[BIOINFO 2024]

PAPipe: A Pipeline for Comprehensive Population Genetic Analysis

Nayoung Park[1], Hyeonji Kim[1], Jeongmin Oh[1] Jinseok Kim[1] Charyeong Heo[1] and Jaebum Kim[1*]

[1] *Department of Biomedical science and engineering, Konkuk University*
** Corresponding author: jbkim@konkuk.ac.kr*

The rapid advancement of next-generation sequencing (NGS) technologies has increased the availability of population genetic variant data. This surge has spurred the development of diverse population analysis tools, enhancing our comprehension of population structure and evolution. However, current tools for analyzing population genetic variant data often require disparate environments, parameters, and input data formats. This complexity can create a significant barrier for researchers unfamiliar with bioinformatics, limiting the adoption of these valuable tools. To address this challenge, we have developed PAPipe, an automated and comprehensive pipeline designed to perform nine widely used population genetic analyses using population NGS data. PAPipe integrates and serializes crucial steps, including read trimming and mapping, genetic variant calling, data filtering, and format conversion. Furthermore, it incorporates nine essential population genetic analyses: principal component analysis, phylogenetic analysis, population tree analysis, population structure analysis, linkage disequilibrium decay analysis, selective sweep analysis, population admixture analysis, sequentially Markovian coalescent analysis, and fixation index analysis. PAPipe features a user-friendly web interface, enabling intuitive parameter setting and result browsing. This interface significantly enhances user convenience and data usability, making complex analyses accessible to a broader range of researchers. PAPipe generates comprehensive results that provide valuable insights into population genetics, bridging the gap between advanced analytical techniques and practical application. By streamlining the analysis process and providing an accessible platform, PAPipe aims to democratize population genetic research, enabling researchers from various backgrounds to leverage the power of NGS data. PAPipe is freely available at https://github.com/jkimlab/PAPipe.