International Journal of Advanced Technology in Engineering and Science www.ijates.com
Volume No.02, Issue No. 08, August 2014 ISSN (online): 2348 – 7550

A GENERAL PURPOSE SOFTWARE SUITE FOR MULTIPLE SEQUENCE ALIGNMENT

Farhana Sarkhwas[1] **, R.K. Kamat**[2]

[1,2]*Department of Computer Science, Shivaji University, Kolhapur(India)*

ABSTRACT

There have been many software programs implemented for the inference of multiple sequence alignments of protein and DNA sequences. Every year there is development of new bioinformatical tools without taking into consideration the fact that how to make easy to run and study these tools. Consequently, the design of an integrative environment for bioinformatical tools is a very important and challenging task to ease the job of the biologists. The current bioinformatical suite written in java which is described in this paper executes four different representative programs for multiple sequence alignment. The Graphical User Interface (GUI) of the suite will be quite user friendly by prompting dialog box, provide help file for execution, physicochemical properties of sequences and saves results.

Keywords : Multiple Sequence Alignment, GUI, Java, Statistical Properties, Algorithms

I. INTRODUCTION

Multiple Sequence Alignment has been an active area of study in the field of computer science and bioinformatics for which a maximum number of algorithms and software tools have been developed quite fast and independently without much concerning the need for standardization. Thousands of biological sequences are available on which biologists have to perform multiple sequence alignment and extract the statistical properties or representation techniques. The biologists normally have to download each tool of their requirement to solve their purpose, hence to get the desired result have to depend on more than one tool. Secondly, have to face the difficulty of executing the software, platform dependency, converting the biological sequences to other formats as per required etc. Biologists are still struggling with the GUI’s of the tools to execute the tools and get the known results.

This study will discuss a general purpose design for bioinformatical suite for analysis of multiple sequence alignment. The proposed suite will be an integration of STRAP, Geneious, DNAMAN and ClustalX (Table 1). These tools are publically free and are platform independent and execute the maximum features related to multiple sequence alignment which will solve the problem of bioinformaticians. The tools to be included in the suite will analyze the biological data for analysis, representation and statistical properties of multiple sequence alignment.

The current suite has one package containing the tools which exhibit the most related properties of multiple sequence alignment. This suite will have a very user friendly environment. Just with a single click the tool gets executed, thus saves the biologists to learn the execution steps of the software. The GUI will show WYSIWYG (what you see is what you get) features.

Similar study was carried out by (Zainuddin et.al., 2011) [1] to support analysis and learning at an undergraduate level for linux operating system. Few bioinformatical suites are available like KDE Bioscience [2], The BioTools Suite [3] and SuiteMSA [4] which are used for sequence analysis comparison, benchmarking and graphical representation. But, the current study aims at developing a plugin for tool included in the suite also will execute small java program for calculating the statistical properties of proteins and showing the graphical representation of amino acids.

Name	Description	Sequence Type*	Alignment Type**	Link	Year	Author	File Format Supported
Geneious	Progressive & Iterative alignment;	Both	Both	www.geneious.com	2008	A.J. Drummond	Fasta, pdb,pir,aln, embl,gb,gde
DNAMAN	Multiple Sequence Alignment	Both	Both	www.lynnon.com	2005	Lynnon	Fasta,gb,pdb, gde
Strap	Multiple Sequence Alignment Tool	Both	Both	http://www.charite.de/bioinf/strap	2004	Christoph Gille	Fasta,aln,pdb ,gb,
Clustal X	Progressive Alignment	Both	Both	www.clustal.org	1994	Julie Thompson & Toby Gibson,	Aln,fasta,pdb ,pir,embl,gb, gde,swp

***Sequence Type TABLE 1: Basic Information of the four Multiple Sequence Alignment Tools**

: Protein or nucleotide. ****Alignment Type:** Local or global

II.OBJECTIVE

The objective of the current study is to give one single platform to the bioinformaticians to perform various functions and representation techniques for multiple sequence alignment. The design of the suite will standardize the performance to analyze the biological data.

III. SYSTEM REQUIREMENTS

The GUI of the suite will be designed using the java language and will be tested on windows platform. The basic hardware requirement of the system for the suite would be 512MB of RAM and 1GB of storage. The main window of the GUI shows the icons of the tools to be executed (Fig.1). It will have a special option i.e. Tool

Help which will graphically explain the functions (load biological data, perform multiple sequence alignment, phylogenetic tree etc.) provided by the tools related to multiple sequence alignment.

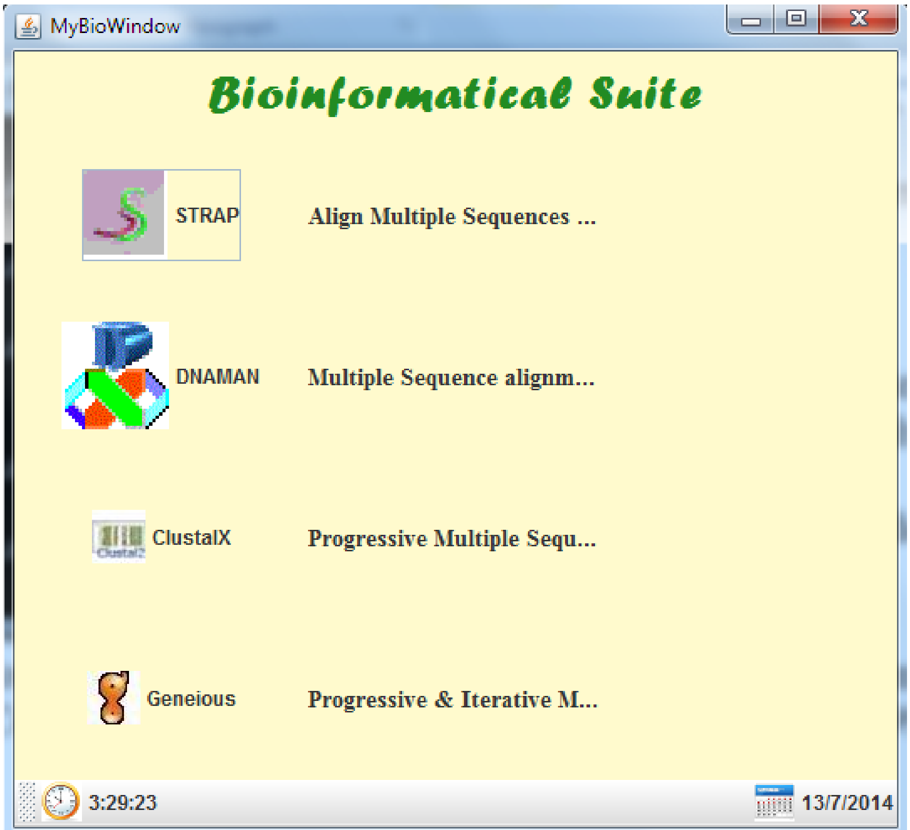

Fig.1 : Main GUI screen of the Bioinformatical Suite

IV. CONCLUSION

The proposed suite is a collection of multiple sequence alignment tools will work on windows operating system with the java environment. The development of the suite is in the early stages showing the framework, execution of tools and a program to calculate and save the statistical properties of protein sequence. The current study is just the beginning towards the final state of the Bioinformatical suite.

REFERENCES

[1] Zainuddin, S.; Hashim, A.R.M.; Shukur, Z.; Keong, L.K.; Hamid, M.N.; Hussein, Z.A.M., A Proposed Design for a suite of Bioinformatics analysis Tool, 2011, *Information Technology Journal 10(2)*, 434-439.

[2] Lu, Q.; Hao, P.; Curcin, V.; He, W.; Li, Y.Y.; Luo, Q.M.; Guo, Y.K.; Li, Y.X., KDE Bioscience: platform for bioinformatics analysis workflows, 2006, *JBiomedInformation, 39(4)*, 440-450.

[3] Wishart D.S. and Fortin,S. The BioTools suite A Comprehensive Suite of Platform-Independent Bioinformatics Tools, 2001, *Mol Biotechnol, 19(1)*, 59-77

[4] Catherine L. A., Cory L. S. and Etsuko N. M. ,SuiteMSA: visual tools for multiple sequence alignment comparison and molecular sequence simulation, 2011, *Bioinformatics, 12(184)*, 1471-2105