Differential gene expression in *Lavandula angustifolia* Mill. under adaptation *ex vitro*

Tsyupka V.A., Kuleshova O.N., Emirsaliev A.O., Zhdanova I.V., Grebennikova O.A., Mitrofanova I.V.

Federal State Funded Institution of Science "The Labor Red Banner Order Nikita Botanical Gardens – National Scientific Center of the RAS", Nikita, Yalta, Russia

E-mail: *valentina.brailko@yandex.ru*

INTRODUCTION

Propagation of commercial lavender (***Lavandula angustifolia* Mill.**) cultivars *in vitro* has a number of advantages over seed or vegetative (cuttings) propagation: obtaining a uniform, genetically stable and healthy planting material. At the same time, the effectiveness of this process depends on the subsequent adaptation of the obtained plants to the open-field conditions.

Figure 1. Lavender cultivar 'Prima' *ex situ*

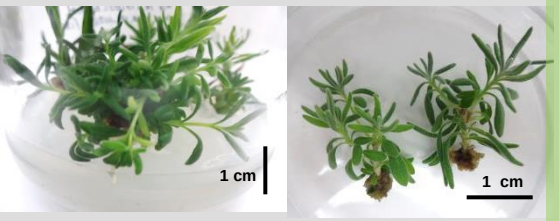

Figure 2. Lavender cultivar 'Prima' *in vitro* (third passage)

MATERIALS AND METHODS

The research was carried out on the 'Prima' cultivar grown in the collection of the Nikita Botanical Gardens (***Fig. 1***). Leaves were collected from plants grown *in vitro* (***Fig. 2***) before transferring them to *in vivo* conditions and from plants after 7 days adaptation in a multilevel plant growth chamber "Conviron" (Canada). Lavender plants were cultured *in vitro* on MS culture medium supplemented with 0.5 mg/L BAP or 0.5 mg/L kinetin and 0.025 mg/L NAA at a temperature of 22-25°C, 14-16-hour photoperiod, light intensity – 25.0-37.5 µmol m^{-2} s^{-1}. To induct rhizogenesis (after third passage) MS medium with different content of auxins, IBA or NAA (0.5-1.0 mg/L) was used. To analyze possible molecular mechanisms, which are the base of the adaptive processes in plants, at the time of transferring from *in vitro* growth to *in vivo* growth the samples of total RNA were isolated from leaves by the NucleoSpin RNA Plant protocol (NucleoSpin, Germany) and 24 barcoded RNA-Seq libraries were created using Illumna TruSeq Stranded mRNA Library Prep Kit. Sequencing was made on a high-performance sequencer Illumina NextSeq 550 using the NextSeq 500 HighOutputv2 Kit with an estimated capacity of at least 90 million reads. As a result of sequencing, 3 libraries of transcriptomic data were obtained for each of the states (*in vitro* and *in vivo*). For the obtained data, pre-preparation, *de novo* transcriptome assembly, transcript-level expression analysis, search for protein-coding regions and their annotation using the SwissProt Viridiplantae database, and evaluation of differential expression were made. The transcripts of up-regulation and down regulation have been identified. The following software packages were used: FastP v. 0.19.5, Trinityv.2.11.0, Kallistov.0.46.1, TransDecoder v.5.5.0, BLASTv.2.11.0, edgeRv.3.32.1.

Figure 3. Lavender cultivar 'Prima' *in vivo* (7 days adaptation under an insulator) (a), plants *in vitro* before planting in the substrate (b)

RESULTS

The following steps of primary bioinformatics processing have been done: grouping of sequences according to their barcodes, obtained data transformation into FASTQ format, data quality assessment. Total amount of the obtained data was **22,01 Gb**, an average part of high quality sequences (Mean Quality Score > 30) was **90,7%** of all data from nine trials.

For further analysis we used readings prepared using **Fastp** software. Data statistical analysis before and after preprocessing (***Table 1, Fig. 4***) let us to conclude that total length of readings was **21,9G b.p.** at the length of the read sequences **35-76 b.p**. before the preprocessing and **20,6G b.p** at the length of the read sequences **15-76 b.p.** after processing. Most of nucleotides reads were of sufficient quality.

Table 1. Results of the general transcriptome assembly

Total trinity 'genes'	Total trinity transcripts	Percent GC	Stats based on ALL transcript contigs				Stats based on ONLY LONGEST ISOFORM per 'GENE'			
			Contig N50	Median contig length	Average contig	Total assembled bases	Contig N50	Median contig length	Average contig	Total assembled bases
91157	224883	42.37	1781	828	1154.22	259563766	1398	410	777.85	70906692

**ValueN50 is adequate (~ 1000 – 2000 bases) for transcriptome assemblies*

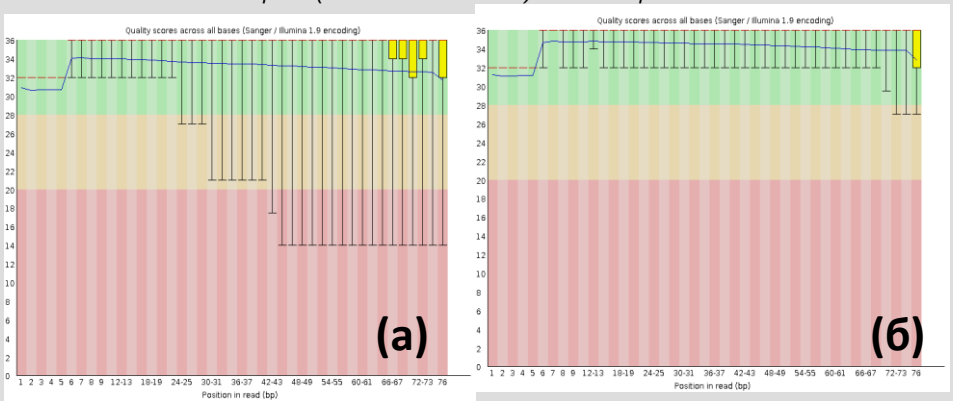

Figure 4. Aggregated statistics of the quality indexes in each position for each nucleotide of the sequence in all readings obtained with FastQC software for the library L8 of direct reading before (a) and after preprocessing (b) with Fastp software.

**X-axis is nucleotide position in reading, Y-axis is the quality index (Q), which is inverse proportional to the probability of the nucleoitide reading mistake during the sequencing, and represented in the decimal logarithmic scale. Values Q > 20, correspond to more than 99% nucleoitide reading accuracy (Q >30 is more than 99,9% reading accuracy). Blue line indicates the mean Q value for each nucleotide in all readings, red line corresponds to Q median for highlighted range on the nucleotide position axis.*
For further analysis we used the libraries prepared with FastP software.

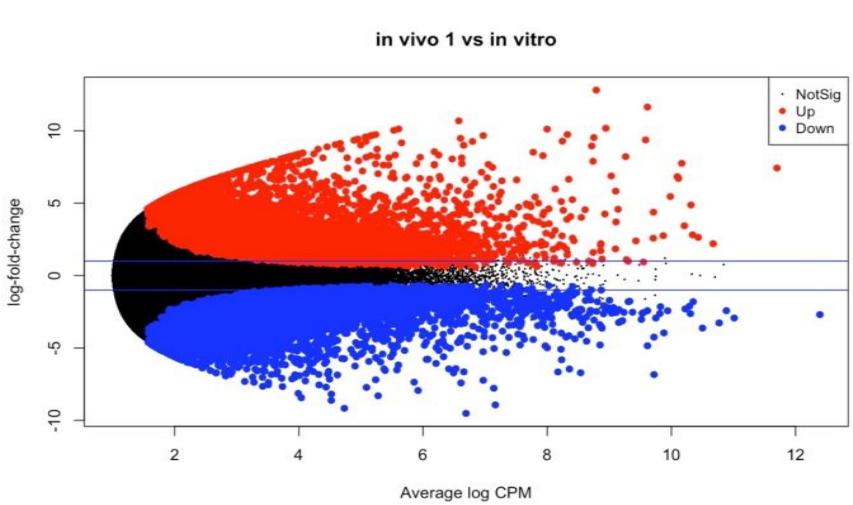

Figure 5. Grouping of differentially expressed genes (Mean-Difference) *in vitro - in vivo*

Based on the obtained values the matrix was made and then analyzed with the software **edgeR v.3.32.1** with the following settings: **Data normalization method – TPM, Test Method - Exact Tests, P adjustment method – False discovery rate (cutoff 0,1)**. The total number of the identified meaningful genes: **13309**, **Up-regulation: 5610, Down-regulation: 7699** (***Fig.5***). Then we made the list of transcripts, which are the most significant for Up-regulation and Down-regulation at the each stage. For the obtained transcriptomes the automatical prediction of the proteins coding areas and their annotation were made. Prediction of functions and metabolic paths by the **data base KEGG** was made using the **web-service BlastKOALA** with the settings: taxonomy group – Plant, KEGG GENES database – family_eukaryotes (*Fig.6*). For all stages of culture and adaptation, proteins coding particular metabolic paths and functional hierarchies were identified and annotated (***Fig.6, Fig.7***).

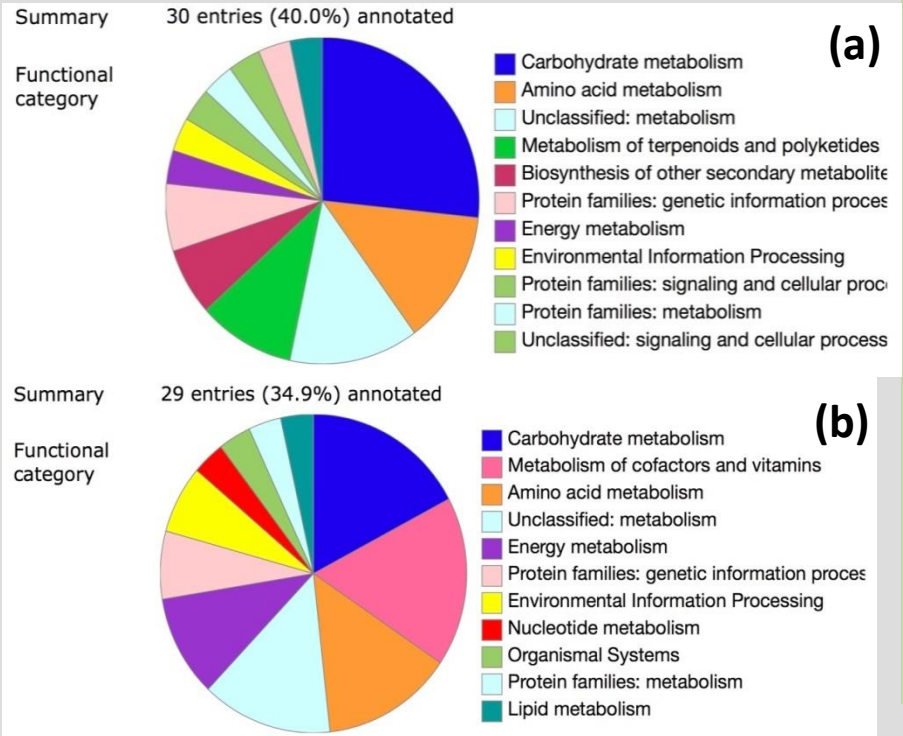

Figure 6. Functional categories of the protein-coding genes of *in vitro* and *in vivo* state predicted in up-regulation (a) и down-regulation (b) and annotated by the data base KEGG Plant.

Based on the obtained transcriptional data, primers for the most significant transcripts of the predicted protein coding sequences in Up-regulation and Down-regulation associated with adaptation processes were developed. They were used for the subsequent assessment of the comparative expression of several genes in vitro and in vivo for 7 days adaptation under a perforated isolator, 14 days adaptation under a perforated isolator, 30 and 60 days adaptation.

The relative level of gene expression was determined by qPCR method, analyzed against the background of the actin reference gene (ACT2). To select working primers, a screening was carried out by amplifying the samples of 7 days adaped plants on cDNA.

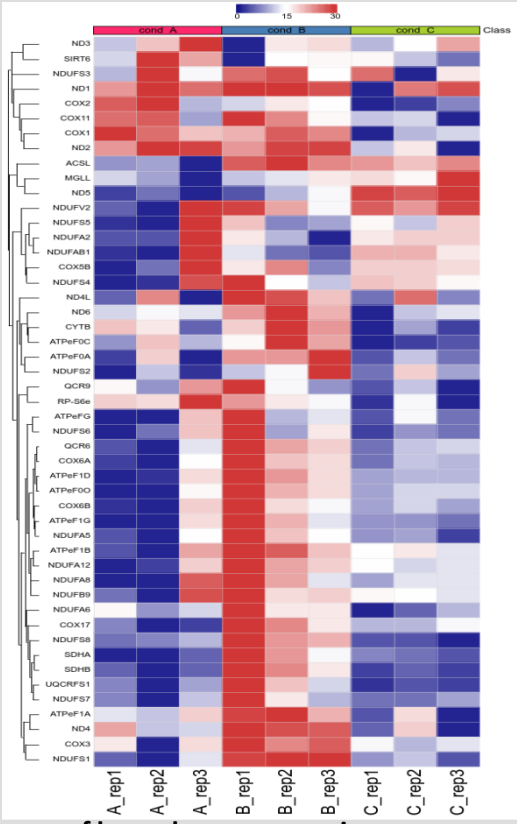

Figure 7. Heat map of lavender co-expression genes, associated with hydrothermal stress under *ex vitro* adaptation: A – *in vitro*, B – *in vivo* adaptation for 7 days, C - *in vivo* adapation for 60 days

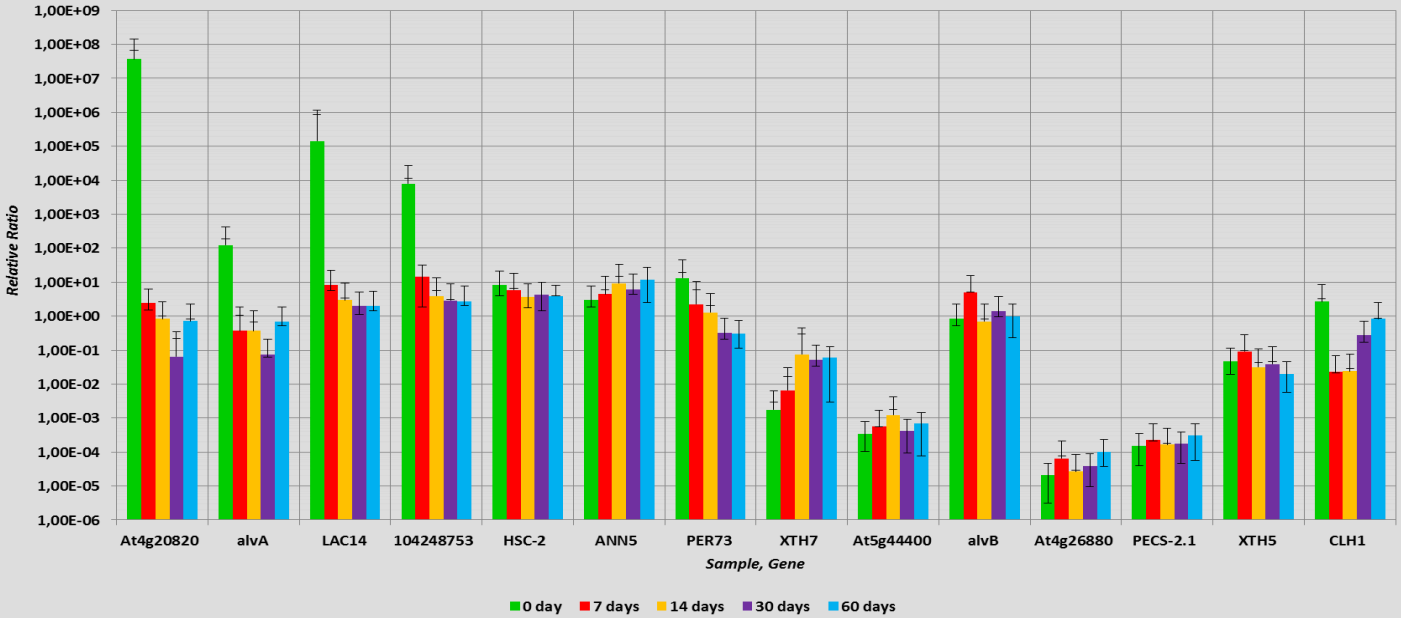

Figure 8. Relative expression of the studied genes, rationed to actin expression (Ralative Ratio in logarithmic terms according to the data of Rel Quant analysis)

The detection results made it possible to select pairs of primers that form the expected amplification product: 3.1.a (for the ACT2 gene), 1.1.u (gene At4g20820), 3.2.u (without annotation, alvA), 5.1.u (LAC14), 6.2.u (At5g44400), 9.1.u (alvB), 11.2.u (GeneID 104248753), 17.2.u (At4g26880), 20.1.u (PECS-2.1), 36.1.u (HSC-2), 44.2.u. (ANN5), 45.1.u. (PER73), 1.2.d (XTH7), 4.2.d (XTH5), 5.2.d (CLH1).

We found two differentially expressed genes that do not have annotations in the databases (however, they are presented in other plants); because of their expression in lavender plants under in vivo adaptation, it was supposed that they are involved in the processes of antioxidant protection, signaling functions; they have been given the working names alvA, alvB (from "adaptation lavandula in virto"). Considering the relatively low level of signaling systems gene expression during 30- and 60-day adaptation, we can reasonably suppose that the adaptation processes had been completed, the plants were ready for planting and successful cultivation in the open field.

The relative level of gene expression was determined by the qPCR method, analyzed against the background of the actin reference gene (ACT2). It was determined that the most highly expressed plant genes in vitro are: At4g20820, LAC 14, id 104248753, alvA, PER73, HSC-2, CLH1, ANN5. At the early stages of in vivo adaptation, the relative expression of the genes alvB, 104248753, LAC 14, At4g20820, PER73, and HSC-2 was higher. A linear increase in the relative gene expression under gradual adaptation in vivo was determined for ANN5, XTH7. For genes At5g44400, At4g26880, PECS-2.1, XTH5, a low (trace) relative level of expression was noted.

CONCLUSION:

It has been demonstrated that under in vitro conditions lavender plants express genes responsible for cell wall biosynthesis, growth, oxidative and temperature stress. Regulation of the adaptation processes is due to the activity of peroxidase, oxidoreductase, heat shock proteins, thickening of cell walls, stomata activity. Considering relatively low level of the signal system genes expression during 30- and 60-days adaptation, we suppose that adaptation processes had been completed and plants were ready to planting and successful cultivation in the open field.

The studies were funded by research grant No. 19-76-00023 of the Russian Science Foundation and SA No. 0829-2019-0038 of the FSFIS "NBG-NSC" on the basis of the Unique Scientific Installation "FITOBIOGEN" FSFIS "NBG-NSC".