

Allpaths on HPC

What is Allpaths?

ALLPATHS-LG is a short read assembler, and it works on both small and large (mammalian size) genomes. To use it, you should first generate ~100 base Illumina reads from two libraries: one from ~180 bp fragments, and one from ~3000 bp fragments, both at about 45x coverage. Sequence from longer fragments will enable longer-range continuity.

More info on QUAST:

[Allpaths-Homepage](#)

Versions Available:

The following versions are available on the cluster:

- Allpaths -LG

How to load Quast?

To load Allpaths on the HPC, use the following commands:

```
module avail bio/allpaths1g
#This will load allpaths-1g
```

To verify if the module and dependencies are loaded correctly, use the following command.

```
module list
```

This should list all the software and dependencies that are loaded.

How to use Quast?

There are two ways to use most of the software on the cluster.

The Interactive Way:

To run the Allpaths in an interactive session, jump into a base terminal of a compute node using the following command:

```
srun -p main --qos main -n 1 -c 16 --mem 40G --pty bash
```

Now, load the Allpaths module and start your job interactively on the terminal.

```
#This command checks if allpaths is correctly installed and runs a  
#test script  
ParallelSortTest X=4 P=16
```

Note: This is preferred usage in case of short runtime jobs and debugging. After you exit the terminal session, your job will be terminated unlike batch method.

The Script method:

To use this method, the user will need to write a batch script and submit it to a slurm.

To copy the example file, use the following command,

```
#Make a test dir on home  
mkdir ~/ap-test && cd ~/ap-test  
  
#get the input files from the web  
wget ftp://ftp.broadinstitute.org/pub/crd/ALLPATHS/Release-LG/test.genome.tar.gz  
  
#unpack the data with tar  
tar -xvf *.tar.gz  
  
#go to the extracted directory  
cd * && mkdir script.sbatch
```

All the test data required will be extracted and ready to process. The sbatch script should be formatted like

```
#!/bin/bash
#SBATCH -J ap-test
#SBATCH -n 1
#SBATCH -c 24
#SBATCH --mem=32G
#SBATCH --time=00:15:00
#SBATCH -p main
#SBATCH --qos main
#SBATCH -e err.err
#SBATCH -o out.out

module load bio/allpaths*

# Move to test data directory
cd $SLURM_SUBMIT_DIR
# Run Allpaths -LG the test files here
sh prepare.sh
sh assemble.sh
#Feel free to look at above pre-written script for all-paths.
```

All the output files will be under the job submit directory.

Where to find help?

If you are stuck on some part or need help at any point, please contact OIT at the following address.

<https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp>