**Course: Advanced Bioinformatics**

**Module title: Introduction to BioJava**

**Module no. : 192**

This module depicts the light on BioJava.

Collection of Java objects that represent and manipulate Biological Data

Not a program, rather a programming library

Open source (LGPL): open for all development, even commercial.

Not `sticky' or `viral

BioJava is an open source project dedicated to providing Java tools for processing biological data. BioJava is a set of library functions written in the Java programming language for manipulating sequences, protein structures, file parsers, CORBA interoperability, DAS, access to AceDB, dynamic programming, and simple statistical routines. BioJava supports a huge range of data, starting from DNA and protein sequences to the level of 3D protein structures. The BioJava libraries are useful for automating many daily and mundane bioinformatics tasks such as to parsing a PDB file, interacting with Jmol and many more. This Application programming interface (API) provides various file parsers, data models and algorithms to facilitate working with the standard data formats and enables rapid application development and analysis. These libraries have also been used in development of various extended analysis tools, for example:

* MUSI: an integrated system for identifying multiple specificity from very large peptide or nucleic acid data sets.
* JEnsembl: a version-aware Java API to Ensembl data systems.
* Expression profiling of signature gene sets with trinucleotide threading
* Resolving the structural features of genomic islands: a machine learning approach
* Utility library for structural bioinformatics

BioJava – History

Started at EBI/Sanger in 1998 by Matthew Pocock and Thomas Down

In the year 2008, BioJava's first Application note was published. It was migrated from its original CVS repository to Git hub on April 2013. In October 2012, the most recent paper on BioJava was published. As of November 2012 Google Scholar counts more than 130 citations.

Features

BioJava provides software modules for many of the typical tasks of bioinformatics programming. These include:

Accessing nucleotide and peptide sequence data from local and remote databases

Transforming formats of database/ file records

Protein structure parsing and manipulation

Manipulating individual sequences

Searching for similar sequences

Creating and manipulating sequence alignments

1100+ classes, 130,000 + lines of code.

Uses CVS version control, JUnit testing, ANT builds.

It now has a fairly stable API. 76 packages!

Where to Find

Home Page: www.biojava.org

BioJava in Anger: http://www.biojava.org/docs/bj\_in\_anger/

Mailing Lists:

biojava-l@biojava.org

biojava-dev@biojava.org

Nightly Builds: http://www.derkholm.net/autobuild/

Download: www.biojava.org/download/

Binaries, source & docs

Compiling: Requires ANT build tool

http://jakarta.apache.org/ant/