***Idotea baltica* RNA-seq Preprocessing**

Trimmomatic 0.33

**RUNNING**

* Designed to work with Illumina sequences
* Adapter clipping is done first as identifying adapters from whole sequences is easier than if they were trimmed first
* Then trimming the ends based on quality since Illumina reads tend to be of lower quality on the ends

Create a batch jobs file, using example on CSC and command:

trimmomatic PE -threads 8 -phred33 \

2\_150911\_BC7F8NANXX\_P2211\_101\_1.fastq.gz \ 2\_150911\_BC7F8NANXX\_P2211\_101\_2.fastq.gz \

out\_fw\_paired.fq.gz out\_fw\_unpaired.fq.gz \

out\_rev\_paired.fq.gz out\_rev\_unpaired.fq.gz \ ILLUMINACLIP:/appl/bio/trimmomatic/adapters/TruSeq3-PE-2.fa:2:30:10 \ LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

This will perform the following (in that order):

* Remove adapters (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10)
* Remove leading low quality or N bases (below quality 3) (LEADING:3)
* Remove trailing low quality or N bases (below quality 3) (TRAILING:3)
* Scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15 (SLIDINGWINDOW:4:15)
* Drop reads below the 36 bases long (MINLEN:36)

Run batch jobs

sbatch trimmomatic.sh

squeue -l -u ketaya #check status

\*\*\* ~2.5 hours (8 CPUs per task, 8GB memory)

Stats located in file err\_6983454.txt

Input Read Pairs: 296,173,922

Both Surviving: 262,020,548 (88.47%)