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

Individual Level Graph Pangenome

In this function, an ultrafast and memory-efficient tool **HISAT2** (Kim, D., *et al.*, 2019) is integrated for construccting individual level graph pangenomes and aligning sequencing reads. See https://daehwankimlab.github.io/hisat2/manual for details.

Input

- **Input reference genome file:** Input reference genome file for primary path of graph pangenome in FASTA format
- **Input VCF file:** Input VCF file containing variant information to be integrated into the primary path of graph pangenome in VCF format
- **Input FASTQ file:** Cleaned single-end or paired-end RNA-seq reads in FASTQ format

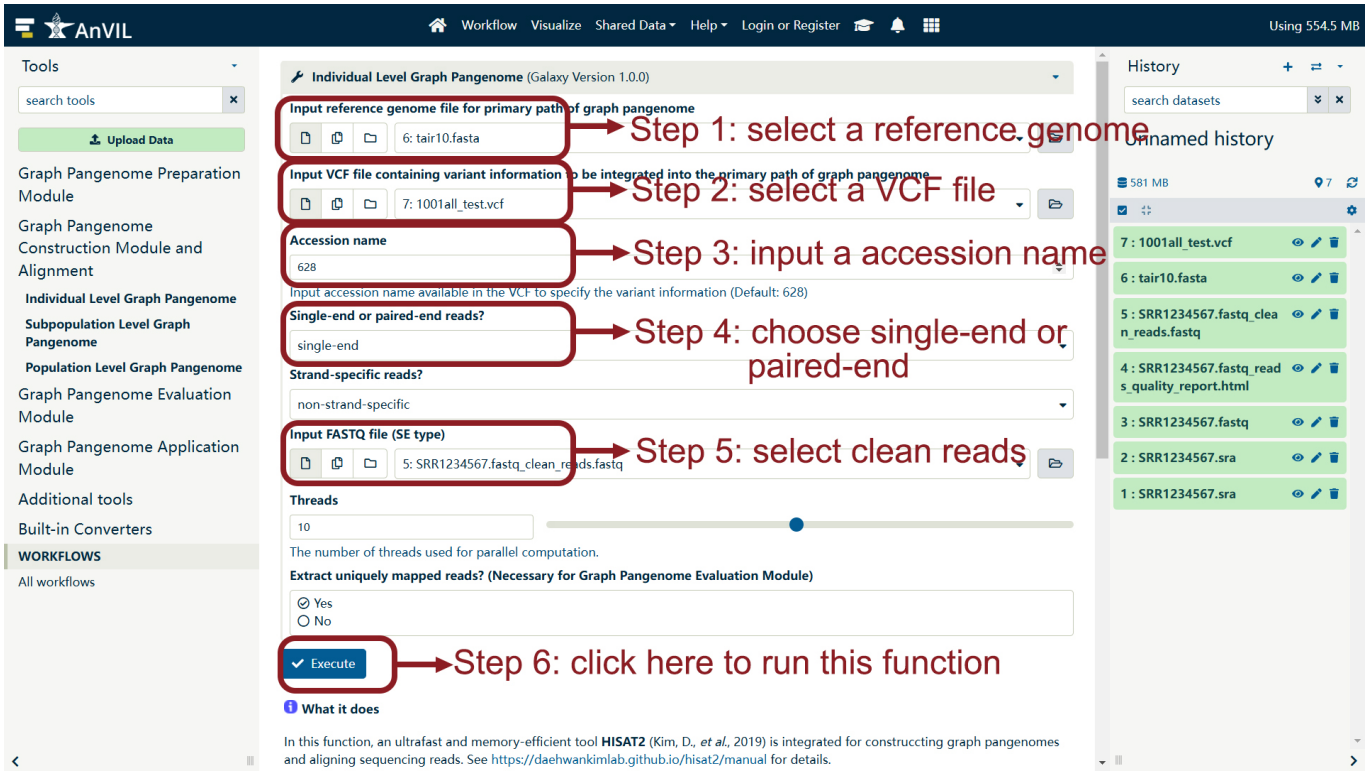
Parameters

- **Accession name:** Input accession name available in the VCF to specify the variant data (Default: 628)
- **Threads:** The number of threads used for parallel computation (Default: 10)

Output

- **HISAT2 alignment report in TXT format**

- **HISAT2 alignment result in BAM format**



Subpopulation Level Graph Pangenome

In this function, an ultrafast and memory-efficient tool **HISAT2** (Kim, D., *et al.*, 2019) is integrated for constructing subpopulation level graph pangenomes and aligning sequencing reads. See https://daehwankimlab.github.io/hisat2/manual for details.

Input

- **Input reference genome file:** Input reference genome file for primary path of graph pangenome in FASTA format
- **Input VCF file:** Input VCF file containing variant information to be integrated into the primary path of graph pangenome in VCF format
- **Input accession name list:** Input accession name list (TXT file) available in the VCF file to specify the variant data
- **Input FASTQ file:** Cleaned single-end or paired-end RNA-seq reads in FASTQ format

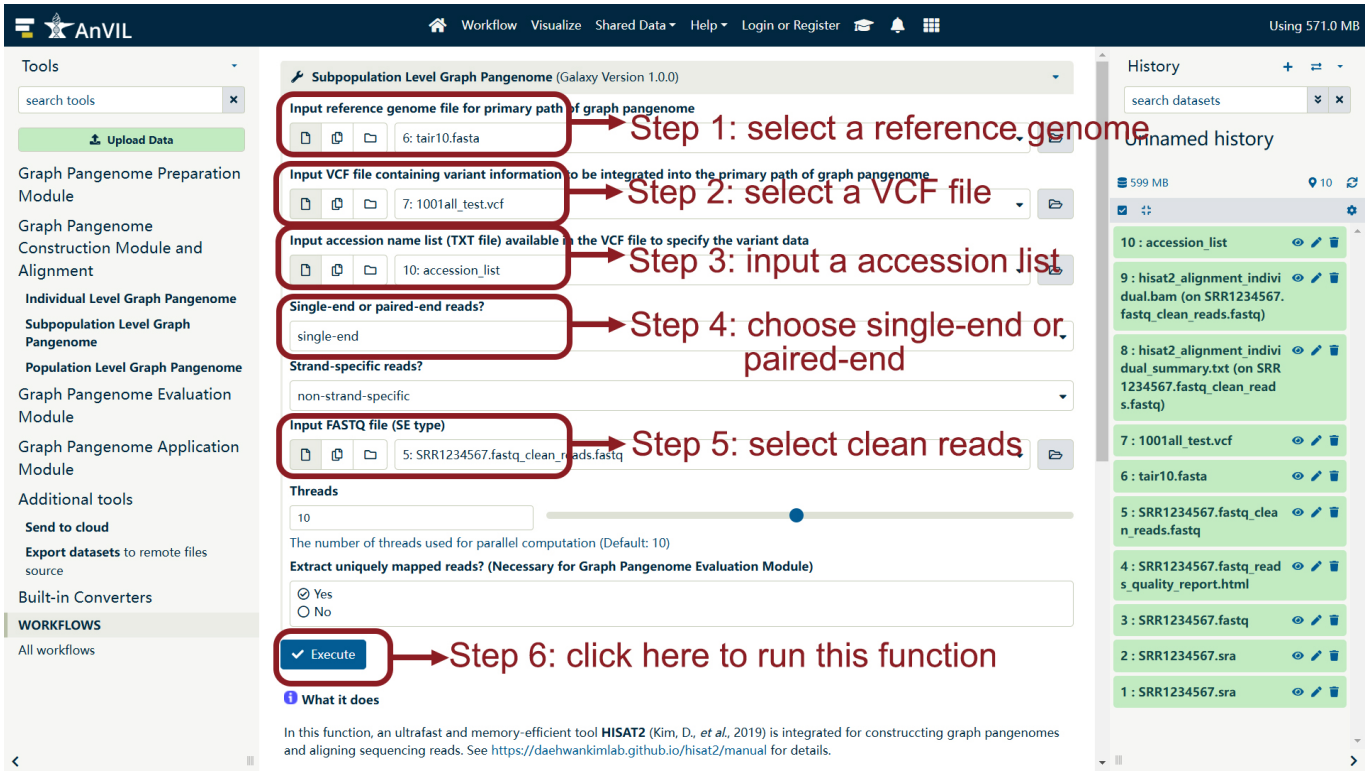
Parameters

- **Threads:** The number of threads used for parallel computation (Default: 10)

Output

- **HISAT2 alignment report in TXT format**

- **HISAT2 alignment result in BAM format**



Population Level Graph Pangenome

In this function, an ultrafast and memory-efficient tool **HISAT2** (Kim, D., *et al.*, 2019) is integrated for constructing population level graph pangenomes and aligning sequencing reads. See https://daehwankimlab.github.io/hisat2/manual for details.

Input

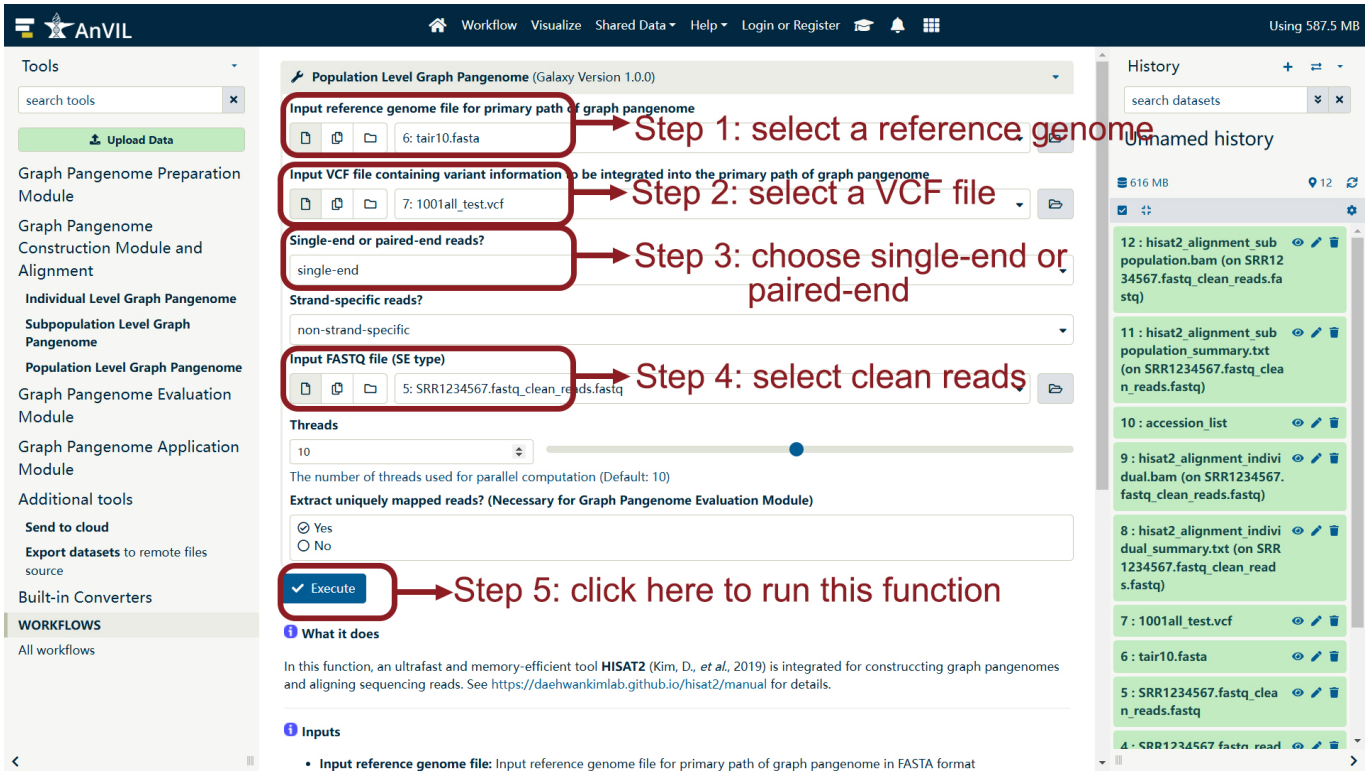
- **Input reference genome file:** Input reference genome file for primary path of graph pangenome in FASTA format
- **Input VCF file:** Input VCF file containing variant information to be integrated into the primary path of graph pangenome in VCF format
- **Input FASTQ file:** Cleaned single-end or paired-end RNA-seq reads in FASTQ format

Parameters

- **Threads:** The number of threads used for parallel computation (Default: 10)

Output

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


AnVIL
Workflow
Visualize
Shared Data
Help
Login or Register
Using 587.5 MB
Tools
search tools
Upload Data
Graph Pangenome Preparation Module
Graph Pangenome Construction Module and Alignment
Individual Level Graph Pangenome
Subpopulation Level Graph Pangenome
Population Level Graph Pangenome
Graph Pangenome Evaluation Module
Graph Pangenome Application Module
Additional tools
Send to cloud
Export datasets to remote files source
Built-in Converters
WORKFLOWS
All workflows
Population Level Graph Pangenome (Galaxy Version 1.0.0)
Input reference genome file for primary path of graph pangenome
6: tair10.fasta
Step 1: select a reference genome
Input VCF file containing variant information to be integrated into the primary path of graph pangenome
7: 1001all_test.vcf
Step 2: select a VCF file
Single-end or paired-end reads?
single-end
Step 3: choose single-end or paired-end
Strand-specific reads?
non-strand-specific
Input FASTQ file (SE type)
5: SRR1234567.fastq_clean_reads.fastq
Step 4: select clean reads
Threads
10
The number of threads used for parallel computation (Default: 10)
Extract uniquely mapped reads? (Necessary for Graph Pangenome Evaluation Module)
Yes
No
Execute
Step 5: click here to run this function
What it does
In this function, an ultrafast and memory-efficient tool HISAT2 (Kim, D., et al., 2019) is integrated for constructing graph pangenomes and aligning sequencing reads. See https://daehwankimlab.github.io/hisat2/manual for details.
Inputs
Input reference genome file: Input reference genome file for primary path of graph pangenome in FASTA format
History
search datasets
Unnamed history
616 MB
12
12 : hisat2_alignment_subpopulation.bam (on SRR1234567.fastq_clean_reads.fastq)
11 : hisat2_alignment_subpopulation_summary.txt (on SRR1234567.fastq_clean_reads.fastq)
10 : accession_list
9 : hisat2_alignment_individual.bam (on SRR1234567.fastq_clean_reads.fastq)
8 : hisat2_alignment_individual_summary.txt (on SRR1234567.fastq_clean_reads.fastq)
7 : 1001all_test.vcf
6 : tair10.fasta
5 : SRR1234567.fastq_clean_reads.fastq