

# Trabajo 1 Laboratorio de Bioinformática - Comparación de secuencias

*Manuel Villalobos Cid*

*03 de Octubre de 2016*

## 1 Instrucciones

Las actividades **prácticas** deberán ser efectuadas por cada uno de los alumnos. En grupos de máximo tres estudiantes, desarrollarán un informe utilizando el sistema de composición de textos denominado LaTeX según un **formato** especificado.

El plazo máximo de entrega es el día **sábado 15 de octubre** a las **23:55 horas**. El informe deberá ser cargado en formato **pdf** en la sección para Laboratorio de Bioinformática en Moodle habilitada para esta tarea.

## 2 Actividad introductoria

### 2.1 Alineamiento de secuencias

- Describa los tipos de **secuencias** que se pueden encontrar en Bioinformática.
- Defina los conceptos de **alineamiento global** y **alineamiento local**, detallando: diferencias entre ambos tipos, aplicaciones, algoritmos principales y herramientas disponibles.
- Defina **homología de secuencias**: **ortología** y **paralogía**. Señale ejemplos.
- ¿Qué son los alineamientos múltiples? Describa los tipos de enfoques que se aplican para resolver el problema.

### 2.2 Basic Local Alignment Search Tool (BLAST)

- Describa las principales características y aplicaciones de BLAST.
- Diríjase a <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- Caracterice cada uno de los cinco programas disponibles en BasicBLAST: **nucleotide blast**, **protein blast**, **blastx**, **tblastn**, **tblastx**.
- Describa las aplicaciones disponibles en **specialized BLAST**.
- Ingrese a **nucleotide blast**. Familiarícese con los parámetros en pantalla.
- Describa los parámetros encontrados en la sección **algorithm parameters**.

## 3 Actividad práctica

### 3.1 Basic Local Alignment Search Tool (BLAST)

- Descargue los siguientes **archivos**. Caracterice las secuencias utilizando R.
- En la herramienta **nucleotide blast** compare el archivo **secuencia.fasta** consigo mismo. Analice resultados.
- Compare el archivo **secuencia.fasta** con **rattusnor...fasta**. Analice resultados.

- Cargue el archivo **secuencia.fasta**. Seleccione en el parámetro “organismo” la opción **Rattus norvegicus (taxid:10116)**. Compare secuencias y explique los resultados. ¿Existe alguna relación en estos resultados con la etapa anterior? Comente.

### 3.2 Alineamiento local en R

- Defina estas dos secuencias y compárelas entre sí utilizando BLAST. Pruebe con distintos algoritmos (megablast, discontinuous megablast y blastn). Analice resultados.

```
seq1<-"GAATTCCTACTACGAAGAATTCCTACTACGAAACTACGAAAATTCCTACTACGA"
seq2<-"GAATTCCTACTACGGAATTCCTCCCTCCATAATTCCTACTACGA"
```

- Utilizando la librería **Biostring** de Bioconductor aplique el algoritmo de **Smith-Waterman** para comparar las secuencias anteriores. Asegúrese de emplear los mismos parámetros de algoritmo que BLAST.
- Construya un **script** o **función** en R que implemente el algoritmo **Smith-Waterman**. Aplíquelo sobre las dos secuencias anteriores y sobre otras secuencias de mayor tamaño. Como referencia puede utilizar los archivos que se encuentran en la carpeta **Pruebas**.
- Compare los resultados obtenidos por **BLAST**, **Smith-Waterman** de la librería Biostring y su **implementación propia**.

Cada uno de los resultados de las actividades debe ser incluido en el informe de laboratorio.