

# Tarea 1 de Bioinformática

La tarea que se os pide para evaluar la primera parte de la asignatura consiste en crear un documento .ipynb que ejecute un análisis y recoja los resultados y las conclusiones pertinentes. Podría ser también un *script* de R, si lo preferís. Podéis tomar como punto de partida el documento Ejemplo.ipynb, siempre que lo apliquéis a unos datos de partida diferentes (ver más abajo). O bien, podéis realizar cualquier otro tipo de análisis utilizando las herramientas que hemos visto en clase. A continuación os doy sugerencias. Se valorará el esfuerzo de realizar un análisis bioinformático diferente al del ejemplo, pero no es necesario.

## Objetivos

En la evaluación del trabajo valoraré los aspectos siguientes, en orden de importancia (los tres primeros son esenciales):

1. Los análisis són **reproducibles** (es decir, puedo obtenerlos yo también ejecutando vuestro código).
2. Los análisis demuestran la utilidad de alguna herramienta o recurso estudiado en la asignatura.
3. De los resultados se extrae alguna conclusión correcta acerca del objeto de estudio.
4. La bibliografía incluye referencias a los algoritmos y recursos utilizados.
5. Tanto la redacción como el código son claros y coherentes.
6. La longitud del trabajo no debe ser excesiva.
7. El análisis incluye algún cálculo o proceso adicional no incluido en el ejemplo.
8. La discusión propone qué otros análisis podrían ayudar a resolver las preguntas pendientes.

## El análisis de ejemplo

En `Ejemplo.ipynb` se realizan algunas búsquedas de blastp en la base de datos swissprot, disponible en [el ambiente de trabajo enlazado](#). La proteína de partida es la secuencia de un enzima de *Arabidopsis thaliana* que participa en la síntesis de clorofila. La introducción y las conclusiones dan algo de contexto y justifican el análisis. Los resultados de las búsquedas se analizan superficialmente para extraer algunas conclusiones.

La tarea puede consistir simplemente en repetir el análisis para otra proteína de partida, ajustando los parámetros si hace falta y adaptando la introducción, las conclusiones y las referencias. Algunas sugerencias:

- Proteínas que participan en neuroreceptores de acetilcolina, como [ACM1](#), [ACHA3](#) o [ACHA7](#), cuyos homólogos aparecieron por primera vez en algún ancestro común a cnidarios y cordados, hace ya bastantes años, según Viscardi *et al.* (2021).

- Las heliorodopsinas, como por ejemplo [esta, de \*Metanobacterium formicicum\*](#): un tercer tipo de rodopsinas descubierto en 2018 (Pushkarev et al. 2018) cuya función y distribución taxonómica sigue despertando interés.

## Otros análisis posibles

- Preparar un alineamiento de una familia de proteínas.
- Generar un modelo oculto de Markov de un dominio proteico.
- Atribuir especies a un conjunto de secuencias de rDNA.
- Comprobar si mediante PSI-BLAST se detecta homología entre las rodopsinas de tipo 1 y las heliorodopsinas.
- Descargar algunos datos interesantes de la European Nucleotide Archive, con alguna finalidad concreta.
- etc.

## Formato de entrega

El trabajo debe ser entregado a través del Aula Virtual, en principio como cuaderno jupyter (archivo `.ipynb`). Si se utilizan archivos adicionales (por ejemplo, fasta) no incluidos en el ambiente de computación disponible en MyBinder, entonces tanto el cuaderno *jupyter* con el código y el texto como dichos archivos adicionales deben ser incluidos en un archivo `.tar` o `.zip`. En cualquier caso, no se aceptarán archivos en formatos cerrados o privativos.

Si para realizar tu trabajo hubieras necesitado instalar algún programa adicional, será necesario que documentes qué programa es y qué versión has instalado.

## Recursos

Todos los ambientes de MyBinder de las prácticas así como [este](#), en el que tenéis además la base de datos swissprot, están a vuestra disposición. Si necesitáis un ambiente diferente, con otros paquetes preinstalados o lo que sea, avisadme. Y si preferís ejecutar el análisis en vuestro propio ordenador y necesitáis ayuda para instalar R, jupyter o cualquier otro programa, me podéis consultar.

## Uso del Jupyter Notebook dentro de Jupyter Lab

La interfaz de Jupyter Lab es muy intuitiva y la hemos estado practicando. Aún así, es muy posible que se nos olvide cómo realizar ciertas acciones, como por ejemplo: eliminar o añadir una celda de texto o de código, borrar todos los resultados de la ejecución anterior de una o más celdas de código, etc. En [este enlace](#) encontrarás toda la documentación sobre Jupyter Lab. Para cuestiones más relacionadas con el cuaderno (el documento `.ipynb`), puedes darle un vistazo a [esta documentación](#).

Recordad a los bloques de texto reconocen la sintaxis [markdown](#), se les da formato mediante el botón "Run" y pueden ser editados de nuevo con un doble clic. Por favor, reservad los bloques de código para el código.

## Recursos sobre R

Hay mucha información online. Yo recomiendo estos dos enlaces:

- [Curso básico de R de Software Carpentry](#)
- [Guía de R para piratas](#)

## Sobre BLAST

He incluido en [el espacio de trabajo](#) los archivos `blastp_help.txt` y `psiblast_help.txt`, con toda la ayuda de los programas `blastp` y `psiblast`. Ahí encontrarás fácilmente, por ejemplo, los campos de información disponibles para personalizar los resultados en formato de tabla. El *script* `blast2fasta_exemple.R` contiene instrucciones para guardar los resultados de una ejecución de BLAST en un archivo fasta.

## Gestión de archivos remotos

No olvides que todo tu trabajo habrá desaparecido cuando se cierre la sesión en MyBinder, intencionada o accidentalmente, si no lo has descargado en tu propio ordenador. Recuerda que el entorno MyBinder se ejecuta remotamente. Descargar con frecuencia tu cuaderno jupyter de trabajo te protegerá de las posibles interrupciones en la conexión al servidor.

Una alternativa es que instales un Jupyter Lab en tu propio ordenador. [Aquí](#) tienes bastante información al respecto. Avísame si crees que puedo ayudarte.

## Referencias

- Lucas Henriques Viscardi, Danilo Oliveira Imparato, Maria Cátira Bortolini, Rodrigo Juliani Siqueira Dalmolin, Ionotropic Receptors as a Driving Force behind Human Synapse Establishment, Molecular Biology and Evolution, Volume 38, Issue 3, March 2021, Pages 735–744, [doi:10.1093/molbev/msaa252](https://doi.org/10.1093/molbev/msaa252).
- Pushkarev, A., Inoue, K., Larom, S. et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nature 558, 595–599 (2018). <https://doi.org/10.1038/s41586-018-0225-9>

In [ ]: