***Idotea baltica* Assembly Completeness**

BUSCO v1.1

BLAST+ v2.2.31

HMMER v3.1b2

EMBOSS v6.5.7

PYTHON 3.4.0

**INSTALLING**

Change directory to USERAPPL and create new directory

cd $USERAPPL

mkdir busco

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

cd busco

wget http://busco.ezlab.org/files/BUSCO\_v1.1.tar.gz

Unzip the package

tar -zxvf BUSCO\_v1.1.tar.gz

Download and unzip the lineage Arthropod file

wget http://busco.ezlab.org/files/arthropoda\_buscos.tar.gz

tar -zxvf arthropoda\_buscos.tar.gz

Load programs

module load biokit

module load python-env/3.4.0

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

**RUNNING**

Run on raw assembly (115,931 contigs)

python3 $USERAPPL/busco/BUSCO\_v1.1.py -o raw \

-in trinity.K25.raw.fasta -l arthropoda -m trans -c 8

\*\*\* ~15min

**RESULTS**

Shown on screen (raw assembly):

Total complete BUSCOs found in assembly (<2 sigma) : 759 (667 duplicated).

Total BUSCOs partially recovered (>2 sigma) : 386

Total groups searched: 2675

Total BUSCOs not found: 863

Total running time: 922.7215542793274 seconds

Short summary file (raw assembly):

#Summarized BUSCO benchmarking for file: trinity.K25.raw.fasta

#BUSCO was run in mode: trans

Summarized benchmarks in BUSCO notation:

C:53%[D:24%],F:14%,M:32%,n:2675

Representing:

759 Complete Single-copy BUSCOs

667 Complete Duplicated BUSCOs

386 Fragmented BUSCOs

863 Missing BUSCOs

2675 Total BUSCO groups searched

\*\*\*

**Run again with ‘good’ contigs (40,122 contigs)**

python3 $USERAPPL/busco/BUSCO\_v1.1.py -o good \

-in trinity.K25.good.fasta -l arthropoda -m trans -c 8

\*\*\* ~15min

Shown on screen (‘good’ assembly):

Total complete BUSCOs found in assembly (<2 sigma) : 762 (563 duplicated).

Total BUSCOs partially recovered (>2 sigma) : 320

Total groups searched: 2675

Total BUSCOs not found: 1030

Total running time: 789.9900279045105 seconds

Short summary file (‘good’ assembly):

#Summarized BUSCO benchmarking for file: trinity.K25.good.fasta

#BUSCO was run in mode: trans

Summarized benchmarks in BUSCO notation:

C:49%[D:21%],F:11%,M:38%,n:2675

Representing:

762 Complete Single-copy BUSCOs

563 Complete Duplicated BUSCOs

320 Fragmented BUSCOs

1030 Missing BUSCOs

2675 Total BUSCO groups searched

**Run again with ‘good’+length filtered (26,569 contigs)**

python3 $USERAPPL/busco/BUSCO\_v1.1.py -o good\_L600 \

-in trinity.K25.good.L600.fasta -l arthropoda -m trans -c 8

\*\*\* ~15min

Shown on screen:

Total complete BUSCOs found in assembly (<2 sigma) : 744 (555 duplicated).

Total BUSCOs partially recovered (>2 sigma) : 304

Total groups searched: 2675

Total BUSCOs not found: 1072

Total running time: 719.1178307533264 seconds

Short summary file:

#Summarized BUSCO benchmarking for file: trinity.K25.good.L600.fasta

#BUSCO was run in mode: trans

Summarized benchmarks in BUSCO notation:

C:48%[D:20%],F:11%,M:40%,n:2675

Representing:

744 Complete Single-copy BUSCOs

555 Complete Duplicated BUSCOs

304 Fragmented BUSCOs

1072 Missing BUSCOs

2675 Total BUSCO groups searched

**Determine how many copies for duplicated BUSCOs**

python BUSCOdup.py full\_table\_busco.txt BUSCO\_dups.txt

244 duplicated BUSCOs have 3 copies

423 duplicated BUSCOs have 2 copies