

## Symbiogenetics

UDC 631.461.52:575.1:577.257.065

doi: 10.15389/agrobiology.2017.5.995rus

doi: 10.15389/agrobiology.2017.5.995eng

### COMPARATIVE PHYLOGENETIC ANALYSIS OF SYMBIOTIC GENES OF DIFFERENT NODULE BACTERIA GROUPS USING THE METATREES METHOD

E.S. KARASEV, E.P. CHIZHEVSKAYA, B.V. SIMAROV, N.A. PROVOROV,  
E.E. ANDRONOV

All-Russian Research Institute for Agricultural Microbiology, Federal Agency of Scientific Organizations, 3, sh. Podbel'skogo, St. Petersburg, 196608 Russia, e-mail evgenii1991.karasev@gmail.com (corresponding author)

ORCID:

Karasev E.S. [orcid.org/0000-0003-4240-9587](https://orcid.org/0000-0003-4240-9587)

Provorov N.A. [orcid.org/0000-0001-9091-9384](https://orcid.org/0000-0001-9091-9384)

Chizhevskaya E.P. [orcid.org/0000-0002-7715-8696](https://orcid.org/0000-0002-7715-8696)

Andronov E.E. [orcid.org/0000-0002-5204-262X](https://orcid.org/0000-0002-5204-262X)

The authors declare no conflict of interests

Acknowledgements:

The experiments were carried out using equipment of the ARRIAM Center of Genomic Technologies, Proteomics and Cell Biology.

Supported by Russian Science Foundation, project № 14-26-00094П

Received December 12, 2016

#### Abstract

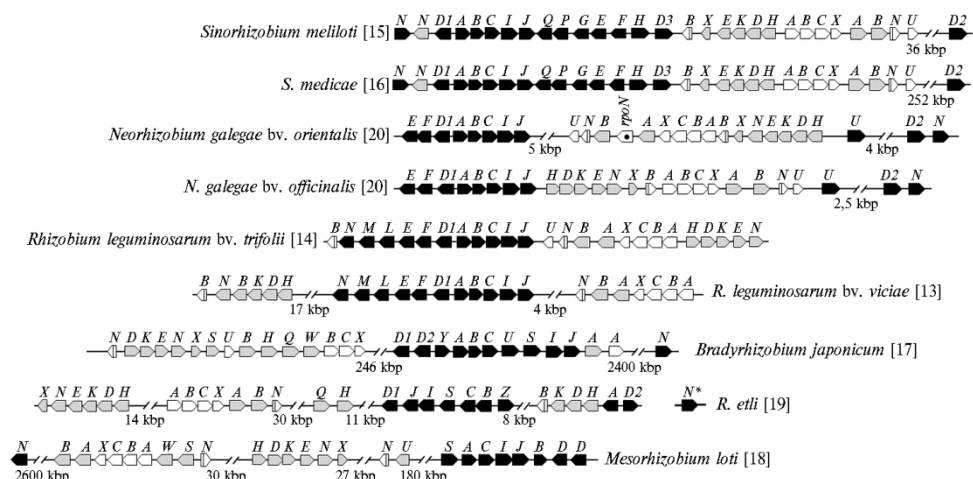
We applied modified phylogenetic analysis method of building meta-trees to study the evolutionary patterns of various groups of symbiotic genes (*nod*-genes control the formation of nodules and *nif/fix*-genes control symbiotic nitrogen fixation). The method consists in the pairwise comparison of topologies and construction combined dendrograms («meta-trees»), where the relative position of the two trees is a measure of corresponding gene phylogenies congruence. Homologues of 18 symbiotic genes (*nodABCDIJN*, *nifABDEHKN*, *fixABC*, *fdxB*) are present in each test organism (9 strains belong to the genera *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, *Sinorhizobium* and *Neorhizobium*). These genes were selected for the implementation of this method, as well as the gene 16S rRNA as a traditional taxonomic chromosome marker. We constructed and compared phylogenetic trees for all these genes and then calculated the pairwise similarity coefficients for their topologies. According to the obtained data we built a meta-tree, and there were two statistically distinct gene clusters identified within. Cluster 1 includes mainly *nif*- and *fix*-genes and cluster 2 — mostly *nod*-genes, that is related to the data of separate localization of these gene groups in the rhizobial genomes. The exception is the arrangement of *nifB* and *fixC* genes with *nodA* in cluster 2, as well as co-localization of *nodI* and *nifD* in cluster 1. During the identified clusters structure analysis we found strong relationship between the relative gene position and the characteristics of their genome localization in nodule bacteria. Importantly, the differences between clusters 1 and 2 are not expressed less clearly than the differences between *nod*- and *nif/fix*-gene groups. It is obvious that clusters 1 and 2 of our meta-tree reflect primarily different mechanisms of nodulation evolution and symbiotic nitrogen fixation associated with the independent origin of the relevant gene groups, and possibly, with their separate horizontal transfer between different groups of rhizobia. Further study of the symbiotic gene evolution in nodule bacteria requires improvements used phylogenetic analysis techniques, including separate analysis meta-tree for rhizobia, representing different stages of symbiosis evolution.

Keywords: phylogenetic analysis, meta-trees, nodule bacteria, symbiotic genes

Soil bacteria of the *Rhizobiales* order are characterized by the ability to enter into symbiosis with leguminous plants, providing them with products for atmospheric nitrogen fixation, which is accompanied by the formation of nodules on the roots, which provide conditions for effective nitrogen fixation and assimilation of ammonium formed as a result of nitrogenase reaction. However, this group of organisms is of interest not only for microbiology, but also for evolutionary genetics of symbiosis. The analyzed model makes it possible to study the mechanisms of evolution of bacterial genes, which control interactions with different groups of eukaryotes [1].

The formation of nodules and symbiotic nitrogen fixation are complex multi-stage processes that are controlled by the genes of both partners. The symbiotic genes of rhizobia are divided into two main groups: virulence genes (*nod* genes), responsible for signal dialogue with the host plant, as a result of which nodules [2] are formed on its roots, and the genes necessary for symbiotic fixation of atmospheric nitrogen — the *nif* genes (control the synthesis and regulation of nitrogenase activity) and *fix* genes (a heterogeneous group of genes involved in the work of the nitrogenase complex, primarily in its supply with electrons and energy) [3].

In most nodule bacteria, *nod* and *nif/fix* genes are assembled into groups located in the genome as separate clusters, the structure of which varies in different species. For example, in nodule bacteria of alfalfa (*Sinorhizobium meliloti* and *S. medicae*), pea (*Rhizobium leguminosarum* bv. *viciae*), clover (*R. leguminosarum* bv. *trifolii*), goat's-rue (*Neorhizobium galegae*) and soybean (*Bradyrhizobium japonicum*), the main *nod* genes are closely linked, while the *nodABC* genes are organized into one operon (Fig. 1). In many rhizobia, the structural genes of *nifHDK* nitrogenase are one operon and are closely linked to the *nifEN* genes. Such organization of nitrogen fixation genes is characteristic for representatives of *Sinorhizobium*, *R. leguminosarum*, *N. galegae* and *Mesorhizobium loti*. However, in *B. japonicum*, an additional copy of the *nifH* gene is located outside the *nifHDK* operon, and the *nifA* gene is located outside the *nif/fix* cluster together with the *nod* genes.



**Fig. 1. 1. Location of symbiotic genes in various species of nodule bacteria:** ■ — *nod* genes, □ — *nif* genes, □ — *fix* genes, ▮ — *fdx* genes, □ — non-symbiotic genes; the *nodN* gene in the genome of *R. etli* (marked with an asterisk) has a chromosomal localization. The schemes are based on the data of genome sequencing of the studied rhizobial strains [4-11].

In evolutionarily advanced species of nodule bacteria (for example, *R. leguminosarum*), symbiotic genes are located on high-molecular plasmids [12]. At the same time, their chromosomal localization is characteristic for the ancestral form — *B. japonicum*, and also for *M. loti* [13-14]. There are rhizobia, in which only a few symbiotic genes can be located on the chromosome. This is typical, for example, for *R. etli*: most symbiotic genes are located on the plasmid, and the *nodN* gene is located on the chromosome [15] (see Fig. 1).

For a long time, the molecular mechanisms of the evolution of the legumes-rhizobia symbiosis, numbering 60-70 million years [16], remained unclear. Only as a result of the methods of high-performance sequencing, when the rapid accumulation of data on the composition and structure of the genomes of vari-

ous bacterial species began, the pangenome and metagenomic analysis managed to reconstruct the long path, which micro- and macrosymbionts passed from recruitment into bacterial cells of the nitrogen fixation genes (*nif* genes) to the development of modern highly specialized legumes-rhizobia symbiotic systems. This process included several stages, in which the groups of genes under consideration were obtained from different sources: *nif* genes and most *fix* genes were obtained by ancestral slow-growing rhizobia (*Bradyrhizobium*) from their free-living ancestors (*Rhodopseudomonas*) through vertical inheritance, while the genes for synthesizing lipo-chitooligosaccharide Nod factors responsible for the formation of nodules (*nod* genes) were obtained from fungi or from gram-positive bacteria during horizontal transfer [17]. The later rapidly growing rhizobia (*Rhizobium*, *Sinorhizobium*, *Neorhizobium*) obtained *nod* and *nif/fix* genes from ancestral rhizobia during horizontal gene transfer, which led to the formation of special plasmids (*pSym*) containing both groups of *sym* genes.

The connection of the phylogeny of *sym* genes with the mechanisms of their origin and evolution, as well as with localization in the rhizobia genomes, has not been adequately studied. For its study, we used a new method of phylogenetic analysis with the construction of “meta-trees” — combined dendrograms, in which the relative position of two trees serves as a measure of the congruence of phylogeny of individual genes. Previously, this method was developed and used to analyze alternative phylogeny obtained in the study of functionally different genes in various groups of eukaryotes (yeasts, fish), in which horizontal gene transfer is limited [18].

In our study, this method was applied to the phylogenetic analysis of two functionally different groups of genes (*nod* and *nif/fix*), which are of unequal origin and distributed among taxonomically unrelated groups of rhizobia through both vertical inheritance and horizontal gene transfer. The identified features of the phylogeny topologies constructed for the nodule formation genes and symbiotic nitrogen fixation make it possible to significantly supplement the earlier concept of the directions and mechanisms of the evolution of the gene symbiotic systems in nodule bacteria.

The purpose of the work was to use meta-trees to analyze the contribution of horizontal transfer to the process of evolutionary assembly of the symbiotic gene cluster in rhizobia.

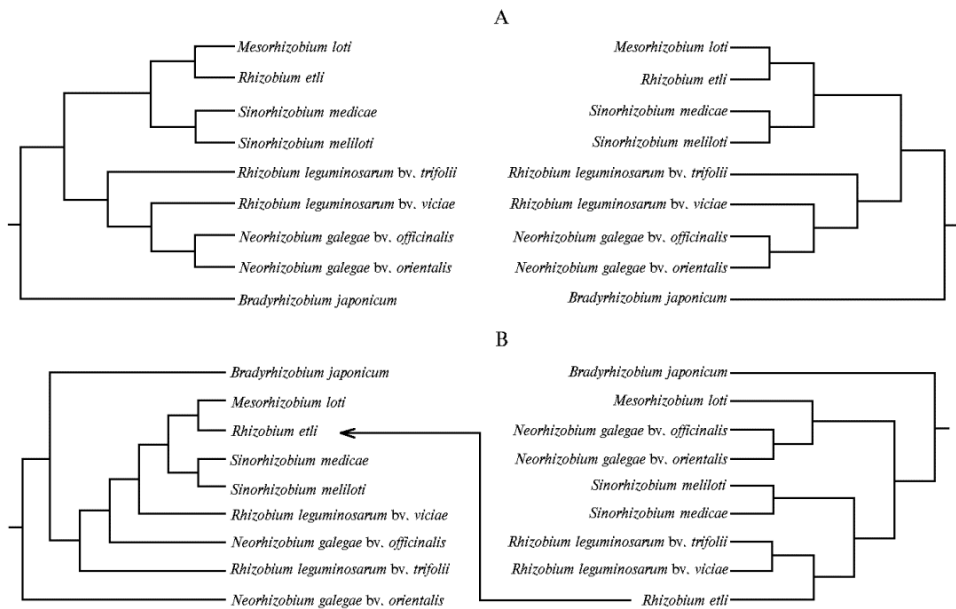
*Techniques.* Nucleotide sequences of symbiotic genes from the full genomic sequences of 9 representatives of nodule bacteria were used for the analysis: *R. leguminosarum* bv. *viciae* 3841 (GenBank GI:115259115) [4], *R. leguminosarum* bv. *trifolii* WSM2304 (GI:209537694) [5], *S. meliloti* 1021 (GI:25168258) [6], *S. medicae* WSM419 (GI:150031715) [7], *B. japonicum* USDA110 (GI:47118316) [8], *M. loti* MAFF303099 (GI:47118328) [9], *R. etli* CFN42 (GI:89213252) [10], as well as *N. galegae* bv. *officinalis* HA-MBI1141 (GI:659665307) and *N. galegae* bv. *orientalis* HAMB1540 (GI:659657635) [11].

Cluster analysis was performed in the computer program MEGA 5.1 (<http://www.megasoftware.net/>). The alignment of nucleotide sequences was performed using the ClustalW algorithm; the mathematical model when composing the tree for each gene under study was p-distance, the clustering mechanism — Neighbor-Joining. The numerical expression of similarities of phylogenetic tree topologies was calculated using statistical approaches (implemented using the program presented at <http://www.mas.ncl.ac.uk/~ntmwn/compare2trees/index.html>) [19]. To construct the meta-tree and analyze the reliability of the clustering, the dissimilarity factors were used, which were calculated as the difference of unit (100%, minuend) and the corresponding similarity factor (subtrahend).

*Results.* At the initial stage of the work, phylogenetic trees were con-

structured for each of the following 18 genes: virulence genes *nodA*, *nodB*, *nodC* (responsible for the synthesis of the core part of the Nod factor) [20–22], *nodD* (in the presence of multiple copies of *nodDI*, it is a constitutively expressed gene of the flavonoid-sensitive transcriptional activator of *nod* genes) [23], *nodI*, *nodJ* (genes of membrane transporters of the Nod factor) [24], *nodN* (encodes the dehydratase enzyme) [25]; genes of symbiotic nitrogen fixation *fixA*, *fixB*, *fixC* (components of the electron transport chain of nitrogenase) [26], *nifA* (gene of the transcriptional regulator of *nif* genes), *nifB*, *nifN*, *nifE* (responsible for the synthesis of the Fe-Mo cofactor of nitrogenase), *nifD*, *nifH*, *nifK* (structural components of nitrogenase) [27], *fdxN* (ferredoxin). The choice was due to the fact that homologs of these genes are present in all representatives of nodule bacteria selected for analysis. In addition, a traditional chromosomal taxonomic marker, the 16S rRNA gene, was included in the study due to the conservation of its nucleotide sequence and low frequency of horizontal transfer.

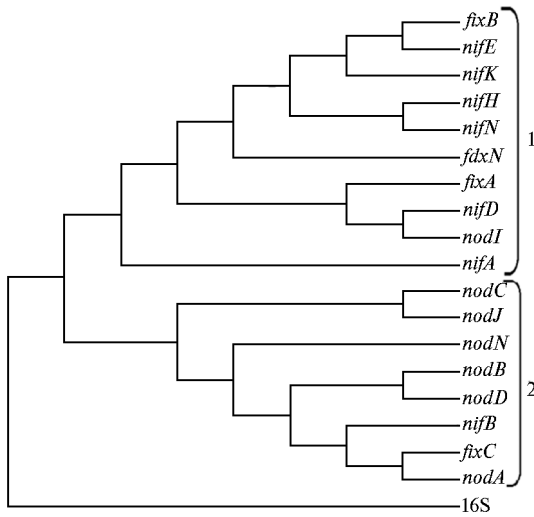
The methodology for comparing topologies is based on the idea that horizontal transfer of genes from one species of rhizobia to another during evolution introduces significant differences in the topology of the phylogeny of these genes. In the case when the topology of the trees of two genes does not differ, it is assumed that they either do not participate in the horizontal transfer at all, or are transferred together (Fig. 2, A).



**Fig. 2. Comparison of the topology of phylogenetic trees for the *nifH* (left) and *nifN* (right) genes with a similarity of 100 % (A) and for the genes *fdxN* (left) and 16S rRNA (right) with a similarity of 50.5 % (B). The arrow shows the most noticeable difference in topology: the transfer of *Rhizobium etli* from the *R. leguminosarum* cluster on the 16S rRNA gene tree to the *M. loti* cluster on the *fdxN* tree.**

At the same time, the discrepancy in the tree topology based on the analysis of the sequences of the same set of strains is a result of the horizontal transfer of symbiotic genes (see Fig. 2, B). According to these differences in the topology, we can assume how the horizontal gene transfer occurred during evolution. For example, it can be seen that in the tree for the 16S-rRNA gene in one cluster there are two biovars *R. leguminosarum* and *R. etli*. Since the 16S rRNA gene is a chromosomal marker, it can be said that clustered together representatives are evolutionarily similar. At the same time, on the tree of the plas-

mid marker *fdxN*, rhizobia *R. etli* grouped in the same cluster with *M. loti*, although *R. etli* and *M. Loti* belong to different families – *Rhizobiaceae* and *Phyllobacteriaceae*. This fact allows assuming that there was a horizontal transfer of the *fdxN* gene between the nodule bacteria *M. loti* and *R. etli*, or these species obtained this gene from a single source during the horizontal transfer.



**Fig. 3. The main clusters of the meta-tree according to the matrix of similarities of topologies in phylogenetic trees constructed on the basis of the data on the nucleotide polymorphism of symbiotic genes: 1 – *nif/fix* cluster, 2 – *nod* cluster.**

concentrated in cluster 1, the group of *nod* genes – in cluster 2, while the chromosomal marker (16S rRNA gene) did not enter any of the clusters (Table).

**Evaluation of the reliability of the meta-tree clustering in accordance with the average distance inside and between clusters/groups of genes using Student's *t*-test**

Comparison group	TD	SD	SE	Reliability of differences according to the Student's <i>t</i> -test at the significance level <i>p</i>
Inside the cluster 1	0.197	0.094	0.013	< 0.001
Inside the cluster 2	0.216	0.104	0.020	< 0.01
Between the clusters	0.269	0.107	0.012	
Inside the <i>nif/fix</i> group	0.194	0.084	0.011	< 0.001
Inside the <i>nod</i> group	0.201	0.082	0.018	< 0.01
Between the groups of symbiotic genes	0.270	0.099	0.011	

Note. TD – average difference of topologies, SD – standard deviation, SE – standard error. Comparison of the average distance between tree topologies within one cluster/group, as well as between the trees of one cluster/group and each tree of another cluster/group was performed. The structure of the clusters is shown in Fig. 3.

Such clustering may indicate that in the evolutionary history of the group of genes, *nif/fix* genes are not related to each other and to chromosomal genes. It is determined by their independent origin and separate horizontal transfer. Indeed, the difference between the groups of *nod* and *nif/fix* genes is expressed not less clearly than the dissimilarity between clusters 1 and 2. However, exceptions were found: the *nodI* gene, which is located in the *nif/fix* cluster 1 and is transferred along with the *nifD* gene, as well as the *nifB* and *fixC* genes that enter the *nod* cluster 2 together with *nodA*.

As noted above, in the genomes of many nodule bacteria, symbiotic genes are grouped into clusters. In this regard, the joint horizontal transfer of groups of *nod* and *nif/fix* genes may be due to the characteristics of their cluster arrangement in the genome (see Fig. 1). However, it is not possible to explain

Then we compared all the 19 phylogenetic trees obtained in order to identify differences in their topologies. The bioinformatic analysis of the obtained data made it possible to calculate the coefficients of pairwise similarity of topologies for all trees, ranging from 1 (complete congruence of two phylogenies) to 0 (complete absence of congruence). Then the constructed matrix of similarities was converted into a matrix of differences by subtracting the corresponding similarity factors from unit. According to this matrix, a meta-tree was constructed (Fig. 3).

It can be seen that on the meta-tree two clusters were reliably identified: the group of *nif/fix* genes was predominantly

the obtained picture only by the specificity of the arrangement of genes on genetic maps. So, in cluster 1, all the structural genes of nitrogenase were identified, with the genes *nifH*, *nifK*, *nifE* and *nifN* being grouped together. Such clustering of genes corresponds to their location in the genomes, since in all studied organisms the *nifHDK* genes constitute one operon, to which the genes *nifE* and *nifN* are adjacent. The only exception is the localization of the *nifN* gene outside the *nif/fix* zone in nodule bacteria of alfalfa (*Sinorhizobium meliloti*). Together with the same *nod* genes, *nodN* was detected in cluster 2, but it was located separately inside the cluster. It should be noted that the localization of *nodN* in the genome may be different. In some representatives, it is located in the same cluster with the *nodABC* genes (*S. meliloti*, *S. medicae*, *R. leguminosarum*); in others, it is located separately from the common *nod* genes. For example, in *R. ga-legae*, the *nodN* is located in another *nod* cluster, and in *B. japonicum*, *M. loti* and *R. etli* — outside the cluster of symbiotic genes. In addition, in *R. galegae*, the *nodN* gene has a chromosomal localization.

An interesting result of the analysis was the entry of the *nifB* and *fixC* genes into cluster 2 and their grouping together with the *nodA* gene. It is now known that the genes responsible for the synthesis of the core Nod factor (*nodABC*) are of a non-rhizobia origin. Thus, Hirch *et al.* [28] analyzed the sequences of bacterial and fungal genes encoding enzymes with catalytic centers similar to those of the *nodABC* gene products. It was suggested that the *nodB* and *nodC* genes could be obtained by rhizobia during the horizontal transfer from gram positive bacteria in which they participate in the biogenesis of the murine cell wall. In addition, these genes could be obtained by nodule bacteria from fungi actively synthesizing chitin-like substances. It is known that many fungi enter into close symbiotic relationships with bacteria, in case of which horizontal gene transfer is quite possible. However, for the *nodA* gene, homologous sequences in other organisms were not found. Thus, the *nodA* gene has a different origin than *nodB* and *nodC* [28]. In this regard, its clustering, together with the *nifB* and *fixC* genes, can indicate their initially simultaneous horizontal transfer from unidentified organisms.

It is important to note that the specifics of genes clustering on a meta-tree should be influenced not only by their location in modern forms of rhizobia, but also by evolutionary events that took place before, including independent rearrangements of genomes (see Fig. 1), what is more, both separate genes and genes containing parts of the genome could change their localization. The example of such rearrangements are the inversion of the *nif/fix* zone in *N. galegae* bv. *orientalis* [18] and the translocation of *nifHDKEN* genes in *R. leguminosarum* bv. *viciae* [11]. There is no doubt that genomic rearrangements have a significant influence on the order of clustering of genes, but the degree of their influence within the framework of the applied model is not clear enough. A significant role in the occurrence of differences in the topologies of the two genes is probably played by the unequal rate of their evolution.

The results obtained in our study do not agree with the data obtained by the methods of molecular phylogeny. Previously, it was shown that the polymorphism of *nod* genes was formed independently of the divergence of the core elements of the bacterial genome, but correlates with the taxonomy of hosts, whereas the polymorphism of the *nif* genes correlates with the divergence of the core part of the bacterial genome, but is not associated with the divergence of the host [29]. In our study, the meta-tree method showed the independence of the evolutionary history of both *nod* and *nif/fix* genes from the chromosomal marker (16S rRNA gene). Perhaps the method used by us is more sensitive to the influence of evolutionary factors that are not related to the horizontal gene

transfer and not taken into account in traditional phylogenetic approaches.

Thus, the results of phylogenetic analysis of *nod* and *nif/fix* genes, which we carried out using the method of constructing and analyzing meta-trees, confirm the previously identified differences in the evolutionary histories of these genes. The detection of two clusters, one of which consists primarily of *nif/fix* genes (cluster 1), and the other of *nod* genes (cluster 2), may indicate the fact that the compared groups of genes originated at different stages in the evolution of rhizobia (in connection with which they occupy different sites on chromosomes or plasmids of rhizobia), as well as the expression of these genes at different stages of symbiosis development (infection of root hairs and release into plant cytoplasm for *nod* genes, symbiotic fixation of nitrogen — for *nif/fix* genes). The presence of some *nod* genes in cluster 1 and some *nif/fix* genes in cluster 2 may be due to the fact that in the late stages of the evolution of rhizobia, the whole complex of *sym* genes was transferred as a whole in the composition of *Sym* plasmids or genomic islands. We cannot exclude the fact that the absence of correlation between the structures of clusters identified in the analysis of the meta-tree and the location of the studied genes on genetic maps is a consequence of intensive intragenomic rearrangements typical for many rhizobia and different rates of evolution of the primary structure in the studied genes.

Further analysis of the role of these evolutionary mechanisms requires a modification of the approach we used, in particular, of the separate analysis of the structures of the meta-trees constructed using samples of strains that occurred at early stages of rhizobia evolution (when horizontal gene transfer was limited due to the chromosomal localization of *sym* genes in *Bradyrhizobium*) or at later stages (when horizontal gene transfer was most intense and most likely captured the entire complex of *sym* genes located on the *Rhizobium* and *Sinorhizobium* plasmids). An important purpose of improving the method used is also to develop statistical and bioinformational criteria for supporting clusters that are detected on meta-trees, since in the first stage of the work we used only standard biometric approaches to solve this problem.

Thus, the phylogenetic analysis of *nod* and *nif/fix* genes, carried out using the construction and comparison of meta-trees, confirmed the previously identified differences in the evolutionary histories of these genes. The presence of two clusters in the meta-tree, one of which combines mainly *nif/fix* genes, and the other — *nod* genes, may be due to the fact that the stages of evolution of the rhizobia, at which the compared groups occurred, did not coincide, and also due to the expression of these genes at different stages of symbiosis development. The absence of correlation between the structure of clusters in the meta-tree and the results of genetic mapping may be due to intra-genomic rearrangements and the unequal rate of evolution of nucleotide sequences of genes. To improve the method used, it is necessary to develop statistical and bioinformational criteria for supporting clusters that are detected on meta-trees.

## REFERENCES

1. Provorov N., Tikhonovich I., Andronov E., Belimov A., Borisov A., Vorob'ev N., Dolgikh E., Zhernakov A., Zhukov V., Kimeklis A., Kopat' V., Kurchak O., Onishchuk O., Safronova V., Sulima A., Chizhevskaya E., Chirak E., Shtark O. *Geneticheskie osnovy evolyutsii bakterii — simbiotov rastenii* [Evolution of bacterial symbionts of plants — genetic aspects]. St. Petersburg, 2016 (in Russ.).
2. Douni Dzh. V knige: *Rhizobiaceae* [In: *Rhizobiaceae*]. St. Petersburg, 2002: 417-434 (in Russ.).
3. Kaminski P., Batut Zh., Boistard P. V knige: *Rhizobiaceae* [In: *Rhizobiaceae*]. St. Petersburg, 2002: 465-492 (in Russ.).
4. Young J., Crossman L., Johnston A. The genome of *Rhizobium leguminosarum* has

- recognizable core and accessory components. *Genome Biol.*, 2006, 7(4): R34 (doi: 10.1186/gb-2006-7-4-r34).
5. Reeve W., O'Hara G., Chain P. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM2304, an effective microsymbiont of the South American clover *Trifolium polymorphum*. *Stand Genomic Sci.*, 2010, 2(1): 66-76 (doi: 10.4056/sigs.44642).
  6. Barnett M., Fisher R., Jones T. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *PNAS USA*, 2001, 98(17): 9883-9888 (doi: 10.1073/pnas.161294798).
  7. Reeve W., Chain P., O'Hara G. Complete genome sequence of the *Medicago* microsymbiont *Ensifer (Sinorhizobium) medicae* strain WSM419. *Stand Genomic Sci.*, 2010, 2(1): 77-86 (doi: 10.4056/sigs.43526).
  8. Kaneko T., Nakamura Y., Sato S. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.*, 2002, 9(6): 189-197 (doi: 10.1093/dnares./9.6.189).
  9. Kaneko T., Nakamura Y., Sato S. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.*, 2000, 7(6): 331-338 (doi: 10.1093/dnares./7.6.331).
  10. Girard M., Flores M., Brom S. Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. *phaseoli*. *J. Bacteriol.*, 1991, 173(8): 2411-2419.
  11. Österman J., March J., Laine P.K. Genome sequencing of two *Neorhizobium galegae* strains reveals a *noeT* gene responsible for the unusual acetylation of the nodulation factors. *BMC Genomics*, 2014, 15: 500 (doi: 10.1186/1471-2164-15-500).
  12. Khaines M., Finan T. V knige: *Rhizobiaceae* [In: *Rhizobiaceae*]. St. Petersburg, 2002: 41-62 (in Russ.).
  13. Göttfert M., Röthlisberger S., Kündig C. Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. *J. Bacteriol.*, 2001, 183(4): 1405-1412 (doi: 10.1128/JB.183.4.1405-1412.2001).
  14. Sullivan J., Ronson C. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *PNAS USA*, 1998, 95(9): 5145-5149.
  15. González V., Santamaría R., Bustos P. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. *PNAS USA*, 2006, 103(10): 3834-3839 (doi: 10.1073/pnas.0508502103).
  16. Vorob'ev N., Provorov N. *Ekologicheskaya genetika*, 2013, 11: 73-85 (in Russ.).
  17. Provorov N., Andronov E. Evolution of root nodule bacteria: reconstruction of the speciation processes resulting from genomic rearrangements in a symbiotic system. *Microbiology*, 2016, 83(2): 131-139 (doi: 10.1134/S0026261716020156).
  18. Nye T. Trees of trees: an approach to comparing multiple alternative phylogenies. *Syst. Biol.*, 2008, 57(5): 785-794 (doi: 10.1080/10635150802424072).
  19. Nye T., Lio P., Gilks W. A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics*, 2005, 22(1): 117-119.
  20. Atkinson E., Palcic M., Hindsgaul O., Long S. Biosynthesis of *Rhizobium meliloti* lipooligosaccharide Nod-factors: NodA is required for an N-acyltransferase activity. *PNAS USA*, 1994, 91(18): 8418-8422.
  21. John M., Rohrig H., Schmidt J. *Rhizobium* NodB protein involved in nodulation signal synthesis is chitooligosaccharide deacetylase. *PNAS USA*, 1993, 90: 625-629.
  22. Debelle F., Rosenberg C., Denarie J. The *Rhizobium*, *Bradyrhizobium* and *Azorhizobium* NodC proteins are homologous to yeast chitin synthases. *Mol. Plant-Microbe Interact.*, 1992, 5: 443-446.
  23. Peck M., Fisher R., Long S. Diverse flavonoids stimulate NodD<sub>1</sub> binding to *nod* gene promoters in *Sinorhizobium meliloti*. *J. Bacteriol.*, 2006, 188(15): 5417-5427 (doi: 10.1128/JB.00376-06).
  24. Evans I., Downie J. The NodI product of *Rhizobium leguminosarum* is closely related to ATP-binding bacterial transport proteins: nucleotide sequence of *nodI* and *nodJ* genes. *Gene*, 1986, 43: 95-101.
  25. Baev N., Shultze M., Barlier I. *Rhizobium nodM* and *nodN* genes are common nod genes: *nodM* encodes functions for efficiency of *nod* signal production and bacteroid maturation. *J. Bacteriol.*, 1992, 174(23): 7555-7565.
  26. Putnoky P., Grosskopf E., Cam D. *Rhizobium fix*-genes mediate at least two communication steps in symbiotic nodule development. *J. Cell Biol.*, 1988, 106: 597-607.
  27. Arnold W., Rump A., Klipp W. Nucleotide sequence of 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumonia*. *J. Mol. Biol.*, 1988, 203: 715-738.
  28. Hirsch A., Lum M., Downie J. What makes the rhizobia-legume symbiosis so special? *Plant Physiol.*, 2001, 127: 1484-1492 (doi: 10.1104/pp.010866).
  29. Provorov N., Vorob'ev N. *Geneticheskie osnovy evolyutsii rastitel'no-mikrobnogo simbioza* [Genetic aspects of plant-microbe symbiosis evolution]. St. Petersburg, 2012 (in Russ.).