Summary results DE analysis Miscanthus 2014 AU-IBERS

Sabine K. Schnabel

June 15, 2015

Description of the data

- Counts table for the 2014 experiment at AU-IBERS (produced by IGATS)
- RNAseq data from Miscanthus mapped to Sorghum Bicolor
- Originally 33032 genes in the table → after data cleaning [1] 16343 genes in table
- 96 samples were taken from this green house experiment in two harvests (May 31, 2014 and June 15, 2014). The experiment ran from May 5, 2014 to June 18, 2014 (application of drought treatment as of May 12, 2014).
- Two treatments: control and drought (check %); five genotypes (WAT03 -sacchariflorus; WAT04 -sacchariflorus; WAT09 -giganteus; WAT10 -sinensis; WAT11 -sinensis).
- The greenhouse design was not officially randomized, but it seems fairly "random". We have information on the row/column position and a block number.
- Replications: each genotype was harvest replicated 9-10 times per harvest which leads to a replication of about 4-5 samples per harvest/genotype/treatment combination.

Plot of the experimental design?

[1] Genes with no expression in any of the samples were removed. Additionally all genes with a mean expression for the non-zero counts of 10 or less were removed.

DE analysis using DESeq2

Before starting the statistical analysis using `DESeq2` we ran some check on whether the greenhouse design or sampling scheme (through 2 harvests) might have an effect on the expression (analysis). To this end we used `glm.nb` including the design terms and checked for the significance of these factors. Based on the results we decided to include `Harvest` into the model in order to control for it. I also for `block` or for `row/column`, but due to difficulties with the fitting procedure and less definite results, we decided to keep `Harvest`. (UPDATE: after discussion on June 15, 2015 we identified that it might be a good idea to include `section` of the greenhouse into the model. This still needs to be done.)

- The differential expression analysis was done with the `R` packages `DESeq2`[2] to be downloaded from Bioconductor (on `R3.2.0` using `DESeq2_1.8.1`).
- The (gene-wise) full model is stated as follows:
  $\texttt{counts} \sim \texttt{Harvest} + \texttt{Treatment} + \texttt{Genotype} + \texttt{Treatment} \times \texttt{Genotype}$
- Based on this model we proceeded to make a series of LRT (likelihood ratio tests) in order to test for the different terms in the model.
  1. LRT test for the interaction: full model vs reduced model (excluding the interaction)
  2. LRT test for a general effect of treatment: full model vs reduced model (only with `Genotype` and `Harvest` in it)
  3. LRT for main effect of the treatment: full model (with only the main effects and design effect) vs reduced model (only with `Genotype` and `Harvest` in it)

Results for (1) LRT for interaction We find 61 (out of 16343) differentially expressed genes (with an adjusted p-value of p=0.0001). The smallest adjusted p-value that we found is in the magnitude of $3*10^{-17}$. Most DE genes are on chromosomes 1 and 2.

Results for (2) LRT for a general treatment effect This analysis was mainly done for a general check. We found 1311 differentially expressed genes (where it cannot be determined if the DE comes from the main effect or the interaction). Genes are not reported here.

[2]Michael I Love, Wolfgang Huber and Simon Anders (2014): Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biology

Results for (3) LRT for treatment main effect We find 1219 differentially expressed genes. The results are exported to a table. 18 of those genes have been found as differentially expressed for interaction, therefore we exclude those from the list.

	genes(interaction)	adjusted p-value
5	Sobic.001G313100.v2.1	0.0000000000
39	Sobic.005G165800.v2.1	0.0000000000
2	Sobic.001G142800.v2.1	0.0000000000
26	Sobic.003G203600.v2.1	0.0000000000
57	Sobic.009G199600.v2.1	0.0000000002
34	Sobic.004G223500.v2.1	0.0000000017
46	Sobic.006G201700.v2.1	0.0000000199
4	Sobic.001G218500.v2.1	0.0000000295
33	Sobic.004G223400.v2.1	0.0000000536
49	Sobic.007G131000.v2.1	0.0000002357
13	Sobic.002G107800.v2.1	0.0000003188
55	Sobic.009G129300.v2.1	0.0000006427
16	Sobic.002G176500.v2.1	0.0000010731
47	Sobic.006G215400.v2.1	0.0000015078
28	Sobic.003G251700.v2.1	0.0000025794
27	Sobic.003G237600.v2.1	0.0000028060
45	Sobic.006G146000.v2.1	0.0000028060
48	Sobic.006G260300.v2.1	0.0000032211
35	Sobic.004G269100.v2.1	0.0000037527
17	Sobic.002G176700.v2.1	0.0000045595
24	Sobic.003G003800.v2.1	0.0000060320
59	Sobic.010G015800.v2.1	0.0000075508
21	Sobic.002G387000.v2.1	0.0000156941
53	Sobic.009G081500.v2.1	0.0000156941
42	Sobic.006G055400.v2.1	0.0000168083
60	Sobic.010G134000.v2.1	0.0000171320
52	Sobic.009G062900.v2.1	0.0000176362
61	Sobic.010G178800.v2.1	0.0000176362
1	Sobic.001G142200.v2.1	0.0000178267
54	Sobic.009G122000.v2.1	0.0000178267
32	Sobic.004G189900.v2.1	0.0000201983
56	Sobic.009G190500.v2.1	0.0000201983
40	Sobic.005G169200.v2.1	0.0000238157
41	Sobic.006G021900.v2.1	0.0000255084
37	Sobic.004G347500.v2.1	0.0000255119
38	Sobic.005G152200.v2.1	0.0000255119
19	Sobic.002G237900.v2.1	0.0000280657
31	Sobic.004G166700.v2.1	0.0000280657
51	Sobic.008G147900.v2.1	0.0000280657
7	Sobic.001G368600.v2.1	0.0000360883
25	Sobic.003G073700.v2.1	0.0000360883
44	Sobic.006G132300.v2.1	0.0000400168
29	Sobic.003G302100.v2.1	0.0000432178
22	Sobic.002G405400.v2.1	0.0000447936
12	Sobic.002G090500.v2.1	0.0000468262
30	Sobic.003G371100.v2.1	0.0000483192
10	Sobic.001G543500.v2.1	0.0000485755
58	Sobic.009G255100.v2.1	0.0000636227
18	Sobic.002G218100.v2.1	0.0000648755
23	Sobic.002G416700.v2.1	0.0000659518
50	Sobic.007G208700.v2.1	0.0000659518
8	Sobic.001G391400.v2.1	0.0000723126
20	Sobic.002G321300.v2.1	0.0000760341
9	Sobic.001G511000.v2.1	0.0000824074
6	Sobic.001G326900.v2.1	0.0000882305
14	Sobic.002G107900.v2.1	0.0000882305
15	Sobic.002G165400.v2.1	0.0000882305
36	Sobic.004G299600.v2.1	0.0000882305
43	Sobic.006G116000.v2.1	0.0000882305
3	Sobic.001G184300.v2.1	0.0000929545
11	Sobic.002G020700.v2.1	0.0000955330