**Course: Advanced Bioinformatics**

**Module title: BioPython Basics**

**Module no. : 201**

**What is Biopython?**

The Biopython Project is an international association of developers of freely available Python ([http://www.python.org](http://www.python.org/)) tools for computational molecular biology. Python is an object oriented, interpreted, flexible language that is becoming increasingly popular for scientific computing. Python is easy to learn, has a very clear syntax and can easily be extended with modules written in C, C++ or FORTRAN.

The Biopython web site ([http://www.biopython.org](http://www.biopython.org/)) provides an online resource for modules, scripts, and web links for developers of Python-based software for bioinformatics use and research. Basically, the goal of Biopython is to make it as easy as possible to use Python for bioinformatics by creating high-quality, reusable modules and classes. Biopython features include parsers for various Bioinformatics file formats (BLAST, Clustalw, FASTA, Genbank,...), access to online services (NCBI, Expasy,...), interfaces to common and not-so-common programs (Clustalw, DSSP, MSMS...), a standard sequence class, various clustering modules, a KD tree data structure etc. and even documentation.

Basically, we just like to program in Python and want to make it as easy as possible to use Python for bioinformatics by creating high-quality, reusable modules and scripts.

**What can I find in the Biopython package**

The main Biopython releases have lots of functionality, including:

* The ability to parse bioinformatics files into Python utilizable data structures, including support for the following formats:
  + Blast output – both from standalone and WWW Blast
  + Clustalw
  + FASTA
  + GenBank
  + PubMed and Medline
  + ExPASy files, like Enzyme and Prosite
  + SCOP, including ‘dom’ and ‘lin’ files
  + UniGene
  + SwissProt
* Files in the supported formats can be iterated over record by record or indexed and accessed via a Dictionary interface.
* Code to deal with popular on-line bioinformatics destinations such as:
  + NCBI – Blast, Entrez and PubMed services
  + ExPASy – Swiss-Prot and Prosite entries, as well as Prosite searches
* Interfaces to common bioinformatics programs such as:
  + Standalone Blast from NCBI
  + Clustalw alignment program
  + EMBOSS command line tools
* A standard sequence class that deals with sequences, ids on sequences, and sequence features.
* Tools for performing common operations on sequences, such as translation, transcription and weight calculations.
* Code to perform classification of data using k Nearest Neighbors, Naive Bayes or Support Vector Machines.
* Code for dealing with alignments, including a standard way to create and deal with substitution matrices.
* Code making it easy to split up parallelizable tasks into separate processes.
* GUI-based programs to do basic sequence manipulations, translations, BLASTing, etc.
* Extensive documentation and help with using the modules, including this file, on-line wiki documentation, the web site, and the mailing list.
* Integration with BioSQL, a sequence database schema also supported by the BioPerl and BioJava projects.

**Installing Biopython**

All of the installation information for Biopython was separated from this document to make it easier to keep updated.

The short version is go to our downloads page (<http://biopython.org/wiki/Download>), download and install the listed dependencies, then download and install Biopython. Biopython runs on many platforms (Windows, Mac, and on the various flavors of Linux and Unix). For Windows we provide pre-compiled click-and-run installers, while for Unix and other operating systems you must install from source as described in the included README file. This is usually as simple as the standard commands:

python setup.py build

python setup.py test

sudo python setup.py install

(You can in fact skip the build and test, and go straight to the install – but its better to make sure everything seems to be working.)

The longer version of our installation instructions covers installation of Python, Biopython dependencies and Biopython itself. It is available in PDF (<http://biopython.org/DIST/docs/install/Installation.pdf>) and HTML formats (<http://biopython.org/DIST/docs/install/Installation.html>).