

MReport: análisis combinado y automatizado de datos 16S y secuenciación dirigida

Joaquín Pereira^{1*}, Alvaro Gonzalez-Revello^{1,2,3*}, Yamila Figueroa², Ana Umpiérrez², Santiago Licandro¹, Claudia Piccini², Pablo Zunino², José Sotelo-Silveira¹, Pablo Smircich¹

* Ambos autores contribuyeron de forma equivalente al trabajo

¹ Departamento de Genómica. IIBCE, MEC

² Departamento de Microbiología. IIBCE, MEC

³ Departamento de Ciencia y Tecnología de Alimentos. Facultad de Veterinaria, UdelaR
Correo: pereirajoaquin199@gmail.com

MReport es un pipeline bioinformático que tiene como objetivo la detección de microorganismos patógenos empleando datos provenientes de estrategias de metagenómica y secuenciación dirigida en forma combinada y con mínima intervención por parte del usuario. Para esto se usan como entradas principales las secuencias provenientes de ARNr 16S y de PCR dirigidos a factores de virulencia de interés en formato FASTQ. El pipeline se compone por dos flujos de trabajo, los cuales se subdividen en distintos módulos, que son organizados con el gestor SnakeMake para facilitar su automatización, reproducibilidad y escalabilidad. Cada flujo de trabajo está diseñado para trabajar respectivamente con las secuencias de ARN16S o de los factores de virulencias de interés. Según las necesidades del usuario estos se pueden ejecutar en forma independiente o conjunta, de manera parcial o completa.

El flujo de trabajo encargado del análisis metagenómico incorpora mayormente funciones del software especializado QIIME2 para el filtrado, clusterización, y asignación taxonómica de las secuencias de ARN16S. Las principales salidas de esta sección del pipeline incluyen: secuencias clusterizadas, tablas y gráficos de asignación taxonómica, árboles filogenéticos, curvas de rarefacción, análisis de componente principal, análisis de abundancia diferencial, entre otros. Por otro lado, la sección del pipeline orientada a la detección de factores de virulencia emplea el software Blastn para realizar un mapeo de secuencias y diversos programas desarrollados en Python para su posterior procesamiento. Las principales salidas de esta sección incluyen: tablas de mapeo Blast filtradas y diversas visualizaciones de la abundancia y cobertura de secuencias.

Luego, los gráficos y tablas generados por ambos flujos de trabajo se integran en un módulo diseñado específicamente para la elaboración de reportes interactivos. Además, actualmente se está desarrollando un wrapper del pipeline el cual tiene como objetivo prevenir posibles errores y facilitar su ejecución por parte del usuario.

Financiación: ANII - Fondo Sectorial de Salud Animal (FSSA_1_2019_1_160114)