

# 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 Evaluation Module

This module assesses the performance of graph pangenomes against the SLR strategy with six read mapping-relevant measurements: *Unique alignment rate, Multiple alignment rate, Error mapping rate, F1 score, Recall and Precision.*

Tools	Description	Input	Output	Time (test data)	Reference
<b>Basic Alignment Statistics</b>	Obtain the basic alignment information among multiple read-genome alignment results	HISAT2 alignment report in TXT format and graph pangenome files	Basic alignment statistics matrix in TXT format	~1 mins	/

Tools	Description	Input	Output	Time (test data)	Reference
<b>Mapping Error Measurement</b>	Perform RNA-seq reads simulation and measure mapping errors	Ground truth files and graph pangenome files	Measurement of mapping errors in CSV format	~10 mins	<a href="#">Flux</a> and <a href="#">HISAT2</a>
<b>F1 Score Calculation</b>	Perform RNA-seq reads simulation and calculate F1 score	Graph pangenome files	Calculation of F1 score in CSV format	~10 mins	<a href="#">Flux</a> and <a href="#">HISAT2</a>

## Basic Alignment Statistics

This function is designed to obtain the basic alignment information among multiple read-genome alignment results.

### Input

- **Input TXT files:** Input HISAT2 alignment reports in TXT format
- **Input FASTA files:** Input reference genome annotation file for preliminary path in FASTA format

### Output

- Basic alignment statistics matrix in TXT format

## Mapping Error Measurement

This function is designed to perform RNA-seq reads simulation and measure mapping errors.

### Input

- **Ground Truth Setting**
- **Input FASTA files:** Input the ground truth reference genome in FASTA format
- **Input CHAIN files:** Input the chain file of ground truth genome (GroundTruth to ReferenceGenome) in CHAIN format
- **Graph Pangenome Information**

- **Input FASTA files:** Input the reference genome file that have been used for graph pangenome construction in FASTA format
- **Input VCF files:** Input the vcf file that have been used for graph pangenome construction in VCF format
- **Input GTF files:** Input the exon annotation file of reference genome that have been used for graph pangenome construction in GTF format
- **RNA-Seq Reads Simulation**
- **Input VCF files:** Input the vcf file containing the accession name for pseudo genome construction in VCF format

## Output

- Measurement of mapping errors in CSV format

## F1 Score Calculation

This function is designed to perform RNA-seq reads simulation and calculate F1 score.

## Input

- **Graph Pangenome Information**
- **Input FASTA files:** Input the reference genome file that have been used for graph pangenome construction in FASTA format
- **Input VCF files:** Input the vcf file that have been used for graph pangenome construction in VCF format
- **Input GTF files:** Input the exon annotation file of reference genome that have been used for graph pangenome construction in GTF format
- **RNA-Seq Reads Simulation**
- **Input VCF files:** Input the vcf file containing the accession name for pseudo genome construction in VCF format

## Output

- Calculation of F1 score in CSV format