

## Supercomputing puts Indian agriculture on fast-track mode

12 December 2013 | Features | By Rahul Koul Koul

### Supercomputing puts Indian agriculture on fast-track mode



Nobel laureate and noted scientist, Dr Walter Gilbert once remarked that most of the biological investigations in 21st century will be in silico. Proving him right, the advent of genetic engineering and genomic approaches, have opened new vistas for increasing the productivity and quality attributes of bio-systems. During the last one decade, genomics has witnessed an information explosion and creation of databases, not amenable to traditional analytical approaches. Here came the role of bioinformatics that emerged as an inter-disciplinary programme, linking computational and mathematical sciences with life sciences.

ASHOKA is the name of India's first supercomputing hub for Indian Agriculture has been established at Centre for Agricultural Bioinformatics (CABin), that stands for advanced super-computing hub for OMICS knowledge in agriculture. The CABin is a part of the Indian Agricultural Statistics Research Institute (IASRI), New Delhi, involved in the development of partnerships at various levels among national and international organizations in the field of bioinformatics and related fields. The main centre for CABin will have dedicated high speed connectivity to domain centres working in different domains of agriculture such as crops, animals, fisheries and agricultural microbes. Further with the active support from agriculture ministry and fruitful partnership with the CDAC, Pune, the IASRI is setting up a nationwide grid of supercomputers for agri-biotech research, agricultural planning and research. Aim is to bring the biologists, statisticians and computer scientists together from the point of view of system biology approach and effective problem solving. This computational facility will not only for researchers of the Indian Council of Agricultural Research (ICAR) but all agri- scientists across the nation.

Although, activities related to bioinformatics were initiated at different ICAR institutions at small scale in isolated mode, hardly any coordinated efforts were made to integrate these activities at national level in the field of agriculture. Therefore, this supercomputing environment is being developed for high performance computing in the field of agricultural bioinformatics and computational biology under a sub-project "Establishment of National Agricultural Bioinformatics Grid (NABG) in ICAR" of the National Agricultural Innovation Project (NAIP), Indian Council of Agricultural Research (ICAR) New Delhi. The facility is set up in a state-of-art data centre and two super-computers of this hub are listed at rank 11 and 24 in the Indian Institute of Sciences (IISc) list of top super-computers of India.

This super-computing hub consists with hybrid architecture of high performance computing having (i) 256 nodes Linux cluster with two masters with 3072 cores and 38 Tera Flops computing, (ii) 16 nodes windows cluster with one master, (iii) 16 nodes GPU cluster with one master with 192 CPUs + 8192 GPUs and (iv) SMP based machine with 1.5 TB RAM. Also, this hub has approximately 1.5 Peta Byte storage divided in to three different types of storage architecture i.e. Network Attached Storage (NAS), Parallel File System (PFS) and Archival. This hub also consists of super-computing systems ( 16 node Linux cluster with one master with 40 TB storage) at National Bureau of Plant Genetic Resources (NBPGR) New Delhi, and Lucknow, National Bureau of Agriculturally Important Microbes (NBAIM) Mau and National Bureau of Agriculturally Important Insects (NBAII), Bangalore which forms a National Agricultural Bioinformatics Grid in the country. Number of computational biology and agricultural bioinformatics software/workflow/pipelines along with National Biological Computing Portal are in the process of development, which will provide seamless access to these biological computing resources to the biological researchers across the country.

According to Dr Anil Rai, head and principal scientist, CABin, "We are trying to build this system so that scientists don't have to come here every-time for the analysis but to ensure that they can carry out the same while sitting on their desktops. For that we are building national bioinformatics portal which is almost 80 percent ready. There is a provision for monitoring of the data results by respective scientists regularly and even sms alerts to provide quick info on progress is also there. This system will support computational requirements of the biotechnological research in the country. This will also bridge the gap between genomic information and knowledge, utilizing statistical and computational sciences. Further, this will help in establishment of large genomic databases, data warehouse, software & tools, algorithms, genome browsers with high-end computational power to extract information and knowledge from cross-species genomic resources."

Dr Dinesh Kumar, senior scientist (biotechnology), CABin believes that it is a right step forward."It will open up new vistas for downstream research in bioinformatics ranging from modelling of cellular function, genetic networks, metabolic pathways, validation of drug targets to understand gene function and culminating in the development of improved varieties and breeds for enhancing agricultural productivity to many folds," he mentioned in a visibly excited tone.

### **Creation of a national level grid**

It is proposed to link all projects and activities in the field of agricultural bioinformatics in the country to the National Agricultural Bioinformatics Grid (NABG), comprising six national Institutes under the aegis of CABin. This network of institutions is based on two-tier architecture and it will be able to bridge the gap between genomic information and knowledge, utilizing statistical and computational sciences. Also, this model will help in the development of partnerships at various levels among national and international organizations. Further, this will also establish functional linkages among researchers and scientists in the field of bioinformatics and related fields. CABin will also facilitate the opening up of new vistas for downstream research in bioinformatics ranging from modelling of cellular function, genetic networks, metabolic pathways, validation of drug targets to understand gene function and culminating in the development of improved varieties and breeds for enhancing agricultural productivity. It will have high-end super computational facilities for voluminous data warehousing, processing and interpretation.

Dr U C Sud, director, IASRI feels that the move will bring about new paradigms in agriculture. "The infrastructure build by ICAR over the period of time has resulted in huge transformation. We have gained enormous expertise in the area of statistics and data analytics. The creation of NABG will further help in streamlining of huge set of data that has been accumulated through research efforts of Indian scientists. I believe that the results shall be overwhelming."

The second tier of this grid is five Domain Bioinformatics Centres (DBC) connected through dedicated high-speed connectivity with CABin. These centres will cater to the needs of the institutions working in different domains of agriculture such as crops, animals, fisheries, agricultural microbes and agricultural insects. The DBC's will provide services and support to institutions in their respective fields of research. The centres are also responsible for information generation and knowledge delivery from CABin to respective domain institutions for conducting experiments. The DBC's will maintain and

verify data quality of genomic information from all sources in their respective domains. The domain centre of this architecture will be responsible for the collection of raw genomic data from laboratories of their respective domain and share information with CABin. The in silico findings in genomic data will be tested and validated through wet lab experiments at the respective domain institutions.

The CABin is at IASRI, New Delhi. The DBC's are in NBPGR, New Delhi (Crops), NBAGR, Karnal(Animals) NBFGR, Lucknow(Fisheries) and NBAIM Mau (Agricultural Microbes) and NBAIL, Bangalore (Insects). The CABin and DBC's of National Agricultural Bioinformatics Grid are connected through high speed Wide Area of Network (WAN) for seamless information flow. Initially it is proposed to connect the main centre by 64 Mbps and DBC' 32 Mbps bandwidths.

Since establishment of National Agricultural Bioinformatics Grid (NABG) in ICAR needs state of art civil work, air-conditioning systems, security systems, designing of High Performance Computing Centre (HPC), procurement of advance hardware and software for supercomputing and their integration suitable for various application areas of bioinformatics. This requires intensive experience and expertise in different fields of science and technologies. Therefore, it was decided to invite International Competitive Bid (ICB) to carry out the task for implementation of this complete High Performance Computing (HPC) Systems which includes computing resources, data centre for environmental and security maintenance and monitoring high end storage systems and other related facilities. Apart from this, this system needs to be configured, tuned and optimized for undertaking biological computational research in advance area of agricultural bioinformatics by the researchers through on-line Web portal for routine jobs. In this all available applications and software needs to be configured and ported on parallel processing environment for efficient utilization of computing resources. Further, to specific project based problems tools and techniques will be developed and implemented.

### **The technical facilitator**

In order to access the grid services, general scientific user communities need to deal with the underlying system which makes them to be burdened with the computational aspect rather than their research area, as most of the scientific users often lack specialized expertise to deal with the complex HPC environments. Therefore, the web based interface which shall be developed by Centre for Development of Advanced Computing (C-DAC), Pune, will be an obvious choice for the end user due to the familiarity with regular internet usage. Web browser-based portal user interfaces provide access to a large variety of resources, services, applications, and tools. Additionally, grid-enabled portals can deliver complex grid solutions to users, wherever, they have access to a web browser running on the internet without need to download or install specialized software or take care of setting up networks, firewalls, and port policies.

"The CABin shall help various domain centres for taking up research and development activities in their respective areas. These institutions will also be responsible for information collection and generation of knowledge of the respective domain which in turn shall be a much valuable input to the Agricultural scientific community of the nation", mentioned Dr Pradeep K Sinha, senior director (HPC and R&D), C-DAC.

Bioinformatics group of C-DAC has been working in the areas of genome annotation and comparative genomics since the last decade. The group participated in the Critical Assessment of Function Prediction (CAFA) challenge. The method proposed by C-DAC is based on similarity search and is titled 'Functional Annotation using Similarity Search' (FASS). It ranked 8th and was rated amongst the top teams across the globe. The work has been published in the March issue of the highly reputed journal Nature Methods (2013) titled 'A large-scale evaluation of computational protein function prediction'.

Mr. Goldi Misra, head, HPC Solutions Group, C-DAC, "The network of institutions based on two-tier architecture in bioinformatics will bridge the gap between genomic information and knowledge, utilizing statistical and computational sciences. It will also open up new vistas for downstream research ranging from modelling of cellular function, genetic networks, metabolic pathways, validation of drug targets to understand gene function and culminating in the development of improved varieties and breeds for enhancing agricultural productivity."

There is no two thoughts about the fact that the project is of a great national importance and the innovative approach shall bear fruits in near future.