UDC.358:577.212.3:577.218

doi: 10.15389/agrobiology.2015.3.278rus

doi: 10.15389/agrobiology.2015.3.278eng

«NEXT GENERATION SEQUENCING» FOR STUDYING TRANSCRIPTOME PROFILES OF TISSUES AND ORGANS OF GARDEN PEA (*Pisum sativum* L.)

(review)

V.A. ZHUKOV, O.A. KULAEVA, A.I. ZHERNAKOV, I.A. TIKHONOVICH

All-Russian Research Institute for Agricultural Microbiology, Federal Agency of Scientific Organizations, 3, sh. Podbel'skogo, St. Petersburg, 196608 Russia, e-mail zhukoff01@yahoo.com

Acknowledgements:

Authors are thankful to E.E. Andronov (ARRIAM, St. Petersburg) for his consultations and fruitful discussions on the issues of «next generation sequencing», to M.N. Povydysh (SPCPA, St. Petersburg) for his assistance in preparing this manuscript, to N.I. Ershov (ICG SB RAN, Novosibirsk) for his assistance in the analysis using bioinformatics methods, and to P. Winter and GenXPro GmbH employees (Frankfurt, Germany) for their assistance in RNA-sequencing.

Supported by Russian Science Foundation (grant № 14-24-00135)

Received February 2, 2015

Abstract

The term «Next Generation Sequencing» refers to modern technologies that help to obtain information about the nucleotide composition of tens and hundreds of millions of sequences in one experiment. NGS technologies are used to solve a wide range of problems (genome sequencing, gene expression assays, development of molecular markers, metagenomic studies of microbial communities, epigenetic studies etc.). One of the major applications of the NGS methods is concerned with analysis of gene expression by sequencing of transcriptome (the whole set of transcribed RNA). The review considers the approaches used for total gene expression analysis by «Next Generation Sequencing» — RNAseq (RNA sequencing) and its modification MACE (Massive Analysis of cDNA Ends). In this modification, developed by GenXPro GmbH (Frankfurt am Main, Germany), for each cDNA molecule only a 100-500 bp fragment (which is adjacent to the 3'-end of the transcript or, in another version, to its 5'-end) is subjected to sequencing; thus, the resolution of the method is increased by several times. In this way, MACE can capture the transcripts with low expression level, which correspond to the key regulatory genes forming the basis of biological processes. Also the review describes functional analysis of RNA sequencing, including the identification of biological patterns based on the detection of differentially expressed genes. An important step of this work is a hierarchical clustering of detected transcripts in accordance with the principles of gene ontology. The genes and gene products interact with each other to form a structured regulatory network, but the identification and analysis of regulatory networks is a complex task that requires the development of mathematical methods and the accumulation of data on gene expression, localization of gene products and their functional annotation. The review presents case studies of transcriptional profiles of the tissues and organs of pea (*Pisum sativum* L.), including those using the MACE technique. Thus, the use of NGS for gene expression studies is, at the moment, the optimal approach for studying the transcriptional profiles of any objects. The combination of NGS and potential of modern computational biology opens up new opportunities for studying the transcriptomes, including those of non-model species, that ensures progressive advance in many areas of biological science.

Keywords: plant genetics, «Next Generation Sequencing», RNA sequencing, gene expression, garden pea.

The term Next Generation Sequencing refers to modern technologies that help to obtain information about the nucleotide composition of tens and hundreds of millions of sequences in one experiment. NGS technologies include 454-sequencing (pyrosequencing) [1], the Illumina/Solexa method [2], the SOLiD (Sequencing by Oligonucleotide Ligation and Detection) method [3], and ion semiconductor sequencing (Ion Semiconductor Sequencing) [4]. Any of these technologies has its own advantages and disadvantages [5], differing in the length of single reads, their number, error frequency, workflow rate and the cost per 1 nucleotide. The variety of NGS methods makes it possible for the researchers to

choose the technology best suited to address specific scientific problems, on the one hand, and encourages the competition between manufacturers, contributing to the rapid evolutionary development of sequencing techniques, on the other hand.

Study of gene expression using NGS. NGS technologies are used to solve a wide range of problems (genome sequencing, gene expression assays, development of molecular markers, metagenomic studies of microbial communities, epigenetic studies etc.) [6, 7]. One of the major applications of the NGS method is related to the analysis of gene expression by sequencing of transcriptome (the whole set of transcribed RNA) [8, 9]. Currently, RNA-sequencing (RNA-seq) complements and is gradually replacing the method of gene expression microarray (microarray technology) [10, 11]. The reason for this is the advantages of RNA-seq. These include the low level of «background noise» and, as a consequence, higher sensitivity making possible to detect up to 90 % of the expressed genes [12, 13]. Moreover, RNA-seq enables the analysis of the expression of any gene, including the ones with unknown sequencing prior to the experiment (in contrast to microarrays constructed based on known sequences of transcripts), which is especially important for non-model objects with poorly studied genomes. Finally, RNA-seq enables the study of alternative splicing and allele-specific gene expression [14, 15]. The cost of the «next generation sequencing» methods is constantly decreasing, which is their additional advantage and makes them more attractive for use [16]. However, there are certain difficulties (explainable by the relative novelty of NGS), which are mainly associated with the processing and interpretation of large amounts of information from each experiment, which makes it necessary to equip research centers with powerful computers and involve bioinformatic experts.

Peculiarities of NGS use for the analysis of gene expression. The genome sequencer generates millions and billions of individual sequences (reads, from the English *read*). The ideology of gene expression analysis using RNA sequencing (so-called digital expression) is based on the assumption of the proportionality of the particular transcript (i.e., the number corresponding reads) and the level of the corresponding gene expression. Properly speaking, this is not absolutely true, as the sequencing efficiency depends on the complexity of «nucleotide context» (for example, the presence of homopolymers, repetitions, palindromes, AT- or GC-rich sites, etc.) [17, 18], but the error margin is low and it is usually neglected [19, 20].

Libraries for sequencing are usually prepared based on cDNA sections fragmented at random (for example, under the exposure to ultrasound). These cDNA fragments are ligated with adapter sequences and subjected to sequencing.

The resulting «raw» sequencing data require special processing consisting usually of four stages. At the first stage, poor quality and adapter sequence reads are deleted from the analysis. The second stage is read mapping to the reference genome or transcriptome, i.e. the finding of correspondence between the reads and their corresponding transcripts [21]. Mapping is the key process requiring special attention. It is complicated by the presence of gene splice variants, paralogous sequences, repetitions and allelic polymorphism. At the third stage, normalization of the number of reads mapped to each transcript relative to the total number of the resulting reads of the sample and their computation are performed. In this way, the RPKM (reads per kilobase per million of mapped reads) value is determined [22]. RPKM reflects the relative transcript molar concentration and is a measure of the expression of specific genes [23]. Finally, at the fourth stage of analysis, a statistical test is performed to identify the transcripts that show a statistically significant difference in level of expression under the comparison of the analyzed samples. All the above stages of analysis are performed using various software packages, both free (Bowtie2, edgeR) [24, 25] and

commercial ones (CLC Genomics Workbench, CLC bio, Denmark).

MACE modification. Analysis of gene expression using RNA-seq involves sequencing of all mRNA fragments isolated from the sample (Fig. 1). Despite the high performance of sequencers, at quantitative analysis, the number of reads corresponding to rare transcripts is low as well, which results in the inability to detect a statistically significant difference in the expression level of

transcripts. An approach known as MACE (Massive Analysis of cDNA Ends) developed by GenXPro GmbH (Frankfurt, Germany) became a new modification of RNA-seq method (Fig. 2). According to MACE, for each cDNA molecule, only a 100-500 bp fragment which is adjacent to the 3'-end of the transcript (in another version, to its 5'-end) is subjected to sequencing; thus, the resolution of the method is increased by several times, and the similar number of

Fig. 1. General scheme of RNA sequencing (RNA-seq). Based on the scheme from GenXPro GmbH web-site (Frankfurt am Main, Germany) (<http://www.genxpro.info>).

Fig. 2. General scheme of RNA sequencing using MACE (Massive Analysis of cDNA Ends): A — binding of cDNA derived from polyadenylated mRNA with streptavidin beads; B — cDNA ultra-sound fragmentation; C — primer ligation for sequencing to the point of fragmentation and sequencing with Illumina HiSeq2000 (Illumina, CIIIA); D — quantitative analysis of resulting sequencing fragments. Based on the scheme from GenXPro GmbH web-site (Frankfurt am Main, Germany) (<http://www.genxpro.info>).

reads results in more accurate information. In this way, MACE can capture the transcripts with low expression level (encoding the receptors or transcription factors, and antisense transcripts). It is often not possible to detect these transcripts by other means of massive analysis (using RNA-seq or microarray), even though they correspond to key regulatory genes, and therefore they should be subject to scrutiny in the investigation of the molecular mechanisms forming the basis for biological processes.

Annotation of transcripts (i.e. assigning a biological function based on the homology with the genes for which such a function is known) is a significant problem in the analysis of sequencing data using MACE. Usually, a transcripts portion sequenced using MACE corresponds to the 3'-untranslated region and thus it is of high variability. For the objects with poorly studied genome or transcriptome, annotation by comparing sequence databases (e.g., gene ontology) is impossible [26-28]; first, a reference transcriptome with full annotated transcripts should be created and then reads obtained by MACE should be mapped.

Functional analysis of RNA sequencing. Identification of differentially expressed genes has, as its direct continuation, the functional analysis of the products encoded by these genes. Detection of biological regularities within the results of RNA sequencing is performed using several approaches. Hierarchical clustering of the revealed transcripts according to the principles of gene ontology is an important step [26-28] which can detect groups of genes involved in specific cellular processes. A number of tools have been developed for this analysis, most of which are free, such as AgriGO [29], Blast2GO [30]. A similar approach is obtained by applying transcriptomic data to the known metabolic and signaling pathways. Using the MapMan [31] and Reactome [32] resources it is possible to study the location and role of the products coded for by so-called genes of interest in a variety of metabolic pathways.

The genes and gene products interact with each other to form a structured regulatory network. Analysis and construction of regulatory networks is an uncommon and complex task that requires the development of mathematical methods and the accumulation of data on gene expression, localization of gene products and their functional annotation resulted in the development of a number of resources that make gene interaction studies possible. To expand the capacity of such analysis in legumes, the LegumeGRN resource was created [33] with which gene networks based on available data both on the transcriptome of alfalfa, lotus, and soybeans [34-37] and own results, can be constructed. Using LegumeGRN, it is possible to detect the group of genes responding consistently to certain effects (co-expressed) which may indicate the involvement of these genes in the same process, and to establish the association of the expression of certain genes with transcription factors.

Studies of transcriptional profiles of the tissues and organs of pea (*Pisum sativum* L.). Pea, being one of the most important legumes in the world [38] is insufficiently studied related to the molecular genetic and genomic issues. Sequencing of pea genome is planned but has not been implemented yet (http://www.coolseasonfoodlegume.org/pea_genome). Therefore, it seems appropriate to study the structure of the transcriptome as the most active part of genome with very different composition in different tissues and organs. For example, the symbiotic genes controlling the development of nitrogen-fixing nodules and arbuscular mycorrhiza in legumes are expressed predominantly in the underground part of the plant (in roots and nodules) [39]. In some studies, RNA sequencing of organs and tissues of the aerial part of pea has been performed already [40-42], but the underground part (roots and nodules) is still neglected by researchers.

A group of authors collaborating with the Center of Biologically Active Compounds and Their Use (Moscow, Russia) have performed transcriptome sequencing in roots and nodules of pea line SGE using Illumina Genome Analyzer II X [43]. The «raw» data obtained (over 112 million 36 bp reads) were processed properly. After deleting of low-quality reads using the Trinity (<http://trinityrnaseq.sourceforge.net/>) program assembler [44], 50,703 contigs were assembled that currently constitute the most comprehensive reference transcriptome of pea roots and nodules. Some transcripts have been revealed to be represented by two or more contigs. By mapping of different sample reads on this transcriptome using the Bowtie2 program (<http://bowtie-bio.sourceforge.net/bowtie2/>) [24] and by statistical analysis using edgeR [25], 2,629 contigs corresponding to the genes with the level of expression significantly higher in the nodules than in roots were found, and 7,441 contigs were revealed corresponding to genes with the expression, on the contrary, specific for the roots compared to nodules [43].

Fig. 3. Graphical representation of the results of gene co-expression analysis in pea (*Pisum sativum* L.) when exposed to cadmium. Groups of genes associated with specific transcription factors are shown. Central node corresponds to a transcription factor with radial nodes as genes associated with it. The MACE (Massive Analysis of cDNA Ends) method was used.

The possibility of gene expression analysis in pea using MACE was also evaluated by the authors [45]. Transcriptome sequencing of the roots of pea exposed to toxic heavy metal cadmium was performed in cooperation with GenX-

Pro GmbH (Frankfurt, Germany). For the four samples analyzed, 37,216 contigs were obtained and annotated based on the comparison with the reference transcriptome of the underground part of the plant. Preliminary analysis revealed the difference in gene expression in response to cadmium in two lines pea that were contrast in the resistance to this heavy metal. Using the LegumeGRN resource, we found an association of certain transcription factors belonging to the GATA, bZIP, bHLH, and WRKY families with the genes in which expression is altered under the exposure to cadmium (Fig. 3) [45].

Thus, over the past few years, RNA sequencing took its rightful place among the modern methods of massive analysis of gene expression, competing with the microarray technology successfully. The cost of gene expression analysis using NGS is steadily decreasing but so far the microarray method in plants with well-studied genome such as *Arabidopsis thaliana* (L.) Heynh. and *Medicago truncatula* Gaertn. is cheaper as twice (not including the cost of microarray development and construction). For non-model species, in particular for pea with its poorly studied genome, total gene expression analysis using NGS is the most appropriate one since, in addition to the analysis of differential expression, the result of this work is the information about the organization of the transcriptome of the studied object.

There are examples of ESR-based creation of microarrays for the analysis of pea apical meristem [46, 47] and the examples of the use of microarrays created based on the sequences of genes *M. truncatula* for pea [48]. However, as discussed above, the resolution of the analysis using NGS, especially when using MACE, exceeds the resolution of microarray technology greatly.

Currently, single examples of using MACE in the study of gene expression in tomato [49], and in fly larvae [50] are known. The results of this study also suggest the possibility of successful gene expression analysis in pea using the described approach and seem to be of great value, given the importance of pea as a crop.

Summing up, we would like to note that due to the development of sequencing technologies, accumulation of information about the structure and organization of genomes and transcriptomes in a wide variety of organisms acquires a large scale. Modern sequencing equipment makes it possible to identify the millions and billions of nucleotide sequences, so it is particularly important not to lose the biological meaning of the research within the great amount of data.

Thus, the use of Next Generation Sequencing (NGS) for gene expression studies is, at the moment, the optimal approach for studying the transcriptional profiles of any objects. The combination of NGS and potential of modern computational biology opens up new opportunities for studying the transcriptomes, including those of non-model species, that ensures progressive advance in many areas of biological science.

REFERENCES

1. Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Research*, 2001, 11: 3-11 (doi: 10.1101/gr.150601).
2. Mardis E.R. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 2008, 9: 387-402 (doi: 10.1146/annurev.genom.9.081307.164359).
3. Pandey V., Nutter R.C., Prediger E. Applied biosystems SOLiD™ system: ligation-based sequencing. In: *Next-generation genome sequencing: towards personalized medicine*. M. Janitz (ed.). WileyVCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2008: 29-41 (ISBN: 978-3-527-32090-5).
4. Rusk N. Torrents of sequence. *Nat. Methods*, 2011, 8(1): 44 (doi: 10.1038/nmeth.f.330).
5. Metzker M.L. Sequencing technologies — the next generation. *Nat. Rev. Genet.*, 2010, 11(1): 31-46 (doi: 10.1038/nrg2626).

6. Shendure J., Ji H. Next-generation DNA sequencing. *Nat. Biotechnol.*, 2008, 26(10): 1135-1145 (doi: 10.1038/nbt1486).
7. Knief C. Analysis of plant microbe interactions in the era of next generation sequencing technologies. *Front. Plant Sci.*, 2014, 5: 216 (doi: 10.3389/fpls.2014.00216).
8. Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 2009, 10(1): 57-63 (doi: 10.1038/nrg2484).
9. Ozsolak F., Milos P.M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.*, 2011, 12(2): 87-98 (doi: 10.1038/nrg2934).
10. 't Hoen P.A., Ariyurek Y., Thygesen H.H., Vreugdenhil E., Vossen R.H., de Menezes R.X., Boer J.M., van Ommen G.J., den Dunnen J.T. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucl. Acids Res.*, 2008, 36(21): e141 (doi: 10.1093/nar/gkn705).
11. Marioni J.C., Mason C.E., Mane S.M., Stephens M., Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, 2008, 18(9): 1509-1517 (doi: 10.1101/gr.079558.108).
12. Wilhelm B.T., Marguerat S., Watt S., Schubert F., Wood V., Goodhead I., Penkett C.J., Rogers J., Bahler J. Dynamic repertoire of an eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 2008, 453(7199): 1239-1243 (doi: 10.1038/nature07002).
13. Wang E.T., Sandberg R., Luo S., Khrebtkova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 2008, 456(7221): 470-476 (doi: 10.1038/nature07509).
14. Wang X., Sun Q., McGrath S.D., Mardis E.R., Soloway P.D., Clark A.G. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS ONE*, 2008, 3(12): e3839 (doi: 10.1371/journal.pone.0003839).
15. Wahlstedt H., Daniel C., Enstero M., Ohman M. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome Res.*, 2009, 19(6): 978-986 (doi: 10.1101/gr.089409.108).
16. Mardis E.R. A decade's perspective on DNA sequencing technology. *Nature*, 2011, 470(7333): 198-203 (doi: 10.1038/nature09796).
17. Nakamura K., Oshima T., Morimoto T., Ikeda S., Yoshikawa H., Shiwa Y., Ishikawa S., Linak M.C., Hirai A., Takahashi H., Altaf-Ul-Amin M., Ogasawara N., Kanaya S. Sequence-specific error profile of Illumina sequencers. *Nucl. Acids Res.*, 2011, 39(13): e90 (doi: 10.1093/nar/gkr344).
18. Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A., Swerdlow H.P., Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 2012, 13: 341 (doi:10.1186/1471-2164-13-341).
19. Garg R., Patel R.K., Jhanwar S., Priya P., Bhattacharjee A., Yadav G., Bhatia S., Chattopadhyay D., Tyagi A.K., Jain M. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.*, 2011, 156(4): 1661-1678 (doi: 10.1104/pp.111.178616).
20. Jain M. Next-generation sequencing technologies for gene expression profiling in plants. *Brief. Funct. Genomics*, 2012, 11(1): 63-70 (doi: 10.1093/bfpg/elt038).
21. Trapnell C., Salzberg S.L. How to map billions of short reads onto genomes. *Nat. Biotechnol.*, 2009, 27(5): 455-457 (doi: 10.1038/nbt0509-455).
22. Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B. Mapping and quantifying mammalian transcriptomes by RNAseq. *Nat. Methods*, 2008, 5(7): 621-628 (doi: 10.1038/nmeth.1226).
23. Wagner G.P., Kin K., Lynch V.J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.*, 2012, 131(4): 281-285 (doi: 10.1007/s12064-012-0162-3).
24. Langmead B., Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.*, 2012, 9(4): 357-359 (doi: 10.1038/nmeth.1923).
25. Robinson M.D., McCarthy D.J., Smyth G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010, 26(1): 139-140 (doi: 10.1093/bioinformatics/btp616).
26. Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. Gene ontology: tool for the unification of biology. *Nat. Genet.*, 2000, 25(1): 25-29 (doi: 10.1038/75556).
27. Gene Ontology Consortium. The Gene Ontology in 2010: extensions and refinements. *Nucl. Acids Res.*, 2010, 38(Suppl. 1): D331- D335 (doi: 10.1093/nar/gkp1018).
28. Blake J.A. Ten quick tips for using the gene ontology. *PLoS Comput. Biol.*, 2013, 9(11): e1003343 (doi: 10.1371/journal.pcbi.1003343).
29. Du Z., Zhou X., Ling Y., Zhang Z., Su Z. agriGO: a GO analysis toolkit for

- the agricultural community. *Nucl. Acids Res.*, 2010, 38(Suppl. 2): W64-W70 (doi: 10.1093/nar/gkq310).
30. Conesa A., Götz S., García-Gymez J.M., Terol J., Talón M., Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 2005, 21(18): 3674-3676 (doi: 10.1093/bioinformatics/bti610).
 31. Usadel B., Nagel A., Thimm O., Redestig H., Blaesing O.E., Palacios-Rojas N., Selbig J., Hannemann J., Piques M.C., Steinhauser D., Scheible W.-R., Gibon Y., Morcuende R., Weicht D., Meyer S., Stitt M. Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol.*, 2005, 138(3): 1195-1204 (doi: 10.1104/pp.105.060459).
 32. Croft D., Mundo A.F., Haw R., Milacic M., Weiser J., Wu G., Caudy M., Garapati P., Gillespie M., Kamdar M.R., Jassal B., Jupe S., Matthews L., May B., Palatnik S., Rothfels K., Shamovsky V., Song H., Williams M., Birney E., Hermjakob H., Stein L., D'Eustachio P. The Reactome pathway knowledgebase. *Nucl. Acids Res.*, 2014, 42(D 1): D472-D477 (doi: 10.1093/nar/gkt1102).
 33. Wang M., Verdier J., Benedito V.A., Tang Y., Murray J.D., Ge Y., Becker J.D., Carvalho H., Rogers C., Udvardi M., He J. LegumeGRN: a gene regulatory network prediction server for functional and comparative studies. *PLoS ONE*, 2013, 8(7): e67434 (doi: 10.1371/journal.pone.0067434).
 34. He J., Benedito V.A., Wang M., Murray J.D., Zhao P.X., Tang Y., Udvardi M.K. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*, 2009, 10: 441 (doi: 10.1186/1471-2105-10-441).
 35. Libault M., Farmer A., Joshi T., Takahashi K., Langley R.J., Franklin L.D., He J., Xu D., May G., Stacey G. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.*, 2010, 63(1): 86-99 (doi: 10.1111/j.1365-3113.2010.04222.x).
 36. Severin A.J., Woody J.L., Bolon Y.T., Joseph B., Diers B.W., Farmer A.D., Muehlbauer G.J., Nelson R.T., Grant D., Specht J.E., Graham M.A., Cannon S.B., May G.D., Vance C.P., Shoemaker R.C. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.*, 2010, 10: 160 (doi: 10.1186/1471-2229-10-160).
 37. Verdier J., Torres-Jerez I., Wang M., Andriankaja A., Allen S.N., He J., Tang Y., Murray J.D., Udvardi M.K. Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *Plant J.*, 2013, 74(2): 351-362 (doi: 10.1111/tpj.12119).
 38. Food and agriculture organization corporate statistical database. FAOSTAT, 2014 (<http://faostat.fao.org>).
 39. Journet E.P., van Tuinen D., Gouzy J., Crespeau H., Carreau V., Farmer M.J., Niebel A., Schiex T., Jaillon O., Chatagnier O., Godiard L., Micheli F., Kahn D., Gianinazzi-Pearson V., Gamas P. Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. *Nucl. Acids Res.*, 2002, 30(24): 5579-5592 (doi: 10.1093/nar/gkf685).
 40. Franssen S.U., Shrestha R.P., Bräutigam A., Bornberg-Bauer E., Weber A.P.M. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics*, 2011, 12: 227 (doi: 10.1186/1471-2164-12-227).
 41. Kaur S., Pembleton L.W., Cogan N.O., Savin K.W., Leonforte T., Paull J., Matern M., Forster J.W. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics*, 2012, 13: 104 (doi: 10.1186/1471-2164-13-104).
 42. Duarte J., Rivière N., Baranger A., Aubert G., Burstin J., Cornet L., Lavaud C., Lejeune-Hénaut I., Martinant J.P., Pichon J.P., Pilet-Nayel M.L., Boutet G. Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics*, 2014, 15: 126 (doi: 10.1186/1471-2164-15-126).
 43. Zhukov V.A., Zhernakov A.I., Ershov N.I., Shtratnikova V.A., Pekov Yu.A., Malakho S.G., Borisov A.Yu., Tikhonovich I.A. *Tezisy dokladov VI s"ezda Vavilovskogo obshchestva genetikov i selektsionerov (VOGiS) i assotsirovannykh geneticheskikh simpoziumov* [Proc. VI Congress of Vavilov Society of Geneticists and Breeders and associated genetic Symposia]. Rostov-na-Donu, 2014: 72.
 44. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 2011, 29(7): 644-652 (doi: 10.1038/nbt.1883).
 45. Kulaeva O.A., Tsyganov V.E. *Tezisy dokladov VI s"ezda Vavilovskogo obshchestva genetikov i selektsionerov (VOGiS) i assotsirovannykh geneticheskikh simpoziumov* [Proc. VI

Congress of Vavilov Society of Geneticist and Breeders and associated genetic Symposia]. Rostov-na-Donu, 2014: 194.

46. Wong C.E., Bhalla P.L., Ottenhof H., Singh M.B. Transcriptional profiling of the pea shoot apical meristem reveals processes underlying its function and maintenance. *BMC Plant Biol.*, 2008, 8: 73 (doi: 10.1186/1471-2229-8-73).
47. Liang D., Wong C.E., Singh M.B., Beveridge C.A., Phipson B., Smyth G.K., Bhalla P.L. Molecular dissection of the pea shoot apical meristem. *J. Exp. Bot.*, 2009, 60(14): 4201-4213 (doi: 10.1093/jxb/erp254).
48. Fondevilla S., Küster H., Krajinski F., Cubero J.I., Rubiales D. Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology. *BMC Genomics*, 2011, 12: 28 (doi: 10.1186/1471-2164-12-28).
49. Fragkostefanakis S., Simm S., Paul P., Bublak D., Scharf K.D., Schleiff E. Chaperone network composition in *Solanum lycopersicum* explored by transcriptome profiling and microarray meta-analysis. *Plant Cell Environ.*, 2015, 38(4): 693-709 (doi: 10.1111/pce.12426).
50. Zajac B.K., Amendt J., Horres R., Verhoff M.A., Zehner R. De novo transcriptome analysis and highly sensitive digital gene expression profiling of *Calliphora vicina* (Diptera: Calliphoridae) pupae using MACE (Massive Analysis of cDNA Ends). *Forensic Sci. Int. Genet.*, 2015, 15: 137-146 (doi: 10.1016/j.fsigen.2014.11.013).