**Downloading the structural RNAs files from Rfam database to filter from sRNA-seq**

Help: <https://docs.rfam.org/en/latest/sequence-extraction.html>

-STEP1: install easel (conda install easel)

-STEP2: download all Rfam fasta files from the database

wget ftp://ftp.ebi.ac.uk/pub/databases/Rfam/CURRENT/fasta\_files/\* .

-STEP3: unzip and combine into one file

gunzip \*.gz

cat \*.fa > Rfam.fa

-STEP4: index the fata file

esl-sfetch --index Rfam.fa

*If the indexing does not work because of repetitive fasta sequences, try:*

awk 'BEGIN{RS=">"; FS="\n"; ORS=""}

(FNR==1){next}

{ name=$1; seq=$0; gsub(/(^[^\n]\*|)\n/,"",seq) }

!(seen[name,seq]++){ print ">" $0 }' file1.fasta file2.fasta file3.fasta ...

-STEP5: create the .sql files that fetches the regions of interest

SELECT concat(fr.rfamseq\_acc,'/',seq\_start,'-',seq\_end)

FROM full\_region fr, rfamseq rf, taxonomy tx, family f

WHERE

rf.ncbi\_id = tx.ncbi\_id

AND f.rfam\_acc = fr.rfam\_acc

AND fr.rfamseq\_acc = rf.rfamseq\_acc

AND tx.ncbi\_id = 4577 ## ncbi ID of the organism or interest, this is Zea mays

AND f.type LIKE '%rRNA%' ## type of features to fetch

AND is\_significant = 1;

and gs.version=14.0;

-STEP6: Create .sql files for all type of structural RNAs you want to filter together

- STEP7: Get the corresponding accessions from mysql, replacing query.sql by the name of the previously created files. If using several files, run this command each time and concatenate the outputs

mysql -urfamro -hmysql-rfam-public.ebi.ac.uk -P4497 --skip-column-names --database Rfam < query.sql > accessions.txt

- STEP8: Extract the fasta sequences of the chosen accessions from the Rfam database

esl-sfetch -f path/to/Rfam.fa /path/to/accessions.txt > Rfam\_ncRNAs.fa

- STEP9: Build bowtie2 indexes on the newly created fasta file and map the trimmed library to it, then use the unmapped reads, where $threads is the number of threads to use, ${ref}/zm\_structural\_RNAs is the path and name of the fasta file and later of the bowtie 2 indexes, ${name}.fastq.gz is the name of your trimmed fastq file to filter and filtered\_${name}.fastq.gz is the results file you get.

bowtie2-build --threads $threads ${ref}/zm\_structural\_RNAs.fa ${ref}/zm\_structural\_RNAs

bowtie2 --very-sensitive -p $threads -x ${ref}/zm\_structural\_RNAs -U ${name}.fastq.gz | samtools view -@ $threads -f 0x4 | samtools fastq -@ $threads | gzip > filtered\_${name}.fastq.gz

- My files for maize:

Whole Rfam database and index (end of STEP4):

/grid/martienssen/home/jcahn/nlsas/Genomes/Rfam/Rfam.fa

/grid/martienssen/home/jcahn/nlsas/Genomes/Rfam/Rfam.fa.ssi

Mysql files (end of STEP6):

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm.rRNA.sql

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm.snoRNA.sql

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm.tRNA.sql

List of accessions (end of STEP7):

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm.accessions.txt

Final fasta file (end of STEP8):

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.fa

Bowtie2 indexes (end of STEP9):

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.1.bt2

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.2.bt2

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.3.bt2

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.4.bt2

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.rev.1.bt2

/grid/martienssen/home/jcahn/nlsas/Genomes/Zea\_mays/structural\_RNA/zm\_structural\_RNAs.rev.2.bt2