In silico comparative genomics

Research activities pursued in this area have a direct bearing on genomics projects and have led to several joint publications with laboratory staff. Activities include the evaluation of algorithms for orthology detection and phylogenomic studies of protein family orthologs related to the plant abiotic stress responsive genes.

Capacity building

Over the past few years, there has been increased interest to work in ICRISAT's Bioinformatics Unit by students from universities and institutions in India. Almost all projects are in the area of comparative biology and phylogenomics. These students are working towards a degree in bioinformatics or biotechnology. In-house capacity is also being strengthened with the help of private sector partners and through participation in Community open source software development initiatives.

Linkages and partnerships

ICRISAT cannot do it alone, so seeks out consultants and third parties that can provide the required expertise in bioinformatics. One such company is the Tata Consultancy Services (TCS), a leading global information technology consulting, services and business process outsourcing organization. Given our successful subcontracting of development of parallelized software for the high performance computer at ICRISAT, we have wanted to extend our linkages with companies like these. We have been working with scientists of the Advanced Technology Centre, the R&D wing for bioinformatics at TCS, to build partnerships in the areas of software platform development, high-throughput comparative biology and systems biology.

Highlights of achievements

The Bioinformatics Unit at CIMMYT – a sister CGIAR center – conducted an evaluation of various LIMS applications in early 2006 and regarded ICRISAT's LIMS to be superior. Since then the application has been downloaded almost 500 times. Besides, the iMAS software has seen 123 downloads and several national program users. Other software both statistical as well as for sequence analysis are being downloaded and used by various users from state, national and international universities and research centers. The unit is also an active node in the challenge program led initiatives in the area of information systems and molecular breeding information management platform along with sister centers as well as ARI/ARCs.

About ICRISAT

ICRISAT
Science with a human face

The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) is a non-profit, non-political organization that does innovative agricultural research and capacity building for sustainable development with a wide array of partners across the globe. ICRISAT's mission is to help empower 644 million poor people to overcome hunger, poverty and a degraded environment in the dry tropics through better agriculture. ICRISAT belongs to the Alliance of Centers of the Consultative Group on International Agricultural Research (CGIAR).

Company Information

ICRISAT-Patancheru
(Headquarters)
Patancheru 502 324
Andhra Pradesh, India
Tel +91 40 30713071
Fax +91 40 30713074
icrisat@cgiar.org

ICRISAT-Liaison Office
CG Centers Block
NASC Complex
Dev Prakash Shastri Marg
New Delhi 110 012, India
Tel +91 11 32472306 to 08
Fax +91 11 25841294

ICRISAT-Nairobi
(Regional hub ESA)
PO Box 39063, Nairobi, Kenya
Tel +254 20 7224550
Fax +254 20 7224001
icrisat-nairobi@cgiar.org

ICRISAT-Niamey
(Regional hub WCA)
BP 12404, Niamey, Niger (Via Paris)
Tel +227 20722529, 20722725
Fax +227 20734329
icrisatsc@cgiar.org

ICRISAT-Bamako
BP 320
Bamako, Mali
Tel +223 20 223375
Fax +223 20 228683
icrisat-w-mali@cgiar.org

ICRISAT-Bulawayo
Matopos Research Station
PO Box 776,
Bulawayo, Zimbabwe
Tel +263 83 8311 to 15
Fax +263 83 8253/8307
icrisatzw@cgiar.org

ICRISAT-Lilongwe
Chitedze Agricultural Research Station
PO Box 1096
Lilongwe, Malawi
Tel +265 1 707297/071/067/057
Fax +265 1 707298
icrisat-malawi@cgiar.org

ICRISAT-Maputo
c/o IIAM, Av. das FPLM No 2698
Caixa Postal 1906
Maputo, Mozambique
Tel +258 21 461657
Fax +258 21 461581
icrisatmoz@panintra.com

www.icrisat.org

Jan 2009

Bioinformatics

A platform for ICRISAT's global research needs

ICRISAT

INTERNATIONAL CROPS RESEARCH INSTITUTE FOR THE SEMI-ARID TROPICS
Science with a human face

Background

ICRISAT's Global Theme on Biotechnology employs a range of modern genomic technologies to enhance the efficiency and effectiveness of crop improvement. The rate-limiting step in genomics is no longer data generation but rather the speed at which, data is captured, validated, analyzed and turned into useful knowledge. The role of Bioinformatics is to remove this rate-limiting step through the development of software platforms for handling large volumes of data generated, and facilitate its analysis. The term 'platform' implies all those vital services and technologies that are needed to support genomics research projects at all of ICRISAT locations. Bioinformatics at ICRISAT is for both research and research support.

Research efforts focus on two major areas. One relates to infrastructure development, which includes: (a) the development of appropriate information systems for data capture, storage, retrieval and dissemination; validating the large volumes of data with special emphasis on data quality, and (b) the development of analysis tools to accelerate research efforts in molecular marker discovery, annotation and comparative genomics.

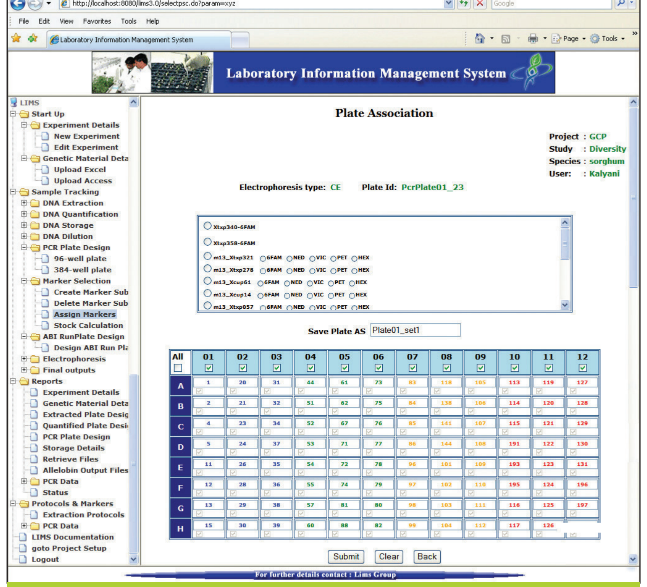
The hardware infrastructure in Bioinformatics has been steadily growing since 2004. There are three middle level servers, one Paracel Linux cluster comprising five dual AMD Opteron processors (for high-end computing jobs), four workstations and 18 Pentium IV PCs. The open source movement and its philosophy have largely driven platform software development at Bioinformatics. Almost all the software development is done using the freely available Perl or Java programming languages. Since academic and research institutions have taken to the open source culture there are a number of bioinformatics software packages available in the public domain. All the data analysis pipelines constructed at ICRISAT have been built upon freely available software. All tools developed here are put back into the public domain for others to use and extend.

Activities

Bioinformatics ativities are grouped under four sections: infrastructure development, comparative genomics, capacity building, and linkages and partnerships.

Infrastructure development

(a) A Laboratory Information Management System (LIMS) , meets the needs of a moderately high throughput molecular genotyping facility. This software project was initiated in 2004, and the beta version 2.7 is currently in use. The application is continually updated to keep pace with changing laboratory needs. Functional modules include sample tracking, job scheduling that allows users to book time slots on available sequencing machines, report generation etc. Users access the LIMS through their browsers. All genotyping information generated is centrally stored. The LIMS interfaces with laboratory instruments through file exchange, keeping track of the sample from its source to the final output data. The application was successfully transferred to the Biosciences Eastern and Central Africa (BeCA) facility at Nairobi and the International Institute for Tropical Agriculture (IITA)-Ibadan and has been customized by a private sector partner for their use.


http://localhost:8080/lims3.0/selectpsc.do?param=xyz
Google
File Edit View Favorites Tools Help
Laboratory Information Management System
Page Tools

Laboratory Information Management System

- LIMS
  - Start Up
    - Experiment Details
      - New Experiment
      - Edit Experiment
    - Genetic Material Deta
      - Upload Excel
      - Upload Access
  - Sample Tracking
    - DNA Extraction
    - DNA Quantification
    - DNA Storage
    - DNA Dilution
    - PCR Plate Design
      - 96-well plate
      - 384-well plate
    - Marker Selection
      - Create Marker Sub
      - Delete Marker Sub
      - Assign Markers
      - Stock Calculation
    - ABI RunPlate Design
      - Design ABI Run Pla
    - Electrophoresis
    - Final outputs
  - Reports
    - Experiment Details
    - Genetic Material Deta
    - Extracted Plate Desig
    - Quantified Plate Desi
    - PCR Plate Design
    - Storage Details
    - Retrieve Files
    - Allelobin Output Files
    - PCR Data
    - Status
  - Protocols & Markers
    - Extraction Protocols
    - PCR Data
  - LIMS Documentation
  - goto Project Setup
  - Logout

Plate Association

Project : GCP
Study : Diversity
Species : sorghum
User: : Kalyani

Electrophoresis type: CE   Plate Id: PcrPlate01_23

○ Xtxp340-6FAM
○ Xtxp358-6FAM
○ m13_Xtxp321 ○6FAM ○NED ○VIC ○PET ○HEX
○ m13_Xtxp278 ○6FAM ○NED ○VIC ○PET ○HEX
○ m13_Xcup61 ○6FAM ○NED ○VIC ○PET ○HEX
○ m13_Xcup14 ○6FAM ○NED ○VIC ○PET ○HEX
○ m13_Xtxp057 ○6FAM ○NED ○VIC ○PET ○HEX

Save Plate AS Plate01_set1

All □	01 ☑	02 ☑	03 ☑	04 ☑	05 ☑	06 ☑	07 ☑	08 ☑	09 ☑	10 ☑	11 ☑	12 ☑
A	1	20	31	44	61	73	83	118	105	113	119	127
B	2	21	32	51	62	75	84	138	106	114	120	128
C	4	23	34	52	67	76	85	141	107	115	121	129
D	5	24	37	53	71	77	86	144	108	191	122	130
E	11	26	35	54	72	78	96	101	109	193	123	131
F	12	28	36	55	74	79	97	102	110	195	124	196
G	13	29	38	57	81	80	98	103	111	116	125	197
H	15	30	39	60	88	82	99	104	112	117	126	

Submit Clear Back

For further details contact : Lims Group

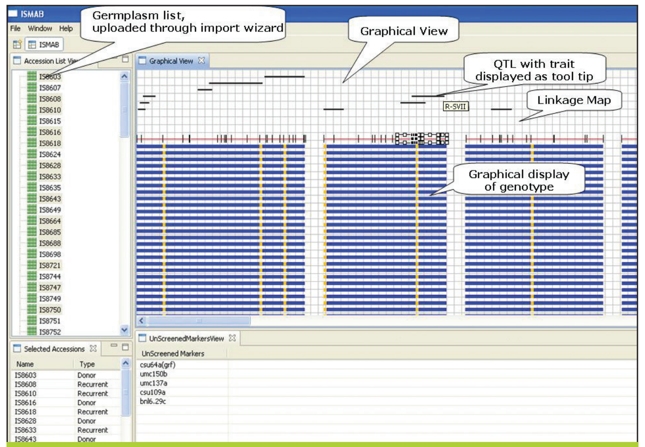
LIMS User Interface. The functional modules are arranged in a hierarchical fashion on the left. Each submodule represents a step of the genotyping workflow. The UI shown here allows assignment of molecular markers to a PCR plate.

(b) The Integrated ICRISAT Crop Resources Information System (ICRIS) database is available to ICRISAT users on the institute Intranet as well as password protected access from http://icris.icrisat.org/ICRIS/. The database currently stores genotype, marker and phenotype information. A suite of web pages allow the user to both browse and query the database, and retrieve results as germplasm x marker or germplasm x trait matrices. Users are provided with a number of data export formats that will allow downloaded data to be immediately used with relevant analysis tools. Submission of data to the database by data generators is enabled through templates.

(c) The Information System for Marker Assisted Breeding (ISMAB) is a standalone system being developed to integrate information relevant to the breeding program that may come from local and/or external databases. The system has been designed with four basic components. These include a Molecular Breeding Design Tool (MBDT) that will assist breeders in selecting parental germplasm based on phenotypic data, allow user to choose markers and design crosses and the target genotype. The second component interfaces with the LIMS to carry out the sample tracking for germplasm. The third component is the MOlecular Selection Tool (MOSEL) that facilitates the selection of the most promising lines in terms of closeness to the target genotype. The fourth component is the loading of data into the central database, for future use and dissemination.

Mr Mallikarjun, programmer, demonstrates links to statistical software from the iMAS home page.

In the first year of its development, the first component of the system has been designed and developed. The MBDT module currently allows the import of files consisting of genotype data, linkage maps or QTL data through a file import wizard. The user can choose colors to display allele information in the graphical genotypes, select foreground markers based on a segment of interest. The module allows the selection of donor and recurrent and design of the desired genotype.



IS8603
IS8607
IS8608
IS8610
IS8615
IS8616
IS8618
IS8624
IS8628
IS8633
IS8635
IS8643
IS8649
IS8664
IS8685
IS8688
IS8698
IS8721
IS8744
IS8747
IS8749
IS8750
IS8751
IS8752

Name	Type
IS8603	Donor
IS8608	Recurrent
IS8610	Recurrent
IS8616	Donor
IS8618	Recurrent
IS8628	Donor
IS8633	Recurrent
IS8643	Donor

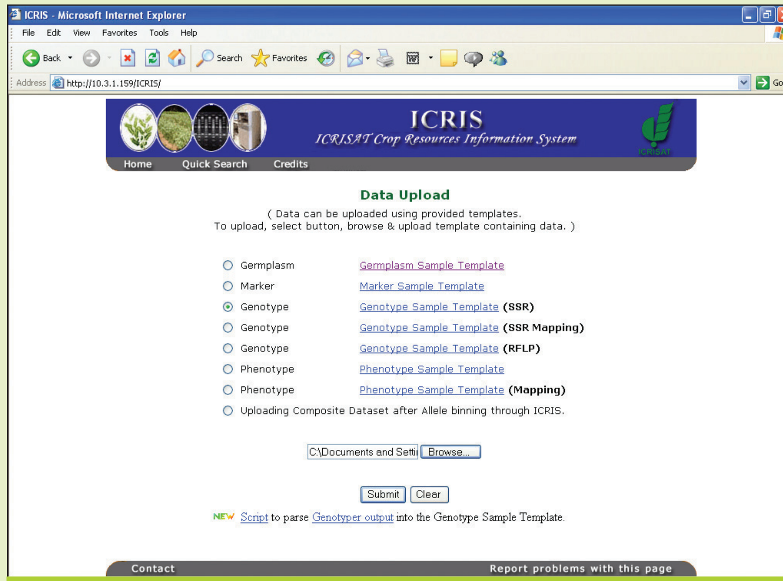
UnScreened Markers
csu64a(grf)
umc150b
umc137a
csu109a
bnl6.29c

The first component of ISMAB. Developed using the eclipse Graphical Editing Framework, the tool will allow visual analytics for genotype and associated phenotype data.

(d) The integrated decision support system, called iMAS, facilitates marker-assisted plant breeding by integrating freely available quality software needed for experimental design, QTL mapping,and visualization, providing simple-to-understand-and-use online decision guidelines that help the user interpret results. Into its fifth year of the software life cycle the standalone application serves as a useful training tool.

(e) High performance computing toolbox: The Paracel Linux cluster hosts pipelines (a series of different software through which data is pipelined) and standalone software analysis tools relevant to comparative genomics and population genetics. Several workflow/pipelining environments are available in the public domain and one such environment has been implemented on the Paracel HPC at ICRISAT. The Pasteur Institute Software Environment (PISE) allows integration of in-house scripts/tools along with external tools that a user may need for his analysis. Ease of use is achieved through the creation of a graphical user interface (GUI) for all of the programs/scripts available in the environment and the chaining together of scripts to facilitate automation and analysis. Over 40 analysis tools are currently available for sequence analysis through this environment. Besides pipelines for the analysis of Next Generation Sequencing (NGS) data is also being implemented and made available through the HPC.

All software developed are available for download from the GT-Biotechnology software downloads page (http://www.icrisat.org/gt-bt/softwares_downloads.htm).


ICRIS - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://10.3.1.159/ICRIS/
Go
ICRIS
ICRISAT Crop Resources Information System
ICRISAT
Home Quick Search Credits
Contact
Report problems with this page

Data Upload

(Data can be uploaded using provided templates.
To upload, select button, browse & upload template containing data.)

○ Germplasm   Germplasm Sample Template
○ Marker   Marker Sample Template
● Genotype   Genotype Sample Template (SSR)
○ Genotype   Genotype Sample Template (SSR Mapping)
○ Genotype   Genotype Sample Template (RFLP)
○ Phenotype   Phenotype Sample Template
○ Phenotype   Phenotype Sample Template (Mapping)
○ Uploading Composite Dataset after Allele binning through ICRIS.

C:\Documents and Setti Browse...

Submit Clear

NEW Script to parse Genotyper output into the Genotype Sample Template.

The ICRIS User Interface displays available genotype and phenotype data templates, though which the user can upload data.