

# Letra Obligatorio

Daniela Costa, Guillermo Lamolle

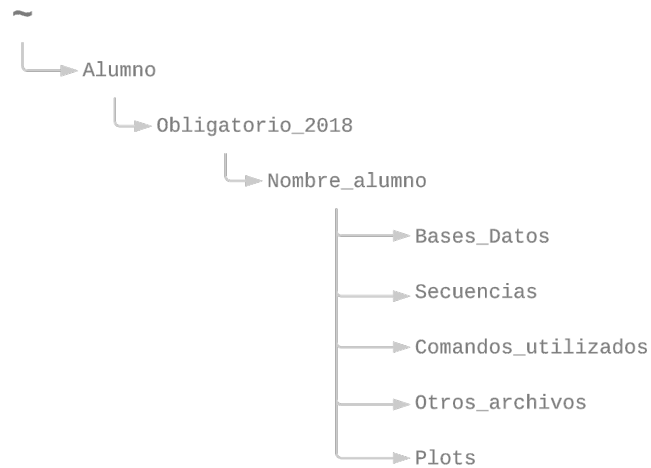
Noviembre 2018

Se le asignarán los nombres de dos bacterias en las que se basará su trabajo.

1. Describir las bacterias, especificando:

- Ubicación taxonómica
- Estilo de vida
- Otras particularidades, por ejemplo:
  - Si es patógena
  - En qué ambientes vive, etc.

2. Dada la siguiente estructura de directorios:



(a) Utilice la línea de comandos para recrear la estructura de directorios mostrada en la figura.

- (b) Cambiar el nombre del directorio `Nombre_alumno` por su nombre, utilizando el formato `Apellido_N`, donde N es la inicial de su nombre. Por ejemplo, si el nombre es José Rodríguez, el directorio se llamará `Rodriguez_J`
3. Obtener los archivos correspondientes a las secuencias nucleotídicas del genoma de ambas bacterias y las secuencias de los genes codificantes.
    - (a) Mediante los métodos adecuados, obtener:
      - i. Cantidad de secuencias en cada genoma
      - ii. Largo de cada genoma
      - iii. Contenido GC promedio de cada genoma
    - (b) Elija una de las bacterias y:
      - i. Obtenga la composición nucleotídica de cada gen.
      - ii. Obtenga la composición aminoacídica de cada gen.
      - iii. Guardar en un objeto tipo `data.frame` el contenido GC medio, y en cada posición del codón de las secuencias codificantes.
      - iv. Obtenga **UN** gráfico donde se relacionen los contenidos GC de las tres posiciones (eje y) con el contenido GC medio (eje x) de los genes. Utilice un color diferente para cada posición del codón.
  4. ¿Cómo podría obtener la variación en el contenido GC a lo largo del genoma? Describa brevemente el método.
  5. ¿En qué se basa el método que utiliza `oriloc()` para determinar los sitios de origen y término de la replicación?
  6. Explique en qué consiste el índice *RSCU* para el uso de codones
  7. Determinar si las bacterias asignadas contienen el gen que codifica la bacterioferritina (*bfr*). Para esto, utilice la base de datos de secuencias aminoacídicas asignada, considerando los siguientes criterios:
    - Porcentaje de identidad superior al 50%
    - La longitud del alineamiento debe ser mayor al 50% de la longitud del *subject*.

### IMPORTANTE

Coloque cada archivo en la carpeta adecuada de la estructura de directorios mostrada en la figura:

- Guarde los gráficos obtenidos en formato **pdf** en la carpeta **Plots**
- Coloque los archivos correspondientes a la base de datos del ejercicio 3 en el directorio **Bases\_Datos**
- Coloque los archivos que contengan las respuestas a las preguntas y códigos utilizados en la carpeta **Comandos\_utilizados**

Deberá entregar a Bedelía:

- Una copia impresa de los plots, códigos utilizados y respuestas.
- La estructura de directorios completa (desde **Nombre\_alumno**) en un CD con los archivos en sus respectivas carpetas. Rotule el CD con nombre, número de estudiante y el texto "Obligatorio Bioinformática 1".

Incluya solamente los códigos relevantes y cuide la presentación del trabajo. La claridad de la presentación del trabajo influirá en la nota final. Como siempre, si hay una pregunta que requiere comandos, en la respuesta deben ponerse LAS DOS COSAS (los comandos, y la respuesta propiamente dicha).

**Ejemplo:**

Pregunta: ¿cuánto es el promedio de 2, 3 y 4?

Respuesta: `mean(c(2,3,4))`

El promedio es 3.

Si se ponen solo los comandos o solo la respuesta, aunque estén correctos, la nota será menor, en algún caso mucho menor.