***Idotea baltica* RNA-seq *de novo* Assembly**

Trinity 2.1.0

Relative to 25-mers, 20-mers will reduce specificity and increase sensitivity

Batch jobs file using 20-mers (trinity\_20.sh)

#SBATCH --cpus-per-task=8

#SBATCH -t 12:00:00

#SBATCH --mem=88000

module load biokit

Trinity --seqType fq \

--max\_memory 80G \

--left /wrk/ketaya/trinity/trinity\_rcor\_norm/insilico\_read\_normalization/out\_fw\_paired.cor.fq.normalized\_K25\_C50\_pctSD200.fq \

--right /wrk/ketaya/trinity/trinity\_rcor\_norm/insilico\_read\_normalization/out\_rev\_paired.cor.fq.normalized\_K25\_C50\_pctSD200.fq \

--CPU $SLURM\_CPUS\_PER\_TASK \

--KMER\_SIZE 20 \

--grid\_conf $TRINITY\_HOME/hpc\_conf\_taito.slurm

\*\*\* NOTE! memory should be at least 10x cpus

**RUNNING**

Run (job 7664532 17.11.2015 13.20)

sbatch trinity\_20.sh

\*\*\* ~5hrs (8 cores, 80GB RAM)

**OUPUT**

Output Statistics found in file ‘Trinity.timimg’

Statistics:

===========

Trinity Version: v2.1.0

Compiler: GCC

Trinity Parameters: --seqType fq --max\_memory 80G --left /wrk/ketaya/trinity/trinity\_rcor\_norm/insilico\_read\_normalization/out\_fw\_paired.cor.fq.normalized\_K25\_C50\_pctSD200.fq --right /wrk/ketaya/trinity/trinity\_rcor\_norm/insilico\_read\_normalization/out\_rev\_paired.cor.fq.normalized\_K25\_C50\_pctSD200.fq --CPU 8 --KMER\_SIZE 20 --grid\_conf /appl/bio/trinity/trinityrnaseq-2.1.0/hpc\_conf/taito.slurm

Paired mode

Input data

Left.fasta 0 MByte

Right.fasta 0 MByte

Number of unique KMERs: 86883676

Number of reads: 0 Output data

Trinity.fasta 49 MByte

Runtime

=======

Start: Tue Nov 17 17:57:31 EET 2015

End: Tue Nov 17 18:46:32 EET 2015

Trinity 2941 seconds

Inchworm 0 seconds

Chrysalis 2941 seconds

Butterfly 0 seconds

Rest 0 seconds

Timing stats are messed up when restarting a Trinity run

(Didn’t give it enough time to begin with so I had to resubmit the batch job, it used the old Trinity out dir)

Nx Stats

################################

## Counts of transcripts, etc.

################################

Total trinity 'genes': 77981

Total trinity transcripts: 78037

Percent GC: 39.23

########################################

Stats based on ALL transcript contigs:

########################################

Contig N10: 2620

Contig N20: 1858

Contig N30: 1434

Contig N40: 1097

Contig N50: 826

Median contig length: 342

Average contig: 577.05

Total assembled bases: 45031014

#####################################################

## Stats based on ONLY LONGEST ISOFORM per 'GENE':

#####################################################

Contig N10: 2616

Contig N20: 1858

Contig N30: 1434

Contig N40: 1097

Contig N50: 826

Median contig length: 341

Average contig: 576.94

Total assembled bases: 44990508