



Universidad Miguel Hernández de Elche

# **Easymap: un programa que facilita la cartografía de mutaciones mediante secuenciación masiva**

**Samuel Daniel Lup**

Tutores:

José Luis Micol Molina

David Wilson Sánchez

Unidad de Genética

Instituto de Bioingeniería

Máster en Biotecnología y Bioingeniería

Curso académico 2015-2016

JOSÉ LUIS MICOL MOLINA, Catedrático de Genética de la Universidad Miguel Hernández de Elche, y

DAVID WILSON SÁNCHEZ, Contratado posdoctoral de Genética de la Universidad Miguel Hernández de Elche,

HACEMOS CONSTAR:

Que el presente trabajo ha sido realizado bajo nuestra dirección y recoge fielmente la labor realizada por Samuel Daniel Lup como Trabajo de Fin del Máster en Biotecnología y Bioingeniería. Las investigaciones reflejadas en esta memoria se han desarrollado íntegramente en la Unidad de Genética del Instituto de Bioingeniería de la Universidad Miguel Hernández de Elche.

David Wilson Sánchez

José Luis Micol Molina

Elche, 30 de junio de 2017.

## I.- RESUMEN Y PALABRAS CLAVE

Las búsquedas de mutantes basadas en la genética directa han resultado útiles para la identificación de muchos genes y continúan siendo herramientas poderosas para la disección de la función y las interacciones génicas en los organismos modelo. La secuenciación masiva ha revitalizado las tediosas estrategias genéticas para la identificación de mutaciones causantes de un fenotipo de interés. La cartografía mediante secuenciación combina la secuenciación masiva con las estrategias clásicas de análisis de ligamiento y logra la identificación rápida de mutaciones puntuales. Existen programas que analizan los datos de secuenciación masiva y determinan la posición de la mutación que causa un fenotipo de interés. Sin embargo, estos programas son poco amigables, requieren *software* adicional para completar sus análisis o son relativamente caros. Hemos creado Easymap, un programa que automatiza el flujo de trabajo desde los datos brutos hasta las mutaciones candidatas. Easymap ofrece dos tipos de cartografía: de segregantes agrupados para las mutaciones puntuales y de secuencias señalizadas para las inserciones grandes, como las que causan los transposones y el ADN-T. Hemos cartografiado con Easymap mutaciones de Arabidopsis a partir de lecturas de secuenciación masiva simuladas.

**Palabras clave:** Easymap; cartografía mediante secuenciación; análisis de ligamiento; bioinformática.

Forward genetic screens have identified many genes and continue to be powerful tools for the dissection of gene action and interactions in model species. Moreover, massive sequencing has revitalized the time-consuming genetic approaches to identify the mutation causing a phenotype of interest. Mapping-by-sequencing combines next-generation sequencing with classical mapping strategies and allows rapid identification of point mutations. Programs are available that can analyze whole-genome sequencing data to map the position of the causal mutations for a specific phenotype, but these programs are complicated to install or use, require additional software to perform their analyses, or require the user to purchase expensive licenses. We created the Easymap program, which simplifies the data analysis workflow from raw reads to candidate mutations. Two main workflow types are available: bulked segregant mapping for point mutations, and tagged-sequence mapping for large insertions such as transposons or T-DNAs. We successfully tested Easymap as a tool to identify Arabidopsis mutations from simulated massive sequencing reads.

**Keywords:** Easymap; mapping-by-sequencing; linkage analysis; bioinformatics.