Reviewer 1:

“Were any different correlation approaches considered and did they provide similar or different results from the Gini correlation analysis. The authors claim it is not often applied and they are correct. This correlation analysis has some positive aspects when dealing with non-normally distributed data. The question is when other techniques are applied, does the strength/direction, # connections in the correlation network shown in Figure 6 change?”

Reviewer 2:

“Probably it’s better to split (N vs. N+1) model to V3 vs. E4 model and E4 vs. R3 model, which would provide more information and made the MS more understandable. The author may pinpoint the specific stage in which gene expression has good association with final ED of mature biomass and could serve as good marker. Because lignin pathway genes are all expressed relatively low in V3 stage, even the initial lignin gene expression in V3 showed correlation with the final lignin accumulation in R3 stage, I still doubt if it could affect final lignin accumulation in shoot. It may so happen that the gene expression in V3 has similar trend of the gene expression in E4. And in E4 stage, the high expression of lignin pathway genes (Figure 7) could affect the final lignin accumulation. Could the authors find some indicators to show which association is primary association? It would be an interesting question to be discussed in the MS.”

Discussion with Fan on Jan 18, 2021

Actions

1) Ensure data are correct (file available in “data and Analysis” folder). Please note the need to make sure that there are not rounding errors with information loss in the files. This is especially important for non-parametric correlation analysis.

2) Make table of correlations among cell wall components

* GCC vs SCC vs PCC
* P-values for each

[Previous plan to address comment of Reviewer 2:

3) Visualization of gene expression vs cell wall composition data: NvN, NvN+1, NvN+2 – Do any developmental stages or genotypes dominate?

3.5) Redo GE vs CW correlations with subsets of data.

* Separate developmental stages and segments
* Separate genotypes

]

New suggested plan to address comment of Reviewer 2:

3Alt)

Q: Is there a way to see which combination of relationships (subset of samples, genotypes) gives the greatest number of high correlations between PhP gene expression and FA, pCA, lignin, and (negatively) for ED. THis soulds like a machine learning problem to me, but maybe this is something that is fast for you to set up now? It might be faster than to go through visualizing the data and manually trying each combination.

A: For your idea about running the correlations on subsets of samples or genotypes and find out which sample subset tend to always have the highest correlation scores, I think it is do-able and answers the reviewer questions. We may want to define the sample subset set first and run them all together. It will work like this: assuming we have 10 sample subset, we calculate coefficient for all and find the maximum. Then we count how many times each subset is the max for all gene-CW component pairs. Output will be bargraph with 10 sample subsets and the number is how many times the are the maximum.

vs = CORRELATION

+ = VERTICAL CONCATENATION

Sample subsets

1) Whole data:

DONE: NGE vs NCW:

V3 vs V3 +

E4 vs E4 +

R3 vs R3 +

S1 vs S1 +

S2 vs S2 +

S3 vs S3

N(GE) vs N + 1(CW):

V3 vs E4 +

E4 vs R3 +

S1 vs S2 +

S2 vs S3

N(GE) vs N + 2(CW):

V3 vs R3 +

S1 vs S3

2) Stages only

* NGE vs NCW:
* V3 vs V3 +
* E4 vs E4 +
* R3 vs R3
* NGE vs N+1CW:
* V3 vs E4 +
* E4 vs R3
* NGE vs N+2CW: V3 vs R3

3) \*\*Stage pairs:

* NGE vs N+1CW:
* V3 vs E4
* NGE vs N+1CW:
* E4 vs R3

2) Segments only; NvsN;

N(GE) vs N(CW):

S1 vs S1 +

S2 vs S2 +

S3 vs S3

N(GE) vs N+1(CW);

S1 vs S2 +

S2 vs S3

N(GE) vs N+2 (CW):

S1 vs S3

3-5) Genotypes only (AP13, VS16, A4)

A4

N(GE) vs N(CW):

V3 vs V3 +

E4 vs E4 +

R3 vs R3 +

S1 vs S1 +

S2 vs S2 +

S3 vs S3

A4

N(GE) vs N + 1(CW):

V3 vs E4 +

E4 vs R3 +

S1 vs S2 +

S2 vs S3

A4

N(GE) vs N + 2(CW):

V3 vs R3 +

S1 vs S3

AP13

N(GE) vs N(CW):

V3 vs V3 +

E4 vs E4 +

R3 vs R3 +

S1 vs S1 +

S2 vs S2 +

S3 vs S3

AP13

N(GE) vs N + 1(CW):

V3 vs E4 +

E4 vs R3 +

S1 vs S2 +

S2 vs S3

AP13

N(GE) vs N + 2(CW):

V3 vs R3 +

S1 vs S3

VS16

N(GE) vs N(CW):

V3 vs V3 +

E4 vs E4 +

R3 vs R3 +

S1 vs S1 +

S2 vs S2 +

S3 vs S3

VS16

N(GE) vs N + 1(CW):

V3 vs E4 +

E4 vs R3 +

S1 vs S2 +

S2 vs S3

VS16

N(GE) vs N + 2(CW):

V3 vs R3 +

S1 vs S3