RNA-seq analysis in Common Bean (*Phaseolus vulgaris* L.) under phosphorous stress

Introduction

Common beans are the important legume crop worldwide, and especially in developing countries, they serve as main protein source in people's diet (Hernández et al. 2007). They are mostly grown in tropical regions of Latin America and Eastern Africa (Diaz et al. 2017). However, the productivity of common beans in those areas is often limited by the low input of phosphorous(P) nutrient. In Lain America, 60% of the common beans cultivated land is under low P stress (Diaz et al. 2017)where farmers also have financial burden to buy P (Rao et al. 2016). P is one of necessary macro nutrients for plant growth. Due to its immobility and partial solubility, P always accumulates in the shallow soil layers, which make plants can only absorb a small portion of P in the soil (Batjes 2011). Previous studies showed some genotypes of common beans have adapted to low P environment by modifying their root systems and presented higher P use-efficiency to possibly maintain grain yield (Rao et al. 2016). Thus, to study the genetic mechanism of those adapted genotype may provide us some insights of how to enhance tolerance to P stress in common bean.

Here, we used two previous identified contrasting common bean genotypes regarding to P-use-efficiency: IAC Imperador is P efficient and responsive while DOR364 is P inefficient and unresponsive (da Silva et al. 2014). The aim of this study is to identify different gene expression profiles between genotypes and P application rates. We grew two common bean genotypes in the hydroponic system with two different P concentration: P-restricted concentration 4.00 mg L^{-1} and a P-control concentration 8.00 mg L^{-1}.(da Silva et al. 2019) We conducted RNA-seq analysis to detect how common bean change to respond to induced P stress within or among genotypes and we also wanted to investigate if there was the interaction of genotype and treatment effect.

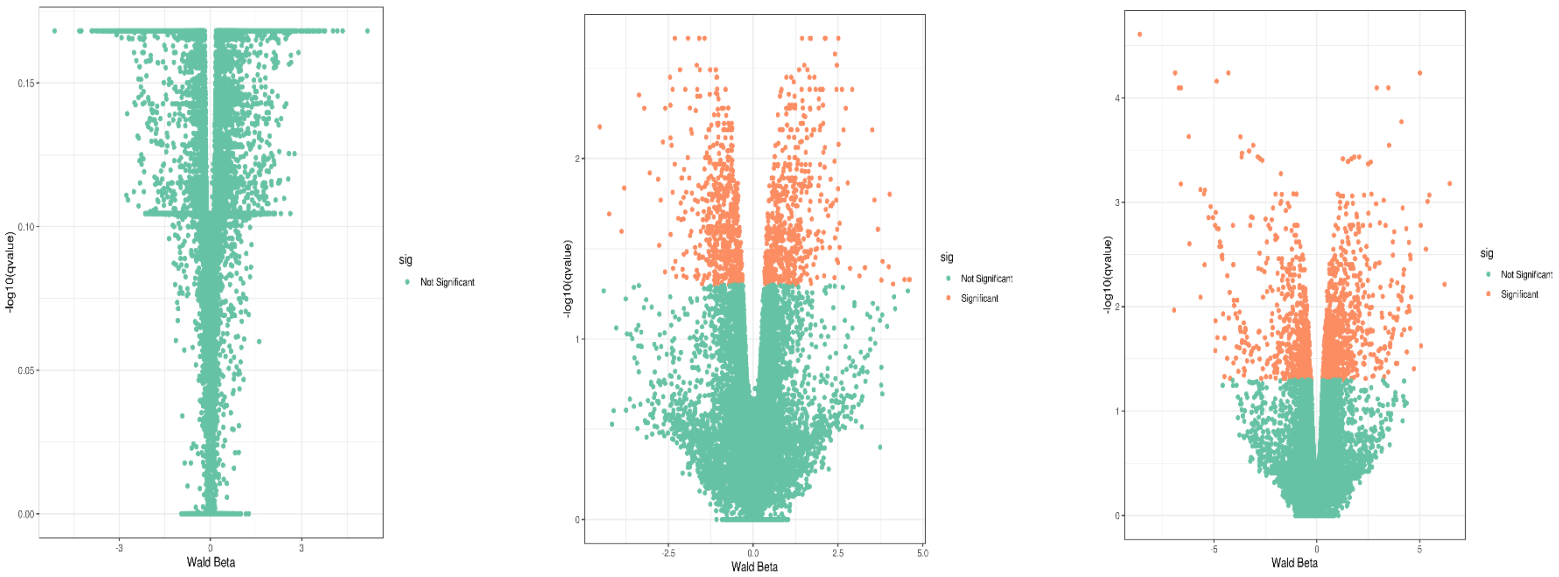
Materials and Methods

Dataset

We applied 2x2 factorial experimental design with two factors: genotype and P treatment. We have two common bean genotypes: DOR364 and IAC Imperador and two P treatments: control (no P stress) and P-restriction stress. After combined genotype with treatment, we have possible 4 conditions. In each condition, we extracted total RNAs from root systems of 3 biological samples when common beans were in the full flowering stage. The 12 RNA libraries were prepared and sequenced with the Illumina Hiseq 2500 sequencer. The RNA seq data was uploaded into NCBI website. The SRA accession is PRJNA498535 (https://www.ncbi.nlm.nih.gov/sra/PRJNA498535).

Differential gene expression analysis

The RNA-seq data was downloaded from above website address using SRATools v2.9.1-



a b c

centos_linux64 and *Phaseolus vulgaris* cds gene sequence was obtained from *Phaseolus vulgaris* genome folder (*Phaseolus vulgaris* v2.1) on the Phytozome v.10 database. We used kallisto v 0.43.1 to quantify counts of transcripts (Bray et al. 2016). The obtained counts were analyzed with Sleuth v 0.30 program for expression analysis (Pimentel et al. 2017). We parsed gene annotation information for each transcript ID to get consistent result that gene and transcript results compatible with each other. To visualize gene expression profile in volcano plot, we got beta value (regression coefficient) from Wald test to estimate fold change and the Benjamin-Hochberg (Benjamini and Hochberg 1995) corrected false discovery rate (FDR) was obtained from likelihood ratio test. We declare significant differential expression transcripts when FDA < 0.05.

We made five different comparisons of differential expression test for treatment effect, genotype effect, and genotype by treatment (interaction) effect. Comparison 1: DOR364 at restrictive P level vs DOR364 at control P level. Comparison 2: IAC Imperador at restrictive P level vs IAC Imperador at control P level. Comparison 3: IAC Imperador vs DOR364 at restrictive P level. Comparison 4: IAC Imperador vs DOR364 at control P level. Comparison 5: IAC Imperador x restrictive P level vs IAC Imperardor x control P level vs DOR364 x restrictive P level vs DOR364 control P level.

Results

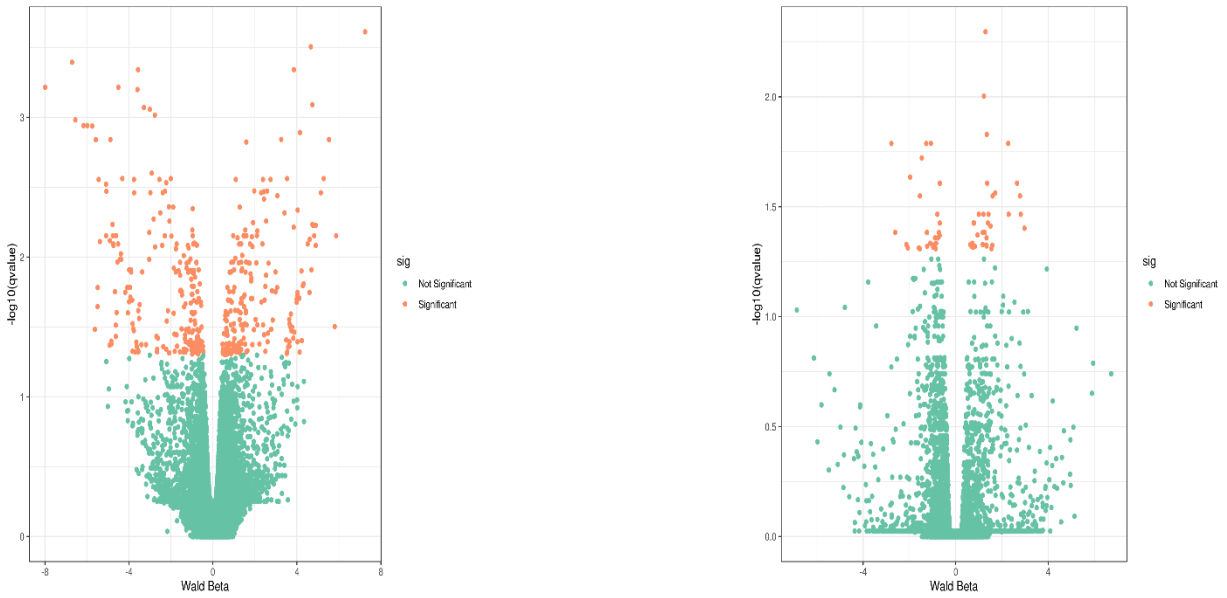


Figure 1. Volcano plots of 5 comparisons. Dots in green represent non-significant differentially expressed transcripts; Dots in red represent significant differentially expressed transcripts. a) Comparison 1: DOR364 at restrictive P level vs DOR364 at control P level. b) Comparison 2: IAC Imperador at restrictive P level vs IAC Imperador at control P level. c) Comparison 3: IAC Imperador vs DOR364 at restrictive P level.d) Comparison 4: IAC Imperador vs DOR364 at control P level. e) Comparison 5: IAC Imperador x restrictive P level vs IAC Imperardor x control P level vs DOR364 x restrictive P level vs DOR364 control P level.

Summary of results

The genotype of DOR364 did not have any significant deferentially expressed (DE) transcripts between treatments which confirmed that DOR 364 was P-unresponsive genotype (Fig 1.a, Comparison 1). While the P-responsive genotype IAC Imperador, exhibited 1355 significant DE transcripts accounting for 3% of total expressed transcripts between treatments (Fig 1.b, Comparison 2). We found 1694 (4.58%) significant DE transcripts under P-restriction environment and 453 (1.22%) significant DE transcripts under P-control environment between genotypes(Fig 1.c, Fig 1.d, Comparison 3, Comparison 4). For testing for interaction effect of genotype by treatment, we identified 63 (0.17%) significant DE transcripts (Fig 1.d, Comparison 5).

Conclusion

The pipeline I wrote for RNA-seq DE expression analysis is basically performed well and can automatically generate volcano plot and significant DE expression transcript and gene table. But there is still a lot to improve of this pipeline: (1) I did not set a threshold of Wald beta value to filter out low fold change transcripts for volcano plot. I will explore this dataset carefully and select a biological or statistically meaningful change value for beta threshold. (2) I will perform gene ontology enrichment analysis for identified significant DE genes/transcripts. Because it is hard to interpret so many genes at once.It would be great to classify genes into functional categories. (3) When using sleuth, we need to create a full model and a reduced model based on which variable we want to test. I wrote a function

DE_sleuth() to get DE results by just inputting which the data, variable you want test, and Wald test coefficient. But this function only for single variable. When we have two or three variables and want to test the interaction effect among variables, it wont work. I am going to write a function that can deal with interaction effect of two or three variables.

Reference

Batjes, N H. 2011. "Global Distribution of Soil Phosphorus Retention Potential." *World Soil Information* 06 (February): 42.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society . Series B (Methodological)* 57 (1): 289–300.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34 (5): 525–27.

Diaz, Lucy M., Jaumer Ricaurte, Cesar Cajiao, Carlos H. Galeano, Idupulapati Rao, Steve Beebe, and Bodo Raatz. 2017. "Phenotypic Evaluation and QTL Analysis of Yield and Symbiotic Nitrogen Fixation in a Common Bean Population Grown with Two Levels of Phosphorus Supply." *Molecular Breeding* 37 (6).

Hernández, Georgina, Mario Ramírez, Oswaldo Valdés-López, Mesfin Tesfaye, Michelle A. Graham, Tomasz Czechowski, Armin Schlereth, et al. 2007. "Phosphorus Stress in Common Bean: Root Transcript and Metabolic Responses." *Plant Physiology* 144 (2): 752–67.

Pimentel, Harold, Nicolas L. Bray, Suzette Puente, Páll Melsted, and Lior Pachter. 2017. "Differential Analysis of RNA-Seq Incorporating Quantification Uncertainty." *Nature Methods* 14 (7): 687–90. https://doi.org/10.1038/nmeth.4324.

Rao, Idupulapati M., John W. Miles, Stephen E. Beebe, and Walter J. Horst. 2016. "Root Adaptations to Soils with Low Fertility and Aluminium Toxicity." *Annals of Botany* 118 (4): 593–605.

Silva, Daiana Alves da, José Antônio de Fátima Esteves, Ueliton Messias, Aurélio Teixeira, João Guilherme Ribeiro Gonçalves, Alisson Fernando Chiorato, and Sérgio Augusto Morais Carbonell. 2014. "Efficiency in the Use of Phosphorus by Common Bean Genotypes." *Scientia Agricola* 71 (3): 232–39.

Silva, Daiana Alves da, Siu Mui Tsai, Alisson Fernando Chiorato, Sónia Cristina da Silva Andrade, José Antonio de Fatima Esteves, Gustavo Henrique Recchia, and Sérgio Augusto Morais Carbonell. 2019. "Analysis of the Common Bean (Phaseolus Vulgaris L.) Transcriptome Regarding Efficiency of Phosphorus Use." *PLoS ONE* 14 (1): 1–27.