***Idotea baltica* Assembly Evaluation**

TransRate v1.0.1

**INSTALLING**

Change directory to USERAPPL and create new directory

cd $USERAPPL

mkdir transrate

Go to the *transrate* directory and download the installation package

cd transrate

wget https://bintray.com/artifact/download/blahah/generic/transrate-1.0.1-linux-x86\_64.tar.gz

Unzip the package

tar -zxvf transrate-1.0.1-linux-x86\_64.tar.gz

Tell the shell the location of the executable

export PATH=${PATH}:${USERAPPL}/transrate/transrate-1.0.1-linux-x86\_64

**NOTE!** This command must be run every time logging into Taito

**RUNNING**

Read mapping (assembly in FASTA, reads in FASTQ)

transrate --assembly trinity\_20.fa,trinity\_25.fa,trinity\_30.fa \

--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 \

--threads 16

\*\*\* ~30min

Output in assemblies.csv

25-mers has the highest optimized score

Run Nx stats on optimized 25-mers

module load biokit

$TRINITY\_HOME/util/TrinityStats.pl good.trinity\_25.fa

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

## Counts of transcripts, etc.

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

Total trinity 'genes': 30147

Total trinity transcripts: 40122

Percent GC: 39.70

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

Stats based on ALL transcript contigs:

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

Contig N10: 3853

Contig N20: 3000

Contig N30: 2450

Contig N40: 2053

Contig N50: 1710

Median contig length: 900

Average contig: 1184.48

Total assembled bases: 47523534

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

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

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

Contig N10: 3659

Contig N20: 2846

Contig N30: 2282

Contig N40: 1871

Contig N50: 1536

Median contig length: 729

Average contig: 1028.17

Total assembled bases: 30996227

\*\*\* Only 1/3 of contigs kept

**Rerun with optimized assembly**

transrate --assembly good.trinity\_25.fa \

--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 \

--threads 16

Smallest length is still 201 so it doesn’t filter by length