Genome *de novo* sequencing, assembly and functional annotation of pathogenic fungi *Armillaria borealis*

V. Akulova[1]*, V. Sharov[1], Yu. Putintseva[1], N. Oreshkova[1, 2], S. Feranchuk[1, 3, 4],
D. Kuzmin[1], I. Pavlov[1, 2], K. Krutovsky[1, 5, 6, 7]

[1] *Siberian Federal University, Krasnoyarsk, Russia*

[2] *V.N. Sukachev Institute of Forest SB RAS, Krasnoyarsk, Russia*

[3] *Irkutsk National Research Technical University, Irkutsk, Russia*

[4] *Limnological Institute SB RAS, Irkutsk, Russia*

[5] *Georg-August University of Göttingen, Göttingen, Germany*

[6] *Vavilov Institute of General Genetics RAS, Moscow, Russia*

[7] *Texas A&M University, College Station, Texas, USA*

* *e-mail: vfedotova@sfu-kras.ru*

Motivation and Aim: The forest decline is observed almost everywhere as a result of negative anthropogenic and climatic effects, often aggravated by pests, fungi and other phytopathogens. The pollution, increased average annual temperature, decreased precipitation, more frequent droughts and other climatic extremes can weaken trees and make fungi much more destructive. The forest conservation has become a very serious problem, since the scale of tree death caused by the phytopathogenic fungi is enormous. *Armillaria borealis* (Marxm. & Korhonen) is a fungi from the Physalacriaceae family (Basidiomycota) widely distributed in Siberia and the Far East and is also causing the root rot disease that weakens and often kills woody plants. Our goal was to *de novo* sequence, assemble and characterize the genome of *Armillaria borealis* and to generate data that can be used to identify the fungi virulence factors.

Methods and Algorithms: The fungi material was collected from active mycelia of *A. borealis* taken from the *Abies sibirica* trees died in 2015. DNA was sequenced using the 250-bp insert paired-end libraries on the Illumina MiSeq platform at the Laboratory of Forest Genomics of the Siberian Federal University. A *de novo* genome assembly was performed using the SPAdes genome assembler. Protein coding regions were identified in the genome using Exonerate. The EVidenceModeler and Augustus software were used to predict genes using gene models. Finally, the functional annotation was done using predictions as well as protein and transcript alignments and assignments based on PFAM, InterPro and GO ontology.

Results: The *A. borealis* genome assembly contained ~79 Mbp and was comparable with 60 and 84 Mbp for the *A. ostoyae* and *A. gallica* genomes, respectively. The N50 for contigs equaled 15659 bp. Functional annotation revealed 6703 protein coding genes, which was also comparable with 7797 and 8261 in *A. ostoyae* and *A. gallica*, respectively, and provided important data for further comparative analysis.

Conclusion: We are currently reconstructing metabolic pathways of *Armillaria* core genes and pathogenicity. This genome study provides much needed knowledge regarding the woody plant fungal pathogenicity, and useful insights towards identifying specific genes associated with pathogenesis and other metabolic functions.