11th Conference on Dynamical Systems Applied
to Biology and Natural Sciences DSABNS 2020
Trento, Italy, February 4-7, 2020

HOMEOBOX GENES: INVESTIGATING THE DEVELOPMENT OF *PINUS SYLVESTRIS* (SCOTS PINE)

Guseva T.A.*[1],
Biriukov V.V.[1,2] and Sadovsky M.G.[1,3]

[1]Siberian Federal University, Krasnoyarsk, Russia

[2]Laboratory of Genomic Research and Biotechnology,
Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch
of the Russian Academy of Sciences",
Krasnoyarsk, Russia

[3]ICM SB RAS, Krasnoyarsk, Russia

dianema2010@mail.ru (*corresponding author),
biryukov.vv@ksc.krasn.ru, msad@icm.krasn.ru

Homeobox genes play an important role in the processes of the organism developing and have been discovered in plants, fungi, and vertebrates [1]. The aim of the work is to examine the development of Scots pine (*Pinus sylvestris* L.) at the homeobox-containing genes level using the differential gene expression approach.

Previously, the WUSCHEL- related homeobox gene family in *Pinus pinaster* was analyzed [2].

To date, there are a great number of tools and methods for the de novo transcriptome analysis [3]. *P. sylvestris* sequencing data deposited in the NCBI BioProject database under accession number PRJNA531617 (SRR8996768-SRR8996761) provided by the Norwegian Institute of Bioeconomy Research was used as a source of this study [4]. *P. sylvestris* transcriptome was assembled using reads from five tissue types (needle, phloem, vegetative bud, embryo and megagametophyte) from six non-related individuals of *P. sylvestris*. For the *de novo* transcriptome assembly reads were preliminarily trimmed with the Q¿ 30 using Trimmomatic-0.36 and assembled using Trinity software (version 2.8.4). HMMER software (version 3.2.1) was used for the homeobox domain identification in the assembled transcripts using a hidden Markov model of homeodomain which was downloaded from the PFAM database (Accession number PF00046, ID Homeodomain). To estimate transcript abundance and perform differential expression (DE) analysis, align_and_estimate_abundance perl script, RSEM, as well as bowtie programs (version 1.2.3), were used. Also, cross-sample TMM normalization was performed. Finally, the EdgeR package (R version 3.5.0, Bioconductor version 3.8) was used for statistical analysis and identifying significantly differentially expressed transcripts.

Assembled transcriptome comprises of 775,502 transcripts with the mean GC content of 40.19%. The N50 value for transcripts was 1,273 bp, while the median contig length was 360 bp. 417 homeobox-containing transcripts were selected for further analysis,254 transcripts were annotated using the UniProt database (e-value ¡1e-3).

 ISBN: 978-989-98750-7-4

46 statistically significantly differentially expressed trinity genes were identified (adjusted p-value¡0.03) (figure 1) and visualized as a heatmap. Five resulting clusters of DE transcripts were analyzed. In the first cluster, two transcripts encoding proteins involved in early and late embryo development were found, which were upregulated at embryo and megagametophyte stages and downregulated at other stages. The second cluster comprises transcripts encoding transcriptional factors (TF) involved in seed and leaf development as well as in cell differentiation, downregulated in phloem tissue. In the third cluster, transcripts annotated as proteins responsible for multicellular organism development, phloem histogenesis, auxin-mediated morphogenesis, plant organ development, regulation of leaf morphology and involved in meristem formation were observed. These transcripts are downregulated in needle and megagametophyte tissues. Transcripts in cluster four included proteins involved in the detection and cellular response to the cytokine stimulus and leaf morphogenesis. Transcripts of this cluster are significantly upregulated in needles and downregulated in megagametophyte tissues. Cluster five comprises five transcripts encoding proteins determining cotyledon morphogenesis, response to gibberellin, cell differentiation and positive regulation of transcription. These transcripts are upregulated in bud tissue.

Thus, the presented results gain a more comprehensive understanding of the *P. sylvestris* development process.

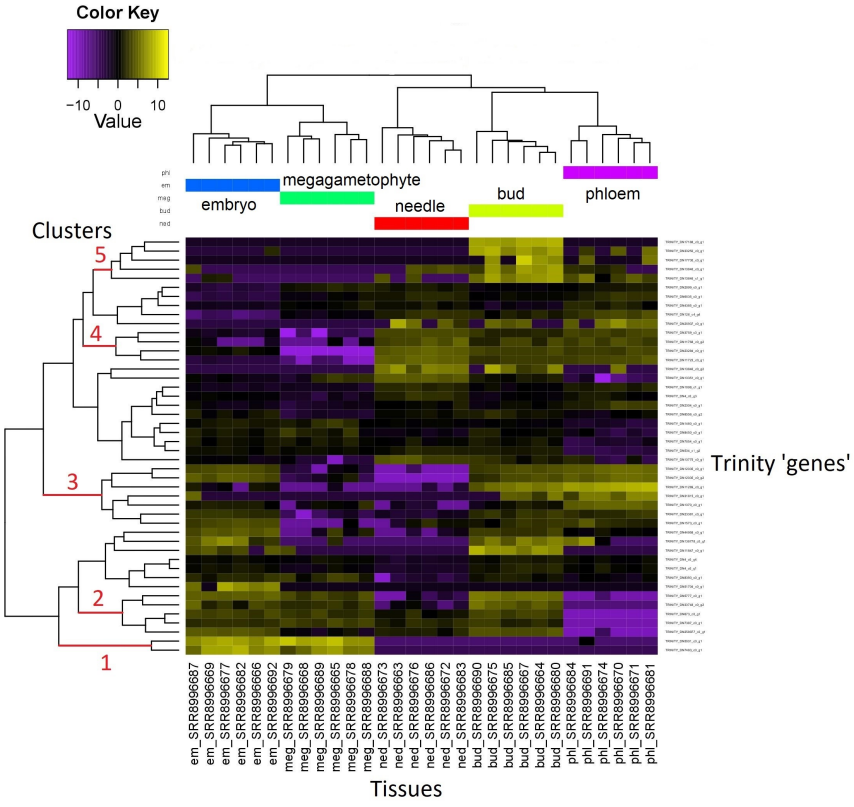

Figure 1: Heatmap of gene expression, five tissues of *Pinus sylvestris*

 ISBN: 978-989-98750-7-4

References

[1] Viola, Ivana L., and Daniel H. Gonzalez. (2016). *Structure and evolution of plant homeobox genes. Plant Transcription Factors*, Academic Press, 101-112. https://doi.org/10.1016/B978-0-12-800854-6.00006-3

[2] Alvarez, Jos M., et al. (2018). *Analysis of the WUSCHEL-RELATED HOMEOBOX gene family in Pinus pinaster: new insights into the gene family evolution*. Plant physiology and biochemistry, 123, 304-318. https://doi.org/10.1016/j.plaphy.2017.12.031

[3] De Heredia, Unai Lpez, and Jos Luis Vzquez-Poletti. (2016). *RNA-seq analysis in forest tree species: bioinformatic problems and solutions*. Tree Genetics & Genomes, 12(2), 30. https://doi.org/10.1007/s11295-016-0995-x

[4] Ojeda, Dario I., et al. (2019). *Utilization of tissue ploidy level variation in de novo transcriptome assembly of Pinus sylvestris*. G3: Genes, Genomes, Genetics, 9(10), 3409-3421. https://doi.org/10.1534/g3.119.400357

 ISBN: 978-989-98750-7-4