

## Herramienta bioinformática para la identificación de motivos conservados en promotores bacterianos

Arroyo-Mateo, Antonio<sup>1,2</sup>; Delgado-Martín, Belén<sup>3</sup>; Díaz-Martínez, Luis<sup>3</sup>; Rodríguez-Moreno, Luis<sup>1,2</sup>, Ramos, Cayo<sup>1,2</sup>.

<sup>1</sup>Área de Genética, Facultad de Ciencias, Campus Teatinos, Universidad de Málaga, E-29010 Málaga

<sup>2</sup>Departamento de Microbiología y Protección de Cultivos, Instituto de Hortofruticultura Subtropical y Mediterránea «La Mayora», Extensión Campus Teatinos, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), E-29010 Málaga

<sup>3</sup>Unidad de Bioinformática, Universidad de Málaga, E-29010 Málaga

Los factores de transcripción (FTs) son proteínas que se unen a secuencias específicas de ADN llamadas elementos reguladores, que se encuentran generalmente en las regiones promotoras de los genes. Al unirse a estos elementos, pueden activar o reprimir la transcripción génica. En el caso de FTs de amplio espectro, la identificación bioinformática de sus secuencias de unión en los promotores de un genoma permite predecir el conjunto de genes cuya regulación esté relacionada. Recientemente hemos desarrollado una herramienta bioinformática que permite el análisis y la búsqueda de motivos conservados en las regiones promotoras de los genes desregulados (DEGs) que se obtienen al llevar a cabo técnicas transcriptómicas como el RNAseq. Se trata de un flujo (*pipeline*) de *scripts* encadenados capaces de ejecutarse sucesivamente y de manera automatizada, gracias a un gestor de flujos desarrollado en la Unidad de Bioinformática de la Universidad de Málaga. El flujo usa como entrada listas de DEGs provenientes de experimentos llevados a cabo en una cepa bacteriana cuyo genoma esté secuenciado. El trabajo se ejecuta de forma paralela para cada una de las listas de DEGs. El primer paso consiste en la generación de un archivo de anotación de los genes de la lista a partir de un archivo de anotación del genoma de la bacteria. A partir del archivo de anotación, se obtiene otro archivo con las coordenadas genómicas de cada uno de los genes y se genera un archivo con las coordenadas de sus promotores, que sirven para obtener las secuencias de los genes y sus promotores en un archivo multifasta. A continuación, el flujo busca en las secuencias promotoras motivos almacenados en una base de datos de la propia plataforma. También permite la búsqueda de motivos de forma dirigida usando un archivo con matrices que representa cada uno de los motivos que queremos buscar en nuestras secuencias. Esta herramienta bioinformática está resultando de gran utilidad para la búsqueda de dominios de unión de TFs conservados, y servirá como punto de partida para el estudio y descubrimiento de nuevos mecanismos de regulación en bacterias.