Bioinformatics Tools

M.B. Priyadarshi

National Bureau of Plant Genetic Resources, New Delhi-12

Abstract: In broad spectrum, bioinformatics deals with DNA sequence analysis, searching databases, annotation, protein structure prediction, primer designing etc. To accomplish these activities over the period of time need based tools were developed. Internet has large number of bioinformatics tools scattered, therefore in today's scenario big enterprises such as pharmaceuticals firms are employing bioinformaticians to do activities related to sequence data. Bioinformatics tools are designed to extract meaningful information from the huge dataset of sequences. They help to do similarity searches, assembling short sequence reads and developing consensus sequences, finding genes and markers in silico and in performing comparative analysis of different genomes.

In this article an effort has been made to compile brief information of ten bioinformatics tools (DNASTAR, pDraw32, Sequencher, MacVector, Bioedit, DNA Dynamo, CLC Free Workbench, SNS-Align, UGene, Emboss) which are very useful for researchers. In general, all these tools are accessible on internet. Some of them are free, some are paid and some are open source cross-platform softwares. These tools are providing facility for DNA sequence analysis, database searching, viewing graphical representation, editing and annotation.

1. INTRODUCTION

Bioinformatics is a new stream of Science consisting of Biological and Computer Science information. According to B. Jayaram et.al. "It is an interdisciplinary area of Science & Technology encompassing a systematic development and application of IT solutions to handle biological information by addressing biological data collection and warehousing data mining, database searches, analyses and interpretation, modeling and product design." According to Mark Gerstein of Yale University, "Bioinformatics is conceptualizing biology in terms of molecules (in the sense of physical chemistry) and then applying "informatics" techniques (derived from disciplines such as applied math, computer science, and statistics) to understand and organize the information associated with these molecules, on a large scale.

In broad spectrum, bioinformatics deals with DNA sequence analysis, searching databases, annotation, protein structure prediction, primer designing etc. To accomplish these activities over the period of time need based tools were developed. Internet has large number of bioinformatics tools scattered, therefore in today's scenario big enterprises such as pharmaceuticals firms are employing bioinformaticians to do activities related to sequence data.

Bioinformatics tools are designed to extract meaningful information from the huge dataset of sequences. They help to do similarity searches, assembling short sequence reads and developing consensus sequences, finding genes and markers *in silico* and in performing comparative analysis of different genomes.

In this article an effort has been made to compile brief information of ten bioinformatics tools which are very useful for researchers.

2. BIOINFORMATICS TOOLS

2.1 DNASTAR

[8] DNASTAR is global company which sells software of bioinformatics in more than 90 countries worldwide. It was established in 1984. Its primary software product, Lasergene, is used by molecular biology researchers for DNA and protein sequence assembly and analysis. The first version of Lasergene was released in 1986. Gradually with time new capabilities were added in the software and Lasergene has become famous in academic, government and pharmaceuticals world. This software supports various types of cloning techniques such as virtual restriction enzyme cloning, TOPO cloning, and cloning using the Gateway technology. By June 2007 new software entitled "ArrayStar" was released for microarray gene expression analysis. Presently Latest version of Array star in Version 4.0 which was released in 2010. It includes QSeq, which is a optional module. It helps scientist to visualize and anlyze RNA-Seq na ChiP-Seq data. It is paid software. It is available for Windows and Macintosh operating systems. *http://www.dnastar.com/*

2.2 pDraw32

[4] Pdraw32 is shareware tool and may be used unlicensed. It can be used for annotation, DNA cloning, sequence editing, DNA analysis and enzyme selection. It also provides graphical and textual outputs and can draw XY plots and virtual gel plot. It has facility of TM calculator. To run pDraw32 minimum requirement in case of windows is 'Windows 98' operating

system. It can also be installed on Unix and Mac operating system. *http://www.acaclone.com/*

2.3 Sequencher

[6] It is widely used DNA analysis program. It provides latest Network file System (NFS) technology which gives access to superior algorithms, hence offering an easy-to- use point and clickable interface. It is best software for labs working with DNA dataset. It is used across in a wide range of disciplines including genetics for SNP discovery, systematic for evolutionary analysis, forensics for human identification, epidemiology for vaccine development and treatment selection and mutation detection for clinical HIV analysis. This program is available in both Macintosh (OSX) and Windows versions. It is freeware. *http://genecodes.com/*

2.4 MacVector

[7] MacVector is developed by MacVector, Inc. for Macintosh application for biologist, to do various kinds of editing, designing, searching and analysis. MacVector can perform graphical sequence editing, cloning, annotation, primer designing, DNA sequence analysis, protein analysis, database searching, multiple sequence alignment and sequence assembly. Graphical representation developed in MacVector can be saved and can be opened into MS Powerpoint for further modifications.

It provides a range of analysis, which help in decision making. In addition to open reading frame analysis, the program provides various methods to use base or codon composition, which helps to determine whether, open reading frames has the characteristics of protein coding region. In addition to that nucleic acid subsequence analysis, can be done to find motif in the sequence. MacVector13 is latest version and requires Mac OS x10.6 or above with minimum 8 GB of RAM. It does not run on Mac OS 9 or Microsoft Windows based PC computers. *http://www.macvector.com/index.html/*

2.5 Bioedit

[5] Bioedit is a sequence analyzing program written by graduate student. It is intended to supply a single program that can handle most simple sequence and alignment editing and manipulation functions that researchers are likely to do on a daily basis as well as few basic sequences analyses. It is free tool. It is available for Windows operating system. http://www.mbio.ncsu.edu/BioEdit/bioedit.html

2.6 DNA Dynamo

[9] DNA Dynamo software is developed by "Blue Tractor Software Ltd." It is UK based software development company and independent Software Vendor (ISV). This software can be used for sequence navigation, restriction site analysis, abi/scf chromatogram, sequence aligment and editing. It has got

searchable vector data, maps and virtual subcloning features. It requires JAVA 1.5 or later to run. It is available for Windows, Linux and Macintosh Operating system. *http://www.bluetractorsoftware.co.uk/DynamoDemo.htm*

2.7 CLC Free Workbench

[10] It is free software package supporting number of basic bioinformatics features such as creating, editing, sequence alignment, analyzing restriction site, phylogenetics, searching genbank and translation of advanced DNA Sequence to protein. It can handle various file formats. By using CLC Sequence Viewer one can import and export both data and graphical files. It is available in Windows, Mac OSX and Linux. *http://www.clcbio.com/index.php?id=28*

2.8 SnS-Align

[11] SNS-Align (structure and sequence alignment) is developed by Dr. Rakesh K Mishra, Coordinator, Bioinformatics Center, Center for Cellular and Molecular Biology, Hyderabad, India. It is a tool developed in Perl to align protein sequence by combining sequence and structure information at an instance. It is available for windows and Linux. Its primary requirement is Perl installation with Tk module and Emboss. It is available for Windows and Linux Operating systems. It is free software. *http://www.ccmb.res.in/bic/software_snsAlign.php*

2.9 Ugene

[12] Ugene is free open-source cross platform bioinformatics software. It is available for Windows, Macintosh and Linux operating systems. The latest version is Unipro UGENE 1.14. It consists of several features for editing, analyzing, annotation, viewing and searching sequences. In addition to this it has got 3D structure viewer for PDB and MMDB file formats. It has capabilities of protein secondary structure prediction with GOR 1V and PSIPRED algorithms.

It has feature of constructing dot plots for nucleic acid sequence. Dot plots helps in identification of difference between sequences as mutations, invertions, insertions deletions and low-complexity region. *http://ugene.unipro.ru/*

2.10 Emboss

[13] EMBOSS is "The European Molecular Biology Open Software Suite". EMBOSS is a free open source software analysis package specially developed for the needs of the molecular biology users. It is a comprehensive set of sequence analysis programs. It can be used for searching sequence patterns, protein motif identification, codon usage analysis and sequence alignment. It runs on Unix, Windows and Macintosh operating systems. Overall, it is a toolkit for creating robust bioinformatics application. http://emboss.sourceforge.net/

3. CONCLUSION AND SUMMARY

[2] In a developing country like India, bioinformatics has a key role to play in areas like agriculture where it can be used for increasing the nutritional content, increasing the volume of the agricultural produce and implanting disease resistance etc. In the pharmaceutical sector, it can be used to reduce the time and cost involved in drug discovery process particularly for third world diseases, to custom design drugs and to develop personalized medicine. Bioinformatics need based tools will help biologist and researchers to get significant biological insight as well as a global perspective to draw principles in biology

REFERENCES

[1] http://en.wikipedia.org/wiki/Bioinformatics.

[2] Jayaram B., Bhushan K. "Bioinformatics for Better Tomorrow". www.scfbio-iitd.res.in

[3] http://en.bio-soft.net/index.html

[4] http://www.acaclone.com/

[5] http://www.mbio.ncsu.edu/BioEdit/bioedit.html

[6] Sequencher® version 5.2 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com

[7] http://www.macvector.com/index.html

[8] http://www.dnastar.com/

[9] http://www.bluetractorsoftware.co.uk/DynamoDemo.htm

[10] http://www.clcbio.com/index.php?id=28

[11] http://www.ccmb.res.in/bic/software_snsAlign.php

[12] http://ugene.unipro.ru

[13] http://emboss.sourceforge.net