Writing Sequence Files

Talk to a Teacher
http://spoken-tutorial.org
National Mission on Education through ICT
http://sakshat.ac.in

Snehalatha Kaliappan
IIT Bombay

15 September 2015

Learning Objectives

Learning Objectives

- **Create Sequence Record Objects**

Learning Objectives

- **Create Sequence Record Objects**
- **Write sequence files**

Learning Objectives

- **Create Sequence Record Objects**
- **Write sequence files**
- **Convert between file formats**

Learning Objectives

- **Create Sequence Record Objects**
- **Write sequence files**
- **Convert between file formats**
- **Sort records in a file by length**

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 14.10**

System Requirements

- **Ubuntu OS version 14.10**
- **Python version 2.7.8**

System Requirements

- Ubuntu OS version 14.10
- Python version 2.7.8
- IPython interpreter version 2.3.0

System Requirements

- Ubuntu OS version 14.10
- Python version 2.7.8
- IPython interpreter version 2.3.0
- Biopython version 1.64

SeqIO Functions

SeqIO Functions

- **Read contents of the file: Parse and Read**

SeqIO Functions

- **Read contents of the file: Parse and Read**
- **Writing sequences to a file: write**

SeqIO Functions

- **Read contents of the file: Parse and Read**
- **Writing sequences to a file: write**
- **Interconvertion between various file formats: convert**

Sequence Record Objects

Sequence Record Objects

- **It is the basic data type for the sequence input/output interface**

Sequence Record Objects

- **It is the basic data type for the sequence input/output interface**
- **A sequence is associated with higher level features such as identifiers and descriptions**

Limitations of Convert function

- **Writing some formats requires information which other files formats don't contain**

- **For example: Can convert a GenBank file to a FASTA file, can't do the reverse**
  Can convert a FASTQ file to a FASTA file, can't do the reverse

Summary

- **Create Sequence Record Objects**
- **Write sequence files using write function of Sequence Input/Output module**
- **Convert between sequence file formats using convert function**
- **Sort records in a file by length**

Assignment

Assignment

- **Extract the gene HIV1gp3 at positions 4587 to 5165 from the genomic sequence of HIV**
- **The file HIV.gb is included in code files of this tutorial**

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**