[BIOINFO 2024]

A Snakemake-based bacterial whole genome comparison pipeline for multi-group clinical isolates

Hoeyoung Kim[1], Jinki Yeom[1,2,*], and Kwangsoo Kim[3,4*]

[1]*Interdisciplinary Program in Bioinformatics, Seoul National University*
[2]*Department of Biomedical Science, College of Medicine, Seoul National University*
[3]*Department of Medicine, College of Medicine, Seoul National University,*
[4]*Department of Transdisciplinary Medicine, Institute of Convergence Medicine with Innovative Technology, Seoul National University Hospital*
**Corresponding authors: jinki.yeom@snu.ac.kr, kwangsookim@snu.ac.kr*

Bacterial clinical isolates are often grouped based on various traits, such as infection pathways, virulence levels, or antibiotic resistance profiles. These traits are due to the adaption and evolution during the infection process, acquiring novel genes through horizontal gene transfer. Thus, discovering group-specific pathogenic marker genes and protein products are crucial for understanding the specific virulence and infection mechanisms. Comparative genomics based on whole genome sequencing is actively conducted using multi-step bioinformatics approaches, requiring various software tools. Therefore, there is an increasing need for a pipeline that can streamline such steps. Here, we developed ABComp (Assembly Polishing and Bacterial Whole-genome Comparison Pipeline for Multi-group Clinical Isolates) to compare genomes and identify pathogenic markers of clinical isolate groups in an efficient and user-friendly manner, utilizing the Snakemake workflow management system. ABComp consists of two main modules—assembly polishing and downstream comparative analysis—allowing users to initiate with either raw reads or analysis-ready genomes. Finally, we enable protein sequence alignments to identify group-specific pathogenic markers. By applying ABComp to carbapenem-resistant *Acinetobacter baumannii* (CRAB), we identified two pathogenic markers, 3-oxoacyl-ACP reductase FabG and antibiotic biosynthesis monooxygenase, specific to a highly virulent and multi-drug-resistant group. Our pipeline demonstrates the potential for automating the discovery of critical markers linked to virulence and antibiotic resistance.