Sequencing and assembly of mitochondrial genomes in three conifer species *Larix sibirica*, *Pinus sibirica* and *Pinus sylvestris*

E. Bondar[1]*, A. Kirichenko[1], V. Sharov[1], Yu. Putintseva[1], N. Oreshkova[1, 2],
S. Feranchuk[3, 4], Yu. Konstantinov[5], V. Shmakov[5], V. Belkov[5], D. Kuzmin[1],
S. Sadovsky[1], K. Krutovsky[1, 6, 7, 8]

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

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

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

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

[5] *Siberian Institute of Plant Physiology and Biochemistry SB RAS*

[6] *George-August University of Göttingen, Göttingen, Germany*

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

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

* e-mail: bone-post@ya.ru

Motivation and Aim: Conifers are ancient group of plants of great economic value and ecological importance represented by more than 600 species including Siberian larch (*Larix sibirica*), Siberian pine (*Pinus sibirica*) and Scots pine (*Pinus sylvestris*) that are keystone species in Siberian boreal forest. Due to the huge sizes both nuclear and mitochondrial genomes are still understudied in conifers. Only a few nuclear and mitochondrial genomes in conifers have been sequenced and published so far. We recently sequenced and assembled nuclear and mitochondrial genomes in two Siberian conifers – *L. sibirica* and *P. sibirica* and in *P. sylvestris*.

Methods and Algorithms: Sequencing was performed using Illumina HiSeq2000 and DNA extracted from purified mitochondria isolated from needles. CLC Assembly Cell was used to assemble paired-end reads into contigs, and the gap-filling program Sealer was used to build scaffolds. To select mitochondrial sequences from nuclear and plastid sequences, assembled contigs were mapped to the nucleotide database of complete and partial plant mitochondrial genomes. To further verify putative mitochondrial contigs in *L. sibirica*, *k*-means clustering in R was used, and the obtained assembly was compared to the whole genome sequencing data. RepeatModeler and RepeatMasker were used to identify repeats.

Results: The Siberian pine mitochondrial genome assembly contained 3.22 Mbp with N50 of 598,637 bp and average GC of 41.78 %. The Scots pine assembly contained 2.34 Mbp with N50 of 289,815 bp and average GC of 42.31 %. The Siberian larch assembly contained 53 scaffolds with a total length of 11.37 Mbp, N50 of 578,009 bp and the largest scaffold of 1,423,047 bp.

Conclusion: The Siberian and Scots pine mitochondrial genome lengths were similar to those earlier studied in other conifers – spruces *Picea glauca* and *P. abies*, while the mitochondrial genome of Siberian larch was much larger and similar to an annual flowering plant *Silene conica* (11.3 Mbp). It does not contain many genes, and repeated and low complexity regions cover only 14.46 % of the sequence. The verification and detailed annotation of the assembled mitochondrial genomes are in progress.