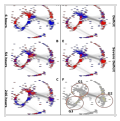

Aug 24, 2019

Analysis of co-expression networks among Populus spp. involved in the response to phosphate deficiency and aluminium toxicity

Thiago Bergamo Cardoso[1], Renan Terassi Pinto[1], Luciano Vilela Paiva[1]

[1]Federal University of Lavras

1 *Works for me* dx.doi.org/10.17504/protocols.io.6q7hdzn

Thiago Cardoso
Federal University of Lavras

ABSTRACT

It is reasonable that some molecular responses to aluminium toxicity and phosphate starvation stresses may be co-opted, as these stresses co-exist in the same environment and the most directly affected organ is the root in both cases .Some findings pointing to co-opted responses have already been reported, like the exudation of organic acids by root tips, that chelates Al^{3+} ions and at the same time increase phosphate availability, activation of the same transcriptional regulator triggered by both stresses and possibilities of hormone signalling crosstalk.

There are already reports in the literature of applications of molecular knowledge about adaptation to acid soils in commercially relevant crops. Here we focus on studyng the broad transcriptomic response to aluminium and phosphate starvation stresses in *Populus ssp.* as a model plant for woody crops, due to its genetic, genomic and transcriptomic data availability.

In view of the above, this work highlights the transcriptional responses to aluminum toxicity and phosphate starvation in *Populus*spp. by the use of public microarray data and points to connections between the responses to the two stresses through gene co-expression networks. This approach is justified by the importance of understanding the relationship between the responses of these two major coexisting stresses in acidic soils and the lack of studies that address this relationship, as well as the responses to each stress in a specific way, in broad transcriptomic data.

Analysis of data from microarray experiments

1 The first step was to download the data and perform the necessary analyzes to determine which genes were differentially expressed.

1.1 The microarray data used in this work were obtained from the NCBI database using accession numbers GSE19297 and E-MTAB-3934

1.2 Microarray analyses were carried out with the Affymetrix GeneChip poplar genome array (v1.1 from the *P. trichocarpa*genome project),data were analysed using the free statistical software R (R version 3.3.2) following the protocol described by Janz. The R package "*affy*" was used to normalize the probes using the "*rma*" function, available in the Bioconductor software. The $\log_2$expression values of transcripts that were present in all replicates, obtained by the "*mas5calls*" function, in at least one of the conditions were used for the subsequent analyses. **Supplementary table S1.xlsx**

1.3 Student's t-test ($p<0.05$) was used to identify the genes that were differentially expressed in treatments with aluminium as well in doses of phosphate.The genes considered as differentially expressed were those genes that showed (1) expression in all replicates in at least one library and (2) abundance of the transcript in at least one treatment significantly different from that in another treatment according to Student's t-test, that is, differential expression in at least one library.

Statistics_DE_P_Al.xlsx

1.4 For annotation of differentially expressed transcripts according to the *Populus trichocarpa* genome, the *Blastx* tool was used to align the probes with the CDS database of *P. trichocarpa*. An e-value < 1e-5 and a maximum of 10 alignments was used for each probe, and the best alignment was selected. anotacao.xlsx

Construction of gene co-expression networks

2

2.1 To identify the proximity of the gene expression profile in the different treatments, the biological coefficient of variability (BCV) was determined for each treatment and an MDS plot was plotted using the EdgeR package.

2.2 To verify the carrelation of the expression profile of the genes two analyses were performed, a Spearman correlation to identify the relationship between genes and the p-value of this number to determine if this correlation was statistically significant. The p-value was determine using the formula in the excel P = IDIST{ABS[r/SQRT({1-r*r}/{n-2})],[n-2],2} as describe by Usadel, where the number of samples is indicated by n, and r is the observed co-expression score, only correlations with p-value <0.05 were used

2.3 Spearman's correlation with $\rho \geq 0.8$ was selected because unlike the value used in the literature, $\rho \geq 0.7$, the goal was to focus only in the genes that showed the strongest correlation.

2.4 Only genes that satisfy the two previous characteristics were used. Network_080_DE_Al_P_pvalue.csv

Network_080_DE_Al_P_pvalue.csv

Enrichment of gene co-expression networks

3 To establish the type of regulation in relation to stress for each dataset, the following procedures were used: (1) for the aluminium stress-related libraries, the mean values of the biological triplicates of each gene in the different libraries were estimated, and subsequently, the values of the aluminium stress-related libraries were divided by the values of the library for the same stress exposure time but in the control condition; (2) for the phosphate-related libraries, again, the mean expression values for each gene among the biological triplicates of all libraries were estimated, and then the values of the medium phosphate concentration library were divided by the values of the high phosphate concentration library (to characterize the stress), and likewise, the low phosphate concentration values were divided by the high phosphate concentration values (severe stress).

3.1 Using the values obtained in the calculations described above, it was possible to generate the filters used in the work: positive regulation (coefficient > 1), negative regulation (coefficient < 1), transcriptional variation higher than 20% (0.83 ≤ coefficient ≤ 1.2) and transcriptional variation 2-fold up or down(0.5 ≤ coefficient ≤ 2.0). **Supplementary table S2.xlsx**

3.2 For each gene co-expression network constructed, the *P. trichocarpa* proteins corresponding to the probes used in the respective network were analysed in the *TF prediction server* software of the *PlantTFDB* database for predicting possible transcription factors present in the network in addition to the family of transcription factors to which it belongs. In addition, the *Blast2GO* software was used to identify enriched *Gene Ontology* classes in the data for the distinct groups of gene co-expression networks. **Supplementary table S3.xlsx**