Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp.41—46

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

BioZephyr: A Computational Biology Workbench

Ankit Shukla[1,2], Menorca Chaturvedi[1,2]*, Ramyash Pal[2], Ankur Mohan[2] and Santosh Kumar[2]

[1]SRM University, Chennai, India
[2]BCS-InSilicoBiology, Lucknow, U.P., India
**Corresponding Author Email: menorca.chaturvedi@ymail.com*

Abstract

Various web servers and workbenches have been created up till now for solving different biological problems related to nucleotide and protein sequence analysis . However, in order to find solutions to several miniature problems/functions, a need was felt to either refine existing programs, or use separate online/offline tools for different problems. Several of these offline downloadable tools require the user's system to have pre-installed packages, for instance, PHP-MySQL and often face compatibility issues with operating system versions SYNERGY DE A Professional Series Workbench. It involves a lot of time and effort to download and install these tools. Taking these problems into consideration, we have developed a computational biology workbench, which serves as a platform for carrying out multiple essential functions for nucleotide and protein sequence analysis, with utmost ease in navigability and functionality. The aim of our initiative is to provide a one-point-solution to the user, for carrying out all the fundamental functions which would aid in a project, without having to surf different websites for different tools. This would prevent compatibility issues of results obtained from different servers. The present work was carried out to create an online user-friendly workbench BioZephyr by using Perl, BioPerl modules and CGI.

Keywords: Computational Biology workbench, Perl, BioPerl, CGI, online tools/server.

Introduction

This study has been undertaken keeping in mind the following issues: incompatibility of results obtained from tools developed by different groups(such as file formats), the need to download and configure softwares on a system for using various tools, the lack of a workbench consisting of simple as well as complex sequence analysis tools.

BioZephyr strives to overcome these problems and stand out among the various workbenches which have been developed in the past. Since an array of problems can be solved on this platform, the output of an analysis using one tool of Biozephyr workbench can be used as the input for another, without the need of changing file formats. It provides the user with a host of tools, including a user friendly file format converter, which not only displays the converted file as the output, but also stores the output as a text file on the user's system. Besides this, the complex biological functions like Translation, Transcription and ORF Finder can be solved in a jiffy using the tools provided on this workbench. Another useful tool is the Restriction Cleavage Site Finder, which matches the site to the enzyme, using the Restriction Enzyme Database as a reference. To make the work even simpler, tools for calculating the melting point of short nucleotides, length of nucleotides, GC content and molecular weight of nucleotide sequences have been provided on Biozephy workbench. Tools, currently under development include GBParse and RemoteBlast. GBParse would provide the user with customized results from GenBank, that is, the user could pre-select the categories of GenBank format which user would like to view for all the sequences uploaded. RemoteBlast, built using BioPerl modules [3]Peter Schattner, BioperlTutorial, permits the user to carry out similarity searches with the aid of NCBI Server [4] National Center for Biotechnology Information. BioZephyr workbench, in its initial stages, has been tested on local server.CGI has been used to connect the Perl scripts [1] James Tisdall , 'Beginning Perl for Bioinformatics' and HTML, enabling the scripts to run within the browser efficiently. We shall soon be hosting it on the web server, with its own domain.

Materials and Methodology

BioZephyr has been designed using Perl, BioPerl and CGI. Perl was our preferred choice of scripting language, as it enables compact codes to be written for complex problems. CGI, or, Common Gateway Interface, has been used to create an interactive web site, connecting Perl to HTML, and adding onto its functionality. Scripts for Transcription, Translation, Reverse Complementary, Pattern Matching, ORF Finder (Open Reading Frame) and Restriction Cleavage Site Finder have been written in CGI Perl. BioPerl modules like Bio::SeqIO and Bio::Tools::Run::RemoteBlast [3]Peter Schattner, BioperlTutorial] were used to create the FileFormat Converter and RemoteBlast tools. The FileFormat Converter was designed using CGI Perl scripting [1] James Tisdall , 'Beginning Perl for Bioinformatics' and BioPerl module [3] Peter Schattner, BioperlTutorial to interconvert Fasta, GenBank and EMBL file formats. Dreamweaver 8.0 has been used to design the webpage of BioZephyr web site and PHP pages. The workbench was validated on WAMP Server (Windows, Apache, MySQL, PHP), which consists of independent packages installed on the system.

Results and Discussion

BioZephyr is an easy-to-use workbench, having a collection of fundamental tools, required while working on Bioinformatics. Figure-1 shows the homepage designed for BioZephyr workbench that would be soon hosted with its own domain name

Figure 1: Showing the homepage of BioZephyr Workbench.

The FileFormat Converter is a very useful tool, which aids the user to interconvert formats for Fasta, GenBank and EMBL files. The user needs to upload the file containing the sequence or paste the sequence in the text box provided followed by selecting a conversion option from the drop down menu located on the web page. In case there is a mismatch between the type of file uploaded and the conversion selected, the user will immediately be shown an error. If the selection is correct, the user's screen will show the contents of the converted file, while simultaneously saving the result as a text file on the system. Thus, there remains no need of manually downloading the result file. This is demonstrated as a screenshot in Figure 2, shown below.

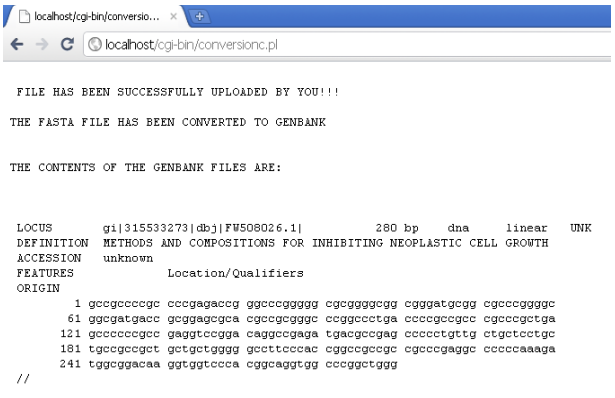

Figure 2: Showing the result of File Format Converter.

The GBParse tool, built using CGI Perl and Bioperl, provides the user with customized results from GenBank, that is, the user could pre-select the categories of

GenBank format which he would like to view for all the sequences uploaded. The 'ORF Finder' is another useful tool and displays the six-frame translation of a protein sequence as shown in Figure 3

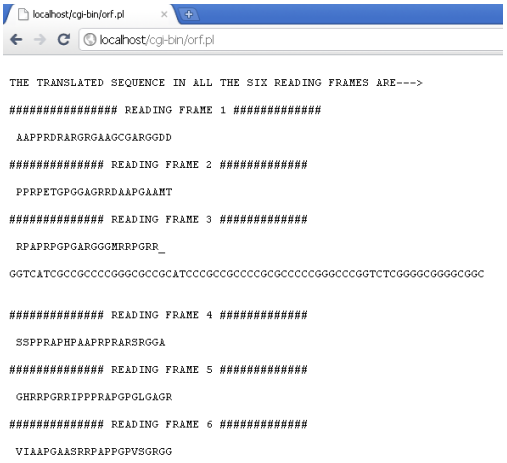

Figure 3: Showing the results of ORF Finder tool.

Apart from this, tools like Melting Point Calculator, Molecular weight Finder, GC Content and Length Finder were developed using CGI Perl scripts. Simple in nature, yet, extremely useful in carrying out sequence analysis, is the 'Melting Point Calculator'. This tool comes in handy when the user needs to calculate the melting point of short nucleotide sequences [7] DNA interactions of monofunctional organometallic ruthenium(II) antitumor complexes in cell-free media.

Novakova O Et. al. The user simply needs to enter the nucleotide sequence, and the melting point is displayed on the screen. A screenshot of one such result is shown here in Figure 4.

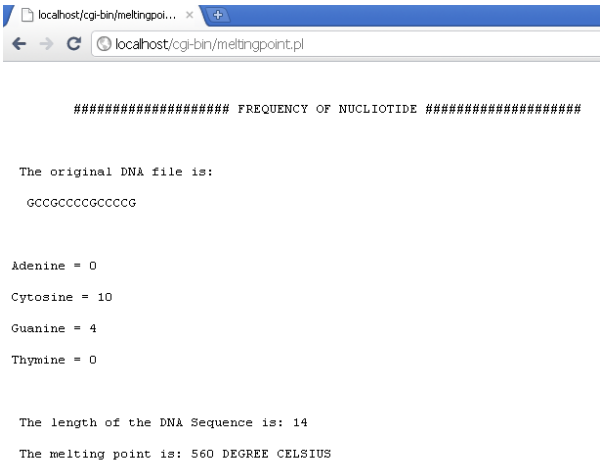

Figure 4: Showing the frequency of nucleotides, length and Melting point.

These tools, though simple in nature, are not readily available on a single platform. Hence, the effort was made to provide a multifunctional workbench to cater to all the needs of a sequence analyser.

Conclusion

This workbench which incorporates a user friendly environment along with specific online web server tools for finding solutions to a large spectrum of small biological problems/functions related to sequence analysis and would be of immense utility in the areas of computational biology, sequence analysis and bioinformatics/biotechnology . BioZephyr will help save time and effort of the user in using multiple tools/web servers. Hence, this effort provides a multifunctional workbench to cater to all the needs of the user.

BioZephyr is also expected to offer basal support in terms of curtailing the efforts in finding adequate platform for PCR design, combining multiple sequence alignments, prediction of RNA Secondary structures, downloading pdb files and other visualization tools which are currently under development.

Acknowledgement

We wish to thank Mr. Ramyash Pal at BCS-Insilico Biology, Lucknow for his help and guidance throughout the designing and development process. We also thank Dr. A.K. Shukla, SGPGI, Lucknow for proofreading this manuscript and giving us his valuable time and suggestions.

References

[1] James Tisdall ,'Beginning Perl for Bioinformatics'

[2] SWAMI: Integrating Biological Databases and Analysis Tools Within User Friendly Environment (2007) Rami Rifaieh, Roger Unwin, Jeremy Carver, and Mark A. Miller, 4th International Workshop, DILS Data Integration in the Life Sciences, Philadelphia, PA,USA, June 2007, pp.48-58, LNBI4544, ISBN 978-3-540-73254-9.

[3] The Biology Workbench: http://workbench.sdsc.edu

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

[5] Peter Schattner, *BioperlTutorial* - a tutorial for Bioperl: www.bioperl.org/Core/Latest/bptutorial.html

[6] The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information. http://www.ncbi.nlm.nih.gov/

[7] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped

BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

[8] Pruess M., Fleischmann W., Kanapin A., Karavidopoulou Y., Kersey P., Kriventseva E., Mittard V., Mulder N., Phan I., Servant F., and Apweiler R. (2003) The Proteome Analysis database: a tool for the *in silico* analysis of whole proteomes. *Nucleic Acids Research 31: 414-417.*

[9] DNA interactions of monofunctional organometallic ruthenium(II) antitumor complexes in cell-free media. Novakova O, Chen H, Vrana O, Rodger A, Sadler PJ, Brabec V. Biochemistry. 2003 Oct 7;42(39):11544-54.

[10] SYNERGY DE A Professional Series Workbench: http://synergyde.com/products/development_tools/workbench_on_unix.aspx