ASHOKA is a Super-Computing hub of computational biology and bioinformatics to support biotechnological research in agriculture. This hub consists of hybrid hardware architecture of Linux and Windows based clusters, distributed over six national institutions across the country to form National Agricultural Bioinformatics Grid (NABG). These geographical distributed clusters with massive storage capacity are connected through high bandwidth i.e. MPLS/NKN connectivity. This network of institutions will bridge the gap of Omics information and knowledge within their respective domain. It will open up new vistas for downstream research in agriculture including modeling of molecular structures and functions, study of genetic networks, understanding of metabolic pathways etc. at molecular level. This may culminate in the development of improved varieties/breeds/commodities for enhancing agricultural production and productivity.
The ASHOKA, state-of-art High Performance Super-Computing for Indian Agriculture has been established under a National Agricultural Innovation Project (NAIP), ICAR, New Delhi, sub project Establishment of National Agricultural Bioinformatics Grid in ICAR (NABG). This Super-Computing hub consists of (i) 256 nodes Linux cluster with two masters, (ii) 16 nodes Windows cluster with one master, (iii) 16 nodes GP-GPU cluster with one master and (iv) one SMP of 64 cores with 1.5 TB RAM at IASRI, New Delhi. Apart from this, five Mini Super-Computers at five domain institutions i.e. 16 nodes Linux based cluster with one master at NBPGR New Delhi, NBAGR Karnal, NBGFR Lucknow, NAIM Mau and NBALI Bangalore form a National Agricultural Bioinformatics Grid (NABG) in the country. This national grid has more than 1.5 Peta Byte storage divided into three different types of architecture namely Network Attached Storage (NAS), Parallel File System (PFS) and Archival.

### Hardware Configuration
- 256 Nodes Linux Based Cluster (3072 cores) with two masters
- 16 Nodes GP-GPU Based Linux Cluster (192 CPU + 8192 GPU cores) with one master
- 16 Nodes Windows Based Cluster (192 cores) with one master
- 16 Nodes Linux Based Cluster (192 cores) with one master at each of the five domains of NABG
- One Symmetric Multi-Processing (SMP) for high memory intensive job (64 cores with 1.5 TB memory)

### Storage Capacity
- 250 TB of Network Attached Storage (NAS)
- 250 TB of Parallel File System (PFS) storage
- 200 TB of Archival storage
- 25 TB of NAS and 50 TB of PFS at each of the five domains of NABG

### Networking
- 4 x QDR based Infiniband Interconnect switches
- 10G/100G/1000G Ethernet switches
- MPLS/NKN Connectivity

Out of these nine Super-Computers of NABG, two Super-Computers of this hub have been listed in the list of top Super-Computers of India. Number of computational biology and agricultural bioinformatics software/workflows/pipelines have been deployed on these Super-Computing systems and made available to the researchers of the country. A National Agricultural Biocomputing Portal has been developed, which will provide seamless access to this Super-Computing facilities to the biological researchers across the country.

### Ranks among top supercomputers of India
(Source: http://topsupercomputers-india.iisc.ernet.in/jspjs/june2013/index.html)
Open Source Software/Tools

Fifty two open source software/tools are configured on this HPC environment to carry out various biological data analysis. These software/tools were identified based on online survey conducted among researchers from National Agricultural Research and Education System (NARES) institutions.

**ANVAYA:** It consists of workflows/pipelines running on HPC environment which are capable of executing complex computational genome analysis by making optimal utilization of the computing environment. Comprehensive analysis of heterogeneous genomic data requires a flexible platform for running complex queries which are capable of integrating and analyzing large amount of genomic data through pipeline environments. ANVAYA provides an interface to bioinformatics tools and databases in a workflow environment which is helpful in executing the set of software/tools in series or in parallel. Rules engine of ANVAYA defines rules for logical connection among the existing tools. It also offers novel functionalities to carry out exhaustive comparative analysis via custom tools with new functionalities and built-in PERL parsers. This has been developed by C-DAC Pune.

Commercial Software

**CLC Genomics Workbench:** It is a comprehensive software, which performs DNA, RNA, Protein sequence analysis, SNP detection and identification of genomic rearrangement of novel transcripts/exons, de-novo assembly etc. CLC Genomics Server is an advanced software solution offering secure, powerful and flexible bioinformatics computing. CLC Genomics Server as a core gives a unique and stable software architecture that makes it possible to apply a range of bioinformatics analysis services on HPC.

**Discovery Studio:** This software is a suite of life science molecular design solutions for computational biologists and computational chemists. It makes easier to examine the properties of large and small molecules, study systems, identify leads and optimize candidates. This tool will help in rapidly automate routine tasks, integrate third party applications and even deploy models. This will accelerate scientific innovation through rapid integration and deployment of critical software solutions for molecular modeling and simulation.
Web portal is referred as a Web site or “gateway” that provides multiple services/features like Resources, Tools, Database, News, Discussion Forums etc. National Agricultural Biocomputing Portal has been developed to provide a single point of access to High Performance Computing (HPC) resources and tools. These computational resources includes different clusters configured under National Agricultural Bioinformatics Grid (NABG) and various bioinformatics applications/software/tools which have been installed on these clusters. This portal is designed for seamless integration with grid computing architecture and providing services such as application services, grid information services, user authentication services, data management services, e-mail notification services etc. The portal is developed using open source tools comprising of Linux, Apache, MySQL, and PHP/PERL.

The various components of this computational facilities are being managed and monitored using browser based software. Round the clock helpdesk support is also made available to address the issues related to operational management of this facility. Different set of automation tools are configured to manage these computational resources.