De novo transcriptome analysis of two early developmental stages of shiitake mushroom

L.M. Gapa[1,2], V.Y. Gorshkov[1,2], R.F. Gubaev[1,2], N.E. Gogoleva[1,2], E.P. Vetchinkina[3], Y.V. Gogolev[1,2]

[1]*Kazan Institute of Biochemistry and Biophysics RAS, Lobachevsky str., 2/31, Kazan, Russia, 420111*

[2]*Kazan Federal University, Kremlyovskaya str., 18, Kazan, Russia, 420111*

[3]*Institute of Biochemistry and Physiology of Plants and Microorganisms RAS, Prospekt Entuziastov, 13, Saratov, Russia, 410049*

gapalm@yandex.ru

Lentinula edodes, also known as shiitake, is an edible basidiomycota mushroom cultivated all around the globe (1). Little is known about the mechanisms controlling its morphogenesis and fruiting. In attempt to describe the shiitake physiological processes, which may take part in early stages of morphogenesis, the transcriptomic analysis of white non-pigmented mycelium and brown mycelial mat of shiitake was performed. Since there is no well-annotated genome available for *L. edodes*, *de novo* transcriptome was assembled and used as a reference in order to reveal the differentially expressed genes (DEGs) and classify them into functional categories. Such approach allowed us to find physiological processes that took part in the early stages of morphogenesis.

Total RNA was extracted in two biological replicates for each of four different developmental stages: white non-pigmented mycelium, brown mycelial mat, primordium, and fruiting body. Different physiological stages were used in order to reconstruct the most enriched pool of transcripts. RNA samples were used for preparation of the cDNA libraries. HiSeq 2500 platform was used for high throughput sequencing of the cDNA libraries. Eight sets of the obtained sequence data (bcl files) were converted to fastq files using bcl2fastq. Since the bad quality reads and rRNA-corresponding reads may lead to significant biases during transcriptome assembly and DEG analysis, inappropriate reads were removed with Trimmomatic and SortMeRna packages (2,3).

About 120 million of filtered reads were used in order to reconstruct the reference transcriptome using Trinity assembler (4). Four different assemblies were generated by changing the assembler's parameters: k-mer length and contig reconstruction algorithm. An optimal assembly was chosen based on the following features: number of transcripts, N50 length. We also have performed assembly quality control using BUSCO (5), which searches products of the orthologous genes from OrthoDB among assembled transcripts. Thus, we have chosen the best assembly, which contained 16231 contigs. The chosen assembly was filtered by EvidentalGene utility (https://sourceforge.net/projects/evidentialgene/) in order to remove non-coding and repeated

transcripts, since such transcripts may significantly impact assembly annotation as well as the following identification of DEGs.

Filtered contigs were annotated using BLAST+ (6) by comparing assembled transcripts against protein datasets: NCBI fungi protein data base and UniRef90 data base. InterProScan was used for domain search and classification of protein superfamilies. As a result, 94% of transcripts were annotated.

Two sets in two biological replicates of filtered reads that corresponded to white mycelium and brown mycelial mat were aligned along the assembled transcripts with Bowtie2 ungapped aligner (7). Tuxedo (8) pipeline utilities (Cufflinks, Cuffmerge, Cuffdiff) were used in order to count RPKM metrics for each gene in each condition and then to identify differentially expressed ones. Eight thousand DEGs were revealed in the brown mat stage compared to white mycelium.

Since there is no genomic or transcriptomic resources for shiitake that may facilitate functional analysis of DEGs, we additionally compared assembled transcripts by means of BLAST+ with the proteins of closely related to *L. edodes* organisms, for which the metabolic pathways and gene modules were constructed in KEGG and BioCyc. The shiitake`s DEGs with assigned protein IDs of *Agaricus bisporus* (closest relative species with classified proteins in KEGG) and *Penicillum rubens* (fungi with interactive metabolic map available on BioCyc resource) were then classified into metabolic pathways and functional categories. By using custom scripts, we performed hypergeometric testing in order to discover KEGG gene categories significantly enriched with DEGs. Furthermore, repressed/activated DEGs were visualized in metabolic maps and additional pathway enrichment analysis was carried out using BioCyc resource.

The application of the bioinformatic approaches mentioned above allowed us to reconstruct and annotate the reference transcriptome, that wouldl be further used for the description of all morphological stages of shiitake mushroom. In this study, we have obtained the information on a broad number of physiological processes differentially regulated during two early stages of morphogenesis in shiitake. It has been demonstrated that the transition from the stage of white mycelium to brown mycelial mat stage is related to the activation of biosynthesis of cell wall components and pigments as well as of regulatory systems. In turn, processes related to cell division (replication, protein synthesis), antibiotic biosynthesis were repressed. Detailed information on the morphogenesis-related processes in shiitake will be discussed during the presentation.

This study was supported by the Russian Science Foundation (№ 15-14-10022) and Russian Foundation For Basic Research (№ 15-04-02926).

1. A.C Campbell. (1999) The commercial exploitation of the white rot fungus *Lentinula edodes* (shiitake), *International Biodeterioration & Biodegradation*. **43**:101–107

2. A.M.Bolger *et al*. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*. **30**: 2114–2120.

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

4. B.J.Haas *et al*. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, *Nat Protoc*. **8**:1494–1512.

5. F.A.Simão *et al* (2015). BUSCO : assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics.* **31**:3210–3212.

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

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

8. C.Trapnell *et al*. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, *Nat Protoc*. **7**:562–578.