

BecA-ILRI Bioinformatics Platform

Capacity Building

Bioinformatics is a relatively new specialist area. We need to raise awareness of the field here in Africa and expose local scientists to its potential. The West has spent millions of dollars sequencing the genomes of humans, animals, plants and parasites and the resulting data are freely available on the Internet. This is a vast body of knowledge that local scientists can use to solve their specific problems or to answer research questions. All scientists in Africa need to make use of these data are a computer, good Internet access and bioinformatics skills.

Raising awareness and building capacity in bioinformatics in Africa is a major focus of the bioinformatics Hub. This is accomplished through regular training courses and research projects undertaken at the BecA Hub. In association with the Swedish University of Agricultural Sciences and the Linnaeus Centre for Bioinformatics, in Uppsala, Sweden, ILRI and the BecA Hub run an annual two-week "Introduction to Bioinformatics" training course. Since 2005 we have trained more than 200 students from more than seven eastern and central Africa countries in the field of bioinformatics.



For further information please contact:

Bioinformatics Group Leader:

Etienne de Villiers: e.villiers@cgiar.org

Unix specialist

Alan Orth: a.orth@cgiar.org

BecA-ILRI Bioinformatics Platform



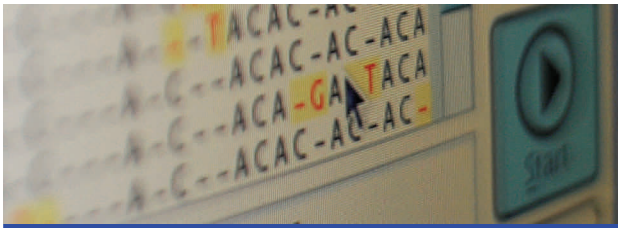
ILRI
INTERNATIONAL
LIVESTOCK RESEARCH
INSTITUTE

biosciences
eastern and central africa

mobilizing biosciences for Africa's development

The BecA Hub/ILRI Bioinformatics Platform provides advanced computational capabilities in Bioinformatics to all scientists at the BecA Hub to uncover the wealth of biological information hidden in the mass of DNA sequences, structure, literature and other biological data and to obtain a clearer insight into the fundamental biology of organisms and to use this information to enhance the standard of life for mankind.

The initial focus of the platform is to be an easy-to-use platform for advanced sequence analysis. We make available high-speed, specialized hardware and sophisticated commercial and academic software. To make these useful, we maintain the latest copies of all major sequence databases. Investigators can use these components to perform one-time projects or to build up their own ongoing analysis pipeline.



Services

We provide scientific and technical staff that can assist with or collaborate on individual projects. As part of the Bioinformatics Platform, we also have access to the advice and expertise of associated ILRI staff members. The Bioinformatics Platform is a specialist node of The European Molecular Biology Network (EMBNet). EMBNet is a consortium of collaborating bioinformatics groups located in Europe and a number of groups outside Europe. Each group is represented by a 'node', and is responsible for the maintenance of local services for their users (e.g. education, training, software, database distribution, technical support, helpdesk). The combined expertise of the nodes allows EMBNet to provide services to the bioinformatics community that encompasses more than can be provided by a single node.

Hardware

The Bioinformatics Platform is hosted on a High Performance Computer (HPC) platform located on the ILRI campus.



The HPC platform is a 66 CPU Linux cluster, based on AMD 64-bit Opteron technology and build with ROCKS version 5.2. Each compute node contains two 64-bit Opteron processors with 2 GB of memory. In total the system has 72 GB of memory and more than 3 TB of disk space. Platform LSF scheduling software manages job scheduling.

Paracel BlastMachine runs parallelized versions of NCBI BLAST, PSI-BLAST, and MegaBLAST. It utilises 20 of the 66 CPUs on the HPC. The BlastMachine accelerates the BLAST algorithms by dividing a job across multiple processors and can divide both the database and the query among groups of processors.

Fast Internet access to advanced research and education networks such as Internet2 and GEANT.

Some things you can do using the bioinformatics platform:

- Quickly run very large BLAST jobs on our 20-CPU BlastMachine cluster.
- Next Generation sequence data analysis pipelines including:
 - re-sequencing and mapping reads to reference genome(s),
 - *de novo* assembly of genomes,
 - detection of foreign genomic material in the presence of host genome
- Develop and run parallel MPI jobs on the 66 CPU Linux cluster, or use it as a compute farm for serial jobs.
- Sophisticated, commercial-quality sequence and EST assembly.
- We maintain the latest copies of all major sequence databases for each of our platforms - you can set up your own pipeline to run a set of queries at regular intervals against the latest data.
- Choose from our growing list of academic software packages supported on our machines, including packages for genetic analysis and structural biology.
- Web-based and UNIX login access to the various machines. We can host large-scale projects. We are working on technologies to allow remote use of our machines.

