**1A. Quality Control: FastQC and Trimmomatic**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=16

#SBATCH --mem-per-cpu=4G

#SBATCH --time=6:00:00

#SBATCH --output=fastqc.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="fastqc"

#SBATCH -p intel

module load fastqc

fastqc -t 16 -f fastq F1\_1\_R\_S145\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F1\_1\_R\_S145\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F1\_2\_R\_S148\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F1\_2\_R\_S148\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F1\_3\_R\_S151\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F1\_3\_R\_S151\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F1\_4\_R\_S154\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F1\_4\_R\_S154\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F1\_YA\_R\_S157\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F1\_YA\_R\_S157\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F2\_1\_R\_S146\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F2\_1\_R\_S146\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F2\_2\_R\_S149\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F2\_2\_R\_S149\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F2\_3\_R\_S152\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F2\_3\_R\_S152\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F2\_4\_R\_S155\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F2\_4\_R\_S155\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F2\_YA\_R\_S158\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F2\_YA\_R\_S158\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F3\_1\_R\_S147\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F3\_1\_R\_S147\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F3\_2\_R\_S150\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F3\_2\_R\_S150\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F3\_3\_R\_S153\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F3\_3\_R\_S153\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F3\_4\_R\_S156\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F3\_4\_R\_S156\_L007\_R2\_001.fastq.gz

fastqc -t 16 -f fastq F3\_YA\_R\_S159\_L007\_R1\_001.fastq.gz

fastqc -t 16 -f fastq F3\_YA\_R\_S159\_L007\_R2\_001.fastq.gz

**1B. Quality Control: Trimmomatic**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=25

#SBATCH --mem-per-cpu=4G

#SBATCH --time=6:00:00

#SBATCH --output=trimm2.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="TrimmTest"

#SBATCH -p intel

module load trimmomatic

trimmomatic PE -threads 25 -phred33 ./F1\_1\_R\_S145\_L007\_R1\_001.fastq.gz F1\_1\_R\_S145\_L007\_R2\_001.fastq.gz F1\_1\_R1\_paired.fastq.gz F1\_1\_R1\_unpaired.fastq.gz F1\_1\_R2\_paired.fastq.gz F1\_1\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F1\_2\_R\_S148\_L007\_R1\_001.fastq.gz F1\_2\_R\_S148\_L007\_R2\_001.fastq.gz F1\_2\_R1\_paired.fastq.gz F1\_2\_R1\_unpaired.fastq.gz F1\_2\_R2\_paired.fastq.gz F1\_2\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F1\_3\_R\_S151\_L007\_R1\_001.fastq.gz F1\_3\_R\_S151\_L007\_R2\_001.fastq.gz F1\_3\_R1\_paired.fastq.gz F1\_3\_R1\_unpaired.fastq.gz F1\_3\_R2\_paired.fastq.gz F1\_3\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F1\_4\_R\_S154\_L007\_R1\_001.fastq.gz F1\_4\_R\_S154\_L007\_R2\_001.fastq.gz F1\_4\_R1\_paired.fastq.gz F1\_4\_R1\_unpaired.fastq.gz F1\_4\_R2\_paired.fastq.gz F1\_4\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F1\_YA\_R\_S157\_L007\_R1\_001.fastq.gz F1\_YA\_R\_S157\_L007\_R2\_001.fastq.gz F1\_YA\_R1\_paired.fastq.gz F1\_YA\_R1\_unpaired.fastq.gz F1\_YA\_R2\_paired.fastq.gz F1\_YA\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F2\_1\_R\_S146\_L007\_R1\_001.fastq.gz F2\_1\_R\_S146\_L007\_R2\_001.fastq.gz F2\_1\_R1\_paired.fastq.gz F2\_1\_R1\_unpaired.fastq.gz F2\_1\_R2\_paired.fastq.gz F2\_1\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F2\_2\_R\_S149\_L007\_R1\_001.fastq.gz F2\_2\_R\_S149\_L007\_R2\_001.fastq.gz F2\_2\_R1\_paired.fastq.gz F2\_2\_R1\_unpaired.fastq.gz F2\_2\_R2\_paired.fastq.gz F2\_2\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F2\_3\_R\_S152\_L007\_R1\_001.fastq.gz F2\_3\_R\_S152\_L007\_R2\_001.fastq.gz F2\_3\_R1\_paired.fastq.gz F2\_3\_R1\_unpaired.fastq.gz F2\_3\_R2\_paired.fastq.gz F2\_3\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F2\_4\_R\_S155\_L007\_R1\_001.fastq.gz F2\_4\_R\_S155\_L007\_R2\_001.fastq.gz F2\_4\_R1\_paired.fastq.gz F2\_4\_R1\_unpaired.fastq.gz F2\_4\_R2\_paired.fastq.gz F2\_4\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F2\_YA\_R\_S158\_L007\_R1\_001.fastq.gz F2\_YA\_R\_S158\_L007\_R2\_001.fastq.gz F2\_YA\_R1\_paired.fastq.gz F2\_YA\_R1\_unpaired.fastq.gz F2\_YA\_R2\_paired.fastq.gz F2\_YA\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F3\_1\_R\_S147\_L007\_R1\_001.fastq.gz F3\_1\_R\_S147\_L007\_R2\_001.fastq.gz F3\_1\_R1\_paired.fastq.gz F3\_1\_R1\_unpaired.fastq.gz F3\_1\_R2\_paired.fastq.gz F3\_1\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F3\_2\_R\_S150\_L007\_R1\_001.fastq.gz F3\_2\_R\_S150\_L007\_R2\_001.fastq.gz F3\_2\_R1\_paired.fastq.gz F3\_2\_R1\_unpaired.fastq.gz F3\_2\_R2\_paired.fastq.gz F3\_2\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F3\_3\_R\_S153\_L007\_R1\_001.fastq.gz F3\_3\_R\_S153\_L007\_R2\_001.fastq.gz F3\_3\_R1\_paired.fastq.gz F3\_3\_R1\_unpaired.fastq.gz F3\_3\_R2\_paired.fastq.gz F3\_3\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F3\_4\_R\_S156\_L007\_R1\_001.fastq.gz F3\_4\_R\_S156\_L007\_R2\_001.fastq.gz F3\_4\_R1\_paired.fastq.gz F3\_4\_R1\_unpaired.fastq.gz F3\_4\_R2\_paired.fastq.gz F3\_4\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

trimmomatic PE -threads 25 -phred33 ./F3\_YA\_R\_S159\_L007\_R1\_001.fastq.gz F3\_YA\_R\_S159\_L007\_R2\_001.fastq.gz F3\_YA\_R1\_paired.fastq.gz F3\_YA\_R1\_unpaired.fastq.gz F3\_YA\_R2\_paired.fastq.gz F3\_YA\_R2\_unpaired.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

**2A. Build A. pisum Genome index on Hisat2**

module load hisat2

hisat2-build GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.fna Al4f\_genomic\_index

**2B. Build Buchnera Genome index on Hisat2**

module load hisat2

hisat2-build GCF\_000009605.1\_ASM960v1\_genomic.fna Buch\_genomic\_index

**3A. Hisat2 to A. pisum**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=16

#SBATCH --mem-per-cpu=4G

#SBATCH --time=12:00:00

#SBATCH --output=hisat2\_al4f.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="HISAT2\_Al4f"

#SBATCH -p intel

module load hisat2

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F1\_1\_R1\_paired.fastq.gz -2 ./F1\_1\_R2\_paired.fastq.gz -U ./F1\_1\_R1\_unpaired.fastq.gz,./F1\_1\_R2\_unpaired.fastq.gz -S F1\_1.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F1\_2\_R1\_paired.fastq.gz -2 ./F1\_2\_R2\_paired.fastq.gz -U ./F1\_2\_R1\_unpaired.fastq.gz,./F1\_2\_R2\_unpaired.fastq.gz -S F1\_2.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F1\_3\_R1\_paired.fastq.gz -2 ./F1\_3\_R2\_paired.fastq.gz -U ./F1\_3\_R1\_unpaired.fastq.gz,./F1\_3\_R2\_unpaired.fastq.gz -S F1\_3.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F1\_4\_R1\_paired.fastq.gz -2 ./F1\_4\_R2\_paired.fastq.gz -U ./F1\_4\_R1\_unpaired.fastq.gz,./F1\_4\_R2\_unpaired.fastq.gz -S F1\_4.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F1\_YA\_R1\_paired.fastq.gz -2 ./F1\_YA\_R2\_paired.fastq.gz -U ./F1\_YA\_R1\_unpaired.fastq.gz,./F1\_YA\_R2\_unpaired.fastq.gz -S F1\_YA.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F2\_1\_R1\_paired.fastq.gz -2 ./F2\_1\_R2\_paired.fastq.gz -U ./F2\_1\_R1\_unpaired.fastq.gz,./F2\_1\_R2\_unpaired.fastq.gz -S F2\_1.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F2\_2\_R1\_paired.fastq.gz -2 ./F2\_2\_R2\_paired.fastq.gz -U ./F2\_2\_R1\_unpaired.fastq.gz,./F2\_2\_R2\_unpaired.fastq.gz -S F2\_2.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F2\_3\_R1\_paired.fastq.gz -2 ./F2\_3\_R2\_paired.fastq.gz -U ./F2\_3\_R1\_unpaired.fastq.gz,./F2\_3\_R2\_unpaired.fastq.gz -S F2\_3.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F2\_4\_R1\_paired.fastq.gz -2 ./F2\_4\_R2\_paired.fastq.gz -U ./F2\_4\_R1\_unpaired.fastq.gz,./F2\_4\_R2\_unpaired.fastq.gz -S F2\_4.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F2\_YA\_R1\_paired.fastq.gz -2 ./F2\_YA\_R2\_paired.fastq.gz -U ./F2\_YA\_R1\_unpaired.fastq.gz,./F2\_YA\_R2\_unpaired.fastq.gz -S F2\_YA.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F3\_1\_R1\_paired.fastq.gz -2 ./F3\_1\_R2\_paired.fastq.gz -U ./F3\_1\_R1\_unpaired.fastq.gz,./F3\_1\_R2\_unpaired.fastq.gz -S F3\_1.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F3\_2\_R1\_paired.fastq.gz -2 ./F3\_2\_R2\_paired.fastq.gz -U ./F3\_2\_R1\_unpaired.fastq.gz,./F3\_2\_R2\_unpaired.fastq.gz -S F3\_2.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F3\_3\_R1\_paired.fastq.gz -2 ./F3\_3\_R2\_paired.fastq.gz -U ./F3\_3\_R1\_unpaired.fastq.gz,./F3\_3\_R2\_unpaired.fastq.gz -S F3\_3.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F3\_4\_R1\_paired.fastq.gz -2 ./F3\_4\_R2\_paired.fastq.gz -U ./F3\_4\_R1\_unpaired.fastq.gz,./F3\_4\_R2\_unpaired.fastq.gz -S F3\_4.sam

hisat2 -p 16 --dta -x ./Al4f\_genomic\_index -1 ./F3\_YA\_R1\_paired.fastq.gz -2 ./F3\_YA\_R2\_paired.fastq.gz -U ./F3\_YA\_R1\_unpaired.fastq.gz,./F3\_YA\_R2\_unpaired.fastq.gz -S F3\_YA.sam

**3B. Hisat to Buchnera**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=16

#SBATCH --mem-per-cpu=4G

#SBATCH --time=12:00:00

#SBATCH --output=hisat2buch.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="HISAT2buch"

#SBATCH -p intel

module load hisat2

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F1\_1\_R1\_paired.fastq.gz -2 ./F1\_1\_R2\_paired.fastq.gz -U ./F1\_1\_R1\_unpaired.fastq.gz,./F1\_1\_R2\_unpaired.fastq.gz -S F1\_1Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F1\_2\_R1\_paired.fastq.gz -2 ./F1\_2\_R2\_paired.fastq.gz -U ./F1\_2\_R1\_unpaired.fastq.gz,./F1\_2\_R2\_unpaired.fastq.gz -S F1\_2Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F1\_3\_R1\_paired.fastq.gz -2 ./F1\_3\_R2\_paired.fastq.gz -U ./F1\_3\_R1\_unpaired.fastq.gz,./F1\_3\_R2\_unpaired.fastq.gz -S F1\_3Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F1\_4\_R1\_paired.fastq.gz -2 ./F1\_4\_R2\_paired.fastq.gz -U ./F1\_4\_R1\_unpaired.fastq.gz,./F1\_4\_R2\_unpaired.fastq.gz -S F1\_4Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F1\_YA\_R1\_paired.fastq.gz -2 ./F1\_YA\_R2\_paired.fastq.gz -U ./F1\_YA\_R1\_unpaired.fastq.gz,./F1\_YA\_R2\_unpaired.fastq.gz -S F1\_YABuch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F2\_1\_R1\_paired.fastq.gz -2 ./F2\_1\_R2\_paired.fastq.gz -U ./F2\_1\_R1\_unpaired.fastq.gz,./F2\_1\_R2\_unpaired.fastq.gz -S F2\_1Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F2\_2\_R1\_paired.fastq.gz -2 ./F2\_2\_R2\_paired.fastq.gz -U ./F2\_2\_R1\_unpaired.fastq.gz,./F2\_2\_R2\_unpaired.fastq.gz -S F2\_2Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F2\_3\_R1\_paired.fastq.gz -2 ./F2\_3\_R2\_paired.fastq.gz -U ./F2\_3\_R1\_unpaired.fastq.gz,./F2\_3\_R2\_unpaired.fastq.gz -S F2\_3Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F2\_4\_R1\_paired.fastq.gz -2 ./F2\_4\_R2\_paired.fastq.gz -U ./F2\_4\_R1\_unpaired.fastq.gz,./F2\_4\_R2\_unpaired.fastq.gz -S F2\_4Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F2\_YA\_R1\_paired.fastq.gz -2 ./F2\_YA\_R2\_paired.fastq.gz -U ./F2\_YA\_R1\_unpaired.fastq.gz,./F2\_YA\_R2\_unpaired.fastq.gz -S F2\_YABuch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F3\_1\_R1\_paired.fastq.gz -2 ./F3\_1\_R2\_paired.fastq.gz -U ./F3\_1\_R1\_unpaired.fastq.gz,./F3\_1\_R2\_unpaired.fastq.gz -S F3\_1Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F3\_2\_R1\_paired.fastq.gz -2 ./F3\_2\_R2\_paired.fastq.gz -U ./F3\_2\_R1\_unpaired.fastq.gz,./F3\_2\_R2\_unpaired.fastq.gz -S F3\_2Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F3\_3\_R1\_paired.fastq.gz -2 ./F3\_3\_R2\_paired.fastq.gz -U ./F3\_3\_R1\_unpaired.fastq.gz,./F3\_3\_R2\_unpaired.fastq.gz -S F3\_3Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F3\_4\_R1\_paired.fastq.gz -2 ./F3\_4\_R2\_paired.fastq.gz -U ./F3\_4\_R1\_unpaired.fastq.gz,./F3\_4\_R2\_unpaired.fastq.gz -S F3\_4Buch.sam

hisat2 -p 16 --dta -x ./Buch\_genomic\_index -1 ./F3\_YA\_R1\_paired.fastq.gz -2 ./F3\_YA\_R2\_paired.fastq.gz -U ./F3\_YA\_R1\_unpaired.fastq.gz,./F3\_YA\_R2\_unpaired.fastq.gz -S F3\_YABuch.sam

**4. SamTools Sort**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=4

#SBATCH --mem-per-cpu=16G

#SBATCH --time=16:00:00

#SBATCH --output=samtoolsSort\_Al4f.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="SamtoolsSort\_Al4f"

#SBATCH -p intel

module load samtools

echo ""

echo "F1\_1"

samtools view -Su F1\_1.sam | samtools sort -o F1\_1\_sorted

echo ""

echo "F1\_2"

samtools view -Su F1\_2.sam | samtools sort -o F1\_2\_sorted

echo ""

echo "F1\_3"

samtools view -Su F1\_3.sam | samtools sort -o F1\_3\_sorted

echo ""

echo "F1\_4"

samtools view -Su F1\_4.sam | samtools sort -o F1\_4\_sorted

echo ""

echo "F1\_YA"

samtools view -Su F1\_YA.sam | samtools sort -o F1\_YA\_sorted

echo ""

echo "F2\_1"

samtools view -Su F2\_1.sam | samtools sort -o F2\_1\_sorted

echo ""

echo "F2\_2"

samtools view -Su F2\_2.sam | samtools sort -o F2\_2\_sorted

echo ""

echo "F2\_3"

samtools view -Su F2\_3.sam | samtools sort -o F2\_3\_sorted

echo ""

echo "F2\_4"

samtools view -Su F2\_4.sam | samtools sort -o F2\_4\_sorted

echo ""

echo "F2\_YA"

samtools view -Su F2\_YA.sam | samtools sort -o F2\_YA\_sorted

echo ""

echo "F3\_1"

samtools view -Su F3\_1.sam | samtools sort -o F3\_1\_sorted

echo ""

echo "F3\_2"

samtools view -Su F3\_2.sam | samtools sort -o F3\_2\_sorted

echo ""

echo "F3\_3"

samtools view -Su F3\_3.sam | samtools sort -o F3\_3\_sorted

echo ""

echo "F3\_4"

samtools view -Su F3\_4.sam | samtools sort -o F3\_4\_sorted

echo ""

echo "F3\_YA"

samtools view -Su F3\_YA.sam | samtools sort -o F3\_YA\_sorted

**5. StringTie**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=16

#SBATCH --mem-per-cpu=4G

#SBATCH --time=16:00:00

#SBATCH --output=Stringtie\_Al4f.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="Stringtie\_Al4f"

#SBATCH -p intel

module load stringtie

echo ""

echo "F1\_1"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F1\_1.gtf F1\_1\_sorted

echo ""

echo "F1\_2"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F1\_2.gtf F1\_2\_sorted

echo ""

echo "F1\_3"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F1\_3.gtf F1\_3\_sorted

echo ""

echo "F1\_4"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F1\_4.gtf F1\_4\_sorted

echo ""

echo "F1\_YA"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F1\_YA.gtf F1\_YA\_sorted

echo ""

echo "F2\_1"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F2\_1.gtf F2\_1\_sorted

echo ""

echo "F2\_2"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F2\_2.gtf F2\_2\_sorted

echo ""

echo "F2\_3"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F2\_3.gtf F2\_3\_sorted

echo ""

echo "F2\_4"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F2\_4.gtf F2\_4\_sorted

echo ""

echo "F2\_YA"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F2\_YA.gtf F2\_YA\_sorted

echo ""

echo "F3\_1"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F3\_1.gtf F3\_1\_sorted

echo ""

echo "F3\_2"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F3\_2.gtf F3\_2\_sorted

echo ""

echo "F3\_3"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F3\_3.gtf F3\_3\_sorted

echo ""

echo "F3\_4"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F3\_4.gtf F3\_4\_sorted

echo ""

echo "F3\_YA"

stringtie -p 16 -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -o F3\_YA.gtf F3\_YA\_sorted

**6. Create StringTie Merge txt file**

F1\_1.gtf

F1\_2.gtf

F1\_3.gtf

F1\_4.gtf

F1\_YA.gtf

F2\_1.gtf

F2\_2.gtf

F2\_3.gtf

F2\_4.gtf

F2\_YA.gtf

F3\_1.gtf

F3\_2.gtf

F3\_4.gtf

F3\_YA.gtf

**7. Run StringTie Merge**

stringtie --merge -G GCF\_005508785.1\_pea\_aphid\_22Mar2018\_4r6ur\_genomic.gff -p 16 -o F#\_#\_merged.gtf F#\_#\_gtf\_list.txt

**8. Create Ballgown Folder**

mkdir ballgown

cd ballgown/

mkdir F1\_1

mkdir F1\_2

mkdir F1\_3

mkdir F1\_4

mkdir F1\_YA

mkdir F2\_YA

mkdir F2\_4

mkdir F2\_3

mkdir F2\_2

mkdir F2\_1

mkdir F3\_1

mkdir F3\_2

mkdir F3\_3

mkdir F3\_4

mkdir F3\_YA

**9. StringTie Abundance**

#!/bin/bash -l

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=8

#SBATCH --mem-per-cpu=8G

#SBATCH --time=16:00:00

#SBATCH --output=StringtieAbundance\_Al4f.stdout

#SBATCH --mail-type=ALL

#SBATCH --job-name="StringtieAbundance\_Al4f"

#SBATCH -p intel

module load stringtie

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F1\_1/F1\_1.gtf ./F1\_1\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F1\_2/F1\_2.gtf ./F1\_2\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F1\_3/F1\_3.gtf ./F1\_3\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F1\_4/F1\_4.gtf ./F1\_4\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F1\_YA/F1\_YA.gtf ./F1\_YA\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F2\_1/F2\_1.gtf ./F2\_1\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F2\_2/F2\_2.gtf ./F2\_2\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F2\_3/F2\_3.gtf ./F2\_3\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F2\_4/F2\_4.gtf ./F2\_4\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F2\_YA/F2\_YA.gtf ./F2\_YA\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F3\_1/F3\_1.gtf ./F3\_1\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F3\_2/F3\_2.gtf ./F3\_2\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F3\_3/F3\_3.gtf ./F3\_3\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F3\_4/F3\_4.gtf ./F3\_4\_sorted

stringtie -e -B -p 8 -G ./F#\_#\_merged.gtf -o ./ballgown/F3\_YA/F3\_YA.gtf ./F3\_YA\_sorted

**10. Download and run prepDE.py script**

<http://ccb.jhu.edu/software/stringtie/index.shtml?t=manual>

python prepDE.py

**11. Create "sampleList.txt”**

sample treatment batch

F1\_1 1 1

F1\_2 2 1

F1\_3 3 1

F1\_4 4 1

F1\_YA YA 1

F2\_1 1 2

F2\_2 2 2

F2\_3 3 2

F2\_4 4 2

F2\_YA YA 2

F3\_1 1 3

F3\_2 2 3

F3\_3 3 3

F3\_4 4 3

F3\_YA YA 3

**12. Run DEBrowser in R Studio**

Load: > library(debrowser)

Start: > startDEBrowser()

Data Prep:

Count Data File: gene\_count\_matrix.csv

Metadata File: sampleList.txt

Upload Summary

Filter>Low Count Filtering>Max 10

Batch Effect Corrections and Normalization>TMM/none

Download csv (use this in GSEA)

Go to DE Analysis

Conditon 1: F1\_1, F2\_1, F3\_1

Conditoin 2: F1\_2, F2\_2, F3\_2

DE Method: EdgeR

Normalization: TMM

Start DE: padj = 0.01; foldchange = 2

Test = exactTest

Download Figs and Table

Repeat DE Browser for:

* 1. 1v2, 2v3, 3v4, 4vYA
  2. 1 v rest, 2 v rest, 3 v rest, 4 v rest, YA v rest
  3. Gene\_count & transcript\_count

**13. Gene Set Enrichment Analysis**

To convert files to run GSEA:

1. Go to <http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html>
2. Download: javaGSEA/DesktopApplication/2GB (for 64-bit Java only)
   1. May need to update java
3. Go to Documentation 🡪 user guide
4. Need 2 files:
   1. Expression dataset (this will be a modified version of gene\_count\_matrix.xls)
      1. Note used normalized version of it downloaded from DEBrowser
         1. Not raw read counts
      2. <http://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats> (how to format from xls to gct)
      3. Switch Column A and B so LOC in A and MSTRG in B
         1. Change headers to NAME, Description
      4. Add two rows to top, in 1: #1.2; in 2: number of rows of data (-3 top row)
         1. Row 2, col B: # of samples (ie: 15)
      5. Error to correct: any novel gene (N/A in column A) needs to be deleted
      6. Save as tab delimited text, and with .gct ending
         1. “AL4f\_norm.gct” 🡪 RNAseqCluster folder
   2. Gene set
      1. “apiGeneList.gmt” download to Desktop 🡪 RNAseqCluster folder

To Run GSEA:

1. Open GSEA
2. Load Data: select: 1-“.gct” & 4-“.gmt” & 1-“.cls”
3. Run GSEA: Set Parameters
   1. Expression datset: Al4f\_norm.gct
   2. Gene sets database: Any of the 4 .gmt
   3. # perms: 1000
   4. Phenotype Labels: cls 🡪 1st vs 2nd
      1. Has to be pairwise, but can be multiple files
   5. Collapse data set to gene symbols: False
   6. Permutation type: gene\_set
   7. Analysis name: unique per run
   8. Leave next few alone
   9. Min size: 2
   10. Save results: make dir inside my data GSEA (need to change from default)
   11. Leave rest as default
4. Run (repeat add. comps: 1v2, 2v3, 3v4, 4vYA, 1vRest. 2vRest, 3vRest, 4vRest, YAvRest)
5. Analyze Results:
   1. Click on Success (or go to Results folder and open index)
   2. Goes to html w/ report X /123 enriched in A, Y/123 enriched in B
      1. Detailed [enrichment results in html](about:blank) format
      2. Goes to html with table of upreg paths, can click on Details
      3. Html to specific names of genes in that path that are up reg
      4. Can copy gene set name and search in KEGG Pathway
      5. <https://www.genome.jp/kegg/pathway.html>
         1. Prefix: api Enter: gene set 🡪 Go