

Steps:

1. Download elegans and morhua reads, reference genomes, and annotations
2. FastQC (and multi) for read quality
3. Trim with score 20 selection
4. Redo FastQC (and multi)
5. Create genome indices and align with Hisat2
6. Quantify with featureCounts (DGE, DTU, DEU) + transcript table

```
##### Step 1 #####
```

```
# C elegans reads - download all available reads
```

```
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/055/SRR11996355/SRR11996355.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/056/SRR11996356/SRR11996356.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/052/SRR11996352/SRR11996352.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/049/SRR11996349/SRR11996349.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/051/SRR11996351/SRR11996351.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/048/SRR11996348/SRR11996348.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/054/SRR11996354/SRR11996354.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/053/SRR11996353/SRR11996353.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR119/050/SRR11996350/SRR11996350.fastq.gz
```

```
# C elegans reference genome and annotations
```

```
wget  
https://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS19/species/caenorhabditis\_elegans/PRJNA13758/caenorhabditis\_elegans.PRJNA13758.WBPS19.genomic.fa.gz  
wget  
https://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS19/species/caenorhabditis\_elegans/PRJNA13758/caenorhabditis\_elegans.PRJNA13758.WBPS19.annotations.gff3.gz
```

```
# G morhua reads - download only BaP exposure reads (22 paired-end reads, 44 total files)  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/008/SRR6296808/SRR6296808_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/001/SRR6296821/SRR6296821_1.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/009/SRR6296809/SRR6296809_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/000/SRR6296810/SRR6296810_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/006/SRR6296816/SRR6296816_1.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296823/SRR6296823_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/004/SRR6296804/SRR6296804_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296822/SRR6296822_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/007/SRR6296817/SRR6296817_1.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/001/SRR6296821/SRR6296821_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296833/SRR6296833_1.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296823/SRR6296823_1.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296793/SRR6296793_2.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296822/SRR6296822_1.fastq.gz  
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296792/SRR6296792_2.fastq.gz
```

```
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/004/SRR6296804/SRR6296804_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/008/SRR6296828/SRR6296828_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/001/SRR6296791/SRR6296791_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296803/SRR6296803_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/009/SRR6296829/SRR6296829_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296802/SRR6296802_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/006/SRR6296816/SRR6296816_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296793/SRR6296793_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/008/SRR6296828/SRR6296828_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296792/SRR6296792_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/007/SRR6296817/SRR6296817_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/009/SRR6296829/SRR6296829_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296833/SRR6296833_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296832/SRR6296832_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/006/SRR6296796/SRR6296796_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/008/SRR6296808/SRR6296808_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/005/SRR6296815/SRR6296815_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/007/SRR6296797/SRR6296797_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/009/SRR6296809/SRR6296809_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/008/SRR6296798/SRR6296798_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/007/SRR6296797/SRR6296797_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/003/SRR6296803/SRR6296803_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296832/SRR6296832_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/005/SRR6296815/SRR6296815_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/008/SRR6296798/SRR6296798_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/002/SRR6296802/SRR6296802_2.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/001/SRR6296791/SRR6296791_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/006/SRR6296796/SRR6296796_1.fastq.gz
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR629/000/SRR6296810/SRR6296810_1.fastq.gz
```

```
# G morhua reference genome and annotations
wget
https://ftp.ensembl.org/pub/release-115/fasta/gadus_morhua/dna/Gadus_morhua.gadMor3.0.dn
a.toplevel.fa.gz
wget
https://ftp.ensembl.org/pub/release-115/gff3/gadus_morhua/Gadus_morhua.gadMor3.0.115.gff3.
gz
```

```
##### Step 2 #####

```

```
module load fastqc/0.12.1
module load multiqc/1.27.1
```

```
# C elegans
```

```

fastqc -c 8 -o ~/sharedscratch/reads/celegans_reads/
~/sharedscratch/reads/celegans_reads/*.fq.gz
multiqc -o ~/sharedscratch/reads/celegans_reads/trimmed_qc
~/sharedscratch/reads/celegans_reads/

# G morhua

fastqc -c 8 -o ~/sharedscratch/reads/gmorhua_reads/
~/sharedscratch/reads/gmorhua_reads/*.fq.gz
multiqc -o ~/sharedscratch/reads/gmorhua_reads/trimmed_qc
~/sharedscratch/reads/gmorhua_reads/

# examine the produced html files for read quality
# we found that these reads were of excellent quality and may not needed trimming
# but we performed trimming because it is good practice

##### Step 3 #####
# create a metadata file containing each SRR file ID (e.g. SRR11996348). For example, C elegans would have 1 column with 9 rows. G morhua would have 1 column with 22 rows (omit pairs).
# extract correct library name from metadata.txt file, using SLURM_ARRAY_TASK_ID variable

module load trimgalore/0.6.10
lib=$(cut -f 1 metadata.txt | head -n $SLURM_ARRAY_TASK_ID | tail -n 1)
# run trim_galore
mkdir -p trimmed
trim_galore -j 8 -q 20 --illumina -o trimmed ~/sharedscratch/reads/celegans_reads/${lib}.fastq.gz
# path must change according to names (as with the whole repository)

lib=$(cut -f 1 metadata_cod.txt | head -n $SLURM_ARRAY_TASK_ID | tail -n 1)
# run trim_galore
mkdir -p trimmed_cod
trim_galore -j 8 -q 20 --illumina --paired -o trimmed
~/sharedscratch/reads/gmorhua_reads/paired_reads_not_trimmed/${lib}_1.fastq.gz
~/sharedscratch/reads/gmorhua_reads/paired_reads_not_trimmed/${lib}_2.fastq.gz

##### Step 4 #####
# C elegans
fastqc -c 8 -o ~/sharedscratch/reads/celegans_reads/trimmed_qc
~/sharedscratch/reads/celegans_reads/trimmed_qc/*.fq.gz
multiqc -o ~/sharedscratch/reads/celegans_reads/trimmed_report
~/sharedscratch/reads/celegans_reads/trimmed_qc/

```

```

# G morhua

fastqc -c 8 -o ~/sharedscratch/reads/gmorhua_reads/trimmed_qc
~/sharedscratch/reads/gmorhua_reads/trimmed_qc/*.fq.gz
multiqc -o ~/sharedscratch/reads/gmorhua_reads/trimmed_report
~/sharedscratch/reads/gmorhua_reads/trimmed_qc

# view trimmed html files to inspect if quality has improved
# use trimmed_qc reads in next steps

##### Step 5 #####
module load samtools/1.19.2
module load hisat2/2.2.1 # building indexes (should be 8 index (.ht2) files for both)
hisat2-build Celegansgenome.fa ~/sharedscratch/references/Celegans/
hisat2-build Gmorhuagenome.fa ~/sharedscratch/references/Gmorhua/

lib=$(cut -f 1 metadata.txt | head -n $SLURM_ARRAY_TASK_ID | tail -n 1)
hisat2 -p 8 -x ~/sharedscratch/references/Celegans/genomeindex -U
~/sharedscratch/reads/celegans_reads/trimmed/${lib}_trimmed.fq.gz | samtools sort -@8 -O
bam -o ${lib}.bam

# output should be 1 .bam file and SLURM report per sample
# Note that option -U is specified because the data is unpaired -> if unsure for your data,
# perform unpaired and paired alignment and compare alignment rates

lib=$(cut -f 1 metadata_cod.txt | head -n $SLURM_ARRAY_TASK_ID | tail -n 1)
hisat2 -p 8 -x ~/sharedscratch/references/Cod/genomeindexCod -1
~/sharedscratch/reads/gmorhua_reads/trimmed/${lib}_1_val_1.fq.gz -2
~/sharedscratch/reads/gmorhua_reads/trimmed/${lib}_2_val_2.fq.gz | samtools sort -@8 -O
bam -o ${lib}v2.bam

# Note that we use -1 and -2 to specify each end of the paired end reads, otherwise the same
# (no -U)

##### Step 6 #####
# look at featureCounts -h to see options
# DGE first
module load subread/2.0.6

```

```

featureCounts -a ~/sharedscratch/references/Celegans/Celegansannotations.gff3 -t mRNA -g Parent -s 0 -T 8 -M --fraction -o ~/sharedscratch/references/Celegans/genecounts
~/sharedscratch/reads/celegans_reads/bam/SRR*

featureCounts -a ~/sharedscratch/references/Cod/Codannotations.gff3 -t mRNA -g Parent -s 2
-M --fraction -T 8 -p --countReadPairs -B -C -o
~/sharedscratch/references/Cod/genecount_morhua_test
~/sharedscratch/reads/gmorhua_reads/bam/SRR*

# DTU

featureCounts -a ~/sharedscratch/references/Celegans/Celegansannotations.gff3 -t mRNA -g ID
-s 0 -M --fraction -T 8 -o ~/sharedscratch/references/Celegans/dtuelegans
~/sharedscratch/reads/celegans_reads/bam/SRR*

featureCounts -a ~/sharedscratch/references/Cod/Codannotations.gff3 -t mRNA -g ID -s 2 -M
--fraction -T 8 -p --countReadPairs -B -C -o ~/sharedscratch/references/Cod/dtumorhua
~/sharedscratch/reads/gmorhua_reads/bam/SRR*

# later, it will be helpful to have a copy of all transcripts found in the organisms (from the
# reference genome), so we can create a table for transcripts matched with their genes

awk '$3=="mRNA"' ~/sharedscratch/references/Celegans/Celegansannotations.gff3 | cut -f 9 |
cut -d ';' -f 2,3 | tr ';' '\t' > transcript2genelegans.txt

awk '$3=="mRNA"' ~/sharedscratch/references/Cod/Codannotations.gff3 | cut -f 9 | cut -d ';' -f
2,3 | tr ';' '\t' > transcript2genemorhua.txt

# DEU
# We need to create reference gtfs that contain flattened exon bins first, but they're based on
# the same reference genomes -> we just need the gtf version instead
module load bedtools2

wget
https://ftp.ensembl.org/pub/release-115/gtf/caenorhabditis_elegans/Caenorhabditis_elegans.WBcel235.115.gtf.gz
gzip Caenorhabditis_elegans.WBcel235.115.gtf.gz

awk '$3=="exon"' Caenorhabditis_elegans.WBcel235.115.gtf | sed 's/;/ *//g' > exons_elegans.gtf

bedtools sort -i exons_elegans.gtf | bedtools merge -s -d 1 -c 7,9 -o last -i - -bed >
exons_elegans.flattened.txt

# add in required columns for GTF format

```

```

awk -F"\t' '{print $1, "WormBase", "exon", $2, $3, ".", $4, ".", $5}' OFS="\t"
exons_elegans.flattened.txt > exons_elegans.flattened.gtf
wc -l exons_elegans.flattened.gtf
153463 exons_elegans.flattened.gtf
# this is how many exon bins we have in our new gtf reference
# in case there are any empty rows
awk '($4 <= 0 || $5 <= 0 || $4 >= $5) {print NR, $0}' exons_elegans.flattened.gtf

# same for Cod
wget
https://ftp.ensembl.org/pub/release-115/gtf/gadus_morhua/Gadus_morhua.gadMor3.0.115.gtf.gz
gzip Gadus_morhua.gadMor3.0.115.gtf.gz

awk '$3=="exon"' Gadus_morhua.gadMor3.0.115.gtf | sed 's/.*//g' > exons_morhua.gtf
bedtools sort -i exons_morhua.gtf | bedtools merge -s -d 1 -c 7,9 -o last -i - -bed >
exons_morhua.flattened.txt

# add in required columns for GTF format

awk -F"\t' '{print $1, "Ensembl", "exon", $2, $3, ".", $4, ".", $5}' OFS="\t"
exons_morhua.flattened.txt > exons_morhua.flattened.gtf

##### now to quantify in featureCounts using these bins as annotations #####
featureCounts -a exons_elegans.flattened.gtf -T 8 -f -t exon -g gene_id -s 0 -M --fraction -O -o
exoncountsele_test.txt ~/sharedscratch/reads/celegans_reads/bam/SRR*

featureCounts -a exons_morhua.flattened.clean.gtf -T 8 -f exon -g gene_id -s 2 -M --fraction -T
8 -p --countReadPairs -B -C -O -o exoncountsморhua_test
~/sharedscratch/reads/gmorhua_reads/bam2/SRR*

```