Transcriptome analysis of pectobacteria-infected tobacco plants

R.F. Gubaev, V.Y. Gorshkov, O.E. Petrova, A.G. Daminova, N.E. Gogoleva, O.I. Parfirova, Y.V. Gogolev

Kazan Institute of Biochemistry and Biophysics RAS, Lobachevsky str., 2/31, Kazan, Russia, 420111, rim.gubaev@kibb.knc.ru

Plant-microbe interactions are associated with physiological alterations of both host and the pathogen that lead to the formation of the pathosystem. Many pathogen-induced physiological changes of the host plant may be related to activation of defense responses as well as susceptible ones. Defense responses are aimed on the protection of the plant against pathogen propagation. In turn, susceptible responses were shown to be involved the transformation of host environment in a way beneficial for the pathogen.

Necrotrophic pectobacteria, including *Pectobacterium atrosepticum*, are considered as brute force pathogens that cause soft-rot diseases by producing plant cell wall degrading enzymes (1). However the virulence of pectobacteria largely depends on biotrophic stages of the life cycle, when pathogen may manipulate the host. Neither defensive nor susceptible responses of the plants during pectobacteria-induced pathogenesis were characterized. The application of modern omics technologies combined with methods of computational biology is an effective approach for the description of complex biological systems, including pathological ones. In our study, we performed transcriptome profiling of the infected plants in order to describe their pathogen-induced responses. The main advantage of such technique is the ability to separate host and pathogen nucleotide sequences by means of bioinformatic approaches.

Total RNA was extracted from intact and infected plants and corresponding cDNA libraries were prepared. HiSeq 2500 platform was used for high throughput sequencing of cDNA libraries. The obtained bcl files were converted to fastq files using bcl2fastq tool. Bad quality reads as well as rRNA corresponding reads were removed using Trimmomatic and SortMeRna software respectively (2, 3). The quality of pre- and post-processed reads was assessed using FastQC tool. Coding sequences of tobacco genome were used as reference. Since tobacco genome is of the bad quality, redundant and non-sense transcripts were

removed using EvidentialGene package in order to avoid biases during the following steps of analysis. Sorted reads were aligned to tobacco reference transcriptome with Bowtie2 (4). Read counting per transcript per million reads was performed using RSEM package (5).

One of the most important problem in the transcriptome profiling of dynamically developing pathosystems is related to the possible variability of biological replicates. In order to verify relative similarity of biological replicates, hierarchical clustering of CPM values obtained for each gene of each sample was performed in RStudio. Replicates that showed relative similarity were chosen for further analysis of the differentially expressed genes (DEGs) with edgeR package. 3580 up- and 3350 down-regulated genes were revealed in infected plants compared to the control ones.

In order to facilitate and increase the efficiency of the interpretation of transcriptome data, it is reasonable to classify large number of DEGs into functional categories like metabolic pathways and modules. In our experiments, such kind of classification was challenging, since the databases that allow the performance of gene categorization contain information only for a limited number of organisms, and the tobacco is not among them. In order to overcome this problem, we used BLAST+ (6) to compare the tobacco protein-coding sequences with those of potato (closest relative of tobacco), which are classified into biological categories in different databases (KEGG, MapMan, BioCyc). The genes and their expression level values of tobacco were fused to IDs of corresponding proteins of potato. Using a custom script the tobacco genes with assigned potato protein IDs were classified in KEGG metabolic pathways and modules. The merged DEG list was then split into subsets of genes either up- or down-regulated in the course of infection. Hypergeometric testing was applied in order to identify pathways that were significantly enriched with up- or down-regulated gene subsets.

In order to clarify and extend the obtained information on the physiological alterations of the infected plants, we classified DEGs into functional categories represented in MapMan and BioCyc sources. Such an approach allowed us to find additional pathways that were not represented in KEGG and to visualize DEGs on metabolic diagrams and maps.

It has been demonstrated that the *P. atrosepticum*-caused infection is associated with responses related to biotic stresses as well as general metabolism. The infection was

accompanied by the suppression of the processes responsible for plant growth and development as well as the inhibition of general metabolic pathways (glycolysis, photosythesis, biosynthesis of lipids and amino acids). Such kind of reactions likely reflect the deviation of the plant resources normally spent for growth and development to the interaction with the pathogen. This is in accordance with the activation of defense-related reactions (secondary metabolism, biosynthesis of ethylene and antimicrobial proteins) in infected plants. Furthermore, pectobacteria-induced infection was related to the activation of jasmonate (JA)- and repression abscisic acid (ABA)-mediated hormonal systems. These modulations in hormonal signaling to our mind reflect the susceptible responses of the plants to pectobacteria because of the following reasons. First, ABA and ABA-mediated responses were demonstrated to negatively effect pectobacteria virulence (our unpublished data). Second, pectobacteria were demonstrated to activate JA-induced responses in order to colonize their hosts (7). Another set of susceptible responses may be related to the activation of genes that encode enzymes involved in degradation of biopolymers (proteins and carbohydrates) and transport of sugars (invertases). These processes may form a pool of low molecular weight metabolites that increase pathogen's proliferation and progression of the infection. Additionally, plant cell wall modification was demonstrated to be the part of susceptible responses to pectobacteria (8). The transcriptome analysis allowed us to identify the pathogen-induced genes of plant cell wall category, which products are involved in cell wall loosening presumably increasing pathogen's fitness under *in planta* conditions.

Thus, the application of methods of computational biology allowed us to portray physiological features of the infected plants. By using a combination of alternative databases and automated algorithms, large amounts of pectobacteria-regulated processes of plants that may be considered to be either susceptible or defense responses, were identified.

This study was supported by the Russian Science Foundation (№ 15-14-10022).

1. A.O. Charkowski et al. (2014). Genomics of Plant-Associated Bacteria: The Soft Rot Enterobacteriaceae, *Genomics of Plant-Associated Bacteria* **1:**37–58

2. A.M.Bolger et al. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, btu170.

3. E.Kopylova et al. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data, *Bioinformatics*. **28**:3211–3217.

4. B.Langmead, S.Salzberg (2012) Fast gapped-read alignment with Bowtie 2, *Nature Methods,* **9:**357–359.

5. B.Li, C.N.Dewey (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, **12:**323.

6. C.Camacho et al. (2009) BLAST+: architecture and applications, *BMC bioinformatics*, **10:**421.

7. P.Panda et al. (2016) *Pectobacterium atrosepticum* and *Pectobacterium carotovorum* harbor distinct, independently acquired integrative and conjugative elements encoding coronafacic acid that enhance virulence on potato stems, *Frontiers in microbiology*, **7:**1–13.

8. V.Y.Gorshkov et al. (2016) Pathogen-induced conditioning of the primary xylem vessels–a prerequisite for the formation of bacterial emboli by *Pectobacterium atrosepticum*. *Plant Biology*, **18:**609-617.