The Effects of Citrus Greening on the Genetic Expression in Healthy Citrus Sinensis Trees

Paige Applegate, Brian Barker, Lindsay Dobrie

Abstract

Citrus Greening (Candidatus Liberibacter asiaticus) is a citrus plant disease. The disease causes fruit to disfigure, green, and bitter, rendering the product unsuitable for sales. Citrus Greening refers to either Candidatus Liberibacter asiaticus (CLas) or Citrus tristeza virus (CTV). A tree is infected by a carrying insect called the Asian citrus psyllid (Diaphorina citri Kuwayama or ACP). This disease has affected millions of acres of citrus crops across the world. Since there is no cure for it, most trees with it die within a few years. To help resist greening, vulnerable citrus species, such as the Citrus sinensis used in this study, are often grafted to more resistant trees. Here we show that diseased grafts treated with either CLas or CTV can be differentiated from healthy grafts using gene expression analysis after about eight months. This is useful as the early symptom of Citrus Greening, yellowing of leaves, is difficult to differentiate from other possible culprits like nutrient deficiency. Further spreading of the disease can devastate entire citrus markets, so early identification is crucial in preventing widespread economic damage. Our analysis confirms a large array of similar analyses (many of which can be found in the *References* section), and further indicates the value of genetic analysis in combating Citrus Greening.

Introduction

Citrus tristeza virus (CTV) has ravaged the citrus industry in recent years (Moreno, 2008). There are multiple strains as the virus has spread throughout different species and continents, however, it has three major syndromes known as tristeza, stem pitting, and seedling yellows (Moreno, 2008; Dawson, 2013). Tristeza is debilitating, the stem pitting causes severe growth stunting, and the seedling yellows is a more rare manifestation of the virus (Moreno, 2008; Dawson, 2013). There are multiple ways to treat and prevent infection, however, there is no cure to the virus. One of the more common treatments involves grafting citrus varieties onto a rootstock that is tolerant to the virus (Cristofani-Yaly, 2007; Gandía, 2007).

While there are common syndromes, symptoms can vary depending on both virus strain and host species (Cheng, 2016). It is best to detect infection as early as possible to prevent the disease from spreading to other trees, preferably before the tree shows symptoms (Chin, 2020). However, there is currently no foolproof way to do this.

This study graft inoculated *Citrus sinensis* (navel orange) trees. The rootstock was infected using Candidatus Liberibacter asiaticus, Spiroplasma citri, or was given one of two isolates of citrus tristeza virus to infect *Citrus sinensis* (navel orange) trees (2). The control trees

were graphed with healthy plants. Leaf samples were then collected from the trees for RNA extraction and gene expression analysis across 2, 4, and 8 months.

The purpose of this project was to answer this question: are there any genetic signs indicating the stage of citrus greening infection by comparing infected trees to healthy samples? The hypothesis is that the stage of development is apparent by comparing the genetic data of infected and healthy samples. The data used can be found here and at the link in the second reference. Examining the genetic expression and running clustering analyses on the results showed that, in month 8, the expression between the healthy and infected trees began to differ. While this result is not very concrete, it does confirm that citrus greening can be detected in a plant early on. Further testing and improvement on the study needs to be done, however, knowing citrus greening begins to be detectable at around 8 months can help set a time frame for future studies. The earlier the disease is detected, the quicker farmers can either treat or remove the tree to prevent neighboring infections.

Methods

Initially, the methods of analysis were run on the complete dataset. However, in order to gain clearer insight on the data, the analysis was rerun separating the samples by the month they were taken. The main dataset used in the analysis contained the gene expression count data across all samples, and was analyzed using the following methods in R.

The differentially expressed genes across all samples were discovered using DESeq2 (Love et al., 2021). The analysis was grouped by the sample, month, and treatment type. Once the differentially expressed genes were found, clustProfiler was used to run an enrichment analysis (Wang et al, 2021) to trace the gene ontologies. However, the researchers initially mapped the genes to a genome (CPBD, 2021) that did not have a clear gene mapping with any clustProfiler genome, and enrichment was unsuccessful. Additionally, enrichment using gProfiler2 and GenomicSuperSignature was similarly unsuccessful.

General patterns across all samples could still be determined using clustering on the top 5000 differentially expressed genes. PAM Clustering (Maechler, 2021) was run selecting a k=4, ConsensusClusterPlus (Wilkerson, 2010) was run selecting a k=8, and hierarchical clustering using hclust (Pathak, 2018) was run selecting a k=6. While the PAM clustering was automatic, the second two cluster counts were selected by determining the significance of increasing and decreasing the cluster counts.

Once much of the aforementioned information was gleaned from the dataset, it was visualized using various plots. Dendrograms, PCA plots, and Heatmaps were important for displaying the cluster analysis data. Volcano plots, PCA plots, and Heat Maps were used to display the differential gene expression analysis results, and were especially useful at highlighting important differentially expressed genes. A t-SNE plot was used to compare the control and experimental groups.

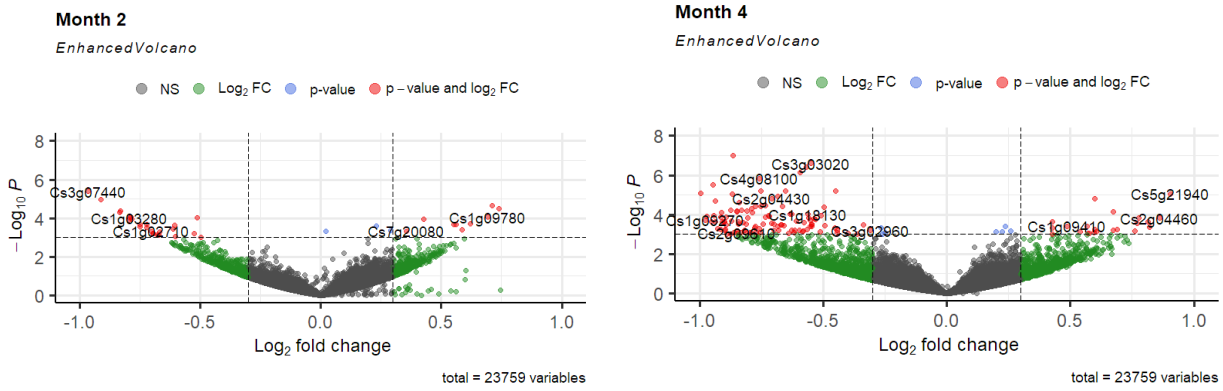
In order to determine the significance of the clusters that were calculated and displayed, a chi square and p-adjusted chi square analyses were run on the clustering results. While this

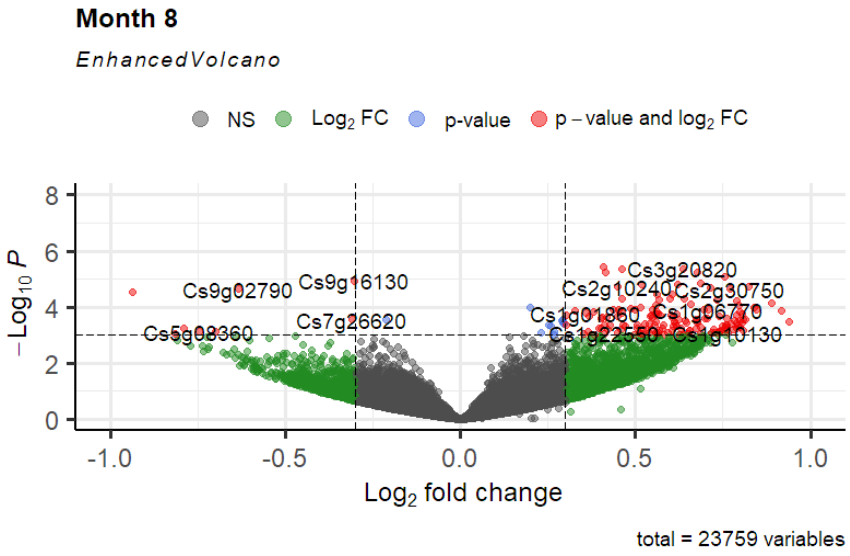
determined significance, earlier on various normalization methods were used in order to reduce outliers in the data.

The code used in this analysis can be found here and in the github repository link in the first reference.

Results

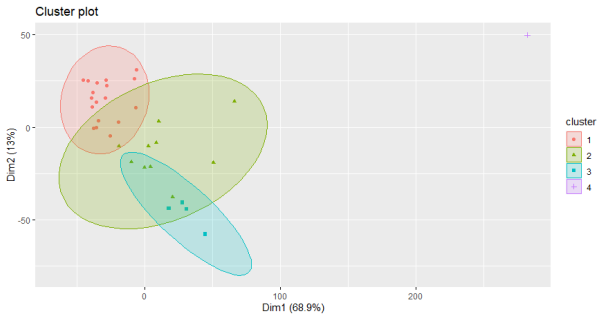
Through this project we were able to answer our original question, can one tell the stage of development of citrus greening by comparing the infected trees to the healthy sample.

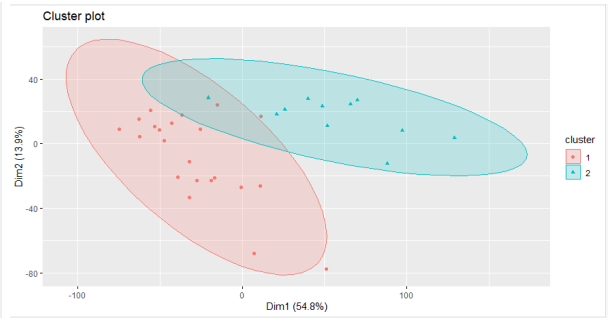




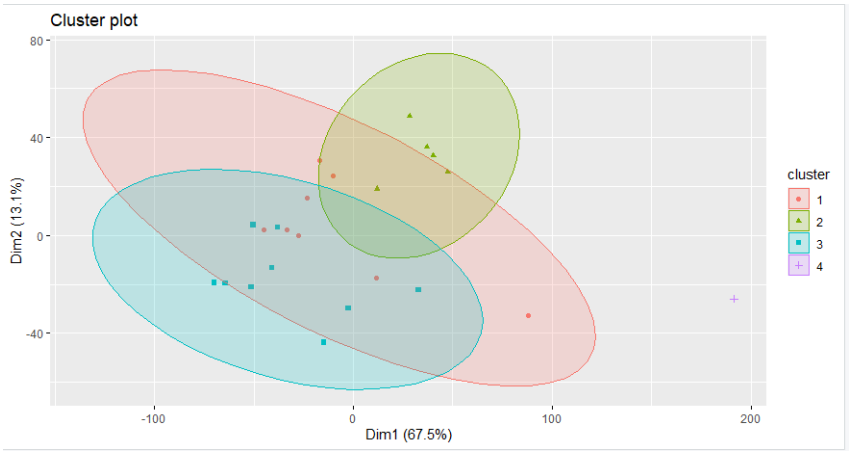
As seen in the volcano plots, the gene expression variation increases with time. While that alone is not enough to determine whether disease treatment directly affects expression, it does illustrate the expected increase in variance as the treatments have more opportunity to influence the plants.

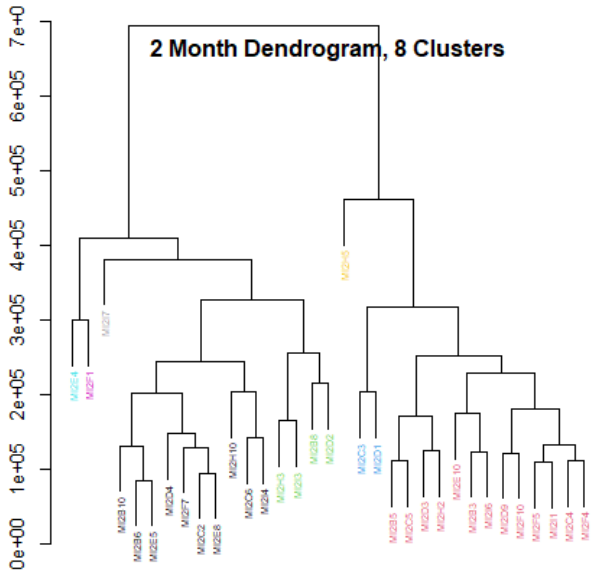
Month 2 **Month 4**

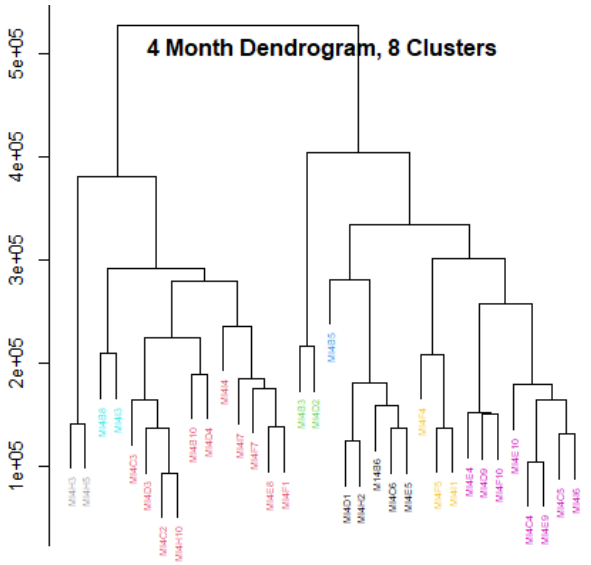

Cluster plot
Dim2 (13%)
Dim1 (68.9%)
cluster
1
2
3
4

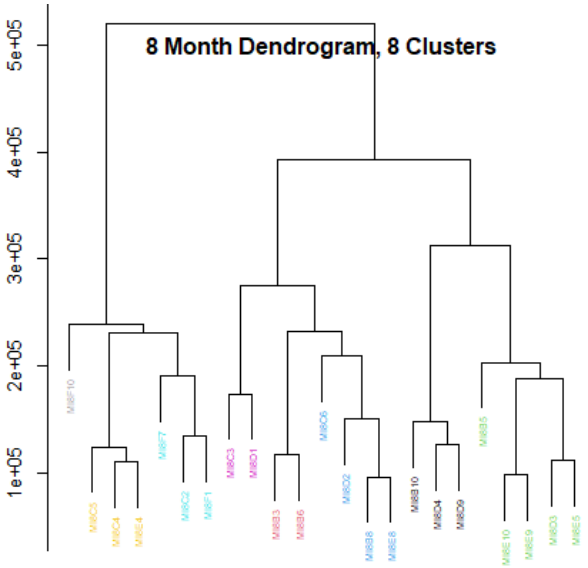

Cluster plot
Dim2 (13.9%)
Dim1 (54.8%)
cluster
1
2

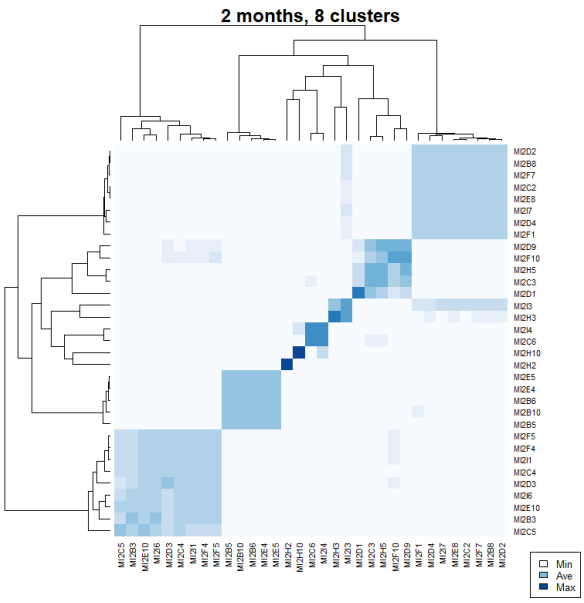
Month 8


Cluster plot
Dim2 (13.1%)
Dim1 (67.5%)
cluster
1
2
3
4

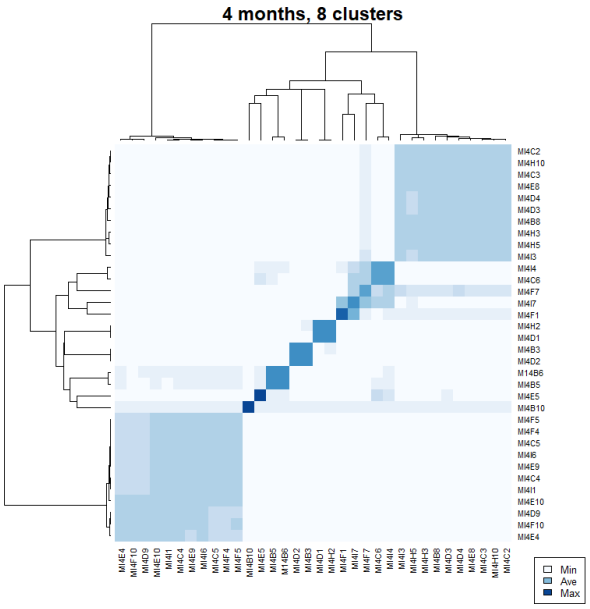

2 Month Dendrogram, 8 Clusters


4 Month Dendrogram, 8 Clusters

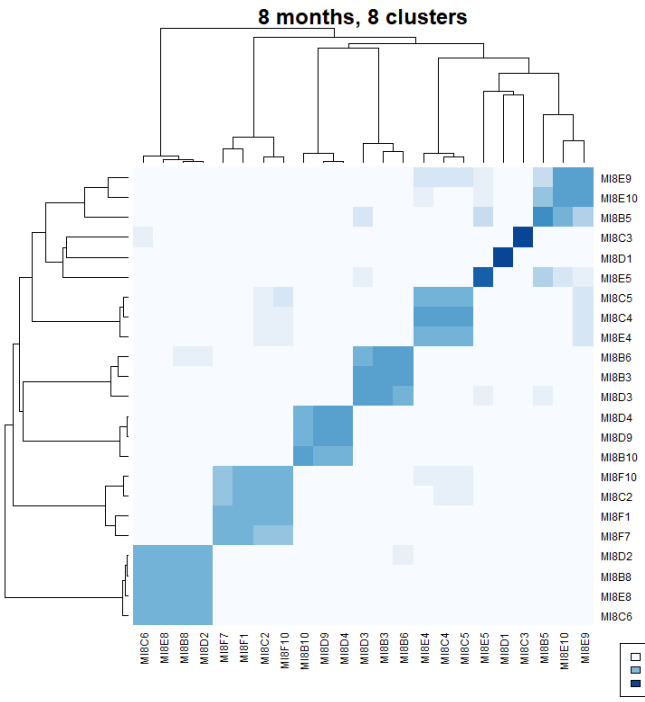




Chi-squared P-value: 0.2797

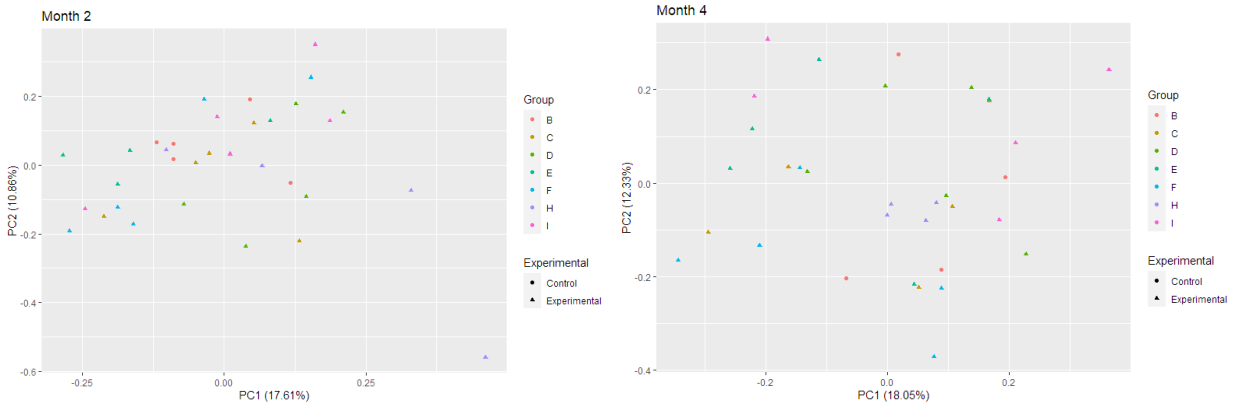


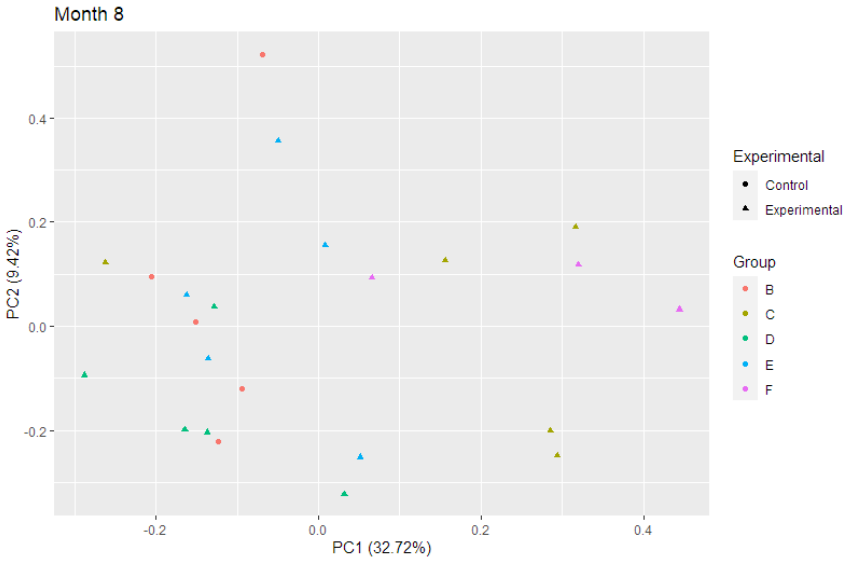
Chi-squared P-value: 0.1448



Chi-squared P-value: 0.02974

Clustering across the month groups reveals increasingly significant variation in gene expression. The predictive ability of differential gene expression analysis is low for the early month samples, as expected. By month eight, however, clustering methods are able to reliably differentiate samples by their gene expression, matching the treatment groups with a p-value of 0.02974. This can be seen clearly in the dendrogram and heatmaps. What's more, the most significant results are found when clustering in eight groups which match the various treatment groups, implying that each variation in treatment had significant effects on gene expression.





PCA analysis reveals an increase in grouping by treatment group around month eight. Samples from month 2 are largely clustered together, likely due to insufficient time to diverge. Month 4 samples have a more even spread, but no clear groups can be discerned. By month eight, however, treatment groups tend to cluster near each other, indicating that treatment has a large effect on gene expression.

Conclusion

This analysis suggests that our hypothesis, the stage of development can be told by comparing the infected and healthy samples, is true. We found that Citrus sinensis treated with CLas or CTV can be reliably differentiated from healthy specimens using RNA analysis. In earlier stages this is not apparent, but by month eight, gene expression between the different groups begins to diverge. PCA analysis, the volcano plots, and the clustering algorithms all indicate an increase in gene expression variance over time, with the clustering algorithms exposing the significant relationship between treatment group and gene expression.

Initially, the analysis was done on the entire data set, without separating groups by month. This led to confusing and unreadable results. Instead of continuing with this data, we decided to re-run much of our analyses separating the data by the month the samples were taken (2, 4, 8), which yielded clearer data.

In total there were 92 samples. Each sample was either infected or uninfected (experimental or control), and all plants were grafted. The experimental samples were infected with one of four diseases (Candidatus Liberibacter asiaticus, Spiroplasma citri, or one of two Citrus Tristeza Viruses). In a future analysis of the data, it would be interesting to see how, month to month, the different diseases individually progressed compared to the control group.

References

1. https://github.com/Brian-Barker/Bioinformatics
2. McNeil, C., Araujo, K., Godfrey, K., & Slupsky, C. (n.d.). *Gene expression profiles of Washington navel oranges (Citrus sinensis (L) Osbeck) graft innoculated with Candidatus Liberibacter asiaticus, Spiroplasma citri, or two citrus tristeza isolates*. National Center for Biotechnology Information. Retrieved December, 2021, from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE167171.
3. Moreno, P.; Ambros, S.; Albiach-Marti, M. R.; Guerri, J.; Pena, L. Citrus Tristeza Virus: A Pathogen That Changed the Course of the Citrus Industry. Mol. Plant Pathol. 2008, 9 (2), 251– 268 link
4. Dawson, W.; Garnsey, S.; Tatineni, S.; Folimonova, S.; Harper, S.; Gowda, S. Citrus Tristeza Virus-Host Interactions. Front. Microbiol. 2013, 4, 88, link
5. Cristofani-Yaly, M.; Berger, I. J.; Targon, M. L. P. N.; Takita, M. A.; Dorta, S. de O.; Freitas-Astúa, J.; Souza, A. A. de; Boscariol-Camargo, R. L.; Reis, M. S.; Machado, M. A. Differential Expression of Genes Identified from Poncirus Trifoliata Tissue Inoculated with CTV through EST Analysis and in Silico Hybridization. Genet. Mol. Biol. 2007, 30, 972– 979, link
6. Cheng, C.; Zhang, Y.; Zhong, Y.; Yang, J.; Yan, S. Gene Expression Changes in Leaves of Citrus Sinensis (L.) Osbeck Infected by Citrus Tristeza Virus. J. Hortic. Sci. Biotechnol. 2016, 91 (5), 466– 475, link
7. Gandía, M.; Conesa, A.; Ancillo, G.; Gadea, J.; Forment, J.; Pallás, V.; Flores, R.; Duran-Vila, N.; Moreno, P.; Guerri, J. Transcriptional Response of Citrus Aurantifolia to Infection by Citrus Tristeza Virus. Virology 2007, 367 (2), 298– 306, link
8. Liu, Y.; Wang, G.; Wang, Z.; Yang, F.; Wu, G.; Hong, N. Identification of Differentially Expressed Genes in Response to Infection of a Mild Citrus Tristeza Virus Isolate in Citrus Aurantifolia by Suppression Subtractive Hybridization. Sci. Hortic. 2012, 134, 144– 149, link
9. Bowman, K.; Albrecht, U. Comparison of Gene Expression Changes in Susceptible, Tolerant and Resistant Hosts in Response to Infection with Citrus Tristeza Virus and Huanglongbing. Journal of Citrus Pathology 2015, 2, 2, link
10. Chin, E. L.; Ramsey, J. S.; Mishchuk, D. O.; Saha, S.; Foster, E.; Chavez, J. D.; Howe, K.; Zhong, X.; Polek, M.; Godfrey, K. E.; Mueller, L. A.; Bruce, J. E.; Heck, M.; Slupsky, C. M. Longitudinal Transcriptomic, Proteomic, and Metabolomic Analyses of Citrus Sinensis (L.) Osbeck Graft-Inoculated with “Candidatus Liberibacter Asiaticus”. J. Proteome Res. 2020, 19 (2), 719– 732, link
11. Love, M. I., Anders, S., & Huber, W. (2021, October 26). *Analyzing RNA-seq data with DESeq2*. Bioconductor. Retrieved from http://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html#count-matrix-input.

12. Yu, G., Wang, L.-G., Hu, E., Chen, M., & Dall'Olio, G. (n.d.). *clusterProfiler*. Bioconductor. Retrieved December 2021, from https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html.
13. Citrus Genomics & Genetic Improvement Lab, Huazhong Agricultural University. (2021, June). *CPDB: Citrus Pan-genome to Breeding Database*. Citrus Pan-genome to Breeding Database. Retrieved December 2021, from http://citrus.hzau.edu.cn/.
14. Maechler, M., Rousseeuw, P., Struyf, A., & Hubert, M. (2021, April 16). *"Finding Groups in Data": Cluster Analysis extended Rousseeuw et al.* The Comprehensive R Archive Network. Retrieved December 2021, from https://cran.r-project.org/web/packages/cluster/.
15. Wilkerson, D. M, Hayes, Neil D (2010). "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking." *Bioinformatics*, **26**(12), 1572-1573. http://bioinformatics.oxfordjournals.org/content/26/12/1572.abstract.
16. Pathak, M. (2018, July 24). *Hierarchical clustering in R: Dendrograms with hclust*. DataCamp Community. Retrieved December 2021, from https://www.datacamp.com/community/tutorials/hierarchical-clustering-R.