Lab Notebook

October 1st Thursday (Group): Project ideas

Our primary focus is mainly on health and diseases. Since Nina is on pre-med track, she wants to do something related to her med career and is very interested in cancer. She found a dataset called the Cancer Genome Atlas Program. The files that have the sequence reads are controlled, which means that we don't have the permission to access. We guess that every dataset that includes patients' specific DNAs or RNAs are closed. https://portal.gdc.cancer.gov/repository
But if we select sequence reads, all will be controlled and need PIV to access. One question she had was: how cancer therapy targest specific protein, and how we can prevent from having cancer if we know there is an oncogene.

October 2nd Friday (Group): Project ideas

Since cancer dataset not available, we discussed something new:
Yijin is interested in diabetes. She knows that there are two types of diabetes and is interested in Type I Diabetes. She also wants to explore the gender difference. She found two articles: one tells a genetic mouse model for metastatic lung cancer with gender differences in survival (https://www.nature.com/articles/1210493); the other is a pan-cancer analysis (https://www.nature.com/articles/s41586-020-1969-6)
Qizhe is interested in how environment can possibly change the expression of certain gene. We can compare the genome and transcriptome in the dataset of drosophila from different states in the America (from Florida all the way to Maine).

October 6th Tuesday (Group): Project ideas

We decided that we would switch back to drosophila and the topics we are interested in are written on our group padlet: https://rochester.padlet.org/amandalarracuente1/vn9qrh946neaa55k
Two of those topics are circadian rhythms and long non-coding RNA(lncRNA) on cancer. We also would like to see the difference among different species of Drosophila. Can we tell a specie just by the expression of some certain genes? For example, D. bifurca is known for its longest sperm cell, what are the genes that control this phenomenon? We were also able to share our individual strengths. All of us are good at looking for relevant articles and seeking information in a timely fashion.

October 7th Wednesday (Group): 12:30 - 1:00 pm Zoom Meeting with Xiaomi

Xiaomi gave us an example of alcohol tolerance of Drosophila. We are now interested in how alcohol affects Drosophila.

October 9th Friday (Nina Hu): 12:15 - 1:00 pm Zoom Meeting with Dr. Larracuente

Links shared by Professor:

https://www.ncbi.nlm.nih.gov/sra
http://dgrp2.gnets.ncsu.edu/data.html
https://pubmed.ncbi.nlm.nih.gov/26503115/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479727/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5089932/

October 13th Tuesday (Yijin Zhao): Module 10: Library Module 1

October 15th Thursday (Qizhe Zhang): Module 10: Lecture videos & quizzes

October 16th Friday (Nina Hu): Module 10: Lab worksheet

October 20th Tuesday (Qizhe Zhang): Group proposal outline

Questions we have before meeting tomorrow (Nina Hu):

1. What do you think about the scope of our project? We used the example proposal as a guideline to draft this proposal outline. You mentioned on Tuesday that this group actually changed their project later on. Can we ask why they did that?
2. You mentioned that we wouldn't be able to do a formal differential expression analysis. So we wonder whether this descriptive one will still be meaningful to do for our project?
3. Can we ask the advantage of using a splice-aware aligner compared to Bowtie2? None of us had experience or even heard of these two packages before. I guess we probably need to discuss it in the independent proposal?
4. For Objective 2, do you think comparing the different isoforms that we found in objective 1 would work or the current one is fine? What's the difference between the transcript variant we found on NCBI nucleotide and the different isoforms in RNAseq data? We're thinking that more isoforms are included in the RNAseq data and the transcript variants we found on NCBI Nucleotide are just some of them confirmed by papers?
5. For Objective 3, we didn't find RNAseq data for other species. I guess we might have done something wrong when searching on SRA. Could you please show us the steps to search the datasets again? Really sorry about this…
6. For the species we'd like to choose in Objective 3, we want to do species that are closely related to D. melanogaster. I even wonder whether we can draw a phylogenetic tree based on that. Maybe I'm thinking it in the wrong way.
7. We thought BLAST is used for comparing. But we're unsure whether BLAST can be used to compare the RNAseq expression data among different species.

8. Regarding the RNAseq datasets, we chose “L3 salivary glands” and “L3 digestive system” for the dataset in Objective 1 (links in the project outline), are these adults or larvae? What does L3 mean here? Can we ask what a normal RNAseq dataset looks like? For the link you shared with us on Tuesday, the tissues that we can find for D. melanogaster include “mixed digestive system”, “heads”, “female ovaries”, “male testes”, “accessory glands”, and “imaginal discs”, we wonder whether you have any recommendations/suggestions on choosing datasets/tissues? You also mentioned “carcass” in the comments...
9. For the independent proposal due next Tuesday, the example proposal talks about expected results, analysis, and potential pitfalls. Did the author come out with those by his/her understanding of the proposal outline and the literature cited? We feel that we’re not that knowledgeable to be this specific…

October 22nd Thursday (Yijin Zhao): Journal Club Paper 1 & Module 11: PubMed activity

Feedback from Professor on Slack (Nina Hu):

Just some feedback that I can provide quickly now and we can discuss more in our meeting: 1. Scope is good. Just like in research you’d do in a lab, sometimes you will shift your approach as you start to do analyses (e.g. switch datasets because a dataset wasn’t good, read a paper that gave you a better idea of a contrast to make). That is totally ok if it happens. 2) You can still do a differential expression analysis but likely not what we do in class because it is not appropriate for all comparisons. More on this later. 3) Splice aware means that the aligner knows that some of your reads might overlap exon-exon junctions in the mRNA that don’t exist in the genome because an intron is there. About splice-aware aligners, in Module 9, check out slides 17-29 (and in video 2). In class we map to a transcriptome in the lab with kallisto. For projects many people will also need to learn tools that we didn’t have a lab for. We can help you with that. Check out the documentation in the meantime though. 4) I’m not totally sure what you are asking. All predicted transcripts for genes that have multiple isoforms will have a confidence level associated with the type of evidence supporting its existence. It’s easier to navigate this information in Flybase than on ncbi. Go to flybase and search for Adh. 5) I will help you with this. We’ll go to NCBI SRA, enter search term, narrow search to RNA on left side of screen, and select species on the right side of the screen. Unfortunately, I can’t record breakout rooms but can record the main room. 6) Yes, you can draw a phylogeny. Drosophila simulans is a human commensal species like Drosophila melanogaster and I imagine is also parasitized by wasps. I don’t think that D. yakuba is but as much as we use that species I actually don’t know its ecology (we can look it up). I wanted to know the motivation for your species choice and if it was guided by literature. 7. BLAST can be used for searching for sequence homology. You already have Adh sequences annotated for those species accessible through flybase and ncbi. 8) L3 means 3rd larval instar. This is the stage where larvae crawl around eating voraciously before they pupate. Check out the Drosophila primer on blackboard and in the project instructions for more information. I’m not sure what you mean by normal RNAseq dataset so let’s talk about that in person. Let your tissue choice be guided by literature. We can talk about that too, but I’d need a little more time to do some research. Carcass is the remainder of the tissue after things have been dissected. So if someone dissects heads, the carcass is the rest of the body. You have to look at the primary research article methods to know what to expect in carcass. 9) No worries! I’m not expecting you to be experts or very knowledgable on the topic. Yes that group read a bunch of papers before writing that proposal. But what I’m looking for is what you expect to find. You wrote down objectives/goals and as you did this, you likely thought about what you might find. So if you learned, for example, that a specific isoform is related to a phenotype you are interested in based on literature and you chose species that have and do not have that phenotype, then you might make a prediction about the relative expression of that isoform in species with the phenotype.

October 23rd Friday (Qizhe Zhang): Module 11: Lectures and quizzes

Meeting notes for group (Nina Hu)

When we extract mRNAs, we reverse transcribe them into cDNAs

We have reads that exons join together
Reads that mapped to the exon-exon junction won’t be able to map well because there is an intron there
So we use splice-aware aligners for mapping mRNAs back to the reference genome.
Splice-aware aligners know to take reads that don’t map to the genome and have one read map to different places, because they overlap exon-exon junctions
Bowtie2 by itself, map reads to the transcriptome
Look at different isoforms, different transcripts that encode different protein isoforms.
It’s easier to use Hisat2, takes the read, map it to the genome, allows the exon-exon junction in your read so it splits them up
HTseq count, takes bam file that you get from Hisat2, it will take in a gff file which lists all of the features in the genome
In the gff file, it will say the transcript ID, exon 1 coordinates, exon 2 coordinate
It summarizes the counts by feature, summarizing all of the counts for a gene, it will take all of things that are the parts for that gene, and output a count summary (total count)
Summarize by exons, different transcript by different exons
The variants found on NCBI nucleotides can be isoforms
FBgn0000055 is the gene ID, the unique identifier when you use the gff file, you will refer to this number
Flybase search adh, go to the first one and the summary section has info about the fly
The yellow(orange) are the transcripts, the purple are the protein isoforms
Grey areas are UTRs (transcribed but not translated)
Purple isoforms match the yellow
Isoforms of A
Dataset keep L3
Isoforms in different tissues paper
See expression data, tissues

October 27th Tuesday (Nina Hu): Independent proposal
October 29th Thursday (Nina Hu): Journal club Paper 2
November 3rd Tuesday (Yijin Zhao): Project presentation preparation: text on the slides
November 5th Thursday (Qizhe Zhang): Project presentation: picture on the slides
November 6th Friday (Group): Group presentation
November 10th Tuesday (Nina Hu): Midterm: question 1-4
November 12th Thursday (Qizhe Zhang): Journal club Paper 3
November 13th Friday (Yijin Zhao): Module 11 Lab challenge
November 17th Tuesday (Qizhe Zhang): Midterm: question 5-7
November 19th Thursday (Group): Journal club Paper 4
November 20th Friday (None present): Absent due meeting with PI
After class: download SRRs, samtools, igv, hisat2 on local machine (Group)

Manual of Hisat2

http://daehwankimlab.github.io/hisat2/manual/
http://gensoft.pasteur.fr/docs/hisat2/2.0.1-beta/manual.inc.html
https://chipster.csc.fi/manual/hisat2-paired-end.html

Two worksheets we found that we thought were useful:

http://dna.colorado.edu/ShortRead/2019/6_RNA-seq/6_worksheet_6.1_HISAT2.pdf
https://wikis.utexas.edu/display/bioiteam/Mapping+with+HISAT2

A blog article from Medium website that we thought explained well:

https://medium.com/@shilparaopradeep/guide-to-hisat2-for-rna-seq-reads-alignment-against-human-population-81428813ce4b

We also found something interesting:

https://bioconductor.org/packages/devel/bioc/vignettes/Rhisat2/inst/doc/Rhisat2.html

We tried this as a group:

https://decodebiology.github.io/bioinfotutorials/rnaseq_tutorial.html

November 24th Tuesday (Group)

Read the manual for HTseq count, DEseq2, and download GFF files

December 1st Tuesday

Re-Download SRRs on Bluehive (Yijin Zhao)

SRR070408

slurm file: download_sra.sh

```
#!/bin/bash
#SBATCH -J download
#SBATCH -e download.err
#SBATCH -o download.out
#SBATCH --mem-per-cpu=8000
#SBATCH -t 24:00:00
#SBATCH -n 4
#SBATCH -p standard
##This is an example slurm script for moving read files to Bluehive. You will need to comment out lines you do not need
# for example, if you have paired end reads, comment out single end read comman. You will also need to replace read IDs
# and paths to directories
module load sratoolkit
dir_path="/scratch/bio257_2020/Users/group2_adh"
#replace group_name with your own
samplename="SRR070408"
#replace with your unique identifier for the reads
#if paired end reads
fastq-dump --split-files --gzip --outdir "$dir_path" "$samplename"
#if single end reads
#fastq-dump --gzip --outdir "$dir_path" "$samplename"
#delete the temp file generated in your home directory
```

```
rm /home/whu12/ncbi/public/sra/SRR*
#replace your_netid with your own
```

command: sbatch download_sra.sh

download.err:

```
rm: cannot remove '/home/whu12/ncbi/public/sra/SRR*': No such file or directory
```

download.out:

```
Read 29852257 spots for SRR070425
Written 29852257 spots for SRR070425
```

files:

SRR070408_1.fastq.gz
SRR070408_2.fastq.gz

SRR070425

slurm file: d425

```
#!/bin/bash
#SBATCH -J download
#SBATCH -e d425.err
#SBATCH -o d425.out
#SBATCH --mem-per-cpu=8000
#SBATCH -t 24:00:00
#SBATCH -n 4
#SBATCH -p standard
##This is an example slurm script for moving read files to Bluehive. You will need to comment out lines y
# for example, if you have paired end reads, comment out single end read comman. You will also need to re and paths to directories
module load sratoolkit
dir_path="/scratch/bio257_2020/Users/group2_adh"
#replace group_name with your own
samplename="SRR070425"
#replace with your unique identifier for the reads
#if paired end reads
fastq-dump --split-files --gzip --outdir "$dir_path" "$samplename"
#if single end reads
#fastq-dump --gzip --outdir "$dir_path" "$samplename"
#delete the temp file generated in your home directory
rm /home/whu12/ncbi/public/sra/SRR*
#replace your_netid with your own
```

command: sbatch d425

d425.err:

```
rm: cannot remove '/home/whu12/ncbi/public/sra/SRR*': No such file or directory
```

d425.out:

```
Read 29852257 spots for SRR070425
Written 29852257 spots for SRR070425
```

files:

SRR040425_1.fastq.gz
SRR040425_2.fastq.gz

SRR6161775 (D. willistoni)

slurm file: d775

```
#!/bin/bash
#SBATCH -J download
#SBATCH -e d775.err
#SBATCH -o d775.out
#SBATCH --mem-per-cpu=8000
#SBATCH -t 24:00:00
#SBATCH -n 4
#SBATCH -p standard
```

##This is an example slurm script for moving read files to Bluehive. You will need to comment out lines y
for example, if you have paired end reads, comment out single end read comman. You will also need to re and paths to directories
and paths to directories
module load sratoolkit
dir_path="/scratch/bio257_2020/Users/group2_adh"
#replace group_name with your own
samplename="SRR6161775"
#replace with your unique identifier for the reads
#if paired end reads
fastq-dump --split-files --gzip --outdir "$dir_path" "$samplename"
#if single end reads
#fastq-dump --gzip --outdir "$dir_path" "$samplename"
#delete the temp file generated in your home directory
rm /home/whu12/ncbi/public/sra/SRR*
#replace your_netid with your own

command: sbatch d775

d775.err:

rm: cannot remove '/home/whu12/ncbi/public/sra/SRR*': No such file or directory

d775.out:

Read 27550939 spots for SRR6161775
Written 27550939 spots for SRR6161775

Files:

SRR6161775_1.fastq.gz
SRR6161775_2.fastq.gz

SRR6177779 (D. sechellia)

slurm file: d779

#!/bin/bash
#SBATCH -J download
#SBATCH -e d779.err
#SBATCH -o d779.out
#SBATCH --mem-per-cpu=8000
#SBATCH -t 24:00:00
#SBATCH -n 4
#SBATCH -p standard
##This is an example slurm script for moving read files to Bluehive. You will need to comment out lines y
for example, if you have paired end reads, comment out single end read comman. You will also need to r
and paths to directories
module load sratoolkit
dir_path="/scratch/bio257_2020/Users/group2_adh"
#replace group_name with your own
samplename="SRR6161779"
#replace with your unique identifier for the reads
#if paired end reads
fastq-dump --split-files --gzip --outdir "$dir_path" "$samplename"
#if single end reads
#fastq-dump --gzip --outdir "$dir_path" "$samplename"
#delete the temp file generated in your home directory
rm /home/whu12/ncbi/public/sra/SRR*
#replace your_netid with your own

command: sbatch d779

d779.err:

rm: cannot remove '/home/whu12/ncbi/public/sra/SRR*': No such file or directory

d779.out:

Read 21546665 spots for SRR6161779

Written 21546665 spots for SRR6161779

SRR6161782

slurm file: d782

```
#!/bin/bash
#SBATCH -J download
#SBATCH -e d782.err
#SBATCH -o d782.out
#SBATCH --mem-per-cpu=8000
#SBATCH -t 24:00:00
#SBATCH -n 4
#SBATCH -p standard
##This is an example slurm script for moving read files to Bluehive. You will need to comment out
lines y$
# for example, if you have paired end reads, comment out single end read comman. You will also
need to re$
# and paths to directories
module load sratoolkit
dir_path="/scratch/bio257_2020/Users/group2_adh"
#replace group_name with your own
samplename="SRR6161782"
#replace with your unique identifier for the reads
#if paired end reads
fastq-dump --split-files --gzip --outdir "$dir_path" "$samplename"
#if single end reads
#fastq-dump --gzip --outdir "$dir_path" "$samplename"
#delete the temp file generated in your home directory
rm /home/whu12/ncbi/public/sra/SRR*
#replace your_netid with your own
```

command: sbatch d782

d782.err:

```
rm: cannot remove '/home/whu12/ncbi/public/sra/SRR*': No such file or directory
```

d782.out:

```
Read 26508281 spots for SRR6161782
Written 26508281 spots for SRR6161782
```

files:

SRR6161782_1.fastq.gz
SRR6161782_2.fastq.gz

Download reference genomes and GFF files for the three Drosophila Species (Qizhe Zhang)

D. melanogaster:
/scratch/bio257_2020/Example_data/reference.genome/dmel-all-chromosome-r6.13.fasta
GFF: https://www.ncbi.nlm.nih.gov/genome/?term=drosophila+melanogaster

Command:

```
scp dmel.gff whu12@bluehive.circ.rochester.edu:/scratch/bio257_2020/Users/group2_adh
```

D. sechellia: https://www.ncbi.nlm.nih.gov/genome/?term=drosophila%20sechellia

command:

```
tar -cf s.tar sechellia
scp s.tar whu12@bluehive.circ.rochester.edu:/scratch/bio257_2020/Users/group2_adh
cd /scratch/bio257_2020/Users/group2_adh
tar -xf s.tar
```

folder: sechellia

files:

GCF_000005925.1_dwil_caf1_genomic.fna
GCF_000005925.1_dwil_caf1_genomic.gff

D. willistoni: https://www.ncbi.nlm.nih.gov/genome/?term=drosophila+willistoni

Command:

tar -cf w.tar willistoni
scp w.tar whu12@bluehive.circ.rochester.edu:/scratch/bio257_2020/Users/group2_adh
cd /scratch/bio257_2020/Users/group2_adh
tar -xf w.tar

folder: willistoni

files:

GCF_000005925.1_dwil_caf1_genomic.fna
GCF_000005925.1_dwil_caf1_genomic.gff

Indexed D. melanogaster genome (Nina Hu):

command: hisat2-build -p 24 dmel-all-chromosome-r6.13.fasta dm_indexed.fa

```
Settings:
  Output files: "dm_indexed.fa.*.ht2"
  Line rate: 6 (line is 64 bytes)
  Lines per side: 1 (side is 64 bytes)
  Offset rate: 4 (one in 16)
  FTable chars: 10
  Strings: unpacked
  Local offset rate: 3 (one in 8)
  Local fTable chars: 6
  Local sequence length: 57344
  Local sequence overlap between two consecutive indexes: 1024
  Endianness: little
  Actual local endianness: little
  Sanity checking: disabled
  Assertions: disabled
  Random seed: 0
  Sizeofs: void*:8, int:4, long:8, size_t:8
Input files DNA, FASTA:
  dmel-all-chromosome-r6.13.fasta
Reading reference sizes
  Time reading reference sizes: 00:00:02
Calculating joined length
Writing header
Reserving space for joined string
Joining reference sequences
  Time to join reference sequences: 00:00:02
  Time to read SNPs and splice sites: 00:00:00
Using parameters --bmax 26732442 --dcv 1024
  Doing ahead-of-time memory usage test
  Passed!  Constructing with these parameters: --bmax 26732442 --dcv 1024
Constructing suffix-array element generator
Building DifferenceCoverSample
  Building sPrime
  Building sPrimeOrder
  V-Sorting samples
  V-Sorting samples time: 00:00:12
  Allocating rank array
  Ranking v-sort output
  Ranking v-sort output time: 00:00:02
  Invoking Larsson-Sadakane on ranks
  Invoking Larsson-Sadakane on ranks time: 00:00:02
```

  Sanity-checking and returning
Building samples
Reserving space for 12 sample suffixes
Generating random suffixes
QSorting 12 sample offsets, eliminating duplicates
QSorting sample offsets, eliminating duplicates time: 00:00:00
Multikey QSorting 12 samples
  (Using difference cover)
  Multikey QSorting samples time: 00:00:00
Calculating bucket sizes
Splitting and merging
  Splitting and merging time: 00:00:00
Split 1, merged 4; iterating...
Splitting and merging
  Splitting and merging time: 00:00:00
Avg bucket size: 1.58414e+07 (target: 26732441)
Converting suffix-array elements to index image
Allocating ftab, absorbFtab
Entering GFM loop
Getting block 1 of 9
Getting block 2 of 9
Getting block 3 of 9
  Reserving size (26732442) for bucket 2
  Reserving size (26732442) for bucket 1
Getting block 4 of 9
  Reserving size (26732442) for bucket 3
  Reserving size (26732442) for bucket 4
Getting block 5 of 9
  Calculating Z arrays for bucket 4
  Calculating Z arrays for bucket 2
  Calculating Z arrays for bucket 1
Getting block 8 of 9
  Reserving size (26732442) for bucket 8
Getting block 9 of 9
Getting block 6 of 9
  Entering block accumulator loop for bucket 4:
  Reserving size (26732442) for bucket 6
Getting block 7 of 9
  Entering block accumulator loop for bucket 1:
  Calculating Z arrays for bucket 8
  Entering block accumulator loop for bucket 2:
  Calculating Z arrays for bucket 3
  Reserving size (26732442) for bucket 5
  Reserving size (26732442) for bucket 7
  Calculating Z arrays for bucket 6
  Reserving size (26732442) for bucket 9
  Entering block accumulator loop for bucket 8:
  Entering block accumulator loop for bucket 6:
  Calculating Z arrays for bucket 5
  Calculating Z arrays for bucket 9
  Calculating Z arrays for bucket 7
  Entering block accumulator loop for bucket 3:
  Entering block accumulator loop for bucket 5:
  Entering block accumulator loop for bucket 9:
  Entering block accumulator loop for bucket 7:
  bucket 1: 10%
  bucket 9: 10%
  bucket 8: 10%
  bucket 6: 10%
  bucket 7: 10%
  bucket 3: 10%

bucket 1: 20%
bucket 5: 10%
bucket 2: 10%
bucket 9: 20%
bucket 4: 10%
bucket 1: 30%
bucket 6: 20%
bucket 9: 30%
bucket 8: 20%
bucket 3: 20%
bucket 1: 40%
bucket 7: 20%
bucket 5: 20%
bucket 2: 20%
bucket 9: 40%
bucket 1: 50%
bucket 6: 30%
bucket 8: 30%
bucket 3: 30%
bucket 4: 20%
bucket 9: 50%
bucket 1: 60%
bucket 2: 30%
bucket 7: 30%
bucket 8: 40%
bucket 6: 40%
bucket 9: 60%
bucket 5: 30%
bucket 1: 70%
bucket 3: 40%
bucket 7: 40%
bucket 4: 30%
bucket 9: 70%
bucket 2: 40%
bucket 1: 80%
bucket 8: 50%
bucket 3: 50%
bucket 5: 40%
bucket 1: 90%
bucket 6: 50%
bucket 9: 80%
bucket 7: 50%
bucket 8: 60%
bucket 2: 50%
bucket 1: 100%
Sorting block of length 9586584 for bucket 1
(Using difference cover)
bucket 3: 60%
bucket 5: 50%
bucket 9: 90%
bucket 6: 60%
bucket 4: 40%
bucket 7: 60%
bucket 8: 70%
bucket 2: 60%
bucket 3: 70%
bucket 9: 100%
Sorting block of length 14679497 for bucket 9
(Using difference cover)
bucket 6: 70%

bucket 5: 60%
bucket 8: 80%
bucket 2: 70%
bucket 3: 80%
bucket 7: 70%
bucket 4: 50%
bucket 2: 80%
bucket 8: 90%
bucket 3: 90%
bucket 6: 80%
bucket 5: 70%
bucket 4: 60%
bucket 8: 100%
Sorting block of length 15575379 for bucket 8
(Using difference cover)
bucket 2: 90%
bucket 7: 80%
bucket 3: 100%
Sorting block of length 7717778 for bucket 3
(Using difference cover)
bucket 6: 90%
bucket 4: 70%
bucket 2: 100%
Sorting block of length 19637635 for bucket 2
(Using difference cover)
bucket 5: 80%
bucket 7: 90%
bucket 6: 100%
Sorting block of length 9971415 for bucket 6
(Using difference cover)
bucket 4: 80%
bucket 5: 90%
bucket 4: 90%
bucket 7: 100%
Sorting block of length 17229264 for bucket 7
(Using difference cover)
bucket 4: 100%
Sorting block of length 26175158 for bucket 4
(Using difference cover)
bucket 5: 100%
Sorting block of length 22000306 for bucket 5
(Using difference cover)
Sorting block time: 00:01:27
Returning block of 7717779 for bucket 3
Sorting block time: 00:02:11
Returning block of 9586585 for bucket 1
Sorting block time: 00:02:00
Returning block of 9971416 for bucket 6
Sorting block time: 00:02:56
Returning block of 15575380 for bucket 8
Sorting block time: 00:03:13
Returning block of 14679498 for bucket 9
Sorting block time: 00:03:08
Returning block of 19637636 for bucket 2
Sorting block time: 00:03:00
Returning block of 17229265 for bucket 7
Sorting block time: 00:03:11
Returning block of 22000307 for bucket 5
Sorting block time: 00:03:29
Returning block of 26175159 for bucket 4

Exited GFM loop
fchr[A]: 0
fchr[C]: 41351058
fchr[G]: 71311730
fchr[T]: 101240543
fchr[$]: 142573024
Exiting GFM::buildToDisk()
Returning from initFromVector
Wrote 52172452 bytes to primary GFM file: dm_indexed.fa.1.ht2
Wrote 35643264 bytes to secondary GFM file: dm_indexed.fa.2.ht2
Re-opening _in1 and _in2 as input streams
Returning from GFM constructor
Returning from initFromVector
Wrote 77356861 bytes to primary GFM file: dm_indexed.fa.5.ht2
Wrote 36275454 bytes to secondary GFM file: dm_indexed.fa.6.ht2
Re-opening _in5 and _in5 as input streams
Returning from HierEbwt constructor
Headers:
  len: 142573024
  gbwtLen: 142573025
  nodes: 142573025
  sz: 35643256
  gbwtSz: 35643257
  lineRate: 6
  offRate: 4
  offMask: 0xfffffff0
  ftabChars: 10
  eftabLen: 0
  eftabSz: 0
  ftabLen: 1048577
  ftabSz: 4194308
  offsLen: 8910815
  offsSz: 35643260
  lineSz: 64
  sideSz: 64
  sideGbwtSz: 48
  sideGbwtLen: 192
  numSides: 742568
  numLines: 742568
  gbwtTotLen: 47524352
  gbwtTotSz: 47524352
  reverse: 0
  linearFM: Yes
Total time for call to driver() for forward index: 00:07:09

files:

dm_indexed.fa.1.ht2
dm_indexed.fa.2.ht2
dm_indexed.fa.3.ht2
dm_indexed.fa.4.ht2
dm_indexed.fa.5.ht2
dm_indexed.fa.6.ht2
dm_indexed.fa.7.ht2
dm_indexed.fa.8.ht2

Indexed D. willistoni genome:

hisat2-build -p 24 willistoni/GCF_000005925.1_dwil_caf1_genomic.fna wil_i.fa

Settings:
  Output files: "wil_i.fa.*.ht2"
  Line rate: 6 (line is 64 bytes)

```
  Lines per side: 1 (side is 64 bytes)
  Offset rate: 4 (one in 16)
  FTable chars: 10
  Strings: unpacked
  Local offset rate: 3 (one in 8)
  Local fTable chars: 6
  Local sequence length: 57344
  Local sequence overlap between two consecutive indexes: 1024
  Endianness: little
  Actual local endianness: little
  Sanity checking: disabled
  Assertions: disabled
  Random seed: 0
  Sizeofs: void*:8, int:4, long:8, size_t:8
Input files DNA, FASTA:
  willistoni/GCF_000005925.1_dwil_caf1_genomic.fna
Reading reference sizes
  Time reading reference sizes: 00:00:04
Calculating joined length
Writing header
Reserving space for joined string
Joining reference sequences
  Time to join reference sequences: 00:00:03
  Time to read SNPs and splice sites: 00:00:00
Using parameters --bmax 41926881 --dcv 1024
  Doing ahead-of-time memory usage test
  Passed!  Constructing with these parameters: --bmax 41926881 --dcv 1024
Constructing suffix-array element generator
Building DifferenceCoverSample
  Building sPrime
  Building sPrimeOrder
  V-Sorting samples
  V-Sorting samples time: 00:00:14
  Allocating rank array
  Ranking v-sort output
  Ranking v-sort output time: 00:00:02
  Invoking Larsson-Sadakane on ranks
  Invoking Larsson-Sadakane on ranks time: 00:00:03
  Sanity-checking and returning
Building samples
Reserving space for 12 sample suffixes
Generating random suffixes
QSorting 12 sample offsets, eliminating duplicates
QSorting sample offsets, eliminating duplicates time: 00:00:00
Multikey QSorting 12 samples
  (Using difference cover)
  Multikey QSorting samples time: 00:00:00
Calculating bucket sizes
Splitting and merging
  Splitting and merging time: 00:00:00
Split 1, merged 6; iterating...
Splitting and merging
  Splitting and merging time: 00:00:00
Split 1, merged 1; iterating...
Splitting and merging
  Splitting and merging time: 00:00:00
Avg bucket size: 3.19443e+07 (target: 41926880)
Converting suffix-array elements to index image
Allocating ftab, absorbFtab
Entering GFM loop
Getting block 1 of 7
```

Reserving size (41926881) for bucket 1
Getting block 2 of 7
Getting block 3 of 7
Reserving size (41926881) for bucket 3
Reserving size (41926881) for bucket 2
Getting block 5 of 7
Reserving size (41926881) for bucket 5
Getting block 7 of 7
Getting block 4 of 7
Getting block 6 of 7
Reserving size (41926881) for bucket 4
Reserving size (41926881) for bucket 7
Reserving size (41926881) for bucket 6
Calculating Z arrays for bucket 3
Calculating Z arrays for bucket 2
Entering block accumulator loop for bucket 3:
Calculating Z arrays for bucket 4
Calculating Z arrays for bucket 5
Entering block accumulator loop for bucket 2:
Calculating Z arrays for bucket 7
Entering block accumulator loop for bucket 4:
Entering block accumulator loop for bucket 5:
Entering block accumulator loop for bucket 7:
Calculating Z arrays for bucket 6
Entering block accumulator loop for bucket 6:
Calculating Z arrays for bucket 1
Entering block accumulator loop for bucket 1:
bucket 7: 10%
bucket 1: 10%
bucket 3: 10%
bucket 6: 10%
bucket 2: 10%
bucket 5: 10%
bucket 4: 10%
bucket 7: 20%
bucket 1: 20%
bucket 3: 20%
bucket 2: 20%
bucket 7: 30%
bucket 5: 20%
bucket 6: 20%
bucket 1: 30%
bucket 4: 20%
bucket 7: 40%
bucket 3: 30%
bucket 1: 40%
bucket 2: 30%
bucket 7: 50%
bucket 6: 30%
bucket 5: 30%
bucket 1: 50%
bucket 7: 60%
bucket 4: 30%
bucket 3: 40%
bucket 7: 70%
bucket 1: 60%
bucket 2: 40%
bucket 6: 40%
bucket 7: 80%
bucket 1: 70%

bucket 5: 40%
bucket 4: 40%
bucket 6: 50%
bucket 7: 90%
bucket 3: 50%
bucket 1: 80%
bucket 2: 50%
bucket 7: 100%
Sorting block of length 40244787 for bucket 7
(Using difference cover)
bucket 5: 50%
bucket 1: 90%
bucket 6: 60%
bucket 4: 50%
bucket 3: 60%
bucket 2: 60%
bucket 1: 100%
Sorting block of length 32299224 for bucket 1
(Using difference cover)
bucket 5: 60%
bucket 6: 70%
bucket 3: 70%
bucket 4: 60%
bucket 2: 70%
bucket 6: 80%
bucket 3: 80%
bucket 5: 70%
bucket 2: 80%
bucket 6: 90%
bucket 4: 70%
bucket 3: 90%
bucket 5: 80%
bucket 2: 90%
bucket 6: 100%
Sorting block of length 35204375 for bucket 6
(Using difference cover)
bucket 3: 100%
Sorting block of length 17045368 for bucket 3
(Using difference cover)
bucket 4: 80%
bucket 2: 100%
Sorting block of length 35623357 for bucket 2
(Using difference cover)
bucket 5: 90%
bucket 4: 90%
bucket 5: 100%
Sorting block of length 24941802 for bucket 5
(Using difference cover)
bucket 4: 100%
Sorting block of length 38251109 for bucket 4
(Using difference cover)
Sorting block time: 00:01:33
Returning block of 17045369 for bucket 3
Sorting block time: 00:02:44
Returning block of 32299225 for bucket 1
Sorting block time: 00:02:14
Returning block of 24941803 for bucket 5
Sorting block time: 00:02:59
Returning block of 35204376 for bucket 6
Sorting block time: 00:02:54
Returning block of 35623358 for bucket 2

Sorting block time: 00:03:29
Returning block of 40244788 for bucket 7
Sorting block time: 00:02:59
Returning block of 38251110 for bucket 4
Exited GFM loop
fchr[A]: 0
fchr[C]: 70183145
fchr[G]: 111837001
fchr[T]: 153469783
fchr[$]: 223610028
Exiting GFM::buildToDisk()
Returning from initFromVector
Wrote 80874733 bytes to primary GFM file: wil_i.fa.1.ht2
Wrote 55902512 bytes to secondary GFM file: wil_i.fa.2.ht2
Re-opening _in1 and _in2 as input streams
Returning from GFM constructor
Returning from initFromVector
Wrote 216134265 bytes to primary GFM file: wil_i.fa.5.ht2
Wrote 56767354 bytes to secondary GFM file: wil_i.fa.6.ht2
Re-opening _in5 and _in5 as input streams
Returning from HierEbwt constructor
Headers:
    len: 223610028
    gbwtLen: 223610029
    nodes: 223610029
    sz: 55902507
    gbwtSz: 55902508
    lineRate: 6
    offRate: 4
    offMask: 0xfffffff0
    ftabChars: 10
    eftabLen: 0
    eftabSz: 0
    ftabLen: 1048577
    ftabSz: 4194308
    offsLen: 13975627
    offsSz: 55902508
    lineSz: 64
    sideSz: 64
    sideGbwtSz: 48
    sideGbwtLen: 192
    numSides: 1164636
    numLines: 1164636
    gbwtTotLen: 74536704
    gbwtTotSz: 74536704
    reverse: 0
    linearFM: Yes
Total time for call to driver() for forward index: 00:08:15

files:

wil_i.fa.1.ht2
wil_i.fa.2.ht2
wil_i.fa.3.ht2
wil_i.fa.4.ht2
wil_i.fa.5.ht2
wil_i.fa.6.ht2
wil_i.fa.7.ht2
wil_i.fa.8.ht2

Indexed D. sechellia genome:

```
hisat2-build -p 24 sechellia/GCF_004382195.1_ASM438219v1_genomic.fna se_i.fa
Settings:
  Output files: "se_i.fa.*.ht2"
  Line rate: 6 (line is 64 bytes)
  Lines per side: 1 (side is 64 bytes)
  Offset rate: 4 (one in 16)
  FTable chars: 10
  Strings: unpacked
  Local offset rate: 3 (one in 8)
  Local fTable chars: 6
  Local sequence length: 57344
  Local sequence overlap between two consecutive indexes: 1024
  Endianness: little
  Actual local endianness: little
  Sanity checking: disabled
  Assertions: disabled
  Random seed: 0
  Sizeofs: void*:8, int:4, long:8, size_t:8
Input files DNA, FASTA:
  sechellia/GCF_004382195.1_ASM438219v1_genomic.fna
Reading reference sizes
  Time reading reference sizes: 00:00:03
Calculating joined length
Writing header
Reserving space for joined string
Joining reference sequences
  Time to join reference sequences: 00:00:02
  Time to read SNPs and splice sites: 00:00:00
Using parameters --bmax 28906315 --dcv 1024
  Doing ahead-of-time memory usage test
  Passed!  Constructing with these parameters: --bmax 28906315 --dcv 1024
Constructing suffix-array element generator
Building DifferenceCoverSample
  Building sPrime
  Building sPrimeOrder
  V-Sorting samples
  V-Sorting samples time: 00:00:13
  Allocating rank array
  Ranking v-sort output
  Ranking v-sort output time: 00:00:02
  Invoking Larsson-Sadakane on ranks
  Invoking Larsson-Sadakane on ranks time: 00:00:02
  Sanity-checking and returning
Building samples
Reserving space for 12 sample suffixes
Generating random suffixes
QSorting 12 sample offsets, eliminating duplicates
QSorting sample offsets, eliminating duplicates time: 00:00:00
Multikey QSorting 12 samples
  (Using difference cover)
  Multikey QSorting samples time: 00:00:00
Calculating bucket sizes
Splitting and merging
  Splitting and merging time: 00:00:00
Split 1, merged 8; iterating...
Splitting and merging
  Splitting and merging time: 00:00:00
Split 1, merged 0; iterating...
Splitting and merging
  Splitting and merging time: 00:00:00
```

Avg bucket size: 2.56945e+07 (target: 28906314)
Converting suffix-array elements to index image
Allocating ftab, absorbFtab
Entering GFM loop
Getting block 1 of 6
  Reserving size (28906315) for bucket 1
Getting block 3 of 6
  Reserving size (28906315) for bucket 3
Getting block 2 of 6
  Reserving size (28906315) for bucket 2
Getting block 4 of 6
Getting block 5 of 6
  Reserving size (28906315) for bucket 5
  Reserving size (28906315) for bucket 4
  Calculating Z arrays for bucket 1
  Entering block accumulator loop for bucket 1:
Getting block 6 of 6
  Reserving size (28906315) for bucket 6
  Calculating Z arrays for bucket 3
  Calculating Z arrays for bucket 2
  Calculating Z arrays for bucket 5
  Calculating Z arrays for bucket 6
  Calculating Z arrays for bucket 4
  Entering block accumulator loop for bucket 2:
  Entering block accumulator loop for bucket 3:
  Entering block accumulator loop for bucket 5:
  Entering block accumulator loop for bucket 6:
  Entering block accumulator loop for bucket 4:
  bucket 1: 10%
  bucket 6: 10%
  bucket 5: 10%
  bucket 3: 10%
  bucket 2: 10%
  bucket 4: 10%
  bucket 1: 20%
  bucket 6: 20%
  bucket 5: 20%
  bucket 2: 20%
  bucket 3: 20%
  bucket 4: 20%
  bucket 1: 30%
  bucket 6: 30%
  bucket 2: 30%
  bucket 1: 40%
  bucket 5: 30%
  bucket 6: 40%
  bucket 4: 30%
  bucket 3: 30%
  bucket 1: 50%
  bucket 5: 40%
  bucket 6: 50%
  bucket 3: 40%
  bucket 1: 60%
  bucket 2: 40%
  bucket 4: 40%
  bucket 1: 70%
  bucket 5: 50%
  bucket 3: 50%
  bucket 6: 60%
  bucket 4: 50%
  bucket 2: 50%

bucket 1: 80%
bucket 6: 70%
bucket 5: 60%
bucket 3: 60%
bucket 4: 60%
bucket 1: 90%
bucket 2: 60%
bucket 6: 80%
bucket 5: 70%
bucket 1: 100%
Sorting block of length 28073245 for bucket 1
(Using difference cover)
bucket 3: 70%
bucket 4: 70%
bucket 2: 70%
bucket 6: 90%
bucket 5: 80%
bucket 3: 80%
bucket 4: 80%
bucket 2: 80%
bucket 6: 100%
Sorting block of length 26520870 for bucket 6
(Using difference cover)
bucket 5: 90%
bucket 3: 90%
bucket 4: 90%
bucket 2: 90%
bucket 5: 100%
Sorting block of length 23743199 for bucket 5
(Using difference cover)
bucket 3: 100%
Sorting block of length 27331034 for bucket 3
(Using difference cover)
bucket 4: 100%
Sorting block of length 26329400 for bucket 4
(Using difference cover)
bucket 2: 100%
Sorting block of length 22169262 for bucket 2
(Using difference cover)
Sorting block time: 00:03:32
Returning block of 22169263 for bucket 2
Sorting block time: 00:03:36
Returning block of 23743200 for bucket 5
Sorting block time: 00:03:54
Returning block of 26520871 for bucket 6
Sorting block time: 00:03:55
Returning block of 27331035 for bucket 3
Sorting block time: 00:04:06
Returning block of 28073246 for bucket 1
Sorting block time: 00:03:58
Returning block of 26329401 for bucket 4
Exited GFM loop
fchr[A]: 0
fchr[C]: 44434093
fchr[G]: 76977519
fchr[T]: 109584654
fchr[$]: 154167015
Exiting GFM::buildToDisk()
Returning from initFromVector
Wrote 55640802 bytes to primary GFM file: se_i.fa.1.ht2

Wrote 38541760 bytes to secondary GFM file: se_i.fa.2.ht2
Re-opening _in1 and _in2 as input streams
Returning from GFM constructor
Returning from initFromVector
Wrote 69130217 bytes to primary GFM file: se_i.fa.5.ht2
Wrote 39191866 bytes to secondary GFM file: se_i.fa.6.ht2
Re-opening _in5 and _in5 as input streams
Returning from HierEbwt constructor
Headers:
  len: 154167015
  gbwtLen: 154167016
  nodes: 154167016
  sz: 38541754
  gbwtSz: 38541755
  lineRate: 6
  offRate: 4
  offMask: 0xfffffff0
  ftabChars: 10
  eftabLen: 0
  eftabSz: 0
  ftabLen: 1048577
  ftabSz: 4194308
  offsLen: 9635439
  offsSz: 38541756
  lineSz: 64
  sideSz: 64
  sideGbwtSz: 48
  sideGbwtLen: 192
  numSides: 802954
  numLines: 802954
  gbwtTotLen: 51389056
  gbwtTotSz: 51389056
  reverse: 0
  linearFM: Yes
Total time for call to driver() for forward index: 00:08:14

files:

se_i.fa.1.ht2
se_i.fa.2.ht2
se_i.fa.3.ht2
se_i.fa.4.ht2
se_i.fa.5.ht2
se_i.fa.6.ht2
se_i.fa.7.ht2
Se_i.fa.8.ht2

Hisat2 & Samtools (Nina Hu)

Download Hisat2 from: https://cloud.biohpc.swmed.edu/index.php/s/fE9QCsX3NH4QwBi/download

file: hisat2-2.2.1

command:

```
tar -cf hisat2.tar hisat2-2.2.1
scp hisat2.tar whu12@bluehive.circ.rochester.edu:/scratch/bio257_2020/Users/group2_adh
tar -xf hisat2.tar
cd hisat2-2.2.1
make
```

Set path variable every time before submitting each slurm file

```
export PATH=$PATH:/scratch/bio257_2020/Users/group2_adh/hisat2-2.2.1
```

SRR070408

command:

```
module load samtools
module load python3
hisat2 -p 4 --phred33 -x dmelanogaster/m_indexed/dm_indexed.fa -1 rnaseq/SRR070408_1.fastq.gz -2
rnaseq/SRR070408_2.fastq.gz | samtools view -@ 4 -b -h -F 4 - | samtools sort -@ 4 - -o bam/b408.mapped.sorted.bam
19516639 reads; of these:
  19516639 (100.00%) were paired; of these:
    2776903 (14.23%) aligned concordantly 0 times
    15048887 (77.11%) aligned concordantly exactly 1 time
    1690849 (8.66%) aligned concordantly >1 times
    ----
    2776903 pairs aligned concordantly 0 times; of these:
      157738 (5.68%) aligned discordantly 1 time
    ----
    2619165 pairs aligned 0 times concordantly or discordantly; of these:
      5238330 mates make up the pairs; of these:
        2765539 (52.79%) aligned 0 times
        1997491 (38.13%) aligned exactly 1 time
        475300 (9.07%) aligned >1 times
92.91% overall alignment rate
[bam_sort_core] merging from 12 files and 4 in-memory blocks...
```

file:

b408.mapped.sorted.bam

command:

```
samtools view -@ 2 -b -h -q 10 b408.mapped.sorted.bam | samtools sort -@ 2 - -o b408.unique.mapped.sorted.bam
[W::bgzf_read_block] EOF marker is absent. The input is probably truncated
[bam_sort_core] merging from 0 files and 4 in-memory blocks...
```

file:

b408.unique.mapped.sorted.bam

command:

```
samtools index b408.mapped.sorted.bam
samtools index b408.unique.mapped.sorted.bam
```

file:

b408.mapped.sorted.bam.bai
b408.unique.mapped.sorted.bam.bai

SRR070425

command:

```
hisat2 -p 4 --phred33 -x dmelanogaster/m_indexed/dm_indexed.fa -1 rnaseq/SRR070425_1.fastq.gz -2
rnaseq/SRR070425_2.fastq.gz | samtools view -@ 4 -b -h -F 4 - | samtools sort -@ 4 - -o bam/b425.mapped.sorted.bam
29852257 reads; of these:
  29852257 (100.00%) were paired; of these:
    7106990 (23.81%) aligned concordantly 0 times
    17831204 (59.73%) aligned concordantly exactly 1 time
    4914063 (16.46%) aligned concordantly >1 times
    ----
    7106990 pairs aligned concordantly 0 times; of these:
      66624 (0.94%) aligned discordantly 1 time
    ----
    7040366 pairs aligned 0 times concordantly or discordantly; of these:
      14080732 mates make up the pairs; of these:
        8531284 (60.59%) aligned 0 times
        3951521 (28.06%) aligned exactly 1 time
        1597927 (11.35%) aligned >1 times
85.71% overall alignment rate
[bam_sort_core] merging from 24 files and 4 in-memory blocks...
```

file:
b425.mapped.sorted.bam
command:
samtools view -@ 2 -b -h -q 10 b425.mapped.sorted.bam | samtools sort -@ 2 - -o b425.unique.mapped.sorted.bam
[W::bgzf_read_block] EOF marker is absent. The input is probably truncated
[bam_sort_core] merging from 0 files and 4 in-memory blocks...
file:
b425.unique.mapped.sorted.bam
command:
samtools index b425.mapped.sorted.bam
samtools index b425.unique.mapped.sorted.bam
file:
b425.mapped.sorted.bam.bai
b425.unique.mapped.sorted.bam.bai

SRR6161775 (D. willistoni)

hisat2 -p 4 --phred33 -x dwillistoni/d_indexed/wil_i.fa -1 rnaseq/SRR6161775_1.fastq.gz -2 rnaseq/SRR6161775_2.fastq.gz |
samtools view -@ 4 -b -h -F 4 - | samtools sort -@ 4 - -o bam/b775.mapped.sorted.bam
command:
27550939 reads; of these:
  27550939 (100.00%) were paired; of these:
    2979709 (10.82%) aligned concordantly 0 times
    22539644 (81.81%) aligned concordantly exactly 1 time
    2031586 (7.37%) aligned concordantly >1 times

    2979709 pairs aligned concordantly 0 times; of these:
      375029 (12.59%) aligned discordantly 1 time

    2604680 pairs aligned 0 times concordantly or discordantly; of these:
      5209360 mates make up the pairs; of these:
        1786242 (34.29%) aligned 0 times
        2176490 (41.78%) aligned exactly 1 time
        1246628 (23.93%) aligned >1 times
96.76% overall alignment rate
[bam_sort_core] merging from 16 files and 4 in-memory blocks...
file: b775.mapped.sorted.bam
command:
samtools view -@ 4 -b -h -q 10 b775.mapped.sorted.bam | samtools sort -@ 4 - -o b775.unique.mapped.sorted.bam
[W::bgzf_read_block] EOF marker is absent. The input is probably truncated
[bam_sort_core] merging from 0 files and 4 in-memory blocks...
file: b775.unique.mapped.sorted.bam
command:
samtools index b775.mapped.sorted.bam
samtools index b775.unique.mapped.sorted.bam
file:
b775.mapped.sorted.bam.bai
b775.unique.mapped.sorted.bam.bai

SRR6161779 (D. sechelia)

hisat2 -p 4 --phred33 -x dsechelia/s_indexed/se_i.fa -1 rnaseq/SRR6161779_1.fastq.gz -2 rnaseq/SRR6161779_2.fastq.gz |
samtools view -@ 4 -b -h -F 4 - | samtools sort -@ 4 - -o bam/b779.mapped.sorted.bam
21546665 reads; of these:
  21546665 (100.00%) were paired; of these:
    7533263 (34.96%) aligned concordantly 0 times
    13550351 (62.89%) aligned concordantly exactly 1 time

463051 (2.15%) aligned concordantly >1 times

7533263 pairs aligned concordantly 0 times; of these:
199973 (2.65%) aligned discordantly 1 time

7333290 pairs aligned 0 times concordantly or discordantly; of these:
14666580 mates make up the pairs; of these:
10520857 (71.73%) aligned 0 times
3539683 (24.13%) aligned exactly 1 time
606040 (4.13%) aligned >1 times
75.59% overall alignment rate
[bam_sort_core] merging from 8 files and 4 in-memory blocks...

file:

b779.mapped.sorted.bam

command:

samtools view -@ 4 -b -h -q 10 b779.mapped.sorted.bam | samtools sort -@ 4 - -o b779.unique.mapped.sorted.bam
[W::bgzf_read_block] EOF marker is absent. The input is probably truncated
[bam_sort_core] merging from 0 files and 4 in-memory blocks...

file: b779.unique.mapped.sorted.bam

command:

samtools index b779.mapped.sorted.bam
samtools index b779.unique.mapped.sorted.bam

file:

b779.mapped.sorted.bam.bai
b779.unique.mapped.sorted.bam.bai

SRR6161782 (D. melanogaster)

hisat2 -p 4 --phred33 -x dmelanogaster/m_indexed/dm_indexed.fa -1 rnaseq/SRR6161782_1.fastq.gz -2 rnaseq/SRR6161782_2.fastq.gz | samtools view -@ 4 -b -h -F 4 - | samtools sort -@ 4 - -o b782.mapped.sorted.bam

command:

26508281 reads; of these:
26508281 (100.00%) were paired; of these:
1799219 (6.79%) aligned concordantly 0 times
24061761 (90.77%) aligned concordantly exactly 1 time
647301 (2.44%) aligned concordantly >1 times

1799219 pairs aligned concordantly 0 times; of these:
389394 (21.64%) aligned discordantly 1 time

1409825 pairs aligned 0 times concordantly or discordantly; of these:
2819650 mates make up the pairs; of these:
1425957 (50.57%) aligned 0 times
1288996 (45.71%) aligned exactly 1 time
104697 (3.71%) aligned >1 times
97.31% overall alignment rate
[bam_sort_core] merging from 16 files and 4 in-memory blocks...

file:

b782.mapped.sorted.bam

command:

samtools view -@ 4 -b -h -q 10 b782.mapped.sorted.bam | samtools sort -@ 4 - -o b782.unique.mapped.sorted.bam
[W::bgzf_read_block] EOF marker is absent. The input is probably truncated
[bam_sort_core] merging from 0 files and 4 in-memory blocks...

file: b782.unique.mapped.sorted.bam

command:

samtools index b782.mapped.sorted.bam
samtools index b782.unique.mapped.sorted.bam

file:

b782.mapped.sorted.bam.bai
b782.unique.mapped.sorted.bam.bai

HTSeq count (Yijin Zhao)

b408

command:

```
htseq-count -r pos -i "gene_id" -t 'exon' -f 'bam' --max-reads-in-buffer=60000000 --stranded=reverse
bam/b408.unique.mapped.sorted.bam dmelanogaster/dmel-all-r6.30.gtf > b408.unique.count
100000 GFF lines processed.
200000 GFF lines processed.
300000 GFF lines processed.
400000 GFF lines processed.
500000 GFF lines processed.
543992 GFF lines processed.
100000 SAM alignment record pairs processed.
200000 SAM alignment record pairs processed.
300000 SAM alignment record pairs processed.
400000 SAM alignment record pairs processed.
500000 SAM alignment record pairs processed.
600000 SAM alignment record pairs processed.
700000 SAM alignment record pairs processed.
800000 SAM alignment record pairs processed.
900000 SAM alignment record pairs processed.
1000000 SAM alignment record pairs processed.
1100000 SAM alignment record pairs processed.
1200000 SAM alignment record pairs processed.
1300000 SAM alignment record pairs processed.
1400000 SAM alignment record pairs processed.
1500000 SAM alignment record pairs processed.
1600000 SAM alignment record pairs processed.
1700000 SAM alignment record pairs processed.
1800000 SAM alignment record pairs processed.
1900000 SAM alignment record pairs processed.
2000000 SAM alignment record pairs processed.
2100000 SAM alignment record pairs processed.
2200000 SAM alignment record pairs processed.
2300000 SAM alignment record pairs processed.
2400000 SAM alignment record pairs processed.
2500000 SAM alignment record pairs processed.
2600000 SAM alignment record pairs processed.
2700000 SAM alignment record pairs processed.
2800000 SAM alignment record pairs processed.
2900000 SAM alignment record pairs processed.
3000000 SAM alignment record pairs processed.
3100000 SAM alignment record pairs processed.
3200000 SAM alignment record pairs processed.
3300000 SAM alignment record pairs processed.
3400000 SAM alignment record pairs processed.
3500000 SAM alignment record pairs processed.
3600000 SAM alignment record pairs processed.
3700000 SAM alignment record pairs processed.
3800000 SAM alignment record pairs processed.
3900000 SAM alignment record pairs processed.
4000000 SAM alignment record pairs processed.
4100000 SAM alignment record pairs processed.
4200000 SAM alignment record pairs processed.
4300000 SAM alignment record pairs processed.
4400000 SAM alignment record pairs processed.
```

4500000 SAM alignment record pairs processed.
4600000 SAM alignment record pairs processed.
4700000 SAM alignment record pairs processed.
4800000 SAM alignment record pairs processed.
4900000 SAM alignment record pairs processed.
5000000 SAM alignment record pairs processed.
5100000 SAM alignment record pairs processed.
5200000 SAM alignment record pairs processed.
5300000 SAM alignment record pairs processed.
5400000 SAM alignment record pairs processed.
5500000 SAM alignment record pairs processed.
5600000 SAM alignment record pairs processed.
5700000 SAM alignment record pairs processed.
5800000 SAM alignment record pairs processed.
5900000 SAM alignment record pairs processed.
6000000 SAM alignment record pairs processed.
6100000 SAM alignment record pairs processed.
6200000 SAM alignment record pairs processed.
6300000 SAM alignment record pairs processed.
6400000 SAM alignment record pairs processed.
6500000 SAM alignment record pairs processed.
6600000 SAM alignment record pairs processed.
6700000 SAM alignment record pairs processed.
6800000 SAM alignment record pairs processed.
6900000 SAM alignment record pairs processed.
7000000 SAM alignment record pairs processed.
7100000 SAM alignment record pairs processed.
7200000 SAM alignment record pairs processed.
7300000 SAM alignment record pairs processed.
7400000 SAM alignment record pairs processed.
7500000 SAM alignment record pairs processed.
7600000 SAM alignment record pairs processed.
7700000 SAM alignment record pairs processed.
7800000 SAM alignment record pairs processed.
7900000 SAM alignment record pairs processed.
8000000 SAM alignment record pairs processed.
8100000 SAM alignment record pairs processed.
8200000 SAM alignment record pairs processed.
8300000 SAM alignment record pairs processed.
8400000 SAM alignment record pairs processed.
8500000 SAM alignment record pairs processed.
8600000 SAM alignment record pairs processed.
8700000 SAM alignment record pairs processed.
8800000 SAM alignment record pairs processed.
8900000 SAM alignment record pairs processed.
9000000 SAM alignment record pairs processed.
9100000 SAM alignment record pairs processed.
9200000 SAM alignment record pairs processed.
9300000 SAM alignment record pairs processed.
9400000 SAM alignment record pairs processed.
9500000 SAM alignment record pairs processed.
9600000 SAM alignment record pairs processed.
9700000 SAM alignment record pairs processed.
9800000 SAM alignment record pairs processed.
9900000 SAM alignment record pairs processed.
10000000 SAM alignment record pairs processed.
10100000 SAM alignment record pairs processed.
10200000 SAM alignment record pairs processed.
10300000 SAM alignment record pairs processed.
10400000 SAM alignment record pairs processed.
10500000 SAM alignment record pairs processed.

10600000 SAM alignment record pairs processed.
10700000 SAM alignment record pairs processed.
10800000 SAM alignment record pairs processed.
10900000 SAM alignment record pairs processed.
11000000 SAM alignment record pairs processed.
11100000 SAM alignment record pairs processed.
11200000 SAM alignment record pairs processed.
11300000 SAM alignment record pairs processed.
11400000 SAM alignment record pairs processed.
11500000 SAM alignment record pairs processed.
11600000 SAM alignment record pairs processed.
11700000 SAM alignment record pairs processed.
11800000 SAM alignment record pairs processed.
11900000 SAM alignment record pairs processed.
12000000 SAM alignment record pairs processed.
12100000 SAM alignment record pairs processed.
12200000 SAM alignment record pairs processed.
12300000 SAM alignment record pairs processed.
12400000 SAM alignment record pairs processed.
12500000 SAM alignment record pairs processed.
12600000 SAM alignment record pairs processed.
12700000 SAM alignment record pairs processed.
12800000 SAM alignment record pairs processed.
12900000 SAM alignment record pairs processed.
13000000 SAM alignment record pairs processed.
13100000 SAM alignment record pairs processed.
13200000 SAM alignment record pairs processed.
13300000 SAM alignment record pairs processed.
13400000 SAM alignment record pairs processed.
13500000 SAM alignment record pairs processed.
13600000 SAM alignment record pairs processed.
13700000 SAM alignment record pairs processed.
13800000 SAM alignment record pairs processed.
13900000 SAM alignment record pairs processed.
14000000 SAM alignment record pairs processed.
14100000 SAM alignment record pairs processed.
14200000 SAM alignment record pairs processed.
14300000 SAM alignment record pairs processed.
14400000 SAM alignment record pairs processed.
14500000 SAM alignment record pairs processed.
14600000 SAM alignment record pairs processed.
14700000 SAM alignment record pairs processed.
14800000 SAM alignment record pairs processed.
14900000 SAM alignment record pairs processed.
15000000 SAM alignment record pairs processed.
15100000 SAM alignment record pairs processed.
15200000 SAM alignment record pairs processed.
15300000 SAM alignment record pairs processed.
Warning: Mate records missing for 1785879 records; first such record: <SAM_Alignment object: Paired-end read 'SRR070408.13309610' aligned to 3L:[3221168,3221243)/->.
15400000 SAM alignment record pairs processed.
15500000 SAM alignment record pairs processed.
15600000 SAM alignment record pairs processed.
15700000 SAM alignment record pairs processed.
15800000 SAM alignment record pairs processed.
15900000 SAM alignment record pairs processed.
16000000 SAM alignment record pairs processed.
16100000 SAM alignment record pairs processed.
16200000 SAM alignment record pairs processed.
16300000 SAM alignment record pairs processed.
16400000 SAM alignment record pairs processed.

16500000 SAM alignment record pairs processed.
16600000 SAM alignment record pairs processed.
16700000 SAM alignment record pairs processed.
16800000 SAM alignment record pairs processed.
16900000 SAM alignment record pairs processed.
17000000 SAM alignment record pairs processed.
17098310 SAM alignment pairs processed.

file: b408.unique.count

b425

command:

htseq-count -r pos -i "gene_id" -t 'exon' -f 'bam' --max-reads-in-buffer=60000000 --stranded=reverse bam/b425.unique.mapped.sorted.bam dmelanogaster/dmel-all-r6.30.gtf > b425.unique.count
100000 GFF lines processed.
200000 GFF lines processed.
300000 GFF lines processed.
400000 GFF lines processed.
500000 GFF lines processed.
543992 GFF lines processed.
100000 SAM alignment record pairs processed.
200000 SAM alignment record pairs processed.
300000 SAM alignment record pairs processed.
400000 SAM alignment record pairs processed.
500000 SAM alignment record pairs processed.
600000 SAM alignment record pairs processed.
700000 SAM alignment record pairs processed.
800000 SAM alignment record pairs processed.
900000 SAM alignment record pairs processed.
1000000 SAM alignment record pairs processed.
1100000 SAM alignment record pairs processed.
1200000 SAM alignment record pairs processed.
1300000 SAM alignment record pairs processed.
1400000 SAM alignment record pairs processed.
1500000 SAM alignment record pairs processed.
1600000 SAM alignment record pairs processed.
1700000 SAM alignment record pairs processed.
1800000 SAM alignment record pairs processed.
1900000 SAM alignment record pairs processed.
2000000 SAM alignment record pairs processed.
2100000 SAM alignment record pairs processed.
2200000 SAM alignment record pairs processed.
2300000 SAM alignment record pairs processed.
2400000 SAM alignment record pairs processed.
2500000 SAM alignment record pairs processed.
2600000 SAM alignment record pairs processed.
2700000 SAM alignment record pairs processed.
2800000 SAM alignment record pairs processed.
2900000 SAM alignment record pairs processed.
3000000 SAM alignment record pairs processed.
3100000 SAM alignment record pairs processed.
3200000 SAM alignment record pairs processed.
3300000 SAM alignment record pairs processed.
3400000 SAM alignment record pairs processed.
3500000 SAM alignment record pairs processed.
3600000 SAM alignment record pairs processed.
3700000 SAM alignment record pairs processed.
3800000 SAM alignment record pairs processed.
3900000 SAM alignment record pairs processed.
4000000 SAM alignment record pairs processed.

4100000 SAM alignment record pairs processed.
4200000 SAM alignment record pairs processed.
4300000 SAM alignment record pairs processed.
4400000 SAM alignment record pairs processed.
4500000 SAM alignment record pairs processed.
4600000 SAM alignment record pairs processed.
4700000 SAM alignment record pairs processed.
4800000 SAM alignment record pairs processed.
4900000 SAM alignment record pairs processed.
5000000 SAM alignment record pairs processed.
5100000 SAM alignment record pairs processed.
5200000 SAM alignment record pairs processed.
5300000 SAM alignment record pairs processed.
5400000 SAM alignment record pairs processed.
5500000 SAM alignment record pairs processed.
5600000 SAM alignment record pairs processed.
5700000 SAM alignment record pairs processed.
5800000 SAM alignment record pairs processed.
5900000 SAM alignment record pairs processed.
6000000 SAM alignment record pairs processed.
6100000 SAM alignment record pairs processed.
6200000 SAM alignment record pairs processed.
6300000 SAM alignment record pairs processed.
6400000 SAM alignment record pairs processed.
6500000 SAM alignment record pairs processed.
6600000 SAM alignment record pairs processed.
6700000 SAM alignment record pairs processed.
6800000 SAM alignment record pairs processed.
6900000 SAM alignment record pairs processed.
7000000 SAM alignment record pairs processed.
7100000 SAM alignment record pairs processed.
7200000 SAM alignment record pairs processed.
7300000 SAM alignment record pairs processed.
7400000 SAM alignment record pairs processed.
7500000 SAM alignment record pairs processed.
7600000 SAM alignment record pairs processed.
7700000 SAM alignment record pairs processed.
7800000 SAM alignment record pairs processed.
7900000 SAM alignment record pairs processed.
8000000 SAM alignment record pairs processed.
8100000 SAM alignment record pairs processed.
8200000 SAM alignment record pairs processed.
8300000 SAM alignment record pairs processed.
8400000 SAM alignment record pairs processed.
8500000 SAM alignment record pairs processed.
8600000 SAM alignment record pairs processed.
8700000 SAM alignment record pairs processed.
8800000 SAM alignment record pairs processed.
8900000 SAM alignment record pairs processed.
9000000 SAM alignment record pairs processed.
9100000 SAM alignment record pairs processed.
9200000 SAM alignment record pairs processed.
9300000 SAM alignment record pairs processed.
9400000 SAM alignment record pairs processed.
9500000 SAM alignment record pairs processed.
9600000 SAM alignment record pairs processed.
9700000 SAM alignment record pairs processed.
9800000 SAM alignment record pairs processed.
9900000 SAM alignment record pairs processed.
10000000 SAM alignment record pairs processed.
10100000 SAM alignment record pairs processed.

10200000 SAM alignment record pairs processed.
10300000 SAM alignment record pairs processed.
10400000 SAM alignment record pairs processed.
10500000 SAM alignment record pairs processed.
10600000 SAM alignment record pairs processed.
10700000 SAM alignment record pairs processed.
10800000 SAM alignment record pairs processed.
10900000 SAM alignment record pairs processed.
11000000 SAM alignment record pairs processed.
11100000 SAM alignment record pairs processed.
11200000 SAM alignment record pairs processed.
11300000 SAM alignment record pairs processed.
11400000 SAM alignment record pairs processed.
11500000 SAM alignment record pairs processed.
11600000 SAM alignment record pairs processed.
11700000 SAM alignment record pairs processed.
11800000 SAM alignment record pairs processed.
11900000 SAM alignment record pairs processed.
12000000 SAM alignment record pairs processed.
12100000 SAM alignment record pairs processed.
12200000 SAM alignment record pairs processed.
12300000 SAM alignment record pairs processed.
12400000 SAM alignment record pairs processed.
12500000 SAM alignment record pairs processed.
12600000 SAM alignment record pairs processed.
12700000 SAM alignment record pairs processed.
12800000 SAM alignment record pairs processed.
12900000 SAM alignment record pairs processed.
13000000 SAM alignment record pairs processed.
13100000 SAM alignment record pairs processed.
13200000 SAM alignment record pairs processed.
13300000 SAM alignment record pairs processed.
13400000 SAM alignment record pairs processed.
13500000 SAM alignment record pairs processed.
13600000 SAM alignment record pairs processed.
13700000 SAM alignment record pairs processed.
13800000 SAM alignment record pairs processed.
13900000 SAM alignment record pairs processed.
14000000 SAM alignment record pairs processed.
14100000 SAM alignment record pairs processed.
14200000 SAM alignment record pairs processed.
14300000 SAM alignment record pairs processed.
14400000 SAM alignment record pairs processed.
14500000 SAM alignment record pairs processed.
14600000 SAM alignment record pairs processed.
14700000 SAM alignment record pairs processed.
14800000 SAM alignment record pairs processed.
14900000 SAM alignment record pairs processed.
15000000 SAM alignment record pairs processed.
15100000 SAM alignment record pairs processed.
15200000 SAM alignment record pairs processed.
15300000 SAM alignment record pairs processed.
15400000 SAM alignment record pairs processed.
15500000 SAM alignment record pairs processed.
15600000 SAM alignment record pairs processed.
15700000 SAM alignment record pairs processed.
15800000 SAM alignment record pairs processed.
15900000 SAM alignment record pairs processed.
16000000 SAM alignment record pairs processed.
16100000 SAM alignment record pairs processed.
16200000 SAM alignment record pairs processed.

16300000 SAM alignment record pairs processed.
16400000 SAM alignment record pairs processed.
16500000 SAM alignment record pairs processed.
16600000 SAM alignment record pairs processed.
16700000 SAM alignment record pairs processed.
16800000 SAM alignment record pairs processed.
16900000 SAM alignment record pairs processed.
17000000 SAM alignment record pairs processed.
17100000 SAM alignment record pairs processed.
17200000 SAM alignment record pairs processed.
17300000 SAM alignment record pairs processed.
17400000 SAM alignment record pairs processed.
17500000 SAM alignment record pairs processed.
17600000 SAM alignment record pairs processed.
17700000 SAM alignment record pairs processed.
17800000 SAM alignment record pairs processed.
17900000 SAM alignment record pairs processed.
18000000 SAM alignment record pairs processed.
18100000 SAM alignment record pairs processed.
Warning: Mate records missing for 3404295 records; first such record: <SAM_Alignment object: Paired-end read 'SRR070425.5165617' aligned to X:[3250865,3250945)/->.
18200000 SAM alignment record pairs processed.
18300000 SAM alignment record pairs processed.
18400000 SAM alignment record pairs processed.
18500000 SAM alignment record pairs processed.
18600000 SAM alignment record pairs processed.
18700000 SAM alignment record pairs processed.
18800000 SAM alignment record pairs processed.
18900000 SAM alignment record pairs processed.
19000000 SAM alignment record pairs processed.
19100000 SAM alignment record pairs processed.
19200000 SAM alignment record pairs processed.
19300000 SAM alignment record pairs processed.
19400000 SAM alignment record pairs processed.
19500000 SAM alignment record pairs processed.
19600000 SAM alignment record pairs processed.
19700000 SAM alignment record pairs processed.
19800000 SAM alignment record pairs processed.
19900000 SAM alignment record pairs processed.
20000000 SAM alignment record pairs processed.
20100000 SAM alignment record pairs processed.
20200000 SAM alignment record pairs processed.
20300000 SAM alignment record pairs processed.
20400000 SAM alignment record pairs processed.
20500000 SAM alignment record pairs processed.
20600000 SAM alignment record pairs processed.
20700000 SAM alignment record pairs processed.
20800000 SAM alignment record pairs processed.
20900000 SAM alignment record pairs processed.
21000000 SAM alignment record pairs processed.
21100000 SAM alignment record pairs processed.
21200000 SAM alignment record pairs processed.
21300000 SAM alignment record pairs processed.
21400000 SAM alignment record pairs processed.
21500000 SAM alignment record pairs processed.
21575736 SAM alignment pairs processed.

file: b425.unique.count

b775

command:

```
htseq-count -r pos --idattr=ID -f 'bam' --max-reads-in-buffer=60000000 --stranded=reverse bam/ b775.unique.mapped.sorted.bam
dwillistoni/GCF_000005925.1_dwil_caf1_genomic.gff > b775.unique.count
100000 GFF lines processed.
200000 GFF lines processed.
253251 GFF lines processed.
100000 SAM alignment record pairs processed.
200000 SAM alignment record pairs processed.
300000 SAM alignment record pairs processed.
400000 SAM alignment record pairs processed.
500000 SAM alignment record pairs processed.
600000 SAM alignment record pairs processed.
700000 SAM alignment record pairs processed.
800000 SAM alignment record pairs processed.
900000 SAM alignment record pairs processed.
1000000 SAM alignment record pairs processed.
1100000 SAM alignment record pairs processed.
1200000 SAM alignment record pairs processed.
1300000 SAM alignment record pairs processed.
1400000 SAM alignment record pairs processed.
1500000 SAM alignment record pairs processed.
1600000 SAM alignment record pairs processed.
1700000 SAM alignment record pairs processed.
1800000 SAM alignment record pairs processed.
1900000 SAM alignment record pairs processed.
2000000 SAM alignment record pairs processed.
2100000 SAM alignment record pairs processed.
2200000 SAM alignment record pairs processed.
2300000 SAM alignment record pairs processed.
2400000 SAM alignment record pairs processed.
Warning: Mate records missing for 197025 records; first such record: <SAM_Alignment object: Paired-end read
'SRR6161775.8795726' aligned to NW_002031144.1:[2943,2994)/->.
2500000 SAM alignment record pairs processed.
2600000 SAM alignment record pairs processed.
2665247 SAM alignment pairs processed.
```

file: b775.unique.count

b779

command:

```
htseq-count -r pos --idattr=ID -f 'bam' --max-reads-in-buffer=60000000 --stranded=reverse bam/b779.unique.mapped.sorted.bam
dsechellia/GCF_004382195.1_ASM438219v1_genomic.gff > b779.unique.count
100000 GFF lines processed.
200000 GFF lines processed.
300000 GFF lines processed.
343153 GFF lines processed.
100000 SAM alignment record pairs processed.
200000 SAM alignment record pairs processed.
300000 SAM alignment record pairs processed.
400000 SAM alignment record pairs processed.
500000 SAM alignment record pairs processed.
600000 SAM alignment record pairs processed.
700000 SAM alignment record pairs processed.
800000 SAM alignment record pairs processed.
Warning: Mate records missing for 53683 records; first such record: <SAM_Alignment object: Paired-end read
'SRR6161779.4535750' aligned to NC_045949.1:[6441439,6441818)/+>.
900000 SAM alignment record pairs processed.
939100 SAM alignment pairs processed.
```

file: b779.unique.count

b782

command:

```
htseq-count -r pos -i "gene_id" -t 'exon' -f 'bam' --max-reads-in-buffer=60000000 --stranded=reverse
bam/b782.unique.mapped.sorted.bam dmelanogaster/dmel-all-r6.30.gtf > b782.unique.count
100000 GFF lines processed.
200000 GFF lines processed.
300000 GFF lines processed.
400000 GFF lines processed.
500000 GFF lines processed.
543992 GFF lines processed.
100000 SAM alignment record pairs processed.
200000 SAM alignment record pairs processed.
300000 SAM alignment record pairs processed.
400000 SAM alignment record pairs processed.
500000 SAM alignment record pairs processed.
600000 SAM alignment record pairs processed.
700000 SAM alignment record pairs processed.
800000 SAM alignment record pairs processed.
900000 SAM alignment record pairs processed.
1000000 SAM alignment record pairs processed.
1100000 SAM alignment record pairs processed.
1200000 SAM alignment record pairs processed.
1300000 SAM alignment record pairs processed.
Warning: Mate records missing for 89122 records; first such record: <SAM_Alignment object: Paired-end read
'SRR6161782.5916278' aligned to 2L:[3700729,3700836)/+>.
1400000 SAM alignment record pairs processed.
1468636 SAM alignment pairs processed.
```

File: b782.unique.count

December 3rd Thursday (Group)

Journal club paper 5

Kallisto Index transcriptome (Qizhe Zhang)

module load kallisto/0.43.1

D. melanogaster:

command: kallisto index -i kallisto/dm.kindex dmelanogaster/dmel-all-transcript-r6.13.fasta

```
[build] loading fasta file /scratch/bio257_2020/Example_data/reference.genome/dmel-all-transcript-r6.13.fasta
[build] k-mer length: 31
[build] warning: clipped off poly-A tail (longer than 10)
        from 8 target sequences
[build] counting k-mers ... done.
[build] building target de Bruijn graph ...  done
[build] creating equivalence classes ...  done
[build] target de Bruijn graph has 68636 contigs and contains 33122453 k-mers
```

file: dm.kindex

D. sechellia:

command:

kallisto index -i kallisto/ds.kindex dsechellia/GCF_004382195.1_ASM438219v1_rna.fna

```
quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
```

[index] k-mer length: 31
[index] number of targets: 26,658
[index] number of k-mers: 32,107,150
[index] number of equivalence classes: 48,105
[quant] running in paired-end mode
[quant] will process pair 1: SRR6161779_1.fastq.gz
SRR6161779_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 21,546,665 reads, 14,427,997 reads pseudoaligned
[quant] estimated average fragment length: 372.979
[em] quantifying the abundances ... done
[em] the Expectation-Maximization algorithm ran for 1,294 rounds
[bstrp] number of EM bootstraps complete: 100

file: ds.kindex

D. willistoni:

command: kallisto index -i kallisto/dw.kindex dwillistoni/GCF_000005925.1_dwil_caf1_rna.fna

quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 19,567
[index] number of k-mers: 26,373,002
[index] number of equivalence classes: 32,955
[quant] running in paired-end mode
[quant] will process pair 1: SRR6161775_1.fastq.gz
SRR6161775_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 27,550,939 reads, 18,160,041 reads pseudoaligned
[quant] estimated average fragment length: 309.566
[em] quantifying the abundances ... done
[em] the Expectation-Maximization algorithm ran for 1,188 rounds
[bstrp] number of EM bootstraps complete: 100

file dwil.kindex

408

command:

kallisto quant -b 100 -i kallisto/dm.kindex -o kallisto/k408 -plaintext -t 8 rnaseq/SRR070408_2.fastq.gz rnaseq/SRR070408_2.fastq.gz

quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 30,482
[index] number of k-mers: 33,122,453
[index] number of equivalence classes: 57,757
[quant] running in paired-end mode
[quant] will process pair 1: SRR070408_2.fastq.gz
SRR070408_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 19,516,639 reads, 17,992,347 reads pseudoaligned
[quant] estimated average fragment length: 0
[em] quantifying the abundances ... done
[em] the Expectation-Maximization algorithm ran for 1,312 rounds
[bstrp] number of EM bootstraps complete: 100

folder: k408

425

command:

```
kallisto quant -b 100 -i kallisto/dm.kindex -o kallisto/k425 -plaintext -t 8 rnaseq/SRR070425_2.fastq.gz
rnaseq/SRR070425_2.fastq.gz
quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 30,482
[index] number of k-mers: 33,122,453
[index] number of equivalence classes: 57,757
[quant] running in paired-end mode
[quant] will process pair 1: SRR070425_2.fastq.gz
                             SRR070425_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 29,852,257 reads, 25,761,510 reads pseudoaligned
[quant] estimated average fragment length: 0
[   em] quantifying the abundances ... done
[   em] the Expectation-Maximization algorithm ran for 1,219 rounds
[bstrp] number of EM bootstraps complete: 100
```

folder: k425

775

command:

```
kallisto quant -b 100 -i kallisto/dw.kindex -o kallisto/k775 -plaintext -t 8 rnaseq/SRR6161775_1.fastq.gz
rnaseq/SRR6161775_2.fastq.gz
quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 19,567
[index] number of k-mers: 26,373,002
[index] number of equivalence classes: 32,955
[quant] running in paired-end mode
[quant] will process pair 1: SRR6161775_1.fastq.gz
                             SRR6161775_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 27,550,939 reads, 18,160,041 reads pseudoaligned
[quant] estimated average fragment length: 309.566
[   em] quantifying the abundances ... done
[   em] the Expectation-Maximization algorithm ran for 1,188 rounds
[bstrp] number of EM bootstraps complete: 100
```

folder: k775

779

Command:

```
kallisto quant -b 100 -i kallisto/ds.kindex -o kallisto/k779 -plaintext -t 8 rnaseq/SRR6161779_1.fastq.gz
rnaseq/SRR6161779_2.fastq.gz
quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 26,658
[index] number of k-mers: 32,107,150
[index] number of equivalence classes: 48,105
[quant] running in paired-end mode
```

[quant] will process pair 1: SRR6161779_1.fastq.gz
SRR6161779_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 21,546,665 reads, 14,427,997 reads pseudoaligned
[quant] estimated average fragment length: 372.979
[em] quantifying the abundances ... done
[em] the Expectation-Maximization algorithm ran for 1,294 rounds
[bstrp] number of EM bootstraps complete: 100

file: k779

782

command:

kallisto quant -b 100 -i kallisto/dm.kindex -o k782 -plaintext -t 8 rnaseq/SRR6161782_1.fastq.gz rnaseq/SRR6161782_2.fastq.gz
quant: invalid option -- 'p'

[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 30,482
[index] number of k-mers: 33,122,453
[index] number of equivalence classes: 57,757
[quant] running in paired-end mode
[quant] will process pair 1: SRR6161782_1.fastq.gz
SRR6161782_2.fastq.gz
[quant] finding pseudoalignments for the reads ... done
[quant] processed 26,508,281 reads, 18,996,763 reads pseudoaligned
[quant] estimated average fragment length: 322.872
[em] quantifying the abundances ... done
[em] the Expectation-Maximization algorithm ran for 1,174 rounds
[bstrp] number of EM bootstraps complete: 100

folder: k782

StringTie to obtain TPMs (Nina Hu)

export PATH=$PATH:/scratch/bio257_2020/Users/group2_adh/stringtie-2.1.4.Linux_x86_64

command:

stringtie -p 24 -G dmelanogaster/dmel-all-r6.30.gtf -A stringtie/a408.tab -o stringtie/a408.out bam/b408.unique.mapped.sorted.bam
stringtie -p 24 -G dmelanogaster/dmel-all-r6.30.gtf -A stringtie/a425.tab -o stringtie/a425.out bam/b425.unique.mapped.sorted.bam
stringtie -p 24 -G dmelanogaster/dmel-all-r6.30.gtf -A stringtie/a782.tab -o stringtie/a782.out bam/b782.unique.mapped.sorted.bam
stringtie -p 24 -G dwillistoni/GCF_000005925.1_dwil_caf1_genomic.gff -A stringtie/a775.tab -o stringtie/a775.out
bam/b775.unique.mapped.sorted.bam
stringtie -p 24 -G dsechellia/GCF_004382195.1_ASM438219v1_genomic.gff -A stringtie/a779.tab -o stringtie/a779.out
bam/b779.unique.mapped.sorted.bam

December 4th Friday

BLAST (Yijin Zhao)

command:

module load blast
makeblastdb -in FBtr0100589 -dbtype nucl -parse_seqids -out m89
Building a new DB, current time: 12/05/2020 23:10:05
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m89
New DB title: FBtr0100589
Sequence type: Nucleotide
Deleted existing Nucleotide BLAST database named /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m89
Keep MBits: T
Maximum file size: 1000000000B

Adding sequences from FASTA; added 1 sequences in 0.00978899 seconds.
makeblastdb -in FBtr0100590 -dbtype nucl -parse_seqids -out m90
Building a new DB, current time: 12/05/2020 23:11:24
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m90
New DB title: FBtr0100590
Sequence type: Nucleotide
Deleted existing Nucleotide BLAST database named /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m90
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00878692 seconds.
makeblastdb -in FBtr0100591 -dbtype nucl -parse_seqids -out m91
Building a new DB, current time: 12/05/2020 23:11:46
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m91
New DB title: FBtr0100591
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00783896 seconds.
makeblastdb -in FBtr0100593 -dbtype nucl -parse_seqids -out m93
Building a new DB, current time: 12/05/2020 23:12:35
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m93
New DB title: FBtr0100593
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00784111 seconds.
makeblastdb -in FBtr0100594 -dbtype nucl -parse_seqids -out m94
Building a new DB, current time: 12/05/2020 23:12:48
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/m94
New DB title: FBtr0100594
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00694203 seconds.
makeblastdb -in sXM_002030726.2 -dbtype nucl -parse_seqids -out s1
Building a new DB, current time: 12/05/2020 23:13:01
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s1
New DB title: sXM_002030726.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.010752 seconds.

makeblastdb -in sXM_002030727.2 -dbtype nucl -parse_seqids -out s2
Building a new DB, current time: 12/05/2020 23:13:07
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s2
New DB title: sXM_002030727.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00701618 seconds.
makeblastdb -in sXM_002031657.2 -dbtype nucl -parse_seqids -out s3
Building a new DB, current time: 12/05/2020 23:16:39
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s3
New DB title: sXM_002031657.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00769806 seconds.
makeblastdb -in sXM_002032726.2 -dbtype nucl -parse_seqids -out s4

Building a new DB, current time: 12/05/2020 23:16:53
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s4
New DB title: sXM_002032726.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.007025 seconds.
makeblastdb -in sXM_002035803.2 -dbtype nucl -parse_seqids -out s5
Building a new DB, current time: 12/05/2020 23:17:06
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s5
New DB title: sXM_002035803.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00757408 seconds.
makeblastdb -in sXM_002035804.2 -dbtype nucl -parse_seqids -out s6
Building a new DB, current time: 12/05/2020 23:17:33
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s6
New DB title: sXM_002035804.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00700307 seconds.
makeblastdb -in sXM_032714437.1 -dbtype nucl -parse_seqids -out s7
Building a new DB, current time: 12/05/2020 23:17:51
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s7
New DB title: sXM_032714437.1
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00737691 seconds.
makeblastdb -in sXM_032727775.1 -dbtype nucl -parse_seqids -out s8
Building a new DB, current time: 12/05/2020 23:18:04
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/s8
New DB title: sXM_032727775.1
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.007725 seconds.
makeblastdb -in XM_002060947.3 -dbtype nucl -parse_seqids -out w1
Building a new DB, current time: 12/05/2020 23:22:14
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w1
New DB title: XM_002060947.3
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00813699 seconds.
makeblastdb -in XM_002061733.3 -dbtype nucl -parse_seqids -out w2
Building a new DB, current time: 12/05/2020 23:26:02
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w2
New DB title: XM_002061733.3
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.0108268 seconds.
makeblastdb -in XM_002061958.1 -dbtype nucl -parse_seqids -out w3
Building a new DB, current time: 12/05/2020 23:26:15
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w3
New DB title: XM_002061958.1
Sequence type: Nucleotide
Keep MBits: T

Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00681806 seconds.
makeblastdb -in XM_002061959.1 -dbtype nucl -parse_seqids -out w4
Building a new DB, current time: 12/05/2020 23:26:27
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w4
New DB title: XM_002061959.1
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00692296 seconds.
makeblastdb -in XM_002065338.3 -dbtype nucl -parse_seqids -out w5
Building a new DB, current time: 12/05/2020 23:26:46
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w5
New DB title: XM_002065338.3
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00728297 seconds.
makeblastdb -in XM_002065339.3 -dbtype nucl -parse_seqids -out w6
Building a new DB, current time: 12/05/2020 23:26:59
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w6
New DB title: XM_002065339.3
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00749087 seconds.
makeblastdb -in XM_002065887.3 -dbtype nucl -parse_seqids -out w7
Building a new DB, current time: 12/05/2020 23:27:18
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w7
New DB title: XM_002065887.3
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00811911 seconds.
makeblastdb -in XM_002066421.2 -dbtype nucl -parse_seqids -out w8
Building a new DB, current time: 12/05/2020 23:27:33
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w8
New DB title: XM_002066421.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00795293 seconds.

makeblastdb -in XM_002073145.2 -dbtype nucl -parse_seqids -out w9
Building a new DB, current time: 12/05/2020 23:28:04
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w9
New DB title: XM_002073145.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00675392 seconds.
makeblastdb -in XM_015178175.2 -dbtype nucl -parse_seqids -out w10
Building a new DB, current time: 12/05/2020 23:28:15
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w10
New DB title: XM_015178175.2
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00663614 seconds.
makeblastdb -in XM_023181840.1 -dbtype nucl -parse_seqids -out w11
Building a new DB, current time: 12/05/2020 23:28:26

New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/transcripts/w11
New DB title: XM_023181840.1
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.00688386 seconds.

Let FBtr0100589 be the query

```
blastn -db m89 -query FBtr0100589 -outfmt 7 -out m89.out
blastn -db m90 -query FBtr0100589 -outfmt 7 -out m90.out
blastn -db m91 -query FBtr0100589 -outfmt 7 -out m91.out
blastn -db m93 -query FBtr0100589 -outfmt 7 -out m93.out
blastn -db m94 -query FBtr0100589 -outfmt 7 -out m94.out
blastn -db s1 -query FBtr0100589 -outfmt 7 -out s1.out
blastn -db s2 -query FBtr0100589 -outfmt 7 -out s2.out
blastn -db s3 -query FBtr0100589 -outfmt 7 -out s3.out
blastn -db s4 -query FBtr0100589 -outfmt 7 -out s4.out
blastn -db s5 -query FBtr0100589 -outfmt 7 -out s5.out
blastn -db s6 -query FBtr0100589 -outfmt 7 -out s6.out
blastn -db s7 -query FBtr0100589 -outfmt 7 -out s7.out
blastn -db s8 -query FBtr0100589 -outfmt 7 -out s8.out
blastn -db w1 -query FBtr0100589 -outfmt 7 -out w1.out
blastn -db w2 -query FBtr0100589 -outfmt 7 -out w2.out
blastn -db w3 -query FBtr0100589 -outfmt 7 -out w3.out
blastn -db w4 -query FBtr0100589 -outfmt 7 -out w4.out
blastn -db w5 -query FBtr0100589 -outfmt 7 -out w5.out
blastn -db w6 -query FBtr0100589 -outfmt 7 -out w6.out
blastn -db w7 -query FBtr0100589 -outfmt 7 -out w7.out
blastn -db w8 -query FBtr0100589 -outfmt 7 -out w8.out
blastn -db w9 -query FBtr0100589 -outfmt 7 -out w9.out
blastn -db w10 -query FBtr0100589 -outfmt 7 -out w10.out
blastn -db w11 -query FBtr0100589 -outfmt 7 -out w11.out
cat m89.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m89
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100589     FBtr0100589     100.000 1057    0       0       1       1057    1       1057    0.0
1953
# BLAST processed 1 queries
cat m90.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m90
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100589     FBtr0100590     99.693  976     2       1       83      1057    26      1001    0.0
1784
```

cat m91.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m91
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
1 hits found
FBtr0100589 FBtr0100591 100.000 1057 0 0 1 1057 1 1057 0.0
1953
BLAST processed 1 queries
cat m93.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m93
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
1 hits found
FBtr0100589 FBtr0100593 99.693 976 2 1 83 1057 26 1001 0.0
1784
BLAST processed 1 queries
cat m94.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m94
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
2 hits found
FBtr0100589 FBtr0100594 99.693 976 2 1 83 1057 256 1231 0.0
1784
FBtr0100589 FBtr0100594 100.000 92 0 0 1 92 1 92 4.21e-46
171
BLAST processed 1 queries
cat s1.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s1
# 0 hits found
# BLAST processed 1 queries
cat s2.out
BLASTN 2.10.0+

# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s2
# 0 hits found
# BLAST processed 1 queries
cat s3.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s3
# 0 hits found
# BLAST processed 1 queries
cat s4.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s4
# 0 hits found
# BLAST processed 1 queries
cat s5.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s5
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
1 hits found
FBtr0100589 XM_002035803.2 97.406 1041 20 1 6 1046 1 1034 0.0
1766
BLAST processed 1 queries
cat s6.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s6
# 0 hits found
# BLAST processed 1 queries
cat s7.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;

name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s7
# 0 hits found
# BLAST processed 1 queries
cat s8.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s8
# 0 hits found
# BLAST processed 1 queries
cat w1.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w1
# 0 hits found
# BLAST processed 1 queries
cat w2.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w2
# 0 hits found
# BLAST processed 1 queries
cat w3.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w3
# 0 hits found
# BLAST processed 1 queries
cat w4.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w4
# 0 hits found

BLAST processed 1 queries
cat w5.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w5
# 0 hits found
# BLAST processed 1 queries
cat w6.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w6
# 0 hits found
# BLAST processed 1 queries
cat w7.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w7
# 0 hits found
# BLAST processed 1 queries
cat w8.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w8
# 0 hits found
# BLAST processed 1 queries

cat w9.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w9
# 0 hits found
# BLAST processed 1 queries
cat w10.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;

name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w10
# 0 hits found
# BLAST processed 1 queries
cat w11.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: w11
# 0 hits found
# BLAST processed 1 queries
blastn -db s1 -query sXM_002030726.2 -outfmt 7 -out s1s.out
blastn -db s2 -query sXM_002030726.2 -outfmt 7 -out s2s.out
blastn -db s3 -query sXM_002030726.2 -outfmt 7 -out s3s.out
blastn -db s4 -query sXM_002030726.2 -outfmt 7 -out s4s.out
blastn -db s5 -query sXM_002030726.2 -outfmt 7 -out s5s.out
blastn -db s6 -query sXM_002030726.2 -outfmt 7 -out s6s.out
blastn -db s7 -query sXM_002030726.2 -outfmt 7 -out s7s.out
cat s1s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s1
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
1 hits found
XM_002030726.2 XM_002030726.2 100.000 943 0 0 1 943 1 943 0.0
1742
BLAST processed 1 queries
cat s2s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s2
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
1 hits found
XM_002030726.2 XM_002030727.2 78.591 738 153 4 68 801 48 784 3.42e-140
483
BLAST processed 1 queries

cat s3s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s3
0 hits found
BLAST processed 1 queries
cat s4s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s4
0 hits found
BLAST processed 1 queries
cat s5s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA

Database: s5
0 hits found
BLAST processed 1 queries
cat s6s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s6
0 hits found
BLAST processed 1 queries
cat s7s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s7
0 hits found
BLAST processed 1 queries
cat s8s.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: s8
0 hits found
BLAST processed 1 queries
blastn -db w1 -query XM_002060947.3 -outfmt 7 -out w1w.out
blastn -db w2 -query XM_002060947.3 -outfmt 7 -out w2w.out
blastn -db w3 -query XM_002060947.3 -outfmt 7 -out w3w.out
blastn -db w4 -query XM_002060947.3 -outfmt 7 -out w4w.out
blastn -db w5 -query XM_002060947.3 -outfmt 7 -out w5w.out
blastn -db w6 -query XM_002060947.3 -outfmt 7 -out w6w.out
blastn -db w7 -query XM_002060947.3 -outfmt 7 -out w7w.out
blastn -db w8 -query XM_002060947.3 -outfmt 7 -out w8w.out
blastn -db w9 -query XM_002060947.3 -outfmt 7 -out w9w.out
blastn -db w10 -query XM_002060947.3 -outfmt 7 -out w10w.out
blastn -db w11 -query XM_002060947.3 -outfmt 7 -out w11w.out
cat w1w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w1
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score
1 hits found
XM_002060947.3 XM_002060947.3 100.000 1659 0 0 1 1659 1 1659 0.0 3064
BLAST processed 1 queries
cat w2w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w2
0 hits found
BLAST processed 1 queries
cat w3w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w3
0 hits found
cat w4w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA

Database: w4
0 hits found
BLAST processed 1 queries
cat w5w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w5
0 hits found
BLAST processed 1 queries
cat w6w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w6
0 hits found
BLAST processed 1 queries
[whu12@bluehive transcripts]$ cat w7w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w7
0 hits found
BLAST processed 1 queries
cat w8w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w8
0 hits found
BLAST processed 1 queries
cat w9w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w9
0 hits found
BLAST processed 1 queries
cat w10w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w10
0 hits found
BLAST processed 1 queries
cat w11w.out
BLASTN 2.10.0+
Query: XM_002060947.3 PREDICTED: Drosophila willistoni zinc-type alcohol dehydrogenase-like protein C1773.06c (LOC6637963), transcript variant X1, mRNA
Database: w11
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score
1 hits found
XM_002060947.3 XM_023181840.1 100.000 1587 0 0 73 1659 193 1779 0.0 2931
BLAST processed 1 queries
blastn -db w1 -query XM_002060947.3 -outfmt 7 -out w1w.out
blastn -db m89 -query FBtr0100590 -outfmt 7 -out m89-2.out

```
blastn -db m90 -query FBtr0100590 -outfmt 7 -out m90-2.out
blastn -db m91 -query FBtr0100591 -outfmt 7 -out m91-2.out
cat m89-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m89
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100590     FBtr0100589     99.693  976     2       1       26      1001    83      1057    0.0
1784
# BLAST processed 1 queries
cat m90-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: m90
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100590     FBtr0100590     100.000 1001    0       0       1       1001    1       1001    0.0
1849
# BLAST processed 1 queries
cat m91-2.out
# BLASTN 2.10.0+
# Query: FBtr0100591 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617643,14618070..14618474,14618526.
.14618902); ID=FBtr0100591; name=Adh-RF;
dbxref=FlyBase_Annotation_IDs:CG3481-RF,FlyBase:FBtr0100591,REFSEQ:NM_001032098,REFSEQ:NM_001032098; score=15;
score_text=Strongly Supported; MD5=5ad8501fe5600bed3c74c20f78fb9a19; length=2080; parent=FBgn0000055; release=r6.13;
species=Dmel;
# Database: m91
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100591     FBtr0100591     100.000 2080    0       0       1       2080    1       2080    0.0
3842
# BLAST processed 1 queries
blastn -db m93 -query FBtr0100591 -outfmt 7 -out m93-2.out
blastn -db m94 -query FBtr0100591 -outfmt 7 -out m94-2.out
cat m93-2.out
# BLASTN 2.10.0+
# Query: FBtr0100591 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617643,14618070..14618474,14618526.
.14618902); ID=FBtr0100591; name=Adh-RF;
dbxref=FlyBase_Annotation_IDs:CG3481-RF,FlyBase:FBtr0100591,REFSEQ:NM_001032098,REFSEQ:NM_001032098; score=15;
score_text=Strongly Supported; MD5=5ad8501fe5600bed3c74c20f78fb9a19; length=2080; parent=FBgn0000055; release=r6.13;
species=Dmel;
# Database: m93
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100591     FBtr0100593     99.850  1999    2       1       83      2080    26      2024    0.0
3674
# BLAST processed 1 queries
cat m94.out
# BLASTN 2.10.0+
# Query: FBtr0100589 type=mRNA;
loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100589;
name=Adh-RC;
dbxref=FlyBase_Annotation_IDs:CG3481-RC,FlyBase:FBtr0100589,REFSEQ:NM_001032095,REFSEQ:NM_001032095;
score=15; score_text=Strongly Supported; MD5=02c66ae3a7e25f9cad76813bf677789b; length=1057; parent=FBgn0000055;
release=r6.13; species=Dmel;
```

Database: m94
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score
2 hits found
FBtr0100589 FBtr0100594 99.693 976 2 1 83 1057 256 1231 0.0 1784
FBtr0100589 FBtr0100594 100.000 92 0 0 1 92 1 92 4.21e-46 171
BLAST processed 1 queries
blastn -db s1 -query FBtr0100591 -outfmt 7 -out s1-2.out
cat s1-2.out
BLASTN 2.10.0+
Query: FBtr0100591 type=mRNA; loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617643,14618070..14618474,14618526..14618902); ID=FBtr0100591; name=Adh-RF; dbxref=FlyBase_Annotation_IDs:CG3481-RF,FlyBase:FBtr0100591,REFSEQ:NM_001032098,REFSEQ:NM_001032098; score=15; score_text=Strongly Supported; MD5=5ad8501fe5600bed3c74c20f78fb9a19; length=2080; parent=FBgn0000055; release=r6.13; species=Dmel;
Database: s1
0 hits found
BLAST processed 1 queries
blastn -db s2 -query FBtr0100591 -outfmt 7 -out s2-2.out
cat s2-2.out
BLASTN 2.10.0+
Query: FBtr0100591 type=mRNA; loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617643,14618070..14618474,14618526..14618902); ID=FBtr0100591; name=Adh-RF; dbxref=FlyBase_Annotation_IDs:CG3481-RF,FlyBase:FBtr0100591,REFSEQ:NM_001032098,REFSEQ:NM_001032098; score=15; score_text=Strongly Supported; MD5=5ad8501fe5600bed3c74c20f78fb9a19; length=2080; parent=FBgn0000055; release=r6.13; species=Dmel;
Database: s2
0 hits found
BLAST processed 1 queries
blastn -db s3 -query FBtr0100591 -outfmt 7 -out s3-2.out
cat s3-2.out
BLASTN 2.10.0+
Query: FBtr0100591 type=mRNA; loc=2L:join(14615552..14615643,14616303..14616437,14616503..14616907,14616978..14617643,14618070..14618474,14618526..14618902); ID=FBtr0100591; name=Adh-RF; dbxref=FlyBase_Annotation_IDs:CG3481-RF,FlyBase:FBtr0100591,REFSEQ:NM_001032098,REFSEQ:NM_001032098; score=15; score_text=Strongly Supported; MD5=5ad8501fe5600bed3c74c20f78fb9a19; length=2080; parent=FBgn0000055; release=r6.13; species=Dmel;
Database: s3
0 hits found
BLAST processed 1 queries
blastn -db s3 -query FBtr0100590 -outfmt 7 -out s3-2.out
cat s3-2.out
BLASTN 2.10.0+
Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100590; name=Adh-RE; dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099; score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055; release=r6.13; species=Dmel;
Database: s3
0 hits found
BLAST processed 1 queries
blastn -db s1 -query FBtr0100590 -outfmt 7 -out s1-2.out
car s1-2.out
-bash: car: command not found
cat s1-2.out
BLASTN 2.10.0+
Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402); ID=FBtr0100590; name=Adh-RE; dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099; score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055; release=r6.13; species=Dmel;
Database: s1
0 hits found
BLAST processed 1 queries
blastn -db s2 -query FBtr0100590 -outfmt 7 -out s2-2.out

```
cat s2-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s2
# 0 hits found
# BLAST processed 1 queries
blastn -db s3 -query FBtr0100590 -outfmt 7 -out s3-2.out
cat s3-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s3
# 0 hits found
# BLAST processed 1 queries
blastn -db s4 -query FBtr0100590 -outfmt 7 -out s4-2.out
cat s4-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s4
# 0 hits found
# BLAST processed 1 queries
blastn -db s5 -query FBtr0100590 -outfmt 7 -out s5-2.out
cat s5-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s5
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 1 hits found
FBtr0100590     XM_002035803.2  96.995  965     21      2       26      990     78      1034    0.0
1615
# BLAST processed 1 queries
[whu12@bluehive transcripts]$ blastn -db s6 -query FBtr0100590 -outfmt 7 -out s6-2.out
[whu12@bluehive transcripts]$ cat s6-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s6
# 0 hits found
# BLAST processed 1 queries
blastn -db s7 -query FBtr0100590 -outfmt 7 -out s7-2.out
cat s7-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s7
# 0 hits found
# BLAST processed 1 queries
blastn -db s8 -query FBtr0100590 -outfmt 7 -out s8-2.out
```

```
cat s8-2.out
# BLASTN 2.10.0+
# Query: FBtr0100590 type=mRNA; loc=2L:join(14616267..14616437,14616503..14616907,14616978..14617402);
ID=FBtr0100590; name=Adh-RE;
dbxref=FlyBase_Annotation_IDs:CG3481-RE,FlyBase:FBtr0100590,REFSEQ:NM_001032099,REFSEQ:NM_001032099;
score=15; score_text=Strongly Supported; MD5=556ae0fba390fa23aeb53d06250e7387; length=1001; parent=FBgn0000055;
release=r6.13; species=Dmel;
# Database: s8
# 0 hits found
# BLAST processed 1 queries
blastn -db w1 -query XM_002035803.2 -outfmt 7 -out w1-2.out
Command line argument error: Argument "query". File is not accessible:  `XM_002035803.2'
[whu12@bluehive transcripts]$ blastn -db w1 -query sXM_002035803.2 -outfmt 7 -out w1-2.out
[whu12@bluehive transcripts]$ cat w1-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w1
# 0 hits found
# BLAST processed 1 queries
blastn -db w2 -query sXM_002035803.2 -outfmt 7 -out w2-2.out
cat w2-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w2
# 0 hits found
# BLAST processed 1 queries
blastn -db w3 -query sXM_002035803.2 -outfmt 7 -out w3-2.out
blastn -db w4 -query sXM_002035803.2 -outfmt 7 -out w4-2.out
blastn -db w5 -query sXM_002035803.2 -outfmt 7 -out w5-2.out
blastn -db w6 -query sXM_002035803.2 -outfmt 7 -out w6-2.out
blastn -db w7 -query sXM_002035803.2 -outfmt 7 -out w7-2.out
blastn -db w8 -query sXM_002035803.2 -outfmt 7 -out w8-2.out
blastn -db w9 -query sXM_002035803.2 -outfmt 7 -out w9-2.out
blastn -db w10 -query sXM_002035803.2 -outfmt 7 -out w10-2.out
blastn -db w11 -query sXM_002035803.2 -outfmt 7 -out w11-2.out
cat w2-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w2
# 0 hits found
# BLAST processed 1 queries
[whu12@bluehive transcripts]$ cat w3-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w3
# 0 hits found
# BLAST processed 1 queries
[whu12@bluehive transcripts]$ cat w4-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w4
# 0 hits found
# BLAST processed 1 queries
[whu12@bluehive transcripts]$ cat w5-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w5
# 0 hits found
# BLAST processed 1 queries
[whu12@bluehive transcripts]$ cat w6-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w6
# 0 hits found
# BLAST processed 1 queries
cat w7-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w7
# 0 hits found
```

```
# BLAST processed 1 queries
cat w8-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w8
# 0 hits found
# BLAST processed 1 queries
cat w9-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w9
# 0 hits found
# BLAST processed 1 queries
cat w10-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w10
# 0 hits found
# BLAST processed 1 queries
cat w11-2.out
# BLASTN 2.10.0+
# Query: XM_002035803.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase (LOC6611292), mRNA
# Database: w11
# 0 hits found
# BLAST processed 1 queries
blastn -db w1 -query XM_002030726.2 -outfmt 7 -out w1-3.out
Command line argument error: Argument "query". File is not accessible:  `XM_002030726.2'
blastn -db w1 -query sXM_002030726.2 -outfmt 7 -out w1-3.out
blastn -db w2 -query sXM_002030726.2 -outfmt 7 -out w2-3.out
cat w1-3.out
# BLASTN 2.10.0+
# Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
# Database: w1
# 0 hits found
# BLAST processed 1 queries
cat w2-3.out
# BLASTN 2.10.0+
# Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
# Database: w2
# 0 hits found
# BLAST processed 1 queries
blastn -db w3 -query sXM_002030726.2 -outfmt 7 -out w3-3.out
blastn -db w4 -query sXM_002030726.2 -outfmt 7 -out w4-3.out
blastn -db w5 -query sXM_002030726.2 -outfmt 7 -out w5-3.out
blastn -db w6 -query sXM_002030726.2 -outfmt 7 -out w6-3.out
blastn -db w7 -query sXM_002030726.2 -outfmt 7 -out w7-3.out
blastn -db w8 -query sXM_002030726.2 -outfmt 7 -out w8-3.out
blastn -db w9 -query sXM_002030726.2 -outfmt 7 -out w9-3.out
blastn -db w10 -query sXM_002030726.2 -outfmt 7 -out w10-3.out
blastn -db w11 -query sXM_002030726.2 -outfmt 7 -out w11-3.out
cat w3-3.out
# BLASTN 2.10.0+
# Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
# Database: w3
# 0 hits found
# BLAST processed 1 queries
cat w4-3.out
# BLASTN 2.10.0+
# Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
# Database: w4
# 0 hits found
# BLAST processed 1 queries
cat w5-3.out
# BLASTN 2.10.0+
# Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
# Database: w5
# 0 hits found
# BLAST processed 1 queries
cat w6-3.out
# BLASTN 2.10.0+
```

Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: w6
0 hits found
BLAST processed 1 queries
cat w7-3.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: w7
0 hits found
BLAST processed 1 queries
cat w8-3.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: w8
0 hits found
BLAST processed 1 queries
cat w9-3.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: w9
0 hits found
BLAST processed 1 queries
cat w10-3.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: w10
0 hits found
BLAST processed 1 queries
cat w11-3.out
BLASTN 2.10.0+
Query: XM_002030726.2 PREDICTED: Drosophila sechellia alcohol dehydrogenase 1 (LOC6605950), mRNA
Database: w11
0 hits found
BLAST processed 1 queries

files:

m89.ndb; m89.nhr; m89.nin; m89.nog; m89.nos; m89.not; m89.nsq; m89.ntf; m89.nto
m90.ndb; m90.nhr; m90.nin; m90.nog; m90.nos; m90.not; m90.nsq; m90.ntf; m90.nto
m91.ndb; m91.nhr; m91.nin; m91.nog; m91.nos; m91.not; m91.nsq; m91.ntf; m91.nto
m93.ndb; m93.nhr; m93.nin; m93.nog; m93.nos; m93.not; m93.nsq; m93.ntf; m93.nto
m94.ndb; m94.nhr; m94.nin; m94.nog; m94.nos; m94.not; m94.nsq; m94.ntf; m94.nto
s1.ndb; s1.nhr; s1.nin; s1.nog; s1.nos; s1.not; s1.nsq; s1.ntf; s1.nto
s2.ndb; s2.nhr; s2.nin; s2.nog; s2.nos; s2.not; s2.nsq; s2.ntf; s2.nto
s3.ndb; s3.nhr; s3.nin; s3.nog; s3.nos; s3.not; s3.nsq; s3.ntf; s3.nto
s4.ndb; s4.nhr; s4.nin; s4.nog; s4.nos; s4.not; s4.nsq; s4.ntf; s4.nto
s5.ndb; s5.nhr; s5.nin; s5.nog; s5.nos; s5.not; s5.nsq; s5.ntf; s5.nto
s6.ndb; s6.nhr; s6.nin; s6.nog; s6.nos; s6.not; s6.nsq; s6.ntf; s6.nto
s7.ndb; s7.nhr; s7.nin; s7.nog; s7.nos; s7.not; s7.nsq; s7.ntf; s7.nto
s8.ndb; s8.nhr; s8.nin; s8.nog; s8.nos; s8.not; s8.nsq; s8.ntf; s8.nto
w1.ndb; w1.nhr; w1.nin; w1.nog; w1.nos; w1.not; w1.nsq; w1.ntf; w1.nto
w2.ndb; w2.nhr; w2.nin; w2.nog; w2.nos; w2.not; w2.nsq; w2.ntf; w2.nto
w3.ndb; w3.nhr; w3.nin; w3.nog; w3.nos; w3.not; w3.nsq; w3.ntf; w3.nto
w4.ndb; w4.nhr; w4.nin; w4.nog; w4.nos; w4.not; w4.nsq; w4.ntf; w4.nto
w5.ndb; w5.nhr; w5.nin; w5.nog; w5.nos; w5.not; w5.nsq; w5.ntf; w5.nto
w6.ndb; w6.nhr; w6.nin; w6.nog; w6.nos; w6.not; w6.nsq; w6.ntf; w6.nto
w7.ndb; w7.nhr; w7.nin; w7.nog; w7.nos; w7.not; w7.nsq; w7.ntf; w7.nto
w8.ndb; w8.nhr; w8.nin; w8.nog; w8.nos; w8.not; w8.nsq; w8.ntf; w8.nto
w9.ndb; w9.nhr; w9.nin; w9.nog; w9.nos; w9.not; w9.nsq; w9.ntf; w9.nto
w10.ndb; w10.nhr; w10.nin; w10.nog; w10.nos; w10.not; w10.nsq; w10.ntf; w10.nto
w11.ndb; w11.nhr; w11.nin; w11.nog; w11.nos; w11.not; w11.nsq; w11.ntf; w11.nto
m89.out; m90.out; m91.out; m93.out; m94.out
s1.out; s2.out; s3.out; s4.out; s5.out; s6.out; s7.out; s8.out
w1.out; w2.out; w3.out; w4.out; w5.out; w6.out; w7.out; w8.out; w9.out; w10.out; w11.out
s1s.out; s2s.out; s3s.out; s4s.out; s5s.out; s6s.out; s7s.out; s8s.out
w1w.out; w2w.out; w3w.out; w4w.out; w5w.out; w6w.out; w7w.out; w8w.out; w9w.out; w10w.out; w11w.out

Alignment files:

FBtr0100589 FBtr0100589 100.000 1057 0 0 1 1057 1 1057 0.0 1953
FBtr0100589 FBtr0100590 99.693 976 2 1 83 1057 26 1001 0.0 1784
FBtr0100589 FBtr0100591 100.000 1057 0 0 1 1057 1 1057 0.0 1953
FBtr0100589 FBtr0100593 99.693 976 2 1 83 1057 26 1001 0.0 1784
FBtr0100589 FBtr0100594 99.693 976 2 1 83 1057 256 1231 0.0 1784
FBtr0100589 FBtr0100594 100.000 92 0 0 1 92 1 92 4.21e-46 171
FBtr0100589 XM_002035803.2 97.406 1041 20 1 6 1046 1 1034 0.0 1766

>FBtr0100589:1-1057
CATGCATTATTGTCTCAGTGCAGTTGTCAGTTGCAGTTCAGCAGACGGGCTAACGAGTAC
TTGCATCTCTTCAAATTTACTTAATTGATCAAATCGAAAGAGCCTGCTAAAGCAAAAAAG
AAGTCACCATGTCGTTTACTTTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAG
GCATTGGTCTGGACACCAGCAAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCC
TCGACCGCATTGAGAACCCGGCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGA
CCGTCACCTTCTACCCCTATGATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGA
AGACCATCTTCGCCCAGCTGAAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGG
ACGATCACCAGATCGAGCGCACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGA
CGGCCATTCTGGACTTCTGGGACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACA
TTGGATCCGTCACTGGATTCAATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGG
CCGCCGTGGTCAACTTCACCAGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGG
CTTACACTGTGAACCCCGGCATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGT
TGGATGTTGAGCCTCAGGTTGCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGG
CCTGCGCCGAGAACTTCGTCAAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAAC
TGGACTTGGGCACCCTGGAGGCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAG
AAGTGATACTCCCAAAAAAAAAAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCA
GAAGATATTCACGCAAGGCAATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAAT
ACGATAATAAAACTTTCCATGAAAAATATGGAAAAAT

>FBtr0100590:83-1057
CTTTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCA
GCAAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACC
CGGCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCT
ATGATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGC
TGAAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGC
GCACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCT
GGGACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGAT
TCAATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCA
CCAGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCG
GCATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGG
TTGCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCG
TCAAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGG
AGGCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAAAA
AAAAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCAGAAGATATTCACGCAAGG
CAATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAATACGATAATAAAACTTTCC
ATGAAAAATATGGAAAAAT

>FBtr0100591:1-1057
CATGCATTATTGTCTCAGTGCAGTTGTCAGTTGCAGTTCAGCAGACGGGCTAACGAGTAC
TTGCATCTCTTCAAATTTACTTAATTGATCAAATCGAAAGAGCCTGCTAAAGCAAAAAAG
AAGTCACCATGTCGTTTACTTTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAG
GCATTGGTCTGGACACCAGCAAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCC
TCGACCGCATTGAGAACCCGGCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGA
CCGTCACCTTCTACCCCTATGATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGA
AGACCATCTTCGCCCAGCTGAAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGG
ACGATCACCAGATCGAGCGCACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGA
CGGCCATTCTGGACTTCTGGGACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACA
TTGGATCCGTCACTGGATTCAATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGG
CCGCCGTGGTCAACTTCACCAGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGG
CTTACACTGTGAACCCCGGCATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGT
TGGATGTTGAGCCTCAGGTTGCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGG
CCTGCGCCGAGAACTTCGTCAAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAAC
TGGACTTGGGCACCCTGGAGGCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAG
AAGTGATACTCCCAAAAAAAAAAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCA
GAAGATATTCACGCAAGGCAATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAAT

ACGATAATAAAACTTTCCATGAAAAATATGGAAAAAT
>FBtr0100593:83-1057
CTTTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCA
GCAAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACC
CGGCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCT
ATGATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGC
TGAAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGC
GCACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCT
GGGACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGAT
TCAATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCA
CCAGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCG
GCATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGG
TTGCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCG
TCAAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGG
AGGCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAAAA
AAAAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCAGAAGATATTCACGCAAGG
CAATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAATACGATAATAAAACTTTCC
ATGAAAAATATGGAAAAATATATGAAAATTGAGAAATCCAAAAAACTGATAAACGCTCTA
CTTAATTAAAATAGA
>FBtr0100594:83-1057
AATTGATCAAGCAGCGCTGCCGTCGCCGGCTGAGCAGCCTGCGTACATAGCCGAGATCGC
GTAACGGTAGATAATGAAAAGCTCTACGTAACCGAAGCTTCTGCTGTACGGATCTTCCTA
TAAATACGGGGCCGACACGAACTGGAAACCAACAACTAACGGAGCCCTCTTCCAATTGAA
ACAGATCGAAAGAGCCTGCTAAAGCAAAAAAGAAGTCACCATGTCGTTTACTTTGACCAA
CAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCAGCAAGGAGCT
GCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACCCGGCTGCCAT
TGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCTATGATGTGAC
CGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGCTGAAGACCGT
CGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGCGCACCATTGC
CGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCTGGGACAAGCG
CAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGATTCAATGCCAT
CTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCACCAGCTCCCT
GGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCGGCATCACCCG
CACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGGTTGCCGAGAA
GCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCGTCAAGGCTAT
CGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAGGCCATCCA
GTGGACCAAGCACTG
>FBtr0100594:1-92
CATGCATTATTGTCTCAGTGCAGTTGTCAGTTGCAGTTCAGCAGACGGGCTAACGAGTAC
TTGCATCTCTTCAAATTTACTTAATTGATCAA
>XM_002035803.2:1-1034
ATTATTGTCTCAGTGCAGTTGTCAGTTGCAGTTCAGCAGACCGGCTAACGAGTACTTGCA
TCTCTTCAAATTTACTTAATTGATCAAATCGAAAGAGCCTGCTAAAGCAAAAAAGAAGTC
ACCATGGCGTTTACTTTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATT
GGTCTGGACACCAGCAAGGAGCTGCTCAAGCGCGACCTGAAGAACCTGGTGATCCTCGAC
CGCATTGAGAACCCTGCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTC
ACCTTCTACCCCTATGATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACC
ATCTTCGCCAAGCTGAAGACCGTCGATGTCCTGATCAACGGAGCTGGCATCCTGGACGAT
CACCAGATCGAGCGCACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCC
ATTCTGGACTTCTGGGACAAGCGCAAGGGTGGTCCCGGTGGTATCATCTGCAACATTGGA
TCCGTCACTGGATTCAATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCC
GTGGTCAACTTCACCAGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACCGCTTAC
ACCGTGAACCCCGGCATCACCCGCACCACCCTGGTGCACAAGTTCAACTCCTGGCTGGAT
GTTGAGCCCCAGGTTGCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCATTGGCCTGC
GCCGAGAACTTCGTCAAGGCCATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGAC
TTGGGCACCCTGGAGGCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTG
ATACTCCCAAGAAACACATAACATTAGTTCATAGGGTTCTGCGAACCACAAGATATTCAC
GCAAGGCAATAAGGCTGATTCGAAGCACACTCACATTCTTCCCCTAATACGATAATAAAA
CTCTCCATGAAAAA
>FBtr0100594:83-1057
AATTGATCAAGCAGCGCTGCCGTCGCCGGCTGAGCAGCCTGCGTACATAGCCGAGATCGC
GTAACGGTAGATAATGAAAAGCTCTACGTAACCGAAGCTTCTGCTGTACGGATCTTCCTA
TAAATACGGGGCCGACACGAACTGGAAACCAACAACTAACGGAGCCCTCTTCCAATTG--
------AAACAGATCGAAAGAGCCTGCTAAAGCAAAAAAGAAGTCACCATGTCGTTTACT
TTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCAGC
AAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACCCG
GCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCTAT
GATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGCTG
AAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGCGC
ACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCTGG

GACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGATTC
AATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCACC
AGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCGGC
ATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGGTT
GCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCGTC
AAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAG
GCCATCCAGTGGACCAAGCACTG----------

>XM_002035803.2:1-1034

----------ATTATT--GTCTC----------AGTGCAGTTGTCAGTT
GCAGTTCA--GCA-----G-ACCGGC---TAACGAGTACT-TGCATCTCTTCAAATTTAC
TTAATTGATCAAATCGAAAGAGCCTGCTAAAGCAAAAAAGAAGTCACCATGGCGTTTACT
TTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCAGC
AAGGAGCTGCTCAAGCGCGACCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACCCT
GCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCTAT
GATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCAAGCTG
AAGACCGTCGATGTCCTGATCAACGGAGCTGGCATCCTGGACGATCACCAGATCGAGCGC
ACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCTGG
GACAAGCGCAAGGGTGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGATTC
AATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCACC
AGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACCGCTTACACCGTGAACCCCGGC
ATCACCCGCACCACCCTGGTGCACAAGTTCAACTCCTGGCTGGATGTTGAGCCCCAGGTT
GCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCATTGGCCTGCGCCGAGAACTTCGTC
AAGGCCATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAG
GCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAGA---
----AACACATAACATTAGTTCATAGGGTTCTGCGAACCACAAGATATTCACGCAAGGCA
ATAAGGCTGATTCGAAGCACACTCACATTCTTCCCCTAATACGATAATAAAACTCTCCAT
GAAAAA----------

>FBtr0100594:1-92

----------CATGCATTATT--GTCTC----------AGTGCAGTTGTCAGTT
GCAGTTCA--GCA-----G-ACGGGC---TAACGAGTACT-TGCATCTCTTCAAATTTAC
TTAATTGATCAA----------

GCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCGTC
AAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAG
GCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAAAAAA
AAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCAGAAGATATTCACGCAAGGCA
ATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAATACGATAATAAAACTTTCCAT
GAAAAATATGGAAAAATATATGAAAATTGAGAAATCCAAAAAACTGATAAACGCTCTACT
TAATTAAAATAGA
>FBtr0100589:1-1057

------------CATGCATTATT--GTCTC--------------AGTGCAGTTGTCAGTT
GCAGTTCA--GCA-----G-ACGGGC---TAACGAGTACT-TGCATCTCTTCAAATTTAC
TTAATTGATCAAATCGAAAGAGCCTGCTAAAGCAAAAAAGAAGTCACCATGTCGTTTACT
TTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCAGC
AAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACCCG
GCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCTAT
GATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGCTG
AAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGCGC
ACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCTGG
GACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGATTC
AATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCACC
AGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCGGC
ATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGGTT
GCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCGTC
AAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAG
GCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAAAAAA
AAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCAGAAGATATTCACGCAAGGCA
ATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAATACGATAATAAAACTTTCCAT
GAAAAATATGGAAAAAT----------

>FBtr0100590:83-1057

--CT
TTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCAGC
AAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACCCG
GCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCTAT
GATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGCTG
AAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGCGC
ACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCTGG
GACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGATTC
AATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCACC
AGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCGGC
ATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGGTT
GCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCGTC
AAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAG
GCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAAAAAA
AAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCAGAAGATATTCACGCAAGGCA
ATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAATACGATAATAAAACTTTCCAT
GAAAAATATGGAAAAAT----------

>FBtr0100591:1-1057

------------CATGCATTATT--GTCTC--------------AGTGCAGTTGTCAGTT
GCAGTTCA--GCA-----G-ACGGGC---TAACGAGTACT-TGCATCTCTTCAAATTTAC
TTAATTGATCAAATCGAAAGAGCCTGCTAAAGCAAAAAAGAAGTCACCATGTCGTTTACT
TTGACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATTGGTCTGGACACCAGC
AAGGAGCTGCTCAAGCGCGATCTGAAGAACCTGGTGATCCTCGACCGCATTGAGAACCCG
GCTGCCATTGCCGAGCTGAAGGCAATCAATCCAAAGGTGACCGTCACCTTCTACCCCTAT
GATGTGACCGTGCCCATTGCCGAGACCACCAAGCTGCTGAAGACCATCTTCGCCCAGCTG
AAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGATCACCAGATCGAGCGC
ACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACTTCTGG
GACAAGCGCAAGGGCGGTCCCGGTGGTATCATCTGCAACATTGGATCCGTCACTGGATTC
AATGCCATCTACCAGGTGCCCGTCTACTCCGGCACCAAGGCCGCCGTGGTCAACTTCACC
AGCTCCCTGGCGAAACTGGCCCCCATTACCGGCGTGACGGCTTACACTGTGAACCCCGGC
ATCACCCGCACCACCCTGGTGCACACGTTCAACTCCTGGTTGGATGTTGAGCCTCAGGTT
GCCGAGAAGCTCCTGGCTCATCCCACCCAGCCCTCGTTGGCCTGCGCCGAGAACTTCGTC
AAGGCTATCGAGCTGAACCAGAACGGAGCCATCTGGAAACTGGACTTGGGCACCCTGGAG
GCCATCCAGTGGACCAAGCACTGGGACTCCGGCATCTAAGAAGTGATACTCCCAAAAAAA
AAAAAAAACATAACATTAGTTCATAGGGTTCTGCGAACCAGAAGATATTCACGCAAGGCA

ATAAGGCTGATTCGATGCACACTCACATTCTTCTCCTAATACGATAATAAAACTTTCCAT
GAAAAATATGGAAAAAT----------

```
MAXIMUM LIKELIHOOD TREE
-----------------------

Log-likelihood of the tree: -2006.1835 (s.e. 34.7977)
Unconstrained log-likelihood (without tree): -3428.9395
Number of free parameters (#branches + #model parameters): 14
Akaike information criterion (AIC) score: 4040.3671
Corrected Akaike information criterion (AICc) score: 4040.7177
Bayesian information criterion (BIC) score: 4111.7790

Total tree length (sum of branch lengths): 0.0693
Sum of internal branch lengths: 0.0000 (0.0107% of tree length)

WARNING: 4 near-zero internal branches (<0.0008) should be treated with caution
         Such branches are denoted by '**' in the figure below

NOTE: Tree is UNROOTED although outgroup taxon 'FBtr0100594_83-1057' is drawn at root
Numbers in parentheses are SH-aLRT support (%) / ultrafast bootstrap support (%)

+------------------------------------------------------------FBtr0100594_83-1057
|
|        +----------------------XM_002035803.2_1-1034
|     +**| (0/55)
|     |  +**FBtr0100594_1-92
|  +**| (0/62)
|  |  |  +**FBtr0100589_1-1057
|  |  +**| (0/42)
|  |     +**FBtr0100591_1-1057
+**| (0/44)
|  +**FBtr0100590_83-1057
|
+**FBtr0100593_83-1057
```

December 8th Tuesday

We received the feedback from Xiaomi regarding the values of TPM on StringTie. So we created the python codes below to calculate the TPM for each of the two datasets (Qizhe Zhang):

```
ifile=open("a775.tab", "r")
lines=ifile.readlines()
s187=0
s801=0
s760=0
s015=0
s501=0
s502=0
s963=0
s232=0
s233=0
s282=0
for line in lines:
        if "6639187" in line:
                read=line.split("\t")
                s187+=float(read[len(read)-1])
        if "6638801" in line:
                read=line.split("\t")
                s801=float(read[len(read)-1])
```

```
		if "6638760" in line:
			read=line.split("\t")
			s760+=float(read[len(read)-1])
		if "6643015" in line:
			read=line.split("\t")
			s015+=float(read[len(read)-1])
		if "6643501" in line:
			read=line.split("\t")
			s501+=float(read[len(read)-1])
		if "6643502" in line:
			read=line.split("\t")
			s502+=float(read[len(read)-1])
		if "6637963" in line:
			read=line.split("\t")
			s963+=float(read[len(read)-1])
		if "6642232" in line:
			read=line.split("\t")
			s232+=float(read[len(read)-1])
		if "6642233" in line:
			read=line.split("\t")
			s233+=float(read[len(read)-1])
		if "6650282" in line:
			read=line.split("\t")
			s282+=float(read[len(read)-1])
print("6639187: ", s187)
print("6638801: ", s801)
print("6638760: ", s760)
print("6643015: ", s015)
print("6643501: ", s501)
print("6643502: ", s502)
print("6637963: ", s963)
print("6642232: ", s232)
print("6642233: ", s233)
print("6650282: ", s282)
ifile=open("a779.tab", "r")
lines=ifile.readlines()
s292=0
s293=0
s014=0
s950=0
s951=0
s899=0
for line in lines:
		if "6611292" in line:
			read=line.split("\t")
			s292+=float(read[len(read)-1])
		if "6611293" in line:
			read=line.split("\t")
			s293=float(read[len(read)-1])
		if "6608014" in line:
			read=line.split("\t")
			s014+=float(read[len(read)-1])
		if "6605950" in line:
			read=line.split("\t")
			s950+=float(read[len(read)-1])
		if "6605951" in line:
			read=line.split("\t")
			s951+=float(read[len(read)-1])
		if "6606899" in line:
			read=line.split("\t")
			s899+=float(read[len(read)-1])
```

```
print("6611292: ", s292)
print("6611293: ", s293)
print("6608014: ", s014)
print("6605950: ", s950)
print("6605951: ", s951)
print("6606899: ", s899)
```

We also made a blast of proteins (Nina Hu):

```
module load blast
makeblastdb -in Dm -input_type fasta -dbtype prot -title Dmel -parse_seqids -out Dm_adh
Building a new DB, current time: 12/08/2020 02:57:41
New DB name:   /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/Dm_adh
New DB title:  Dmel
Sequence type: Protein
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 6 sequences in 0.0203412 seconds.
makeblastdb -in Ds -input_type fasta -dbtype prot -title Dsec -parse_seqids -out Ds_adh
Building a new DB, current time: 12/08/2020 02:57:55
New DB name:   /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/Ds_adh
New DB title:  Dsec
Sequence type: Protein
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 8 sequences in 0.0198698 seconds.
makeblastdb -in Dw -input_type fasta -dbtype prot -title Dwil -parse_seqids -out Dw_adh
Building a new DB, current time: 12/08/2020 02:58:49
New DB name:   /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/Dw_adh
New DB title:  Dwil
Sequence type: Protein
Deleted existing Protein BLAST database named /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/Dw_adh
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 10 sequences in 0.019052 seconds.
makeblastdb -in Dm -dbtype prot -parse_seqids -out dmel
Building a new DB, current time: 12/08/2020 03:06:19
New DB name:   /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/dmel
New DB title:  Dm
Sequence type: Protein
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 6 sequences in 0.024251 seconds.
makeblastdb -in Ds -dbtype prot -parse_seqids -out dsec
Building a new DB, current time: 12/08/2020 03:11:27
New DB name:   /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/dsec
New DB title:  Ds
Sequence type: Protein
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 8 sequences in 0.0198531 seconds.
makeblastdb -in Dw -dbtype prot -parse_seqids -out dwil
Building a new DB, current time: 12/08/2020 03:11:35
New DB name:   /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/dwil
New DB title:  Dw
Sequence type: Protein
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 10 sequences in 0.022248 seconds.
blastp -db dsec -query Dm_3 -outfmt 7 -out ds.blast.out
blastp -db dwil -query Dm_3 -outfmt 7 -out dw.blast.out
cat dm.blast.out
# BLASTP 2.10.0+
# Query: NP_477209.2 type III alcohol dehydrogenase [Drosophila melanogaster]
# Database: dmel
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
# 5 hits found
NP_477209.2     NP_001027267.1  37.143  35      19      2       131     164     12      44      1.8
18.1
```

NP_477209.2 NP_001027268.1 37.143 35 19 2 131 164 12 44 1.8
18.1
NP_477209.2 NP_001027269.1 37.143 35 19 2 131 164 12 44 1.8
18.1
NP_477209.2 NP_001027270.1 37.143 35 19 2 131 164 12 44 1.8
18.1
NP_477209.2 NP_001027266.1 37.143 35 19 2 131 164 12 44 1.8
18.1
BLAST processed 1 queries
module load samtools
cat ds.blast.out
BLASTP 2.10.0+
Query: NP_477209.2 type III alcohol dehydrogenase [Drosophila melanogaster]
Database: dsec
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
2 hits found
NP_477209.2 XP_002031693.1 41.176 34 15 1 423 451 329 362 0.17
22.3
NP_477209.2 XP_002035839.1 37.143 35 19 2 131 164 12 44 2.9
18.1
BLAST processed 1 queries
cat dw.blast.out
BLASTP 2.10.0+
Query: NP_477209.2 type III alcohol dehydrogenase [Drosophila melanogaster]
Database: dwil
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
4 hits found
NP_477209.2 XP_002073181.1 29.730 37 25 1 423 458 330 366 0.25
21.9
NP_477209.2 XP_002073181.1 35.000 40 15 2 132 162 208 245 2.6
18.9
NP_477209.2 XP_015033661.1 34.286 35 20 2 131 164 10 42 6.8
17.3
NP_477209.2 XP_002066457.1 23.810 21 16 0 212 232 26 46 9.5
16.9
BLAST processed 1 queries
samtools faidx Dm_I NP_001027267.1:12-44 > dm267
samtools faidx Dm_H NP_001027268.1:12-44 > dm268
samtools faidx Dm_F NP_001027269.1:12-44 > dm269
samtools faidx Dm_E NP_001027270.1:12-44 > dm270
samtools faidx Dm_C NP_001027266.1:12-44 > dm266

makeblastdb -in Dm_3 -input_type fasta -dbtype prot -parse_seqids -out Dm_3adh

Building a new DB, current time: 12/08/2020 03:40:25
New DB name: /gpfs/fs2/scratch/bio257_2020/Users/group2_adh/blast/Dm_3adh
New DB title: Dm_3
Sequence type: Protein
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1 sequences in 0.0083859 seconds.

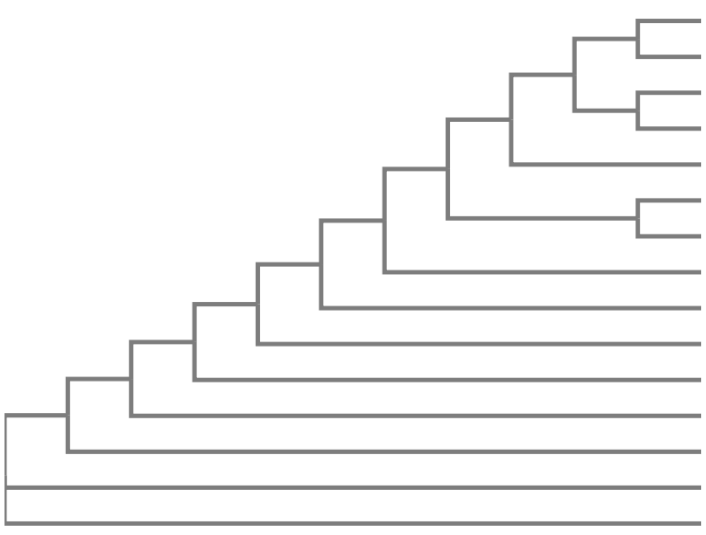
blastp -db Dm_3adh -query Dm_3 -outfmt 7 -out dm3.blast.out
cat dm3.blast.out
BLASTP 2.10.0+
Query: NP_477209.2 type III alcohol dehydrogenase [Drosophila melanogaster]
Database: Dm_3adh
Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end,
evalue, bit score
1 hits found
NP_477209.2 NP_477209.2 100.000 464 0 0 1 464 1 464 0.0
962
BLAST processed 1 queries
samtools faidx Ds_adh3 XP_002031693.1:329-362 > ds693
samtools faidx Ds_839 XP_002035839.1:12-44 > ds839
samtools faidx Dw_adh3 XP_002073181.1:330-366 > dw181

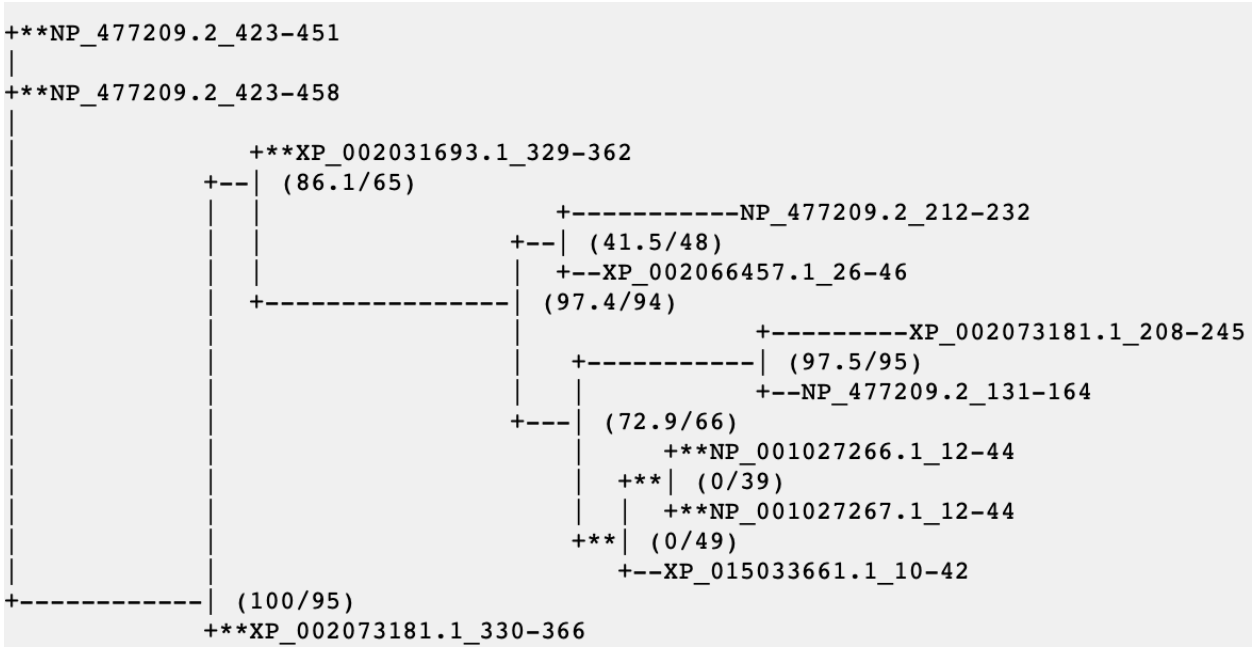
```
samtools faidx Dw_adh3 XP_002073181.1:208-245 > dw181_2
samtools faidx Dw_661 XP_015033661.1:10-42 > dw661
samtools faidx Dw_31kDa XP_002066457.1:26-46 > dw457
cat dm266 dm267 dm268 dm269 dm270 ds693 ds839 dw181 dw181_2 dw457 dw661 > allP
samtools faidx Dm_3 NP_477209.2:131-164 > dm3_1
samtools faidx Dm_3 NP_477209.2:423-451 > dm3_2
samtools faidx Dm_3 NP_477209.2:423-458 > dm3_2
samtools faidx Dm_3 NP_477209.2:212-232 > dm3_3
samtools faidx Dm_3 NP_477209.2:131-164 > dm3_1
samtools faidx Dm_3 NP_477209.2:423-451 > dm3_2
samtools faidx Dm_3 NP_477209.2:423-458 > dm3_3
samtools faidx Dm_3 NP_477209.2:212-232 > dm3_4
cat allP dm3_1 dm3_2 dm3_3 dm3_4 > prot
```

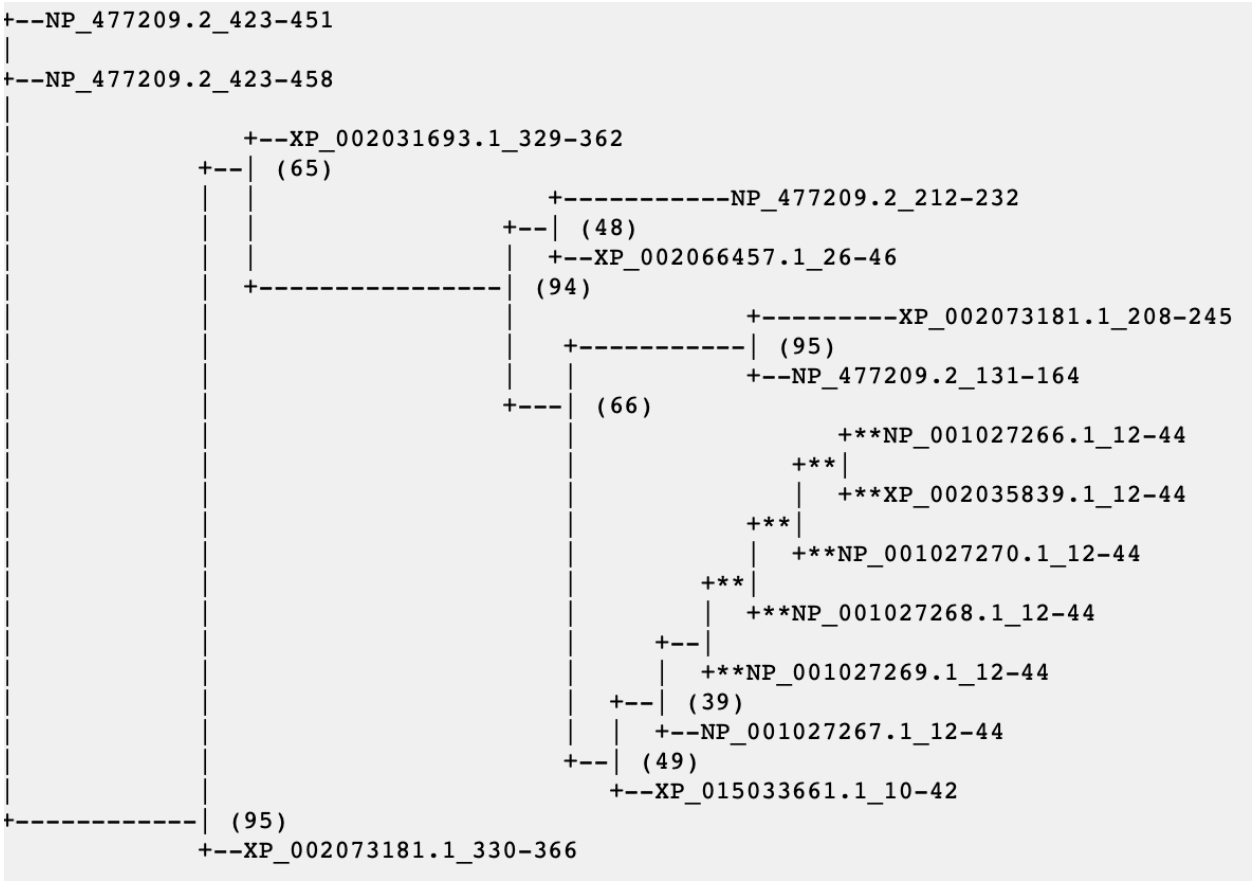
Output:

query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score

NP_477209.2	NP_001027267.1	37.143	35	19	2	131	164	12	44	1.8 18.1
NP_477209.2	NP_001027268.1	37.143	35	19	2	131	164	12	44	1.8 18.1
NP_477209.2	NP_001027269.1	37.143	35	19	2	131	164	12	44	1.8 18.1
NP_477209.2	NP_001027270.1	37.143	35	19	2	131	164	12	44	1.8 18.1
NP_477209.2	NP_001027266.1	37.143	35	19	2	131	164	12	44	1.8 18.1
NP_477209.2	XP_002031693.1	41.176	34	15	1	423	451	329	362	0.17 22.3
NP_477209.2	XP_002035839.1	37.143	35	19	2	131	164	12	44	2.9 18.1
NP_477209.2	XP_002073181.1	29.730	37	25	1	423	458	330	366	0.25 21.9
NP_477209.2	XP_002073181.1	35.000	40	15	2	132	162	208	245	2.6 18.9
NP_477209.2	XP_015033661.1	34.286	35	20	2	131	164	10	42	6.8 17.3
NP_477209.2	XP_002066457.1	23.810	21	16	0	212	232	26	46	9.5 16.9




+**NP_477209.2_423-451
+**NP_477209.2_423-458
+**XP_002031693.1_329-362
(86.1/65)
NP_477209.2_212-232
(41.5/48)
+--XP_002066457.1_26-46
(97.4/94)
XP_002073181.1_208-245
(97.5/95)
+--NP_477209.2_131-164
(72.9/66)
+**NP_001027266.1_12-44
+**| (0/39)
+**NP_001027267.1_12-44
+**| (0/49)
+--XP_015033661.1_10-42
(100/95)
+**XP_002073181.1_330-366


+--NP_477209.2_423-451
+--NP_477209.2_423-458
+--XP_002031693.1_329-362
(65)
NP_477209.2_212-232
(48)
+--XP_002066457.1_26-46
(94)
XP_002073181.1_208-245
(95)
+--NP_477209.2_131-164
(66)
+**NP_001027266.1_12-44
+**|
+**XP_002035839.1_12-44
+**|
+**NP_001027270.1_12-44
+**|
+**NP_001027268.1_12-44
+--|
+**NP_001027269.1_12-44
(39)
+--NP_001027267.1_12-44
(49)
+--XP_015033661.1_10-42
(95)
+--XP_002073181.1_330-366