

## 4 Práctica 4. Reconstrucción filogenética.

### 4.1 Características generales.

#### 4.1.1 Entregables:

- Ésta práctica se lleva a cabo en parejas.
- Un archivo en formato pdf con las respuestas del cuestionario, todas las preguntas deben contener sus respectivas referencias.

### 4.2 Instrucciones.

El objetivo de la construcción de un árbol filogenético es determinar la posible relación evolutiva entre organismos de un grupo determinado. Para ésta práctica vamos a suponer que tenemos secuencias nuevas del gen *rcbL* de la planta de vainilla (*Vanilla sp.*):

```
>Seq1-vanillasp
```

```
ACCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCGGGAGTTCCGCCTGAAGAAGCAGGGGCTGCGGT
```

```
>Seq2-vanillasp
```

```
AGCTGCCGAATCTTCTACTGGTACATGGACAACCTGTGTGGACTGATGGACTGACCAGTCTTGATCGTTACAAAGGGCGAT
```

#### 4.2.1 Obtención de secuencias

- Abre [BLAST](#) y coloca las secuencias en el recuadro de búsqueda. ([clic aquí](#) para ver figura de referencia). En la sección de **Choose Search Set** en **Database**, asegúrate de tener seleccionada la parte de **Standard databases (nr etc.)**: y en el menú desplegable **Nucleotide collection (nr/nt)**. Y en la sección **Program Selection** debes tener seleccionado **Highly similar sequences (megablast)**. Finalmente da click en **BLAST**.

- Una vez que hayas dado clic en **BLAST**, te aparecerá una lista de organismos, cuyas secuencias son parecidas a las que ingresamos como problema. Como podrás notar, las secuencias aparecen en forma de tabla, en donde tenemos algunas columnas importantes como **Scientific Name** que es el nombre científico de los organismos a los que las secuencias se encuentran relacionados, **S value** que es el puntaje de similitud, **E values** que son las probabilidades aproximadas de que las secuencias similares identificadas no estén relacionadas por descendencia evolutiva (es decir, que sean similares solo por casualidad). La selección de secuencias homólogas a menudo se basa en un umbral de valor S y **Accession** que es el número de acceso al NCBI para conocer más del organismo o del proyecto.
- Selecciona 10 secuencias de las que te aparecen en el listado y da clic en **Download** y selecciona **FASTA (complete sequence)** ([clic aquí](#) para ver imagen de ejemplo). Guarda tu archivo como `vanilla-seqs.fasta`, aunque te aparezca como formato `.txt`. Si quieres modifica el encabezado de tus secuencias.
- Descarga el programa [MEGA](#), es el programa que vamos a utilizar para llevar a cabo el alineamiento y la construcción del árbol. [MEGA](#) tiene incluido dos programas para hacer alineamientos `ClustalW` y `MUSCLE`. Al momento de utilizarlos debemos editar algunos parámetros como `GOP` (*gap open penalty*) y `GEP` (*gap extension penalty*). Estos programas tratan de maximizar el alineamiento maximizando el puntaje de “match”. Para obtener la alineación, se introducen espacios (el carácter `-`) correspondientes a eventos de inserción/eliminación ancestrales hipotéticos en las secuencias en los sitios que maximizan la puntuación de alineación. Siempre que se agregue un espacio de forma aislada, la puntuación de alineación se penaliza con la cantidad especificada por el parámetro `GOP`, y cuando se agregan espacios posteriores adyacentes a un espacio existente, la puntuación de alineación se penaliza con la cantidad especificada por el parámetro `GEP`.

## 4.2.2 Alineamiento

- Abre el programa [MEGA](#), da clic en `ALIGN >> Edit/Build Alignment >> Retrieve sequence from a file >> Ok` y seleccionar el archivo que generamos en el paso anterior. Esto te llevará a una nueva ventana con tus secuencias no alineadas, selecciona todas tus secuencias dando `CTRL + A` y da clic en el botón que tiene forma de brazo (corresponde al algoritmo de `MUSCLE`) `>> Align DNA`. Deja los parámetros por default y exporta el alineamiento dando clic en `Data >> Export Alignment >> FASTA format`.

## 4.2.3 Construcción del árbol.

- Da click en `PHYLOGENY >> Construct/test Maximum Likelihood Tree` y selecciona el archivo con tus secuencias alineadas. Te abra un recuadro, selecciona `Nucleotide Sequences`. Te saldrá otro recuadro con la pregunta “Protein-coding nucleotide sequence data?” selecciona “Si/Yes” >> `Standard` >> “Would you like to use currently active data? Yes” y finalmente te abrirá un recuadro con los métodos que puedes seleccionar. Por ejemplo el modelo evolutivo, por default tendrás seleccionado `Jukes-Cantor` y las opciones para inferir el árbol `TREE INFEERANCE OPTIONS` por default te saldra `Nearest Neighbor Interchange (NNI)`. El resto de los parámetros déjalos tal cual y da `ok`. Listo tienes un árbol, guarda la imagen dando clic en `Image` y colocando el siguiente nombre: `arbol_jukes_cantor.png`.
- Vuelve a repetir todo el proceso anterior pero en vez de elegir el modelo evolutivo de Jukes-Cantor elige el modelo de Kimura y guarda la imagen de éste segundo árbol como `arbol_kimura.png`.

## 4.3 Cuestionario

1. Describe en un párrafo las características del género *Vanilla sp.*. Usa términos que tu comprendas.
2. ¿Para qué sirve el gen *rcbL*?
3. ¿Qué significa BLAST y cómo funciona?
4. ¿Por qué es importante alinear antes de llevar a cabo la construcción de un árbol filogenético?
5. ¿Cómo funciona `MUSCLE`? ¿Qué significa que no haya espacios (gaps - ) en el alineamiento?
6. Adjunta las imágenes de los árboles que obtuviste ¿Son iguales?, ¿Qué implicaciones tiene escoger uno u otro modelo evolutivo? De manera general, ¿Qué podrías concluir de cada árbol?