

TRANSCRIPTOME PROFILING OF TOMATO RESPONSE TO PHYTOPLASMA INFECTION

Mitina Irina¹, <https://orcid.org/0000-0001-7103-6084>,
Grajdieru Cristina¹, <https://orcid.org/0000-0003-1560-7924>,
Bahsiev Aighiuni¹, <https://orcid.org/0000-0002-1915-7652>,
Tumanova Lidia¹, <https://orcid.org/0000-0001-5664-871X>,
Zamorzaeva Irina¹, <https://orcid.org/0000-0002-6302-2069>,
Mitin Valentin¹, <https://orcid.org/0000-0001-9328-9672>,
Juříček Miloslav²,

Muller Karel^{2,*}, <https://orcid.org/0000-0002-0817-8810>, * e-mail: muller@ueb.cas.cz

¹*Institute of Genetics, Physiology and Plant Protection of the Moldova State University*

²*Institute of Experimental Botany of the Czech Academy of Sciences*

Summary. Transcriptome sequencing is a valuable tool, which allows the researchers to study the dynamics of gene expression and identify genes, which control economically important traits. In this work, genes involved in tomato response to ‘*Candidatus Phytoplasma solani*’, a bacterial pathogen which can cause severe economic losses, were studied by transcriptome sequencing. As a result, 143 genes were found to be differentially expressed in the infected tomato plants, compared to healthy plants, 81 were upregulated and 62 downregulated. This analysis can shed light on the mechanisms of phytoplasma induced host gene expression regulation and on tomato response to phytoplasma infection.

Keywords: ‘*Candidatus Phytoplasma solani*’, transcriptomics, gene expression, RNA

Introduction

Throughout history, plant diseases have posed significant threats to agricultural output by causing substantial food losses annually while also compromising product quality [1]. The estimated potential yield losses caused by plant pathogens is up to 16% globally and most research in plant pathology aims to reduce yield loss in crops directly or indirectly [2]. In recent years, responding to the needs of ever-growing human population and taking into account climate change, there is an urge for developing effective disease control strategies. One of the ways to do this is to deploy Integrated Crop Management (ICM) strategies, relying less on synthetic inputs for fertilization and crop protection and less on fossil fuel-powered machinery to decrease yield losses due to pest and pathogens and guarantee food security. Another strategy is developing disease resistant cultivars, adapted to future climate changes (reviewed in [3]). Breeding resistant varieties was reported to be the most economical and effective strategy to prevent and control plant diseases [4]. The conventional strategy for resistance breeding has been to introduce disease resistance (R) genes from resistant germplasm into agronomic cultivars, an approach that requires successive generations of recurrent selection for resistance [5, 6]. In recent years, new tools that help plant breeders to detect the genetic diversity and identify key traits, thus speeding up the breeding process, have been developed. Undoubtedly, a tool that revolutionized crop breeding and simplified identification of agronomically important genes is Next Generation Sequencing (NGS). Next-generation sequencing technologies allow the rapid and cost-effective generation of genome-scale data [7]. One of the most common NGS approaches is transcriptome sequencing, which implies assembly of all RNA transcripts from a sample [8]. Transcriptome sequencing allows the researchers to study the dynamics of gene expression and connect the changes in gene expression to phenotypes. In this work, the transcriptome sequencing was used to study the mechanisms of differential response of two tomato varieties to ‘*Candidatus Phytoplasma solani*’ (‘*Ca.P. solani*’)

infection [9]. This pathogen is endemic in Europe and infects a wide range of cultivated plants and weeds [10]. In severe epidemics, yield losses can reach 60% in tomato, 90% in pepper, and 100% in celery [11]. After infection, phytoplasma causes a vast variety of phenotypes in plants, at least some of them are due to the secretion of effector proteins, which induce degradation of plant transcription factors consequently modifying gene expression of the host plants [12]. Understanding differential response of tomato varieties to phytoplasma may shed light on the mechanisms of plant response to this pathogen, and, possibly, to find candidate resistance genes.

Materials and methods

In this work plants of two local tomato varieties – ‘Desteptarea’ and ‘Mary Gratefully’, developed in the Institute of Genetics, Physiology and Plant Protection of Moldova State University were grown in field conditions. Peduncles were collected at mature fruiting stage, snap frozen in liquid nitrogen and stored at -80°C until further processing. Peduncles of three independent plants of each variety both infected with ‘*Ca.P. solani*’ and uninfected were used for RNA extraction. Infection status was initially determined visually and by PCR analysis with previously described primers specific to ‘*Ca.P. solani*’ [13] and later confirmed by sequence analysis. Tissue was grinded in liquid nitrogen, RNA was extracted using GeneJet Plant RNA purification kit (Thermo Fisher Scientific), and RNA quality was checked by agarose gel electrophoresis and NanoDrop Spectrophotometer. Then, RNA was DNase treated with Turbo DNA-free kit (Thermo Fisher Scientific) and sent to sequencing facility of Institute of Experimental Botany, Czech Academy of Sciences for library preparation and next-generation sequencing. The sequencing resulted in at least 40 million 150 bps long read pairs which were quality-filtered using Rcorrector [14] and Trim Galore scripts [15]. Transcripts abundances were determined using Salmon [16] software with parameters --posBias, --seqBias, --gcBias, --numBootstraps 30. Reference index was built from *Solanum tuberosum* CDS dataset, version ITAG4.1. Quality control of quantification and statistical analysis were done using Sleuth package for R [17]. Diagram was created using BioRender. To create GO plot, a list of 143 differentially expressed genes resulted from transcriptomics studies were attributed corresponding Arabidopsis IDs for analysis on DAVID website (<https://david.ncifcrf.gov/>), and the output from DAVID, together with the list of these genes including their q-values and fold-changes were used for analysis and visualization in GPlot package [18] in R.

Results and discussions

To investigate how gene expression in tomato changes in response to ‘*Ca.P.solani*’ infection, the mRNA expression profiles of healthy and infected plants of two varieties (Desteptarea and Mary Gratefully) were analyzed using RNA-seq (RNA sequencing) analysis. Figure 1a illustrates the experimental design, outlining the main steps of RNA-seq analysis. Figure 1b presents the bioinformatics pipeline to detect differentially expressed genes of tomato in response to phytoplasma infection.

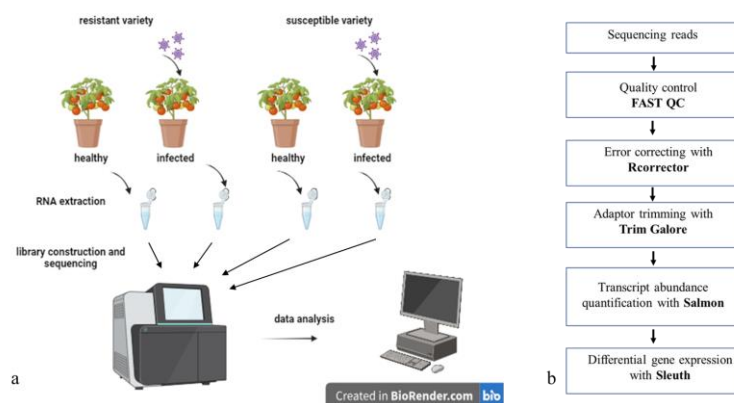


Figure 1. Diagram of the experimental set up. (a) main steps of RNA seq analysis, (b) data analysis pipeline

The distributions of all samples for normalized read counts were compared and are shown in figure 2 as a heatmap (a) and Principal Component Analysis (PCA) plot (b). Despite the relatively poor sample clustering, that can be attributed to the uncontrolled conditions in the field, and the possibility of other stresses apart from phytoplasma infection, 143 genes were significantly ($q\text{-value} \leq 0.05$) differentially expressed in infected samples compared to uninfected plants of both varieties (Figure 3).

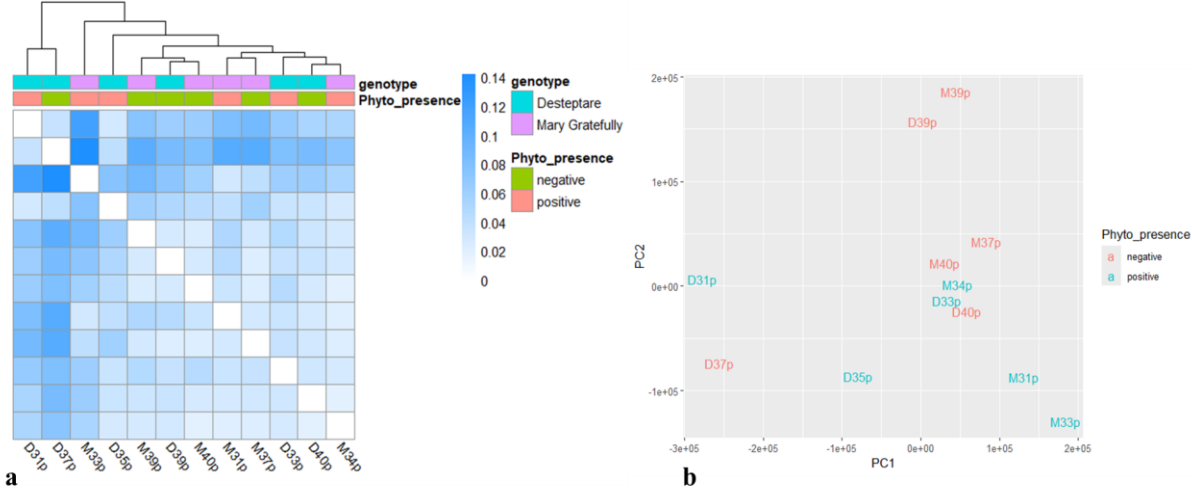


Figure 2. Assessment of transcriptomic data: (a) Heatmap, (b) Principal Component Analysis (PCA)

Of these genes, 81 were upregulated in the infected plants, compared to healthy samples, while 62 genes were downregulated (fold-change ≥ 2).

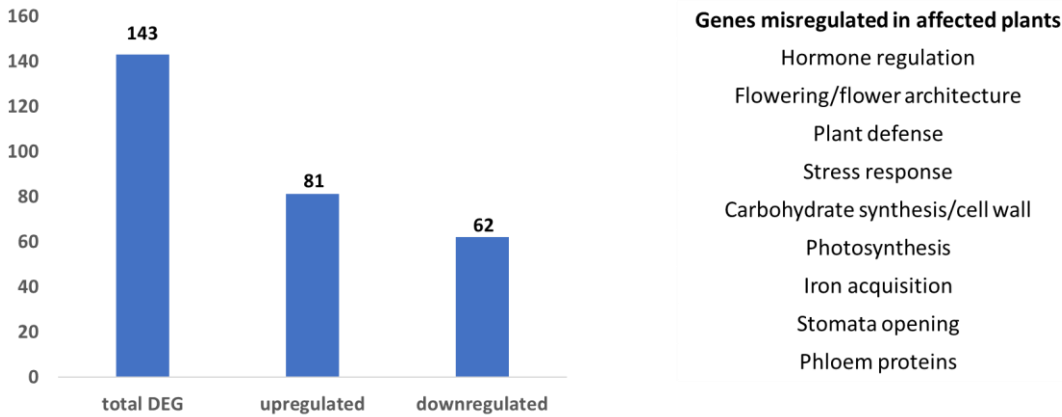


Figure 3. Impact of *Ca. Phytoplasma* infection on tomato transcriptome

Among differentially expressed genes, there are those involved in hormone regulation (Auxin-responsive GH3 family protein gene, TOPLESS 3), flowering control or flower architecture (APETALA2d), plant defense (NADH dehydrogenase genes, disease resistance protein (TIR-NBS-LRR class) family), stomata opening (Stomatal closure-related actin-binding protein 2 gene), etc. This is consistent with the multiple phenotypes that phytoplasma infection produce in tomato, and

with previous research on phytoplasma altering host gene expression. However, functional analysis is needed to elucidate exact role of these genes in tomato response to phytoplasma infection.

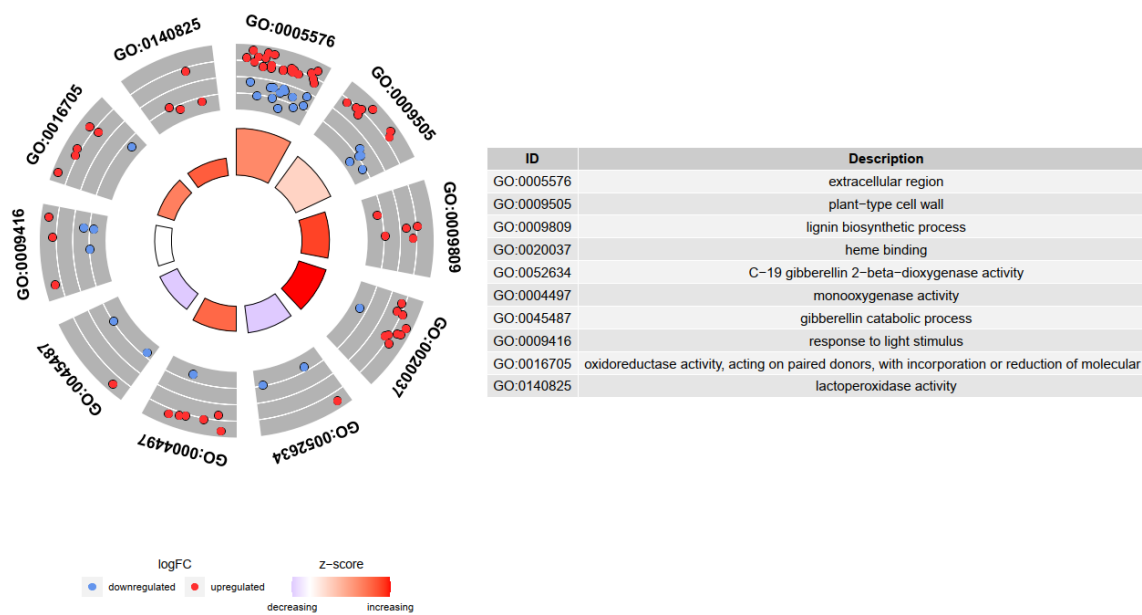


Figure 4. Visualization of functional analysis data

To get some insight into functional analysis of differentially expressed genes, a GO plot combining functional analysis with the expression data was constructed.

The most enriched functional category, in both upregulated and downregulated genes in the infected plants, was the extracellular region, followed by cell wall associated genes. Among differentially expressed genes, those falling in the categories of lignin biosynthesis process and lactoperoxidase activity were exclusively upregulated, those with heme binding activity, monooxygenase activity, oxidoreductase activity and involved in response to light stimulus were mostly upregulated, and those having C-19 gibberellin 2-beta-dioxygenase activity and those involved in gibberellin catabolic process were mostly downregulated.

Together, these data clarify which functional classes of tomato genes are the most affected by phytoplasma infection.

Conclusions

The response of two local tomato varieties ‘Destepare’ and ‘Mary Gratefully’ – to ‘*Ca.P.solani*’ infection was studied by RNA-seq. The samples exhibited high variability, probably due the uncontrolled conditions in the field. In this study, 143 genes were differentially expressed in the infected plants, compared to healthy plants; among them 81 were upregulated and 62 downregulated. This analysis can shed light on the mechanisms of phytoplasma induced host gene expression regulation and on tomato response to phytoplasma infection.

Acknowledgements. This research was supported by Introduction of Transcriptomic Analysis through the Challenge Fund: Advancing Biological Insights on Plant Resistance to ‘*Candidatus* Phytoplasma solani’. UNDP-IRH-00048 Czech Challenge Fund 2023 funded by UNDP and Czech Republic Development Cooperation and Institutional Project, Subprogram 011101 „Genetic and biotechnological approaches to agroecosystem management under conditions of climate change”, funded by the Ministry of Education and Research of the RM.

We would like to thank dr. hab. Nadejda Mihnea for kindly providing plant material.

References

1. Gai, Y.; Wang, H. Plant Disease: A Growing Threat to Global Food Security. In: *Agronomy* 2024, 14, 1615. <https://doi.org/10.3390/agronomy14081615>
2. Ficke A, Cowger C, Bergstrom G, Brodal G. Understanding Yield Loss and Pathogen Biology to Improve Disease Management: Septoria Nodorum Blotch - A Case Study in Wheat. *Plant Dis.* 2018 Apr;102(4):696-707. doi: 10.1094/PDIS-09-17-1375-FE. Epub 2018 Mar 5. PMID: 30673402
3. Richard, B., Qi, A., Fitt, B.D.L. (2022) Control of crop diseases through Integrated Crop Management to deliver climate-smart farming systems for low- and high-input crop production. In: *Plant Pathology*, 71, 187–206. <https://doi.org/10.1111/ppa.13493>
4. Nelson, R., Wiesner-Hanks, T., Wissner, R., and Balint-Kurti, P. (2018). Navigating complexity to breed disease-resistant crops. In: *Nat. Rev. Genet.* 19: 21–33
5. Sanchez-Martin, J., and Keller, B. (2019). Contribution of recent technological advances to future resistance breeding. In: *Theor. Appl. Genet.* 132: 713–732
6. Li, Q., Wang, B., Yu, J., and Dou, D. (2021). Pathogen-informed breeding for crop disease resistance. In: *J. Integr. Plant Biol.* 63: 305–311
7. Zhou, Q., Galindo-González, L., Hwang, S. F., Strelkov, S. E. (2020). Application of genomics and transcriptomics to accelerate development of clubroot resistant canola. In: *Canadian Journal of Plant Pathology*, 43(2), 189–208. <https://doi.org/10.1080/07060661.2020.1794541>
8. Tyagi P., Singh D., Mathur S. et al. Upcoming progress of transcriptomics studies on plants: An overview; *Frontiers in Plant Science*; 13; 2022; <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.1030890>
9. Zamorzaeva I., Mitina I., Bahsiev A., Mitin V., Mihnea N. Impact of ‘candidatus phytoplasma solani’ presence on fruit quality of different moldovan tomato varieties. In: *Phytopathogenic Mollicutes*, 2020
10. Buoso S., Pagliari L., Musetti S. 'Candidatus Phytoplasma solani' interferes with the distribution and uptake of iron in tomato. In: *BMC Genomics*. 2019 Sep 10;20(1):703. doi: 10.1186/s12864-019-6062-x.
11. Navrátil, M., Válová, P., Fialová, R. 2009. The incidence of stolbur disease and associated yield losses in vegetable crops in South Moravia (Czech Republic). In: *Crop Protection*, 28(10) 898–904. <http://www.sciencedirect.com/science/journal/02612194>
12. Omengé KM, Rümpler F, Kathalingam SS, Furch ACU, Theißen G. Studying the Function of Phytoplasma Effector Proteins Using a Chemical-Inducible Expression System in Transgenic In: *Plants. Int J Mol Sci.* 2021 Dec 18;22(24):13582. doi: 10.3390/ijms222413582.
13. Zamorzaeva I., Bahşiev A., Mihnea N. 2016. Testarea primerilor creați pentru diagnosticul molecular al infecției fitoplasmice. In: *Buletinul AȘM. Științele Vieții*. Edit. CSȘDT AȘM. Chișinău. 2: 55-61.
14. Song, L., Florea, L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. In: *GigaSci* 4, 48 (2015). <https://doi.org/10.1186/s13742-015-0089-y>
15. Trim Galore (RRID:SCR_011847)
16. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. In: *Nat. Methods.* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>
17. Pimentel, H., Bray, N., Puente, S. et al. Differential analysis of RNA-seq incorporating quantification uncertainty. In: *Nat Methods* 14, 687–690 (2017). <https://doi.org/10.1038/nmeth.4324>
18. Walter, Wencke, Fátima Sánchez-Cabo, and Mercedes Ricote. GOplot: and R package for visually combining expression data with functional analysis. In: *Bioinformatics*, 2015, 31(17): 2912-2914. doi: 10.1093/bioinformatics/btv300.