



LEARN PYTHON & R FOR BIOINFORMATICS

Prerequisite Terminologies:

In order to have a better understanding of the main topic, you should have the basic concept of the following terms:

- **Introduction to BioPython.**
- **SeqRecord Creating SeqRecords.**
- **FASTA format.**
- **GenBank format.**

Introduction:

The SeqRecord object in BioPython is a module of functions that provide us the ability to create sequence records that resonate with the sequence records of GenBank, EMBL (EBI), and other databases. The SeqRecord object in BioPython is utilized to hold a sequence as a seq object with identifiers (ID and name), description and optionally annotation and sub features related to the input sequence.

Steps:

- Import the following modules:
from Bio.Seq import Seq
from Bio.SeqRecord import SeqRecord
- Create a variable to store the biological sequence (nucleotide or amino acid) as a *Seq* object by calling in the *Seq()* function, as:
random_Seq = Seq("ANY NUCLEOTIDE OR AMINO ACID SEQ")
- Now pass this variable to *SeqRecord()* function to make the *Seq* object a *SeqRecord*, as:
Seq_rec = SeqRecord(random_Seq)

[By using the *print()* and running the above code, you'll notice that it'll only display the sequence on the output screen while the rest of the attributes (i.e., Description, ID, etc.) will not be displayed, since you've not entered those attributes in your code.]

- ★ A sequence record (*SeqRecord*) must have following attributes and their values:
 - ***.eq***
 - ***.id***
 - ***.description***
 - ***.name***
 - ***.letter***
 - ***.annotations***
 - ***.features***
 - ***.dbxrefs***
- ★ To retrieve the values of any of the above attributes, just call in the respective attribute with the variable name and to assign the value to any of the above attributes use the equal operator (=).
- To assign an ID to a particular sequence record object, call in the ***.id*** object with the variable name, as:
seq_rec.id = "ABC123.4"
- To assign a description to a particular sequence record object, call in the ***.description*** object with the variable name, as:
seq_rec.description = "ENTER THE DESCRIPTION OF YOUR SEQ"

- To assign a name to a particular sequence record object, call in the *.name* object with the variable name, as:

seq_rec.name = "ENTER THE NAME OF THE SPECIE THE SEQUENCE BELONGS TO".

- Now to print your SeqRecord on your output screen, call in the *print()* function and use the variable name as parameter, as:

print (seq_rec)

[Now it'll provide you the description, name, ID, sequence, and other attributes you've added in your code on the output screen.]

- To assign a letter annotation to a particular sequence record object, call in the *.letter_annotation ["NAME OF LETTER ANNOTATION"]* object with the variable name, as:

seq_rec.letter_annotation ["NAME OF LETTER ANNOTATION"] = [30, 40, 50, 48]

- To print the letter annotation with its name, call in the *print()* function, as:

print (seq_rec.letter_annotation ["NAME OF LETTER ANNOTATION"])

Summary:

In this video tutorial of BioPython, we learned to create sequence records that resonate the sequence records of GenBank, FASTA, EMBL(EBI), etc, using the SeqRecord module of Bio.SeqRecord class of the BioPython. We also got to know how to retrieve and assign values to the attributes of the particular sequence record object.