

THE CONCEPT OF TAXONOMIC SPACE AND INTEGRAL ESTIMATES OF SHIFT IN THE STRUCTURE OF MICROBIAL COMMUNITY BASED ON ANALYSIS OF 16S rRNA GENE LIBRARIES

E.V. Pershina¹, A.S. Dol'nik², A.G. Pinaev¹, K.A. Loshakova¹, E.E. Andronov¹

¹All-Russian Research Institute for Agricultural Microbiology, Russian Academy of Agricultural Sciences,
3, sh. Podbel'skogo, Pushkin-8, St. Petersburg, 196608 Russia,
e-mail: eeandr@gmail.com;

²Saint Petersburg State University,
28, Universitetskii prosp., Staryi Petergof, St. Petersburg, 199034 Russia,
e-mail: alexander.dolnik@gmail.com

Received September 25, 2012

Summary

The problem of taxonomic structure and dynamics of soil, plant animal and human microbiomes is one of the most intriguing in modern microbiology. High Performance Technologies sequencing of the 16S rRNA gene allows to get much more metagenomic data, but their correct analysis and biological interpretation are still complicated, in particular with regard to the effect of selective amplification with universal primers and proper attribution of the samples. To resolve the problems, we created a special operating environment, the taxonomic space (TS), in which the sequences of 16S rRNA gene are represented by dots, geometric distance between which corresponds to the genetic distance between the sequences. Mapping the 16S-rRNA gene biodiversity data in this TS and evaluation of the microbial community as overorganism, with its integral parameters, have a number of advantages if compared to the traditional approaches. Thus, in the TS where each sequence of the 16S rRNA gene gets its own identifier of the 42 coordinates, the unattributed amplicons in any PCR-library can be analyzed. Although the described TS is not strictly a multi-dimensional mathematical space, in particular, its axes are interdependent, an extremely high correlation coefficients, obtained for genetic distances between sequences and their geometric counterparts, unconditionally, testify in favor of the validity of the use of TS in practice. The development of TS concept is of great importance not only in the analysis of the structure of microbial communities, but also in investigation of 16S rRNA genes evolution. Since the model allows to give a description for any variant, both realized and not yet realized in evolution, the issues related to the origin and divergent evolution of prokaryotes may be investigated, for example, the hypothetical center of origin can be determine, and then the TS will become an evolutionary space. As a model, we used different soil microbiomes in which the changes were induced by environmental conditions (salinity), both natural and simulated. However, the application of this approach can be extended to other complex microbiomes, particularly the microbiota in animals. Moreover, the proposed mathematical method is universal and can be used to study not only biodiversity in prokaryotes, but also the communities of eukaryotic organisms, including animals and plants, with the 18S rRNA gene as a taxonomic marker.

Keywords: microbiom, soil, salinization, 16S rRNA, taxonomic space.

Biodiversity of microbial communities is investigated by different methods the most popular of which is polymorphism of 16S-rRNA genes (1). The technology of high-performance sequencing has provided the great amount of experimental data on biodiversity allowing studying the structure and dynamics of microbial communities (2, 3). However, this has led to some problems related to obtaining, analysis, and biological interpretation of metagenomic data (4). This work reviews two of them.

The first problem is caused by the effect of selective amplification of 16S-rRNA sequences during multimatrix PCR with universal primers (5, 6). The most significant effect is observed in ecosystems with substantial proportion of minor groups of microorganisms (such as soil community), which may lead to PCR-dependent distortions in the structure of microbiome, as well as complete absence of particular nucleotide sequences in PCR product. For example, G.T. Bergmann et al. (7) showed that relatively small affinity of modern universal primers to sequences of the bacterial phyla *Verrucomicrobia* has contributed to underestimation of their abundance in the soil (7). Despite the fact that the problem of selective primer-dependent amplification of 16S-rRNA molecules to some extent affects any modern research, there hasn't yet been suggested any effective solution.

The second problem regards the analysis of taxonomic structure of microbial communities and its biological interpretation. The main difficulty of metagenomic data is associated with significant biodiversity of microbiomes and the presence of taxonomically non-attributable microorganisms. In various estimates, microorganisms not identified at both levels of genus and higher taxa (up to phyla) amount to more than a half of total pool of microorganisms in the community (8). The presence of such elements complicates comparison of microbiomes composition in independent experiments and, in particular, limits the use of molecular techniques in studies including analysis of multiple samples (microbiological soil mapping, monitoring, biogeography of microbiomes, etc.).

To address these problems, the authors have developed the special operational environment for data analysis of nucleotide sequences – taxonomic space (TS) for 16S-rRNA gene. TS is a metric multidimensional space in which 16S-rRNA gene sequences are represented by points, geometric distances between these points correspond to genetic distances between the sequences (percentage of different nucleotide positions). The first version of this space was described previously (9), though the methods used in this version didn't provide high correlation coefficients between genetic distances and their geometric analogs ($r \approx 0.3$).

The present study demonstrates an alternative approach to construction of taxonomic space (TS). The first part of this work shows application of TS for assessing the selective primer-dependent PCR amplification exemplified by three pairs of universal primers for 16S-rRNA gene, the second part is using TS for evaluation of structure and dynamics of microbial communities (in saline soils).

Technique. Assessing the effect of selective primer-dependent PCR amplification, the model object was a sample of sod-podzolic soil collected in July 2007 near the village Belogorka (Leningrad region) from soil horizon A1. In summer 2009, six soil samples were collected along the salinity gradient (T1, T2, T2-3, T3, T4, T5; salt content in T1 – 1.23 %, in T5 – 0.01%) in the region of natural soil salinization (Shingirlau, Kazakhstan) investigated by a scientific expedition. Three other samples of dark brown soil typical for the region (1 – virgin soil NC, 2 and 3 – fallow soil NZ and NK) were collected in the area located 200 km away from the abovementioned region (10). The experiment with induced soil salinization was conducted in containers filled with NK soil whose salinity was changed by introducing the solution with composition similar to that of the sample T1 (per 100 g soil: Cl⁻ — 8 mM, SO₄²⁻ — 12 mM, Na⁺ — 7 mM, K⁺ — 0.5 mM, Mg²⁺ — 2.5 mM, Ca²⁺ — 10 mM) to a final concentration of 3 % (w/w). Calcium

salts and sulphates were introduced separately. The experiment was performed in two technical replicates (control and experiment). In the containers there was maintained constant soil humidity (60 % of full moisture capacity). Samples of the resulting soil were collected from three equidistant points (from each other and from walls of the container) and mixed in equal amounts at the stage of DNA extracts. DNA was isolated from a soil sample (0.2 g) as described (11).

The effect of selective primer-dependent amplification was assessed using the data of amplicon libraries of 16S-rRNA gene created with three pairs of universal PCR primers: fD1/rD1 (27f: 5'-AGAGTTTGATCCTGGCTC-AG-3', 1525r: 5'-AAGGAGGTGATCCAGCC-3') (12); fBD1/rBD1 (642f: 5'-HAATHYGTGCCAGCAGC-3', 1445r: 5'-GTCRTCCYDCCTCCTC-3') (13) и Eu3 (63f: 5'-AGGCCTAACACATGCAAGTC-3', 1494r: 5'-TACGGYTA-CCTTGTTACGAC-3') (14). Amplification was conducted under a standard technique (11), resulting fragments were cloned in pAL-TA vector ("Evrogen", Russia) according to the manufacturer's protocol, the obtained constructs were used to transform competent cells of *Escherichia coli* (DH10B) (15). For each pair of primers there was created a library of 16S-rRNA gene fragments (L1 — Eu3, L2 — fD1/rD1, L3 — fBD1/rBD1). The nucleotide sequence was determined using a primer FGPS (485-292) 5'-CAGCAGCCGCGGTAA-3' (16) in an automatic sequencer SEQ8000 using the protocol recommended by the manufacturer of reagents ("Beckman Culter", USA). Alignment of sequences was performed in the program Clustal X, matrixes of genetic distances and phylogenetic trees were constructed in the program MEGA 5. Taxonomic identification of sequences was carried out on a server RDPII (17).

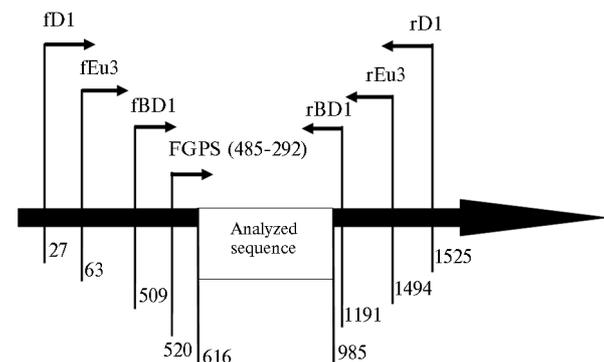


Fig. 1. Position of three pairs of universal primers relative to nucleotide sequence of 16S-rRNA gene of *Escherichia coli* (strain K12).

universal primers for 16S-rRNA gene used in creation of the libraries.

In each amplicon library, there was analyzed almost equal amount of sequences: 33 – in L1, 29 – in L2, 33 – in L3. Among these nucleotide sequences were identified representatives of the bacterial phyla *Proteobacteria* (32 %), *Acidobacteria* (26 %), *Verrucomicrobia* (7 %), *Actinobacteria* (4 %), *Bacteroidetes* (3 %), *Planctomycetes* (3 %), *Chlamydiae* (3%), *Firmicutes* (2%); these data were deposited in GenBank of NCBI (the National Center for Biotechnology Information) with the assigned identifiers HQ412669-HQ412763.

To create libraries of 16S-rRNA gene of microbial communities of saline soils, the samples were subject to amplification with universal primers F515 and R806 to the variable regions V4 (18). The amplification products were pyrosequenced on GS Junior ("Roche", Switzerland) according to the manufacturer's recommendations; the sequences were analyzed in the program QIIME v. 1.5.0 (19). The resulting pool of nucleotide sequences was treated: removed service sequences, filtered, aligned, constructed matrices of genetic distances, determined taxonomic identity by means of default parameters, and performed the cluster analysis with the algorithm of "unweighted unifrac".

The significance of differences was assessed by Student's t-test and Fisher's exact test.

Results. Analysis of taxonomic structure of amplicon libraries. Libraries of 16S-rRNA gene obtained through the use of different universal primers. Figure 1 shows location of three pairs of

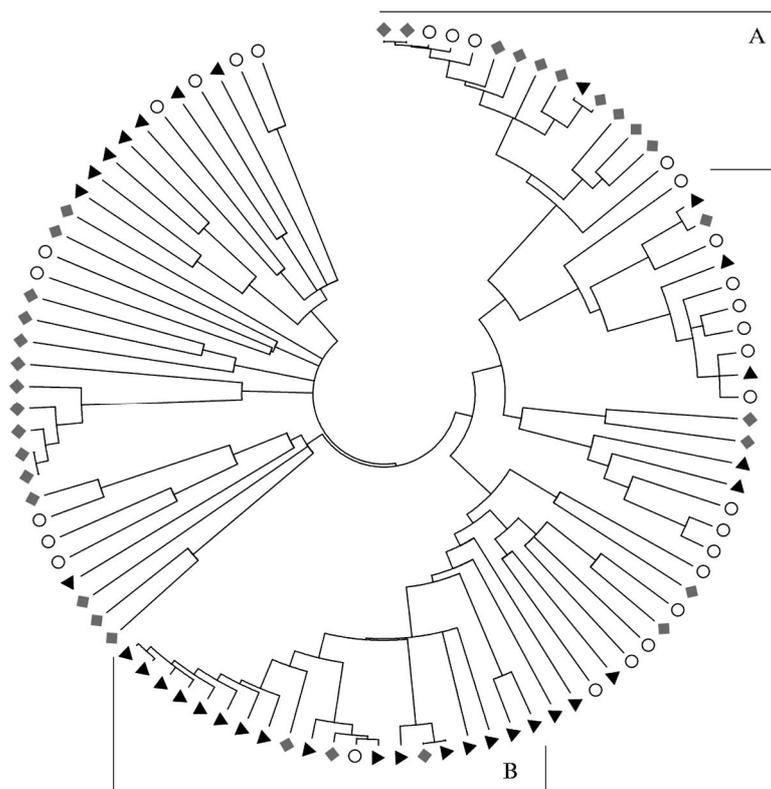


Fig. 2. Phylogenetic associations of 16S-rRNA gene sequences from the amplicon libraries L1 (◆), L2 (○), L3 (▲) identified using three pairs of universal primers: A — sequences from the phylum *Proteobacteria*, B — sequences from the phylum *Actinobacteria*.

The constructed phylogenetic tree showed uneven distribution of sequences from the three libraries about the major

prokaryotic taxa (Fig. 2). Thus, the phylum *Proteobacteria* was represented mainly by nucleotide sequences from L1 library most of which were members of the order *Rhizobiales*. In the phylum *Acidobacteria*, on the contrary, most of the sequences were representatives of libraries L2 and L3. Moreover, these L2 sequences were clustered mainly in Gp3 and Gp2 groups of *Acidobacteria*, while L3 sequences – in group Gp1. The phyla *Verrucomicrobia* and *Chlamydiae* also showed uneven distribution of nucleotide sequences (Table 1).

1. Taxonomic identification of 16S-rRNA gene sequences from investigated amplicon libraries L1, L2, L3 obtained using three pairs of universal primers

Phylum	L1	L2	L3
<i>Proteobacteria</i>	18	8	7
<i>Acidobacteria</i>	4	13	11
<i>Verrucomicrobia</i>	–	1	6
<i>Actinobacteria</i>	4	–	–
<i>Bacteroidetes</i>	2	–	1
<i>Planctomycetes</i>	–	2	1
<i>Chlamydiae</i>	–	–	3
<i>Firmicutes</i>	2	–	–

Note. Dashes – no representatives identified.

the phyla *Firmicutes* and *Bacteroidetes*, and the families (*Balneolaceae* and *Bacillaceae*) (Fig. 3).

The obtained data clearly show the effect of primer-dependent amplification manifested at the level of phyla.

Structure and dynamics of microbial communities in saline soils. Salinization is one of the most significant environmental factors (20). The dynamics of soil microbiome in conditions of natural salinization had been described previously (10). This work was focused on a brief comparative analysis of the structure of soil microbial communities under natural and induced salinization. In all these samples there were identified representatives of 21 bacterial phyla, the most abundant of which were *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. Comparing the samples of natural and “artificial” saline soil, there were revealed some common patterns of changes in structure of these communities: in both cases, there increased the proportion of bacteria from

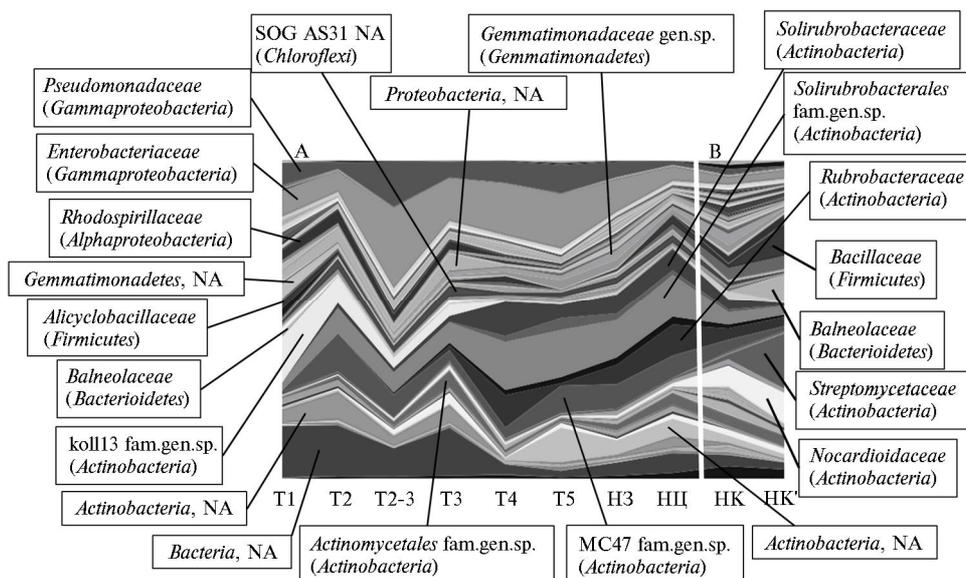


Fig. 3. Taxonomic diversity of soil microbial communities (at the level of family) in conditions of natural (A, T1- NC) and induced salinization (B, NK-NK') identified by sequencing of amplicon libraries of 16S-rRNA gene: T1-T5 — samples of saline soil (natural salinization); NZ, NC, NK — dark brown soil (NC — virgin land, NZ and NK — fallow), NK' — sample of NK after induced salinization with salt composition similar to the variant T1; NA — sequences not attributable to orders.

In these communities, there were also observed successive saline-determinate changes: instead of typical soil *Actinobacteria* (*Rubrobacteriaceae* and *Solirubrobacteriaceae*) there appeared saline-resistant groups of *Actinobacteria* (in natural saline soil – the bacteria not identifiable at the level of family, in the soil with induced salinity – representatives of the families *Nocardioideaceae* and *Streptomycetaceae*) (Fig. 3).

Along with common regularities in development of natural and induced salinization, these soils had some notable differences. Induced salinization caused a significant decrease in diversity of the soil microbial community. In this case, the phyla *Bacteroidetes* and *Firmicutes* were represented only by the two families – *Bacillaceae* and *Balneolaceae*. Salinization also caused a sharp reduce in number of *Actinobacteria* families (Fig. 3, B). On the contrary, in the most saline sample of natural saline soil there was found much greater variety of representatives of these phyla (Fig. 3, A). The dendrogram also shows significant differences in structure of the compared microbiomes – samples T1 and NK' appear in different clusters (Fig. 4).

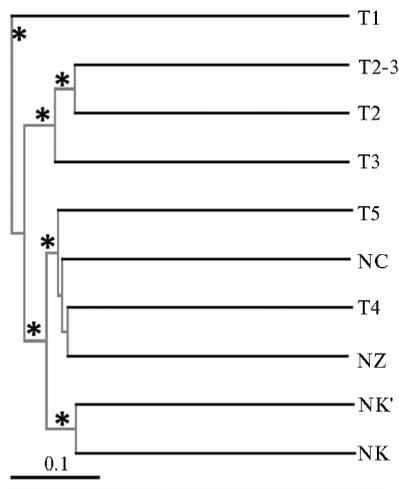


Fig. 4. Cluster dendrogram of soil microbial communities in conditions of natural and induced soil salinization (constructed from the sequencing data of amplicon libraries of 16S-rRNA gene): T1-T5 — samples of saline soil (natural salinization); NZ, NC, NK — dark brown soil (NC — virgin land, NZ and NK — fallow), NK' — sample of NK after induced salinization with salt composition similar to the variant T1. Clustering algorithm “unweighted unfrac”; asterisks label the clusters with reliability more than 80%.

Taxonomic space (TS) and selection of optimum coordinates. A special designed multi-dimensional space – taxonomic space of 16S-rRNA gene (TS) – was applied to assess the effect of selective amplification. Earlier (9), situation of nucleotide sequences in TS had been determined in coordinates of a regular simplex. This work describes coordinates of TS as a system of reference points -

	i	j	A	B	C
i	0	d_{ij}	d_{iA}	d_{iB}	d_{iC}
j		0	d_{jA}	d_{jB}	d_{jC}
A			0	d_{AB}	d_{AC}
B				0	d_{BC}
C					0

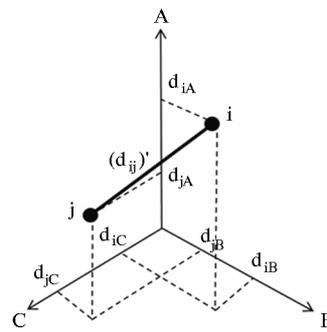


Fig. 5. Conversion of genetic distances into taxonomic distances: i, j — nucleotide sequences of 16S-rRNA gene from the analyzed amplicon library; A, B, C — nucleotide sequences from the database of Ribosomal Database Project (RDP) II selected for construction of taxonomic space (reference points), d — genetic distances between nucleotide sequences (p-distance — proportion of different nucleotide positions in sequences, %), d_i — geometric analogs of genetic distances).

- sequences of 16S-rRNA gene of particular representatives of major bacterial and archaeal taxa. The position of point in TS was set by conversion of genetic distances (proportion of different nucleotide positions, expressed as p-distance) from reference points into geometric coordinates. Thus it became possible to calculate geometric analogs (d_{ij}') for genetic distances (d_{ij}) (Fig. 5).

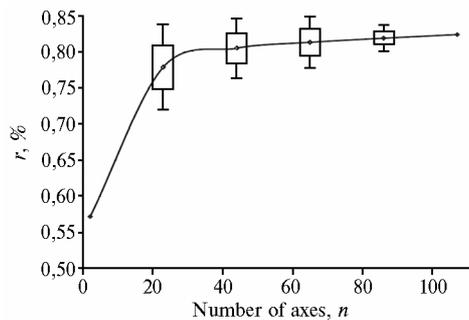


Fig. 6. Correlations between the mean value of correlation coefficient and the number of selected “coordinate axes” — 16S-rRNA gene sequences (total 107) with different abundance of major bacterial and archaeal phyla.

To determine the optimum set of coordinates of TS, there were selected 107 sequences from the database of Ribosomal Database Project II (the largest database of 16S-rRNA gene) considering representativeness among the major bacterial and archaeal phyla. Upon these data then were computed correlation coefficients (geometrical distances) between matrices of genetic distances (d_{ij}) and their analogs calculated in TS (d_{ij}') for all possible sets of coordinates (C_n^{107} , where $n = 2, \dots, 107$). The highest correlation between d_{ij} and d_{ij}' involved a relatively small number of coordinates – from 20 to 45 (Fig. 6). Further constructions were carried out using 42 sequences with best r values. The selected sequences were evenly distributed among 23 bacterial and 2 archaeal phyla; most of the sequences were representatives of the major bacterial phyla: *Proteobacteria* (7 sequences), *Firmicutes* (4), *Actinobacteria* (3), and *Acidobacteria* (3). This list also included members of the phyla *Bacteroidetes*, *Chlamydiae*, *Deinococcus-Thermus*, *Nitrospirae*, *Spirochaetes*, *Thermotogae*, *Chlorobi*, *Chloroflexi*, *Cyanobacteria*, *Gemmatimonadetes*, *Lentisphaerae*, *Planctomycetes*, *Verrucomicrobia*, OP10, TM7, WS3, SR1, *Euryarchaeota*, *Korarchaeota*.

Investigation of amplicon libraries in TS. *View on the effect of primer-dependent amplification within TS.* The amplicon libraries L1, L2, L3 were visualized in TS as a “point cloud” where the effect of selective amplification was expressed by different shape and situation of these clouds reflecting particular variants of used universal primer. The effect was numerically estimated by simplest parameters describing relative position and geometry of the amplicon libraries within TS – its central point (a middle position on each coordinate axis) and variance (measure of spread relative to the center). Differences in position of the amplicon libraries within TS were determined from the calculated distances between all pairs of central points. All differences in spatial distribution of the amplicon libraries were found to be statistically significant in Student's t-test and Fisher's exact test.

For the pair L1-L2: $d = 0.09$, $t = 30$, $f = 33$, for L1-L3: $d = 0.10$, $t = 34$, $f = 36$; for L2-L3: $d = 0.07$, $t = 24$, $f = 29$, where d –

distance between central points of the amplicon libraries, t – number of axes with reliable differences between coordinates of the center point (Student's t -test), f – number of axes with reliable differences in dispersion (Fisher's exact test). Therefore, distances between central points of the amplicon libraries suggested similarity of L2 and L3, and their fundamental difference from L1.

2. Maximum displacement of coordinates of the central point along TS axes in pairs of amplicon libraries of 16S-rRNA gene for different bacterial genera

Genus	(L1-L2)	(L1-L3)	(L2-L3)
<i>Eubacterium</i>	0,011	0,019	0,007
<i>Acidimicrobium</i>	0,000	0,016	0,016
<i>Parachlamydia</i>	0,007	0,018	0,011
<i>Simkania</i>	0,006	0,015	0,009
<i>Holophaga</i>	0,039	0,041	0,002
<i>Acidobacteria</i>	0,033	0,004	0,029
<i>Xiphinematobacter</i>	0,014	0,033	0,020
<i>Korarchaeota NA</i>	0,011	0,007	0,018
<i>Thermotoga</i>	0,015	0,009	0,007
<i>Roseomonas</i>	0,006	0,021	0,015
<i>Caulobacter</i>	0,040	0,044	0,004
<i>Sinorhizobium</i>	0,045	0,048	0,003
<i>Azotobacter</i>	0,004	0,019	0,015
<i>Campylobacter</i>	0,008	0,020	0,011

Note. Amplicon libraries L1, L2, L3 were obtained using different universal primers as described in "Technique".

In other words, using the universal primers D1 and BD1 (unlike Eu3) provided generally similar description of structure of the analyzed microbial community. This conclusion is consistent with history of their creation. It is known that primers fD1/rD1 were among the first used in analysis of microbial diversity (12). Their newer analogs (considering significant changes in databases for more than a decade) – primers fBD1/rBD1 (13). Primers Eu3 are widely used in T-RFLP analysis of microbiomes (14), though constructions of these primers are mostly based on sequences of *Proteobacteria* (21). This fact may explain why the analysis of sequences from Eu3 library revealed a significant number of 16S-rRNA genes of rhizobia – typical representatives of this phylum.

3. Distances (d) between central points of amplicon libraries of 16S-rRNA gene in different samples of soil (natural salinization)

Sample	d	Sample	d
T1—T2	0,07	T2-3—T4	0,11
T1—T2-3	0,15	T2-3—T5	0,01
T1—T3	0,08	T2-3—NC	0,09
T1—T4	0,08	T2-3—NZ	0,12
T1—T5	0,07	T3—T4	0,05
T1—NC ^a	0,19	T3—T5	0,09
T1—NZ	0,07	T3—NC	0,13
T2—T2-3	0,10	T3—NZ	0,07
T2—T3	0,05	T4—T5	0,05
T2—T4	0,04	T4—NC	0,15
T2—T5	0,06	T4—NZ	0,02
T2—NC	0,16	T5—NC	0,17
T2—NZ	0,04	T5—NZ	0,03
T2-3—T3	0,10	NC—NZ	0,17

Note. T1-T5 — samples of saline soils, NZ and NC — dark brown soil (NC — virgin land, NZ — fallow); (^a) — maximum value of the character. Description of used universal primers see in "Technique".

Extended estimate of differences in structure of the amplicon libraries was performed using the table showing displacement of coordinates of the central point along the axes (where displacement corresponds to the difference between two coordinates of central points). The largest shift was observed in the case of the library L1, the phylum *Proteobacteria* (coordinates of *Sinorhizobium* and *Caulobacter*), the minimum shift – in the pair L2-L3 (its maximum was on two *Acidobacteria* – *Xiphinematobacter* and *Acidobacteria NA*) (Table 2). All the detected displaced points fairly well agreed with findings of the conventional approach (phylogenetic tree) whose performance also suggested that major differences in structure of the abovementioned amplicon libraries were associated with the phyla of *Proteobacteria* and *Acidobacteria* (Fig. 2, Table 1).

Structure and dynamics of microbial communities of saline soils estimated by integral parameters of TS. Amplicon libraries of microbiomes of saline soils are represented in TS as "point clouds" constructed similarly to the abovementioned L1, L2, L3. In this case, their relative positions in TS (distances between central points) are considered as a measure of similarity or difference in structure of communities. Such distances were computed for all pairs regarding the variant of natural saline soil, and the maximum value was found in the compared pair of most saline sample (T1) and non-saline soil (NC) (Table 3).

Calculated variance of points about the central point and comparison of mean and variance in Student's t -test and Fisher's exact test confirmed reliability of differences observed in TS. Therefore, TS is fairly applicable for studying the dynamics of microbial

communities.

To describe the dynamics of microbiomes in TS, there were introduced two parameters – displacement vector (direction of succession) and the angle between displacement vectors (measure of similarity of succession processes occurring in different microbial communities). The direction of the displacement vector in community corresponds to the difference between coordinates of its start and end points (central points of communities), module of the vector is equal to the distance between central points, and the angle between the vectors is calculated by scalar multiplication of the vectors.

Processes of soil salinization were estimated in TS regarding the three vectors: NC→T1, NZ→T1 (development of natural soil salinization) and NK→NK' (development of induced soil salinization) (Fig. 7).

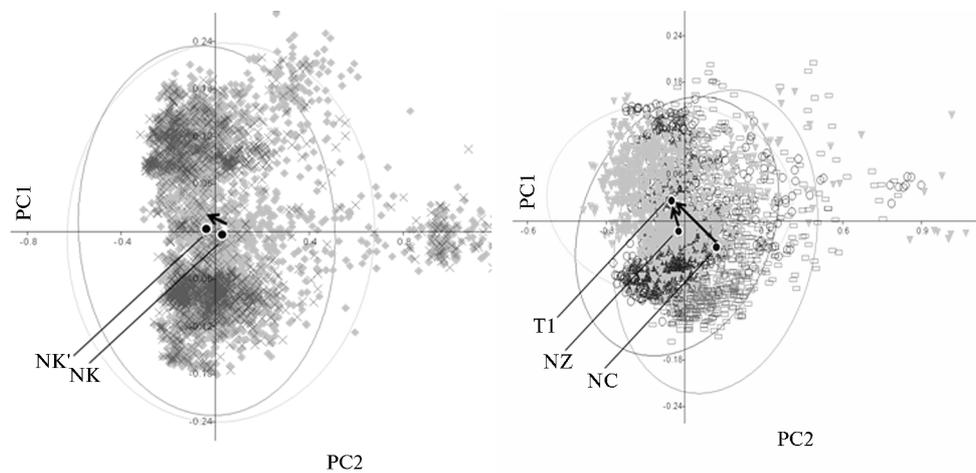


Fig. 7. Integral parameters of microbial communities in taxonomic space (central points and angles between displacement vectors) **constructed using amplicon libraries of 16S-rRNA gene with respect to the conditions of natural (A) and induced (B) soil salinization:** T1, NC, NZ, NK, NK' — respectively, samples of natural saline soil, dark brown soil (NC — virgin land, NZ and NK — fallow), NK' — sample of NK after induced salinization with salt composition similar to the variant T1. PC1, PC2 — coordinates of projection with maximum differences in position of compared aggregates of points.

Analyzing the angles between the vectors in conditions of natural and induced soil salinity, the desired values were found within the sector 90° (42° for pair NK→NK'/NC→T1 and 73° — for NK→NK'/NZ→T1). Assuming the angle in this system as variable from 0° (completely similar development of communities) to 180° (diametrically opposite changes in structure of communities), it can be suggested the presence of similar patterns in development of the studied biomes. This result is in good agreement with the earlier authors' findings (comparative analysis of taxonomic structure of metagenome), where, along with significant differences, samples of saline soil had contained the bacteria from similar orders and families.

This study has revealed an important fact — trends of successions in microbiomes are detectable only in extended analysis of their taxonomic structure while being undetectable in cluster analysis commonly used for such purpose. Using displacement vectors allows several other advantages over conventional methods — simplified procedure of analysis of the dynamics of communities, comparability of communities with different nature, ranking environmental factors by strength of their impact on community.

Certainly, results of the only research don't allow a clear interpretation of angles between the vectors of microbial communities. Their precise estimate can be done upon a series of studies aimed at analysis of factors influencing the microbial community: identification of factors with unidirectional and opposite action, classification of factors by strength of impact on a microbiome, revealing correlations between magnitude of the angle and the value of impact of some particular environmental factor, etc.

Apparently, successful performance of the presented version of TS is associated with enumeration of possible combinations within the pool of coordinate axes. So, high correlation coefficients can only be achieved by using several tens of coordinates, while the previous version of TS (9) operates by only 13 reference points. The proposed alternative model allows revealing the major problems in development of TS — its dimensionality. Finding some compromise between these models is an obvious subject of further research.

Thus, it has been shown applicability of the concept of taxonomic space (TS) for solving the problems related to selective primer-dependent amplification of t16S-rRNA gene and the study of taxonomic structure of microbial communities. The developed new approaches to estimation of biodiversity — mapping biodiversity of 16S-rRNA gene within TS and representation of the microbiome as an overorganism system with inherent integral parameters — may successfully complement conventional methods due to a number of advantages. For example, the concept of TS resolves the currently unclear issue of non-attributed sequences, because in TS each sequence of 16S-rRNA gene is assigned to its own identifier linked to a set of 42 coordinates, which allows its identification in any amplicon library. It ought to be admitted that TS described in this work in strict sense is not a multi-dimensional mathematical space (in particular, axes of TS are interdependent). However, the experience of its practical application has provided extremely high correlation coefficients of pairwise genetic distances between sequences with their geometric analogs, which testifies favorably for applicability of TS. Enumeration of all possible combinations of coordinate axes is the key factor for efficient performance of this method. High correlation coefficients can only be achieved using several tens of coordinates, and, therefore, dimensionality is the major problem in construction of TS. Further development of the concept of TS may give quite important outcomes for investigations of microbial communities and knowing the evolution and history of 16S-rRNA gene. Since this model may describe any variant of its structure (including both evolutionary realized and not yet implemented), it allows discussing the issues related to origin and divergent evolution of prokaryotes, as well as unicellular and multicellular eukaryotes (in TS of 18S-rRNA gene), for example, determining the hypothetical center of origin of taxa considering TS converted into evolutionary space.

REFERENCES

1. Tringe S.G., Hugenholtz P. A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.*, 2008, 5(11): 442-446.
2. Lauber C.L., Hamady M., Knight R., Fierer N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.*, 2009, 15(75): 5111-5120.
3. Lombard N., Prestat E., Elsas J.D.V., Simonet P. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol. Ecol.*, 2011, 78: 31-49.
4. Hunter C.I., Mitchell A., Jones P. McAnulla C., Pesseat S., Scheremetjew M., Hunter S. Metagenomic analysis: the challenge of the data bonanza. *Briefings in Bioinformatics*, 2012, 6(13): 743-746.
5. Acinas S.G., Sarma-Rupavtarm R., Klepac-Ceraj V., Polz M. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries

- constructed from the same sample. *Appl. Environ. Microbiol.*, 2005, 71 (12): 8966-8969.
6. Sipos R., Székely A.J., Palatinszky M., Révész S., Márialigeti K., Nikolausz M. et al. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol. Ecol.*, 2007, 2(60): 341-350.
 7. Bergmann G.T., Bates S.T., Eilers K.G., Lauber C.L., Caporaso J.G., Walters W.A., Knight R., Fierer N. The under-recognized dominance of *Verrucomicrobiin* soil bacterial communities. *Soil Biol. Biochem.*, 2011, 43: 1450-1455.
 8. Sul W.J., Cole J.R., Jesus E.C., Wang Q., Farris R., Fish J.A., Tiedje J.M. Bacterial community comparisons by taxonomy-supervised analysis independent of sequence alignment and clustering. *PNAS USA*, 2011, 108(35): 14637-14642.
 9. Dol'nik A.S., Tamazyan G.S., Pershina E.V., Vyatkina K.V., Porozov Yu.B., Pinaev A.G., Andronov E.E. *Sel'skokhozyaistvennaya Biologiya [Agricultural Biology]*, 2012, 5: 112-120.
 10. Pershina E.V., Tamazyan G.S., Dol'nik A.S., Pinaev A.G., Sergaliev N.Kh., Andronov E.E. *Ekologicheskaya genetika*, 2012, 2: 31-38.
 11. Andronov E.E., Petrova S.N., Chizhevskaya E.P., Korostik E.V., Akhtemova G.A., Pinaev A.G. *Mikrobiologiya*, 2009, 4(78): 525-534.
 12. Weisburg W.G., Barns S.M., Pelletier D.A., Lane D.J. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriology*, 1991, 2(173): 697-703.
 13. Korostik E.V., Pinaev A.G., Andronov E.E. *Ekologicheskaya genetika*, 2006, 4: 32-37.
 14. Singh B., Nazaries L., Munro S., Anderson I., Campbell C. Use of multiplex terminal restriction fragment length polymorphism for rapid and simultaneous analysis of different components of the soil microbial community. *Appl. Environ. Microbiol.*, 2006, 72: 7278-7285.
 15. Maniatis T., Fritch E., Sembruk Dzh. *Metody geneticheskoi inzhenerii. Molekulyarnoe klonirovanie [Methods of Genetic Engineering. Molecular Cloning]*. Moscow, 1984.
 16. Normand P., Orso S., Cournoyer B. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family *Frankiaceae*. *J. Syst. Bacteriol.*, 1996, 46: 1-9.
 17. Wang Q., Garrity G. M., Tiedje J. M., Cole J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 2007, 73(16): 5261-5267.
 18. Bates S.T., Berg-Lyons D., Caporaso J.G., Walters W.A., Knight R., Fierer N. Examining the global distribution of dominant archaeal populations in soil. *ISME J.*, 2010, 5: 908-917.
 19. Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunen ko T., Zaneveld J., Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 2010, 7(5): 335-336.
 20. Lozupone C.A., Knight R. Global patterns in bacterial diversity. *PNAS USA*, 2007, 27(104): 11436-11440.
 21. Marchesi J.R., Sato T., Weightman A.J. Martin T.A., Fry J.C., Hiom S.J., Wade W.G. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl. Environ. Microbiol.*, 1998, 64: 795-799.