Abstract

Symmetry in the Transcriptome and Genome Sequences of Siberian Larch (*Larix sibirica* Ledeb.) †

Vladislav V. Biriukov [1,*], Michael G. Sadovsky [2], Yuliya A. Putintseva [1], Natalia V. Oreshkova [1,3] and Konstantin V. Krutovsky [1,4,5,6]

[1] Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 660036 Krasnoyarsk, Russia; yaputintseva@mail.ru (Y.A.P.); oreshkova@ksc.krasn.ru (N.V.O.); konstantin.krutovsky@forst.uni-goettingen.de (K.V.K.)

[2] Institute of Computational Modelling, Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, 630036 Novosibirsk, Russia; msad@icm.krasn.ru

[3] Laboratory of Forest Genetics and Selection, V. N. Sukachev Institute of Forest, Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

[4] Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany

[5] Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia

[6] Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, USA

* Correspondence: vladislav.v.biriukov@gmail.com

† Presented at Symmetry 2017—The First International Conference on Symmetry, Barcelona, Spain, 16–18 October 2017.

We studied the structuredness and order in draft assemblies of the Siberian larch (*Larix sibirica* Ledeb.) transcriptome and genome. Assembled contigs were converted into the triplet frequency dictionaries followed by their cluster analysis. The transformation maps the contigs into a metric space. We used *K*-means clusterization to analyze the transcriptome and genome structures, which were visualized with the elastic map technique. The observed clusters demonstrated several interesting symmetries, also including those in the statistical and combinatorial properties. For instance, clustering in two classes yields stable classification with a good separability of clusters. We have checked the pattern of the second Chargaff's parity rule implementation for pairs of palindromic triplets within the first and the second classes and between them. It was found that the transcriptome sequences had significantly less discrepancy between the classes compared to genome sequences. Probably, it was due to the presence of contigs in the transcriptome assembly representing RNA sequences transcribed from opposite strands. Unlike the transcriptome, genome contigs demonstrated an inverse pattern of discrepancies regarding the second Chargaff's parity rule. The hypothesis was confirmed by BLAST. In addition, an unexpected octahedral structure exhibiting rotational symmetry was discovered in triplet frequency dictionary ensembles.