicated MS Windows clusters (288 nodes, 48 Intel Xeon Dual Core 2.93GHz and 32GB RAM; 13TB disk array), 5 Linux clusters (124 nodes, AMD Opteron 248 2.2GHz, 32GB RAM, 6TB disk array), 1 hybrid cluster with 80 nodes of Sun V20Z with two AMD Opteron 248 2.2GHz, 2GB RAM, and 300GB HD; 5 MS Windows general purpose/web servers; 6TB file server; 6 general purpose Linux servers; 2 MS SQL servers; 1 workstation dedicated to sequence data processing Dell PWS T5400.

BioHPC: Computational Biology High Performance Computing Made Easily Accessible

BioHPC is a suite of computational biology applications on clusters (BioHPC), with web interface for easy access.

- Provides an interface for job, cluster, data and user administration
- The suite can be easily customized with new applications
- 40,000 job submissions over the last year by 3,500 users from 83 countries
- 130,000 jobs processed since June 2003

Research Highlights: Collaborative Program in Personalized Medicine

Genome wide association study of COPD

- CBSU supported an intercampus Cornell University / Weill Cornell Medical College large scale genome wide association study of Chronic Obstructive Pulmonary Disease (COPD).
- Developed a HIPPA compliant data management system which can be accessed both through web browser and Excel web service client.

Research Highlights: Comparative Genomics and Proteomics Studies

Plant C3/C4 Pathways

- CBSU and 5 groups from Cornell, Yale University and Iowa State University are working together on comparative studies of C3 and C4 photosynthesis pathways, using rice, maize and sorghum as models for the different pathways.

Research Highlights: Support for Multiple Next Generation Sequencing and Genotyping Platforms

- CBSU developed data analysis pipelines for next generation sequencing and genotyping platforms from Roche, Illumina and Affymetrix.
- We have implemented sequence and genotyping data analysis tools on our HPC clusters. We are implementing access for next generation sequencing software application on our BioHPC portal. We are in the process of acquiring a large memory system for genotyping, copy-number-variation study and genome assembly projects.
- We are working together with multiple laboratories to develop novel applications using the next generation sequencing technologies.

Research Algorithms Made Accessible via BioHPC

- CBSU helps Cornell scientists make their algorithms available to the research community through the web interface of BioHPC.
- Establishes web presence and facilitates wide use of new software tools in research.
- CBSU hosts 12 programs developed at Cornell: 7 programs via BioHPC (GIMASAN, SFS, CODE, InStruct, MKPRF, MDV, LOOPP, RepeatFinder) and 5 via their own websites (I-REX, iDeCal, PathogenTracker, PPDDB, URMS).
- 6,500 submissions last year; 31,000 since 2003

Contact Information
Computational Biology Service Unit
Jaroslaw Pillardy, Director
Life Sciences Core Laboratories Center
Cornell University, Ithaca, NY
http://www.cbsu.tc.cornell.edu