Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 135—138

http://www.ripublication.com/jcib.htm

Biojotter: A Biological Notepad

Rahul V.V., Neetu Jabalia, Ashish Gupta and Jitendra Narayan

Amity Institute of Biotechnology, Amity University, Sector-125, Noida, India.
Corresponding Author E-mail: vvrahul11@gmail.com

Abstract

The watershed moment of genomics, a High-throughput sequencing technology promises to change the current scenario of computational genomics and comparative biology by delivering the tens of thousands of genome in the near future. To deal with huge genomic sequences, biologically compatible notepad is highly required therefore, in order to make biological sequence handling easy, an intuitive biotext editor for biological sequences has been developed, named as BioJotter. We have incorporated additional functions to this tool to interpret and visualize the the biological sequence puzzles, and make it easy-to-use and easy to understand interface. This small biological notebook is capable to do the analysis on the run and display the result in user friendly manner. BioJotter is efficient to handle large dataset, and useful for all users ranging from non-biologists to an expert bioinformaticians. This Notepad can be a replacement for commonly used Microsoft Notepad. BioJotter code is written in Perl 5.10 and Perl/Tk modules.

Keywords: Biotext, Tk/Perl, Blast, Notepad, DNA, Sequence, Notebook, Genomics.

Introduction

Comparative genomics analyse characteristics of whole genomes, and mine biologically meaningful information from the large amount of data. Biological sequence manipulation using bioinformatics techniques and tools became a cumbersome task in front of the biological community (Bond and Schüttelkopf, 2009, Abyzov et al., 2005). The Lack of availability of many important sequence manipulation features on a common platform resulting in the slow analysis of the sequences (McInerney et al., 1998, Rice et al., 2000, Roberts et al., 2006). A specific tool is required to retrieve information from these sequences. BioJotter has been developed to tackle these problems. This small notepad with mouse driven function provide an easy to use interface that can handle many tasks of molecular biology in an efficient manner.

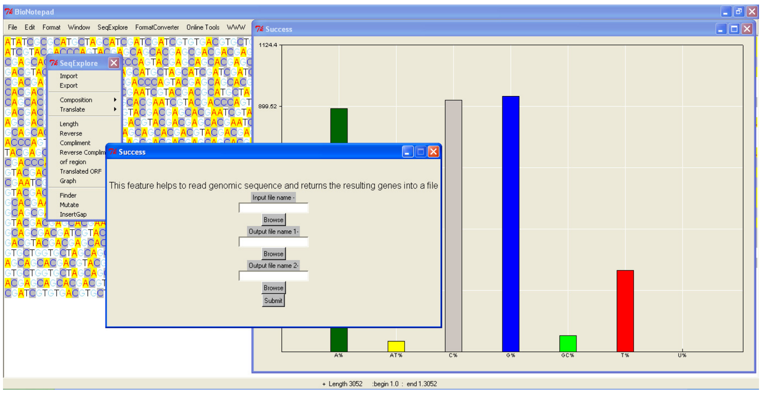
The BioJotter provides sequence computation features that buildup the sequence manipulation an easier task; its main advantage is its user friendly graphical user interface, and direct connectivity with remote servers. The BioJotter has been designed and developed using Perl v.10.5 and various Perl/Tk moduels, which provides an interface for solving the biological jigsaw. Biological sequences contain characters i.e A, T, G, C in their entire line of file. In order to make it more attractive, a special feature introduced in BioJotter by which color of the specific sequence can be changed along with the text area. Pattern, motif, signature, rules are having a lots of importance in computational genetic study, Biojotter have incorporated the feature to highlight the pattern of DNA and Amino acids with different colors. Sequences are provided with default color option which helps to detect the patterns and visualize with an ease. Another feature which is added in BioJotter is, to point out the important location and region of the sequences, is incorporation of click option and track the location with the mouse button. Apart from this the line number feature has been included for better understanding and easily detection of the line of interest. In bioinformatics, genome analysis work random mutation in an area of interested region is required. For that mutator feature along with track and replace option added in this tool, that make sequence handling delightful.

Materials and Methods

BioJotter is currently able to handle RAW and FASTA format with an ease. The format conversion option in this tool converts file to desired format, which can be utilized for further analysis. Using the FormatConvertor, one can convert you genbank, clustalw, and EMBL files to the FASTA fine and start the analysis work with BioJotter. It process the character composition of submitted sequence and display in graphical format. This program helps to start a sequential search of the genome using online available server GenScan and do online BLAST search with an ease. The main modules used in this programme include the LWP::UserAgent; HTTP::Request::Common, Tk::Graph, Tk::NoteBook, Bio::AlignIO, Bio::SeqIO, Tk::widgets qw(Frame Text Label), Tk::BrowseEntry. The Perl module PAR was used to create exe file of the biojotter.

Results and Discussion

Efficacy to handle and visualize the dataset and generate a invaluable biological sense out of it, makes BioJotter a perfect tool for biological experts. It has been tested on Microsoft operating system and found to be effective with larger datasets. Without any delay it also allows the user to do BLAST (Altschul et al., 1990) with online server and display the result, whereas Genscan (Burge and Karlin, 1997) that is being incorporated with this tool save users precious time and do the analysis directly on the server. User can submit multiple FASTA files in case of BLAST and whole genome of an organism in case of Genscan. On the basis of statistical analysis of the residues, it analyze the composition of nucleotide and amino acid residues and creates all the information in colorful graphs[Figure1]. Which makes it more handy to the biologist. To identify ORF regions and for the six frame translation it can also be used. File Format conversion menu allows the user to convert, and work with different formats which include FASTA, Genbank (Benson et al, 2003) , Phylip (Felsenstein et al., 1989) , ClustalW (Larkin et al., 2007), Aln, EMBL (Cochrane. and Galperin, 2010) and DDBJ (Tateno et al., 2002).



This feature helps to read genomic sequence and returns the resulting genes into a file

+ Length 3052 :begin 1.0 : end 1.3052

Figure 1: BioJotter having DNA sequence applied with position based colouring and corresponding statistical graph.The top most windows shows the option of doing blast with the user defined sequence.

With a completely new approach, Biojotter is useful for biologist to manipulate the sequence most effectively. It brings all the frequently used tools and techniques on a common platform which are used by a bioinformatician and biologist. Sequence editing is the primary work in comparative genomics study and biological analysis. That can easily be handled by BioJotter. A number of development plans are already designed which will be available to the next version of BioJotter. The future development plans are as follows:- Multiple Sequence Alignments, Phylogenetic Tree Construction, Protein Structure Visualization, Comparative Genomics section, Sequence Library Function and a History of the biological analysis on this tool.

Acknowledgments

The authors wish to thank Ms. Sunibala Waikhom for all the help shown during the project time. This tool project was supported by Amity Institute of Biotechnology, Amity University, India.

References

[1] Abyzov A., Errami M., Leslin C.M., Ilyin V.A. (2005). Friend, an integrated analytical front-end application for bioinformatics. *Bioinformatics.* 3677-3678.

[2] Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J.(1990). Basic local alignment search tool. *J. Mol. Biol.* 215(3):403–410.

[3] Benson A. D., Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L. (2003). GenBank. *Nucleic Acids Res*. 31(1):23-27.

[4] Bond C.S., Schüttelkopf A.W. (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Biological Crystallography-Acta Cryst*. 510–512.

[5] Burge C., Karlin S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268: 78-94.

[6] Cochrane G.R., Galperin M.Y. (2010). The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. *Nucleic Acids Research.* 38:D1-D4.

[7] Felsenstein J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics.* 5:164-166.

[8] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. (2007). ClustalW and ClustalX version 2. *Bioinformatics.* 23(21): 2947-2948.

[9] McInerney J.O. (1998). GCUA: general codon usage analysis. *Bioinformatics.* 14:372.

[10] Rice P., Longden I., Bleasby A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genetics*. 16:276.

[11] Luthey-Schulten Z. (2006). MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics.* 7:382.

[12] Tateno Y., Imanishi T., Miyazaki S., Fukami-Kobayashi K., Saitou N., Sugawara H., Gojobori T. DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic Acids Res*. 30(1):27–30.