

Identificación de contigs asociados a secuencias de plásmidos a partir de genoma completo de aislamientos de *Klebsiella pneumoniae*.

Diego Talero¹, Emiliano Barreto-Hernandez¹.

¹Centro de Bioinformática - Instituto de Biotecnología - Universidad Nacional de Colombia

Proyecto financiado por Colciencias enmarcado en el Macro-proyecto Diagnóstico molecular de resistencia y virulencia, y seguimiento epidemiológico de bacterias Gram negativas multirresistentes causantes de IAAS, basado en secuenciación de genoma completo (WGS) y datos sociodemográficos y clínicos

dtalero@unal.edu.co

ebarretoh@unal.edu.co

Resumen

Uno de los problemas más frecuentes en salud pública son las Infecciones Asociadas a Atención en Salud (IAAS), la Organización Mundial de la Salud ha publicado una lista de microorganismos de prioridad clínica, dentro de estas todas las Enterobacterias (incluyendo *K pneumoniae*), debido al desarrollo de multirresistencia y capacidades de diseminación alarmantes, características con frecuencia asociadas a los plásmidos. Paralelamente, el desarrollo de tecnologías de secuenciación masiva ha permitido el estudio del “comportamiento” y /o “composición” de los genomas y sus mecanismos de resistencia. El multi-Clasificador propuesto en este trabajo, permite discernir entre secuencias plasmídicas y cromosomales a partir de secuenciación de genoma completo por tecnología Illumina Miseq de cepas de *K pneumoniae*. Este flujo de trabajo escrito en Python 3.7, involucra y unifica el resultado de cinco métodos de identificación de secuencias de plásmidos a partir de genoma completo: PlasmidSPAdes V3.15, PlasFlow V1.1, Homología contra bases de datos de cromosomas y de plásmidos con Blastn V2.5.0 y Clasificación por profundidad de nucleótido con python 3.7. La optimización de cada clasificador débil, la evaluación de distintos conjuntos de clasificadores débiles y la comparación de cinco modelos de unificación; son las bases objetivas que junto al aprendizaje de máquina, dan lugar al Multi-modelo tipo **Staking(SVM)** desarrollado en el instituto, su desempeño de clasificación en base a las métricas precisión, exhaustividad y métrica-F, evidenció un mejor desempeño de clasificación en comparación con los demás modelos de unificación comparados. Este tipo de Multi-modelo permitió aprovechar las características específicas de cada clasificador débil, pues los principios biológicos de clasificación son diversos y complementarios. El desarrollo de este tipo de herramientas de Multi-clasificador requieren para su entrenamiento de una inversión de almacenamiento y procesamiento de máquina considerable, pero su implementación no requiere de una cantidad de recursos importantes.