HW5: RNA-seq

We will be conducting RNA-seq analysis of maize and teosinte. A step-by-step protocol to follow can be found here. This includes both edgeR and DESEQ; we will be using the DESEQ protocol. To save time and computational resources, you are provided with the count tables (attached) that result from mapping reads and counting read abundance. You should be able to do this on your laptop, but you are certainly welcome to use Farm if desired.

The question

Maize was domesticated from teosinte (Zea mays ssp. parviglumis). Piperno et al grew teosinte in conditions (temperature, CO2) emulating climatic conditions around the time of domestication. In these conditions, teosinte began exhibiting a number of maize-like phenotypes. Subsequent experiments with maize showed much less phenotpyic change when grown in ancient conditions, leading to the hypothesis that some of the genetic changes responsible for domestication may have been due to loci showing plastic environmental responses in teosinte, but which have lost plasticity in maize due to selection during domestiation. To test this model of genetic assimilation, we grew maize and teosinte in both modern (ambient temperature, CO2) and ancient (low temperature, CO2)conditions, and assayed gene expression via RNA-sequencing.

The model

We have 2 replicates of each of 4 maize inbred lines in two conditions (4 x 2 x 2 = 16 total samples). You will want to use this R code to appropriately identify the samples and setup the model.

```
rcdata <-read.csv("Maizereadcount.csv",header=T, row.names=1, sep=",")

cols <-
data.frame(genotype=c("RIMMA1","RIMMA1","RIMMA19","RIMMA19","RIMMA140","RIMMA14
0","RIMMA809","RIMMA809","RIMMA1","RIMMA1","RIMMA19","RIMMA19","RIMMA140","RIMM
A140","RIMMA809","RIMMA809"), sample=c("RIMMA1-1","RIMMA1-1","RIMMA19-
1","RIMMA19-1","RIMMA140-1","RIMMA140-1","RIMMA809-1","RIMMA809-1","RIMMA1-
2","RIMMA1-2","RIMMA19-2","RIMMA19-2","RIMMA140-2","RIMMA140-2","RIMMA809-
2","RIMMA809-2"), condition=as.factor(c(rep("265ppm",8), rep("400ppm",8))))

designModel = formula(~ genotype + condition)

dds <- DESeqDataSetFromMatrix(as.matrix(rcdata), colData=cols, design =
designModel)

ddsColl<-collapseReplicates(dds, groupby= dds$sample)

ddsColl<- DESeq(ddsColl)
```

For the teosinte data we do not have inbred lines, and you can use:

```
rcdata <-read.csv("teosintereadcount.csv",header=T, row.names=1, sep=",")

cols <-
data.frame(genotype=c("Pop1","Pop1","Pop1","Pop2","Pop2","Pop2","Pop3","Pop3","
Pop3","Pop4","Pop4","Pop4","Pop1","Pop1","Pop1","Pop2","Pop2","Pop2","Pop3","Po
p3","Pop3","Pop4","Pop4"),condition=as.factor(c(rep("265ppm",12),
rep("400ppm",11))))

designModel = formula(~ genotype + condition)

dds <- DESeqDataSetFromMatrix(as.matrix(rcdata), colData=cols, design =
designModel)

dds <- DESeq(dds)
```

The goal

Identify genes differentially expressed between conditions in teosinte, but **not** differentially expressed in maize. These represent an initial best list of candidate "assimilation" genes.