
Diseñador de sondas de hibridación

Simón Vergara

24 de mayo de 2023

Contents:

1. Probe-Designer	1
1.1. descarga module	1
1.2. test module	2
2. Índices y tablas	3
Índice de Módulos Python	5
Índice	7

1.1 descarga module

Para obtener el ADN que se utilizará como referencia para el diseño de las sondas se va a descargar directamente el archivo genético desde la base de datos ‘Nucleotide’ de NCBI (National Center for Biotechnology Information) del NIH (National Institute of Health, EEUU) (<https://www.ncbi.nlm.nih.gov/nuccore>). Esta base de datos contiene millones de registros de secuencias de nucleótidos con mucha información adicional, como la especie, cromosoma, nombre del gen, exones, la fuente de los datos, etc.

Para llevar a cabo la descarga de los archivos genéticos, se implementó un módulo que permite realizar descargas desde esta base de datos a través de la librería Entrez de BioPython y almacenar los archivos en una carpeta. De esta manera se puede automatizar la obtención de las secuencias de referencia.

Por otra parte se implementó la función que retorna una secuencia a partir de un archivo almacenado localmente, además de crear una copia del archivo y almacenarla. Esta función sólo podrá recibir archivos de secuencia anotados como GenBank y GFF3.

El producto de estas funciones implementadas es la secuencia en formato SeqRecord, una estructura de datos propia de la librería BioPython que permite almacenar grandes secuencias con su respectiva notación, como por ejemplo las posiciones de cada uno de los exones, las transcripciones, secciones no codificantes, entre otros. Además contiene información extra de la secuencia tal como la especie, el cromosoma, la fuente de donde salieron los datos, etc.

`descarga.accnum_to_seqrecord(accession_number)`

Función que recibe un accession number de la base de datos “nucleotide” de NCBI <https://www.ncbi.nlm.nih.gov/nuccore>. Genera el archivo .gbk (GenBank) en la carpeta “files” y retorna la secuencia en formato SeqRecord.

Parámetros

accession_number – El accession number del archivo en GenBank, por ejemplo NG_008617.1

Devuelve

La secuencia en formato SeqRecord con todas sus anotaciones.

`descarga.main()`

Ejecutar directamente el archivo “descarga” requiere entregar el parámetro `–accessionnumber` para descargar una secuencia mediante la función `accnum_to_seqrecord`.

Parámetros

--accessionnumber – El accession number del archivo en GenBank, por ejemplo NG_008617.1

`descarga.parse_file_to_seqrecord(filepath)`

Función que recibe la ruta de un archivo genético anotado y lo retorna como SeqRecord. Genera una copia del archivo en la carpeta files. Solo se aceptan archivos en formato «GenBank» y «GFF3», o sea “.gbk”, “.gb” y “.gff3”. Estos archivos pueden estar comprimidos en formato “.gz”.

Parámetros

filepath – La ruta donde se encuentra el archivo que se quiere utilizar

Devuelve

La secuencia en formato SeqRecord con todas sus anotaciones.

1.2 test module

Este módulo prueba las funciones implementadas en el módulo descarga, usando los métodos para obtener secuencias: mediante un archivo y mediante un accession number. Accede a las ubicaciones de las anotaciones de las secuencias.

CAPÍTULO 2

Índices y tablas

- genindex
- modindex
- search

d

descarga, [1](#)

t

test, [2](#)

A

`accnum_to_seqrecord()` (*en el módulo descarga*), 1

D

`descarga`
 módulo, 1

M

`main()` (*en el módulo descarga*), 1

módulo
 descarga, 1
 test, 2

P

`parse_file_to_seqrecord()` (*en el módulo descarga*),
 2

T

`test`
 módulo, 2