

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 Construction Module and Alignment

This module implements a fast, memory-efficient toolkit HISAT2 to construct graph pangenomes at the individual, subpopulation, or population level. Subsequently, it performs read-genome alignment and gene expression quantification.

Tools	Description	Input	Output	Time (test data)	Reference
Individual Level Graph Pangenome	Construct individual level graph pangenome and perform read-genome alignment	Reference genome in FASTQ format and variation information in VCF format	HISAT2 alignment report in TXT format and alignment result in BAM format	~10 mins	HISAT2

Tools	Description	Input	Output	Time (test data)	Reference
Subpopulation Level Graph Pangenome	Construct subpopulation level graph pangenome and perform read-genome alignment	Reference genome in FASTQ format and variation information in VCF format	HISAT2 alignment report in TXT format and alignment result in BAM format	~10 mins	HISAT2
Population Level Graph Pangenome	Construct population level graph pangenome and perform read-genome alignment	Reference genome in FASTQ format and variation information in VCF format	HISAT2 alignment report in TXT format and alignment result in BAM format	~10 mins	HISAT2
Expression Quantification	Perform expression quantification for read-genome alignment result	Alignment result in BAM format and annotation file in GTF format	Gene read count matrix in TXT format and gene expression quantification matrix in TXT format	~5 mins	StringTie

Individual Level Graph Pangenome

This function is designed to construct individual level graph pangenome and perform read-genome alignment.

Input

- **Input FASTQ files:** Input cleaned sequence reads in FASTQ format
- **Input VCF files:** Input the variation information in VCF format

Output

- HISAT2 alignment report in TXT format

- HISAT2 alignment result in BAM format

Subpopulation Level Graph Pangenome

This function is designed to construct subpopulation level graph pangenome and perform read-genome alignment.

Input

- **Input FASTQ files:** Input cleaned sequence reads in FASTQ format
- **Input VCF files:** Input the variation information in VCF format

Output

- HISAT2 alignment report in TXT format
- HISAT2 alignment result in BAM format

Population Level Graph Pangenome

This function is designed to construct population level graph pangenome and perform read-genome alignment.

Input

- **Input FASTQ files:** Input cleaned sequence reads in FASTQ format
- **Input VCF files:** Input the variation information in VCF format

Output

- HISAT2 alignment report in TXT format
- HISAT2 alignment result in BAM format

Expression Quantification

This function is designed to perform expression quantification for read-genome alignment result.

Input

- **Input BAM files:** Input read-genome alignment results in BAM format
- **Input GTF files:** Input reference genome annotation file in GTF format

Output

- Gene read count matrix in TXT format
- Gene expression quantification matrix in TXT format