**Re-analysis using Trinity, RSEM and edgeR**

**PART ONE**

**First use Trinity to generate transcriptome for each isolate before and after being exposed to sub-lethal concentrations of mefenoxam:**

1. Trinity must be ran separately on each sample because it requires a lot of processing power. However, the three biological replicates must be included in that run. That’s why one must use a txt file indicating which files to use.
2. To run Trinity on sample X, one must have in the same folder or directory, three files: the script file for that sample, the txt file for that sample, and the fastq raw data files from that sample both before and after the acquisition of resistance along with their three biological replicates.
3. Use command: qsub scriptname.sh to run the script.
4. That will generate many files. We are interested in the one named Trinity.fasta
5. That file contains a fasta file with the transcriptome that Trinity generated.

**PART TWO**

**Then, use RSEM to generate transcript count files for each isolate and biological replicate before and after the acquisition of resistance:**

1. The RSEM analysis must be done independently for each biological replicate from each isolate both before and after the acquisition of resistance.
2. To run it, one must have in the same folder or directory, three files: the script for that isolate (and biological replicate), the fasta file obtained from Trinity for that isolate and the fastq raw sequencing file for that isolate (and biological replicate).
3. Use command: qsub scriptname.vi to run the script.
4. Do this with all biological replicates for all isolates and treatments, each in a different folder.
5. In each folder, one must obtain several files, being two of them called ‘.isoforms.results’ and ‘genes.results’. These files contain the read counts for each transcript, and gene, respectively.
6. Change each of these files’ names to indicate from which isolate, biological replicate and treatment, they belong.
7. Put all the files of one isolate in a new single folder.
8. In this file, create a txt file with the list of the .isoforms.results sample files for one isolate, name it **Transcript\_exp\_files**
9. Then, to create a matrix with all read counts for that isolate one must be in the folder containing all count files and:
   1. Run /hpcfs/apps/trinityrnaseq/2.4.0/util/abundance\_estimates\_to\_matrix.pl --est\_method RSEM --out\_prefix Trinity\_Isolate\_transc --quant\_files Transcript\_exp\_files
   2. Check matrix with the command

head -n20 Trinity\_Isolate\_transc.TMM.EXPR.matrix

* 1. Repeat the same but with all genes’ files.

1. Do this for all samples.
2. Create a txt named **‘Isolate\_List\_samples’** that contains two columns like this:
   1. Condition unique\_replicate\_name
   2. Condition unique\_replicate\_name
   3. Condition unique\_replicate\_name
   4. And so on…
3. Before running the DE analysis, check if your biological replicates make sense between them. To do this:
   1. Enter R again an install the package ‘qvalue’ using Bioconductor too.
   2. Run /hpcfs/apps/trinityrnaseq/2.4.0/Analysis/DifferentialExpression/PtR -m Trinity\_Isolate\_transc.counts.matrix -s Isolate\_List\_samples --log2 --compare\_replicates
   3. Download the pdfs named Isolate\_0-0.rep\_compare.pdf and Isolate\_100-100.rep\_compare.pdf, and see if the replicates make sense.
   4. Run /hpcfs/apps/trinityrnaseq/2.4.0/Analysis/DifferentialExpression/PtR -m Trinity\_Isolate\_transc.counts.matrix -s Isolate\_List\_samples --log2 --sample\_cor\_matrix
   5. See pdf named Trinity\_Isolate\_transc.counts.matrix.log2.sample\_cor\_matrix.pdf
   6. Then Run /hpcfs/apps/trinityrnaseq/2.4.0/Analysis/DifferentialExpression/PtR -m Trinity\_Isolate\_transc.counts.matrix -s Isolate\_List\_samples --log2 --CPM --prin\_comp 3

**PART THREE**

**Run EdgeR to determine differential expression for specific transcripts and/or genes:**

1. Load R using the command module load R/3.3.2mro
2. Write R to enter the console.
3. Install the package ‘edgeR’ by running these two commands:
   * 1. source("https://bioconductor.org/biocLite.R")
     2. biocLite("edgeR")
4. Quit the console using q()
5. For the DE analysis, stay in the folder from step 17 and run:
   1. */hpcfs/apps/trinityrnaseq/2.4.0/Analysis/DifferentialExpression/run\_DE\_analysis.pl --matrix Trinity\_Isolate\_transc.counts.matrix --samples\_file* ***Isolate\_List\_samples*** *--method edgeR --output edgeR\_transc\_Isolate*
   2. Enter the output folder, files with '\*.DE\_results' are the results.
   3. To see volcano plots download the file called: *Trinity\_Isolate\_transc.counts.matrix.Isolate\_0-0\_vs\_Isolate\_100-100.edgeR.DE\_results.MA\_n\_Volcano.pdf*
   4. Copy the files **Trinity\_Isolate\_transc.TMM.EXPR.matrix** and **Isolate\_List\_samples** files to that folder.
   5. Run this for selecting the DE genes using a P (FDR actually) of 0.001 and a fold change (C) of 2^(2) or 4-fold: */hpcfs/apps/trinityrnaseq/2.4.0/Analysis/DifferentialExpression/analyze\_diff\_expr.pl --matrix Trinity\_Isolate\_transc.TMM.EXPR.matrix --samples US23\_List\_samples -P 1e-3 -C 2*
   6. Use command *wc -l diffExpr.P1e-3\_C2.matrix* to count them.
   7. The file that contains these selected DE transcripts is called: *Trinity\_Isolate\_transc.counts.matrix.Isolate\_0-0\_vs\_Isolate\_100-100.edgeR.DE\_results.P1e-3\_C2.DE.subset*
   8. Run this to improve the heatmap’s colors */hpcfs/apps/trinityrnaseq/2.4.0/Analysis/DifferentialExpression/PtR -m diffExpr.P1e-3\_C2.matrix.log2.centered.dat -s Isolate\_List\_samples --gene\_dist euclidean --sample\_dist euclidean --heatmap --heatmap\_scale\_limits "-4,4"*
   9. It generates the PDF called: *diffExpr.P1e-3\_C2.matrix.log2.centered.dat.genes\_vs\_samples\_heatmap.pdf*
   10. Repeat for the gene level.

**PART FOUR**

**To annotate the DEG:**

1. Load each isolate’s transcriptome to Galaxy.
2. At FASTA manipulation, use FASTA to tabular to convert them to tabular, use 2 columns.
3. Then upload a list of the IDs of DE transcripts, just in one column and upload it as Tabular too.
4. Use the tool Compare data sets at Join, filter and so on to compare the fasta with it by comparing the first column which is the one of the IDs.
5. Check if that file has all the transcripts.
6. Use FASTA manipulation to do Tabular to Fasta and change that one to Fasta by using the column 1 as name and column 3 as sequence.
7. Download that list of FASTA of DE transcripts.
8. Upload it to BLast2Go and run all posible analyses there.

**Note:** For further info on how to use Trinity please see these links:

<https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Transcript-Quantification>

<https://github.com/trinityrnaseq/BernWorkshop2016/wiki/Day_1_Morning_DenovoAssembly>