

Genomes of three conifer species: *Larix sibirica*, *Pinus sibirica* and *Pinus sylvestris*

Yu. Putintseva^{1*}, V. Sharov¹, D. Kuzmin¹, N. Oreshkova^{1,2}, S. Feranchuk^{3,4}, V. Biryukov¹, S. Novikova¹, K. Miroshnikova¹, S. Makolov¹, M. Sadovsky¹, K. Krutovsky^{1,5,6,7}

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 George-August University of Göttingen, Göttingen, Germany

⁶ Vavilov Institute of General Genetics RAS, Moscow, Russia

⁷ Texas A&M University, College Station, USA

* e-mail: yaputintseva@mail.ru

Key words: conifers, *de novo* assembly, genome, Siberian larch, Siberian pine, Scots pine

Motivation and Aim: The enormous size of conifer genomes makes the task of sequencing them extremely difficult because of both amount of sequencing data needed for good coverage and huge computer resources needed to process them. However, the reference genomes of conifers are much needed for further studies of evolutionary, biochemical and physiological processes in these organisms. Only a few nuclear genomes of conifers have been sequenced and published so far. We recently sequenced and assembled nuclear genomes in three conifer species – *Larix sibirica*, *Pinus sibirica* and *P. sylvestris*.

Methods and Algorithms: Sequencing was performed using the Illumina HiSeq2000 and MiSeq platforms. We developed an original stepwise method of *de novo* assembly by parts (sets), which allowed us to bypass the limitations of modern assemblers associated with a huge amount of data being processed. We performed the scaffolding with program BESST and scaffolding with RNA reads using the RaScaf program. The gap-filling program Sealer was used also to improve the assembly. RepeatModeler and RepeatMasker were used to identify repeats. Genome annotation was performed using the MAKER2 pipeline.

Results: The Siberian larch genome assembly contained 12.34 Gbp with N50 of 6,443 bp and average GC of 35.41 %. The Siberian pine draft assembly contained 13.56 Gbp with N50 of 6,920 bp and average GC of 36.6 %. The Scots pine draft assembly contained 14.79 mln scaffolds with a total length of 7,8 Gbp, N50 of 654 bp.

Conclusion: Using the new stepwise *de novo* assembling method, the genome of Siberian larch, *Larix sibirica* Ledeb. (12.34 Gbp) was for the first time completely assembled *de novo*. It is the first genome assembly for any larch species. We also present draft genome assemblies for two pine species – *P. sibirica* and *P. sylvestris*.

Acknowledgements: This work was funded by a research grant no. 14.Y26.31.0004 from the Government of the Russian Federation.