

Introduction to Biopython

Talk to a Teacher

<http://spoken-tutorial.org>

National Mission on Education through ICT

<http://sakshat.ac.in>

Script: Snehalatha Kaliappan

IIT Bombay

15 May 2015



Learning Objectives



Learning Objectives

- ▶ Important features of Biopython



Learning Objectives

- ▶ **Important features of Biopython**
- ▶ **Download and installation for linux OS**



Learning Objectives

- ▶ Important features of Biopython
- ▶ Download and installation for linux OS
- ▶ Translation of a DNA sequence to a protein sequence using Biopython tools



Pre-requisites



Pre-requisites

- ▶ **Familiar with Undergraduate Biochemistry or Bioinformatics**



Pre-requisites

- ▶ **Familiar with Undergraduate Biochemistry or Bioinformatics**
- ▶ **Basic Python programming**



Pre-requisites

- ▶ Familiar with Undergraduate Biochemistry or Bioinformatics
- ▶ Basic Python programming
- ▶ Refer to Python Spoken Tutorials at <http://spoken-tutorial.org>



System Requirements



System Requirements

- ▶ **Ubuntu OS version 12.04**



System Requirements

- ▶ **Ubuntu OS version 12.04**
- ▶ **Python version 2.7.3**



System Requirements

- ▶ **Ubuntu OS version 12.04**
- ▶ **Python version 2.7.3**
- ▶ **IPython interpreter version 0.12.1**



System Requirements

- ▶ **Ubuntu OS version 12.04**
- ▶ **Python version 2.7.3**
- ▶ **IPython interpreter version 0.12.1**
- ▶ **Biopython version 1.58**



About Biopython



About Biopython

- ▶ **Biopython is a collection of modules for computational biology**



About Biopython

- ▶ Biopython is a collection of modules for computational biology
- ▶ It can perform most basic to advanced tasks required for bioinformatics



Biopython functionality for bioinformatics



Biopython functionality for bioinformatics

- ▶ **Parsing (Extracting information) from various file formats (FASTA, Genbank etc)**



Biopython functionality for bioinformatics

- ▶ Parsing (Extracting information) from various file formats (**FASTA**, **Genbank** etc)
- ▶ Download data from database websites (NCBI, ExPASy)



Biopython functionality for bioinformatics

- ▶ Parsing (Extracting information) from various file formats (**FASTA**, **Genbank** etc)
- ▶ Download data from database websites (**NCBI**, **ExPASy**)
- ▶ Run bioinformatic algorithms such as **BLAST**



Biopython functionality for bioinformatics



Biopython functionality for bioinformatics

- ▶ Tools for performing common operations on sequences (complements, transcription, translation etc)



Biopython functionality for bioinformatics

- ▶ Tools for performing common operations on sequences (complements, transcription, translation etc)
- ▶ Code for dealing with alignments



Biopython functionality for bioinformatics

- ▶ Tools for performing common operations on sequences (complements, transcription, translation etc)
- ▶ Code for dealing with alignments
- ▶ Code to split up tasks into separate processes



Download



Download

- ▶ Biopython package is not part of Python distribution, so it needs to be downloaded independently



Download

- ▶ Biopython package is not part of Python distribution, so it needs to be downloaded independently



<http://biopython.org/wiki/Download>



Installation on Linux OS



Installation on Linux OS

- ▶ Install Python, IPython and Biopython packages using Synaptic Package Manager



Installation on Linux OS

- ▶ Install Python, IPython and Biopython packages using Synaptic Package Manager
- ▶ Prerequisite software will be installed automatically



Installation on Linux OS

- ▶ Install Python, IPython and Biopython packages using Synaptic Package Manager
- ▶ Prerequisite software will be installed automatically
- ▶ Additional packages must be installed for graphic outputs and plots



Task

- ▶ (1) Create a sequence object for the coding DNA strand
- ▶ (2) Transcription of coding DNA strand to mRNA
- ▶ (3) Translation of mRNA to a protein sequence



Sequence Object

- ▶ **'ATGTTACACTCCCGATGA'**
- ▶ **Create a sequence object for the above coding DNA strand**



Summary

- ▶ **Important features of Biopython**
- ▶ **Download and installation on Linux OS**
- ▶ **Convert the DNA sequence into sequence object**



Summary

- ▶ Transcription of the DNA sequence to mRNA (**'transcribe'** method)
- ▶ Translation of mRNA to protein sequence (**'translate'** method)



Assignment



Assignment

- ▶ Translate the DNA sequence to a protein sequence
- ▶ **'ATGGCCCTATAGTGTCTAAGCTAG'**
- ▶ The output shows an internal stop codon
- ▶ Translate the DNA strand till first in frame stop codon



About the Spoken Tutorial Project

- ▶ Watch the video available at http://spoken-tutorial.org/What_is_a_Spoken_Tutorial
- ▶ It summarises the Spoken Tutorial project



About the Spoken Tutorial Project

- ▶ Watch the video available at http://spoken-tutorial.org/What_is_a_Spoken_Tutorial
- ▶ It summarises the Spoken Tutorial project
- ▶ If you do not have good bandwidth, you can download and watch it



Spoken Tutorial Workshops

The Spoken Tutorial Project Team

- ▶ Conducts workshops using spoken tutorials
- ▶ Gives certificates to those who pass an online test
- ▶ For more details, please write to contact@spoken-tutorial.org



Acknowledgements

- ▶ Spoken Tutorial Project is a part of the Talk to a Teacher project
- ▶ It is supported by the National Mission on Education through ICT, MHRD, Government of India
- ▶ More information on this Mission is available at

<http://spoken-tutorial.org/NMEICT-Intro>

