Abstract

Symmetry in a Four-Cluster Triangle Structure of the Siberian Stone Pine (*Pinus sibirica* Du Tour) Transcriptome [†]

Serafima V. Novikova [1,*], Michael G. Sadovsky [2], Yulia 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, Buesgenweg 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: kempfana@mail.ru

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

The RNA isolated from three different tissues of Siberian stone pine—cambium, needles and buds—were sequenced and assembled for each tissue separately. The transcriptomes were converted into the triplet frequency dictionaries (FD) of two types. The first type FD allows to detect the strand specificity of clusters via the Chargaff's second parity rule. The clusters were identified using *K*-means applied to the FD of contigs. We observed a four-cluster triangle structure in the distribution of the FDs. The observed symmetry of the clusters was apparently based on the strand specificity of contigs.

The second FD type was developed separately only for well annotated and sufficiently long contigs for each transcriptome. Each sequence has been tiled using a 300 bp long moving window with a 8 bp step and converted into a triplet FD. Then, each FD has been labelled by phase index. The fragments falling out of a coding region were labelled by 0. Otherwise, they were labelled according to the reading frame shift from 1 to 3, respectively. All the FDs were clusterized by *K*-means, yielding a four-cluster pattern, where the clusters included the FDs with the same phase index. A highly symmetric four-cluster triangle pattern was observed.

Conflicts of Interest: The authors declare no conflict of interest.