

Review Article

BIOINFORMATICS SOFTWARE FROM INDIA: CURRENT STATUS AND CHALLENGES

Deepthi D. Deobagkar*

ISRO Science and Technology Cell, Savitribai Phule Pune University, Pune 411007, India

Abstract: Bioinformatics software and visualisation tools have been a key factor in the rapid and phenomenal advances in genomics, proteomics, medicine, drug discovery, systems approaches and in fact in every area of new development. Indian scientists have also made a mark in a few specific areas. India has an advantage of an early start and extensive and organised network in the Bioinformatics education and research with substantial inputs from the Indian government. India has a strong hold in computation and IT and has a pool of bright and young talent with demographic dividend along with experienced and excellent mentors and researchers. Although small in number and scale, Bioinformatics Industry also has a presence and is making its mark in India. There are a number of high throughput and extremely useful resources available which are critical in biological data analysis and interpretation. This has made a paradigm shift in the way research can be carried out and discoveries can be made in any area of biological, biochemical and chemical research. This article summarises the current status and contributions from India in the development of software and web servers for Bioinformatics applications.

Keywords: Applications; Big data; Web servers; Domain expertise; Interdisciplinary approach

Note - Coloured Figures available on Journal Website in "Archives" Section

1. Introduction

There is an astronomical accumulation of genomic, metagenomic, proteomic as well as metabolomic data, expression profiles and annotation which is increasing at a rapid rate. The storage, organization, systematization and integration of this vast information and data into biological networks as well as database systems and their wide utilization by the scientific research community necessitates the use of computer programming algorithms, analysis tools, services, and workflow systems. Therefore, software and analysis tools, and bioinformatics services and workflow have been the

main fields and core targets of bioinformatics since its emergence. Specialised workflows and platforms and high throughput tools that are available today have completely transformed the way high throughput research can be carried out and these have become necessary and integral part of the research methodology in most disciplines of modern explorations. In the early days of bioinformatics software, the user had to be familiar with the tools and techniques and computer languages. Over the years the tools and approaches have become user friendly wherein the applications are so designed that the user only has to possess core and domain knowledge and training and familiarity but does not need to be concerned with the back end, system architecture and the computational jargon which goes in the design of complex and multifaceted and integrated applications.

Corresponding Author: **Deepthi Deobagkar**
E-mail: deepthi.deobagkar@gmail.com

Received: September 1, 2017

Accepted: September 20, 2017

Published: September 24, 2017

Bioinformatics deals with the application development which can be application software, or simply applications, which are often called productivity programs or end-user programs because they enable the user to complete tasks, such as creating docking software, modelling software's, genomics analysis tools which can analyse metagenomics, transcriptomics, proteomics, prediction software like cancer detection, antibody antigen interaction prediction, protein structure prediction and the list is ever expanding since bioinformatics deals with a vast amount of data from multiple sources and angles. These application software are specific to the task they were designed for and can be as simple as nucleotide frequency calculation application or as complex as Microarray analysis, transcriptomics or genomics tools or system approaches for complex pattern searches.

2. Big data and challenges

There is phenomenal accumulation of large datasets and genomic and structural information which poses many challenges and necessitates the development of contemporary path breaking tools for annotation, systematic analysis and knowledge discovery. The sequence data is voluminous and NCBI: Genbank records 234997362623 bases as sequences deposited and 2164683993369 bases in WGS (June 2017). With the expected complete sequencing of more than 100,000 human genomes and massive enviro- and metagenomic analysis pipelines and gut microbiomes, this is promising to increase. The pattern and gene identification, omics discoveries, evolutionary and functional perspectives, molecular modelling and proteomics, network and systems analysis will demand the development and use of more comprehensive and user-friendly platforms and pipelines. All of these large scale, genome-derived, molecular sequence analyses of raw "Big Data" are impossible to be analyzed unless supported by strong inputs in computational, domain specific approaches to enable rapid discoveries in medicine, agriculture, basic biology, protein and gene function and in all areas of discovery.

3. Bioinformatics in India

The access to internet is on rise. In 2014, number of internet users in India was around >190 million which is now estimated to be 450 million. There have been good developments in Bioinformatics in India. This growth is in all the fields like research,

creation of new software, development of databases and web-servers and its applications. India has a strong hold in computation and IT and has a pool of bright and young talent with demographic dividend along with experienced and excellent mentors and researchers. It also has an advantage of an early start (1986 with DBT, then NBTB) and extensive and organised network in the Bioinformatics education (Centres of excellence, Bioinformatics centres and facilities – more than 170 supported by Department of Biotechnology, Government of India) and research and training with substantial inputs from the Indian government. Other government agencies like Department of Health, agriculture, Department of Electronics and Telecommunication, etc also have strong infrastructure, training programs and inputs in research support. Although small in number and scale, the Bioinformatics Industry is also making its mark felt in India. Bioinformatics is one of the fastest-expanding fields in India's biotechnology sector today. There are more than 200 companies in Bangalore, Hyderabad, Pune, Chennai, and Delhi that are in some way involved in bioinformatics. Large IT organizations such as Intel, IBM, and Wipro also have their presence into this sector in India. More than 350 college-level institutes across India now offer degrees in biotechnology, bioinformatics, and biological sciences. This educational base has led to the trained manpower and if appropriately utilised by imparting quality education and skill sets this can change the future course for India's bioinformatics industry.

The core R&D strength in Indian biotechnology is its relatively well-educated and skilled personnel, with a strong base of English-speaking scientists who are well versed in basic as well as technological aspects. There is availability of scientists and mentors in all disciplines of biology and biotechnology such as drug designing and pharmacy, vaccines and pharmacogenomics, drug discovery and chemoinformatics, medicine, health and genomics, proteomics and molecular modelling, biodiversity and metagenomics, agriculture and plant sciences, brain and behaviour, evolution and phylogenetics, etc. Thus, the country has the scientific skills that encompass capabilities for handling all aspects of biological information acquisition, processing, analysis, and interpretation. India's well-known software skills can be another key advantage in global bioinformatics. Both in computation and IT, Indian professional have made

worldwide and remarkable contributions. Lack of government interference is another advantage and the government agencies can become strong facilitator. India can perform data analysis related to genomic sequencing, functional genomics, proteomics and most other disciplines and this can enhance the research and discovery output along with possible commercial applications.

4. Bioinformatics software and web servers developed in India

There has been small but significant contribution in specific areas such as protein structure prediction, molecular modeling, protein-protein interactions, genomics and transcriptomics, drugs and vaccine design, etc. Many web servers have been developed which are well cited and have excellent applications. Research centers and Institutes have made significant contributions in this aspect in focused areas (Figure 1 and 2). The contribution from India for development of web servers (Table 1) and software (Table 2) is summarized (2014-2017). As an example, contributions from a few Institutes have been listed as a case study in table 3. There are other DBT supported centers and establishments (BTIS net website) which have been included in the table 1 and 2. An attempt has been made to list all the software and web servers which are currently functional and operational in the area of

Bioinformatics (2014-2017). It is noteworthy that some of them are extensively used with very good citations. The software and web servers which can be developed can find extensive applications in imparting quality education, literature and data searches, in discovery of unknown regions of genomes and their functions, in linking medicinal systems from traditional medicine, recent discoveries in modern medical and pharmaceutical research to genome and OMIM (Online Mendelian Inheritance in Man), neurological and cognitive functions, functional genomics and proteomics, pattern searches and many other areas. These can help by online, continuous and disruptive searches and continuous updating along with compilation and collation of multiple datasets from numerous platforms to generate information and knowledge and aid researchers in state of art applications and high throughput discovery mode. It is important to note that these can have extensive use and applications in both basic research and also find commercial applications.

5. Challenges

It is important to consolidate and take bold and visionary steps which will position the scientific community to take up the challenges by forming and consolidating networks of the strong IT and computational industries as well as training and

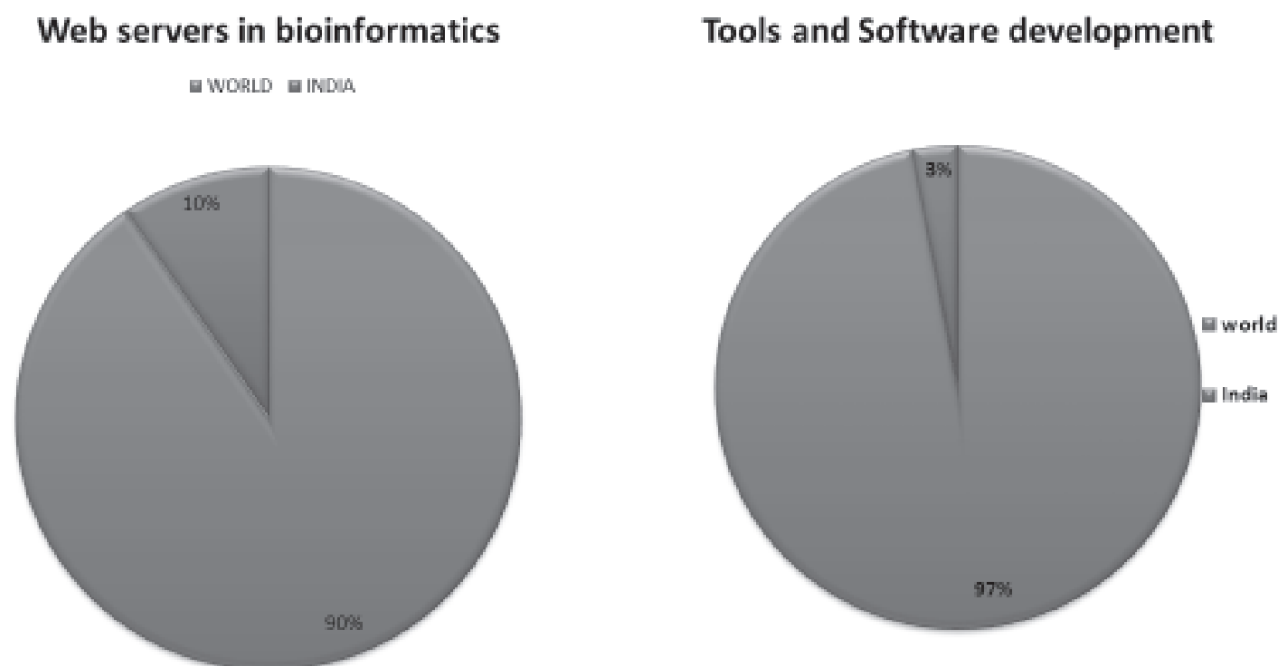


Figure 1: Web servers and tools and bioinformatics software: contribution from India (2014-2017)

Table 1
Contributions from India in development of web server which are available as web server or in public domain

NEMiD	Web-based curated microbial diversity database with geo-based plotting	www.mlabnehu.info/nemid/
CLAP	Web-server for automatic classification of proteins with special reference to multi-domain proteins.	http://nslab.mbu.iisc.ernet.in/clap
AntiAngioPred	Server for Prediction of Anti-Angiogenic Peptides	http://clri.res.in/subramanian/tools/antiangiopred
TSpred	Web server for the rational design of temperature-sensitive mutants	http://mspc.bii.a-star.edu.sg/TSpred/
WImpiBLAST	Web interface for mpiBLAST to help biologists perform large-scale annotation using high performance computing	http://wimpiblast.nabi.res.in
CHEXVIS	Tool for molecular channel extraction and visualization	http://ogl.serc.iisc.ernet.in/chexvis/
SynRi	R and Shiny based application platform for cyanobacterial genome analysis.	https://github.com/NFMC/SynRi
VIRmiRNA	Comprehensive resource for experimentally validated viral miRNAs and their targets.	http://crdd.osdd.net/servers/virmiRNA
SVM-PB-Pred	SVM based protein block prediction method using sequence profiles and secondary	http://bioinfo.bdu.ac.in/~sompbpred
LPIcom	Web server developed for understanding protein-ligand interaction for almost all ligands available in Protein Data Bank.	http://crdd.osdd.net/nagharva/lpicom
EGFRpred	QSAR based model for discriminating EGFR inhibitors and non-inhibitors using Random forest	http://crdd.osdd.net/oscadd/egfrpred/
MtbVeb	Designing Vaccines against Existing and Emerging Strains of <i>Mycobacterium tuberculosis</i>	http://crdd.osdd.net/nagharva/mtbveb/
GREAM	Role of repetitive sequences in mammalian genomes by offering easy selection of more interesting repeats in various contexts/regions	http://resource.ibab.ac.in/GREAM/
MATEPRED	Web server to detect riboswitches in genomic sequences.	http://service.iiserkol.ac.in/~riboscan/
HPVbase	Viral integrations, methylation patterns and microRNAs aberrant expression	http://crdd.osdd.net/servers/hpvbase/
deePred	Prediction of donor splice sites in eukaryotic species	http://cabgrid.res.in:8080/sspred/
EbolaVCR	Resource for designing therapeutics against Ebola Virus	http://crdd.osdd.net/oscadd/ebola/
MP3	Software tool for the prediction of pathogenic proteins in genomic and metagenomic data	http://metagenomics.iiserb.ac.in/mp3/index.php
RStrucFam	Web server to associate structure and cognate RNA for RNA-binding proteins from sequence information.	http://caps.ncbs.res.in/rstrucfam
BIS-GOAT	Breed identification of goat server	http://nabg.iasri.res.in/bisgoat/
Hemo-PI	Identify the peptides in hemolyte	http://crdd.osdd.net/nagharva/hemopi/
MATEPRED	SVM-Based Prediction Method for Multidrug and Toxin Extrusion Proteins.	http://www.bioinformatics.org/maepred_hos/
AVP-IC50 Pred:	Multiple machine learning techniques-based prediction of peptide antiviral activity in terms of half maximal inhibitory concentration (IC50).	http://crdd.osdd.net/servers/ic50avp
RV-Typer:	Web Server for Typing of Rhinoviruses Using Alignment-Free Approach.	http://bioinfo.net.in/RV-Typer/home.html

contd. table 1

Folding RaCe	Robust method for predicting changes in protein folding rates upon point mutations.	http://www.iitm.ac.in/bioinfo/proteinfolding/foldingrace.html .
PreFRP	Prediction and visualization of fluctuation of residues in proteins.	http://www.mpi.edu.in/prefrp/link.html .
RNApin	Protein-interacting nucleotides in a RNA sequence using composition profile of tri-nucleotides	http://crdd.osdd.net/haghava/rnapin/
SeeHaBITaT	Server on bioinformatics applications for Tosporviruses and other species.	http://bit.srmuniv.ac.in/ .
PICI:	Web server with a multi-parametric algorithm for identifying interaction sites within protein complexes.	http://pici.bicpu.edu.in/
RM2TS	Web server to convert a tertiary structure to an alphanumeric string and to predict the tertiary structure from the sequence of a protein	http://www.scfbio-iitd.res.in/software/proteomics/rm2ts.jsp
MICdb3.0	Comprehensive resource of microsatellite repeats from prokaryotic genomes.	http://www.cdfd.org.in/micas .
MFPPi	Multi FASTA ProtParam Interface.	http://insilicogenomics.in/mfppcalc/mfppi.html
ProInflam	Web server for the prediction of proinflammatory antigenicity of peptides and	http://metabiosys.iiserb.ac.in/proinflam/
PluriPred:	Web server for predicting proteins involved in pluripotent network.	http://bicresources.jcbose.ac.in/ssaha4/pluripred/help.php
SMEpred	Workbench: A web server for predicting efficacy of chemically modified siRNAs.	http://bioinfo.intech.res.in/manojk/snepred
ProCarDB	Database of bacterial carotenoids.	bioinfo.intech.res.in/servers/procar/b/
siRNAmo	Database of experimentally validated chemically modified siRNAs.	http://crdd.osdd.net/servers/sirnamod .
modPDZpep	Web resource for structure based analysis of human PDZ-mediated interaction networks.	http://www.nii.ac.in/modPDZpep.html
IRESPred	Web Server for Prediction of Cellular and Viral Internal Ribosome Entry Site (IRES).	http://bioinfo.net.in/IRESPred/
NAPS	Network Analysis of Protein Structures.	http://cabgrid.res.in/NAPS/ .
HSsplice	Donor splice sites using support vector machine	http://nplb.ncl.res.in
NPLB	<i>de novo</i> promoter architectures from genome-wide transcription start sites	http://nfmrc.res.in/ckb/index.html .
CKB	The genomic and proteomic information of 74 fully sequenced cyanobacterial genomes	http://proline.biochem.iisc.ernet.in/
PLIC	Protein-ligand interaction clusters	PLIC
AHTPDB	Comprehensive platform for analysis and presentation of antihypertensive peptides	http://crdd.osdd.net/haghava/ahtpdb
dbEM	A database of epigenetic modifiers curated from cancerous and normal genomes.	http://crdd.osdd.net/haghava/dbem
DOCKSCORE	For ranking protein-protein docked poses	http://caps.ncbs.res.in/dockscore/
CicArMiSatDB	The chickpea microsatellite database.	http://cicarmisatdb.icrisat.org
LeishMicrosatDB	Leishmania genomes database	http://bionedinformri.com/leishmicrosat

Table 2
Bioinformatics software for the analysis of biological data: list of software developed in India

C2Analyzer	Co-target-Co-function Analyzer identify the co-targeting miRNA pairs	www.bioinformatics.org/c2analyzer
SPAAN	Software Program for Prediction of Adhesins and Adhesin-Like Proteins	http://sourceforge.net/projects/adhesin/files/
ISMU	NGS tool for SNP search in the genomics, genetics and breeding studies	http://hpc.icrisat.cgiar.org/ISMU
PGTools	A Software Suite for Proteogenomic Data Analysis and Visualization.	http://qcmg.org/bioinformatics/PGTools
BioInt	Seamless integration, efficient extraction and effortless analysis of the data	
VIRmiRNA	Comprehensive resource for experimentally validated viral miRNAs and their targets.	http://crdd.osdd.net/servers/virmirna
IGMAP	An Interactive Mapping and Clustering Platform for Plants.	http://nipgr.res.in/igmap.html
Igloo-Plot	Tool for visualization of multidimensional datasets.	http://metagenomics.atc.tcs.com/IglooPlot/
SInC	An accurate and fast error-model based simulator for SNPs	http://sourceforge.net/projects/sincsimulator
QSPpred	Server for predicting and designing of Quorum sensing peptides	http://crdd.osdd.net/servers/qsppred/
GFF-Ex	A genome feature extraction package.	http://bioinfo.icgeb.res.in/gff
NRfamPred	A proteome-scale two level method for prediction of nuclear receptor proteins and their sub-families.	http://14.139.227.92/mkumar/nrfampred
No3CoGP	Non-conserved and conserved co-expressed gene pairs on microarray data	http://www.bioinformatics.org/no3cogp/
MaxMod	Hidden Markov model based novel interface to MODELLER for improved prediction of protein 3D models	http://www.immt.res.in/maxmod/
ParaPep	Web resource for experimentally validated antiparasitic peptide sequences and their structures	http://crdd.osdd.net/ragh_aaa/parapep/
ChloroSSRdb	Repository of perfect and imperfect chloroplast simple sequence repeats of green plants	www.compulbio.in/chlorosrdb/
PmiRExAt	Plant miRNA expression atlas database and web applications.	http://pmirexat.nabi.res.in
ProNormz	Human proteins and protein kinases normalization	http://www.biominingbu.org/pronormz
ECMIS	Computational approach for the identification of hotspots at protein-protein interfaces	http://caps.ncbs.res.in/download/ECMIS
MetaNET	A web-accessible interactive platform for biological metabolic network analysis.	http://metanet.osdd.net
ProTSAV	A protein tertiary structure analysis and validation server.	http://www.scfbio-iitd.res.in/software
CancerPPD	Database of anticancer peptides and proteins.	http://crdd.osdd.net/raghava/cancerppd/
Woods	A fast and accurate functional annotator and classifier of genomic and metagenomic sequences	http://metagenomics.iiserb.ac.in/woods
MS3ALIGN	An efficient molecular surface aligner using the topology of surface curvature	http://vgl.serc.iisc.ernet.in/ms3align
FROG	Fingerprinting Genomic Variation Ontology	http://ab-openlab.csir.res.in/frog
SBION	A Program for Analyses of Salt-Bridges from Multiple Structure Files	akbanerjee@biotech.burunio.ac.in
VaccineDA	Prediction, design and genome-wide screening of oligodeoxynucleotide-based vaccine adjuvants.	http://crdd.osdd.net/raghava/vaccineda

contd. table 2

MBSTAR	Multiple instance learning for predicting specific functional binding sites in microRNA	http://www.isical.ac.in
Onco-Regulon	An integrated database and software suite for site specific targeting of transcription factors of cancer genes	http://www.scfbio-iitd.res.in/software/onco
PHYSICO2	PWS-analyzer and finds application in sequence-bioinformatics	http://sourceforge.net/projects/physico2
tbvar	A comprehensive genome variation resource for <i>Mycobacterium tuberculosis</i> .	http://genome.igib.res.in/tbva
zflncRNApedia	A Comprehensive Online Resource for Zebrafish Long Non-Coding RNAsRNApedia	http://genome.igib.res.in/zflnc
PRIGSA	Protein repeat identification by graph spectral analysis	http://bioinf.iit.ac.in/PRIGSA/
YeATS	Tool suite for analyzing RNA-seq derived Transcriptome TSCODE2	https://github.com/sanchak/YEA
miRMOD	Tool for identification and analysis of 5' and 3' miRNA modifications in Next Generation Sequencing small RNA data	http://bioinfo.icgeb.res.in/miRMOD
CrisprGE	Central hub of CRISPR/Cas-based genome editing	http://crdd.osdd.net/servers/crisprge/
SIMBA	Web tool for managing bacterial genome assembly generated by Ion PGM sequencing technology.	http://ufmg-simba.sourceforge.net

Table 3
Contributions from different academic institutes: a case study

1. CDFD (Centre for DNA Fingerprinting):

Software developed by CDFD are:

- InSatDb citation: allows users to obtain microsatellites annotated with size (in bp and repeat units); genomic location (exon, intron, up-stream or transposon); nature (perfect or imperfect); and sequence composition (repeat motif and GC%). One can access microsatellite cluster (compound repeats) information, and a list of microsatellites with conserved flanking sequences (microsatellite family or paralogs).
- IMEx citation: tool for extracting Perfect, Imperfect and Compound Microsatellites or Simple Sequence Repeats (SSR's) or Short Tandem Repeats (STR's) from genome sequences.
- SilkSatDb : online relational database that catalogues information about the microsatellite repeats of the silkworm, *Bombyx mori*.
- Wildsilkbases: It is a BLAST searchable catalogue of expressed sequence tags (ESTs) created using several tissue types collected from the wild silk moths during different developmental stages. It is also provided with a total of 60,000 ESTs derived from 3 major wild silk moths, *Antheraea assama*, *Samia cynthia ricini* and *Antheraea mylitta*.
- MycoperonDB : provides operon and transcriptional unit information of different mycobacterial species at one place. At present, this database covers five species from mycobacteria and consist of an *in silico* model of operon organization of 18,053 genes.
- MICdb 3.0: MICAS is an interactive user-friendly web-based analysis server to find non-redundant microsatellites of a selected bacterial or archaeal genome sequence.
- Hansa is a tool to predict the deleterious effects of a mutation by using 10 Neutral- Disease Mis-Sense Mutation Discriminatory (NDMSMD) features. With the help of this tool user can classify the mutation either as "DISEASE" or "NEUTRAL".
- MtbPPI: application of protein functional linkages and largescale data from sources such as microarray, in order to understand latency in *M. tuberculosis*.
- HPFP: Hierarchical Approach to Protein Fold Recognition.

2. IOB (Institute of Bioinformatics)

Human protein research database: provide datasets and also provides adds web server for protein analysis.

- NetPath: 'NetPath' is a curated resource of signal transduction pathways in humans.
- Pancreatic cancer database: web server with tools to analyse the pancreatic cancer dataset.
- PathBuilder: free software builds the metabolic path based on the protein interactions.
- The Human Proteome Map (HPM - <http://humanproteomemap.org>) portal is an interactive resource integrating the massive peptide sequencing result from the draft map of the human proteome project.
- Human Protein Reference Database (HPRD - <http://www.hprd.org>) is a centralized platform that stores and disseminates information such as protein-protein interactions, post-translational modifications, tissue expression, sub-cellular localizations and enzyme-substrate relationships in normal condition.
- Resource of Asian Primary Immunodeficiency Diseases (<http://rapid.rcai.riken.jp/RAPID>): RAPID is a web-based compendium of molecular alterations in primary immunodeficiency diseases. It provides categories and elaborated information about the genes and proteins that are changed or altered for example protein-protein interactions, microarray gene expression profiles and deleterious and novel mutations associated with them.

3 CDAC

- MGA Viwer 25: It is a multiple genome alignment viewer which highlights via pictorial depiction the conserved and variant regions in prokaryotic genomes.
 - Anvaya: Analysis of heterogeneous genomic data needs a platform which is flexible for running related advance which is capable of combining and analysing big data of genomic pipelines. Anvaya developed by CDAC is a software application consisting of interface to Bioinformatics tools and databases in a workflow environment, to execute the set of analysis tools in series or in parallel.
 - GenoPIPE: An automated pipeline for high-through-put comparative genomics, based on the detection of orthologous groups, which serve as the seed for subsequent annotation and analysis. It does provide SNP, paralogs and protein coding regions which can be present in data but are missed by the gene prediction algorithms, thus it serves as supplementary pipeline for prokaryotic genome annotation.
-

- TaxoGrid : Phylogeny on Grid
- BioUtils : An interface to Bioinformatics Utilities
- iMolDock : iMolDock is a cluster-based/ grid-based portal, an interface to Molecular Docking.
- GenomeGRID : GenomeGrid, a grid portal, provides unique solution to highly complicated supercomputing grid with its user-friendly web interface for sequence analysis codes like Smith-Waterman (S-W), FASTA, BLAST, ClustalW and molecular modelling codes like AMBER enabling bioinformatics expertise to use the maximum amount of data.
- GROMAC: it is one of the software which is used for the modelling and MD simulations.

4. IITD (Indian Institute of Technology, Delhi)

BioSuite: integrates the functions of macromolecular sequence and structural analysis, cheminformatics and algorithms for aiding drug discovery. Developed by IITD and TCS (Tata Consultancy Services). Contains 79 different programs into a) Genome and Proteome Sequence analysis, (b) 3D modelling and structural analysis, (c) Molecular dynamics simulations and (d) Drug design.

Dhanvantari: Supercomputing Facility for Bioinformatics & Computational Biology, IIT DELHI. It is developed for the genomics, proteomics and drug discovery. It provides high performing system with good accuracy. Complete genomes can be submitted and can be analysed using this. Permits proteomics analysis where from the amino acid sequence, drug : protein interactions can be modelled.

SANJIVANI: A Complete Drug Designing Software Suite. Hosted by IIT Delhi. It provides facilities like: Search Molecule from the Database OR Upload the Drug Molecule. Single Binding Site Docking and multiple Binding Site Docking. Docking & scoring of the target molecule.

5. CCMB (Centre for Cellular and Molecular Biology):

Software's hosted

- VAMA (Variability Analysis of Multiply-aligned Amino acid sequences)
- DIDV (Digitization of Image and Data Visualization)
- SnS-Align (Structure and Sequence Alignment)
- RISC – Repeat Induced Sequence Change Identifier
- SNPScore
- BRM-Parser (A tool for comprehensive analysis of BLAST and RepeatMasker results)
- cdBEST: chromatin domain Boundary Element Search Tool

Web servers/ Database

- Amino Acid Periodicity (AAP)
- BetaPropellers
- Chameleon Sequences (exact match of probe with target) (CHAMPEP-EXACT)
- Chameleon Sequences (sliding probe over target) (CHAMPEP-SLIDING)
- Turn Combinations (Cturns)
- Disulphide Bond Connectivity Patterns in Protein Tertiary Structure (DBCP)
- Database of Structural Motifs in Proteins (DSMP)
- Heteroatom Groups and Neighbours In Protein Data Bank (HET_GROUP_PDB)
- Intra Chain Disulphide Bridged Peptides (ICDBP)
- Online Resource for Biochemical Pathways (ORBiP)
- Protein Sequence-Structure Analysis Relational Database (PSSARD)
- Coffee Database
- Mulberry Database
- Crocodile Database
- RiDs db: Repeats in diseases database

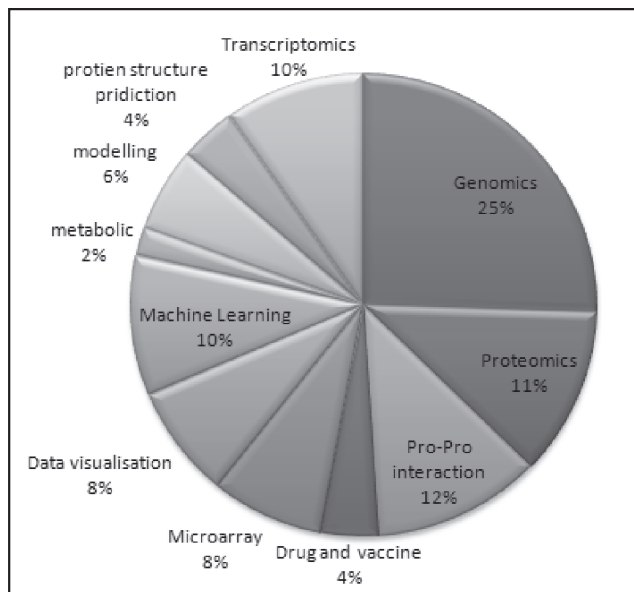


Figure 2: Subject area wise distribution of bioinformatics software from India in 2014-17

research institutes along with state of art biology and computational biology researchers to address the emerging and future challenges in the areas of drug discovery, genomics, metagenomics, protein engineering, systems and network approaches, neurobiology and cognition, agriculture, robotics and electronics among others. With the large pool of human resource, it will be important to strategically utilise the skill development with strong academic input with planning and vision to

drive the engine of software development to make original and path breaking contributions. This can be done by working in a mission mode with time targeted goals in small, focussed multi and interdisciplinary groups which will work by building strategies with a problem-solving approach. This can be in small or large clusters with it will be crucial to address the simple problems and also take up the challenges thrown in by the massive explosion in the genomic and proteomics data and in unravelling the complexities of life. A new approach and fresh outlook for the way forward is critical in order to meet the challenges and make significant contributions.

Acknowledgements

DDD wishes to acknowledge help from Mr. Pawan Mishra, JRF, working in the Ministry of Earth sciences (GOI) project to DDD in compiling the data, tables and figures. DDD is currently an ISRO chair Professor and was a former Director, Bioinformatics Centre, Savitribai Phule Pune University (2001-2003 and 2010- 2017) and Professor of Molecular Genetics at the Zoology Department, SPPU. This article is based on a talk delivered at a national conference on "Breaking barriers through Bioinformatics and Computational biology" at IIT, Delhi on 31st July 2017. The Software list @ DBT website _BTIS net has been referred to in the manuscript.

References

- Gupta, S., Chavan, S., Deobagkar, D.N. and Deobagkar, D.D (2015) Bio/chemoinformatics in India: an outlook. *Briefings in Bioinformatics* 16, 710-731.