



## 1. DATOS BÁSICOS DEL TFG:

**Título:** Poniendo a prueba las nuevas tecnologías para generar genomas perfectos. ¿Tan cerca como pensamos?

### Descripción general (resumen y metodología):

Actualmente, los enormes avances relacionados con técnicas de secuenciación masiva (NGS) han hecho posible una resolución nunca visto anteriormente. En el campo de la genómica, se ha hecho posible obtener el genoma de un organismo, con precios y condiciones muy accesibles. Las técnicas de secuenciación de lecturas largas (long reads) han sido reconocidas como “Método del Año” por la revista Nature en 2022, y el desarrollo tecnológico es constante, con múltiples empresas y nuevos protocolos prometiendo la obtención de “genomas perfectos”. Sin embargo, las long reads presentan problemas metodológicos, que en última instancia afectan a la precisión e integridad de las secuencias obtenidas. Dado que el análisis bioinformático de la mayoría de técnicas NGS/-ómicas requiere en primer lugar el uso de un genoma considerado “de referencia”, posibles errores de secuenciación pueden ser cruciales en estudios de genómica funcional. Por ello, en paralelo al desarrollo de nuevas técnicas de secuenciación y de “ensamblado” de las secuencias para obtener genomas, se han desarrollado herramientas bioinformáticas para corregir las secuencias y asegurar su exactitud. Tradicionalmente, se han basado en aproximaciones “híbridas

**Tipología:** Trabajos experimentales, de toma de datos de campo o de laboratorio.

### Objetivos planteados:

#### \* Objetivo:

En este proyecto, se utilizarán datos genómicos ya publicados y se compararán diferentes soluciones para ensamblar y corregir automáticamente las secuencias.

#### \* Plan de trabajo:

1. Búsqueda bibliográfica de herramientas y datasets disponibles
2. Diseño experimental para el análisis de los datos y benchmarking de las herramientas
3. Ejecución del flujo de trabajo y generación de resultados
4. Preparación de la memoria, exposición del TFG, y posible publicación científica

### Bibliografía básica:

- \*Method of the Year 2022: long-read sequencing. Nat Methods 20, 1 (2023).
- \*Ruiz, José L., et al. "From contigs to chromosomes: automatic Improvement of Long Read Assemblies (ILRA)." bioRxiv (2021): 2021-07.
- \*Zhang, X. et al. Benchmarking of long-read sequencing, assemblers and polishers for yeast genome. Brief Bioinform 23, 1–13 (2022).
- \*Southwood, Dean, et al. "Pyro: A Comprehensive Pipeline for Eukaryotic Genome Assembly." bioRxiv (2023): 2023-04.
- \*Southwood, Dean, et al. "Exhaustive benchmarking of de novo assembly methods for eukaryotic genomes." bioRxiv (2023): 2023-04.

### Recomendaciones y orientaciones para el estudiante:

El proyecto tendrá un fuerte componente formativo en cuanto a su carácter bioinformático. No se espera un gran conocimiento previo, pero sí un interés genuino por las Ciencias de la Salud y por formarse en el campo de la Biología Computacional. Se proveerá la infraestructura computacional necesaria, y se priorizará en todo momento el éxito del TFG del estudiante, así como la publicación

científica del trabajo si fuera posible.

**Plazas:** 1

**2. DATOS DEL TUTOR/A:**

**Nombre y apellidos:** MARÍA CORAL DEL VAL MUÑOZ

**Ámbito de conocimiento/Departamento:** CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

**Correo electrónico:** delval@ugr.es

**3. COTUTOR/A DE LA UGR (en su caso):**

**Nombre y apellidos:**

**Ámbito de conocimiento/Departamento:**

**Correo electrónico:**

**4. COTUTOR/A EXTERNO/A (en su caso):**

**Nombre y apellidos:** Jose Luis Ruiz Rodriguez

**Correo electrónico:** JOSELUIS.RUIZ@CSIC.ES

**Nombre de la empresa o institución:** INSTITUTO DE PARASITOLOGÍA Y BIOMEDICINA LÓPEZ-NEYRA (IPBLN-CSIC)

**Dirección postal:** Parque Tecnológico de Ciencias de la Salud Avenida del Conocimiento, 17 18016 Armilla (Granada)

**Puesto del tutor en la empresa o institución:** Investigador

**Centro de convenio Externo:**

**5. DATOS DEL ESTUDIANTE:**

**Nombre y apellidos:**

**Correo electrónico:**