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

Trinity 2.1.0

Relative to 25-mers, 30-mers will increase specificity and reduce sensitivity

Batch jobs file using 30-mers (trinity\_30.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 30 \

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

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

**RUNNING**

Run (job 7670684 17.11.2015 20.00)

sbatch trinity\_30.sh

\*\*\* ~3hrs (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 30 --grid\_conf /appl/bio/trinity/trinityrnaseq-2.1.0/hpc\_conf/taito.slurm

Paired mode

Input data

Left.fasta 1653 MByte

Right.fasta 1642 MByte

Number of unique KMERs: 90778429

Number of reads: 0 Output data

Trinity.fasta 104 MByte

Runtime

=======

Start: Tue Nov 17 20:06:47 EET 2015

End: Tue Nov 17 23:03:29 EET 2015

Trinity 10602 seconds

Inchworm 542 seconds

Chrysalis 10013 seconds

Butterfly 0 seconds

Rest 47 seconds

Nx Stats

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

## Counts of transcripts, etc.

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

Total trinity 'genes': 90309

Total trinity transcripts: 115561

Percent GC: 39.31

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

Stats based on ALL transcript contigs:

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

Contig N10: 3487

Contig N20: 2564

Contig N30: 1968

Contig N40: 1526

Contig N50: 1154

Median contig length: 359

Average contig: 683.20

Total assembled bases: 78951148

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

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

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

Contig N10: 3104

Contig N20: 2176

Contig N30: 1593

Contig N40: 1152

Contig N50: 814

Median contig length: 313

Average contig: 561.57

Total assembled bases: 50715117