

<https://doi.org/10.17816/ecogen17339-46>

## IDENTIFICATION OF SEQUENCES ENCODING FOR NCR-PEPTIDES AND DEFENSINS IN THE 'META-ASSEMBLY' OF TRANSCRIPTOME OF PEA (*PISUM SATIVUM* L.) NITROGEN-FIXING NODULES

© E.A. Zorin<sup>1</sup>, M.S. Kliukova<sup>1</sup>, O.A. Kulaeva<sup>1</sup>, A.M. Afonin<sup>1</sup>, I.A. Tikhonovich<sup>1,2</sup>, V.A. Zhukov<sup>1</sup>

<sup>1</sup>All-Russia Research Institute for Agricultural Microbiology, Saint Petersburg, Russia;

<sup>2</sup>Saint Petersburg State University, Saint Petersburg, Russia

*Cite this article as:* Zorin EA, Kliukova MS, Kulaeva OA, et al.

Identification of sequences encoding for NCR-peptides and defensins in the 'meta-assembly' of transcriptome of pea (*Pisum sativum* L.) nitrogen-fixing nodules.

*Ecological genetics.* 2019;17(3):39-46. <https://doi.org/10.17816/ecogen17339-46>.

Received: 25.03.2019

Revised: 12.07.2019

Accepted: 24.09.2019

✿ **Background.** The active and careless applying of antibiotics in medicine and agriculture leads to the emergence of resistance to the existing antimicrobial drugs, which reduces the effectiveness of their use. One of the ways to solve this problem is the development of new antibiotics based on plant peptides with antimicrobial activity, for example plant defensins (which identified in all plants) and NCR peptides that are specifically synthesized in nodules of some leguminous plants. **Materials and methods.** In the present study, a meta-assembly of a transcriptome was constructed based on publicly available RNA-sequencing transcriptomes of pea nodules (*Pisum sativum* L.). This meta-assembly was used to search for sequences encoding antimicrobial peptides. **Results.** As a result, 55 and 908 unique sequences encoding defensins and NCR peptides, respectively, were identified. The recognition site for the signal peptidase was predicted and sequences were divided into the signal and mature part of the peptide. Among mature defensins, 22 peptides possess *in silico* predicted antimicrobial activity, and for the NCR peptides family their number was 422. **Conclusion.** Sequences encoding defensins and NCR peptides expressed in nitrogen-fixing pea nodules were identified. They are candidates for testing their antimicrobial activity *in vitro*.

✿ **Keywords:** plant antimicrobial peptides; nitrogen-fixing nodules; rhizobial-legume symbiosis; *P. sativum*; transcriptome; defensins; NCR peptides.

## ИДЕНТИФИКАЦИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, КОДИРУЮЩИХ NCR-ПЕПТИДЫ И ДЕФЕНЗИНЫ, В МЕТАСБОРКЕ ТРАНСКРИПТОМА АЗОТФИКСИРУЮЩИХ КЛУБЕНЬКОВ ГОРОХА ПОСЕВНОГО (*PISUM SATIVUM* L.)

© E.A. Зорин<sup>1</sup>, M.C. Клюкова<sup>1</sup>, O.A. Кулаева<sup>1</sup>, A.M. Афонин<sup>1</sup>, И.А. Тихонович<sup>1,2</sup>, В.А. Жуков<sup>1</sup>

<sup>1</sup>ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Санкт-Петербург;

<sup>2</sup>ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург

*Для цитирования:* Зорин E.A., Клюкова M.C., Кулаева O.A., и др. Идентификация последовательностей, кодирующих NCR-пептиды и дефензины, в метасборке транскриптома азотфиксирующих клубеньков гороха посевного (*Pisum sativum* L.) // Экологическая генетика. — 2019. — Т. 17. — № 3. — С. 39–46. <https://doi.org/10.17816/ecogen17339-46>.

Поступила: 25.03.2019

Одобрена: 12.07.2019

Принята: 24.09.2019

✿ Активное и зачастую бесконтрольное применение антибиотиков в медицине и сельском хозяйстве приводит к возникновению резистентности к используемым веществам, что снижает эффективность их применения. Один из способов решения данной проблемы — разработка новых антибиотиков на основе растительных пептидов, обладающих антимикробной активностью. К таковым относятся дефензины (характерные для всех растений) и NCR-пептиды, специфически синтезируемые в клубеньках некоторых бобовых растений. В настоящем исследовании из доступных данных РНК-секвенирования транскриптома клубеньков гороха посевного (*Pisum sativum* L.) была получена метасборка транскриптома, использованная для поиска последовательностей, кодирующих антимикробные пептиды. В результате было идентифицировано 55 и 908 уникальных последовательностей, кодирующих дефензины и NCR-пептиды соответственно. Последовательности, для которых был предсказан сайт узнавания сигнальной пептидазой, были разделены на сигнальную и зрелую части пептида. Среди зрелых дефензинов антимикробной активностью, предсказанной *in silico*, обладают 22 пептида, среди представителей семейства NCR-пептидов — 422 последовательности. Таким образом, были идентифицированы гены, экспрессирующиеся в азотфиксирующих клубеньках гороха и кодирующие дефензины и NCR-пептиды, являющиеся кандидатами для проверки их антимикробной активности в опытах *in vitro*.

✿ **Ключевые слова:** антимикробные пептиды растений; азотфиксирующие клубеньки; бобово-ризобиальный симбиоз; *P. sativum*; транскриптомика; дефензины; NCR-пептиды.

## BACKGROUND

Formulations that include antibiotics and insecticides of different natures are often used to maintain the yield of agricultural crops by protecting them against pathogenic microorganisms. However, many antibiotics can cause microbial resistance, which significantly reduces their effectiveness and calls for the development of new formulations [1, 2]. Thus, the search for naturally existing molecules that have antimicrobial activity is of great importance.

Plants, as fixed organisms, are unable to escape the effects of biotic and abiotic environmental factors and, instead, have to adapt to them. Therefore, interaction with different symbiotic microorganisms that enhance the adaptive potential of the plant–microbe system is very important for the plant lifecycle [3]. However, only some microorganisms are useful to plants, whereas others demonstrate pathogenic properties; therefore, in the course of evolution, plants have developed multi-faceted and various protection systems against microorganisms. A number of plant systems are used to control microbial life inside the plant, the main one being the immune response, which consists of identification and destruction of pathogenic bacteria and fungi. One component of this system is antimicrobial peptides (AMPs), also known as defensins [4]. Defensins are cysteine-rich peptides consisting of 45–54 amino acids. Defensins of plants, fungi, and mammals are similar in structure and function. Most defensins are involved in the inhibition of fungal infection. Some representatives of the family also demonstrate antibacterial activity and are involved in the development of resistance to some abiotic factors [5]. Defensins and other AMPs (for example, hevein-like peptides and thoinines) monitor microorganisms at the whole-plant level and are produced in all tissues and organs. In addition to whole-plant systems, local systems ensure response to bacterial penetration; for example, these local systems exist in the nitrogen-fixing nodules of leguminous plants attributed to inverted repeat-lacking clade (IRLC), presented by lucerne (*Medicago truncatula* Gaertn.), garden pea (*Pisum sativum* L.), faba bean (*Vicia faba* L.), white clover (*Trifolium repens* L.), etc. [6–9].

It is known that leguminous plants have a selective advantage in their ability to enter into symbiotic relationships with soil bacteria. Bacteria penetrate through the root hairs of plants into the special *de novo* formed organ, the root nodule, in which the processes of endosymbiont differentiation take place, including the conversion of bacteria into a symbiotic form known as bacteroides [10]. This process results in the bacterial conversion (fixation) of atmospheric nitrogen into a form that is available to the host plant [11]. Recently, it has been demonstrated that the process of terminal (irreversible) differentiation of endosymbionts in IRLC leguminous

plants is controlled by a family of defensin-like nodule-specific cysteine-rich (NCR) peptides, which are probably also a part of the immune system of the nodule [7]. It is supposed that NCR peptides, as well as defensins, perform their biological functions by means of demonstration of antimicrobial activity [12–16].

Defensins and NCR peptides vary widely in their nucleotide and amino acid sequences, which allows them to suppress the vital activity of a wide range of microorganisms and do not cause resistance [17]. Because of the growing issue of resistance to antibiotics, the search for and study of these molecules in different plants, both model and non-model, is an important area of research.

In pea plants, in contrast to the *Medicago truncatula* plant model, the search for representatives of the family is complicated because of the absence of genome sequencing data. On the other hand, there is a considerable amount of data on the available transcriptome databases for the organs and tissues of *Pisum sativum*, including the nodules. However, these transcriptome data were assembled by different methods (assembly programs) with different options, which has resulted in the loss of some information. In order to obtain more exhaustive information about pea transcripts, we created a meta-assembly of the nodule transcriptome of *P. sativum* by using published reads of transcriptomes of different pea lines. This meta-assembly is a combination of separately assembled (by means of one assembly program with similar options) transcriptomes of pea nodules of different cultivar. Our analysis aims to improve the effectiveness of searches for target sequences in comparison with previous work, which used data generated by different work groups using different software tools. Our meta-assembly was successfully used for the search for AMPs attributed to the families of defensins and NCR peptides.

## MATERIALS AND METHODS

At the first stage of work, the search for peptides with antimicrobial activity was conducted by means of existing assemblies of transcriptomes of pea nodules available in the NCBI database. Assemblies of the lines Kaspia, Parafield, Cameor, and SGE were used for the work [18–20].

The second stage involved improvement of available assemblies by means of their re-assembly using the latest, updated version of the Trinity program and compilation into meta-assembly.

The quality of reads was estimated by means of FastQC [21]. reads of low quality and adapters were deleted using BBduck from the BBmap package [22]. After filtration of the sequencing data of the pea cultivar, Cameor, nodules, 148,408,730 (89%) reads were left out of 166,803,965 reads; for SGE, 42,054,133

Table 1

**Description of transcriptomes of pea nodules collected within the frame of the work**

Pea line	Contig average length	Percent of transcripts with Open reading frames	Total contigs
SGE	932.3	65.5	48649
Cameor	877.7	47.4	212147
Kaspa	487.1	62.7	113512
Parafield	484.6	61.9	91831

reads out of 52,021,865 were left (81%); for Kaspa, 30,945,747 out of 31,256,637 reads were left (99%); and for Parafield, 20,637,281 out of 20,842,187 reads were left (99%). The assembly of transcriptomes out of the reads that passed processing was made by means of the Trinity program (ver. 2.8.4) [23]. The quality of assemblies was assessed by means of transRate [24] (Table 1).

Structural annotation of assembled contigs was conducted by means of Transdecoder [25]. Assembly of transcriptomes for quantitative analysis of gene expression (quantification) and analysis of single-nucleotide polymorphisms assumes removal of transcripts without the open reading frame or with the open reading frame of less than 100 nucleotides, removal of duplicated transcripts and sequences overlapping for 50% and more with 100% similarity, and removal of all contigs, for which homologous sequences were not found in the available databases. However, the goal of the work was to search for short AMPs, and the abovementioned operations can result in the loss of part of the target sequences, which is crucial in our case. That is why only 100% identical sequences were deleted (397 transcripts, 0.09% of the total), and all contigs containing open reading frames longer than 90 nucleotides were retained. Further analysis was conducted on transcriptomic assembly of the pea nodules of different cultivars, which included 349,953 contigs.

Search of AMPs in the *de novo* meta-assembly of transcriptomes of the pea nodules was conducted using the SPADA program [26], a pipeline that organizes the sequential launch of a number of other tools: Augustus [27], Glimmer [28], GeneWise [29], and Sig-

nalP [30]. Detection of signal sequence in the detected peptides and separation of signal peptides from the mature part were performed using the SignalP program. The presence of four or six cysteines in the mature part of the NCR peptide was validated by means of multiple alignment algorithm implemented in the MAFFT program [31] with option GINSi for more accurate alignment. The physical and chemical properties of the mature peptides were predicted using the service DBAASP [32]. The results were processed by means of a script made in PHP programming language. The antibacterial activity of detected peptides was predicted by means of the web service CAMP3 [33]. Results were analyzed by means of script made in Python programming language.

**RESULTS**

The first round of searches for sequences encoding NCR peptides in previously obtained transcriptomic assemblies of the pea nodules resulted in the identification of 553 *de novo* NCR peptides. 355 additional sequences not detected in the previous assemblies were identified as a result of transcriptome reassembly and repeated search using SPADA. Thus, 908 unique sequences encoding NCR peptides were detected. The boundaries between the signal sequence and the mature part of the peptide were predicted in 838 of these sequences only predicted sequences of mature peptides were used for further analysis.

Physical and chemical properties and antibacterial activity were predicted for the mature part of the 838 NCR peptides using machine learning methods (Tables 2 and 4). Additionally, 55 defensins were identified, 41 of

Table 2

**Prediction of antimicrobial activity of NCR peptides' mature part**

Quantity of peptides	Support vector machines	Random forest	Discriminatory analysis
With antimicrobial activity	328	422	332
Without antimicrobial activity	510	416	506

Table 3

**Prediction of antimicrobial activity of defensins' mature part**

Quantity of peptides	Support vector machines	Random forest	Discriminatory analysis
With antimicrobial activity	15	22	16
Without antimicrobial activity	26	18	25

Table 4

**Amino acid sequences of five NCR peptides having the highest predicted antimicrobial activity**

Sequence	Support vector machines	Random forest	Discriminatory analysis
QACIHDRQCRCTQHTVSKINGFCKCYISNT	0.937	0.893	0.994
IKCDVQADCPKIPNLPFAIYKCKCRLIG	0.951	0.923	0.995
IVVTVDIKEIFKCESHKQCRKQMPNCKRPKIARCVSRTCKCW	0.865	0.698	0.988
KRACTYHHQCSDISCSYGYISLCEKYCHCVKN	0.987	0.860	0.994
KKIKCRTSDCPKQMCRFPTYNKCVRNHCKCVLSRII	0.891	0.751	0.990

Table 5

**Amino acid sequences of five defensins having the highest predicted antimicrobial activity**

Sequence	Support vector machines	Random forest	Discriminatory analysis
KVCEKRSKTWSGFCGKTRNCKKQCINVENAVFGACHRQGFQFACFCYFKC	0.984	0.967	0.998
KTCENLSGTFKGPCIPDGNCNKHCRNNEHLLSGRCRDDFRWCWTRNC	0.826	0.945	0.902
RLCGRTSKTWRGPCYINLSCNAECTMKEQAIFGTCQHFEFCFC	0.664	0.851	0.816
RDCESKSHKFKGTCLSDTNCASVCQTERFTGGHCRGRFRHRCYCTTHC	0.923	0.845	0.917
KFCRSKSRTWSGACIDNSPCSTECKQLEHASHGACRHNGFCFCYFNC	0.913	0.935	0.962

which had a signal sequence and cleavage site between the signal and mature sequences. The results of the predictions of antibacterial activity for 41 defensins are presented in Tables 3 and 5.

**DISCUSSION**

A number of unique mechanisms aimed at protection against pathogens has been formed in plants during the process of evolution. These mechanisms include physical barriers and a wide range of synthesized secondary metabolites and AMPs. One of the largest families of AMPs is the defensins, which are attributed to evolutionary ancient sequences that occur in all major eukaryotic organisms [5]. Except for conservative residues of cysteine, these peptides demonstrate very low similarity of sequence, which explains the wide range of activities (antifungal, antibacterial, and activation by abiotic

factors) demonstrated by these peptides [5]. Besides defensins, several families of defensin-like sequences have been detected in plants, some of which have more specific natures of expression and more focused functionality. Thus, the group of NCR peptides detected in leguminous plants plays an important role in the differentiation of *Bacteroides* and the control of symbiotic development [34]. The goal of this study was to identify and characterize representatives of the family of defensins and NCR peptides in pea.

Optimization of the data processing algorithms, including reassembly of available pea transcriptomes using a unified method by means of the updated version of the Trinity program with their further combination, allowed detection of 908 sequences expressed in the nodules that code NCR peptides and 55 sequences that encode defensins. It is recommended that the complications of

meta-assembly of transcriptomes be reduced by eliminating duplicates, as well as transcripts that do not contain Open reading frames or contain reading frames that are shorter than 90 nucleotides, in order to save time and resources prior to searching for antibacterial peptides by means of SPADA. However, the set of peptides obtained in the search process shall be additionally checked by means of analysis of multiple sequences alignment for the presence of false positive results, which are inevitable in the process of algorithm operation.

The role of defensins in the development of plant–microbe symbiotic interactions is poorly studied. Only a few studies demonstrate involvement of defensins in symbiosis. Comparative transcriptomic analysis of roots and nodules formed in the process of the development of symbiosis by the plant *Datisca glomerata* and by representatives of the genus *Frankia* allowed detection of a group of defensins expressed in the nodules in a special way that differ from the other defensins with the presence of a special terminal domain [35]. Nodule-specific expression of the number of genes encoding defensins was demonstrated in the research into *Medicago truncatula* [34]. Activation of expression of the number of genes coding defensin-like peptides was detected during interaction of *Lotus japonicus* (Regel.) K. Larsen with fungi of the arbuscular mycorrhiza, *Rhizophagus irregularis* [36]. However, the specific role of detected peptides in the development of symbioses is still unknown.

Despite the similarity with defensins, the presence of conservative cysteine pattern, high variability of sequences, possibility of occurrence of antimicrobial activity, which is probably explained by the kinship of these two gene families, the main function of NCR peptides is the differentiation of the non-symbiotic bacterium into the nitrogen-fixing bacteriodes within the symbiotic compartment [37–39]. It is known that NCR peptides, especially ones with a high positive charge, can interact with electronegative bacterial membranes, specifying formation of pores on the surface of bacterial cells. The strong effect of such peptides can result in bacterial cell lysis [16]. However, in case of symbionts, pore formation has a temporary nature, and pores are probably required for penetration of other NCR peptides, which interact with the intracellular targets that modify cell physiology and morphology [15, 40]. It is supposed that an important condition is the availability of protein BacA on the bacterial membrane, which also promotes NCR peptide import inside the cell, where they can be neutralized by means of the bacterial system of protein degradation [41, 42].

In this study, 908 unique sequences encoding NCR peptides were found in *P. sativum*. It should be noted that more than 700 representatives of the family have been described for *M. truncatula* so far, which is not surprising as the pea genome is almost 10 times larger than the *M. truncatula* genome. On the other hand, chimeric sequences can

present among the pea peptide sequences detected in the work because of imperfection of the assembly algorithm; besides, some detected sequences can be the allelic conditions of the same AMP-coding gene. The problem of chimeric sequences can be resolved only after whole-genome sequencing. The problem of allelic variants can be resolved by means of sequencing of the transcriptomes of each of the studied pea lines and assessment the coverage by reads of every peptide. Thus, methods of next-generation sequencing shall be used for the verification of detected peptides and will be the subject of further research.

Representatives with predicted antimicrobial activity were detected among defensins and NCR peptides. Larger amounts of sequences with a high level of probability of antimicrobial activity were detected among NCR peptides in comparison with the identified defensins, which indicates their better representation in the transcriptomic data. Such diversity of NCR peptides in terms of the content and, accordingly, the possibilities and levels of antimicrobial activity in the nodule cells can indicate that NCR peptides inside the nodule functionally substitute defensins and form the local immune system of the plant, which eliminates undesirable bacteria and at the same time promotes differentiation of symbiotic partners.

## CONCLUSION

Identification and characterization of defensins and defensin-like peptides are important for the understanding of their roles in plant protection responses, as well as for the detection of new antimicrobial compounds. This area of research is urgently required as pathogenic bacteria of agricultural plants inflict significant economic damage and affect food safety [43]. The interest in defensins is justified by the possibility of developing new types of protective agents on their basis. As AMPs can seriously vary in different types of plants, the important task is identification of AMP across many species, including agriculturally valuable crops. Leguminous plants are an interesting object for AMP study as they contain both defensins and defensin-like NCR peptides unique to leguminous plants. Because defensins and NCR peptides are numerous and variable in their content, attack different targets in the cells of pathogenic microorganisms, and are jointly produced, microorganisms cannot resist such a cocktail of AMP and obtain resistance to them. The identified AMPs with a wide range of activities can be used to counteract plant pathogens, as well as the pathogens of animals and humans.

Characterization of natural antimicrobial molecules is an important area of research. The example of protein families studied in this work and the description of these molecules contribute much to the understanding of the processes of interaction of plants and bacteria, as well as revealing the mechanisms of different responses to interaction with pathogenic and useful bacteria. Understand-

ing the characteristics of antimicrobial activity allows the construction or modernization of existing peptides, which can be used as new-generation antibiotics, as well as managing the effectiveness of nitrogen-fixing symbiosis.

#### Acknowledgment

The work of antimicrobial sequences detection attributed to NCR peptides family was supported by the Russian Foundation of Basic Research (Grant No. 183400187 мол\_a). Analysis of sequences coding defensins was financed by the State Task of the Rf Ministry of Education and Science (No. 066420190022C-01) for 2019.

#### REFERENCES

- Aslam B, Wang W, Arshad MI, et al. Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist.* 2018;11: 1645-58. <https://doi.org/10.2147/IDR.S173867>.
- Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. *PT.* 2015;40(4):277-283.
- Тихонович И.А., Андронов Е.Е., Борисов А.Ю., и др. Принцип дополнителности геномов в расширении адаптационного потенциала растений // Генетика. — 2015. — Т. 51. — № 9. — С. 973–990. [Tikhonovich IA, Andronov EE, Borisov AY, et al. The principle of genome complementarity in the enhancement of plant adaptive capacities. *Russian Journal of Genetics.* 2015;51(9):831-846. (In Russ.)]. <https://doi.org/10.7868/S001667581509012X>.
- Nawrot R, Barylski J, Nowicki G, et al. Plant antimicrobial peptides. *Folia Microbiol (Praha).* 2014;59(3):181-196. <https://doi.org/10.1007/s12223-013-0280-4>.
- Parisi K, Shafee TMA, Quimbar P, et al. The evolution, function and mechanisms of action for plant defensins. *Semin Cell Dev Biol.* 2019;88:107-118. <https://doi.org/10.1016/j.semcdb.2018.02.004>.
- Kato T, Kawashima K, Miwa M, et al. Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. *Mol Plant Microbe Interact.* 2002;15(2):129-137. <https://doi.org/10.1094/MPMI.2002.15.2.129>.
- Mergaert P, Nikovics K, Kelemen Z, et al. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* 2003;132(1):161-173. <https://doi.org/10.1104/pp.102.018192>.
- Frühling M, Albus U, Hohnjec N, et al. A small gene family of broad bean codes for late nodulins containing conserved cysteine clusters. *Plant Sci.* 2000;152(1):67-77. [https://doi.org/10.1016/S0168-9452\(99\)00219-8](https://doi.org/10.1016/S0168-9452(99)00219-8).
- Crockard A, Bjourson J, Dazzo B, Cooper JE. A white clover nodulin gene, dd23b, encoding a cysteine cluster protein, is expressed in roots during the very early stages of interaction with *Rhizobium leguminosarum biovar trifolii* and after treatment with chitolipooligosaccharide Nod factors. *J Plant Res.* 2002;115(6):439-447. <https://doi.org/10.1007/s10265-002-0053-7>.
- Yeragani VK, Pohl R, Balon R. Lactate-induced panic and beta-adrenergic blockade. *Psychiatry Res.* 1990;32(1): 93-94. [https://doi.org/10.1016/0165-1781\(90\)90139-v](https://doi.org/10.1016/0165-1781(90)90139-v).
- Mylona P, Pawlowski K, Bisseling T. Symbiotic nitrogen fixation. *Plant Cell.* 1995;7(7):869-885. <https://doi.org/10.1105/tpc.7.7.869>.
- Arnold MF, Shabab M, Penterman J, et al. Genome-wide sensitivity analysis of the microsymbiont *Sinorhizobium meliloti* to symbiotically important, defensin-like host peptides. *MBio.* 2017;8(4).pii: e01060-17. <https://doi.org/10.1128/mBio.01060-17>.
- Nagy K, Mikuláss KR, Végh AG, et al. Interaction of cysteine-rich cationic antimicrobial peptides with intact bacteria and model membranes. *Gen Physiol Biophys.* 2015; 34(2):135-144. [https://doi.org/10.4149/gpb\\_2015002](https://doi.org/10.4149/gpb_2015002).
- Ordögh L, Vörös A, Nagy I, et al. Symbiotic plant peptides eliminate *Candida albicans* both *in vitro* and in an epithelial infection model and inhibit the proliferation of immortalized human cells. *Biomed Res Int.* 2014;2014:320796. <https://doi.org/10.1155/2014/320796>.
- Farkas A, Maróti G, Durgó H, et al. Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci USA.* 2014;111(14):5183-5188. <https://doi.org/10.1073/pnas.1404169111>.
- Farkas A, Maróti G, Kereszt A, et al. Comparative analysis of the bacterial membrane disruption effect of two natural plant antimicrobial peptides. *Front Microbiol.* 2017;8:51. <https://doi.org/10.3389/fmicb.2017.00051>.
- Van der Weerden NL, Anderson MA. Plant defensins: common fold, multiple functions. *Fungal Biol Rev.* 2013;26(4):121-131. <https://doi.org/10.1016/j.fbr.2012.08.004>.
- 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(1):611. <https://doi.org/10.1186/s12864-015-1815-7>.
- Alves-Carvalho S, Aubert G, Carrère 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>.
- Zhukov VA, Zhernakov AI, Kulaeva OA, et al. De novo assembly of the pea (*Pisum sativum* L.) nodule transcriptome. *Int J Genomics.* 2015;2015:1-11. <https://doi.org/10.1155/2015/695947>.
- Babraham Bioinformatics. FastQC [Internet]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

22. Joint Genome Institute. BBTools [Internet]. Available from: <https://jgi.doe.gov/data-and-tools/bbtools/>.
23. Grabherr MG, Haas BJ, Yassour M, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* 2011;29(7):644-652. <https://doi.org/10.1038/nbt.1883>.
24. Smith-Unna R, Boursnell C, Patro R, et al. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* 2016;26(8):1134-1144. <https://doi.org/10.1101/gr.196469.115>.
25. Haas BJ, Papanicolaou A, Yassour M, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013;8(8):1494-1512. <https://doi.org/10.1038/nprot.2013.084>.
26. Zhou P, Silverstein KA, Gao L, et al. Detecting small plant peptides using SPADA (small peptide alignment discovery application). *BMC Bioinformatics.* 2013;14(1):335. <https://doi.org/10.1186/1471-2105-14-335>.
27. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005;33:W465-W467. <https://doi.org/10.1093/nar/gki458>.
28. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics.* 2007;23(6):673-679. <https://doi.org/10.1093/bioinformatics/btm009>.
29. Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res.* 2004;14(5):988-995. <https://doi.org/10.1101/gr.1865504>.
30. Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420-423. <https://doi.org/10.1038/s41587-019-0036-z>.
31. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059-3066. <https://doi.org/10.1093/nar/gkf436>.
32. Pirtskhalava M, Gabrielian A, Cruz P, et al. DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res.* 2016;44(D1):D1104-1112. <https://doi.org/10.1093/nar/gkv1174>.
33. Waghugh FH, Barai RS, Gurung P, Idicula-Thomas S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.* 2016;44(D1): D1094-1097. <https://doi.org/10.1093/nar/gkv1051>.
34. Maróti G, Downie JA, Kondorosi É. Plant cysteine-rich peptides that inhibit pathogen growth and control *rhizobial* differentiation in legume nodules. *Curr Opin Plant Biol.* 2015;26:57-63. <https://doi.org/10.1016/j.pbi.2015.05.031>.
35. Demina IV, Persson T, Santos P, et al. Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PLoS ONE.* 2013;8(8):e72442. <https://doi.org/10.1371/journal.pone.0072442>.
36. Handa Y, Nishide H, Takeda N, et al. RNA-seq Transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant Cell Physiol.* 2015;56(8):1490-1511. <https://doi.org/10.1093/pcp/pcv071>.
37. Silverstein KA, Graham MA, VandenBosch KA. Novel paralogous gene families with potential function in legume nodules and seeds. *Curr Opin Plant Biol.* 2006;9(2): 142-6. <https://doi.org/10.1016/j.pbi.2006.01.002>.
38. Mergaert P, Uchiumi T, Alunni B, et al. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc Natl Acad Sci USA.* 2006;103(13):5230-5235. <https://doi.org/10.1073/pnas.0600912103>.
39. Van de Velde W, Zehirov G, Szatmari A, et al. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science.* 2010;327(5969):1122-1126. <https://doi.org/10.1126/science.1184057>.
40. Maróti G, Kondorosi E. Nitrogen-fixing *Rhizobium*-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front Microbiol.* 2014;5:326. <https://doi.org/10.3389/fmicb.2014.00326>.
41. Haag AF, Baloban M, Sani M, et al. Protection of sinorhizobium against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol.* 2011;9(10):e1001169. <https://doi.org/10.1371/journal.pbio.1001169>.
42. Meadows R. How symbiotic bacteria survive host defenses. *PLoS Biol.* 2011;9(10):e1001164. <https://doi.org/10.1371/journal.pbio.1001164>.
43. Mansfield J, Genin S, Magori S, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol.* 2012;13(6):614-629. <https://doi.org/10.1111/j.1364-3703.2012.00804.x>.

## ✉ Authors and affiliations

**Evgeny A. Zorin** — engineer-researcher, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. E-mail: [kjokkjok8@gmail.com](mailto:kjokkjok8@gmail.com).

## ✉ Информация об авторах

**Евгений Андреевич Зорин** — инженер-исследователь, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. E-mail: [kjokkjok8@gmail.com](mailto:kjokkjok8@gmail.com).

## ✿ Authors and affiliations

**Marina Sergeevna Kliukova** — engineer-researcher, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. E-mail: marina.kliukova@gmail.com.

**Olga A. Kulaeva** — PhD, Senior Scientist, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, 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-8, St. Petersburg, Russia. E-mail: afoninalexeym@gmail.com.

**Igor A. Tikhonovich** — Sc.D., Professor PI, Academician of RAS, All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia; Dean of the Faculty, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia. E-mail: arriam2008@yandex.ru.

**Vladimir A. Zhukov** — PhD, Head of the Lab, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. E-mail: vzhukov@arriam.ru.

## ✿ Информация об авторах

**Марина Сергеевна Клюкова** — инженер-исследователь, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. E-mail: marina.kliukova@gmail.com.

**Ольга Алексеевна Кулаева** — канд. биол. наук, старший научный сотрудник, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, 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-8, Санкт-Петербург. E-mail: afoninalexeym@gmail.com.

**Игорь Анатольевич Тихонович** — д-р биол. наук, научный руководитель института, академик РАН, ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург; декан, биологический факультет, ФГБУ ВПО «Санкт-Петербургский государственный университет», Санкт-Петербург. E-mail: arriam2008@yandex.ru.

**Владимир Александрович Жуков** — канд. биол. наук, заведующий лабораторией, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. E-mail: vzhukov@arriam.ru.