

## 3 Práctica 3.

### 3.1 Parte I. Formatos y bases de datos. #

#### 3.1.1 Características generales

1. Práctica que se entrega en **parejas**.
2. Entregables:
  - Un archivo PDF con respuestas que se sube a classroom.

#### 3.1.2 Instrucciones

1. Elige dos artículos de cada color de la siguiente [lista](#), en total serán 6 artículos.
2. Por cada artículo seleccionado crea una [tabla como la siguiente](#) y llénala con la información que se solicita. Para el llenado de la tabla te sugiero que leas los comentarios que hay en la columnas y sigas el ejemplo (fila de color gris), el artículo que se utilizó para el ejemplo se encuentra [aquí](#).

**TIP:** La mayoría de la información que vas a buscar se encuentra en la sección de métodos del artículo.

Puedes hacer tu tabla en una hoja de Google Sheets y poner el enlace en tu reporte (cuidando que pueda tener acceso) o poner tal cual la tabla, como ud prefiera.

3. Elige un artículo de los seis anteriores o busca uno que sea de tu interés y contesta las siguientes preguntas:
  - a. De manera general, ¿De qué trata éste artículo? **TIP:** Esto puedes verlo en la sección del resumen o *Abstract*.
  - b. ¿Cuáles son los objetivos de la investigación?
  - c. ¿Qué tipo de estudio es? (Metagenoma, genómico o metabarcoding)¿Por qué?.

- d. Busca los archivos crudos que se utilizaron en la investigación en una base de datos, y coloca el link. Te recomiendo que descargues los archivos crudos en tu computadora.
- e. Investiga: ¿Cómo puedes saber el número de secuencias que hay en un archivo FASTQ ?

## 3.2 Parte 2. Control de calidad

### 3.2.1 Características generales

Entregables:

- Las preguntas de ambas secciones se entregan en un solo pdf.
- Una carpeta comprimida ( `.zip` **NO** `.rar` ) con los reportes de calidad en `html` .

### 3.2.2 Instrucciones:

4. Genera un reporte de control de calidad utilizando el programa `FASTQC` de:
  - Las secuencias de algún artículo de la lista de artículos del apartado anterior.
  - Las [siguientes secuencias](#).

Recuerda que se tiene un archivo `fastq` con secuencias *forward* y un archivo `fastq` con secuencias *reverse* por cada muestra y que se genera un reporte por cada archivo `fastq` .

5. ¿Qué gráficos/módulos del reporte arrojado por `FASTQC` consideras que son los más importantes? Elige 4 y justifica.
6. ¿Qué concluyes de las secuencias de cada muestra? Escribe una conclusión para las secuencias que tú elegiste y para las secuencias que te otorgué como problema.
7. ¿Por qué es importante conocer el contenido de GC en un genoma? Adjunta una referencia para ésta respuesta. `