PanGraphRNA User Manual

(version 1.0)

- PanGraphRNA is an efficient, flexible and web-based Galaxy platform that can be easily used to construct graph pangenomes from genetic variations at individual, subpopulation, and population levels. It can assist researchers to select appropriate graph pangenomes using various performance metrics for both real and simulation experiments.
- Currently, PanGraphRNA is composed of four functional modules: **Graph Pangenome Preparation Module, Construction Module, Evaluation Module, and Application Moudule**.
- PanGraphRNA was powered with an advanced packaging technology, which enables compatibility and portability.
- PanGraphRNA project is hosted on https://github.com/cma2015/PanGraphRNA
- PanGraphRNA docker image is available at https://hub.docker.com/r/malab/pangraphrna

Graph Pangenome Application Module

This module performs several applications of graph pangenomes, including differential expression analysis and expression-based QTL analysis.

Tools	Description	Input	Output	Time (test data)	Reference
Differential Expression Analysis	Perform differential expression analysis	Gene read count matrix in CSV format	Information of differentially expressed genes in CSV format	~5 mins	DESeq2
Quantitative Trait Locus Identification	Perform identification of expression quantitative trait locus	Variation files and gene expression matrix in TXT format	Identification of expression quantitative trait locus in RData and TXT format	~10 mins	MatrixEQTL

Differential Expression Analysis

This function is designed to perform differential expression analysis.

Input

- **Input CSV files:** Gene read count matrix in CSV format

Output

- Information of differentially expressed genes in CSV format

Quantitative Trait Locus Identification

This function is designed to perform identification of expression quantitative trait locus.

Input

- **Input TXT files:** Input the variation genotype file in TXT format
- **Input TXT files:** Input the variation coordinate file in TXT format
- **Input TXT files:** Input the variation covariates file in TXT format
- **Input TXT files:** Input the gene expression matrix file in TXT format
- **Input TXT files:** Input the gene coordinate file in TXT format

Output

- Identification of expression quantitative trait locus in RData and TXT format