Global Pneumococcal Sequencing Project

wellcome sanger institute

GPS PIPELINE

Quickstart Guide

Requirements

- Compatible with most operating systems: Linux, Windows (running Linux with WSL2), macOS
- Java 11+ or OpenJDK 11+
- Docker or Singularity/Apptainer
- Have at least 16GB of RAM and 50GB of free storage

Setup (Internet connection required)

1. Download or Git Clone the pipeline core files from its GitHub Repository
   a. Download from: https://github.com/sanger-bentley-group/gps-pipeline/releases
   b. To clone, run: `git clone https://github.com/sanger-bentley-group/gps-pipeline.git`
2. Initialise the pipeline after changing directory (`cd`) into the pipeline directory:
   a. Using Docker: `./run_pipeline --init`
   b. Using Singularity: `./run_pipeline --init -profile singularity`
3. This can take a while, as it will download 13GB of container images and 8GB of databases

Run (No internet connection required after initialisation)

1. Run the pipeline with the directory containing your FASTQ files as the input using `--reads`
   a. Using Docker: `./run_pipeline --reads /path/to/reads-dir`
   b. Using Singularity: `./run_pipeline --reads /path/to/reads-dir -profile singularity`
2. Grab a cup of tea and wait

Tip 1:
If you have not specified output path with `--output`, the default is the `output` directory in the pipeline directory.

Tip 2:
Each input sample will generate ~2GB intermediate files on average. You might need to process your samples in batches if the storage space is limited on your system. The `clean_pipeline` helper script of the pipeline may be useful after each successful run.

Documentation

- GitHub Repository: https://github.com/sanger-bentley-group/gps-pipeline

Notice

- The current release of the pipeline only works with Illumina paired-end short reads
- Use a specific version of the pipeline to ensure consistent output for the same study