Contribution ID: **172**   Type: **Oral**

The Nordic BioGrid project –Bioinformatics for the grid

Tuesday 13 April 2010 12:00 (15 minutes)

Life sciences have undergone an immense transformation during the recent years, where advances in ge-nomics, proteomics and other high-throughput techniques produce floods of raw data that need to be stored, analysed and interpreted in various ways. Bioinformatics is crucial by providing tools to efficiently utilize these gold mines of data in order to better understand the roles of proteins and genes and to spark ideas for new experiments.

Conclusions and Future Work

The BioGrid has already contributed to provide computational power for analysis of the medium-chain de-hydrogenase/reductase (MDR) superfamily. The size and complexity of this superfamily has recently been shown to far surpass the means of subclassification that have traditionally been employed for this task. In-stead, more computationally demanding methods must be employed, such as profile Hidden Markov Models, implemented in the HMMer package.

Impact

Regarding databases, the frequently used databases UniProtKB and UniRef have been made available on the distributed and cached storage system within the Nordic grid. A system for database updating has been deployed in a virtual machine hosted by NDGF. The database PairsDB updates have been run on BioGrid & M-grid resources. Further applications are in the pipeline to be gridified including molecule dynamics and phylogeny calculations.

Keywords

bioinformatics, grid, hidden Markov models, large-scale analyses, distributed storage

URL for further information

http://wiki.ndgf.org/display/ndgfwiki/BioGrid

Detailed analysis

BioGrid is an effort to establish a Nordic grid infrastructure for bioinformatics, supported by NDGF (Nordic DataGrid Facility). BioGrid aims both to gridify computationally heavy tasks and to coordinate bioinformatics infrastructure efforts in order to use the Nordic resources more efficiently. Hitherto, the widely used bioin-formatics software packages BLAST and HMMer have been gridified. Furthermore, the multiple sequence alignment programs ClustalW, MAFFT and MUSCLE have been made available on the grid.

Primary authors: Dr BERGLUND SONNHAMMER, Ann-Charlotte (Linnaeus Center for Bioinformatics (LCB), Uppsala University, S-751 05 Uppsala, Sweden); Prof. PERSSON, Bengt (IFM Bioinformatics, Linköping University, S-581 83 Linköping, Sweden; Dept of Cell and Molecular Biology, Karolinska Institutet, S-171 77 Stockholm, Sweden; National Supercomputer Centre (NSC), S-581 83 Linköping, Sweden); Prof. SONNHAMMER, Erik (Stockholm Bioinformatics Center (SBC), Stockholm University, S-106 91 Stockholm, Sweden); Prof. JONASSEN, Inge (CBU, Bergen Centre for Computational Science, N-5020 Bergen, Norway); HEDLUND, Joel (IFM Bioinformatics, Linköping University, S-581 83 Linköping, Sweden); Dr KLEIST, Josva (Nordic Data Grid Facility, Kastruplundgade 22, DK-2770 Kastrup, Denmark); Dr MATTILA, Kimmo (CSC –IT Center for Science Ltd, P.O. Box 405, FI-02101 Espoo, Finland); Dr GRØNAGER, Michael (Nordic Data Grid Facility, Kastruplundgade 22, DK-2770 Kastrup, Denmark); TOURUNEN, Olli (Nordic Data Grid Facility); Dr MÖLLER, Steffen (Institut für Neuro- und Bioinformatik, University of Lübeck, D-23538 Lübeck, Germany)

Presenters: Prof. PERSSON, Bengt (IFM Bioinformatics, Linköping University, S-581 83 Linköping, Sweden; Dept of Cell and Molecular Biology, Karolinska Institutet, S-171 77 Stockholm, Sweden; National Supercomputer Centre (NSC), S-581 83 Linköping, Sweden); HEDLUND, Joel (IFM Bioinformatics, Linköping University, S-581 83 Linköping, Sweden)

Session Classification: Bioinformatics

Track Classification: Experiences from application porting and deployment