ISSN 1022-7954, Russian Journal of Genetics, 2017, Vol. 53, No. 4, pp. 411–419. © Pleiades Publishing, Inc., 2017.
Original Russian Text © N.A. Provorov, E.E. Andronov, O.P. Onishchuk, 2017, published in Genetika, 2017, Vol. 53, No. 4, pp. 401–410.

REVIEW
AND THEORETICAL ARTICLES

Forms of Natural Selection Controlling the Genomic Evolution in Nodule Bacteria

N. A. Provorov*, E. E. Andronov, and O. P. Onishchuk

All-Russia Research Institute for Agricultural Microbiology, St. Petersburg, 196608 Russia

**e-mail: provorovnik@yandex.ru*

Received March 24, 2016; in final form, October 3, 2016

Abstract—The role of different forms of natural selection in the evolution of genomes in root nodule bacteria (rhizobia) is analyzed for the first time. In these nitrogen-fixing symbionts of leguminous plants, two types of genome organization are revealed: (i) unitary type, where over 95% of genetic information is encoded by chromosomes (5.3–5.5 Mb in *Azorhizobium*, 7.0–7.8 Mb in *Mesorhizobium*, 7.3–10.1 Mb in *Bradyrhizobium*); (ii) multipartite type, where up to 50% of genetic information is allocated to plasmids or chromids which may exceed 2 Mb in size and usually control the symbiotic properties (p*Sym*s) in fast-growing rhizobia (*Rhizobium*, *Sinorhizobium*, *Neorhizobium*). Emergence of fast-growing species with narrow host ranges are correlated to the extension of extrachromosomal parts of genomes, including the increase in p*Sym*s sizes (in *Sinorhizobium*). An important role in this evolution is implemented by diversifying selection since the genomic diversity evolved in rhizobia owing to symbiotic interactions with highly divergent legumes. However, analysis of polymorphism in *nod* genes (encoding synthesis of lipo-chitooligosaccharide signaling Nod factors) suggests that the impacts of diversifying selection are restricted to the bacterial divergence for host specificity and do not influence the overall genome organization. Since the extension of rhizobia genome diversity results from the horizontal *sym* gene transfer occurring with low frequencies, we suggest that this extension is due to the frequency-dependent selection anchoring the rare genotypes in bacterial populations. It is implemented during the rhizobia competition for nodulation encoded by the functionally diverse *cmp* genes. Their location in different parts of bacterial genomes may be considered as an important factor of their adaptive diversification implemented in the host-associated microbial communities.

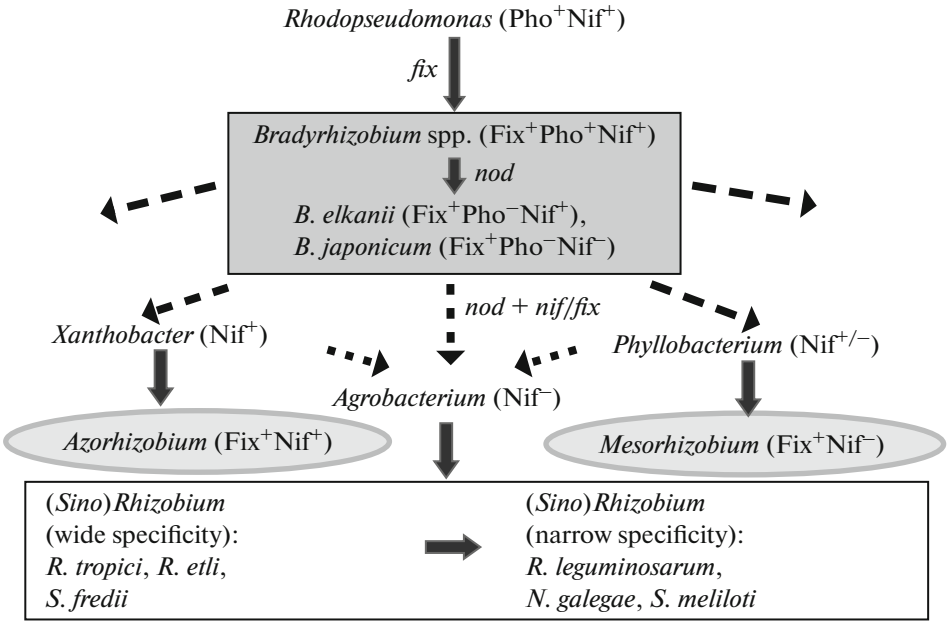
Keywords: nodule bacteria (rhizobia), symbiotic nitrogen fixation, host specificity, genome evolution, unitary and multipartite genomes, plasmids and chromids, frequency-dependent and disruptive selection, competitiveness, theory of symbiogenesis

DOI: 10.1134/S1022795417040123

INTRODUCTION

Symbiotic bacteria are a convenient model for analyzing the driving forces and mechanisms of evolution of prokaryotes, the genomes of which undergo rapid changes during interactions with eukaryotes. The best-developed model is represented by root nodule bacteria (rhizobia), which, despite their deep specialization for nitrogen-fixing symbiosis with legumes, retain a capability of autonomous existence in soil [1]. Comparative genetic analysis of rhizobia demonstrated that this polyphyletic group (at least ten families of α- and β-proteobacteria [2]) underwent a complex evolutionary route from free-living soil diazotrophs to symbiotic specialized nitrogen-fixing organisms [3]. By describing the genomic evolution of rhizobia [3], we demonstrated that it is of combinative type (recombination is the main source of genetic material, including intragenomic reorganizations and horizontal transfer of genes controlling symbiosis, the *sym* genes). Evolution of rhizobia is divergent and leads to speciation of a wide variety of genera, species, and biotypes adapted to diverse ecological niches, which are provided to bacteria by host plants and soil environment.

Despite the fact that the genomes of rhizobia are thoroughly studied, the question of how their complication (macroevolution) is connected to selective pressure in bacteria circulating in the "plant–soil" system remains open. It is logical to assume that disruptive selection, which is associated with divergence of legumes and facilitates coevolution of gene systems in microsymbionts and hosts, is a main factor of evolution of the genomes of rhizobia. In the present study, we will demonstrate that, despite the fact that coevolution of partners occurs very intensely, it cannot be assumed as a main factor of genomic diversification of rhizobia. We will describe an alternative scenario for the origin of their diversity associated with frequency-dependent selection, which acts during competition for root nodule formation and causes fixation of rare



Evolutionary relationships between major groups of rhizobia and related α-proteobacteria. "Primary" rhizobia (*Bradyrhizobium*), which evolved from the free-living diazotrophs (forms related to *Rhodopseudomonas*), autotrophic for carbon (phototrophic fixation of Pho^+) and nitrogen (diazotrophic fixation of Nif^+). Emergence of ability to fix nitrogen *in planta* (Fix^+) in these rhizobia was associated with transfer of *fix* genes from photosynthetic into symbiotic nitrogen-fixing system. The loss of phototrophy in *Bradyrhizobium* spp. was accompanied by the evolution of nodulation genes (*nod*); because of this, rhizobia gained the possibility to actively utilize the products of plant photosynthesis. First level "secondary" rhizobia that emerged during the transfer of *sym* genes (*nif*, *fix*, and *nod* operons), which is marked by dotted arrows, into free-living (*Xanthobacter*) or epiphyte (*Phyllobacterium*) bacteria capable of fixing nitrogen *ex planta* are listed in ovals. The emerging symbionts either maintained this ability (*Azorhizobium*) or lost it (*Mesorhizobium*). Second level "secondary" rhizobia that emerged via transfer of *sym* genes into oncogenic bacteria incapable of nitrogen fixation (related to *Agrobacterium*) are listed in the white rectangle. The presence of plastic multipartite genomes in these bacteria facilitated deep specialization for symbiosis (formation of discrete groups of cross-inoculation) and increasing nitrogen-fixing activity (on the basis of differentiation of bacteria into nonviable intracellular bacteroids).

recombinant genotypes in rhizobia populations. Analysis of selection types specific for symbiosis allowed us to determine ecological and genetic factors of genomic diversification in symbiotic nitrogen-fixing organisms, as well as to approach the analysis of the ratio of adaptive and progressive evolution of symbiosis, including transformations of bacteria into cell organelles.

DYNAMICS OF BACTERIAL GENOME IN A SYMBIOTIC SYSTEM

Comparative analysis of genomic reorganization of rhizobia made it possible to determine the main stages of their evolutionary route (Table 1, figure) [3].

(1) Emergence of slow-growing "primary" (P) rhizobia (*Bradyrhizobium* spp.) from free-living diazotrophs related to *Rhodopseudomonas*. It was based on transition of *fix* genes from the photosynthetic system to control over functioning of nitrogenase encoded by *nif* genes. Ancestral forms of P rhizobia maintained a capability of photo- and diazotrophy, as well as a unitary genome, albeit enlarged (from 5000–5500 kb in *Rhodopseudomonas* to 7300–10100 kb in *Bradyrhizobium*).

(2) Formation of signaling interaction between P rhizobia and plants, which is determined by *nod* genes (from *nodulation*, nodule formation), that encode the synthesis of lipo-chitooligosaccharide Nod factors. The emergence of this synthesis was based on reorganizations of the rhizobia's own genomes and horizontal gene transfer (HGT), which allowed certain genes of chitin-like metabolite synthesis to be acquired by rhizobia from fungi or gram-positive bacteria [4]. Formation of a system of *sym* genes encompassing *nod*, *nif*, and *fix* operons was the result of evolution of P rhizobia. This led to increased genomic plasticity in P rhizobia occurring at both the level of intragenomic rearrangements and HGT, as well as mobility of *sym* genes in populations circulating within the "plant–soil" system.

(3) Transfer of *sym* genes from P rhizobia into unrelated free-living (soil, epiphyte) nitrogen-fixing organisms (e.g., cultures related to *Xanthobacter* and *Phyllobacterium*). These processes led to emergence of first level "secondary" rhizobia (S1), *Azorhizobium* and *Mesorhizobium*, which have unitary genomes.

(4) Transfer of *sym* genes into phytopathogenic bacteria related to *Agrobacterium* that are unable to fix nitrogen led to speciation of *Rhizobium*, *Sinorhizo-*

Table 1. Main stages of rhizobia evolution associated with genomic rearrangements

Stages of evolution	Changes of genome revealed on the level of		Contribution of horizontal gene transfer to genomic evolution	Key traits of emerging forms		
	its general organization	individual genes (gene systems)		photosynthesis	nitrogen-fixing *ex planta*	host specificity*
Emergence of "primary" (P) rhizobia (*Bradyrhizobium* spp.) from free-living diazotrophs (*Rhodopseudomonas*)	Size increases to 7300–10100 kb while maintaining unitary structure	Emergence of a system of *fix* genes, which determine energy supply of nitrogenase *in planta*	Unstudied	+	+	Narrow (formation of stem nodules in aquatic legumes *Aeschynomene*, *Neptunia*, and *Sesbania*)
Species diversification of P rhizobia	Increase in genomic plasticity, transfer of *sym* genes into mobile chromosomal islands	Emergence of a system of *nod* genes, which control the synthesis of lipo-chito-oligosaccharide Nod factors	Acquiring the genes of chitin-like metabolite synthesis (from fungi or gram-positive bacteria)	–	+/–	Wide (formation of root nodules in taxonomically diverse legumes)
Emergence of first level "secondary" rhizobia (S1) (*Azorhizobium*, *Mesorhizobium*) from soil (*Xanthobacter*) or epiphyte diazotrophs (*Phyllobacterium*)	Increase in the size of genome, emergence of *Sym* plasmids (in certain cultures of *Mesorhizobium*)	Decrease of the number of *nif* genes (to 12–13), functionally replaced by *fix* genes (in *Mesorhizobium*)	Migration of symbiotic islands in soil populations (in *Mesorhizobium*)	–	+/–	Same
Emergence of second level "secondary" rhizobia (S2) (*Rhizobium*, *Sinorhizobium*, *Neorhizobium*) from phytopathogens unable to fix nitrogen (*Agrobacterium*)	Formation of multipartite structure (up to 50% of genome, including *sym* genes reassigned into large *Sym* plasmids or chromids)	Decrease in the number of *nif* genes to 7–8, acquiring *nodEF* genes, which control the synthesis of polyunsaturated acyl residues included in Nod factors	Active circulation of plasmid *sym* genes in bacterial populations polymorphic in ecotypes	–	–	Wide (symbiosis of *R. tropici*, *R. etli*, and *S. fredii* with tropical legumes) or narrow (symbiosis of *R. leguminosarum*, *N. galegae*, and *S. meliloti* with legumes of temperate latitudes)

* Rhizobia with wide host specificity form symbiosis with legumes from various tribes or subfamilies (certain strains of *Bradyrhizobium* and *S. fredii* also form nodules with non-legume plant *Parasponia andersonii*, family Cannabaceae); under a narrow specificity, the symbiosis is restricted to legumes from a single tribe or genus.

Table 2. Variation of 24 strains of clover nodule bacteria (*Rhizobium leguminosarum* bv. *trifolii*) in plasmid composition [8]

Incompatibility group	Number of strains containing a plasmid of the group	Size variation, kb	Number of strains containing p*Sym* from the group (sizes of p*Sym*, kb)
a	24	150–500	15 (260–500)
b	24	190–640	8 (270–420)
c	24	350–840	1 (350)
d	23	510–1250	0
e	9	610–1350	0
f	1	1060	0

bium, and *Neorhizobium*, which have multipartite genomes. These second level "secondary" rhizobia (S2) contain large plasmids and chromids, the sizes of which can exceed 2000 kb, comprising up to 50% of the genome.

An important trait of S2 rhizobia is localization of *sym* genes on one of the plasmids (p*Sym*) or chromids. In rhizobia of vicia and clover (*R. leguminosarum* bv. *viciae* and bv. *trifolii*), sizes of p*Sym* vary within a range of 200–550 kb, and they exceed 1100 kb in rhizobia of medick (*S. meliloti*). Analysis of population polymorphism of rhizobia showed p*Sym* to be more uniform in size than "nonsymbiotic" plasmids. For example, in bean rhizobia (*R. leguminosarum* bv. *phaseoli*), the molecular mass of p*Sym* from the soils of North Dakota was 190–260 MDa (variation coefficient C_v = 8.5%), while "nonsymbiotic" plasmids had a mass of 65–700 MDa (C_v = 66.5%) [5]. Sizes of p*Sym* in medick rhizobia (*S. meliloti*) vary within a range of 1100–1680 kb, while other plasmids are much more variable (10–1900 kb [6, 7]).

Similar tendencies of p*Sym* variation are typical for clover rhizobia (*R. leguminosarum* bv. *trifolii*), in which these plasmids have the sizes of 260–500 kb, while *sym* genes could not be determined on megaplasmids more than 1000 kb in size typical of the majority of strains ([8]; Table 2). Apparently, the structures of p*Sym* in S2 rhizobia are controlled by selection more strictly than the structures of "nonsymbiotic" plasmids owing to adaptations to *in planta* niches being a key factor of maintaining these bacteria within ecosystems.

An important tendency of rhizobia evolution is an increase in the density of localization of *sym* genes within the genome, which facilitates their mobility and, thus, high rate of evolution. In P rhizobia (*Bradyrhizobium*), *sym* genes are usually located in several unlinked loci of a chromosome [9]; in S1 rhizobia (*Mesorhizobium*), these genes are located in genomic islands that are able to actively spread within the soil population despite their large size (which can exceed 600 kb) [10]; in S2 rhizobia (*Rhizobium*, *Sinorhizobium*), *sym* genes concentrate in relatively small (<100 kb) clusters on p*Sym* [11].

How are the transformations of rhizobia genomes related to their speciation for symbiosis? It is logical to assume that complexity introduced to the genomes was a result of transfer of *sym* genes into extrachromosomal elements, which determined specific traits of regulation of these genes compared to the autonomous genes. However, this explanation cannot be assumed to be the only one, because complexity introduced in genomes (its division into chromosome, plasmids, and chromids) of the rhizosphere and endophyte nitrogen fixing *Azospirillum* (which, like most rhizobia, are α-Proteobacteria), determined during a comparison with nonsymbiotic *Rhodospirillum* bacteria, is not connected with transition of genes essential for interactions with plants into the extrachromosomal state [12].

The fact that diversification of genus *Bradyrhizobium* (it includes at least 15 species combined into two groups represented by the species *B. japonicum* and *B. elkanii*) occurred with retention of wide and overlapping ranges of host plants is also evidence of the absence of direct connection between the changes in host specificity and structure of rhizobia genomes [2]. A similar tendency is also typical of sister species with narrow specification, *S. meliloti* and *S. medicae*, which are the symbionts of medick (*Medicago*), melilot (*Melilotus*), and fenugreek (*Trigonella*). Comparison of these species with *S. fredii*, which has a wide specification, revealed that both the number of plasmids and the size of p*Sym* increased during the narrowing of the host specificity in *Sinorhizobium*. Transition to a narrower host specificity in *Rhizobium* was accompanied by an increase in the genome size and the number of plasmids; however, the sizes of p*Sym* remained the same [13]. In both genera of rhizobia, genomes of the forms with narrow specification were more homogeneous in size than the genomes of forms with wide specification (Table 3).

DISRUPTIVE SELECTION AND DIVERGENT EVOLUTION

An important factor of rhizobia evolution is disruptive selection, which is determined by the following: availability of a wide range of host plants for these bac-

Table 3. Genomic features of fast-growing rhizobia with contrasting differences in host specificity

Bacterial genus	Specificity*	Species	Genomic characteristics		
			size, kb	number of plasmids	size of p*Sym*, kb**
Rhizobium	Wide	*R. etli, R. tropici*	5034–7080	2–6	300–550
	Narrow	*R. leguminosarum* (bv. *trifolii,* bv. *viciae*)	6873–7751	4–11	200–500
Sinorhizobium	Wide	*S. fredii*	6476–7220	1–5	500–600
	Narrow	*S. meliloti, S. medicae*	6692–6818	2–8	>1100

* Under a wide host specificity, the symbiosis forms with legumes from various tribes and subfamilies; under a narrow specificity, with legumes from a single tribe or genus.

** *Sym* plasmids (p*Sym*) contain genes controlling the main symbiotic functions: nitrogenase synthesis (*nif*); its supply with electrons and equivalent reductants, as well as oxygenic regulation of *nif* genes (*fix*); synthesis of lipo-chito-oligosaccharide Nod factors (*nod*).

teria, differing in specificity of interactions with microsymbionts, and heterogeneity of the soil medium, in which bacteria exist between symbiotic cycles. By performing a metagenomic analysis of the variability of natural *R. leguminosarum* populations based on the *nod* genes [14, 15], we demonstrated that, at the level of metapopulation comprising the *viciae* and *trifolii* biotypes, disruptive selection causes a divergence of subpopulations specialized for legumes from the cross-inoculation groups of vicia and clover. In contrast to the higher organisms, these rhizobia undergo divergent evolution without genetic isolation: intense transfer of *sym* genes occurs between *viciae* and *trifolii* biotypes, which interact with vicia and clover plants growing together [16]. Apparently, the disruptive selection is confined to the *nod* genes within this system and cannot be considered a main speciation factor because of the diverged forms stably maintaining the biotype status. Divergent evolution of rhizobia for host specificity occurs either at the species level (*R. leguminosarum*: *viciae* and *trifolii* biotypes; *N. galegae*: *officinalis* and *orientalis* biotypes; *R. etli*: *phaseoli* and *mimosae* biotypes) or during divergence of sister species (*S. meliloti, S. medicae*) and does not affect the genomic organization.

Disruptive selection in populations of pathogens was previously demonstrated to occur when rare forms (against which the hosts lack the resistance genes) gain selective advantage, thus creating a frequency-dependent selection in microbial populations [17, 18]. It is logical to assume that, in rhizobia, in which early stages of interaction with legumes correspond to the "gene-to-gene" scheme [19], divergence occurs with a frequency-dependent selection that will be discussed in the next section.

An important aim of the evolutionary genetics of symbiosis is an analysis of connection between direction of selection, determined during the studies of symbiont dynamics, and evolution of the primary structure of their genes, evaluated by dN/dS statistics (ratio of nonsynonymous to synonymous substitutions). Such statistics make it possible to determine the effects of directional or stabilizing selection, which are evident from the prevalence of either dN or dS over the threshold values consistent with the hypothesis of neutral evolution for these genes [20].

During the studies of phytopathogenic bacteria and fungi that interact with their hosts via a "gene-to-gene" scheme, it was previously demonstrated that selection that determines the adaptation of microsymbionts to new hosts during the early stages of divergence (acquiring an ability to infect new hosts) acts in a directional form, while at the late stages (increasing the efficiency of reproduction on new hosts) it acts in a stabilizing form [21, 22]. On the basis of the fact that disruptive selection is a form of directional selection, we can assume that numerous sites with increased dN values (which can appear during the emergence of symbiosis with new host plants) exist in the *nod* genes. At the same time, dS prevailing over dN (stabilizing selection) or their equality (neutral evolution) can be expected in *nif* and *fix* genes because of increased nitrogen-fixing activity being an adaptation of the symbionts to acquired host plants. Since in transfer from wide to narrow host specificity deeper genomic transformations are revealed in *Sinorhizobium* than in *Rhizobium* (Table 3), it is logical to assume that the ratios of directional and stabilizing forms of selection in the *sym* genes of bacteria belonging to these genera will be different.

FREQUENCY-DEPENDENT SELECTION AND COMPETITION FOR NODULE FORMATION

Analysis of genomic dynamics of rhizobia demonstrated that recombination, including intragenomic rearrangements and HGT, which normally occur with low frequencies and are not registered under laboratory conditions, play a key role in the evolution of rhizobia [23]. Involvement of rare recombinants in symbiotic evolution may be determined by frequency-dependent selection (FDS), which occurs in microbial populations associated with plants. The first indica-

Table 4. Diversity in organization of unitary and multipartite genomes of rhizobia

Genome types	Bacterial genera (types of symbionts)*	Size of chromosomes, $\times 10^3$ bp	Extrachromosomal replicons (plasmids, chromids**)	
			number	size, kb
Unitary	*Bradyrhizobium* (P)	7.3–10.1	0–4	80–230
	Azorhizobium (S1)	5.3–5.5	0–1	50–200
	Mesorhizobium (S1)	7.0–7.8	0–2	200–400
Multipartite	*Rhizobium* (S2)	4.3–5.1	2–11	200–1350
	Sinorhizobium (S2)	3.6–6.5	2–8	20–2100
	Neorhizobium (S2)	4.6–4.7	1–2	175–1810

* P, primary symbionts that emerged by direct filiations of free-living diazotrophs into symbiotic nitrogen fixing organisms; S1, first level secondary symbionts that emerged by transfer of the system of *sym* genes formed in P rhizobia into soil or epiphyte bacteria capable of nitrogen fixing and related to *Xanthobacter* or *Phyllobacterium*; S2, second level secondary symbionts that emerged by transfer of the system of *sym* genes into phytopathogens related to *Agrobacterium* incapable of nitrogen fixing (figure).
** Chromids containing rRNA and tRNA genes found in goat's rue rhizobia, *Neorhizobium galegae* [41].

tion of its effects was demonstrated during mathematical modeling of competition for infected plants based on empirically determined nonlinear association between the populations of strains in inoculum and nodules [24]. Computer experiments showed that FDS can facilitate fixation of the genotypes occurring with extremely low frequencies (less than 10^{-19}) in rhizobia populations.

Involvement of FDS in the evolution of rhizobia appears to be plausible owing to this type of selection playing an important role in phytopathogenic interactions, which are controlled by "gene-to-gene" systems [25]. In legume–rhizobia symbiosis, these systems were found at the early stages of nodule development that are based on signaling interactions of partners and are highly similar to the systems of plant and phytopathogenic recognition [19].

The new version of the FDS model [26] was based on the hypothesis on migration of rhizobia from the root zone into endosymbiotic (nodule) niches, during which rapid rearrangements of bacterial populations occur, being regulated by the quorum sensing (QS) mechanism that controls bacterial reproduction. Analysis of the QS model (level of migratory activity of strains, FDS pressure) determined on the basis of experimental data demonstrated that genetically modified rhizobia strains with high competitiveness can remain within a population even under a rapidly decreased survivability outside the plant [26].

The effects of FDS were also registered in our studies within the "*R. leguminosarum*–vicia and clover" system [14, 15], where indices of population diversity, which characterize homogeneity of the haplotype distribution among operational taxonomic units, increase during interactions of forms possessing contrasting specificity (bv. *viciae* and bv. *trifolii*) to various plant species. The modeling of the evolution of symbiosis conducted previously [27] demonstrated that such change in the structure of a population can be a result of negative FDS, which facilitates reproduction of rare genotypes of rhizobia in the nodules. Thus, FDS can be viewed as an important factor of diversification of rhizobia because this type of selection facilitates fixation of newly emerging forms, including those possessing an altered genomic organization, in populations.

Competition for nodulation, during which FDS is put into effect, is a subject of intensive genetic research as a key adaptive trait, crucial for practical application of rhizobia. Studies of numerous mutants with decreased competitiveness [28, 29] showed that this trait is controlled by a large number of functionally heterogenic *cmp* genes (from *competitiveness*), localized in different part of the genome of rhizobia. For example, analysis of 378 mutants of medick (*S. meliloti*) rhizobia obtained using a modern modification of the signature-tagged mutagenesis method made it possible to identify over 30 *cmp* genes that determine competitiveness [30]. These genes participate in such cellular functions as signaling (*feuQ*), transport of phosphates (*pstA*, *pstC*), amino acids (*livM*), and heme (*ccmC*), regulation of transcription (*lexA*), reparation or destruction of damaged proteins (*lepA*, *clpA*), formation of polysaccharide capsule (*rkpU*), and synthesis of amino acids and phytohormones (*trpC*, *trpF*, *ilvI*, *ilvD2*, *metA*, and *thiC*). At the same time, no alterations in cultural biochemical and symbiotic traits were found in certain mutants with decreased competitiveness [30, 31], which indicates the presence of genes with competitiveness control as a single function in rhizobia.

It is important to note that the majority of *cmp* genes found in *Rhizobium* and *Sinorhizobium* are localized on chromosomes and are not functionally associated with *sym* genes localized on the plasmids. Therefore, the evolution of these rhizobia, determined by

competition for nodule niches, encompasses not only *nod*, *nif*, and *fix* genes located on p*Sym* and directly participating in the functioning of nodules but also *cmp* genes localized in other region of the genome, which is a factor of its diversification in the system of symbiosis.

CONCLUSIONS

Rhizobia are a unique model for the study of symbiogenic evolution, which is based on integration of forms on different levels of cellular organization. The classic version of the theory of symbiogenesis [32, 33] explains the emergence of eukaryotic cellular organelles (plastids, mitochondria, and their derivatives) from free-living bacteria. However, this theory is usually restricted to late stages of transition of bacteria into organelles that are not suitable for an experimental approach.

How did the early stages of symbiogenesis go? What were the driving forces of this evolution? Numerous data indicate that natural selection was a key factor of symbiogenesis [34], but the mechanisms of its effects in systems of genetic integration of unrelated organisms remain unclear. It is evident that the models of individual (Darwinian) selection are insufficient to describe all processes of symbiogenesis owing to the fact that this selection cannot facilitate the evolution revealed in the symbiotic systems, which is directed at the following: loss of viability (reproductive activity) of the microsymbionts, which in certain cases leads to the altruistic traits forming in microsymbionts for their hosts [35, 36]; deep reduction (including complete elimination) of microbial genome and transfer of significant portions of it to the host [37].

Facultative and ecologically obligate symbioses in which bacteria maintained the capability for autonomous existence provide great opportunities for analysis of driving forces of evolution of the bacterial genome. In rhizobia, this evolution is associated with transition of the genome from unitary to multipartite type (Table 1). Unitary genomes are typical of primary (P) rhizobia that emerged directly from free-living nitrogen-fixing organisms, as well as certain secondary (S) rhizobia that appeared via horizontal transfer of *sym* genes into soil and epiphyte bacteria capable of fixing nitrogen [3]. Multipartite genomes are typical of S rhizobia that emerged via transfer of *sym* genes into phytopathogens related to *Agrobacterium* unable to fix nitrogen, but possessing multipartite genomes. Substantial diversity of genomic architecture, which is modified in S2 rhizobia during transition from wide to narrow symbiotic specification (Table 3), was found in every group of rhizobia (Table 4).

At the same time, the nature of the relationship between complexity introduced in bacterial genomes and evolution of their symbiotic functions controlled by natural selection remains unstudied. Darwinian selection can be considered the main factor of evolution of rhizobia only at its initial stages (emergence of P rhizobia), when individual cells that fix the largest amounts of nitrogen gain more carbon from the plants [38].

We demonstrated that the *nod* genes, which control nodule development, undergo evolution under disruptive selection induced by the host [14, 15]. This type of selection facilitated divergent evolution of rhizobia for host specificity, the peak of which was formation of discrete groups of cross-inoculation. Together with the group selection described in published sources [1, 36], which determined the increase in intensity of nitrogen fixing (controlled by *nif* and *fix* genes), disruptive selection could have facilitated the adaptive evolution of plant–microbe symbiosis directed at the increase in its efficiency (effects on ecological adaptations of the partners).

However, the models of disruptive and group selection are insufficient for explaining transformations of general genomic structure that underlie the macroevolution of symbiotic bacteria. This question can be at least partially solved using the models of frequency-dependent selection, which make it possible to explain the processes of introducing complexity into the genome by involving it in evolution of rarely occurring recombinants.

It is logical to assume that described patterns in the evolution of rhizobia were typical of the early stages of formation of cellular organelles as well. However, the cause of inability to find permanent nitrogen-fixing organelles in eukaryotes remains unclear, although the majority of plants and animals experience strong nitrogen deficiency. Moreover, legumes and certain other flowering plants (*Gunnera*) form intracellular symbiosomes, which are classified as analogs of permanent eukaryotic organelles according to the values of certain indices, in the symbiosis with nitrogen-fixing bacteria [39].

Is it possible to construct plants that permanently maintain nitrogen-fixing organelles? The results of the studies of genomic evolution of the nitrogen-fixing organisms with the best specialization for symbiosis indicate such possibility. Segregation of genes controlling symbiotic and autonomous stages of the life cycle in various genomic structures can indeed be viewed as a prerequisite for genome reduction, which leads to the loss of autonomous existence in bacteria and precedes their transformation into organelles. Cyanobacteria *Nostoc azollae*, which loses a significant portion of its genome [40] owing to the reduction of autonomous phase of the life cycle and formation of vertical inheritance of cyanobionts during the reproduction of the host, water fern *Azolla filiculoides*, illustrates the initial steps of such transformation.

Thus, the studies of genetically specialized symbiotic nitrogen-fixing organisms (rhizobia, cyanobacteria) demonstrated that these organisms are at the early stages of transformation into organelles. Owing to this

fact, the prospects of genetic construction of plants containing such organelles is deemed possible.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation, project no. 14-26-00094 ("Studies of Genomic Dynamics and Disruptive Selection") and no. 14-16-00146 ("Study of Frequency-Dependent Selection").

REFERENCES

1. Provorov, N.A. and Vorob'ev, N.I., *Geneticheskie osnovy evolyutsii rastitel'no-mikrobnogo simbioza* (Evolutionary Genetics of Plant–Microbe Symbiosis), Tikhonovich, I.A., Ed., St. Petersburg: Inform-Navigator, 2012.
2. Berrada, H. and Fikri-Benbrahim, K., Taxonomy of the rhizobia: current perspectives, *Br. Microbiol. Res. J.*, 2014, vol. 4, no. 6, pp. 616–639.
3. Provorov, N.A. and Andronov, E.E., Evolution of root nodule bacteria: reconstruction of the speciation processes resulting from genomic rearrangements in a symbiotic system, *Microbiology* (Moscow), 2016, vol. 85, no. 2, pp. 132–139. doi 10.1134/S0026261716020156
4. Hirsch, A.M., Lum, M.R., and Downie, J.A., What makes the rhizobia–legume symbiosis so special?, *Plant Physiol.*, 2001, vol. 127, pp. 1484–1492.
5. Johnson, M.L. and Berryhill, D.L., Plasmid diversity in North Dakota strains of *Rhizobium phaseoli*, the symbiotic dinitrogen-fixing bacterium of beans, *Proc. North Dakota Acad. Sci.*, 1986, vol. 40, p. 15.
6. Guo, H., Sun, S., Eardly, D., et al., Genome variation in the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti, Genome*, 2009, vol. 52, no. 10, pp. 862–875.
7. Guo, X., Flores, M., Mavingui, P., et al., Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures, *Genome Res.*, 2003, vol. 13, pp. 1810–1817.
8. Mazur, A., Stasiak, G., Wielbo, J., et al., Intragenomic diversity of *Rhizobium leguminosarum* bv. *trifolii* clover nodule isolates, *BMC Microbiol.*, 2011, vol. 11.123. doi 10.1186/1471-2180-11-123
9. Okazaki, S., Noisangiam, R., Okubo, T., et al., Genome analysis of a novel *Bradyrhizobium* sp. DOA9 carrying a symbiotic plasmid, *PLoS One*, 2015, vol. 10, no. 2. e0117392
10. Sullivan, J.T. and Ronson, C.W., Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into phe-tRNA gene, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, vol. 95, pp. 5145–5149.
11. Young, J.P.W., Crossman, L.C., Johnston, A.W.B., et al., The genome of *Rhizobium leguminosarum* has recognizable core and accessory components, *Genome Biol.*, 2006, vol. 7, no. 4, p. R34.
12. Wisniewski-Dyé, F., Lozano, L., Acosta-Cruz, E., et al., Genome sequence of *Azospirillum brasilense* CBG497 and comparative analyses of *Azospirillum* core and accessory genomes provide insight into niche adaptation, *Genes*, 2012, vol. 3, no. 4, pp. 576–602.
13. Black, M., Moolhuijzen, P., Chapman, B., et al., The genetics of symbiotic nitrogen fixation: comparative genomics of 14 rhizobia strains by resolution of protein clusters, *Genes*, 2012, vol. 3, no. 2, pp. 138–166.
14. Andronov, E.E., Onishchuk, O.P., Kurchak, O.N., and Provorov, N.A., Population structure of the clover rhizobia *Rhizobium leguminosarum* bv. *trifolii* upon transition from soil into the nodular niche, *Microbiology* (Moscow), 2014, vol. 83, no. 4, pp. 422–429. doi 10.1134/S0026261714030035
15. Andronov, E.E., Igolkina, A.A., Kimeklis, A.K., et al., Characteristics of natural selection in populations of nodule bacteria (*Rhizobium leguminosarum*) interacting with different host plants, *Russ. J. Genet.*, 2015, vol. 51, no. 10, pp. 949–956. doi 10.1134/S1022795415100026
16. Provorov, N.A., Andronov, E.E., Onishchuk, O.P., et al., Genetic structure of the introduced and local populations of *Rhizobioum leguminosarum* in plant–soil systems, *Microbiology* (Moscow), 2012, vol. 81, no. 2, pp. 224–232. doi 10.1134/S0026261712020129
17. Trotter, M.V. and Spencer, H.G., Frequency-dependent selection and the maintenance of genetic variation: exploring the parameter space of the multiallelic pairwise interaction model, *Genetics*, 2007, vol. 176, no. 3, pp. 1729–1740.
18. Friesen, M.L., Saxer, G., Travisano, M., and Doeveli, M., Experimental evidence for sympartric ecological diversification due to frequency-dependent competition in *Escherichia coli, Evolution*, 2004, vol. 58, no. 2, pp. 245–260.
19. Tikhonovich, I.A., Kozhemyakov, A.P., Ovtsyna, A.O., and Provorov, N.A., Construction of highly efficient symbiotic systems, in *New Approaches and Techniques in Breeding Sustainable Fodder Crops and Amenity Grasses*, S.-Petersburg: All-Russia Research Institute of Plant Industry, 2000, pp. 131–135.
20. Rocha, E.P.C., Maynard Smith, J., Hurst, L.D., et al., Comparisons of dN/dS are time dependent for closely related bacterial genomes, *J. Theor. Biol.*, 2006, vol. 239, pp. 226–235.
21. Gladieux, P., Devier, B., Aguileta, G., et al., Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens, *Infect. Genet. Evol.*, 2013, vol. 17, pp. 123–131.
22. Nystedt, B., Frank, A.C., Thollesson, M., and Andersson, S.G., Diversifying selection and concerted evolution of a type IV secretion system in *Bartonella, Mol. Biol. Evol.*, 2008, vol. 25, no. 2, pp. 287–300.
23. Cervantes, L., Bustos, P., Girard, L., et al., The conjugative plasmid of a bean-nodulating *Sinorhizobium fredii* strain is assembled from sequences of two *Rhizobium* plasmids and the chromosome of a *Sinorhizobium* strain, *BMC Microbiol.*, 2011, vol. 11, p. 149. doi 10.1186/1471-2180-11-149
24. Provorov, N.A. and Vorob'ev, N.I., Population genetics of nodule bacteria: simulation of cyclic processes in bacterial–plant systems, *Russ. J. Genet.*, 1998, vol. 34, no. 12, pp. 1455–1461.

25. Thompson, J.N. and Burdon, J., Gene-for-gene coevolution between plants and parasites, *Nature*, 1992, vol. 360, pp. 121–125.

26. Vorob'ev, N.I. and Provorov, N.A., The quorum sensing and the nodulation competitiveness of rhizobia during infection of leguminous plants, *S-kh. Biol.*, 2015, no. 3, pp. 298–304.

27. Provorov, N.A. and Vorobyev, N.I., Microevolution of nodule bacteria upon generation of mutants with altered survival in the plant–soil system, *Russ. J. Genet.*, 2003, vol. 39, no. 12, pp. 1349–1359. doi 10.1023/B:RUGE.0000009147.28398.90

28. Triplett, E.W. and Sadowsky, M.J., Genetics of competition for nodulation of legumes, *Annu. Rev. Microbiol.*, 1992, vol. 46, pp. 399–428.

29. Onishchuk, O.P. and Simarov, B.V., Genes controlling nodulation competitiveness of nodule bacteria, *Russ. J. Genet.*, 1996, vol. 32, no. 9, pp. 1001–1010.

30. Pobigaylo, N., Szymczak, S., Nattkemper, T.W., and Becker, A., Identification of genes relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants, *Mol. Plant–Microbe Interact.*, 2008, vol. 21, no. 2, pp. 219–231.

31. Onishchuk, O.P., Sharypova, L.A., and Simarov, B.V., Isolation and characterization of the *Rhizobium meliloti* Tn5-mutants with impaired nodulation competitiveness, *Plant Soil*, 1994, vol. 197, pp. 267–274.

32. Merezhkovskii, K.S., *Teoriya dvukh plazm kak osnova simbiogenezisa, novogo ucheniya o proiskhozhdenii organizmov* (The Theory of Two Plasms as the Basis of Symbiogenesis, a New Study on the Origins of Organisms), Kazan: Imp. Univ., 1909.

33. Margulis, L. and Sagan, D., *Acquiring Genomes: A Theory of the Origins of Species*, New York: Basic Books, 2002.

34. Margulis, L., *Symbiosis in Cell Evolution: Life and Its Environment on the Early Earth*, San Francisco: Freeman, 1981.

35. Frank, S.A., Genetics of mutualism: the evolution of altruism between species, *J. Theor. Biol.*, 1994, vol. 170, pp. 393–400.

36. Provorov, N.A. and Vorobyev, N.I., Evolution of host-beneficial traits in nitrogen-fixing bacteria: modeling and construction of systems for interspecies altruism, *Appl. Biochem. Microbiol.* (Dordrecht), 2015, vol. 51, no. 4, pp. 381–387.

37. van der Giezen, M. and Tovar, J., Degenerate mitochondria, *EMBO Rep.*, 2005, vol. 6, pp. 525–530.

38. Udvardi, M.K. and Kahn, M.L., Evolution of the (*Brady*)*Rhizobium*–legume symbiosis: why do bacteroids fix nitrogen?, *Symbiosis*, 1993, vol. 14, pp. 87–101.

39. Udvardi, M. and Poole, P.S., Transport and metabolism in legume–rhizobia symbioses, *Annu. Rev. Plant Biol.*, 2013, vol. 64, pp. 201–225.

40. Ran, L., Larsson, J., Vigil-Stenman, T., et al., Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium, *PLoS One*, 2010, vol. 5, no. 7. e11486

41. Österman, J., Marsh, J., Laine, P.K., et al., 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

Translated by A. Lisenkova