

LEARN PYTHON & R FOR BIOINFORMATICS

Prerequisite Terminologies

Before proceed to learn the main topic, you should be familiar with the followings in order to have thorough understanding

- Accession Number
- Sequence Retrieval using NCBI

FASTA (Format)

FASTA is a text-based format file containing Biological Sequence that is used to organize, sequence and store the Biological Data. It is one of the simplest and widely used formats in Bioinformatics. The Biological Sequence can be either in the form of nucleotides or amino acids in which nucleotides or amino acids are represented using single-letter codes. The format also allows for sequence names and comments to precede the sequences. The first line of the format consists of the description of the sequence and the second line initiates with the sequence.

Syntax

The basic syntax of the typical FASTA is as:

```
>Description of the sequence.....  
Sequence_____]  
_____] ]  
_____] ]  
_____] ]  
_____]
```

- The Description always starts from the ‘>’ sign and usually consists of the **Accession Number** and the **name of the species** of which the sequence is.
- The Sequence is based on the single-letter code denoting either nucleotides or amino acids that have been standardized by IUB/IUPAC.
- Each row consists of 70 to 80 letters, each letter represents the corresponding nucleotide or amino acid.

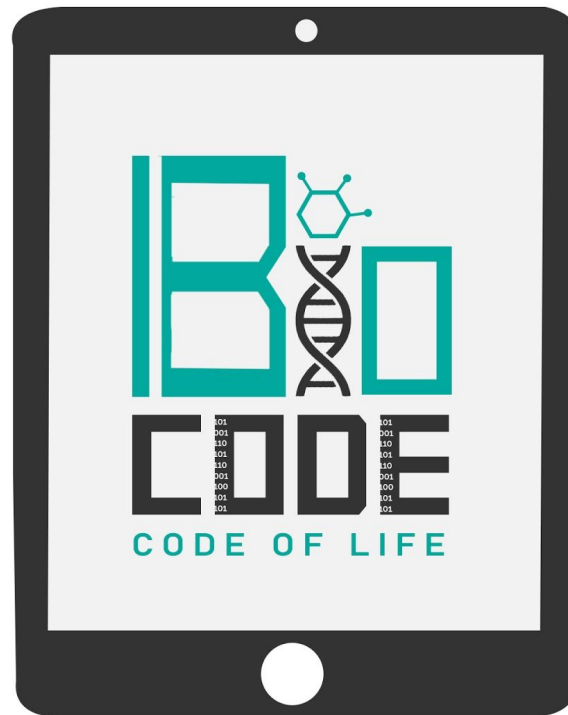
Extension

Like all other formats FASTA also has its own filename extensions in which it is stored, each extension denotes specific type of sequence which are given as:

Extension	Acronym
fasta	generic fasta
fna	fasta nucleic acid
ffn	FASTA nucleotide of gene regions
faa	fasta amino acid
frn	FASTA non-coding RNA

Summary

In this tutorial, we've learnt about the FASTA format that it is the most



LEARN PYTHON & R FOR BIOINFORMATICS

simple and widely used text-based format and what is the syntax of the FASTA and extensions of FASTA. We have used the sequence of mRNA of *Homo sapiens* Lactase (LCT), you can also retrieve the sequence of your requirement.