ANALYSIS OF ALTERNATIVE SPLICING EVENTS IN THE ROOT TIPS AND NODULES OF *PISUM SATIVUM* L.

© E.A. Zorin [1], O.A. Kulaeva [1], A.M. Afonin [1], V.A. Zhukov [1], I.A. Tikhonovich [1, 2]

[1] All-Russia Research Institute for Agricultural Microbiology, Saint Petersburg, Russia;
[2] Saint Petersburg State University, Saint Petersburg, Russia

❧ **Background.** Legumes establish symbioses with nitrogen-fixing bacteria from the Rhizobium group. In exchange for nutrients, bacteria provide fixed nitrogen needed to support plant growth. At the moment, information about the involvement of alternative splicing (AS) in the establishment and maintenance this symbiotic relationships is almost absent, but, as it is a powerful mechanism for the regulation of proteome diversity of the cell, it therefore may participate in cellular response to microsymbionts. **Materials and methods.** Alternative splicing was analyzed using the assembly of "supertranscripts" and alignment of the reads from nodules and root tips to this reference. Target genes expression levels was estimated in tips of non-inoculated roots, and in nodules (2, 4, and 6 weeks post inoculation) with use of RT-qPCR. **Results.** In this study, the analysis of AS events in the nodules and root tips of the pea was carried out. The presence of isoforms of four pea genes (*PsSIP1*, *PsIGN*, *PsWRKY40*, *PsPR-10*) was confirmed and their expression level was estimated. **Conclusion**. Pea nodules were shown to be more enriched with AS events compared to root tips. Among the functional groups of genes that demonstrate AS events, one of the most enriched functional groups is the pathogens stress response. Intron retention probably leads to degradation of the transcript via NMD-system or to change of the protein function, that modulates the activity of genes in nodules.

❧ **Keywords:** pea; symbiosis; alternative splicing; transcriptomics.

АНАЛИЗ СОБЫТИЙ АЛЬТЕРНАТИВНОГО СПЛАЙСИНГА В КОНЧИКАХ КОРНЕЙ И КЛУБЕНЬКАХ *PISUM SATIVUM* L.

© Е.А. Зорин [1], О.А. Кулаева [1], А.М. Афонин [1], В.А. Жуков [1], И.А. Тихонович [1, 2]

[1] ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Санкт-Петербург;
[2] ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург

❧ Горох посевной (*Pisum sativum* L.) способен вступать в симбиотические отношения с азотфиксирующими бактериями из группы ризобий (*Rhizobia*). В обмен на питательные вещества бактерии поставляют фиксированный азот, необходимый для поддержания роста растения. Фиксация азота происходит в специализированных органах — клубеньках. Формирование и функционирование клубенька — сложноорганизованный процесс, требующий тонкой регуляции работы генов. Один из ее механизмов — альтернативный сплайсинг (АС). На данный момент информация о роли АС в установлении и поддержании симбиотических отношений между азотфиксирующими бактериями и бобовыми растениями практически отсутствует, но известно, что это мощный способ регуляции разнообразия протеома клетки, а потому АС наблюдается в реакциях на присутствие микросимбионта. В рамках работы был проведен анализ событий АС в клубеньках и кончиках корней гороха посевного. Показано, что среди функциональных групп генов, демонстрирующих события АС, одной из наиболее представленных является группа, принимающая участие в стрессовых реакциях растения на патогены. Вероятно, транскрипты с удержанным интроном подвергаются воздействию NMD-системы либо меняют свою функцию вследствие изменения первичной и вторичной структуры. Также выявлена *in silico*, экспериментально подтверждена и оценена количественно с применением метода ПЦР в реальном времени представленность мРНК изоформ четырех генов гороха: *PsIGN1*, *PsSIP1*, *PsPR-10*, *PsWRKY40*.

❧ **Ключевые слова:** горох посевной; симбиоз; альтернативный сплайсинг; транскриптомика.

INTRODUCTION

Legumes, including the garden pea (*Pisum sativum* L.), are capable of forming symbioses with an extremely wide range of beneficial soil microorganisms, representing a key example of mutualistic plant—microbial interactions. One of the most ecologically impor-

tant and well-studied of these systems is the symbiosis of leguminous plants with nitrogen-fixing bacteria. This type of interaction is beneficial to both the plant and the bacteria, since the plant receives atmospheric nitrogen, which bacteria can fix, and the bacteria receive nutrients and other metabolites. Nitrogen fixation occurs within special organs called nodules [1]. Currently, owing to the study of mutants of model leguminous plants, the initial stage of nodulation has been best studied, and the use of the genome synteny phenomenon for cloning homologous genes in the garden pea has led to the identification of pea gene sequences responsible for signal perception and specificity of plant interaction and nodule bacteria strains [2]. However, the number of genetic determinants involved in nodulation still exceeds the number of known symbiotic genes of leguminous plants; in addition, very little is known about the role of various regulatory mechanisms in processes of formation and/or functioning of nitrogen-fixing nodules.

One mechanism for the regulation of protein diversity in eukaryotic cells is alternative splicing (AS) [3]. This mechanism is involved in plant responses to biotic and abiotic stress, during which modulation of the quantitative ratio of transcripts in the cell and regulation of gene expression occur [4−6]. AS has been studied far less in plants than in other organisms (e. g., in mammals), and this is especially true for studies involving "non-model" organisms. However, it has been revealed that AS plays a key role in the regulation of the proteome of a plant cell under various conditions, including stress [6−10]. Depending on the mechanism and the final exon−intron structure of the transcript, several types of splicing have been distinguished. The first four are described most completely in the literature: 1) exon skipping, when a whole exon is skipped (that is, it is not included in the mature mRNA, in comparison with the primary transcript containing the exon); 2) intron retention (IR), when an intron is not spliced and remains part of the mature mRNA; 3) alternative donor site, in which a donor site, also known as a 5'-site of splicing, differs from that of mature mRNA; and 4) alternative acceptor, in which a 3'-site of splicing is different in the isoform. On the basis of the structure of the primary transcript, three additional types of AS events are distinguished: 1) alternative splicing sites, when both acceptor and donor sites are changed; 2) a new intron, when a splicing site arises in a previously known exon; and 3) a retained exon (RE), when a new exon replaces a previously annotated exon in mature mRNA [11−13].

The plant's interactions with bacteria (although non-pathogenic, in the case of nodule symbiosis) may have traits in common with stress reactions, so it is highly possible that AS events can occur in the tissues of nitrogen-fixing nodules. Thus, for example, in our previous study, during analysis of the transcriptome of garden pea nodules, detected AS of the gene homologous to the *Lotus japonicus IGN1* gene (ign − ineffective greenish nodules 1) [14]. This gene encodes the protein containing transmembrane domains and ankyrin repeats. In *L. japonicus*, *IGN1* gene expression is not limited to symbiotic organs (transcripts are found in nodules, roots, leaves, and flowers), but in *IGN1* gene mutants, only the functioning of the nodules is impaired, and pale green nodules are formed, in which nitrogen is not fixed [15]. Nodules demonstrate signs of premature aging, starting from the stage when nitrogen fixation begins in normal nodules. The probable function of the IGN1 protein is anchoring other proteins on the cell membrane, and it may be related to the control of plant defense reactions; in the absence of the IGN1 protein, the activity of plant defense systems leads to the death of bacteria and the degradation of symbiotic structures [15]. The longer isoform of the garden pea *IGN1* gene contains the first intron (type "intron retention") and, as a result, can be translated into a protein that does not contain the initial 21 amino acids from the N-terminus [14]. However, since the protein is probably translated from an alternative start codon, it may have a different signal sequence at the N-terminus and a different localization in the cell, and it can be transported to the compartments associated with symbiosis.

Another case of AS participation in nodulation was found by Chao Wong et al. [9], who demonstrated that the mRNA of the *LjSIP1* gene is subject to AS in nodules, thus providing two structurally different isoforms, short (SIP1S) and long (SIP1L). SymRK-interacting protein 1 (SIP1) can interact with the receptor kinase, SymRK, in *L. japonicus* and probably participates in a common symbiotic signaling pathway. The longer version (SIP1L) contains 17 additional amino acids that form a complete heat shock protein 20 (Hsp20)-like domain. It is noteworthy that in *L. japonicus*, the final product of this isoform cannot interact with the receptor kinase. It is important that the short (SIP1S) form was not found in other higher plants, and the long isoform (SIP1L) can interact with the orthologs of the SymRK receptor kinase. Therefore, most likely, the loss of the ability of SIP1L to interact with SymRK in *L. japonicus* is compensated by increased expression of the shorter version (SIP1S) which binds SymRK and plays a role in symbiotic signaling [10].

Thus, in various tissues and organs of plants, including nodules of various legumes, AS events can

be detected. Some genes that have specific expression patterns in these organs may be involved in the establishment of symbiotic relationships between the plant and nitrogen-fixing soil bacteria. This fact constitutes grounds for further research. However, the situation is complicated by the fact that, for the most effective analysis of AS events, the assembled genome of the organism is required as the object of research. The pea genome has not yet been sequenced; however, we have at our disposal data from experiments on sequencing the transcriptome of various tissues of this plant. It is worth noting that work with this kind of data toward this aim requires special methods of computational analysis. This paper describes AS events in nodules and root tips of the garden pea using the "supertranscriptome" method of assembling from the available transcriptome data of nodules and root tips with the subsequent analysis of mapping patterns of reads for the assembled "supertranscripts." The use of reverse transcription PCR and quantitative real-time PCR confirmed the presence of AS and quantified the representation of the mRNA of various isoforms in pea root tips and nodules at three time points after inoculation for four genes: *PsSIP1*, *PsIGN1*, *PsWRKY40*, and *PsPR-10*.

MATERIALS AND METHODS
Bioinformatical methods

The results of RNA-seq experiments were used as a material for computer-aided data analysis, such as assembly and raw reads of the nodules of the SGE line, assembly and raw reads of the root tips of the SGE line [14], assembly of the transcriptome of different pea tissues of the Cameor line [16], and assembly of the transcriptome of various pea tissues of the JI2822, Terese, and JI992 lines (GEUU01.1) [17], as well as Kaspa and Parafield [18]. To achieve this, we used an algorithm recommended by the authors of the Lace tool (available at https://github.com/Oshlack), with minor modifications (Fig. 1). The quality of raw reads was controlled by FastQC [19]. Tails were trimmed, and low-quality reads were removed using the Trimmomatic [20]. Decontamination was performed using the BBtools software package [21]. The reads that passed quality control were aligned to the transcriptome assembly using the STAR program [22], since this algorithm enables detection of the splice sites.

For a more comprehensive and complete analysis of AS, "supertranscripts" should be constructed from the transcriptomes of different tissues. For this task, the assemblies of nodules and root tips were combined. After combining the two assemblies in order to create a set of contigs with more complete exon−intron composition, assembly complexity was reduced using the CD-HIT program [23], by clustering similar transcripts and removing identical ones. Thus, 63,395 of 95,459 contigs remained in the combined data set.

A necessary step in the construction of supertranscripts is the clustering of the initial contigs among themselves. The first stage of clustering is alignment of reads to the assembled transcripts. The STAR program in two-pass mode was used for more accurate alignment and searching for splice sites. The alignments obtained were processed by the Corset utility [24], the result of which were contigs grouped by alignment pattern. In the Lace program [25],
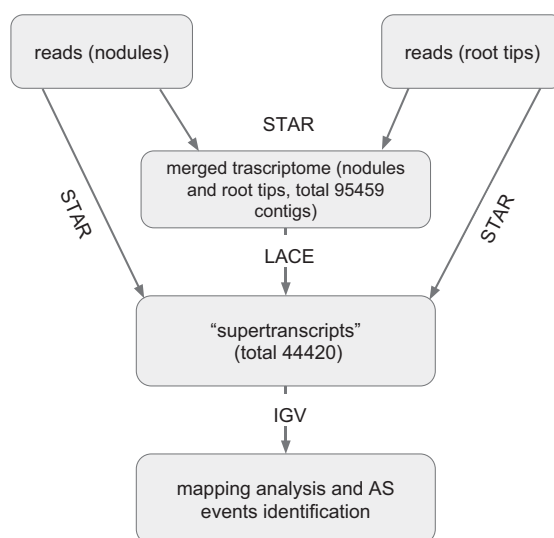

Fig. 1.   Scheme of assembly of "supertranscripts" with the subsequent analysis of alternative splicing events

a "supertranscript" assembly containing "supertranscripts" combining data from nodule and root tissue was constructed from contigs and a file with data on how contigs are clustered among themselves. Alignment visualization and sequence analysis were performed in the Integrative Genomics Viewer genomic browser [26]. Additional validation of the presence of specific isoforms was performed by local alignment of transcript sequences using Blast for transcriptome assemblies of pea lines Kaspa, Parafield, Cameor, JI2822, Terese, and JI992.

Additional functional analysis of sequences demonstrating AS events included an analysis of the open reading frames and the effect exerted by the AS on the organization of the coding sequence. Prediction of open reading frames and translation into the amino acid sequence were performed using the Augustus [29] and ExPASy tools [30]. Protein sequences were aligned using the Smith−Waterman algorithm implemented in BLAST for the UniProt database [31] and using the hidden Markov model algorithm for the Pfam database [32] to predict the potential function of the protein according to homology. Functional analysis of sequences demonstrating AS was performed in the REVIGO program [27], and the annotation of supertranscripts in the Gene Ontology categories was performed using the EggNog database [28].

Biological methods

A line of garden pea (*Pisum sativum* L.) SGE was used as an object [34]. To obtain the biological material from the root tips, the seeds were sterilized in concentrated sulfuric acid for 20 min, followed by washing with distilled water five times. The plants were grown in a Heraeus Voetsch phytotron (Germany) with a 16 h/8 h day/night cycle, at a temperature of 21 °C, with relative humidity of 75%.

To isolate RNA, the tissue of the root tips (5 mm cut from each lateral root) was homogenized using liquid nitrogen. RNA was isolated according to the NucleoSpin miRNA protocol (MachereyNagel, Germany). RNA quality was assessed using gel electrophoresis in a 1.5% agarose gel. The concentration was evaluated on a Shimadzu UV mini1240 spectrophotometer (Shimadzu, Japan). 1.5 μg of RNA was used for the reverse transcription reaction. cDNA was synthesized using ThermoScientific RevertAid First Strand cDNA Syntesis kit (Thermo Fisher Scientific, USA) in a T 100 Thermal Cycler amplifier (Bio-Rad Laboratories, USA) according to the manufacturer's protocol. Samples for the presence of DNA were assessed by PCR with primers to the promoter of the translation initiation factor 4E (number DQ641471 in the NCBI database).

The procedures for seed sterilization, plant inoculation, growing conditions, RNA isolation, and cDNA synthesis of nodules were performed according to methods described previously [35]. Materials for analysis were collected at 2, 4, and 6 weeks after inoculation with *Rhizobium leguminosarum* bv. *viciae* Rlv 3841 strain. The experiment was performed in three biological replicates.

The mRNA representation of the detected genes was estimated using real-time quantitative PCR. The primers for the real-time PCR procedure were selected so that the sequence of the forward primer fell on the exon site upstream the 5'-splicing site for approximately 100 base pairs, and the sequence of the reverse primer fell on the junction of two exons. This selection method enabled estimation of the representation of a transcript subjected to constitutive splicing, that is, without demonstrating the AS pattern, since it was expected that there was no intron in the product (Fig. 2). To estimate the amount of the long isoform (with an intron), prim-
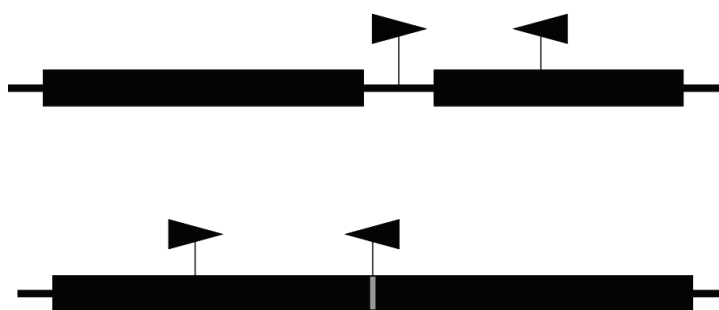

Fig. 2. Layout of primers for quantitative assessment of isoforms. The upper figure is an isoform diagram with an intron (broad blocks are exons, narrow is an intron). The bottom figure is an intron-free isoform diagram (black blocks are exons, gray is the border of exons)

ers were also used, which were developed according to the scheme shown in Fig. 2: the forward primer falls on the middle of the intron, and the reverse one falls on the exon site located downstream the 3'-splicing site.

Statistical data processing

To compare gene expression between samples, One-way Analysis Of Variance was used, followed by the use of the Tukey test. This enabled the identification of which groups differed significantly from each other and to correct for multiple comparisons [33]. Data analysis and visualization was performed in the R programming environment (version 3.4.3).

RESULTS AND DISCUSSION

In total, the supertranscriptome assembly contained 44,420 "supertranscripts" (see Fig. 1). Each "supertranscript" is a set of exons and introns from all transcripts synthesized from one assumed gene from different tissues of one organism. The Augustus utility predicted 40,884 coding sequences (CDS) out of 44,420 "supertranscripts". Repeated alignment of root tips and nodule reads on the "supertranscriptome"

enabled the detection and visualization of AS events. In total, we found 311,477 splicing sites for the nodule tissue. 278,265 sites were represented by GT/AG, 4355 by GC/AG, and 1285 by AT/AC. 27,572 sites were represented by non-canonical dinucleotide sequences (Fig. 3). By analyzing the assembly of the root tips, 47,917 splicing sites were detected, including 40,583 for GT/AG, 912 for GC/AG, 216 for AT/AC, and 6206 non-canonical ones.

An analysis of the exon–intron structure of nodule transcripts using Augustus showed that at least 10,233 transcripts contained one or more introns in their composition. The total number of IR events was 16,149. The total length of all introns was 2,198,563, and the average length was 136 base pairs. The total length of the nucleotide pairs of all introns in the assembly of the root tips was only 928,087 (2.3 times less), which is most likely associated not only with a smaller variety of AS events but also with a low sequencing depth. The average length of an intron was 113, the total number of introns was 8224, and the number of transcripts containing at least one intron was 5977 (Table 1). The alignment analysis revealed
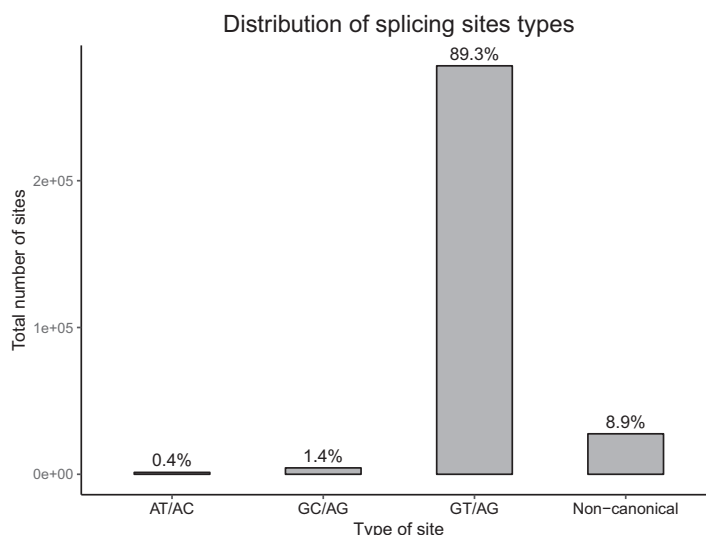

Fig. 3.   Distribution of different types of splice sites in the pea nodule transcriptome

Table 1

Quantitative description of the alternative splicing events in nodules and root tips assemblies

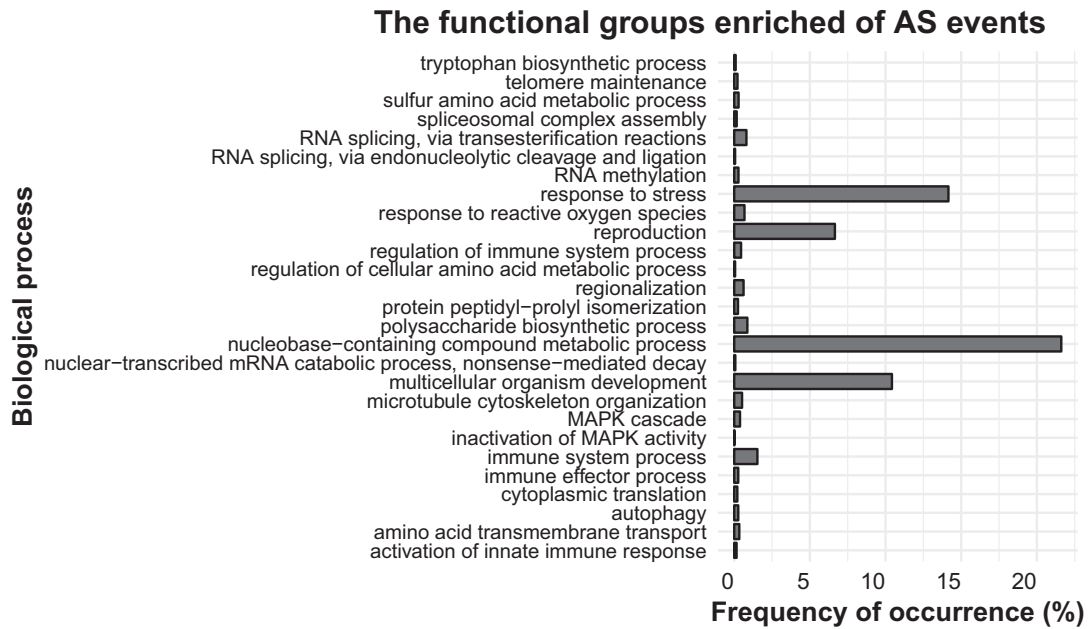| Plant organ	Number of transcripts with >1 intron (the percentage from total number of contigs)	Total number of unspliced introns	Total length of introns (base pairs)	Mean length of intron (base pairs)
Nodules	10 233 (17,5%)	16 119	2 198 563	136
Root tips	5 977 (16%)	8 224	928 087	113

The functional groups enriched of AS events

Fig. 4. Enrichment of various functional groups with alternative splicing events. Frequency is the percentage of total transcripts showing alternative splicing falling into this functional group

5486 transcripts demonstrating AS events. Annotation of the transcripts detected was performed through the EggNog database. 812 of 5,486 unique transcripts were annotated.

Functional analysis of the sequences (in terms of Gene Ontology) with AS showed that the group of genes associated with nucleotide metabolism was enriched by AS events. AS within this group is likely to affect replication and transcription processes, in that it has a general effect on nodule functioning (Fig. 4). The second largest event-rich AS group includes genes responsible for stress response. Further analysis of the AS events was conducted within this cluster. For experimental confirmation and quantitative assessment of mRNA representation, four genes were selected, namely, *PsSIP1*, *PsPR-10*, *PsWRKY40*, and *PsIGN1*.

As a result of the analysis by quantitative real-time PCR, it was found that, at weeks 2 and 4 after inoculation, the expression of both isoforms of the *SIP1* gene was slightly increased, which may indicate the importance of both long and short isoforms and also suggests that in pea, as in *L. japonicus*, the short isoform is a participant in symbiotic signaling and compensates for the inability of the long variant to interact with SymRK (Fig. 5), but these differences were not statistically significant. However, since the expression of the long isoform was also induced in nodules at weeks 2, 4, and 6 after inoculation, it can be assumed that it is capable of performing some other functions important for nodule functioning. The expression pattern of the

pea gene *SIP1* is consistent with the results obtained by Chao Wong et al. in a study performed on *L. japonicus* [9]. For a more detailed description of the *SIP1* gene function, further research is required.

Analysis of transcriptome data, alignment of the mRNA sequences of the *PsIGN1* gene to the database of the Transcriptome Shotgun Assembly (at NCBI website), suggested that the pre-mRNA of this gene is indeed subject to AS. As mentioned above, this gene encodes a protein-containing transmembrane domains and ankyrin repeats. The gene has two isoforms: a longer one, containing the first unexcised intron with a premature stop codon, and a short one, without an intron in this position. Apparently, the long form (with intron) is present in nodules, but is absent in other tissues, as indicated by the analysis of transcriptomic data from the SGE line. According to experimental data, both isoforms are presented in both root tips and nodules, since as a result of reverse transcription PCR and analysis of amplification products using gel electrophoresis, products of expected size were found, amplified both from the primers matched to the intron and from the primers whose position excluded the intron amplification. Moreover, the quantitative analysis of mRNA representation of both isoforms by real-time PCR showed that their ratio in nodules of different ages was at the same level and did not significantly change, indicating the importance of both isoforms (see Fig. 5). Probably, IR in the case of a long isoform leads to the production of the protein from an alternative start co-
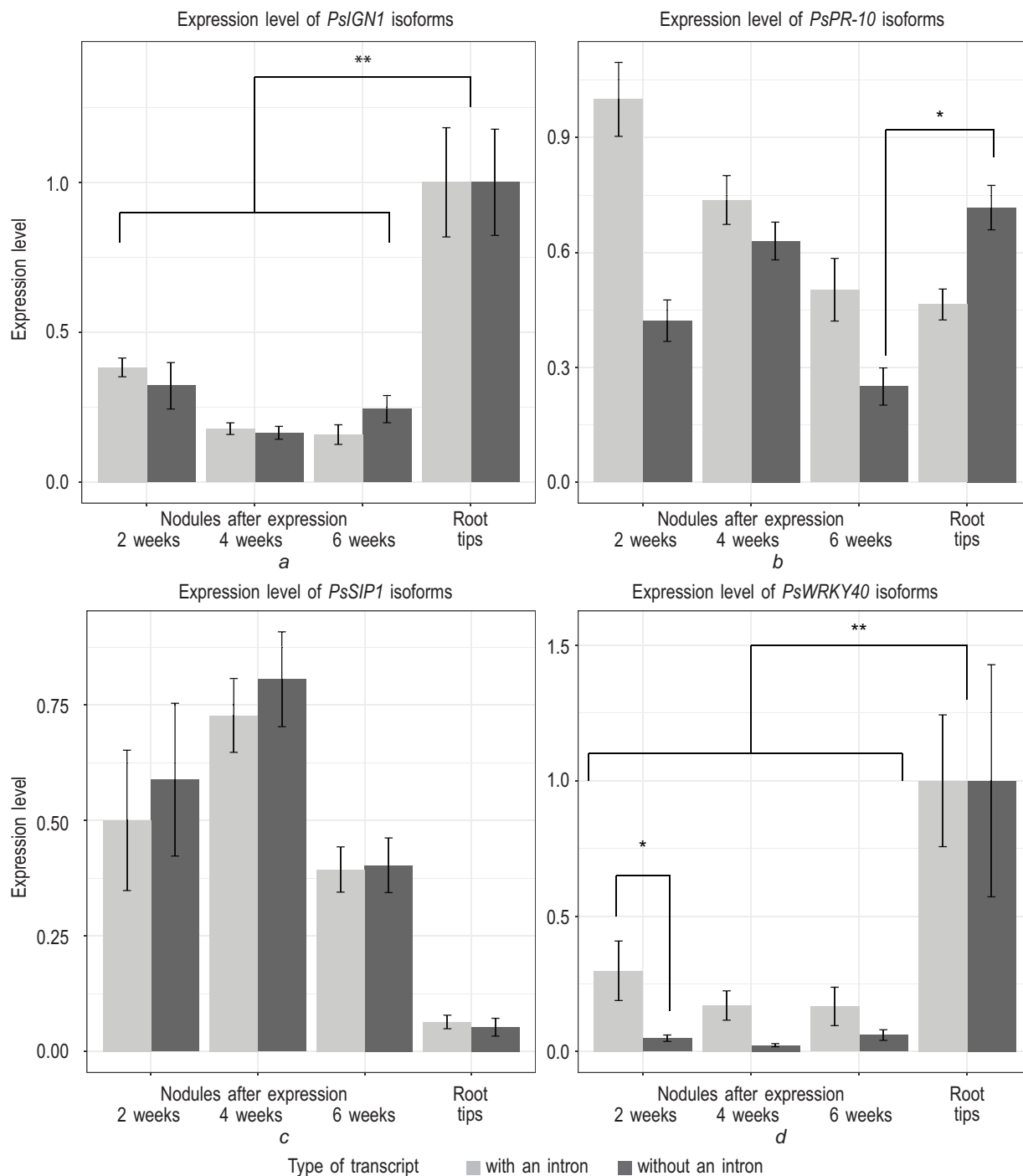
Fig. 5. Comparison of expression of the pea gene isoforms *PsPR-10*, *PsWRKY40*, *PsIGN*, *PsSIP1*. In fig. *a* the lines are connected to the comparison group, demonstrating statistically significant differences (roots relative to nodules in terms of 2, 4, 6 weeks), in fig. *d* significant differences are shown for isoforms in roots relative to nodules at 2, 4, 6 weeks and separately isoforms with intron and without in nodules for 2 weeks (* $p < 0,05$, ** $p < 0,0001$)

don located upstream the retained intron, as a result of which the final protein product acquires a signal sequence, different in its structure, at the N-terminus and, as a result, a different localization in the cell. It is noteworthy that the isoform with retained intron seems to be a characteristic only of pea, since the alignment

of this isoform to *M. truncatula*, *L. japonicus*, and *Glycine max* transcriptomes showed its absence, whereas this isoform is present in transcriptome assemblies of SGE, Kaspa, and Parafield pea lines.

A gene from the WRKY family was also identified among a group of genes exhibiting AS events. WRKY is

a large family of transcription factors with the WRKY-domain. The promoters of many genes of defense systems are enriched with W-box-elements, which serve as target sites of the transcription factors of the WRKY family [36]. Alignment results using BLAST showed that the closest homologues of this gene are *MtWRKY40* (GenBank accession XP_003604245) in *M. truncatula* and *TpWRKY56* (GenBank accession PNY02664.1) in clover, since the percentages of identity of the amino acid sequences of the pea gene product and these homologues are 87% and 88%, respectively. It is known that *WRKY40* positively regulates plant defense reactions by inducing the expression of a number of defense genes, but only when coexpressed with two other family members, *WRKY18* and *WRKY60*, possibly through interaction or formation of heterodimers [38].

Two isoforms of the *PsWRKY40* gene were found in the transcriptome of the garden pea, namely, a long isoform with an unexcised intron located at positions 667 to 825 and containing a premature stop codon in its sequence and a short one from which the intron was excised during splicing. As a result, a full-length protein with a functional WRKY-domain is translated from the short isoform, which retains the ability to bind with target sites and induce the expression of defense genes.

Quantitative analysis revealed that expression of both isoforms of the *PsWRKY40* gene in root tips was at the same level, and in nodules, a sharp decrease in both isoforms was observed at weeks 2, 4, and 6. At week 2, the number of isoforms with retained intron was increased relative to the isoform without intron at a statistically significant level ($p = 0.0166$; see Fig. 5). These data may indicate that, for full functioning of the nodules, the plant needs to suppress the expression of transcription factors that positively regulate the work of defense genes. In addition to lowering the general level of gene expression, one of the suppression methods is an increase in the number of isoforms with retained intron with a premature stop codon by means of AS, which we observed from the real-time PCR results. Transcripts with premature stop codon may become targets of the Nonsense-Mediated Decay (NMD) system. The system identifies aberrant mRNAs that carry premature stop codons in the intron and selects them for further degradation, thereby preventing the accumulation of potentially harmful and dangerous truncated proteins [37, 39]. Such a mechanism is also a simpler and faster (in an evolutionary sense) method of regulating the expression of a gene whose activity must be suppressed than, for example, changing the sequence of a promoter.

Another gene demonstrating AS is *PR-10*. *PR-10* proteins (pathogenesis-related proteins) represent a family of small monomeric proteins characterized by a conformation with an internal hydrophobic pocket formed by seven antiparallel beta-sheets and C-terminal alpha helix. It has recently been established that some proteins from the PR-10 family are involved in nodulation processes [40]. Among other things, proteins of this family have affinity for a wide range of ligands such as cytokinins, gibberellins, fatty acids, and flavonoids. Thus, PR-10 proteins can be involved in hormone-mediated resistance to pathogens or regulation of the development of nodules, but since a specific PR-10 protein can carry in its structure sites for binding different ligands simultaneously, its functions may vary depending on the tissue and the conditions when it is expressed [41]. In the transcriptome of garden pea nodules, two isoforms of the *PsPR-10* gene of *M. truncatula* homologue *MtPhBP* were found (GenBank accession number XP_003600177.1): a long isoform containing an additional intron in positions 241 to 541 with a premature stop codon and a short isoform subject to constitutive splicing and therefore deprived of intron. Thus, the intron-free form is active, and the transcript with the intron is translated into a truncated protein or is degraded by the NMD system even before the translation stage.

A quantitative analysis of the *PsPR-10* gene expression showed that, at weeks 2 and 6, there was a decrease in expression of the functional variant without an intron in nodules, and at weeks 2 and 4, there was an increase in expression of the nonfunctional isoform with the intron in the nodules relative to the root tips (but this difference was not statistically significant; see Fig. 5). This probably indicates that the functional product of this gene may be involved in the stress reactions of the plant in response to infection, and therefore, the functional product is subject to negative regulation. The expression of this isoform may be at a high level during the first 3−4 days after inoculation and decreases at later periods, as in *M. truncatula* (see the expression atlas of MtGEA [42]). An increased expression level of a nonfunctional isoform with an intron retained, but at the level of statistical significance, may indicate that the transcript is not degraded by the NMD system, but acquires a new function (for example, in signaling) because of changes in the secondary and tertiary structures.

As expected, the number of isoforms with excised intron (functional form) of *PsPR-10* and *PsWRKY40* genes was reduced, since the plant probably needs to suppress defense reactions in nodules when interacting

with nitrogen-fixing bacteria, and the isoform with retained intron can be widely represented either because it is not capable of serving as a transcription factor (in case of *PsWRKY40*) and is subject to degradation through the NMD system or because it acquires a new function because of a change in the protein structure (in case of *PsPR-10*).

CONCLUSION

In this work, the method of assembling "supertranscripts" [25] was first used for the analysis of AS in nodules of garden pea. The results obtained enable us to conclude that, in nitrogen-fixing nodules, specific ratios of the isoforms of some genes are registered (*PsPR-10* and *PsWRKY40*), and in general, AS is more intense in comparison with the root tips. We also found that, in garden pea nodules, as in the nodules of *L. japonicus*, there are two isoforms of the *PsSIP1* gene, namely, those with retained intron and without, although it has been previously thought [9] that the presence of the short isoform in nodules is a phenomenon specific to *L. japonicus*. For a more detailed and reliable study of AS in pea, a more complex analysis is required, based on data from genomic and transcriptome sequencing, which the authors are currently working on.

ACKNOWLEDGMENTS

REFERENCES

1. Тихонович И.А., Андронов Е.Е., Борисов А.Ю., и др. Принцип дополнительности геномов в расширении адаптационного потенциала растений // Генетика. — 2015. — Т. 51. — № 9. — С. 973−990. [Tikhonovich IA, Andronov EE, Borisov AY. The principle of genome complementarity in the enhancement of plant adaptive capacities. *Genetika*. 2015;51(9): 973-990. (In Russ.)]. https://doi.org/10.7868/S001667581509012X.

2. Shtark OY, Sulima AS, Zhernakov AI, et al. Arbuscular mycorrhiza development in pea (*Pisum sativum* L.) mutants impaired in five early nodulation genes including putative orthologs of NSP1 and NSP2. *Symbiosis*. 2016;68(1-3):129-144. https://doi.org/10.1007/s13199-016-0382-2.

3. Reddy AS, Marquez Y, Kalyna M, Barta A. Complexity of the alternative splicing landscape in plants. *Plant Cell*. 2013;25(10):3657-3683. https://doi.org/10.1105/tpc.113.117523.

4. Gracz J. Alternative splicing in plant stress response. *Biotechnologia*. 2016;97(1):9-17. https://doi.org/10.5114/bta.2016.57719.

5. Laloum T, Martin G, Duque P. Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci*. 2018;23(2):140-150. https://doi.org/10.1016/j.tplants.2017.09.019.

6. Shang X, Cao Y, Ma L. Alternative Splicing in Plant Genes: A Means of Regulating the Environmental Fitness of Plants. *Int J Mol Sci*. 2017;18(2). https://doi.org/10.3390/ijms18020432.

7. Zhiguo E, Wang L, Zhou J. Splicing and alternative splicing in rice and humans. *BMB Rep*. 2013;46(9):439-447. https://doi.org/10.5483/BMBRep.2013.46.9.161.

8. Zenoni S, Ferrarini A, Giacomelli E, et al. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol*. 2010;152(4):1787-1795. https://doi.org/10.1104/pp.109.149716.

9. Wang C, Zhu H, Jin L, et al. Splice variants of the SIP1 transcripts play a role in nodule organogenesis in *Lotus japonicus*. *Plant Mol Biol*. 2013;82(1-2): 97-111. https://doi.org/10.1007/s11103-013-0042-3.

10. Walters B, Lum G, Sablok G, Min XJ. Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. *DNA Res*. 2013;20(2):163-171. https://doi.org/10.1093/dnares/dss041.

11. Iniguez LP, Ramirez M, Barbazuk WB, Hernandez G. Identification and analysis of alternative splicing events in *Phaseolus vulgaris* and *Glycine max*. *BMC Genomics*. 2017;18(1):650. https://doi.org/10.1186/s12864-017-4054-2.

12. Kelemen O, Convertini P, Zhang Z, et al. Function of alternative splicing. *Gene*. 2013;514(1):1-30. https://doi.org/10.1016/j.gene.2012.07.083.

13. Wang Y, Liu J, Huang BO, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep*. 2015;3(2):152-158. https://doi.org/10.3892/br.2014.407.

14. Zhukov VA, Zhernakov AI, Kulaeva OA, et al. *De Novo* Assembly of the Pea (*Pisum sativum* L.) Nodule Transcriptome. *Int J Genomics*. 2015;2015:695947. https://doi.org/10.1155/2015/695947.

15. Kumagai H, Hakoyama T, Umehara Y, et al. A Novel Ankyrin-Repeat Membrane Protein, IGN1, Is Required for Persistence of Nitrogen-Fixing Symbiosis in Root Nodules of *Lotus japonicus*. *Plant Physiol*.

2007;143(3):1293-1305. https://doi.org/10.1104/pp.106.095356.

16. Alves-Carvalho S, Aubert G, Carrere S, et al. Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 2015;84(1):1-19. https://doi.org/10.1111/tpj.12967.

17. Jiao K, Li X, Guo W, et al. High-Throughput RNA-Seq Data Analysis of the Single Nucleotide Polymorphisms (SNPs) and Zygomorphic Flower Development in Pea (*Pisum sativum* L.). *Int J Mol Sci.* 2017;18(12). https://doi.org/10.3390/ijms18122710.

18. Sudheesh S, Sawbridge TI, Cogan NO, et al. *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics.* 2015;16:611. https://doi.org/10.1186/s12864-015-1815-7.

19. Bioinformatics.babraham.ac.uk [Internet]. FASTQC [cited 2018 Nov 2]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

21. jgi.doe.gov/data-and-tools [Internet]. bbtools; [cited 2018 Nov 2]. Available from: https://jgi.doe.gov/data-and-tools/bbtools/.

22. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. https://doi.org/10.1093/bioinformatics/bts635.

23. Fu L, Niu B, Zhu Z, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150-3152. https://doi.org/0.1093/bioinformatics/bts565.

24. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biol.* 2014;15(7):410. https://doi.org/10.1186/s13059-014-0410-6.

25. Hawkins AD, Oshlack A, Davidson NM. SuperTranscript: a data driven reference for analysis and visualisation of transcriptomes. *BioRxiv.* 2016:077750. https://doi.org/10.1101/077750.

26. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178-192. https://doi.org/10.1093/bib/bbs017.

27. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7):e21800. https://doi.org/10.1371/journal.pone.0021800.

28. Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol.* 2017;34(8):2115-2122. https://doi.org/10.1093/molbev/msx148.

29. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005;33(Web Server issue): W465-467. https://doi.org/10.1093/nar/gki458.

30. Gasteiger E, Gattiker A, Hoogland C, et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003;31(13):3784-3788. https://doi.org/10.1093/nar/gkg563.

31. UniProt C. The universal protein resource (UniProt). *Nucleic Acids Res.* 2008;36(Database issue): D190-195. https://doi.org/10.1093/nar/gkm895.

32. Finn RD, Bateman A, Clements J, et al. Pfam: The protein families database. *Nucleic Acids Res.* 2014;42(D1):D222-D230. https://doi.org/10.1093/nar/gkt1223.

33. spss-tutorials.com [Internet]. ANOVA − Simple Introduction [cited 2018 Nov 2]. Available from: https://www.spss-tutorials.com/anova-what-is-it/.

34. Kosterin OE, Rozov SM. Mapping of the new mutation blb and the problem of integrity of linkage group I. *Pisum Genet.* 1993;25:27-31.

35. Serova TA, Tsyganova AV, Tsyganov VE. Early nodule senescence is activated in symbiotic mutants of pea (*Pisum sativum* L.) forming ineffective nodules blocked at different nodule developmental stages. *Protoplasma.* 2018;255(5):1443-1459. https://doi.org/10.1007/s00709-018-1246-9.

36. Chakraborty J, Ghosh P, Sen S, Das S. Epigenetic and transcriptional control of chickpea WRKY40 promoter activity under *Fusarium* stress and its heterologous expression in *Arabidopsis* leads to enhanced resistance against bacterial pathogen. *Plant Sci.* 2018;276:250-267. https://doi.org/10.1016/j.plantsci.2018.07.014.

37. Birkenbihl RP, Kracher B, Ross A, et al. Principles and characteristics of the *Arabidopsis* WRKY regulatory network during early MAMP-triggered immunity. *Plant J.* 2018;96(3):487-502. https://doi.org/10.1111/tpj.14043.

38. Xu X, Chen C, Fan B, Chen Z. Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell.* 2006;18(5):1310-1326. https://doi.org/10.1105/tpc.105.037523.

39. Rayson S, Arciga-Reyes L, Wootton L, et al. A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in *Arabidopsis thaliana* NMD mutants. *PLoS One*. 2012;7(2):e31917. https://doi.org/10.1371/journal.pone.0031917.

40. Ruszkowski M, Szpotkowski K, Sikorski M, Jaskolski M. The landscape of cytokinin binding by a plant nodulin. *Acta Crystallogr D Biol Crystallogr*. 2013;69(Pt 12):2365-2380. https://doi.org/10.1107/S0907444913021975.

41. Liu JJ, Ekramoddoullah AKM. The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiol Mol Plant Pathol*. 2006;68(1-3):3-13. https://doi.org/10.1016/j.pmpp.2006.06.004.

42. mtgea.noble.org/v3/ [Internet]. The *Medicago truncatula* Gene Expression Atlas (MtGEA) Project [cited 2018 Nov 3]. Available from: https://mtgea.noble.org/v3/.

✿ Information about the authors

Evgeny A. Zorin — Technician, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: kjokkjok8@gmail.com.

Olga A. Kulaeva — PhD, Senior Scientist, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: okulaeva@arriam.ru.

Alexey M. Afonin — Researcher, Laboratory of Genetics of Plant-microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: afoninalexeym@gmail.com.

Vladimir A. Zhukov — PhD, Head of the Lab, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: vzhukov@arriam.ru.

Igor A. Tikhonovich — Sc.D., Professor PI, Academician of RAS, All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia; Dean of the Faculty, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia. E-mail: arriam2008@yandex.ru.

✿ Информация об авторах

Евгений Андреевич Зорин — техник 1-й категории, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: kjokkjok8@gmail.com.

Ольга Алексеевна Кулаева — канд. биол. наук, старший научный сотрудник, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: okulaeva@arriam.ru.

Алексей Михайлович Афонин — инженер-исследователь, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: afoninalexeym@gmail.com.

Владимир Александрович Жуков — канд. биол. наук, заведующий лабораторией, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: vzhukov@arriam.ru.

Игорь Анатольевич Тихонович — д-р биол. наук, научный руководитель института, академик РАН, ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург; декан, биологический факультет, ФГБУ ВПО «Санкт-Петербургский государственный университет», Санкт-Петербург. E-mail: arriam2008@yandex.ru.