

Working directory: /home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus
Output directory: rnabloom/12875770_3

=====
==

JOB-SPECIFIC METADATA

=====
==

Job 12914855 began:
File list is: /home/tfsevers/manifests/20231109_rnabloom_pool.txt
Input filename:
File prefix

=====
==

SLURM CONFIGURATION

=====
==

Job: 12914855
Array Job ID: 12875770
Array Task ID: 3
Job name: rnabloom
Directory: /scratch/tfsevers/data/transcriptomes/resynthesized_bnapus
Nodes: gra1016
Host: gra-login2
Nodes (n): 1
Cores/Node: 32
CPU per task: 32
ntasks: 1
Date:

=====
==

JOB ARRAY SCRIPT

=====
==

Array job command: /home/tfsevers/bin/run_rnabloom_pool.sh

=====
==

```
/home/tfsevers/bin/run_rnabloom_pool.sh
#!/bin/bash
#SBATCH --account=def-jwhitton
#SBATCH --time=71:59:00
#SBATCH --ntasks=1
#SBATCH --array=1
#SBATCH --mem=200G
#SBATCH --cpus-per-task=32
#SBATCH --mail-user=tonyaseverson@gmail.com
#SBATCH --mail-type=ALL
#SBATCH --job-name=rnabloom
#SBATCH -o /home/tfsevers/logs/slurm/rnabloom/%A_%a_%x.out
#SBATCH -e /home/tfsevers/logs/slurm/rnabloom/%A_%a_%x.err
```

```

module --force purge    # Recommended to avoid environment loading
conflicts
module load StdEnv/2023
module load java/17.0.6
module load minimap2/2.26

export LC_COLLATE=C

# Inputs
flist=$1
replicates=$2 # To get the number of array jobs, divide the

# Identify last line of manifest to include in output of `head`
task=$SLURM_ARRAY_TASK_ID
last=$(( ${task} * ${replicates} ))
echo $last

# Outputs
outdir="${SLURM_JOB_NAME}/${SLURM_ARRAY_JOB_ID}_${task}"
mkdir -p $outdir
sample_list="$outdir/pooled_samples.txt"
head -n $last $flist | tail -n $replicates > $sample_list

divider='=====
===== '

# Report input filenames and locations
echo "$CURRENTDATE"
echo "Working directory: $PWD"
echo "Output directory: $outdir"
echo "$divider"

# Job-specific info
echo "JOB-SPECIFIC METADATA"
echo "$divider"
echo "Job $SLURM_JOB_ID began: $CURRENTDATE"
echo "File list is: $flist"
echo "Input filename: $fname"
echo "File prefix $fprfx"
echo "$divider"

# Record Slurm and run setup info
echo "SLURM CONFIGURATION"
echo "$divider"
echo "Job: $SLURM_JOB_ID"
echo "Array Job ID: $SLURM_ARRAY_JOB_ID"
echo "Array Task ID: $SLURM_ARRAY_TASK_ID"
echo "Job name: $SLURM_JOB_NAME"
echo "Directory: $SLURM_SUBMIT_DIR"

```

```

echo "Nodes: $SLURM_JOB_NODELIST"
echo "Host: $SLURM_SUBMIT_HOST"
echo "Nodes (n): $SLURM_JOB_NUM_NODES"
echo "Cores/Node: $SLURM_CPUS_ON_NODE"
echo "CPU per task: $SLURM_CPUS_PER_TASK"
echo "ntasks: $SLURM_NTASKS"
echo "Date: $CURRENTDATE"
echo "${divider}"

# Display original command and script
origCMD=$(scontrol show job $SLURM_JOB_ID | awk -F= '/Command={print $2}')
IFS=' ' read -r -a oJobInfo <<< "$origCMD"
jobScript=${oJobInfo}

echo "JOB ARRAY SCRIPT"
echo "${divider}"
echo "Array job command: $origCMD"
echo "${divider}"
echo "$jobScript"
cat "$jobScript"
echo "${divider}"

# Code
java -Xmx10g -jar ~/bin/thirdparty/RNA-Bloom_v2.0.1/RNA-Bloom.jar
-stranded -ntcard -fpr 0.005 -k 25-75:5 -extend -t
$SLURM_CPUS_PER_TASK -outdir $outdir -rcr -pool "$sample_list"

wait
echo $(date +%F_%T)
=====
==
RNA-Bloom v2.0.1
args: [-stranded, -ntcard, -fpr, 0.005, -k, 25-75:5, -extend, -t, 32,
-outdir, rnabloom/12875770_3, -rcr, -pool, rnabloom/12875770_3/
pooled_samples.txt]

name: rnabloom
outdir: rnabloom/12875770_3
Pooled assembly mode is ON!
Parsing pool reads list file `rnabloom/12875770_3/
pooled_samples.txt`...

K-mer counting with ntCard...
Running command: `ntcard -t 32 -k 25 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k25.hist`...
Unique k-mers (k=25): 1,417,252,736
Unique k-mers (k=25,c>1): 505,722,931

```

```

Running command: `ntcard -t 32 -k 30 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k30.hist`...
Unique k-mers (k=30):      1,639,788,516
Unique k-mers (k=30,c>1): 550,015,161
Running command: `ntcard -t 32 -k 35 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k35.hist`...
Unique k-mers (k=35):      1,834,826,731
Unique k-mers (k=35,c>1): 583,544,336
Running command: `ntcard -t 32 -k 40 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k40.hist`...
Unique k-mers (k=40):      1,999,448,392
Unique k-mers (k=40,c>1): 610,295,110
Running command: `ntcard -t 32 -k 45 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k45.hist`...
Unique k-mers (k=45):      2,134,151,889
Unique k-mers (k=45,c>1): 628,722,368
Running command: `ntcard -t 32 -k 50 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k50.hist`...
Unique k-mers (k=50):      2,240,788,621
Unique k-mers (k=50,c>1): 641,215,375
Running command: `ntcard -t 32 -k 55 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k55.hist`...
Unique k-mers (k=55):      2,325,324,773
Unique k-mers (k=55,c>1): 649,616,363
Running command: `ntcard -t 32 -k 60 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k60.hist`...
Unique k-mers (k=60):      2,387,223,474
Unique k-mers (k=60,c>1): 652,792,521
Running command: `ntcard -t 32 -k 65 -c 65535 -p rnabloom/12875770_3/
rnabloom @rnabloom/12875770_3/rnabloom.ntcard.readslist.txt`...
Parsing histogram file `rnabloom/12875770_3/rnabloom_k65.hist`...
Unique k-mers (k=65):      2,421,804,075
Unique k-mers (k=65,c>1): 650,412,002
Setting k to 60
K-mer counting completed in 55m 6s

```

Bloom filters	Memory (GB)
de Bruijn graph:	7.579159
k-mer counting:	16.580326
paired k-mers (reads):	7.579159
paired k-mers (frags):	7.579159
screening:	7.579159

=====
Total: 46.89696

```
> Stage 1: Construct graph from reads (k=60)
Sampling read lengths (l>=60, n=10000) from each file...
  min   Q1    M    Q3   max   file
  60   143   149   150   150   /home/tfsevers/
scratch/data/transcriptomes/resynthesized_bnapus/bbsplit_cleaned/
10734406/bn3501_2.cleaned.cleaned.fq.gz
  60   141   148   150   150   /home/tfsevers/
scratch/data/transcriptomes/resynthesized_bnapus/bbsplit_cleaned/
10734406/bn3502_2.cleaned.cleaned.fq.gz
  60   142   148   150   150   /home/tfsevers/
scratch/data/transcriptomes/resynthesized_bnapus/bbsplit_cleaned/
10734406/bn3503_2.cleaned.cleaned.fq.gz
  60   143   148   150   150   /home/tfsevers/
scratch/data/transcriptomes/resynthesized_bnapus/bbsplit_cleaned/
10734406/bn3501_1.cleaned.cleaned.fq.gz
  60   141   148   150   150   /home/tfsevers/
scratch/data/transcriptomes/resynthesized_bnapus/bbsplit_cleaned/
10734406/bn3502_1.cleaned.cleaned.fq.gz
  60   143   148   150   150   /home/tfsevers/
scratch/data/transcriptomes/resynthesized_bnapus/bbsplit_cleaned/
10734406/bn3503_1.cleaned.cleaned.fq.gz
Weighted [Q1,M,Q3]: [142, 148, 150]
Absolute [min,max]: [60, 150]
Paired k-mers distance: 72
Max. tip length: 88
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3501_2.cleaned.cleaned.fq.gz`...
Parsed 141,185,293 sequences in 36m 22s
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3502_2.cleaned.cleaned.fq.gz`...
Parsed 131,290,521 sequences in 34m 50s
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3503_2.cleaned.cleaned.fq.gz`...
Parsed 146,924,770 sequences in 38m 46s
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3501_1.cleaned.cleaned.fq.gz`...
Parsed 141,314,013 sequences in 38m 26s
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3502_1.cleaned.cleaned.fq.gz`...
Parsed 131,389,982 sequences in 35m 2s
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
```

```
bn3503_1.cleaned.cleaned.fq.gz`...
Parsed 147,008,170 sequences in 38m 56s
Parsed 839,112,749 sequences from 6 files.
DBG Bloom filter FPR: 0.511 %
Counting Bloom filter FPR: 0.536 %
Reads paired k-mers Bloom filter FPR: 0.0874 %
> Stage 1 completed in 3h 43m 13s
```

```
> Stage 2: Assemble fragments for 1 samples
>> Working on "bn35" (sample 1 of 1)...
Parsing `/home/tfsevers/scratch/data/transcriptomes/
resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3501_2.cleaned.cleaned.fq.gz` and `/home/tfsevers/scratch/data/
transcriptomes/resynthesized_bnapus/bbsplit_cleaned/10734406/
bn3501_1.cleaned.cleaned.fq.gz`...
Fragment Lengths Sampling Distribution (n=9530)
      min    Q1    M    Q3    max
      64    153   205   279   894
Paired k-mers distance: 83
Max. graph traversal depth: 468
Parsed 124,477,590 read pairs.
      connected: 118,319,869 (95.05315%)
      not connected: 6,074,056 (4.8796377%)
      discarded: 83,665 (0.0672129%)
Fragments paired kmers Bloom filter FPR: 0.00636366 %
Screening Bloom filter FPR: 0.0054562273 %
>> Fragments assembled in 3h 8m 6s
```

```
> Stage 2 completed in 3h 8m 6s
```

```
> Stage 3: Assemble transcripts for 1 samples
>> Working on "bn35" (sample 1 of 1)...
Rebuilding graph from assembled fragments...
[14] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e1.nbits`...
[2] Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e1.nbits`...
[20] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.01.nbits`...
[13] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e0.nbits`...
[4] Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e3.nbits`...
[15] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e2.nbits`...
[18] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e5.nbits`...
[9] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.e2.nbits`...
[3] Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e2.nbits`...
[5] Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e4.nbits`...
[6] Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e5.nbits`...
[8] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.e1.nbits`...
[1] Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e0.nbits`...
[19] Parsing `rnabloom/12875770_3/bn35/
```

```
bn35.fragments.long.01.nbits`...
[21] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.01.nbits`...
[17] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e4.nbits`...
[10] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.e3.nbits`...
[12] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.e5.nbits`...
[7] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.e0.nbits`...
[16] Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e3.nbits`...
[11] Parsing `rnabloom/12875770_3/bn35/
bn35.fragments.short.e4.nbits`...
[12] Parsed 43 sequences.
[6] Parsed 187 sequences.
[11] Parsed 864 sequences.
[5] Parsed 5,763 sequences.
[20] Parsed 49,837 sequences.
[9] Parsed 71,589 sequences.
[8] Parsed 123,197 sequences.
[18] Parsed 262,796 sequences.
[7] Parsed 203,846 sequences.
[4] Parsed 61,600 sequences.
[17] Parsed 900,240 sequences.
[16] Parsed 1,174,903 sequences.
[14] Parsed 1,098,117 sequences.
[15] Parsed 1,309,855 sequences.
[13] Parsed 1,305,253 sequences.
[3] Parsed 508,397 sequences.
[21] Parsed 5,762,749 sequences.
[2] Parsed 1,491,879 sequences.
[19] Parsed 1,721,095 sequences.
[1] Parsed 2,155,266 sequences.
DBG Bloom filter FPR: 0.024 %
Reads paired k-mers Bloom filter FPR: 0.00929 %
Fragments paired k-mers Bloom filter FPR: 0.00707 %
Graph rebuilt in 4m 19s
Assembling transcripts...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e5.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e4.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e3.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e2.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e1.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.e0.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.e5.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.e4.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.e3.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.e2.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.e1.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.e0.nbits`...
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e5.nbits`...
```

```
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e4.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e3.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e2.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e1.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.e0.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.long.01.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.fragments.short.01.nbits` ...
Parsing `rnabloom/12875770_3/bn35/bn35.unconnected.01.nbits` ...
Parsed 18,218,163 fragments.
Screening Bloom filter FPR: 0.0011488696 %
Transcripts assembled in 5m 42s
Assembled transcripts at `rnabloom/12875770_3/bn35/
bn35.transcripts.fa`
Reducing redundancy in assembled transcripts...
Parsed 2,149,461 overlap records in 1m 54s
contained reads: 45,609
dovetail reads: 12,524
G: |V|=12,524 |E|=9,368
G: |V|=12,524 |E|=5,298
before: 254,843 after: 206,588
Redundancy reduced in 1m 58s

> Stage 3 completed in 12m 4s
Total runtime: 7h 58m 32s
2023-11-16_13:20:39
```