DOI 10.18699/WIBSB-2018-04

De novo sequencing, assembly and annotation of *Armillaria borealis* genome

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, TX 77843-2138, USA*
** e-mail: vfedotova@sfu-kras.ru*

Key words: genome assembly, functional annotation, fungal pathogenicity

Motivation and Aim: Massive forest decline as a result of negative anthropogenic and climatic effects, often aggravated by pests, fungi and other phytopathogens, has been observed almost everywhere. Environmental changes can weaken trees and make fungi more destructive. Forest conservation has become a serious issue, since the scale of tree death caused by phytopathogenic fungus is enormous. *Armillaria borealis* (Marxm. & Korhonen) is a fungi from the *Physalacriaceae* family 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 sequence *de novo*, assemble and characterize the genome of *Armillaria borealis* and to obtain data that can be used to identify the fungi virulence factors, such as target genes. We also intend to provide population genetics with reference material to study forest populations of *Armillaria spp*.
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 SibFU. To evaluate the completeness of the gene set and assembly, BUSCO was performed using *Basidiomycota* odb9 base. Coding regions were identified in the genome using Exonerate; the EVidenceModeler and Augustus software were used to predict genes. Finally, a functional annotation was done using predictions, protein and transcript alignments, and assignments based on PFAM, InterPro and GO ontology.
Results: The *A. borealis* genome assembly contained ~69 Mbp and was comparable with 60 and 84 Mbp for the *A. ostoyae* and *A. gallica* genomes, respectively. The N50 for contigs equaled 15,659 bp. BUSCO results showed that 94.8 % of reference genes were captured as complete single-copy BUSCOs. Functional annotation revealed 6,703 protein coding genes, which was comparable with 7,797 and 8,261 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 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.
Acknowledgements: This work was supported by research grant No. 14.Y26.31.0004 from the Government of the Russian Federation.