



En este desafío ud trabajará con secuencias de ADN. Puede consultar más información acerca del ADN, sobre cómo está compuesto, y algunas otras consideraciones en el siguiente [link](#).

Es posible representar información acerca de los nucleótidos del ADN representadas como texto: “CTCCCAAACCCAATGTGAACGTTA”. Donde C: Citosina; T. Timina; A: Adenina; G: Guanina.

Ud posee una colección de archivos de ADN en formato FASTA. Una secuencia bajo formato FASTA comienza con una descripción en una única línea (línea de cabecera), seguida por líneas de datos de secuencia. La línea de descripción se distingue de los datos de secuencia por un símbolo '>' (mayor que) en la primera columna. La palabra siguiente a este símbolo es el identificador de la secuencia, y el resto de la línea es la descripción (ambos son opcionales). No debería existir espacio entre el '>' y la primera letra del identificador. Se recomienda que todas las líneas de texto sean menores de 80 caracteres. La secuencia termina si aparece otra línea comenzando con el símbolo '>'; esto indica el comienzo de otra secuencia. Un ejemplo simple de una secuencia en el formato FASTA puede ser:

```
>cytochrome_b
GCGTCGAGCCCCNACCAAGGCGCCCGGAGGACCAACAAAACCTTTTTGTATAC
CCCCTCGCGGGTTTTTATAATCTGAGCCTTCTCGGCGCCTCTCGTAGGCGTTTCGAAAA
TGAATCAAACTTTCAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAA
ATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATT
GCGCCCGCCAGTATTCTGGCGGCATGCCTGTCCGAGCGTCATTTCAACCCTCGAACCCC
```

Un ejemplo de archivo FASTA con secuencias genéticas perteneciente a un ser humano se encuentra en <https://goo.gl/n1S8dN>.

Un investigador de BioIngeniería de la ESPOL, posee una base datos de Genes en Formato FASTA. Los archivos están ubicados en la siguiente dirección: <https://goo.gl/PyeuA> . Este investigador necesita automatizar algunas tareas, las cuales se describen a continuación:

Actividad 1

Se desea obtener un nuevo archivo con la secuencia original de la cadena genética, la cadena reversa, la complementaria y la reversa complementaria. Para conocer cómo realizar esta tarea, puede consultar información en internet. En los siguientes links ([link1](#),

[link2](#)) puede encontrar alguna información útil. Ud debe realizar esto para todos los archivos de genes disponibles en el repositorio de Genes.

Ejemplo del nuevo archivo *Sample1_full.fasta*:

```
>Sample1
GGGGAAAAAAAATTTATATAT

>Sample1_complement
CCCCTTTTTTTTAAATATATA

>Sample1_reverse
TATATATTTAAAAAAAAGGGG

>Sample1_reverse_complement
ATATATAAATTTTTTTTCCCC
```

Fig. 1 Ejemplo de un archivo fasta con cadenas reversa, complementaria y reversa complementaria

Puede encontrar un ejemplo de un archivo completo en <https://goo.gl/DAVUYa>.

Actividad 2

En la actualidad existen herramientas online que pueden identificar a qué especie de ser vivo pertenece una determinada cadena de un gen. El Centro Nacional para la Información Biotecnológica (NCBI) provee una interfaz para realizar esta tarea utilizando genes en formato FASTA a través de su herramienta BLAST.

BLAST (Basic Local Alignment Search Tool), posee una interfaz web que permite realizar identificación de especies. Las siguientes imágenes describen el proceso:

1. Copiar el gen de la cadena original del gen en formato FASTA. Ejm. Fig 2

```

File Edit Format View Help
>Contig1
TTCCGTAGGTGAACCTGCGGAGGGATCATTACCGAGTTTACAACCTCCCAAACCCAATGTG
AACGTTACCAAACCTGTTGCCTCGGCGGGATCTCTGCCCCGGGTGCGTCGCAGCCCCGGAC
CAAGGCGCCCCGCGGAGGACCAACCAAACCTCTTTTTGTATACCCCTCGCGGGTTTTTT
ATAATCTGAGCCTTCTCGGCGCCTCTCGTAGGCGTTTCGAAAATGAATCAAAACTTTCAA
CAACGGATCTCTTGTTCTGGCATCGATGAAGAACGCAGCGAAAATGCGATAAGTAATGTG
AATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCGCCAGTATTCT
GGCGGGCATGCCTGTCCGAGCGTCATTTCAACCCTCGAACCCCTCCGGGGGGTCCGGCGTT
GGGGATCGGCCCTCCCTTAGCGGGTGGCCGTCTCCGAAATACAGTGGCGGTCTCGCCGCA
GCCTCTCTGCGCAGTAGTTTGCACTCGCATCGGGAGCGCGGCGCGTCCACAGCCGTT
AAACACCCAACCTTCTGAAATGTTGACCTCGGATCAGGTAGGAATACCCGCTGAACTTAAG
CATATCAATAA

```

Fig. Ejemplo de un archivo fasta con un Gen desconocido

2. Ir al sitio: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. Seleccionar el link Nucleotide Blast
4. Pegar la cadena FASTA en el área de texto correspondiente. Ejem. Fig.3
5. Seleccionar el botón BLAST al final de la página web

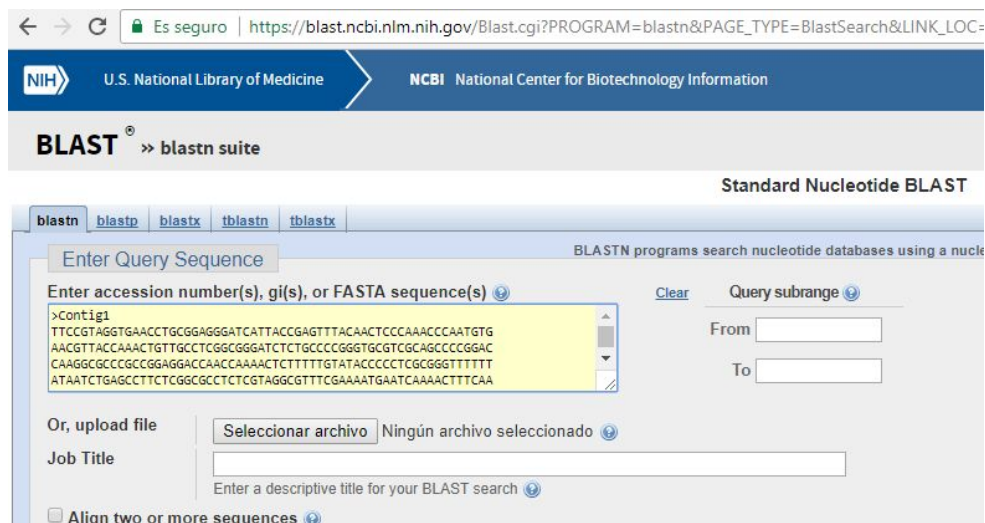


Fig. 3 Interfaz de la herramienta BLAST de NCBI

6. La página se redirecciona, después de algunos segundos, se muestra una pantalla como la de la figura 4.

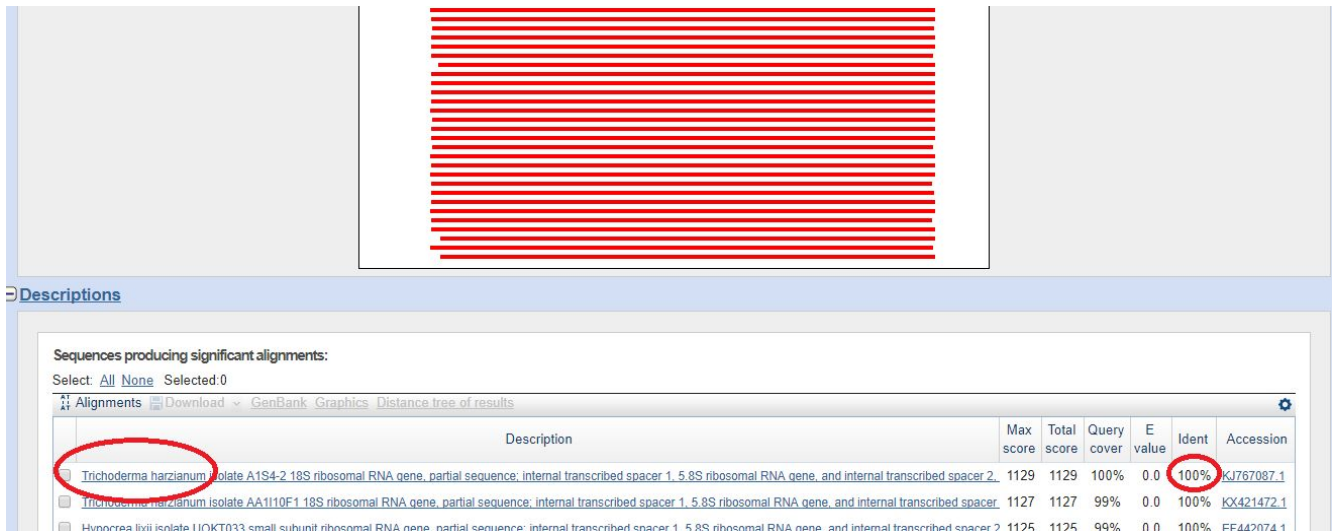


Fig. 4 Resultado de la consulta realizada con BLAST

- En el apartado de Descripción, existe una lista de las posibles especies que contengan el gen consultado en su código genético. En el ejemplo anterior se observa que la especie corresponde a *Trichoderma Harzianum* con una posible identidad del 100%.
- Si necesita observar un video detallado de todo el proceso descrito, lo puede observar en la siguiente dirección: <https://goo.gl/qWSoVJ>.

El investigador desea conocer la lista de especies a la que pertenecen los genes de los que se tiene registros en archivos ubicados en <https://goo.gl/PyeuaA>. Su tarea consiste en automatizar todo el proceso de búsqueda de identidad de especies a través de BLAST. Se conoce de la existencia de una librería llamada BioPython, la cual hace posible realizar la consulta BLAST a través de este lenguaje de programación. Para esta tarea, ud debe identificar las especies utilizando BLAST a través de BioPython.

Actividad 3

Se desea contar con un diagrama de barras o pastel de la cantidad de Nucleótidos (Timina, Adenina, Guanina, Citosina) de las cadenas pertenecientes a cada especie. Colocar 3 ejemplos de estos gráficos en su presentación.



Entregables

Subir en la dirección <https://goo.gl/4j4Tjn> un archivo rar con el formato **NombreGrupo_Paralelo.rar** lo siguiente:

1. Código utilizado y archivos generados.
2. Una presentación en Power Point de sus hallazgos para ser presentada en 5 min.

En la presentación

- Explicar el algoritmo que utilizaron para utilizar Python y Blast.
- Mostrar los gráficos de la actividad 3.