# ANÁLISIS BIOINFORMÁTICO DE ENSAMBLADO DE TRANSCRITO DE NOVO

Inicialmente se generó un ensamblado armado con los FIELDMANUALd conjuntos de muestras, usando el programa FIELDMANUALe.

Resumen de los archivos que se entregan:

Trinity.fasta: Archivo de secuencias en formato FASTA con FIELDMANUALf transcritos resultantes del ensamble de novo.

FIELDMANUALg: Tabla de abundancias por gen y por condición (para FIELDMANUALy genes), obtenida a partir del alineamiento de las secuencias sobre el ensamble “Trinity.fasta”, usando bowtie2 versión 2.3.4.3. Con esta matriz se realizaron los análisis de expresión diferencial usando 2 métodos (EdgeR y NOISeq) que se explicarán en la siguiente sección.

FIELDMANUALw: Tabla resultante de la anotación para cada uno de estos ensamblados con el programa trinotate v3.1.1. En ésta se reportan FIELDMANUALx anotaciones para FIELDMANUALy genes únicos y FIELDMANUALf transcritos únicos que fueron buscados por blastx, blastp, RNAMMER, se buscó en Pfam, SignalP, TmHMM, eggnog, Kegg, gene\_ontology\_blast, gene\_ontology\_pfam.

Y con esta matriz se realizaron los análisis de expresión diferencial usando 4 métodos (EdgeR, NOISeq, DESeq2 y limma) que se explicarán en la siguiente sección.

# ANÁLISIS BIOINFORMÁTICO DE EXPRESIÓN DIFERENCIAL

La tabla de coberturas por gene generada con el programa RSEM v1.3.1, se utilizó como entrada en el sitio <http://www.uusmb.unam.mx/ideamex/> para realizar el Análisis de Expresión Diferencial considerando las siguientes comparaciones:

FIELDMANUALz vs FIELDMANUALZb

FIELDMANUALZb vs FIELDMANUALZd

FIELDMANUALz vs FIELDMANUALZd

Para reportar los resultados se utilizaron como líneas de corte un False Discovery Rate (FDR) o pvalue menor a 0.05, una magnitud de cambio superior a 2 (LogFC >=2) y una cuenta por Millón igual a 2 (CPM=2).

Se aplicaron 4 métodos de análisis de expresión diferencial:

* [edgeR Results](http://www.uusmb.unam.mx/ideamex/avila.dela195537/EdgeR.php)
* [DESeq2 Results](https://bioconductor.org/packages/release/bioc/html/DESeq2.html)
* [limma Results](https://www.bioconductor.org/packages/release/bioc/html/limma.html)
* [NOISeq Results](https://bioconductor.org/packages/release/bioc/html/NOISeq.html)

# RESULTADOS

Los resultados se encuentran disponibles en la liga de IDEAmex: http://www.uusmb.unam.mx/ideamex/FIELDMANUALZf

En la sección de análisis de datos, en la gráfica [Multidimensional Scaling (MDS)](http://www.uusmb.unam.mx/ideamex/ncalderon194014/DataAnalysis_Results/AllConditions_MDSPlotNorm.pdf), podemos observar una muy buena distancia entre los grupos a comparar. En la gráfica de [cajas y bigotes](http://www.uusmb.unam.mx/ideamex/ncalderon194014/DataAnalysis_Results/AllConditions_BoxPlotNorm.pdf), con las muestras ya normalizadas, se observan distribuciones muy diferentes para cada especie, cosa que se corrobora con la [gráfica de densidad](http://www.uusmb.unam.mx/ideamex/ncalderon194014/DataAnalysis_Results/AllConditions_DensitiesPlot.pdf).

Dentro de cada una de las carpetas:

• FIELDMANUALh

• FIELDMANUALi

• FIELDMANUALj

• FIELDMANUALk

Encontrará los archivos que contienen en la primera columna los identificadores de todos transcritos con los valores de expresión diferencial diferente de cero, seguido de los parámetros de cada método seleccionado, y luego las cuentas crudas y normalizadas para cada réplica de cada muestra.

FIELDMANUALl

FIELDMANUALm

FIELDMANUALn

Los archivos siguientes, contienen sólo aquellos genes con una magnitud de cambio mayor a 2 (LogFC >2). y un pvalue < 0.05, para aquellos genes con cuentas por millón > 2 (CPM>2):

FIELDMANUALo

FIELDMANUALp

FIELDMANUALq

Cada método puede generar una lista de genes que pueden coincidir o no, con los otros métodos. En los archivos siguientes están aquellos genes que fueron reportados como diferencialmente expresados por más de un método:

FIELDMANUALr

FIELDMANUALs

FIELDMANUALt

En la sección FIELDMANUALu, se hace un análisis integrativo de los resultados dados por cada método, para cada par de muestras comparadas.

Los archivos Trinity.fasta (ensamblado de novo), FIELDMANUALg (matriz de abundancias crudas) y FIELDMANUALw (anotación automática de los transcritos) pueden ser descargados en la siguiente liga:

​ FIELDMANUALv