International Journal of Computational Research and Development (IJCRD)
Impact Factor: 4.775, ISSN (Online): 2456 - 3137
(www.dvpublication.com) Volume 2, Issue 1, 2017

BLOOMING ASTONISHINGLY OF BIOINFORMATICS AND ITS APPLICATIONS

Saradha Arumugam*, Dr. M. Vasanth & Dr. S. Saravanan*****
CAS in Marine Biology, Faculty of Marine Sciences, Annamalai University, Parangipettai, Tamilnadu

Cite This Article: Saradha Arumugam, Dr. M. Vasanth & Dr. S. Saravanan, "Blooming Astonishingly of Bioinformatics and Its Applications", International Journal of Computational Research and Development, Volume 2, Issue 1, Page Number 14-20, 2017.

Abstract:

Bioinformatics is the discipline that uses computers to store, retrieve, manipulate and distribute information related to biological macromolecules such as RNA, DNA and proteins. It is broad term covering the use of computer algorithms to analyze biological data. For theoretical analysis and therapy design to be carried out effectively, a range of tools and resources need to be made available in both experimental and computational aspects. The Bioinformatics websites were useful resources for determining whether a specific genomic region is disrupted frequently in a particular disease or malignancy and for finding chromosomal cell lines and reagents for regional mapping, for e.g., NCBI. This helps in developing integrated data bases over the internet. This also helps in understanding function of gene and also genome. Various www servers and Internet are quiet potent and playing a dynamic role in the Bioinformatics field. It is destinated to change the face of biomedicine.

Key Words: Bioinformatics, Application, Computational Tools & Website

Introduction:

The term bioinformatics was coined by Paulien Hogeweg in 1979 for the study of informatic processes in biotic systems. The National Center for Biotechnology Informaion (NCBI, 2001) defines bioinformatics as: "Bioinformatics is the field of science in which biology, computer science and information technology merge into a single discipline.There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information." Bioinformatics, sometimes called Computational Biology, is the use of techniques from applied mathematics, informatics, statistics and computer science to solve biological problems. It was fuelled by the need to create huge databases such as GenBank and EMBL and DNA database of Japan to store and compare the sequences data erupting from the human genome and other genome sequences.

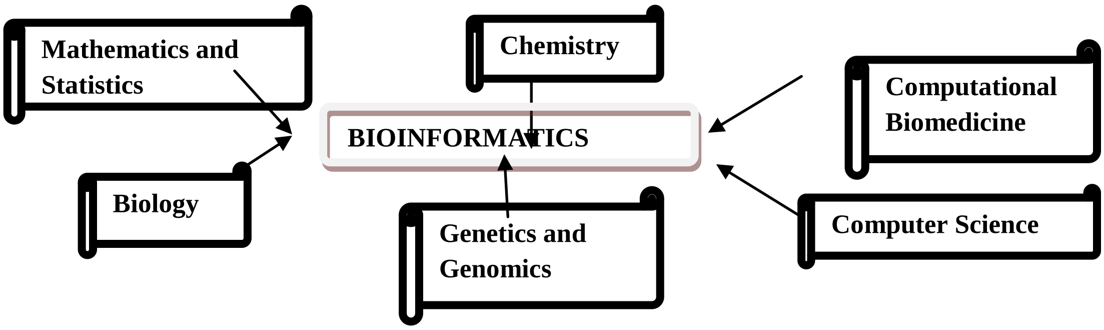

Figure 1: Combination of Tools involved in Bioinformatics

Databases of Bioinformatics:

Databases	Websites
Macromolecular Structure	
Protein Data Bank (PDB)	http://www.rcsb.org/pdb
Nucleic Acids Database (NAD)	http://www.ndbserver.rutgers.edu/
FSSP	http://www2.embl-ebi.ac.uk/dali/fssp
HIV protease database	http://www.ncifcrf.gov/CRYS/HIVdb/NEW DATABASE
SCOP	http://www.scop.mrc-lmb.cam.ac.uk/scop
CATH	http://www.biochem.ucl.ac.uk/bsm/cath
PDBsum	http://www.biochem.ucl.ac.uk/bsm/pdbsum
ReLiBase	http://www2.ebi.ac.uk:8081/home.html
Protein sequence	
SWISS-PROT	http://www.expasy.ch/sprot/sprot-top.html
PIR-International	http://www.mips.biochem.mpg.de/proj/protseqdb

PRF	https://www.prf.or.jp/aisatsu-e.html
OWL	http://www.bioinf.man.ac.uk/
NRDB	http://www.ncbi.nlm.nih.gov/entrez/query.fegi?db=Protein
PROSITE	http://www.expasy.ch/prosite
PRINTS	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/PRINTS.html
Pfam	http://www.sanger.ac.uk/Pfam/
Genome sequences	
GeneCensus	http://www.bioinfo.mbb.yale.edu/genome
Entrez genomes	http://www.ncbi.nlm.nih.gov/entrz/query.fcgi?db=Genome
COGS	http://www.ncbi.nlm.nih.gov/COG
GOLD	http://www. gold.jgi.doe.gov/
TIGR	https://www.hsls.pitt.edu/
Organism Specific Databases	http://www.humgen.nl/orgspecdatabases.html
PEDANT	http://www.pedant.gsf.de/
GeneQuiz	http://www.ebi.ac.uk/genomes/
HAMAP	http://hamap.expasy.org/
ArkDb	http://www.thearkdb.org/arkdb/
UK Cropnet	http://flora.life.nottingham.ac.uk/
Codon Usage Database	https://www.kazusa.or.jp/codon
DNA databases	
DDBJ	http://www.ddbj.nig.ac.jp/
RNA databases	
ssu rRNA	http://bioinformatics.psb.ugent.be/webtools/rRNA/
lsu rRNA	http://bioinformatics.psb.ugent.be/webtools/rRNA/
RDP	https://www.rdp.cme.msu.edu/
Carbohydrate databases	
CarbBank	http://glycomics.ccrc.uga.edu
Model Organism Databases	
i. *Escherchia coli*	
Escherchia coli Stock Centre Database	http://www.cgsc.biology.yale.edu/cgcs.html
Escherchia coli Genome Centre	http://www.genetics.wisc.edu
The *Escherchia coli* index	http://www.web.bham.ac.uk/bcm4ght6/res.html
SGD	http://www.genome.stanford.edu/Saccharomyces
MYGD	http://www.websyr.mips.biochem.mpg.de/proj/yeast
YPD	http://www.proteome.com/YPDhome.html
ii.*Arabidopsis thaliana*	
MATDB	http://www.mips.biochem.mpg.de/proj/thal/
TAIR	http://www.arabidopsis.org/search
KAOS	http://www.kazusa.or.jp/kaos
iii. *Oryza sativa*	
RGP	http://www.rgp.dna.affrc.go.jp/index.html
INE	http://www.rgp.dna.affrc.go.jp/giot/INE.html
iv.*Caenorhabdins elegans*	
WormBase	http://www.proteome.com/databases
ACEDB	http://www.sanger.ac.uk/software/Acedb/
WormPD	http://www.proteome.com/databases
Drosophila melanogaster	
FlyBase	http://www.flybase.bio.indiana.edu/
BDGP	http://www.fruitfly.org/index.html
Mus musculus	
MGI	http://www.informatics.jax.org
Homo sapiens	
GDB	http://www.gdb.org/
Ensembl	http://www.ensembl.ebi.ac.uk
Entre-genome	http://www.ncbi.nlm.nih.gov/cgi-bin/Entre/humsrch?chr=humchir.inf&query

Human genome working draft	http://www.genome.ucsc.edu/
HGP	http://www.sanger.ac.uk/HGP
Pathway databases	
KEGG	http://www.genome.jp/kegg/
Reaction pathway	http://www.reactome.org/
Panther pathway	http://www.pantherdb.org/pathway/
LIGAND	http://www.ligand-expo.rcsb.org/
BRITE	http://www.genome.ad.jp/brite/
Boehringer Mannheim	https://www.boehringer-ingelheim.com/
EcoCyc	http://www. ecocyc.org/
MetaCyc	http://www. metacyc.org/
WIT	http://www.wit.mcs.anl.gov/WIT2/
UM-BBD	https://www.hsls.pitt.edu/obrc/index.php?page=URL1100188151
BRENDA	http://www.brenda.uni-koeln.de
SRS	http://www.srs6.ebi.ac.uk/srs6bin/cgi-bin/wegtz?-page+LinInfo+-id+4FIDs1DbEnd+-page+top
Microarray databases	
ArrayExpress	http://www.ebi.ac.uk/arrayexpress
GEO	http://www.ncbi.nlm.gov/geo
Specific protein family databases	
Proweb	http://www.proweb.org/
TRANSFAC	http://www.transfac.gbf-braunschweig.de/TRANSFAC/index.html
GCRDb	http://www. gpcrdb.org/
PKR	http://www.sdsc.edu/kinases/
ENZYME	http://www.expasy.ch/enyme
MEROPS	http://www.merops.co.uk
ReLib	http://www.relibase.ebi.ac.uk
CAZy	http://www.afmb.cnrs-mrs.fr/~pedro/CAZY/db.html
PROMISE	http://www.bioinf.leeds.ac.uk/promise
IMGT	http://www.imgt.cinesfor:8104
REBASE	http://www.rebase.neb.com/rebase.html
Protein classification databases	
CluSTr	http://www.ebi.ac.uk.clustr
ProtoMap	http://www.protomap.cs.huji.ac.il/
PROTFAM	http://www.speedy.mips.biochem.mpg.de/mips/programs/classification.html
Hits	http://www.lifecenter.sgst.cn/hit/
BioSpace	http://www.biospace.com/
LPFC	https://www.hsls.pitt.edu/obrc/index.php?page=URL1150480174
Enzyme structure databases	http://www.biochem.ucl.ac.uk/bsm/enzymes/
PRESAGE	http://www. presage.berkeley.edu
CDD	https://www.ncbi.nlm.nih.gov/cdd/
Specialized structure databases	
PDIdb	http://melolab.org/pdidb/web/content/links
3DInsight	http://www.integbio.jp/dbcatalog/en/record/nbdc00002?jtpl=56
MolMovDB	http://www.molmovdb.org/

Table 1: List of Databases available in the Webservers

Bioinformatics Centers and Servers:

S.No	Abbreviations	Full form	State/Country Name
1.	ANGIS	Australian National Genomic Information Service	Australia
2.	BEN	Belgian EMBNet Node	Belgium
3.	Biology Workbench		
4.	Brutlag Bioinformatics Group		at Stanford University, USA
5.	CBCG	USDA-ARS Centre for Bioinformatics and Comparative Genomics	UK

6.	EBI	European Bioinformatics Institute	Hinxton, USA
7.	EBS	Edinburg Biocoputing Service	Edinburg, UK
8.	EMBnet and related WWW servers		
9.	ExPASy		Geneva, CH
10.	Genestream II		at IGH, Montpelliar, France
11.	GenomeNet		Japan
12.	IBC	Institute for Biomedical computing	Washington University Seattle,USA
13.	ISCB	International Society for Computational Biology	Bethesda MD, USA.
14.	ISREC	Swiss Institute for Experimental Cancer Research	Switzerland
15.	MBCR	Molecular Biology Computation Resource	Baylor College of Medicine; Houston, USA
16.	MIPS	Munich Information Centre for Protein Sequences	Munich, Germany.
17.	NCBI	National Centre for Biotechnology Information	Bethesda MD, USA.
18.	NCGR	National Centre for Genome Resources	Santa Fe, New Mexico, USA
19.	RIB	Roslin Institute Bioinformatics	Edinburg, UK
20.	Sanger Centre		Hinxton, UK
21.	SDSC	San Diego Supercomputer Centre	USA
22.	SIB	Swiss Institute for Bioinformatics	Geneva, CH
23.	SSS	Sequence and Structure Searching site	Stanford, USA

Table 2: List of Bioinformatic Centres and Servers

Bioinformatics is Being Used in the Following Fields:

- Microbial genome applications
- Molecular medicine
- Personalised medicine
- Preventative medicine
- Gene therapy
- Drug development
- Antibiotic resistance
- Evolutionary studies
- Waste cleanup
- Biotechnology
- Climate change Studies
- Alternative energy sources
- Crop improvement
- Forensic analysis
- Bio-weapon creation
- Insect resistance
- Improve nutritional quality
- Development of Drought resistant varieties
- Vetinary Science

Applications of Bioinformatics in Microbiology:

The major impact of the bioinformatics has been in automating the microbial genome sequencing. The development of integrated databases over the Internet and analysis of genomes to understand gene and genome function. BLAST based database search and Smith-Water-man based gene-pair alignment algorithm and their variations are being used extensively in comparing genes and genomes and have become the first steps to derive the gene-function and the functionality of genomes. Since the sequencing of the first complete microbial genome of *Haemophilus influenzae* in 1995 (fleischmann, 1995]. It also quickened the drug discovery, vaccine design and also the design of anti-microbial agents.

The 16SrRNA approach is rooted in the concept of point mutation of conserved genes due to their slow mutation rate, uses 16SrRNA database and multiple sequence alignment (Chenna *et al*, 2003) and uses neighbor join algorithm [Howe *et al*, 2002] to build an evolutionary tree. Before microbial genomes were sequenced, this technique was considered quantitatively sound, and using 16SrRNA database three distinct domains – bacteria, archaea, and eukaryotes – were identified. Archea domain is hyperthermophilic and its 16SrRNA is somewhat different from 16SrRNA of bacteria.

Application of Bioinformatics in Marine:

Marine is relate to the Ocean which is 3/4th of sea was occupied in the earth, plays a vital role in influence the climate change. The term Marine Bioinformatics refers to use of computer techniques to gather, store, integrate, analyse and disseminate data on the distribution of marine organisms, its description, systematic classification, phylogeny, structure and sequence data on biomolecules including protein (Proteomics) and gene (Genomics) (Vinithkumar, 2004). would come to a standstill.

It offers specialised analytical technology tools for both laboratory and networking technology fields to achieve the goal of finding natural bioactive compounds from marine organisms as potential drugs, antifouling compounds, biomaterials etc. It has highest implication in marine based ecology, biology, biotechnology and molecular biology, which are integrated for the gene based discovery and development of marine drugs. The marine bioinformatics includes marine genomics and marine proteomics (Carte, 1996).

Application of Bioinformatics in Agriculture:

Bioinformatics is a new field of science. But, it is making progress in every field of biotechnology very rapidly. As it has its application in the medicine by providing the genome information of various organisms. Similarly, the field of agriculture has also taken advantage of this field because microorganisms play an important role in agriculture and bioinformatics provides full genomic information of these organisms. The genome sequencing of the plants and animals has also provided benefits to agriculture.

A number of applications of bioinformatics in agriculture in view of functional genomics, data mining techniques, genome-wide association studies, high-performance computing facilities in agriculture and various bioinformatics tools/databases important for breeders, biotechnologists and pathologists. Agricultural genomics leads to the global understanding of plant/animal and pathogen biology and its application would be beneficial for agriculture (Iquebal *et al*, 2015).

Tools of bioinformatics are playing significant role in providing information about the genes present in the genome of these species. These tools have also made possible These tools have also made it possible to predict the function of different genes and factors affecting these genes. The information provided about the genes by the tools makes the scientists to produce enhanced species of plants which have drought, herbicide and pesticide resistance in them. Similarly specific genes can be modified to improve the production of meat and milk. Certain changes can be made in their genome to make them disease resistant e.g., Bt, *Arabidosis thaliana*.

Application of Bioinformatics in Biotechnology:

Bioinformatics has got so many applications in biotechnology field. Some of the important ones include automatic genome sequencing, automatic identification of genes, identification of functions of gens, predicting the 3D structure modelling and pair-wise compairism of genes.

1. Automatic Genome Sequencing: Functional analyses of protein sequences can now be performed on a computer using a variety of software tools that allow the user to exploit the biochemical knowledge accumulated in sequence databases. A high degree of automation is required to cope with the analysis of the huge number of sequences generated by genome sequencing projects, and to ensure consistent and reproducible results, freeing the expert user to verify and refine these analyses and to follow up new discoveries (Andrade *et al.*, 1999). (i) Bioinformatics are used in the development of automated sequencing techniques that use PCR or BAC based gene amplifications, two dimensional electrophoresis and also automated reading of nucleotides. (ii) Bioinformatics are used in joining the sequence of smaller fragments of DNA to form a complete genome sequence. (iii) Bioinformatics are also used in the prediction of promoter region of the genome. (iv) Bioinformatics are also used in the prediction of promoter coding region of the genome.

2. Identification of Genes: A unit of heredity which is transferred from a parent to offspring and is held to determine some characteristic of the offspring. The structure of a gene consists of many elements of which the actual protein coding sequence. Bioinformatics programmes such as GLIMMER and GenBank are used to identify the coding region or open reading frames in the genome.

3. To identify Gene Function: Genome annotation hardly can be any exact definition. But, for the purpose of this discussion, it might be useful to define annotation as a subfield in the general field of genome analysis, which includes more or less anything that can be done with genome sequences by computational means. The “unit” of genome annotation is the description of an individual gene and its protein (or RNA) product and the focal point of each such record is the function assigned to the gene product. The record may also include a brief description of the evidence for this assigned function, e.g. percent identity with a functionally characterized homolog or the boundaries of domains detected in a domain database search. But, there is no room for any details of the analysis. After identifying the open reading frames present in the genome, next step is to annotate the structure and function of the genes. Bioinformatics tool such as sequence search and pair-wise gene alignment technique are used to identify the gene function. The major four algorithms such as BLAST, BLOSUM, ClustalX and SMART are used for the functional annotation of genes.

Application of Bioinformatics in Drug Designing:

The drug is most commonly an organic small molecule that activates or inhibits the function of a biomolecule such as a protein, which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves the design of small molecules that are complementary in shape and charge to the biomolecular target with which they interact and therefore will bind to it. Drug design frequently but not necessarily relies on computer modeling techniques. Furthermore, bioinformatics may also help to analyze pharmacogenomics study (Chang *et al*, 2012) in terms of SNP-SNP interaction (Yang *et al*, 2011) to predict the effectiveness of drugs and disease susceptibility. In future, the bioinformatics and pharmacogenomics studies may become more popular in natural products.

Application of Bioinformatics in Medicine:

More importantly, as doctors and hospitals transition away from paper medical records, this data is increasingly being collected and made available in an electronic format. The availability of large data sets of digital medical information has made possible the use of informatics to improve health care and medical research. Often referred to as “*in silico*” research, informatics offers a new pathway for medical discovery and investigation. The field of bioinformatics has exploded within the past decade to keep pace with advancements in molecular biology and genomics research.

Molecular Medicine:

Molecular medicine is a spacious field, in which chemical, physical, biological and medical techniques are used to depict molecular structures and mechanisms, identify essential molecular and genetic errors of disease, and to develop molecular interventions to correct them. The human genome completion implies that we can explore for the genes directly associated with different diseases and begins to understand the molecular basis of these diseases more clearly. By getting this new knowledge of the molecular mechanisms of disease it will enable better treatments, cures and even preventative tests to be developed.

Application of Bioinformatics in Nanotechnology:

In nanotechnology we can deliver drugs through nano size vehicles. Researchers are trying to develop nanoparticle size of molecules that can deliver drugs directly to diseased cells in our body. For example, we can deliver anticancerous drugs via nano size particles which have been integrated in adenovirus which quickly multiplies in cancerous cells to destroy them. Nanotechnology convergence holds great pledge for molecular diagnosis and individualized therapy of cancer and other human diseases. It also helped us to block the cancer channels (blood vessels) to stop their nutrition and cancerous cells can be starving to death. Similarly, by studying not only human genomes we can discover many interesting regions in different organisms which can be used to make a nanodevice for cancer destruction though biocomputing skills. It is critical to apply nanotechnology to the development of new materials and new techniques that can be utilized for the development of suitable sample preparation steps, such as extraction, concentration and isolation.

Challenges in Bioinformatics:

- ✓ RNA splicing model
- ✓ Protein evolution
- ✓ Drug design
- ✓ Speciation
- ✓ Protein structure prediction
- ✓ Signal transduction pathways
- ✓ Transcription model
- ✓ Gene Ontologies

Why we are Need and Use of Bioinformatics?

The successful sequencing of entire human genome in the Human Genome Project opened up our understating on biological systems in an unprecedented way. This was made possible due to the advancement of molecular biotechnology combined with information technology. Presently, bioinformatics techniques are widely being applied to study the land based animals, plants and humans. Also to an extent in marine based systematics, phylogenetics, genomics and proteomics.

Conclusion:

These bioinformatics applications address the transfer of information within this protocol including systems that generate hypotheses, design experiments, store and organize the data from these experiments in databases. Test the compatibility of the data with models and modify hypotheses. As a result of the massive surge in data and its complexity, many of the challenges in biology have actually become challenges can handle large quantities of data and probe the complex dynamics observed in nature. Further, it is often regarded as the application of computational techniques to understand and organise the information associated with biological macromolecules. The development of new drugs is an expensive process the toxicology profile must be taken into account at an early stage of the drug development. Thus, a highly potent molecule may be rejected at a late stage of the drug discovery process because of unavoidable and unpredictable side-effects. In order to successfully design safer drugs it has to be tested for its activity towards the particular protein target.

Acknowledgement:

We, the authors are thankful to our Dean and Director of CAS in Marine Biology and Authorities of Annamalai University for providing facilities and encouragement.

References:

1. Vinithkumar, N.V. Marine Bioinformatics. OSTI Newsletter, Quarterly, Ocean Science and Technology for Islands, National Institute of Ocean Technology, 2.1, 2004.
2. Carte, B.K. Biomedical potential of marine natural products. Biosciences. 271-286, 1996.
3. M. A. Iquebal, Sarika Jaiswal_ C. S. Mukhopadhyay, Chiranjib Sarkar, Anil Rai, Dinesh Kumar. Applications of Bioinformatics in Plant and Agriculture. The omics of plant sciences. 755-789, 2015.
4. Fleischmann RD, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.Science, 269:496-512, 1995.
5. Chenna R, Sugawara1 H, Koike Lopez TR, Gibson TJ, Higgins DJ, Thompson JD: Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research, 31(13): 3497-3500, 2003.
6. Howe K, Bateman A, Durbin R: Quick tree: building huge neighbor-joining trees of protein sequences. Bioinformatics, 11:1546-1547, 2002.
7. Chang HW, Chuang LY, Tsai MT, Yang CH. The importance of integrating SNP and cheminformatics resources to pharmacogenomics. Curr Drug Metab 13: 991-999, 2012.
8. Yang CH, Chuang LY, Chen YJ, Tseng HF, Chang HW. Computational analysis of simulated SNP interactions between 26 growth factor-related genes in a breast cancer association study. OMICS 15: 399-407, 2011.
9. Miguel A. Andrade, Nigel P.Brown, Christophe Lorey, Sebastian Hoersch, Antoine de Daruvar, Christian reich, Angelo franchini, Javier Tamames, Alfonso Valencia, christos ouzounis, and Chrissander. Automated genome sequence analysis and annotation. Bioinformatics. 15:5, 391-412 1999