# Journal of
# SCIENCE, TECHNOLOGY
# & MANAGEMENT

# SEQUENCE ANALYZER- ANALYZING DNA SEQUENCE USING JAVA PROGRAMS

**J. Christina[1], S. Priya[2], E. Nandha Devi[1]**

### Abstract

BioJava is a Java framework for processing biological data, it's a programming language, coding are readily available for analyzing DNA sequence but many biologists are not familiar with the coding and its usage. Hence, to address this concern, the present project is initiated to evolve a GUI driven Sequence Analyzer, which enables the biologists to perform basic tasks related to DNA like Transcription, Translation, Reverse Complement, Residue Count and Sequence Similarity, without code level programming. This tool will be further developed in future by incorporating BioPython and BioPerl to make it an eminent tool in the biologists' society.

*Keywords:* BIOJAVA, JAVA, DNA, GUI, HTML, Sequence Analyzer, BIOPYTHON, BIOPERL.

## INTRODUCTION

BioJava was developed in 1999; it is an open-source project for processing of biological data in the Java programming language, (Prlić A, 2012). BioJava is a collection of modules which consists of programs, that helps the users to retrieve and analyze DNA, RNA and PROTEIN sequences (http://www.di.unito. it/~botta/didattica/biojavaTutorial.pdf). If a user wants to analyze his/her sequence then one has to download the BioJava jars into a suitable platform (e.g. NetBeans IDE) and run the suitable program to get the desired results. But, the drawback is not many biologists are familiar with using the software and coding's. On the other hand, the generation and analysis of high-throughput sequencing data are becoming a major component of many studies in molecular biology and medical research (Priya Sankar *et al*, 2010). Analyzing the DNA data generated plays a major role in the field of Bioinformatics, sequencing the DNA has become an easy task but analyzing them and getting the information is still a difficult task, many software's are being designing to solve this problem. One such application of BioJava designed to reduce the problem of analyzing DNA data and to extract useful information from the DNA data is Sequence Analyzer. Sequence Analyzer is at its beginning stage of development where only basic operations can be performed, it will be further developed where the output will show the full Annotations of genomes, Phylogenetic analysis and Treeview (Alok J. Saldanha et al, 2004) at just one click for the sequence of his/her choice. Much software's and tools are

1. *Department of Bioinformatics, Stella Maris College (Autonomous), 17, Cathedral Road, Chennai - 86.* sagayanilo @gmail.com
2. *Xinnovem Architecting Discovery, Chennai.*

available for protein annotation but not much for genome annotation, Sequence Analyzer will be proven to be a better genome annotator in near future and of great use to the biological society.

## METHODOLOGY

Sequence Analyzer has been developed using HTML, SQL Server 2005, JAVA, and JSP on NetBeans IDE with GlassFish Server. A snapshot of the tool and its usage is shown below

### RESULTS

*Program Input:* Select a task and upload the DNA sequence.

*Program Output:* Output is generated as per the selected task, for e.g. Transcription ATGC – UACG.

## DISCUSSION

In this post genomic-era, data's are growing enormously from biological sources and the
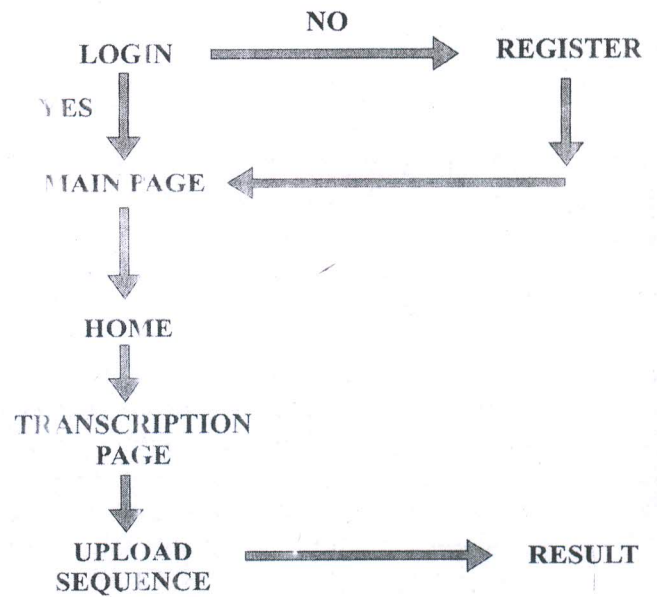
**METHOD TO ACCESS THE TOOL**



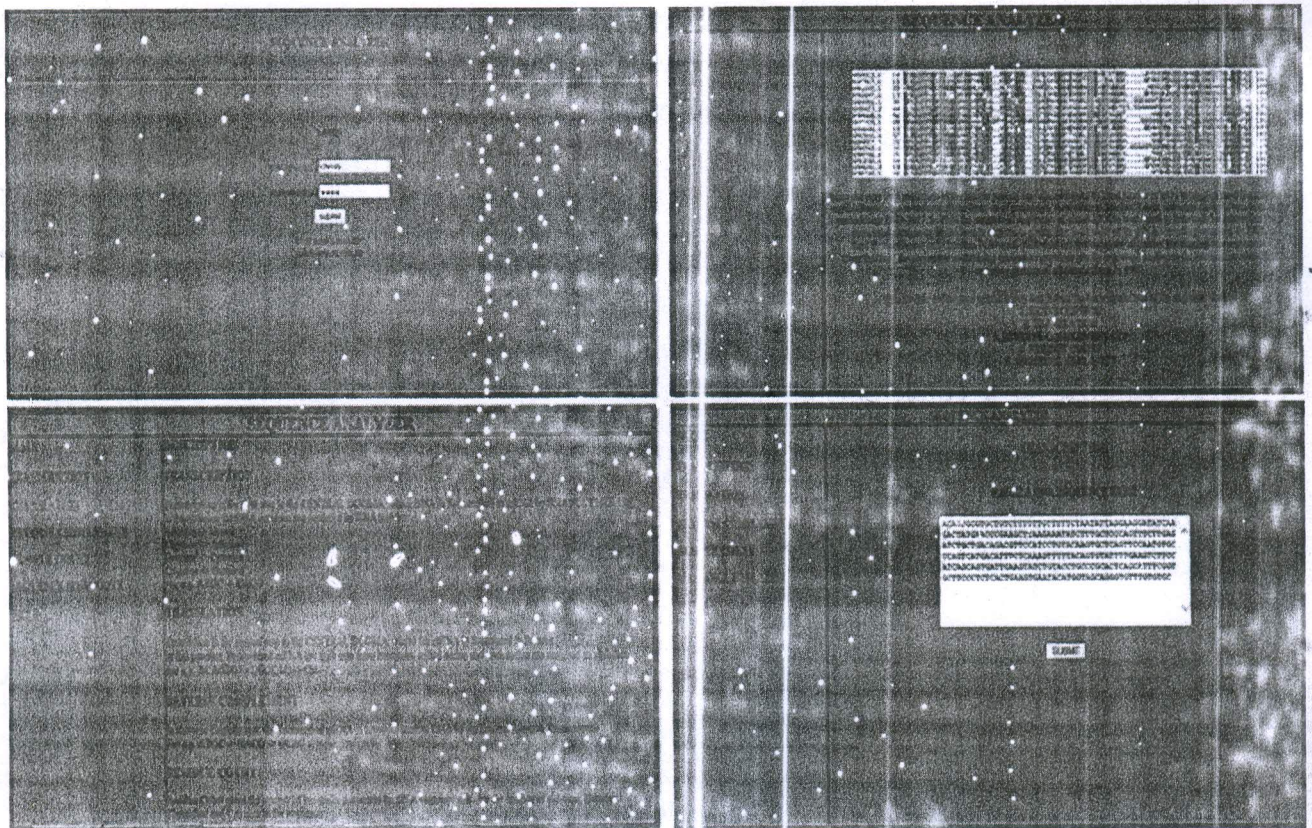Fig. 1: Diagrammatic representation of the method to access the tool.



Fig. 2: This snapshot shows the Login page, Main page, Home page and Transcription page.

demand for tools to interpret these data's are also increasing. In order to fill the gap in interpretation of data's BioProjects were developed. Different views about BioProjects from various authors are discussed in the following. According to (Pjotr Prins *et al*, 2012) open-source software (OSS) encourages computer programmers to reuse software components written by others. In evolutionary bioinformatics, OSS comes in a broad range of programming languages, including C/C++, Perl, Python, Ruby, Java, and R. To avoid writing the same functionality multiple times for different languages, it is possible to share components by bridging computer languages and Bio* projects, such as BioPerl, BioPython, BioRuby, BioJava, and R/Bioconductor. BioJava was spawned in the mid-90s as a result of the computational needs of Matthew Pocock and Thomas Down, two Ph.D. students at the Sanger Institute in Cambridge, England (Steven Meloan, 2004). BioJava will probably remain the toolkit of choice only of well-established Java Programmers (Harry Mangalam, 2002). It's an open source, platform independent, can handle large amount of data, execution is faster, bug free, coding's are very simple to use, visualization is possible, cross – platform and etc.., The recently formed BioInformatics Centre (BIC) at the National University of Singapore (NUS) provides access to various commonly used computational tools available over the World Wide Web (WWW) - using a uniform interface and easy access. They have also come up with a new database tool, BioKleisli©, which allows you to interact with various geographically scattered, heterogeneous, structurally complex and constantly evolving data sources. (Kolatkar P. R. *et al*, 1998). DNA algorithm was implemented using symmetric coding in BioJava and MatLab (Mircea-Florin Vaida et al, 2011).

Pocock used NetBeans IDE for developing BioJava, this information gave me an idea to use NetBeans IDE with GlassFish Server, and it's a platform which serves to develop various appli-cations. HTML (HyperText Markup Language) pages as front end, JSP (Java Servlet Pages), JAVA programs and Microsoft SQL server 2005 as back end, with these materials Sequence Analyzer was developed which does basic five operations, they are Transcription, Translation, Reverse complement, Residue count and Sequence similarity. Therefore the tool has been successfully developed and gives the desired output. This tool will be very useful for the biologists in their research work. This tool will be further developed in future by integrating BIOJAVA, BIOPERL, and BIOPYTHON in this application.

## Acknowledgments

## References

*Alok J. Saldanha et al, Java Treeview—extensible visualization of microarray data, 2004, http://bioinformatics.oxfordjournals.org/content/20/17/3246.full.pdf+html (retrieved May 2013).*

*Andreas Prlić et al, BioJava: an open-source framework for bioinformatics in 2012, 2012, http://bioinformatics.oxfordjournals.org/content/28/20/2693.long (retrieved April 2013).*

*Harry Mangalam, The Bio*Toolkits - A Brief Overview, 2002, http://bib.oxfordjournals.org/content/3/3/296.full.pdf, (retrieved May 2013).*

*Kolatkar P. R. et al., P. R. Development Of Software Tools At Bioinformatics Centre (Bic) At The National University Of Singapore (Nus), 1998, http://psb.stanford.edu/psb-online/proceedings/psb98/kolatkar.pdf. (retrieved May 2013).*

*Mircea Florin Vaida et al, Alternative DNA Security using BioJava. 2011, http://users.utcluj.ro/~vaida/Publications/Vaida_DICTAP_2011_ID96.pdf. (retrieved May 2013).*

*Pjotr Prins et al, Sharing Programming Resources Between Bio* Projects Through Remote Procedure Call and Native Call Stack Strategies,2012, http://link.springer.com/protocol/10.1007/978-1-61779-585-5_21, (retrieved May 2013).*

*Priya Sankar et al, BDTK (Biological Database Tool Kit), November 2010,http://www.bmjournal.in/admin/photo/1362029178BDTK%20-%20November.pdf. (retrieved May 2013).*

*Steven Meloan, BioJava - Java Technology Powers Toolkit for Deciphering Genomic Codes. June 2004, http://www.oracle.com/technetwork/articles/javase/biojava-140011.html (retrieved April 2013).*

*URL -http://www.di.unito.it/~botta/didattica/biojavaTutorial.pdf. (retrieved May 2013).*