



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**
- **Entrez Databases in NCBI**

Introduction:

Bio.Entrez is the submodule within the BioPython package that provides code to access NCBI over the World Wide Web to retrieve various sorts of information. This module provides a number of functions which will return the data as a handle object. This is the standard interface used in Python for reading data from a file and provides methods or offers iteration over the contents line by line.

Steps:

- Import the *Entrez* module from the BioPython package in order to use its functions.

from Bio import Entrez

- Declare a variable (e.g: Entrez.Email) and provide your Email address as:

Entrez.email = "Email address"

Note: You should always provide the actual Email address. Providing fake Email is unethical.

- To see what sort of databases can be accessed through the Entrez module within BioPython, use the '*einfo()* function'.
- To do so, declare a variable (e.g: handle) in which the results will be stored and then pass in the *Entrez.einfo()* function.

handle = Entrez.einfo()

Note: *einfo* function provides field index term counts, last update, and available links for each database.

- Now to read the results that will be received from NCBI database, declare a variable (e.g: record) in which the results will be read, provide the Entrez.read function and pass in the handle variable.

record = Entrez.read(handle)

[Use the print function to print out the results.]

print(record)

[Run the entire code, so you'll see a list of databases that can be accessed through the Entrez module within BioPython.]

- If you want to retrieve the information for a single database, pass in the '*db*' parameter to the *einfo* function in the above code, as:

handle = Entrez.einfo(db = "database name")

[Now run the entire code, so you'll get all the information related to only a single database.]

- You can also do a particular search in a particular database of NCBI through the Entrez module within BioPython, as:

```
handle = Entrez.esearch(db="pubmed", term="cancer genomics")  
record = Entrez.read(handle)  
print(record)  
print(record["IdList"])
```

- Here, we're using the 'esearch' function to search the cancer genomics information in 'PubMed' database of NCBI, reading the results using Entrez.read function and also printing out the IDs that are available for cancer genomics papers in PubMed database.

[Run the above code, so you'll see all the IDs available for cancer genomics papers.]

[This is just the introduction for the Entrez module, we'll discuss the details in the next sections of this video.]

Summary:

In this video tutorial of BioPython, we have learned about the Entrez module within BioPython. We also got to know how to retrieve the information about all the databases of NCBI as well as information for a single database and how to perform a particular search within a single database utilizing the Entrez module within BioPython.