

POSSIBILITIES OF DE NOVO TRANSCRIPTOME SEQUENCING IN PYLOGENETIC RESEARCH ON AN EXAMPLE OF TARAXACUM OFFICINALE (ASTERACEAE)

M.G. Kutsev, M.V. Skaptsov, S.V. Smirnov, T.A. Sinitsyna, A.A. Kechaykin, M.S. Ivanova, A.I. Shmakov
Altai state university, 656049, Barnaul, Lenina prosp., 61
E-mail: m_kutsev@mail.ru, mr.skaptsov@mail.ru, serg_sm@mail.ru, t.sinitsyna@list.ru, alekseikechaikin@mail.ru,
ssbgbot@mail.ru

As a result of the research we obtained a cDNA library, sequences and carried out *de novo* assembly and functional annotation of the transcriptome of systematically complex, almost cosmopolitan *Taraxacum officinale*. Systematics of the species is complicated because of separation of many microspecies and intraspecific taxa, which may be explained by the presence of ploidy races, the phenomenon of apomixis and significant polymorphism. Attempts of separation of microspecies and closely related species based on sequencing of chloroplast and nuclear DNA fragments, did not give acceptable results. We have made the first attempt to analyze the transcriptome for understanding the genome evolution of agamospermous-sexual complex of the species. A total of 84440 reads were obtained with a total 31,540,710 bp. As a result of the *de novo* assembly, we obtained 13902 contigs, with an average GC content equal to 38.1% and a maximum length of 5255 bp. In total, we received 3798 annotated genes. According to the functional annotation based on sequence homology, 2687 contigs were attributed to biological processes (19.32%), 3299 – to molecular functions (23.7%), 2157 – to the cellular component (15.51%) and 7497 contigs – with unknown functions. In the first category “single-organism cellular process”, “response to stimulus”, “photosynthesis-light reaction”, “oxidation-reduction” and “translation” have dominated. In the category of molecular function “nucleic acid binding”, “hydrolase activity”, “transferase” and “oxidoreductase” activities have dominated. In the cellular component category “integral component of membrane”, “chloroplast thylakoid membrane”, “photosystems” and “nucleus” have dominated.

Keywords: *T. officinale*, transcriptome, RNA, cDNA, pyrosequencing.

Citation:

Kutsev, M.G., Skaptsov, M.V., Smirnov, S.V., Sinitsyna, T.A., Kechaykin, A.A., Ivanova, M.S., Shmakov, A.I. (2016). Possibilities of *de novo* transcriptome sequencing in pylogenetic research on an example of *Taraxacum officinale* (Asteraceae). *Biological Bulletin of Bogdan Chmelnytskyi Melitopol State Pedagogical University*, 6 (3), 319–323.

Поступило в редакцію / Submitted: 12.11.2016

Принято к публикации / Accepted: 02.12.2016

crossref <http://dx.doi.org/10.15421/2016101>

© Kutsev, Skaptsov, Smirnov, Sinitsyna, Kechaykin, Ivanova, Shmakov, 2016

Users are permitted to copy, use, distribute, transmit, and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship.



This work is licensed under a Creative Commons Attribution 3.0. License

INTRODUCTION

Transcriptome research opens up new possibilities in phylogenetic studies of evolutionary complex plant groups, monophyletic, hybridogenic, polyploid and apomictic. In the future, a comparative analysis of all expressed genome elements will allow to evaluate not only the taxonomic position, but also an evolution of supra-species taxa. At present a comparative analysis of molecular markers: non-coding sequences - ITS, ETS, or encoding mitochondrial, chloroplast or nuclear - rbcL, matK, trnL-F, rpl and the other, does not always give an idea of the systematic position or an appraisal of evolution course. Transcriptome analysis allows to find new informative markers, to estimate the evolution on basis of accumulation of mutations in paralogous and orthologous genes, or to build the evolution on basis of consensus phylogenetic trees using hundreds of genes obtained after the transcriptome sequencing. The research of expressing gene pool of ancient and modern plant groups, hybrids, polyploids and study of apomict evolution, their ways of variability and speciation are of particular interest.

Analysis of the activity of mobile genetic elements and epigenetic changes is an important mechanism of gene variability and regulation, and also relates to an adaptive divergence to the new conditions (Stapley et al., 2015).

Transposon activity, methylation levels, histone modifications and microRNA can generate heritable genetic variations in response to the environment. These mechanisms, perhaps, are the basis of an evolution of plant groups, which are characterized by asexual reproduction, apomixis, in which new species are formed despite the fact that all individuals in a population are clones. The activity of transposons can accumulate, because there is no recombination, thereby enhancing effect on genome (Arkhipova, Meselson, 2005).

Furthermore, due to the greater speed of accumulation of mutations, sequencing of mobile genetic elements allows to construct phylogenetic relationships in apomictic groups and groups of a hybrid origin, in which classical molecular markers can't detect systematic position with high probability, forming a "comb" in dendrograms. In these difficult plant groups an analysis of many genes allows to investigate single nucleotide polymorphism (SNP), small insertions / deletions (InDel) and determine accurately the mutated genes, the expression of which can affect the anatomical-morphological and physiological changes during the subsequent selection (Kuravadi et al., 2015; Li et al., 2016).

Comparative data obtained after the transcriptome sequencing can be used for searching new conservative markers – single-copy and low-copy genes. Previously for phylogenetic analysis ADH, TPI, GAP3DH, LEAFY, PGK, petD, GBSSI, GPAT, ncpGS, GIGANTEA, GPA1, AGB1, PPR and RBP2 were recommended as such genes. In the case of doubling of such genes they are rapidly eliminated in the genome during subsequent polyploidization. In cases of a presence of few gene copies in polyploids they are complementary and supplement each other (Duarte et al., 2010).

In some cases, when the nuclear markers cannot reliably separate the plant groups or reconstruct their evolution, for example, for Pteridophyta, chloroplast markers or whole genome sequencing of the chloroplast DNA are used. In such cases a direction of the nuclear DNA evolution and a set of hypotheses which are difficult to prove or disprove remain unclear. Comparative analysis of chloroplast and nuclear genome data allows us to reconstruct the complete picture of the evolution.

Thus, Grusz et al. (2016) showed a similar evolutionary rate of chloroplast and nuclear genomes, despite the hypothesis of differences in DNA-polymerases of nuclei and plastids. Another hypothesis of a greater accumulation of mutations in species with long gametophyte stage was partially confirmed for vittaroid ferns, which are characterized by a long vegetation of gametophyte. Thus, transcriptome sequencing opens up new possibilities in phylogenetic research and study of evolutionary processes.

MATERIALS AND METHODS

As an object of investigation we used *Taraxacum officinale*, growing in South-Siberian Botanical Garden. Fresh leaves were homogenized in extraction solution (4M guanidine thiocyanate, 10 mM EDTA, 50 mM HEPES, pH 4.5) and centrifuged. To the supernatant an equal volume of isopropanol was added and centrifuged to precipitate nucleic acids. The RNA was purified using lithium chloride (Barlow et al., 1963). Residual DNA amounts were removed by hydrolysis with DNase I. The quality of RNA was assessed using horizontal electrophoresis in 1.5% agarose gel (Fig. 1). Samples with RNA ratio 28S / 18S not less than 2 : 1 were selected.

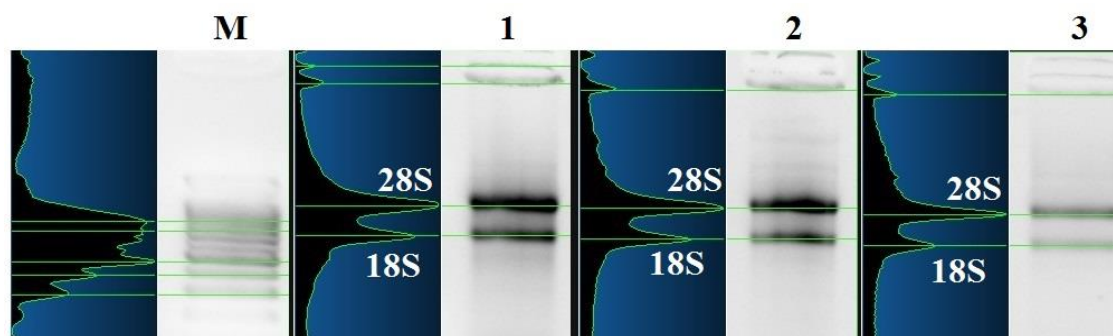


Figure 1. Examples of electrophoresis 28s/18S RNA of *T. officinale* for isolation quality assessment.

cDNA library was prepared using a set of GS FLX Titanium Rapid Library Preparation Kits (Roche 454, Branford, CT). The emulsion PCR and pyrosequencing were performed with Roche454 kits according to the manufacturer's instructions. The sequencing reaction was performed using a sequencer Roche 454 GS Junior. *De novo* assembly, normalization, searching errors and duplicated sequences were carried out using the software Geneious, Biomatters Limited. Searching homologous sequences by the BLAST algorithm and GO (gene ontology) analysis for functional annotation were performed using the software Blast2GO (Conesa et al., 2005).

RESULTS AND DISCUSSION

cDNA library has been derived after the mRNA enrichment by oligo-dT primer as a result of the reverse transcription of the *T. officinale* total RNA. Altogether 84440 reads have been obtained with 31,540,710 bp. The nucleotide sequences have been deposited in NCBI SRA № SRX2299371 database. The average length of reads

was 373 bp, with the peak at 511 bp (Fig. 2 a). After *de novo* assembly 13902 contigs have been obtained, with an average of GC content at 38.1%, the minimum length – 43 bp, the maximum – 5255 bp (Fig. 2b).

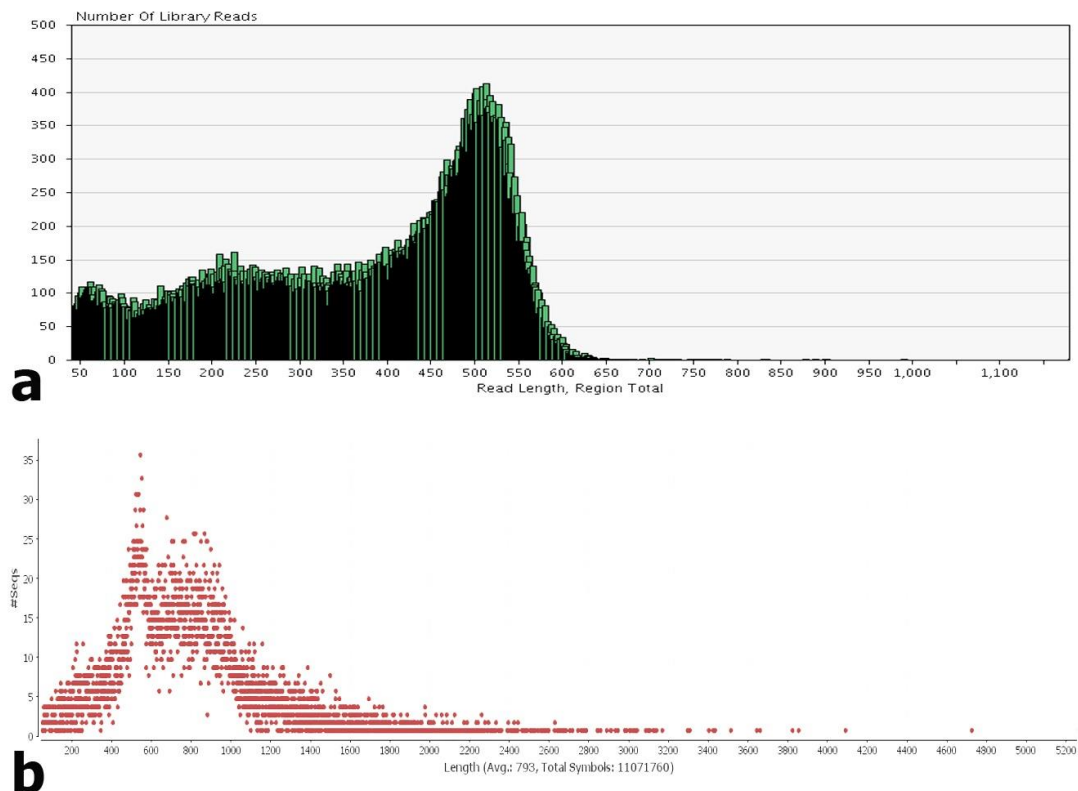


Figure 2. General data of sequencing and *de novo* assembly of the transcriptome. a. Graph of reads lengths; b. Graph of contig distribution.

We have used the public databases to annotate the transcriptome using BLAST algorithms (E value <math><1.0E-3</math>). We have got 16905 annotations in all. Maximal part of annotations has been received from UniProtDB database (99.8%), the remainder has been accounted for of the TAIR, GR Protein and PDB. GO (gene ontology) annotation is an international classification system for the standardization of gene functions, which includes three GO categories: biological processes, molecular functions and cellular components.

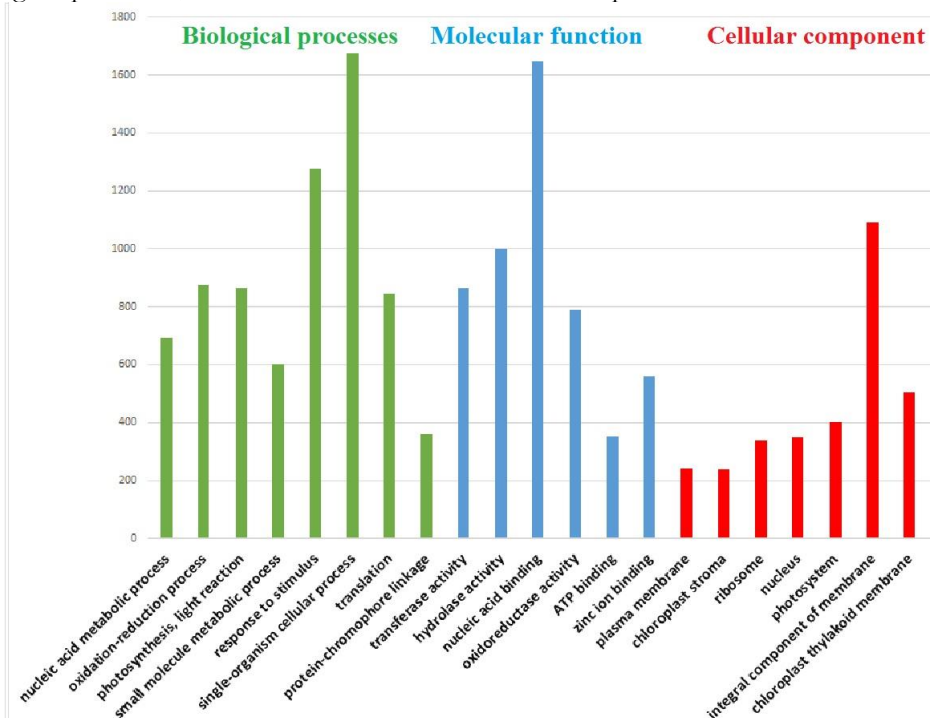


Figure 3. GO classification of the *T. officinale* transcriptome.

On the basis of the sequence homology from the mutual 13902 contigs 2687 contigs have been attributed to biological processes (19.32%), 3299 – to molecular function (23.7%), 2157 – to the cellular component (15.51%). In the first category “single-organism cellular process”, “response to stimulus”, “photosynthesis-light reaction”, “oxidation-reduction” and “translation” have dominated. In the category of molecular function “nucleic acid binding”, “hydrolase activity”, “transferase” and “oxidoreductase” activities have dominated. In the cellular component category “integral component of membrane”, “chloroplast thylakoid membrane”, “photosystems” and “nucleus” have dominated (Fig. 3). We have also obtained 7497 contigs with unknown functions.

We also have made a comparison of the results using BLAST by homology with plant sequences deposited in the NCBI database (Fig. 4). Most of the homologous sequences are referred to the following species: *Cynara cardunculus* (759), *Daucus carota* (636), *Cajanus cajan* (417), *Lactuca sativa* (359), *Taraxacum officinale* (344).

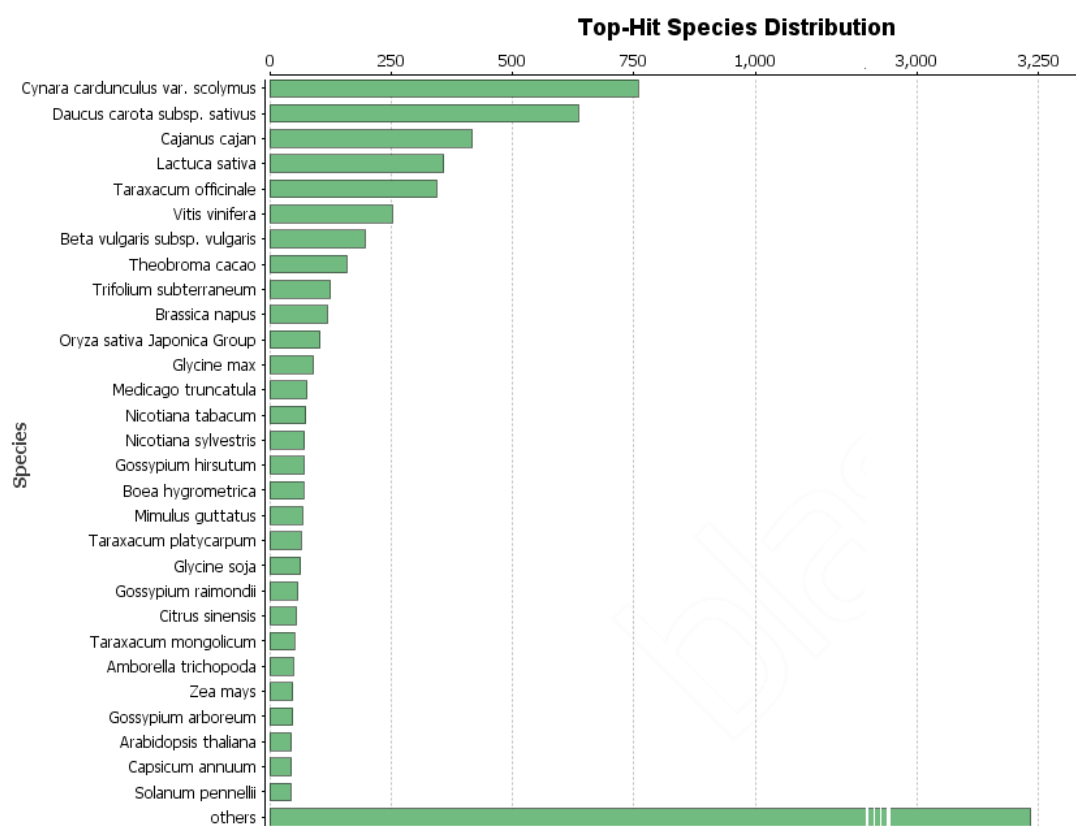


Figure 4. Species distribution by maximal homology of the sequences.

As a result of transcriptome study we may get thousands of sequences of coding genes. Among them, after a comparative analysis we can reveal conserved genes for use in phylogenetic researches. We have received 3798 annotated genes altogether, including more than 600 sequences of retrotransposons. The obtained data allow us to study difficult in terms of the evolution plant groups. Transcriptome analysis of clones from apomictic plant groups lets us to trace the path of variability in absence of reliable anatomical-morphological characters. Assessing the genetic divergence of the *T. officinalis* apomictic populations the transcriptome analysis has revealed that about one-third of inherited divergences have been caused by mobile genetic elements.

The transcriptome analysis has also disclosed differences in the metabolism mechanisms of acyl-lipid and abscisic acid, which may reflect functional differences within the apomictic lines (Ferreira de Carvalho et al., 2016).

Deng et al. (2015) have analyzed a transcriptome of ten orchid species and identified 315 single-copy orthologous genes for use to construct the phylogenetic relationships between the species. The phylogenetic trees have supported the topology on all nodes with almost 100% bootstrap and coordinated with previous phylogenetic studies of Orchidaceae.

Thus, the use of transcriptome for search new molecular markers opens up great opportunities for finding new characters for phylogeny and systematics purposes, as well as for a construction of living systems.

ACKNOWLEDGEMENT

This work was supported by the Russian Science Foundation, project No. 14-14-00472.

REFERENCES

- Arkhipova, I., Meselson, M. (2005). Deleterious transposable elements and the extinction of asexuals. *Bioessays*, 27, P. 76-85.
- Barlow, J.J., Mathias, A.P., Williamson, R., Gammack, D.B. (1963). A simple method for the quantitative isolation of undegraded high molecular weight ribonucleic acid. *Biochem. Biophys. Res. Commun.*, 13, 61-66.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* (Oxford, England), 21(18), 3674–3676.
- Deng, H., Zhang, G., Lin, M., Wang, Y., Liu, Z. (2015). Mining from transcriptomes: 315 single-copy orthologous genes concatenated for the phylogenetic analyses of *Orchidaceae*. *Ecol. Evol.*, 5(17), 3800–3807.
- Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, P.K., Leebens-Mack, J., dePamphilis, C.W. (2010). Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.*, 10, 61.
- Ferreira de Carvalho, J., Oplaat, C., Pappas, N., Derks, M., Ridder, D., Verhoeven, K.J.F. (2016). Heritable gene expression differences between apomictic clone members in *Taraxacum officinale*: Insights into early stages of evolutionary divergence in asexual plants. *BMC Genomics*, 17, 203.
- Grusz, A.L., Rothfels, C.J., Schuettpehl, E. (2016). Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns. *BMC Genomics*, 17, 692.
- Kuravadi, N.A., Yenagi, V., Rangiah, K., Mahesh, H., Rajamani, A., Shirke, M.D., Russiachand, H., Loganathan, R.M., Shankara, L.C., Siddappa, S., Ramamurthy, A., Sathyanarayana, B., Gowda, M. (2015). Comprehensive analyses of genomes, transcriptomes and metabolites of neem tree. *PeerJ*, 3, e 1066.
- Li, D., Zeng, R., Li, Y., Zhao, M., Chao, J., Li, Y., Wang, K., Zhu, L., Tian, W., Liang, G. (2016). Gene expression analysis and *SNP/InDel* discovery to investigate yield heterosis of two rubber tree F1 hybrids. *Sci. Rep.*, 6, 24984.
- Stapley, J., Santure, A.W., Dennis, S.R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol. Ecol.*, 24(9), 2241–2252.