https://doi.org/10.17816/ecogen17323-38

SINORHIZOBIUM MELILOTI: CHROMOSOMAL TYPES AND GENOMIC ISLANDS

© M.E. Cherkasova, V.S. Muntyan, A.S. Saksaganskaia, B.V. Simarov, M.L. Roumiantseva

All-Russia Research Institute for Agricultural Microbiology, Saint Petersburg, Russia

Cite this article as: Cherkasova ME, Muntyan VS, Saksaganskaia AS, et al. *Sinorhizobium meliloti:* chromosomal types and genomic islands. *Ecological genetics*. 2019;17(3):23-38. https://doi.org/10.17816/ecogen17323-38.

Received: *15.02.2019* Revised: *03.07.2019* Accepted: *24.09.2019*

❀ **Background.** Polymorphism analysis was done for the core genome sequences of nodule bacteria of *S. meliloti* species in order to identify chromosomal types and to evaluate the occurrence of accessory elements (genomic islands) in them. **Materials and methods.** Chromosomal studied loci were: *bet*BC (marker M-I) and *SMc*04407-*SMc*04881 (marker M-II) both are related to metabolic processes and stress tolerance, and 16S-23S intergenic sequences (marker M-III) to search phylogenetical distance at intraspecies level. **Results.** Significant differences between the occurrence of alleles of gene-markers M-I/M-II and MIII were determined between strains related to tested the 5 homogenous groups and 9 subgroups of strains differing by geographical region/source (nodule, soil) of isolation, as well as by salt tolerance. Four chromosomal types were identified among tested *S. meliloti* native isolates and a preference occurence of one of the three islands Rm1021 in links with particular chromosomal type was shown. The significant prevalence of strains with particular chromosomal type was shown for *S. meliloti* populations native to centers of alfalfa diversity at the NE of Caucasus, as well as at NE of Kazakhstan (Aral Sea related region), as well as in agrocenoses. **Conclusion.** It was predicted that strains inherited altered markers M-I/M-II may belong to divergent clonal lines occured in both centers of alfalfa diversity, while strains with altered sequences of all three markers could be a representatives of a new *S. meliloti* biovar(s), the formation of which is occurred much more intensively at the modern center of the introgressive hybridization of alfalfa at NE of Kazakhstan.

❀ **Keywords:** *Sinorhizobium meliloti*; chromosomal markers polymorphism; chromosomal types; genomic islands; PCR analysis; salt tolerance.

SINORHIZOBIUM MELILOTI: ХРОМОСОМНЫЕ ТИПЫ И ГЕНОМНЫЕ ОСТРОВА

© М.Е. Черкасова, В.С. Мунтян, А.С. Саксаганская, Б.В. Симаров, М.Л. Румянцева

ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Санкт-Петербург

Для цитирования: Черкасова М.Е., Мунтян В.С., Саксаганская А.С., и др. *Sinorhizobium meliloti*: хромосомные типы и геномные острова // Экологическая генетика. – 2019. – Т. 17. – № 3. – С. 23–38. https://doi.org/10.17816/ecogen17323-38.

Поступила: *15.02.2019* Одобрена: *03.07.2019* Принята: *24.09.2019*

❀ Выполнен анализ полиморфизма последовательностей корового генома клубеньковых бактерий *S. meliloti* с целью выявления хромосомных типов и оценки встречаемости в них геномных островов, рассматриваемых как акцессорные элементы хромосомы. В результате сопряженного анализа генов-маркеров M-I (*bet*BC) и M-II (*SMc*04407-*SMc*04881), продукты которых задействованы в клеточном метаболизме и вовлечены в процессы формирования стрессоустойчивости, а также последовательностей маркера M-III (IGS *rrs-rrl*), используемых в филогенетических исследованиях на уровне вида, были выявлены достоверные различия между пятью типическими группами и девятью подгруппами штаммов, различавшихся по району и источнику выделения, а также по солеустойчивости. Определены четыре хромосомных типа и показана предпочтительность наличия одного из трех островов Rm1021 в каждом из них. Установлены достоверные различия по встречаемости штаммов, имевших определенный хромосомный тип в очагах разнообразия люцерны, расположенных в северных районах Кавказа и Казахстана (Приаралье), а также в агроценозах. Сделано заключение, что штаммы с измененными маркерами M-I/M-II могут относиться к дивергентным клональным линиям, тогда как штаммы с измененным маркером M-III, а также маркерами M-I/M-II могут являться представителями нового(ых) биовара(ов) клубеньковых бактерий люцерны, который(ые) формируется(ются) значительно активнее в современном центре интрогрессивной гибридизации люцерн в Приаралье.

❀ **Ключевые слова:** *Sinorhizobium meliloti*; полиморфизм хромосомных маркеров; хромосомные типы; геномные острова; ПЦР-анализ; солеустойчивость.

BACKGROUND

The genome of nodule bacteria (rhizobia) includes chromosome and plasmids that significantly differ in size. The availability of several replicons in the genome can be explained by the fact that rhizobia are typical soil saprophytes, as well as form non-obligatory, nitrogen-fixing symbiosis with legume plants. In the process of nodule formation, the bacterial cells are exposed to irreversible morpho-physiological and metabolic changes in the microaerophilic and hyperosmotic conditions of the nodule.

Genes are encoding in the bacterial chromosomes and determining vital processes are known as housekeeping genes (or basic/core genes) [1–7]. Orthologs are homologous genes present in all strains of one species/genus, as well as in the appropriate supposed common ancestors [8, 9]. The study of the gene structure of genetically unrelated bacteria strains by the restriction endonuclease profiles (the restriction fragment length polymorphism – RFLP) and/or by nucleotide sequences levels allowed to the detect the alleles with different levels of similarity, and based on that strains were united into groups/clusters [4, 9–11]. Chromosomes of strains from distinct clusters or groups are described as being from different chromosomal lineages [4, 10, 11], chromosomal groups [12], chromosomal genotypes [10, 13], or chromosomal types [10, 11, 14, 15]. It is necessary to note that there is no generally accepted term and above listed terms are often used synonymously, even in one publication [4, 10, 11, 16]. The term "chromosomal types" is used in this work.

The most important requirement for choosing genetic markers for the determination of chromosomal types is a low level of structural polymorphism in the candidate genes and the absence (or low level) of recombination between them [10]; besides that, they should be orthologs [17]. Publication analysis demonstrates that *rrs* (16S rRNA), *rrl* (23S rRNA), *atp*D (ATP synthase beta chain), *gln*II (glutamate-ammonia ligase), *rec*A (protein involved in recombination-dependent repair processes) or *zwf* (glucose-6-phosphate 1-dehydrogenase), *sod* (superoxide dismutase), *lac*Z (β-galactosidase) or *SMc*00019 (conserved hypothetical protein), *tru*A (tRNA-pseudouridine synthase A), and *thr*A (homoserin dehydrogenase) are often used as genetic markers [10, 11, 18, 19]. The authors mostly analyzed two or three core genes from the above-mentioned list; while, there are publications where one genetic marker is used, for example, *gyr*B (DNA gyrase subunit B) [4] or *dna*J (or Hsp40; heat shock protein, size 40 kDa) [20] was used. We analyzed products of the above-mentioned genes by using the database of the Clusters of Orthologous Groups of proteins (**COG** group) [17]. We studied how products of above-mentioned genes are met across 25 known COG groups and found that they are belonged to 8 groups related to three out of four COG clusters. One of the specified clusters includes genes whose products participate in Information storage and processing (groups L, J, and K). Another cluster includes genes whose products are involved in Metabolism process (groups C, E, G, and P), and the third cluster includes genes whose products participate in Cellular processes and signaling (group O). There are cases when in addition to core genes the intergenic sequences *rrs-rrl* (Intergenic Transcribed Spacer, ITS) were used to detect chromosomal types [4, 13, 20, 21]. However, ITS analysis is mostly used in phylogenetic studies of bacteria at the intraspecies level [18, 20–23], as this allows the detection of new phylogenetically distant clusters of strains (lineages/biovars/subspecies) [21, 24, 25].

The research of Escobar-Páramo et al. (2004) is worth to note, as authors propose that strains of *Escherichia coli* with certain chromosomal types contain certain virulence factors [26]. The latter ones are attributed to additional, or accessory, elements of genome, as they are not present in all strains of the same species [2]. In general accessory elements include plasmids, genomic islets, and genomic islands, which all are considered as genetic elements of the auxiliary, flexible [3, 5, 27, 28], or accessory genome (the latter term is used much more often in publication) [1, 4, 5, 7, 27, 28]. As a rule, the genes from the accessory part of the genome are not involved in the control of the vital cell functions but can determine the ability of bacteria to occupy different ecological niches [29, 30]. Genomic islands (GIs) are the most intriguing accessory elements; they are long sequences (up to 700 kbp) of phage origin that are site-specifically inserted in to the gene sequences that encode tRNA [31]. Genomic islands are characterized by a reduced content of guanine and cytosine base pairs (average up to 12%) in comparison with sequences from core part of the genome [31, 32]. The availability of sequences with less stable structure (due to their enrichment with double hydrogen bonds between adenine and thymine base pairs) can affect on the chromosome structure and on gene activity as well [33]. Genomic islands contain mobile genetic elements of different classes and open reading frames (ORFs). A functional analysis of the hypothetical products of open reading frames of GIs showed that their ORFs could be paralogs/orthologs of functionally different genes homologous to those of bacteria from different phyla [28]. This proves that GIs participate in long-distance horizontal gene transfer. It has been demonstrated that strains of one species can significantly differ in GIs presence [34–36]; however, there is currently no data about the dependence of GIs occurrence on the core genome properties for nodule bacteria.

Three long sequences (from 19 to 80 kbp in length) met with GIs characteristics according to the Islander database, were detected in the chromosome of reference strain *Sinorhizobium meliloti* Rm1021, a symbiont of the alfalfa [37]. The adjacent GIs (Sme21T and Sme19T) of similar length are located at the first quarter of chromosome, whereas the third genome island (Sme80S) is located at the last quarter of the chromosome and it is significantly longer than both above indicated (Fig. 1) [28]. Sme80S includes copies of genes relating to symbiotic activity and osmoprotection [36]. Significant differences in occurrence of Sme80S and Sme21T were established for *S. meliloti* native isolates adapted to salt and arid soils accordingly [28].

The aim of this study was to investigate the polymorphism of a set of core chromosome sequences for detec-

tion chromosomal types and to assess the GIs occurrence (considered as accessory elements of the chromosome) in genomes of *S. meliloti* native isolates differing in origin (center of legume diversity) and in source of recovering (nodule or soil), and in phenotype characteristics (salt-tolerant or salt-sensitive).

MATERIALS AND METHODS

Strains. The 218 native isolates of *S. meliloti* belonging to the collection of nodule bacteria strains of the Laboratory of Genetics and Breeding of Microorganisms, ARRIAM, Saint Petersburg, Russia were studied. Strains were recovered from **n**odules of host plants and **t**rapped from soil (hereafter referred to as N-isolates and T-isolates, respectively) collected at the northern part of the Caucasian genetic center of cultivated plants (North Caucasus Center; NCC) and at the northern area of the coastal part of Aral Sea region of Kazakhstan attributed to the up-to-date center of introgressive hybridization of alfalfa (Aral Sea Center; ASC), as well as from soils of agrocenoses next to the Saint Petersburg (NW of Russia) by appropriate techniques [38]. In order to isolate strains from soils, the method of trapping was used, according to which the sterile seedlings of plants were inoculated with soil extracts [38]. In total, 95 and 89 strains of *S. meliloti* were analyzed from NCC and ASC presented by N- and T-isolates (31/64 and 66/23 accordingly), and 34 T-isolates were analyzed from agrocenoses. Taxonomic classification of strains from NCC and ASC to *S. meliloti* species was determined previously [39]; strains from agrocenoses were assigned to *S. meliloti* species according to data proceeded by amplified ribosomal DNA restriction analysis (ARDRA). The strain *S. meliloti* Rm1021 was used as a reference.

Salt tolerance of strains was determined by growth in TY medium (0.6 M NaCl), according to the method described previously [38]. Strains from NCC and ASC had salt-tolerant (normal growth; R) or salt-sensitive phenotypes (poor growth; S) [40]. Salt tolerance of isolates from agrocenoses also determined in current work. The ratios of R/S of strains in populations from NCC and ASC and in the group of strains from agrocenoses were as follows: 74/21, 49/40, and 31/3, respectively.

Marker sequences of chromosome were determined in accordance with the genome-wide sequence of reference strain *S. meliloti* Rm1021 (GenBank NC_003047.1; see Fig. 1). Marker M-I is a 1544 bp sequence that includes the sequence of the 3'-end of the gene *bet*B (143 bp) and the 5'-end of gene *bet*C (1400 bp), and there is 1 bp between genes *bet*C and *bet*B. Genes *bet*B and *bet*C are located in the second quarter of the chromosome and are involved in the formation of the salt-tolerant phenotype, as well as in metabolism of carbon and nitrogen (see Fig. 1). Marker M-II is a locus *msfp* located in the last quarter of the chromosome, close to *ori*C (see Fig. 1). The analyzed sequence of *msfp* (1280 bp) includes the sequence of the 5'-end of the gene *SMc*04407 (532 bp), the sequence between genes *SMc*04407 and *SMc*04881 (518 bp), and the sequence of the 3'-end of the gene *SMc*04881 (230 bp). The product of the first gene is a protein that is probably involved in the transport of an osmoprotectant (proline/betaine); the product of the second gene is involved in the synthesis of a homologue of the VapC-toxin of type II toxin-antitoxin system, and in the formation of bacterial cell resting state in stress conditions [41]. Since M-I and M-II are related to metabolism and stress tolerance, they are denoted and are used further as "metabolic markers."

Marker M-III is intergenic sequence ITS (IGS *rrs-rrl*; see Fig. 1) of ribosomal operons (*rrs-rrl*; 1308 bp), including the 3'end sequence of 16S rRNA gene (21 bp), the sequence between 16S rRNA and $tRNA^{Ile}$ genes (256 bp), the sequence of $tRNA^{Ile}$ gene (77 bp), the sequence between $tRNA^{Ile}$ and $tRNA^{Ala}$ genes (139 bp), the sequence of $tRNA^{Ala}$ gene (76 bp), the sequence between $tRNA^{Ala}$ and 23S rRNA genes (607 bp), and the 5'end of the sequence of 23S rRNA gene (131 bp). Analysis of the marker M-III sequence is using in phylogenetic research of bacteria at the intraspecies level and as a phylogenetic marker.

Analysis of the protein products of gene-markers M-I and M-II demonstrated that they were belonged to the same COG groups (C, E, G, P, and R) as the above-mentioned markers used in papers to determine the chromosomal types of the bacteria strains [4, 10, 11, 18–20].

PCR analysis was conducted for the purpose of strain genotyping. The restriction profiles of PCR-amplified sequences (PCR-RFLP) obtained for each of the studied strains were analyzed as RFLP types or alleles according methods

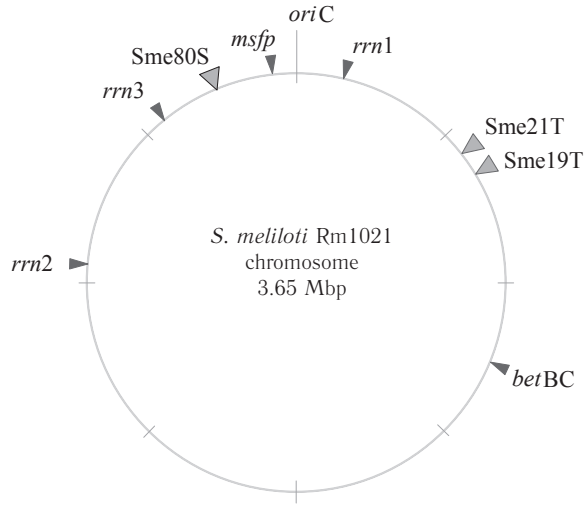


Fig. 1. The scheme of localization of the core marker sequences and genome islands on the chromosome of the reference strain of *Sinorhizobium meliloti* Rm1021: *oriC* 6 – chromosomal origin of replications; ▲ – core markers; *rrn* (1-3) – ribosomal RNA operons containing ITS-sequences; △ – genomic islands

described previously [42, 43]. PCR analysis was conducted using appropriate primers pairs for the specified marker sequences: M-I (*bet*BC) [44], M-II (*msfp*; MSFp-F: CACCAGCGAGAGGAAGAGAC, MSFp-R: GAAACCCTGCGTTTGTTGAT), and M-III (ITS) [39]. PCR was conducted in 20 μl of mixture for amplification containing 25 mM $MgCl_2$ and 1.5 units of Taq DNA-polymerase (Eurogene™ PK113L) at pH 9.1 in a C1000™ Thermal Cycler (Bio-Rad, USA). Primer annealing was performed at 55 °C for 30 response cycles. DNA was extracted by standard method of phenol-chloroform extraction from cell lysates [45]. Products of amplification were restricted using with endonucleases *Msp*I and *Hae*III separately in the case of markers M-I and M-III and endonuclease *Eco*RI in the case of M-II (Thermo Fisher Scientific™ ER0541, ER0151, and ER0271, respectively). Restriction was done according to standard protocol of Thermo Fisher Scientific (temperature: 37 °C; time: 2 h). Electrophoretic separation of the obtained fragments was performed in 3% agarose gel in 0.5-fold TAE.

As a result, PCR-RFLP profiles (hereafter, RFLP types) for each specified marker were obtained for every strain. RFLP types similar to that of the reference strain were designated as type "a", whereas different ones, i. e., divergent ones, were marked with different letters of the alphabet, for example, "b" and "c", etc.

The presence of genomic islands (GI) were determined with PCR using original primers pairs on the external bordered regions [36]. Total DNA of strain Rm1021, that genome sequence contains three islands (Sme19T, Sme21T, and Sme80S) and strain CXM1-105, that genome sequence does not contain islands, were used as controls in the experiments for detection of GIs in native isolates according to method described previously [31].

Statistical processing of data was carried out using the PAST software [46]: criterion χ^2 at $\alpha = 0.05$ (df – number of degrees freedom), Shannon heterogeneity index (H), and evenness index (E). Nei's genetic diversity index was calculated using the formula: $N = (1 - \sum(P_i^2)) \cdot (n/(n-1))$, where P_i = frequency of i-th genotype and n = overall number of detected genotypes [47]. Linkage disequilibrium (LD) was calculated using the Arlequin software (v. 3.5, 2010) [48].

RESULTS AND DISCUSSION

The 218 native isolates of *Sinorhizobium meliloti* originated from geographically far away regions (centers of legume diversity NCC and ASC and agrocenoses) and from different sources (N- and T-isolates) were analyzed. The analysis of the N-isolates from NCC and ASC revealed that predominantly with almost similar frequencies they had R-phenotype (average frequency 0.68). Isolates of R-phenotype was also dominating among the T-isolates from NCC and agrocenoses (frequencies 0.80 and 0.91 respectively), while isolates of S-phenotype was in privilege in group of T-isolates from ASC (frequency 0.65). Between these groups of T-isolates the differences were significant ($\chi^2_{ASC/NCC} = 15.6$; $p = 7.7 \cdot 10^{-5}$; df = 1; $\chi^2_{ASC/AC} = 20.2$; $p = 7.0 \cdot 10^{-6}$; df = 1). Significant differences in the occurrence of strains differed in phenotype were detected for groups of N- and T-isolates from ASC ($\chi^2 = 5.15$; $p = 2.3 \cdot 10^{-2}$; df = 1) but were not detected for similar groups from NCC.

According to the obtained data, the five groups of isolates differed in origins (NCC, ASC and agrocenoses) and sources of isolation (N and T) were formed. Each group was represented by two subgroups of isolates of R or S phenotype, i. e., in total of 10 subgroups (further homogenous groups/subgroups; Fig. 2). The subgroup of T-isolates of the S-phenotype from agrocenoses was represented only by three strains (frequency in the group 0.09; see Fig. 2), it was not included in a comparative statistical analysis.

Polymorphism of sequences of M-I and M-II markers was studied in above-mentioned homogenous groups and subgroups of *S. meliloti* native isolates (see Fig. 2). RFLP types obtained for M-I and M-II markers were similar to that of the reference strain (further the reference type or allele "a"), as well as sequences differed from the reference (divergent types or divergent alleles) were detected. The five divergent alleles for marker M-I and four divergent alleles for M-II were revealed, i. e., the appropriate marker sequences had structural modifications in comparison with reference sequence. Values of Shannon heterogeneity indices (H) were calculated from the occurrence of reference and divergent alleles of M-I and M-II markers, which amounted to 0.89 and 0.97, respectively.

Reference alleles of M-I and M-II markers were more frequent in both subgroups of T-isolates from NCC and in the subgroup of salt-sensitive T-isolates from ASC (see Fig. 2, *a*, *b*). The frequencies of occurrence of these alleles were similar: 0.84, 0.69, and 0.9, respectively for each marker (see Fig. 2, *a*, *b*). However, in cases of salt-tolerant T-isolates from ASC and from agrocenoses the frequencies of reference alleles for M-I or M-II were low, but equal or similar (0.13 and 0.27, respectively; see Fig. 2, *a*, *b*). Significant differences were detected between the subgroups of T-isolates from ASC ($\chi^2 = 12.0$; $p = 5.2 \cdot 10^{-4}$, df 1 и $\chi^2 = 15.0$; $p = 1.1 \cdot 10^{-4}$; df = 1, respectively, for M-I and M-II). The reference allele of M-I was detected with similar frequencies in both subgroups of N-isolates from NCC and among salt-tolerant N-isolates from ASC (average frequency: 0.52), but less often among salt-sensitive N-isolates from ASC (0.44; see Fig. 2, *a*). It is interesting that the reference allele of M-II was detected more often in salt-tolerant than in salt-sensitive N-isolates whether in cases of NCC or ASC (1.3 and 1.3 times, respectively); however, the differences were not significant (see Fig. 2, *b*). The reference alleles of M-I and M-II occurred significantly more often among

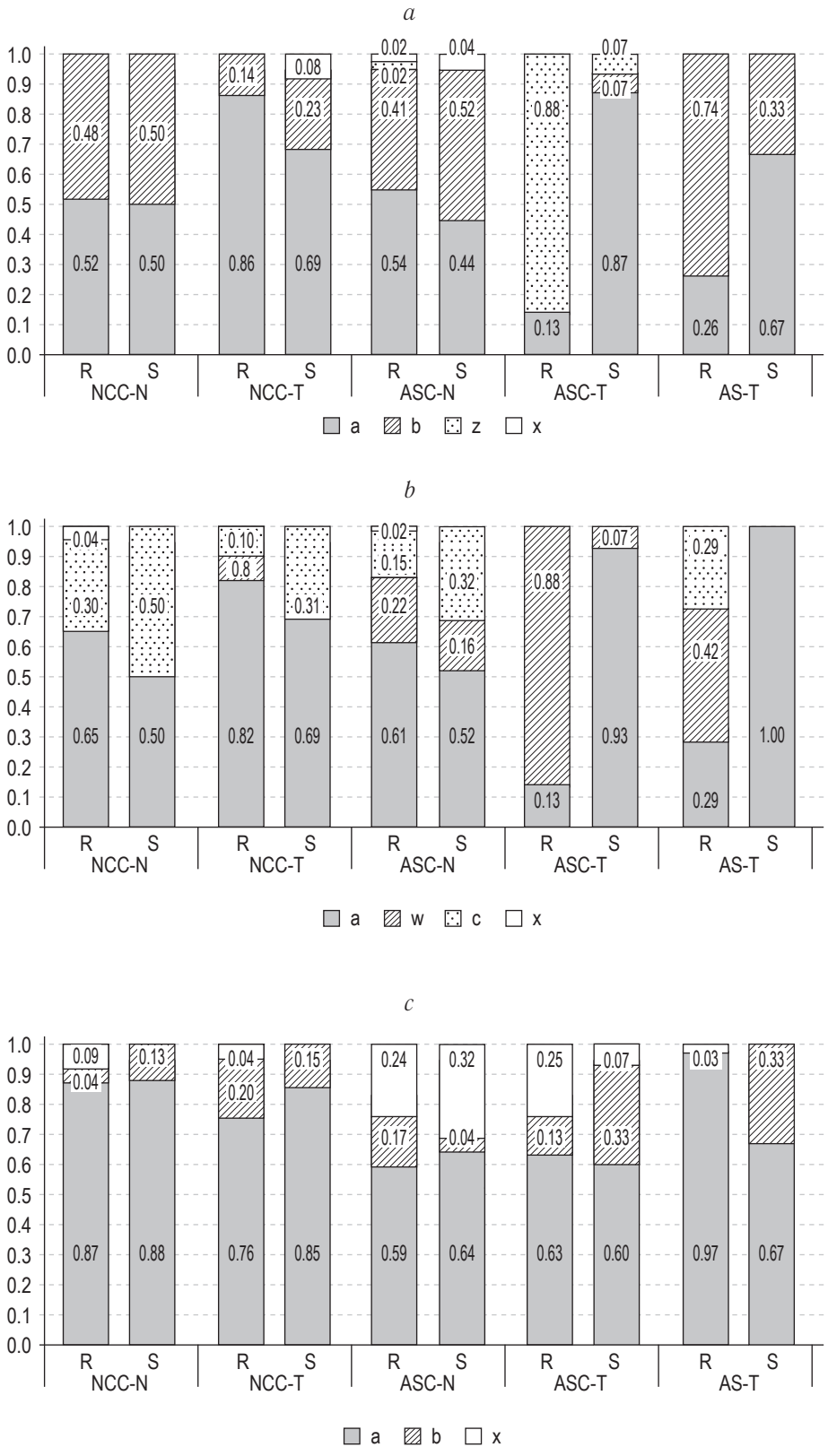


Fig. 2. Frequency of occurrence of RFLP-type M-I (*a*), M-II (*b*) and M-III (*c*) markers in strains of *S. meliloti*: a, b, c, w, z – RFLP types, x – group combining unique divergent RFLP types; R – strains of salt-tolerant phenotype; S – strains of salt-sensitive phenotype; N – strains recovered from nodules; T – strains trapped from soil; NCC – North Caucasian genetic center; ASC – Aral Sea center of alfalfa diversity; AC – agrocenoses

Table 1

Linkage disequilibrium (LD) between markers M-I and M-II

LD		M-II				
		NCC-N	NCC-T	ASC-N	ASC-T	AC-T
M-I	NCC-N	0.674				
	NCC-T		0.639			
	ASC-N			0.376		
	ASC-T				0.830	
	AC-T					0.764

Note. N – strains recovered from nodules; T – strains trapped from soil; NCC – North Caucasus genetic center; ASC – Aral Sea center of alfalfa diversity.

T-isolates of NCC (mostly of salt-tolerant phenotype) and among N-isolates of ASC ($\chi^2 = 25.3$; $p = 4.8 \cdot 10^{-7}$; df = 1; $\chi^2 = 24.1$; $p = 9.2 \cdot 10^{-7}$; df = 1, respectively), as it was determined by comparative analysis of N- and T-groups of isolates with one another.

Divergent alleles "b" of M-I and "c" of M-II were detected in almost all subgroups of strains, except in T-isolates from ASC (see Fig. 2, *a*, *b*). Both alleles mentioned above were found in more than a one fourth part of salt-sensitive T-isolates from NCC, while salt-tolerant isolates from the same region contained these alleles in 2–3 times less often (see Fig. 2, *a*, *b*). Remarkable that allele "b" of M-I was dominant in salt-tolerant T-isolates from agrocenoses (frequency 0.74), while allele "c" of M-II was found half as often (see Fig. 2, *a*, *b*). Both divergent alleles were also identified in N-isolates from ASC and NCC with frequencies were comparable with the occurrence frequencies for reference alleles in corresponding groups of strains ($\chi^2 = 0.68$; $p = 0.9$; df = 3). In addition, allele "c" of M-II was identified more often in salt-sensitive N-isolates from NCC and ASC, but the differences were not significant ($p > 0.5$; see Fig. 2, *a*, *b*). However in the groups of N-isolates from NCC and ASC, alleles "b" of M-I and "c" of M-II were found significantly more frequently than in corresponding groups of T-isolates ($\chi^2 = 12.2$; $p = 6.9 \cdot 10^{-3}$; df = 3; $\chi^2 = 8.7$; $p = 3.3 \cdot 10^{-2}$; df = 3 respectively).

It should also be noted the occurrence of divergent alleles "z" of M-I and "w" of M-II, of which the first one was found in single N- and T-isolates from ASC, while the second one was found in single salt-tolerant and salt-sensitive T-isolates from NCC and ASC, respectively. In contrast with the above, both alleles were occurred with more frequencies in both subgroups of N-isolates (average frequency 0.19), as well in subgroup of salt-tolerant T-isolates from ASG. In the latter case, both alleles occurred with the frequency 0.88 (see Fig. 2, *a*, *b*).

Besides the four types of divergent alleles of markers M-I and M-II considered above, the unique divergent alleles were identified for single strains, which were combined into groups "x" (see Fig. 2, *a*, *b*). Strains that had unique types of marker M-I, were found in both subgroups of N-isolates from ASC, as well as among salt-sensitive T-isolates in NCC. Unique types of M-II marker were observed only in subgroups of salt-tolerant N-isolates in ASC and NCC. Values of χ^2 calculated for all groups of strains harboring above-mentioned alleles of M-I and M-II markers are provided in Supplementary, Table 1.

Thus, the data produced by the analysis of markers M-I and M-II, encoding genes related to metabolic processes and stress tolerance (particular to salinity) showed that soil isolates from different geographical areas inherited predominantly alleles similar to reference. The exception was two subgroups: T-isolates of the salt-tolerant phenotype from ASC and AC, in which divergent alleles of marker M-I, as well as M-II, were dominated with similar frequencies. Groups of N-isolates were significantly more diverse than T-isolates from both NCC and ASC, due to the fact that the frequencies of occurrence of divergent alleles of M-I and M-II were higher and comparable to the frequencies of occurrence for the corresponding reference alleles. The data obtained allowed us to conclude that an increase of genotypic diversity (diversification) of strains isolated from soil (trapping method, see "Materials and methods") apparently occurs under impact of the frequency-dependent disruptive selection in genetic centers of legume diversity and in agrocenoses, whereas genotypic diversity of strains recovered from nodules is under the effect of the frequency-dependent stabilizing selection, that is in agreement with data of [49, 50]. In addition, strains with altered metabolic markers with respect to reference, which are dominating among T-isolates of the salt-tolerant phenotype may be referred to clonal lines from the ASC population, as well as from agrocenoses.

Polymorphism of M-III marker sequence. The reference strain Rm1021 harbored the three identical ITS sequences (GenBank NC_003047.1), all of three have identical RFLP profile, hereinafter type "a" of marker M-III (see Materials and methods, Fig. 1). The 16 divergent types of marker M-III, in addition to the reference type "a", were identified in tested strains. This indicated at the structural differences in ITS sequences of ribosomal operons in native isolates, which is consistent with [39]. The heterogeneity index was significantly higher for M-III marker sequences, than for markers M-I and M-II (H = 1.38, 0.89 and 0.97, correspondingly) of the 218 tested strains. The heterogeneity index calculated for M-III marker sequences was 2.4 times higher for strains from ASC than from NCC (1.51 and 0.63, respectively). Besides, values of *H* were similar in case of groups of N- and T-isolates from NCC (0.52 and 0.62, respectively), but were 1.5 times higher for N-isolates than for T-isolates from ASC (1.55 and 1.06, respectively). The population of strains from ASC is phylogenetically more diverse (according to the values of the index H for marker M-III) and it can be assumed that processes aimed at the formation of new biovars of alfalfa nodule bacteria can occur in this particular region.

Type "a" of M-III was predominant in all subgroups of strains but mostly in rhizobia of salt-tolerant phenotype. It is occurred significantly more often (1.4 times average) in the population from NCC than from ASC ($\chi^2 = 27.2$; $p = 5.5 \cdot 10^{-6}$; df = 3) mostly in the rhizobia of the salt-tolerant phenotype. At the same time, no significant differences between subgroups of N- and T-isolates from evaluated genetic centers were detected ($p > 0.05$; see Fig. 2, *c*). The reference type was dominant among T-isolates originated from agrocenoses (frequency 0.97; see Fig. 2, *c*).

Divergent type "b" of marker M-III was the second-most frequent in all subgroups of strains, with the exception of one salt-sensitive T-isolate from agrocenose. The type "b" of M-III was identified in salt-sensitive N- and T-isolates (average frequency 0.14), but 1.4 times more often in T-isolates and extremely rarely in salt-tolerant N-isolates from NCC (see Fig. 2, *c*). The type "b" was mostly detected in N-isolates of R-phenotype (frequency 0.17), but only in single isolate of S-phenotype from population ASC (see Fig. 2, *c*). With the highest frequency of type "b" of marker M-III was identified in salt-sensitive T-isolates from ASC (frequency 0.33), whereas in salt-tolerant strains it was observed 2.5 times less often (see Fig. 2, *c*). Geographically distinct populations of *S. meliloti* strains from evaluated genetic centers were significantly differ in the occurrence of typical and divergent alleles of marker M-III ($\chi^2 = 15.4$; $p = 4.5 \cdot 10^{-4}$; df = 2).

The remaining 15 divergent types were unique and were identified in one-three strains. Majority of unique types (13 out of 15) were detected in the subgroup of N-isolates and in the subgroup of salt-tolerant T-isolates from ASC (average frequency 0.27; see Fig. 2, *c*). Strains that had specified unique types of marker M-III were combined in groups "x" (see Fig. 2, *c*), which were not further considered.

Thus, divergent sequences of marker M-III were predominant in population of strains from ASC in comparison with strains from other genetic centers, as well as from agrocenoses. The obtained results correlate with previously published data [39, 51, 52] and confirm that, in populations of rhizobia adapted to the condition of extreme salinity, the processes of diversification are active and are probably aimed at the formation of new biovars of alfalfa nodule bacteria.

Analysis of combinations of chromosomal marker sequences. Pairwise analysis of combinations of reference and divergent sequences of markers M-I, M-II, and M-III was conducted in order to detect linkage disequilibrium (LD). As a result, LD was detected between markers M-I and M-II in all examined groups (LD from 0.38 to 0.83; see Table 1) but was not revealed between marker M-III and marker M-I as well with marker M-II as well with the combination of M-I/M-II in all considered groups (data not provided). It should be concluded that evaluated markers could be used to assess diversity of chromosomal types, besides M-I/M-II, can be considered as a joint metabolic marker, while M-III, as a phylogenetic marker should be tested independently.

The 35 out of 510 theoretically possible combinations of markers M-III and M-I/M-II were revealed through the analysis of 218 strains (Supplementary, Table 2). Eight combinations were abundant in both tested geographically distinct populations (NCC and ASC), while four of them were identified also among strains from agrocenoses (data not provided), and other 15 combinations were uniquely identified only in salt-tolerant N-isolates from ASC (Supplementary, see Table 2).

The diversity of subgroups of strains was assessed according to Shannon diversity index calculated by the number of determined combinations. It was established that N-isolates from ASC were significantly more diverse than the corresponding groups of strains from NCC, whereas for subgroups of T-isolates from NCC, ASC and agrocenoses such differences were absent (Supplementary, see Table 2).

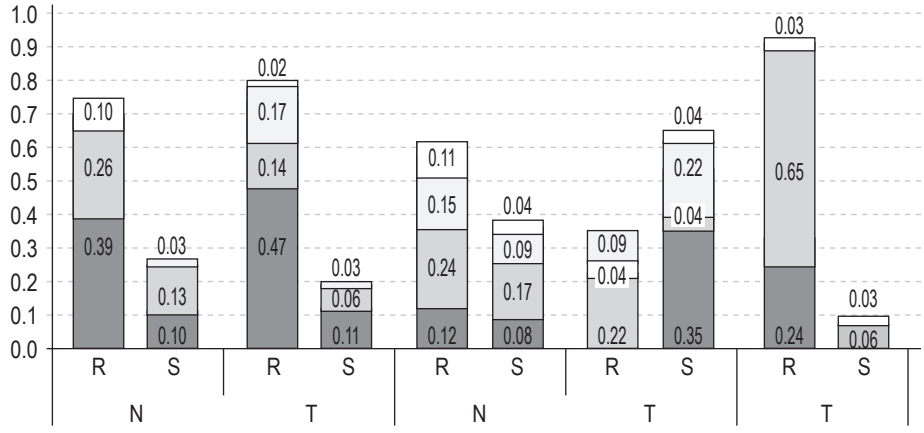
The 35 combinations of marker sequences were grouped into four distinct groups AI, AII, BI and BII based on the type of alleles (reference (A) or divergent (B) alleles of M-III, and reference (I) or and divergent (II) alleles of metabolic marker). These four groups were further considered as chromosomal types. Type AI (considered as a reference type) was represented only by combination of alleles of "a" type ($H_{AI} = 0$), whereas types AII, BI, and BII united correspondingly 8, 11, and 15 different com-

binations (H_{AII} = 1.48; H_{BI} = 1.37; H_{BII} = 2.58). It should be noted that dominant combinations were identified in each type: in type AII there were 2 out of 8 combinations (frequencies 0.29 and 0.44); in type BI the 1 out of 11 combinations (frequency 0.67); while in type BII the 13 of 15 combinations were identified in single strains with total frequency of occurrence amounted to 0.87 (data not provided).

Type AI was mainly revealed among salt-tolerant N- and T-isolates from NCC, it was observed 1.2 times more likely (average frequency 0.43; Fig. 3). The same chromosomal type was detected with similar frequency among salt-sensitive T-isolates (0.35) and was not found in salt-tolerant strains from ASC (see Fig. 3). Type AI was identified in single T-isolates of R-phenotype from agrocenoses (0.24; see Fig. 3). In all other subgroups (salt-sensitive N- and T-isolates from NCC and T-isolates from ASC), type AI was detected with similarly low frequencies (average frequency 0.10). Thus, chromosomal type AI was identified significantly more often among T-isolates from NCC (χ^2 = 6.9; $p = 8.5 \cdot 10^{-3}$; df = 1) that had mostly salt-tolerant phenotype. Comparative analysis of groups of strains showed that chromosomal type AI was found 2.5 and 1.7 times more often among N-isolates than among T-isolates from NCC and ASC, respectively.

Type AII (typical allele M-III and divergent alleles M-I/M-II) was revealed in all examined subgroups. This chromosomal type was characteristic of N- and T-isolates of different phenotypes from ASC (average frequency 0.21), as well as of salt-tolerant N-isolates from NCC (frequency 0.26; see Fig. 3). In the other cases, chromosomal type AII was detected half as often (subgroups of N- and T-isolates of different phenotype from NCC) or was identified in single salt-sensitive T-isolates from both genetic centers of legumes (see Fig. 3). According to the results of comparing the groups of N- and T-isolates from NCC and ASC, it was found that chromosomal type AII was almost 2 times more common among N-isolates than in the corresponding groups of T-isolates in both genetic centers, however, significant differences were obtained only for N-isolates from ASC (χ^2 = 7.4; $p = 6.6 \cdot 10^{-3}$, df = 1). Chromosomal type AII was predominant only for salt-tolerant T-isolates from agrocenoses (frequency 0.65; see Fig. 3).

Type BI (divergent allele M-III and typical alleles M-I/M-II) occurred much less frequently in tested subgroups of strains compared to type AII. The chromosomal type BI was characteristic of strains from three subgroups (with similar frequencies): salt-tolerant T-isolates from NCC, salt-sensitive T-isolates from ASC, and salt-tolerant N-isolates from ASC (average frequency 0.18; see Fig. 3). The same chromosomal type was identified in salt-sensitive N-isolates from ASC with less frequency (0.09; see Fig. 3). Among the others subgroups this chromosomal type was a character of a single isolates (average frequency 0.03) or was not detected (see Fig. 3). Comparison of groups of strains revealed that type BI was significantly more often observed among N-isolates from ASC and among salt-tolerant T-isolates from NCC (χ^2 = 8.1; $p = 4.3 \cdot 10^{-3}$; df = 1).



	NCC				ASC				AC	
□ BII	0.10	0	0.02	0	0.11	0.04	0.09	0.04	0.03	0.03
□ BI	0	0.03	0.17	0.03	0.15	0.09	0.04	0.22	0	0
▒ AII	0.26	0.13	0.14	0.06	0.24	0.17	0.22	0.04	0.65	0
■ AI	0.39	0.10	0.47	0.11	0.12	0.08	0	0.35	0.24	0.06

Fig. 3. The occurrence of distinct chromosomal types among *S. meliloti*. AI, AII, BI, BII – chromosomal types (see text); rest symbols see on Fig. 2

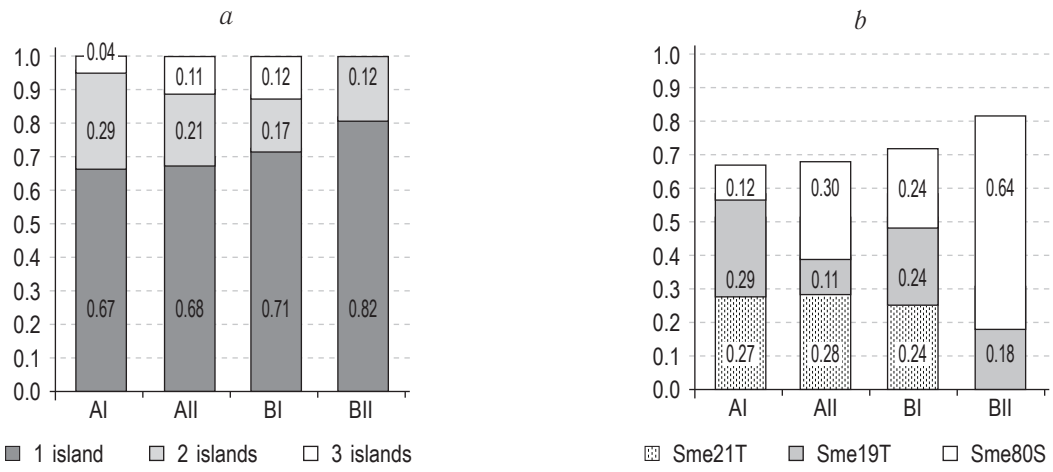


Fig. 4. The occurrence of genomic islands in *S. meliloti* strains harboured different chromosomal types of: *a* – the occurrence of one, two or three genomic islands in appropriate chromosomal types; *b* – the occurrence of each of the islands: Sme19T, Sme21T, Sme80S in appropriate chromosomal type; AI, AII, BI, BII – chromosomal types (see Fig. 3)

Chromosomal type BII (divergent alleles of M-III and alleles M-I/M-II) was identified only for a few strains related to 8 of the 10 studied subgroups (see Fig. 3). More often, type BII was observed in three subgroups of salt-tolerant N- and T-isolates from ASC and in group of N-isolates from NCC (average frequency 0.1; $p > 0.5$). Analysis of the groups of strain demonstrated that type BII was observed mostly in N-isolates from ASC ($p > 0.5$).

A comparative analysis of strains from two genetic centers showed that chromosomal type AI was typical for more than 54% strains of *S. meliloti* from NCC, i. e., 2.4 times more frequently than for strains from ASC ($\chi^2 = 18.6$; $p = 1.6 \cdot 10^{-5}$; df = 1). The indicated type of chromosome was detected mainly in subgroup of salt-tolerant T-isolates from NCC, which contains also strains harboring with chromosomal type BI, characterized by an altered sequence of marker M-III. On the contrary the population of *S. meliloti* strains from ASC more than 76% of strains had chromosomal types differ from type AI. Thus, chromosomal types AII, BI and BII were detected in 1.4, 1.7 and 3.8 times more frequently in strains from ASC than from NCC, respectively. The vast majority of the N-isolates (65%) from ASC had chromosomal type AII or BI, and 15% of salt-tolerant strains harbored the chromosomal type BII. The described differences in the occurrence of chromosomal types between geographically distinct populations of ASC and NCC were significant ($\chi^2 = 20.6$; $p = 1.2 \cdot 10^{-4}$; df = 3). Strains from agrocenoses had mostly chromosomal type AII, characterized by modified metabolic markers.

Occurrence of GI was assessed in groups of *S. meliloti* strains with certain chromosomal types (AI, AII, BI, or BII; Fig. 4). The presence of three GIs (as in the reference strain Rm1021, see Materials and methods), as well as the presence of two or one genomic islands the integration sites of which are similar to those of Rm1021 were analyzed in native isolates. Three GIs were rarely detected in strains harbored chromosomal type AI (frequency 0.04) but were more frequently observed in chromosomal types AII and BI (average frequency 0.11; see Fig. 4, *a*), but it was not a case for the type BII. The two GIs were identified more frequently (1.5 times in average) in strains with chromosomal type AI (average frequency 0.29; see Fig. 4, *a*). The one GI was observed in most strains (average frequency 0.72) but more frequently (1.2 times) in strains with chromosomal type BII (frequency 0.82; see Fig. 4, *a*).

The occurrence of particular GI of Rm1021 (Sme19T, Sme21T и Sme80S) was examined in strains with certain chromosomal types (see Fig. 4, *b*). It was shown that Sme19T and Sme80S were significantly abounded in strains with chromosomal types AI and AII ($\chi^2 = 7.4$; $p = 2.4 \cdot 10^{-3}$; df = 2; see Fig. 4, *b*), while Sme80S occurred significantly more often in strains with chromosomal type BII ($\chi^2 = 7.6$; $p = 5.9 \cdot 10^{-3}$; df = 2). Curious that either of the three genomic islands occurred with equal frequencies in strains harboring chromosomal type BI (0.23; see Fig. 4, *b*).

Thus, significant differences in occurrence of GIs related to accessory elements of the core genome were revealed in native isolates of *S. meliloti* with different chromosomal types. The results are consistent with the data for other nodule bacteria and enterobacteria [9, 26] and testified that certain accessory elements can be associated with certain combinations of core sequences.

CONCLUSION

Linked analysis of core genome marker sequences M-I and M-II, containing genes, whose products are involved in cellular metabolism and in the processes related to stress tolerance and sequences of marker M-III used in phylogenetic studies was carried out. In the

result the effects of ecological and evolutionary factors on the genetic diversity of native isolates of root nodule bacteria *S. meliloti* species recovered from distinct geographical areas and from agrocenoses were evaluated. A high level of polymorphism of sequences of metabolic (M-I and M-II) and phylogenetic (M-III) markers was detected in homogenous groups and in subgroups of isolates differed in origins and source of isolation, as well as in phenotype characteristic. An analysis of the frequencies of the occurrence of different alleles of M-I and M-II markers provide a conclusion that the genotypic diversity of strains isolated from soil (T-isolates) is under the frequency-dependent disruptive selection, which leads to the prevalence of a particular genotype that is most adapted to the agroecological conditions at the corresponding centers of genetic diversity. The high genotypical diversity of strains recovered from nodules (N-isolates) of plants widely abounded in studied genetic centers of legume diversity was due to stabilizing selection, that is agreement with the data of [49, 50].

Assessment of linkage disequilibrium (LD) between marker sequences showed there is no or low levels of genetic recombination between metabolic markers. Based on it and in accordance with the data of [10], selected markers were used to determine their combinations in *S. meliloti* native isolates. Finally strains were grouped according to identified chromosomal type, as four distinct chromosomal types were clarified according to combinations occurred between the reference and divergent alleles of the markers studied. It was shown that geographically distinct populations of *S. meliloti* were significantly differ by the occurrence of chromosomal types, that correlates with the data obtained for other rhizobia species [9, 11, 53]. Strains with altered metabolic markers or which harboured chromosomal type AII were detected in the population of each of the studied gene centers (NCC and ASC), the proportion of which among N-isolates was almost 2 times higher than among T-isolates in both gene centers. It ought to note that strains with chromosomal type AII were detected mainly in soil agrocenoses. Strains contained altered metabolic markers of the core genome can be related to divergent clonal lineage(s), that is correlates with previously obtained data on the presence of divergent chromosomal types among strains from ASC [39, 40, 51], while data concerning the presence of clonal lineage(s) in population from NCC and agrocenoses are obtained for the first time. At about 40% of both N- and T-isolates recovered from the center of introgressive hybridization of alfalfa at the Aral Sea region (ASC) subjected to extreme salinization had had altered phylogenetic marker and, in the vast majority, altered metabolic markers (chromosome types BI and BII). Strains like these could be representatives of new Sinorhizobium biovars, that is in agreement with previously published data [39].

In addition, strains belonging to the same, or closely related biovar were detected exclusively among T-isolates in NCC for the first time, however they were found twice as often among T-isolates from ASC. Clonal lineages of strains detected in both geographically distinct populations, as well as in agrocenoses, and native isolates attributed to the newly formed biovar of Sinorhizobium bacteria are objectives subjected for further molecular-genetic research.

In order to detect the dependence of the occurrence of accessory elements on the inherited characteristics of the core genome, the GIs were analyzed in genomes of strains of *S. meliloti* that differed in combination of sequences of the core genome markers.

This allowed us to determine the preference of occurrence of any genomic island in certain chromosomal types of *S. meliloti*. The obtained data support the hypothesis proposed by the authors [26] that certain chromosomal type can contain certain accessory elements.

Additional information

The work was conducted using equipment of the Core facilities center Genome technologies, proteomics and cellular biology of the All-Russia Research Institute of Agricultural Microbiology (ARRIAM) with financial support of grants of RFBR No. 18-04-01278a (assessment of structural polymorphism of marker M-II) and RSF No. 17-16-01095.

REFERENCES

1. Young JP, Crossman LC, Johnston AW, et al. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* 2006;7(4):R34. https://doi.org/10.1186/gb-2006-7-4-r34.
2. Шестаков С.В. Как происходит и чем лимитируется горизонтальный перенос генов у бактерий // Экологическая генетика. – 2007. – Т. 5. – № 2. – С. 12–24. [Shestakov SV. How does the horizontal gene transfer in bacteria occur and than is it tied up. *Ecological genetics*. 2007;5(2):12-24. (In Russ.)]. https://doi.org/10.17816/ecogen5212-24.
3. Равин Н.В., Шестаков С.В. Геном прокариот // Вавиловский журнал генетики и селекции. – 2013. – Т. 17. – № 4–2. – С. 972–984. [Ravin NV, Shestakov SV. The genome of prokaryotes. *Vavilov journal of genetics and breeding*. 2013;17(4-2):972-984. (In Russ.)]
4. Mauchline TH, Hayat R, Roberts R, et al. Assessment of core and accessory genetic variation in *Rhizobium leguminosarum symbiovar trifolii* strains from diverse locations and host plants using PCR-based methods. *Lett Appl Microbiol*. 2014;59(2):238-246. https://doi.org/10.1111/lam.12270.
5. Тихонович И.А., Андронов Е.Е., Борисов А.Ю., и др. Принцип дополнительности геномов в рас-

SUPPLEMENTARY

Table 1

Values of criterion χ^2 for subgroups of strains of *S. meliloti* different in the frequency of occurrence of RFLP types marker sequences

Marker	Group of strains			Values of χ^2 criterion								
				NCC				ASC				AC
				N		T		N		T		T
				R	S	R	S	R	S	R	S	R
M-I	NCC	N	R		–[I]	10.0 (1)[II]	–	–	–	26.2 (2)	8.0 (2)	3.9 (1)
			S	–		6.0 (1)	–	–	–	12.8 (2)	–	–
		T	R	10.0 (1)	6.0 (1)		–	12.6 (3)	15.5 (2)	50.7 (2)	–	30.4 (1)
			S	–	–	–		–	–	17.2 (3)	–	10.9 (2)
	ASC	N	R	–	–	12.6 (3)	–		–	35.6 (3)	–	8.2 (3)
			S	–	–	15.5 (2)	–	–		28.0 (3)	10.6 (3)	–
		T	R	26.2 (2)	12.8 (2)	50.7 (2)	17.2 (3)	35.6 (3)	28.0 (3)		15.0 (2)	33.5 (2)
			S	8.0 (2)	–	–	–	–	10.6 (3)	15.0		19.1 (2)
	AC	T	R	3.9 (1)	–	30.4 (1)	10.9 (2)	8.2 (3)	–	33.5	19.1 (2)	
M-II	NCC	N	R		–	8.8 (3)	–	–	–	26.1 (3)	–	14.9 (3)
			S	–		8.9 (2)	–	–	–	12.8 (2)	9.3 (2)	–
		T	R	8.8 (3)	8.9 (2)		–	–	8.0 (2)	28.9 (2)	–	23.8 (2)
			S	–	–	–		–	–	17.2 (2)	–	9.1 (2)
	ASC	N	R	–	–	–	–		–	13.1 (3)	–	8.6 (3)
			S	–	–	8.0 (2)	–	–		14.1 (2)	7.8 (2)	–
		T	R	26.1 (3)	12.8 (2)	28.9 (2)	17.2 (2)	13.1 (3)	14.1 (2)		15.0 (1)	–
			S	–	9.3 (2)	–	–	–	7.8 (2)	15.0 (1)		16.8 (2)
	AC	T	R	14.9 (3)	–	23.8 (2)	9.1 (2)	8.6 (3)	–	–	16.8 (2)	
M-III	NCC	N	R		–	–	–	–	–	–	–	–
			S	–		–	–	–	–	–	–	–
		T	R	–	–		–	8.4 (2)	13.2 (2)	–	–	7.0 (2)
			S	–	–	–		–	6.1 (2)	–	–	–
	ASC	N	R	–	–	8.4 (2)	–		–	–	–	13.9 (2)
				–	–	13.2 (2)	6.1 (2)	–		–	8.1 (2)	10.2 (2)
		T	R	–	–	–	–	–	–		–	8.6 (2)
			S	–	–	–	–	–	8.1 (2)	–		12.2 (2)
	AC	T	R	–	–	7.0 (2)	–	13.9 (2)	10.2 (2)	8.6 (2)	12.2 (2)	

Note. R – strains of salt-tolerant phenotype; S – strains of salt-sensitive phenotype; N – strains recovered from nodules; T – strains trapped from soil; NCC – North Caucasian genetic center; ASC – Aral Sea center of alfalfa diversity; AC – agrocenoses. [I] – value of χ^2 less than the critical one (differences are not significant); (..)[II] – the number of degrees of freedom (df).

Table 2

The number of detected combinations of markers M-I, M-II and M-III type in subgroups and groups of *S. meliloti*

Geographical origin			NCC				ASC				AC		Total number of strains
Group of strains			N		T		N		T		T		
Phenotypical subgroup			R	S	R	S	R	S	R	S	R	S	
Statistical properties	sub-groups	Total number of combinations	6	3	9	4	20	11	4	5	4	2	35
		Unique combinations	4	1	6	1	15	6	2	3	3	2	
		Nei index (*N*)	0.76	0.86	0.68	0.84	0.95	0.94	0.75	0.79	0.90	–	
		Shannon index (*H*)	1.28	0.96	1.34	1.16	2.67	2.13	1.07	1.23	1.18	–	
		Evenness index (*E*)	0.60	0.87	0.43	0.80	0.72	0.77	0.73	0.68	0.82	–	
		Total number of strains	24	7	51	13	41	25	8	15	31	3	218
	groups	Total number of combinations	7		10		26		7		5		35
		Unique combinations	3		6		23		4		0		
		Nei index (*N*)	0.76		0.69		0.94		0.88		0.87		
		Shannon index (*H*)	1.34		1.37		2.74		1.60		1.29		
		Evenness index (*E*)	0.54		0.39		0.59		0.70		0.72		
		Total number of strains	31		64		66		23		34		218

Note. R – strains of salt-tolerant phenotype; S – strains of salt-sensitive phenotype; N – strains recovered from nodules; T – strains trapped from soil; NCC – North Caucasus genetic center; ASC – Aral Sea center of alfalfa diversity; AC – agrocenoses.

ширении 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.

6. Chidebe IN, Jaiswal SK, Dakora FD. Distribution and phylogeny of microsymbionts associated with cowpea (*Vigna unguiculata*) nodulation in three agroecological regions of *Mozambique*. *Appl Environ Microbiol*. 2018;84(2). pii:e01712-17. https://doi.org/10.1128/AEM.01712-17.
7. Jiao J, Ni M, Zhang B, et al. Coordinated regulation of core and accessory genes in the multipartite genome of *Sinorhizobium fredii*. *PLoS Genet*. 2018;14(5):e1007428. https://doi.org/10.1371/journal.pgen.1007428.
8. Boussau B, Karlberg EO, Frank AC, et al. Computational inference of scenarios for alpha-proteobac-

terial genome evolution. *Proc Natl Acad Sci USA*. 2004;101(26):9722-9727. https://doi.org/10.1073/pnas.0400975101.
9. Wang X, Liu D, Luo Y, et al. Comparative analysis of rhizobial chromosomes and plasmids to estimate their evolutionary relationships. *Plasmid*. 2018;96-97:13-24. https://doi.org/10.1016/j.plasmid.2018.03.001.
10. Young JP, Wexler M. Sym plasmid and chromosomal genotypes are correlated in field populations of *Rhizobium leguminosarum*. *J Gen Microbiol*. 1988;134:2731-9. https://doi.org/10.1099/00221287-134-10-2731.
11. Stefan A, van Cauwenberghe J, Rosu CM, et al. Genetic diversity and structure of *Rhizobium leguminosarum* populations associated with clover plants are influenced by local environmental variables. *Syst Appl Microbiol*. 2018;41(3):251-259. https://doi.org/10.1016/j.syapm.2018.01.007.
12. Van Berkum P, Badri Y, Elia P, et al. Chromosomal and symbiotic relationships of rhizobia nodulating *Medicago truncatula* and *M. laciniata*. *Appl Environ Microbiol*. 2007;73(23):7597-7604. https://doi.org/10.1128/AEM.01046-07.
13. Проворов Н.А., Андронов Е.Е., Онищук О.П., и др. Генетическая структура интродуцированных и местных популяций *Rhizobium leguminosarum* в системах «растения–почва» // Микробиология. – 2012. – Т. 81. – № 2. – С. 244–253. [Provorov NA, Andronov EE, Onishchuk OP, et al. Genetic structure of the introduced and local populations of *Rhizobium leguminosarum* in plant-soil systems. *Microbiology*. 2012;81(2):224-232. (In Russ.)]
14. Laguerre G, Mavingui P, Allard M-R, et al. Typing of rhizobia by PCR DNA fingerprinting and PCR-restriction fragment length polymorphism analysis of chromosomal and symbiotic gene regions: application to *Rhizobium leguminosarum* and its different biovars. *Appl Environ Microbiol*. 1996;62(6):2029-2036.
15. Guo XW, Zhang XX, Zhang ZM, Li FD. Characterization of *Astragalus sinicus* rhizobia by restriction fragment length polymorphism analysis of chromosomal and nodulation genes regions. *Curr Microbiol*. 1999;39(6):358-364. https://doi.org/10.1007/s002849900472.
16. Efrose RC, Rosu CM, Stedel C, et al. Molecular diversity and phylogeny of indigenous *Rhizobium leguminosarum* strains associated with *Trifolium repens* plants in Romania. *Antonie Van Leeuwenhoek*. 2018;111(1):135-153. https://doi.org/10.1007/s10482-017-0934-3.
17. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 2015;43:D261-269. https://doi.org/10.1093/nar/gku1223.
18. Zhang YM, Tian CF, Sui XH, et al. Robust markers reflecting phylogeny and taxonomy of rhizobia. *PLoS One*. 2012;7(9):e44936. https://doi.org/10.1371/journal.pone.0044936.
19. Guo HJ, Wang ET, Zhang XX, et al. Replicon-dependent differentiation of symbiosis-related genes in *Sinorhizobium* strains nodulating *Glycine max*. *Appl Environ Microbiol*. 2014;80(4):1245-1255. https://doi.org/10.1128/AEM.03037-13.
20. Alexandre A, Laranjo M, Young JP, Oliveira S. *dnaJ* is a useful phylogenetic marker for alphaproteobacteria. *Int J Syst Evol Microbiol*. 2008;58(12):2839-2849. https://doi.org/10.1099/ijs.0.2008/001636-0.
21. Biondi EG, Pilli E, Giuntini E, et al. Genetic relationship of *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains isolated from Caucasian region. *FEMS Microbiol Lett*. 2003;220(2):207-213. https://doi.org/10.1016/S0378-1097(03)00098-3.
22. Escobar-Páramo P, Sabbagh A, Darlu P, et al. Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. *Mol Phylogenet Evol*. 2004;30(1):243-250. https://doi.org/10.1016/S1055-7903(03)00181-7.
23. Matzke NJ. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst Biol*. 2014;63(6):951-970. https://doi.org/10.1093/sysbio/syu056.
24. Dresler-Nurmi A, Fewer DP, Räsänen LA, Lindström K. The diversity and evolution of Rhizobia. In: Pawlowski K. (eds). Prokaryotic symbionts in plants. Springer-Verlag; 2009. P. 3-41.
25. Tounsi-Hammami S, Le Roux C, Dhane-Fitouri S, et al. Genetic diversity of rhizobia associated with root nodules of white lupin (*Lupinus albus* L.) in Tunisian calcareous soils. *Syst Appl Microbiol*. 2019;42(4):448-456. https://doi.org/10.1016/j.syapm.2019.04.002.
26. Escobar-Páramo P, Clermont O, Blanc-Potard AB, et al. Specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol Biol Evol*. 2004;21(6):1085-1094. https://doi.org/10.1093/molbev/msh118.
27. Проворов Н.А. Симбиогенез как эволюция генетических систем открытого типа // Генетика. – 2018. – Т. 54. – № 8. – С. 879–889. [Provorov NA. Symbiogenesis as evolution of open genetic systems. *Russian Journal of Genetics*. 2018;54(8):888-896. (In Russ.)]. https://doi.org/10.1134/S0016675818080106.
28. Румянцева М.Л., Мунтян В.С., Черкасова М.Е., и др. Геномные острова штамма *Sinorhizobium meliloti* Rm1021 — азотфиксирующего симбионта люцерны // Генетика. – 2018. – Т. 54. – № 7. – С. 745-756. [Roumiantseva ML, Muntyan VS, Cherkasova ME, et al. Genomic islands in *Sinorhizobium meliloti* Rm1021, nitrogen-fixing symbiont of alfalfa. *Russian Journal of Genet-*

ics. 2018;54(7):759-769. (In Russ.)]. https://doi.org/10.1134/S0016675818070135.

29. Barcellos FG, Menna P, da Silva Batista JS, Hungria M. Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (*Ensifer*) *fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. *Appl Environ Microbiol*. 2007;73(8):2635-2643. https://doi.org/10.1128/AEM.01823-06.
30. Wielbo J. Rhizobial communities in symbiosis with legumes: genetic diversity, competition and interactions with host plants. *Cent Eur J Biol*. 2012;7(3):363-372. https://doi.org/10.2478/s11535-012-0032-5.
31. Румянцева М.Л., Мунтян В.С., Черкасова М.Е., и др. Сравнительный анализ геномных характеристик у референтных штаммов *Sinorhizobium meliloti* — симбионтов люцерны // Сельскохозяйственная биология. — 2017. — Т. 52. — № 5. — С. 928-939. [Roumiantseva ML, Muntyan VS, Cherkasova ME, et al. A comparative analysis of genomic characters of reference *Sinorhizobium meliloti* strains, the alfalfa symbionts. *Agricultural Biology*. 2017;52(5):928-939. (In Russ.)]. https://doi.org/10.15389/agrobiology.2017.5.928rus.
32. Che D, Hasan MS, Chen B. Identifying pathogenicity islands in bacterial pathogenomics using computational approaches. *Pathogens*. 2014;3(1):36-56. https://doi.org/10.3390/pathogens3010036.
33. Krogh TJ, Møller-Jensen J, Kaleta C. Impact of chromosomal architecture on the function and evolution of bacterial genomes. *Front Microbiol*. 2018;9:2019. https://doi.org/10.3389/fmicb.2018.02019.
34. Dobrindt U, Hochhut B, Hentschel U, Hacker J. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*. 2004;2(5):414-424. https://doi.org/10.1038/nrmicro884.
35. Juhas M, van der Meer JR, Gaillard M, et al. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev*. 2009;33(2):376-393. https://doi.org/10.1111/j.1574-6976.2008.00136.x.
36. Мунтян В.С., Черкасова М.Е., Андронов Е.Е., и др. Встречаемость островов в геномах природных штаммов *Sinorhizobium meliloti* // Генетика. — 2016. — Т. 52. — № 10. — С. 1126—1133. [Muntyan VS, Cherkasova ME, Andronov EE, et al. Occurrence of islands in genomes of *Sinorhizobium meliloti* native isolates. *Russian Journal of Genetics*. 2016;52(10):1015-1022. (In Russ.)]. https://doi.org/10.7868/S0016675816080105.
37. Hudson CM, Lau BY, Williams KP. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res*. 2015;43 (Database issue):D48-53. https://doi.org/10.1093/nar/gku1072.
38. Румянцева М.Л., Симаров Б.В., Онищук О.П., и др. Биологическое разнообразие клубеньковых бактерий в экосистемах и агроценозах. Теоретические основы и методы / Под ред. М.Л. Румянцевой, Б.В. Симарова. — СПб.; Пушкин: Инновационный центр защиты растений, 2011. — 104 с. [Roumiantseva ML, Simarov BV, Onishchuk OP, et al. Biologicheskoe raznoobrazie kluben'kovykh bakterii v ekosistemakkh i agrotsenozakh. Teoreticheskie osnovy i metody. Ed. by M.L. Rumyantseva, B.V. Simarov. Saint Petersburg; Pushkin: Innovatsionnyi tsentr zashchity rastenii; 2011. 104 p. (In Russ.)]
39. Румянцева М.Л., Мунтян В.С., Менгони А., Симаров Б.В. ITS-полиморфизм солеустойчивых и солечувствительных природных штаммов *Sinorhizobium meliloti* — симбионтов люцерны, донника и пажитника // Генетика. — 2014. — Т. 50. — № 4. — С. 400—412. [Roumiantseva ML, Muntian VS, Mengoni A, Simarov BV. ITS-polymorphism of salt-tolerant and salt-sensitive native isolates of *Sinorhizoblum meliloti* — symbionts of alfalfa, clover and fenugreek plants. *Russian Journal of Genetics*. 2014;50(4):348-359. (In Russ.)]. https://doi.org/10.7868/S0016675814040109.
40. Румянцева М.Л., Мунтян В.С. Клубеньковые бактерии *Sinorhizobium meliloti*: солеустойчивость и ее генетическая детерминированность // Микробиология. — 2015. — Т. 84. — № 3. — С. 263—280. [Roumiantseva ML, Muntyan VS. Root nodule bacteria *Sinorhizobium meliloti*: tolerance to salinity and bacterial genetic determinants. *Microbiology*. 2015;84(3):303-318. (In Russ.)]. https://doi.org/10.7868/S0026365615030179.
41. Демидёнок О.И., Гончаренко А.В. Системы токсин-антитоксин бактерий и перспективы их использования в медицине (обзор) // Прикладная биохимия и микробиология. — 2013. — Т. 49. — № 6. — С. 539—546. [Demidenok OI, Goncharenko AV. Bacterial toxin-antitoxin systems and perspectives for their application in medicine. *Applied Biochemistry and Microbiology*. 2013;49(6):535-541. (In Russ.)]. https://doi.org/10.7868/S055510991306007X.
42. Rannala B, Qiu W-G, Dykhuizen DE. Methods for estimating gene frequencies and detecting selection in bacterial populations. *Genetics*. 2000;155(2):499-508.
43. Rasmussen HB. Restriction fragment length polymorphism analysis of PCR-amplified fragments (PCR-RFLP) and gel electrophoresis — valuable tool for genotyping and genetic fingerprinting. *Gel Electrophoresis — Principles and Basics*. 2012. P. 315-334. https://doi.org/10.5772/37724.

44. Румянцева М.Л., Белова В.С., Онищук О.П., и др. Полиморфизм *bet*-генов у штаммов *Sinorhizobium meliloti* из генцентров люцерны // Сельскохозяйственная биология. – 2011. – Т. 46. – № 3. – С. 48–54. [Roumiantseva ML, Belova VS, Onishchouk OP, et al. Polymorphism of *bet*-genes among *Sinorhizobium meliloti* isolates native to gene centers of alfalfa. *Agricultural Biology*. 2011;46(3):48-54. (In Russ.)]
45. Маниатис Т., Фрич Э., Сэмбрук Д. Методы генетической инженерии. Молекулярное клонирование / Пер. с англ. под ред. А.А. Баева, К.Г. Скрябина. – М.: Мир, 1984. – 479 с. [Maniatis T, Fritch EE, Sambrook J. Molecular cloning: a laboratory manual. Translated from English ed. by A.A. Baev, K.G. Skryabin. Moscow: Mir; 1984. 479 p. (In Russ.)]
46. Hammer O, Harper DAT, Ryan PD. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 2001;4(1):1-9.
47. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978;89(3):583-590.
48. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.
49. Проворов Н.А., Андронов Е.Е., Онищук О.П. Формы естественного отбора, определяющего геномную эволюцию клубеньковых бактерий // Генетика. – 2017. – Т. 53. – № 4. – С. 401–410. [Provorov NA, Andronov EE, Onishchuk OP. Forms of natural selection controlling the genomic evolution in nodule bacteria. *Russian Journal of Genetics*. 2017;53(4):411-419. (In Russ.)]. https://doi.org/10.7868/S0016675817040129.
50. Levin BR. Frequency dependent selection in bacterial populations. *Philos Trans R Soc Lond B Biol Sci*. 1988;319(1196):459-472. https://doi.org/10.1098/rstb.1988.0059.
51. Румянцева М.Л., Саксаганская А.С., Мунтян В.С., и др. Структурный полиморфизм генов вирулентности и солеустойчивости *Sinorhizobium meliloti* // Генетика. – 2018. – Т. 54. – № 5. – С. 524–534. [Roumiantseva ML, Saksaganskaia AS, Muntyan VS, et al. Structural polymorphism of *Sinorhizobium meliloti* genes related to virulence and salt tolerance. *Russian Journal of Genetics*. 2018;54(5):525-535. (In Russ.)]. https://doi.org/10.7868/S001667581805003X.
52. Румянцева М.Л., Онищук О.П., Белова В.С., и др. Полиморфизм штаммов *Sinorhizobium meliloti*, выделенных в центрах разнообразия люцерны, различающихся по почвенно-климатическим условиям // Экологическая генетика. – 2009. – Т. 7. – № 2. – С. 19–25. [Roumiantseva ML, Onischuk OP, Belova VS, et al. Polymorphism among *Sinorhizobium meliloti* isolates native to the origins of alfalfa diversity differed in soil-climate characteristics. *Ecological genetics*. 2009;7(2):19-25. (In Russ.)]. https://doi.org/10.17816/ecogen7219-25.
53. Palmer KM, Young JP. Higher diversity of *Rhizobium leguminosarum* biovar viciae populations in arable soils than in grass soils. *Appl Environ Microbiol*. 2000;66(6):2445-2450. https://doi.org/10.1128/AEM.66.6.2445-2450.2000.

❀ Authors and affiliations

Mariia E. Cherkasova – Research Engineer, Laboratory of Genetics and Breeding of Microorganisms. All-Russian Research Institute for Agricultural of Microbiology, Pushkin-8, St. Petersburg, Russia. SPIN: 5341-5736. E-mail: mariiacherkasova@mail.ru.

Victoria S. Muntyan – Junior Researcher, Laboratory of Genetics and Breeding of Microorganisms. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. SPIN: 7138-6763. E-mail: vucovar@yandex.ru.

Alla S. Saksaganskaia – Research Engineer, Laboratory of Genetics and Breeding of Microorganisms. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. SPIN: 5832-1676. E-mail: allasaksaganskaya@mail.ru.

Boris V. Simarov – Doctor of Biology, Professor, Principal Researcher, Laboratory of Genetics and Breeding of Microorganisms. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. SPIN: 6859-1141. E-mail: genet@yandex.ru.

❀ Информация об авторах

Мария Евгеньевна Черкасова — инженер-исследователь, лаборатория генетики и селекции микроорганизмов. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. SPIN: 5341-5736. E-mail: mariiacherkasova@mail.ru.

Виктория Спартаковна Мунтян — младший научный сотрудник, лаборатория генетики и селекции микроорганизмов. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. SPIN: 7138-6763. E-mail: vucovar@yandex.ru.

Алла Сергеевна Саксаганская — инженер-исследователь, лаборатория генетики и селекции микроорганизмов. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. SPIN: 5832-1676. E-mail: allasaksaganskaya@mail.ru.

Борис Васильевич Симаров — д-р. биол. наук, профессор, главный научный сотрудник, лаборатория генетики и селекции микроорганизмов. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. SPIN: 6859-1141. E-mail: genet@yandex.ru.

❀ Authors and affiliations

Marina L. Roumiantseva – PhD (Candidate of Biology), Leading Researcher, Head of the Laboratory, Laboratory of Genetics and Breeding of Microorganisms. All-Russian Research Institute for Agricultural Microbiology, Pushkin-8, St. Petersburg, Russia. SPIN: 5470-9527. E-mail: mroumiantseva@yandex.ru.

❀ Информация об авторах

Марина Львовна Румянцева — канд. биол. наук, ведущий научный сотрудник, заведующая лабораторией, лаборатория генетики и селекции микроорганизмов. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин-8, Санкт-Петербург. SPIN: 5470-9527. E-mail: mroumiantseva@yandex.ru.