Original Article

Microevolution processes are detected in symbiotic microbiomes of Baikal sponges by the methods of fractal theory

Feranchuk S.I.[1,2,*], Potapova U.V.[1], Chernogor L.I.[1], Belkova N.L.[1,3], Belikov S.I.[1]

[1] *Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Ulan-Batorskaya Str., 3, Irkutsk, 664033, Russia*
[2] *Department of Informatics, National Research Technical University, Lermontov Str., 83, Irkutsk, 664074, Russia*
[3] *Scientific Centre for Family Health and Human Reproduction Problems, Timiryaseva Str., 16, Irkutsk, 664033, Russia*

ABSTRACT. In recent years, a large scale ecological crisis has been observed in the Lake Baikal ecosystem. It is clearly shown by several signs, including the mass disease of sponges in the coastal zone. To investigate the causes of the crisis, the composition of symbiotic communities in sponges was investigated in 2015 by sequencing of the 16S rRNA gene in three locations of the lake. The methods of fractal theory were adopted in order to detect a fractal structure in the distribution of the sequencing reads, being considered as fragments of the 16S rRNA gene for individual bacteria within the collected samples. The fractal-like distributions were constructed for the seven most abundant phylotypes, and the observed properties of the distributions reflect microevolution processes within the selected genera and species. The values of the fractal dimension, evaluated for the distributions, are observed to correlate with an anthropogenic load at the place of sample collection, for the *Flavobacterium* and *Synechococcus* genera. The sampling sites were also observed to be associated with the properties of the distributions for chloroplasts of Trebouxiophyceae algae, the endosymbiont of *Lubomirskia baicalensis* sponge.The long-scale time dependency of fractal dimension was also evaluated for the data from temperature detectors in four locations of Lake Baikal. The values of the fractal dimension for fluctuations of temperature are also observed to be associated with an anthropogenic load in the place of measurement. The consistency of both approaches validates the usefulness of fractal-based methods in the interpretation of the experiments designed to study the ecological crisis in Lake Baikal.

Keywords: Lake Baikal, sponges, f fractals, microbiome

1. Introduction

1.1 Motivation

Lake Baikal is a unique example of a large-scale and self-sustaining ecosystem, and the ecological crisis on Lake Baikal observed in recent years needs substantial efforts just to investigate the causes and consequences of the changes in this ecosystem. Endemic sponges in Lake Baikal are clearly vulnerable in this crisis (Bormotov, 2011; Timoshkin et al., 2016; Khanaev et al., 2018). The sponges in Baikal, like all species of the phylum Porifera, always develop in symbiosis with bacterial and micro-algae species (Taylor et al., 2007; Chernogor et al., 2013; Webster and Thomas, 2016). So, the experiments intended to evaluate the composition of microbial communities in sponges from Baikal is one of the most direct ways to investigate processes in the whole lake.

Precedents where disease of sponges or corals were investigated using modern methods of molecular biology are known (Stabili et al., 2012; Prinzón et al.,

2015). And there is a number of studies when a single pathogen which caused the disease was identified (Webster et al., 2002; Luter et al., 2010; Choudhury et al., 2015). But the investigations of microbial communities of diseased sponges in Lake Baikal using similar approaches just demonstrated the complexity of the problem under study (Denikina et al., 2016; Kulakova et al., 2018). So, other wider views of the processes in the Baikal ecosystem and, in particular, in the analysis of microbial communities in sponges, are required.

So-called fractal theory is based on several mathematical notions rooted in geometry and statistics. The approaches of fractal theory can be applied to describe many complex phenomena of different scales, in physics and biology. As counterbalance to this universality, the results obtained by fractal-based methods do not have solid foundations down at microscopic scales. Unlike conventional methods of molecular biology, they can give only hints in the attempts to interpret it in an applied system. And the methods from fractal theory are not very often used in the applied mole-

cular biology and microbiology, partly for the reasons mentioned above.

But in the case of the symbiotic microbiomes of Baikal sponges, even hints would be useful in attempts to explain the causes and consequences of the crisis. So, the presented research has the purpose of applying the methods of fractal theory to the comparative analysis of microbial communities of sponges collected in Lake Baikal in a time of crisis. It was partly based on the methodological study by Feranchuk et al. (2018a) where several features of fractal structure were detected in the rank-abundance distributions of custom microbial communities. Also, the fine-grained temperature measurements in several locations of Lake Baikal were analyzed using methods of fractal theory, and some connections were found with the results obtained from the analysis of microbiomes.

1.2 Fractal theory, advantages and disadvantages

The term 'fractal structure' was introduced in the years 1960-1970 by B. Mandelbrot; in his paper "How Long Is the Coast of Britain" (Mandelbrot, 1967) he applied the concept of non-integer dimension to describe the so-called 'coastline paradox' noticed by L.F. Richardson (1961). The 'coastline paradox' which makes it difficult to measure the length of a coastline in any units of length is demonstrated in Fig. 1. The definitions of the Hausdorff dimension or Minkovski dimension, introduced to mathematics in the early XX-th century, allow the estimation of a non-integer value of dimension for certain specific geometrical objects. And similar methods can be applied to objects arising as natural phenomenons. Citing B. Mandelbrot, "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line".

Objects which can be described as fractals arise from many specific development mechanisms; the precise details of these mechanisms cannot be explained by fractal theory. The features of fractal structure can be detected in any complex real word object if linear dependency is observed in the distribution of the geometrical properties of the object, using logarithmic coordinates, as in the example shown in Fig. 2. In linear coordinates, this dependency would have a form of the so-called "power law" ($y = Ax^D$); the value of D, with negative sign, would be the value of the fractal dimension.

The advantage of the fractal representation of complex geometrical objects is that it provides a way to describe complex structures with a single parameter D, the value of fractal dimension. The exact value of the fractal dimension can be different for the same object, depending on the methods used to calculate this value. But within the frames of any method, the estimates of the value D can be of significance, sufficient to allow this value to be used to compare several objects. There is robust statistical support for the conclusions derived from this comparison (Jelinek and Fernandez, 1998). An alternative interpretation is that the methods of fractal theory provide an efficient way to estimate the number of independent variables in high-dimensional data (Karbauskaite and Dzemuda, 2016).

The features of power-law dependency can be detected in other distributions, not only in geomet-
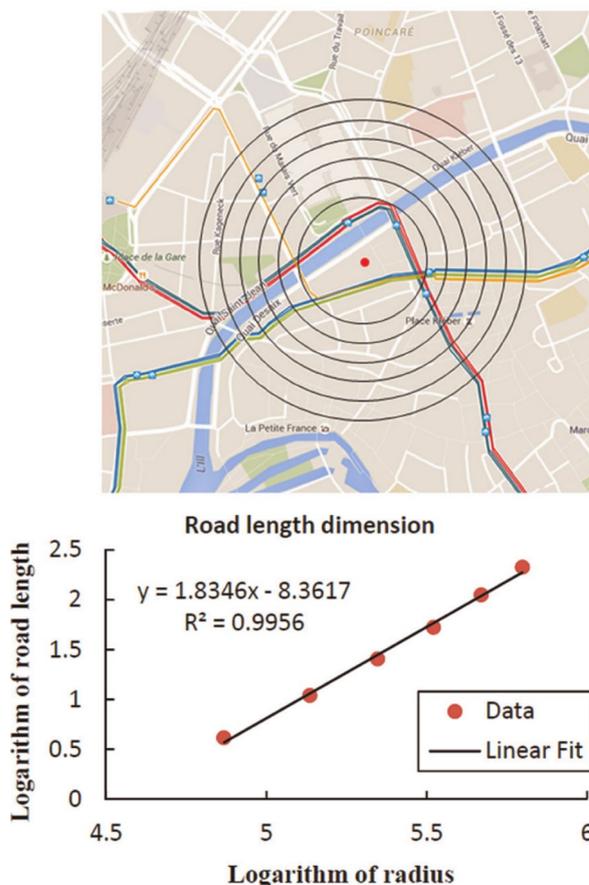
Fig. 2. Detection of fractal structure for the road network in Strasbourg, France. Adopted from (Wang et al., 2017)

Fig. 1. The coast of the Britain measured on several scales
left: Unit = 200 km, length = 2400 km (approx.)
right: Unit = 50 km, length = 3400 km
Reprinted from Wikipedia project page

rical objects. The probability distribution known as Zipf's law, which was observed for the frequencies of words in the texts in natural languages, is also a case of power-law distribution. In Zipf's law, the frequency of a word in some text corpus is inversely proportional to its rank in the frequency table.Another case of power-law distribution, the Pareto distribution used in economics, was derived from an observation by Vilfredo Pareto that 80% of Italy's land was owned by 20% of the population. This observation was then generalized to any kind of property, and in its expanded form, it expresses the power-law dependency in the wealth distribution of society.

The limitations of fractal theory follow from the fact that these methods are almost never supported by a detailed description of the system under study, and any predictions based on these methods are a 'risky business' (Seuront, 2015). In certain cases when the power law is insufficient to explain the distributions obtained by applying fractal techniques to the system under study, a continuous spectrum of exponents can be used; this approach is known as 'multi-fractal analysis' (Harte, 2001). Another more simple expanded form of the power law dependency describing the distributions of income and wealth in the economy, is developed in a series of studies known as econophysics (Yakovenko and Rosser, 2009; Banerjee and Yakovenko, 2010; Xu et al., 2017).

In the discussions proposed by B. Mandelbrot (Mandelbrot, 1960), the cases of power-law distribution were opposed to statistical physics, where the

Boltzmann-Gibbs exponential probability distribution is normally applied. But in econophysics it was demonstrated on several simple models that both distributions can be described within a single framework. And this description can be fitted to the two parts of the distributions which are derived from income statistics in a real-word economy (Fig. 3A).

To explain briefly the models used to derive the unified framework, the Pareto distribution is valid for the higher class of a society where income is proportional to the assets of a household, and the Boltzmann-Gibbs distribution is valid when income is proportional to efforts and time invested in the economic activity. In the first group, but not in the second, the logarithm transformation should be applied to the values of income; this separation can unify the two models for a better description of observed income distributions.

As a result of the fit, two coefficients would consistently describe a power law (Pareto) dependency for the higher class, and an exponential (Boltzmann-Gibbs) dependency for the lower class in the income distributions of a population. Here, an exponent in power law dependency has a meaning similar to fractal dimension, and a factor in exponential dependency has a meaning similar to a temperature parameter in a classical thermodynamics. In this way the income distributions can be explained using only two parameters, but not a continuous spectrum of exponents as in a multifractal approach applied to the same kind of data.

2 Methods

2.1 Processing of 16S rRNA gene sequencing reads of sponge symbionts in Lake Baikal

The results of this study are based on a 16S rRNA microbiome survey of Baikal sponges collected in June 2015 in three locations of the Lake. Several biological replicates of the *L.baicalelenis* sponges, some healthy in appearance, and some diseased, were collected in all three locations; the raw sequencing reads obtained using 454 technology for the V4-V6 region of the 16S rRNA gene are available at NCBI BioProject database under the accession number PRJNA369024. A summary of the archives used in the analysis is shown in Table 1.

All the reads were quality trimmed using the Mothur package (Schloss et al., 2009) and short reads (<200 nt) were filtered. The composition of sponge microbiomes is presented in Fig. 4 as a heatmap chart, for the 25 most abundant phylotypes of bacteria and chloroplasts. The matrix of abundances was calculated using closed-reference OTU picking, implemented in the QIIME1 package (Caporaso et al., 2010), with a database gg_13_7 and Greengenes version of bacterial taxonomy. All environmental sequences were aligned and phylogenetically assigned up to the species level. The taxon identification for phylotype was done according to the degree of similarity at the level of genus (94–97.5%), family (90–94%), order (85–90%), class (80–85%) or phylum (75–80%).
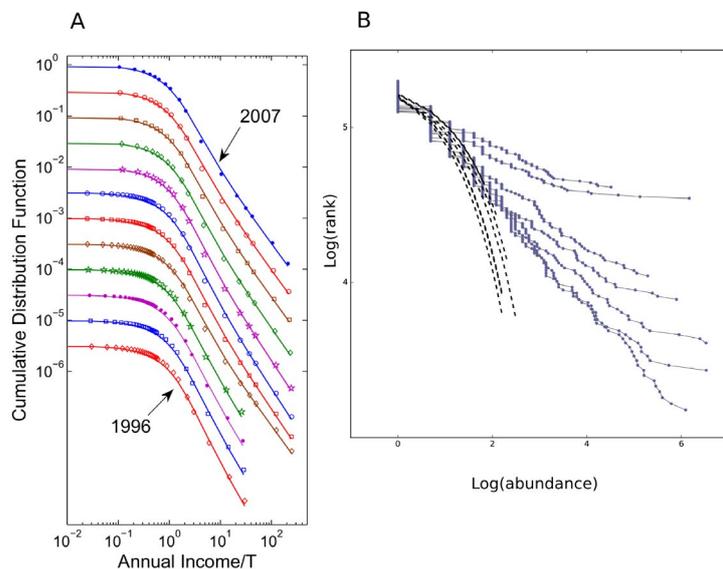
Fig. 3. Cases of dependencies where a description combined from power-law and exponential models could be applied.

A. Cumulative probability distributions of tax returns for USA.

Distributions were constructed from the IRS data (symbols) and their fits with the theoretical distribution, shown in the log–log scale versus the normalized annual income r/T . Plots for different years are shifted vertically for clarity. Reprinted from Banerjee and Yakovenko, 2010.

B. Rank-abundance distributions for sediment microbial communities at a level of genus, shown in the log–log scale. Adopted from Feranchuk et al., 2018a.

Table 1. Integral properties and biodiversity estimates of 23 samples used in the study

Sample	SRA ID	Raw reads	Reads in OTU	OTU Number	Ace	Chao1	Shannon	Simpson	Gini	Singletons	Doubletons
1	2	3	4	5	6	7	8	9	10	11	12
L1	SRR5208569	5253	3833	129	293	248	3.07	0.79	0.93	73	21
L2	SRR5208568	6200	4528	46	82	79	1.47	0.54	0.96	22	6
L3	SRR5208567	7269	4364	107	256	203	2.59	0.72	0.95	61	18
L4	SRR5208566	2762	1454	262	486	477	5.7	0.93	0.73	138	43
L5	SRR5208565	3560	1911	240	359	362	4.86	0.85	0.78	104	43
L6	SRR5208564	3713	1969	235	360	340	4.61	0.85	0.8	103	49
L7	SRR5208563	3888	2145	79	148	135	3.36	0.81	0.89	37	11
L8	SRR5208562	6528	3513	98	208	185	3	0.78	0.93	55	16
OV1	SRR5208561	4445	3747	131	228	202	2.74	0.67	0.93	59	23
OV2	SRR5208560	3338	2783	192	367	345	3.42	0.74	0.89	101	32
OV3	SRR5208559	4787	2019	148	236	231	3.94	0.85	0.86	66	25
OV4	SRR5208558	3484	1830	131	240	229	3.86	0.82	0.87	66	21
OV5	SRR5208557	2970	1277	186	264	276	6.18	0.97	0.66	69	25
OV6	SRR5208556	3068	1757	175	281	285	4.75	0.89	0.81	79	27
OV7	SRR5208555	5062	2687	265	399	359	6.07	0.96	0.76	101	53
OV8	SRR5208554	2584	1003	220	362	375	6.63	0.98	0.62	103	33
T1	SRR5208553	5095	4477	272	582	619	3.34	0.72	0.91	154	33
T2	SRR5208552	5063	4387	282	515	500	3.7	0.75	0.89	142	45
T3	SRR5208551	4912	2011	332	553	509	6.23	0.95	0.73	157	68
T4	SRR5208550	3102	1744	174	350	334	4.7	0.9	0.82	90	24
T5	SRR5208549	5487	2727	252	420	423	5.54	0.95	0.81	116	38
T6	SRR5208548	8843	4822	265	432	423	5.05	0.93	0.86	113	39
T7	SRR5208547	6581	3067	475	764	781	7.31	0.98	0.69	209	70

Collection sites are notated as follows:
L -Listvyanka area, southern Baikal (51.862 N 104.8475 E); OV - Olkhon Gate area, central Baikal (53.0175 N 106.9297 E); T - Turali cape area, northern Baikal (55.2877 N 109.7586 E).
Samples with signs of disease are shown with gray background. Estimates of biodiversity are shown in columns 6-10

2.2 Detection of power-law distributions in microbiology data, for 16S rRNA sequencing of sponge symbionts in Lake Baikal

It has been assumed that the features of fractal structures are present in the patterns of community ecology (Saravia, 2015; Våge and Thingstad, 2015). Cases of power law dependency have been detected in rank-abundance distributions of microbial communities (Feranchuk et al., 2018a), as it is shown in Fig. 3B. The features of fractal structure are expected to be observed in the object under study, whatever the scale. For microbial communities, the fractal structure should be present at all levels of taxonomic hierarchy, even for bacteria within the same species or genus. In the latter case, the properties of the fractal-like patterns should give hints not only about features of microbial community as an ecosystem, but also about micro-evolution processes within the selected species.

The environmental sequences assigned to the same phylotype, as shown on Fig. 4, are close to each other and may even belong to the same species of bacteria, but no-one expects that genomes of individual bacteria in these species be completely identical. Instead, the distributions of proximity between individual bacteria can be investigated using the available sequencing data, attempting to detect a fractal structure in these distributions. So, several microbial genera with different adaptation strategies could be chosen to study in deep the fractal-like properties of symbiotic communities of Baikal sponges.

Seven abundant components of microbiome which were selected for further precise analysis, from the 25 shown on Fig. 4, are specified below. These integral description of these phylotypes are presented in Table 2; the Genbank ID of the 16S rRNA gene sequence selected as a reference for a most abundant OTU within each phylotype, is included in the Table 2.
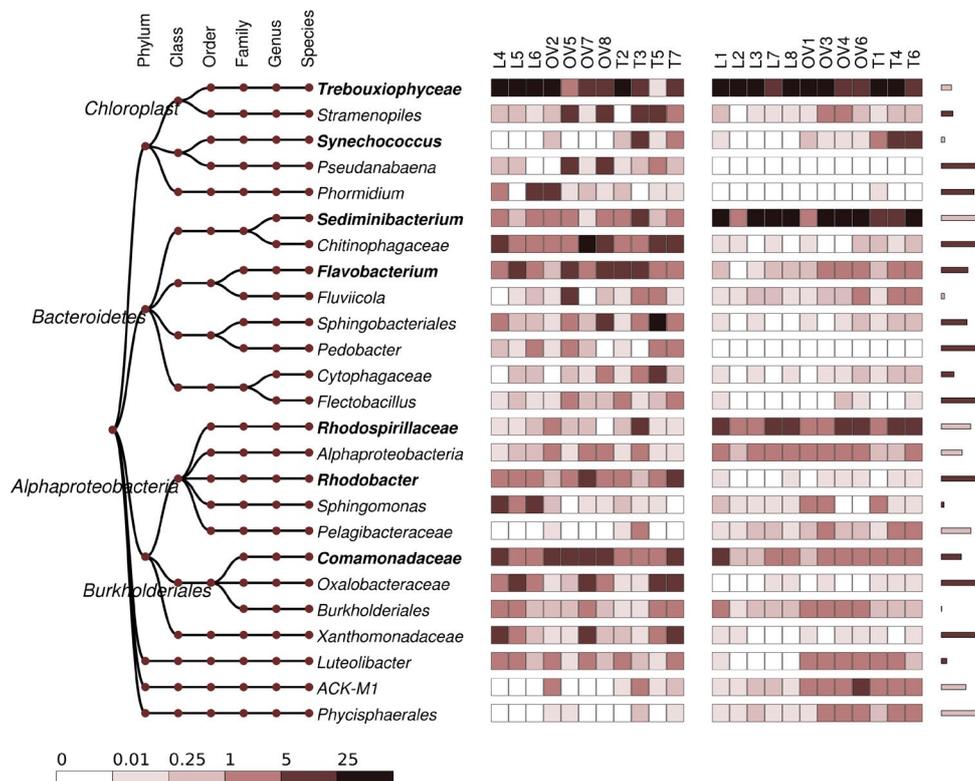
Fig. 4. A composition of microbiomes for healthy and diseased sponges, for the 25 most abundant genera of bacteria and chloroplast. Bars on the right show the significance of difference in abundances between healthy and diseased samples, for each genera. Collection sites are notated as follows: L - Listvyanka, OV - Olkhon Gate, T - Turali cape.

The phylotype annotated to Trebouxiophyceae class represents 16S rRNA gene of chloroplast from unicelluar algae, endosymbiont of *L. baicalensis* sponges (Chernogor et al., 2013). The phylotypes identified at the genus level as *Sediminibacterium* (Chitinophagaceae, Chitinophagales, Chitinophagia, Bacteroidetes) and *Synechococcus* (Synechococcaceae, Synechococcales, Cyanobacteria), as well as the phylotype identified at the level of family and belonged to *Rhodospirillaceae* (Rhodospirillales, Alphaproteobacteria, Proteobacteria) are typical for microbiomes of Baikal sponge. Reference sequences for these representatives were previously described in the sponge samples collected in Lake Baikal (Kaluzhnaya et al., 2011; Gladkikh et al., 2014). Some species of Synechococcus was reported in asso-

ciation with the harmful algal blooms and eutrophication (O'Neil et al., 2012; Berry et al., 2015) and the growth of these representatives of cyanobacteria was detected in several recent measurements in Baikal (Timoshkin et al., 2016). The representatives of family *Comamonadaceae* (Burkholderiales, Betaproteobacteria, Proteobacteria) and F*lavobacterium* (Flavobacteriaceae, Flavobacteriales, Flavobacteriia, Bacteroidetes) are typical for freshwater environments (Bernardet and Bowman, 2006; Willems, 2014) and have been observed in metagenomic surveys on Baikal (Kadnikov et al., 2012; Gladkikh et al., 2014). Some species from both groups were characterized as opportunistic bacteria and pathogens (Horňák and Corno, 2012; Brown et al., 2015; Walsh et al., 2017). The phylotype identi-

Table 2. The specifications of bacterial genera selected for a detailed analysis

Annotation	gg_13_5 ID	Genbank ID	h/d	p-value	% diseased	% healthy	Pareto	Boltzmann (b0)	P/B
1	2	3	4	5	6	7	8	9	10
Trebouxiophyceae	1118847	GU936925.1	h	0.085	31.8	49.4	1.21	10.7	3.1%
Sediminibacterium	840888	FJ800528.1	h	< 0.001	3.2	27.2	1.16	4.9	1.8%
Rhodospirillaceae	848608	FJ800529.1	h	0.001	1.1	5.4	0.86	7.32	3.3%
Synechococcus	550168	GU305759.1	~		0.9	2.0	0.80	8.2	1.8%
Flavobacterium	4324048	JX221823.1	d	0.002	6.8	1.2	0.59	22.0	2.5%
Comamonadaceae	72607	AF289156.1	d	0.008	6.4	2.1	0.92	28.3	2.3%
Rhodobacter	217320	DQ676416.1	d	< 0.001	2.7	< 0.1	1.08	10.7	3.1%

The significant increase of bacteria in healthy (h) or diseased (d) samples is shown on column 4
Columns 8, 9, 10 describe properties of distance-based distributions described below:
column 8 - slope of regression line in power-law part of the distribution;
column 9 - slope of regression line in exponential part of the distribution;
column 10 - threshold of identity which separate power-law part and exponential part in the distribution.

fied as *Rhodobacter* (Rhodobacteraceae, Rhodobacterales, Alphaproteobacteria, Proteobacteria) family was significantly abundant in the diseased samples, in contrast to the other phylotypes from the class Alphaproteobacteria identified as Rhodospirillaceae which was detected preliminary in healthy sponges and described as typical to Baikal. The reference sequence for this genus was obtained from samples in freshwater pond (Briée et al., 2007).

The approach, intended to detect power law distributions for the sequencing reads which are allocated to the same species, is inspired by an algorithm for calculating a Hausdorff-like dimension. One can think about individual sequencing reads as being points in a certain multi-dimensional space. If the dimension of this space was known, it would be possible to construct a system of multi-dimensional spheres around each of the points and to count the number of points within each sphere. The average point count within each sphere could be then transformed to a distribution as a function of sphere radius. Then power law dependency might be observed, and used to evaluate the Hausdorff dimension of this object.

Alternatively, it is possible to calculate a matrix of distances between the points and to count the number of cells in this matrix, where the distances between points are below some threshold. This threshold value is similar in meaning to the sphere radius in the Hausdorff-like approach. So, in the matrix the count of paired distances below the threshold, as a function of threshold value, should provide a distribution where the presence of power law dependency might be expected.

To use the proposed approach in the distributions of individual 16S rRNA genes within each of the selected phylotypes, an Usearch algorithm with 95% identity threshold was applied to screen the sequencing reads in the bulk archives, for each reference. Paired alignments generated in this screening were used to construct the multiple alignments. In these multiple alignments, paired distances between aligned sequencing reads were calculated. These distances might vary from 0 (completely identical) to 5% mutations (the identity threshold in the screening). In the distance matrix, the fraction of paired distances below some threshold was plotted against the value of the threshold; the maximal threshold being 5%, where 100% of distances are included. The obtained distributions are presented in Fig. 5, at both linear and logarithmic scales for threshold of identity.

In the distributions in Fig. 5, the linear dependency is clearly visible in the logarithmic coordinates on both axes, and the presence of linear dependency is strongly confirmed by the statistical tests (p-value $< 10^{-7}$). Also, for several phylotypes, in the right tail of the distribution, a part can be observed which looks like it could be approximated by the Boltzmann-Gibbs model. It can be seen on the right chart of Fig. 5, as a near-linear dependency in appropriate scaling
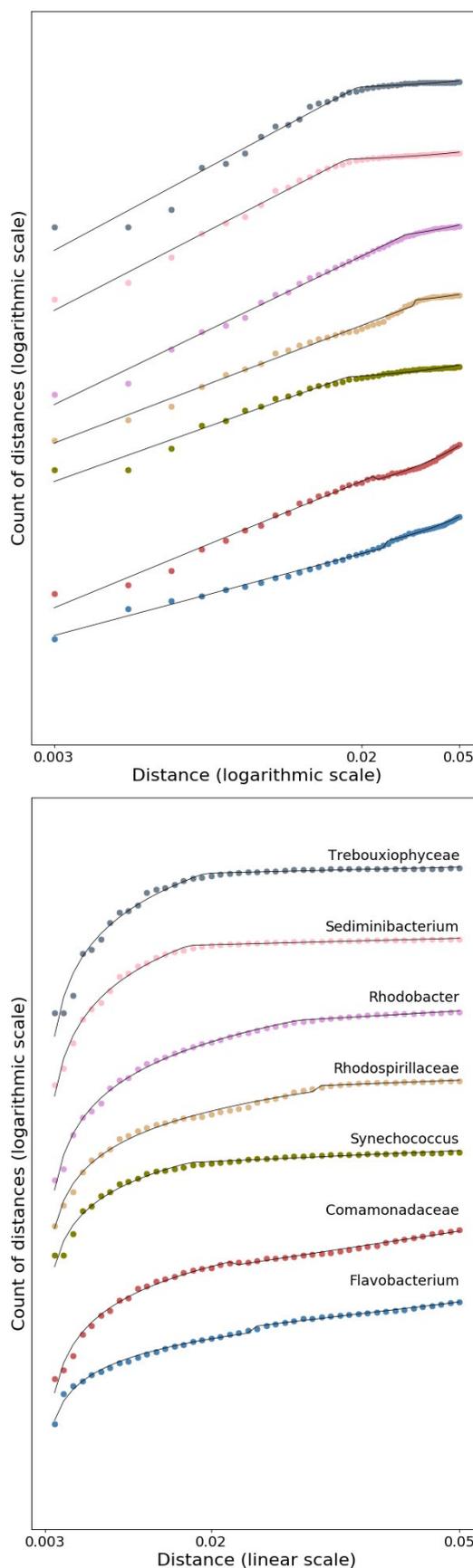
Fig. 5. The distributions of paired distances between fragments obtained by the sequencing of the V4-V6 region of the 16S rRNA genes. The distributions for fragments which have identity >95% to the reference annotated sequences are shown for 7 selected genera. The same distributions are shown in both parts of the chart, at different scales.
X axis - the threshold values of distances, at logarithmic scale (top) and at linear scale (down).
Y axis - the relative count of paired distances below the threshold, at logarithmic scale. The scatter plots are biased along the Y axis to a fixed offset value for each genus.
The thin solid lines show the fit of the combined model used to approximate the distributions.

of the axes. To evaluate the slope of the linear dependency in both parts, the same algorithm was applied to all the distributions. In this algorithm, the point which separates the parts described by the Pareto and by the Boltzmann-Gibbs distributions was selected using as criteria a total mean-square deviation in both parts of the distributions. The pairs of optimal regression lines for each phylotype are also shown in Fig. 5. The slopes of the lines in both parts of the distribution, and in the value of the optimal threshold which separates the two parts, are presented in Table 2.

Applications of fractal theory can provide only some hints about the properties of a complex system. But, as discussed above, the same methodology from fractal theory allow one to compare and interpret the distributions for several objects where a fractal-like structure is observed with sufficient significance. So, the distributions for the seven abundant phylotypes presented in Fig. 5 can be visually separated into three groups. For the Trebouxiophyceae, *Sediminibacterium.* and *Synechococcus* phylotypes, a gradual steady increase of count values is observed in the distributions, with a nearly constant right part for the high identity thresholds. For *Flavobacterium* and *Comamonadacease* phylotypes, an uneven but visible increase can be observed in the right part of the distributions. The phylotypes *Rhodospirillaceae* and *Rhodobacter* can be described as a third group with an intermediate form of the distributions.

These observed "fractal groupings" of the abundant phylotypes can most easily be interpreted in terms of the adaptation strategy of the microorganisms' species and genera. The *Trebouxiophyceae* and *Sediminibacterium* are endemic to Baikal and expected to have a long-term adaptation to their conservative ecological niches. In contrast, the representatives of phylotypes *Flavobacterium* and *Comamonadaceae* have been observed to have highly adaptive survival strategies (Pulkkinen et al., 2009; Jezberová et al., 2017). Cyanobacteria are known to have a stable lineage with consistent survival strategy (Marsac and Houmard, 1993); this hints at the reason behind the observation that the distribution for phylotype *Synechococcus* is close to those of the conservative *Trebouxiophyceae* and *Sediminibacterium*. The *Rhodospirillaceae* and *Rhodobacter* phylotypes belonged to Alphaproteobacteria, their representatives described as phototrophic bacteria, can be expected to have an intermediate rate of adaptation flexibility (Baldani et al., 2014).

These observed "fractal groupings" of the abundant phylotypes can most easily be interpreted in terms of the adaptation strategy of the microorganisms' species and genera. The Trebouxiophyceae and Sediminibacterium are endemic to Baikal and expected to have a long-term adaptation to their conservative ecological niches. In contrast, the representatives of genera *Flavobacterium* and *Comamonadaceae* have been observed to have highly adaptive survival strategies (Pulkkinen et al., 2009; Jezberová et al., 2017). Cyanobacteria are known to have a stable lineage with consistent survival strategy (Marsac and Houmard, 1993); this hints at the reason behind

the observation that the distribution for phylotype Synechococcus is close to those of the conservative Trebouxiophyceae and *Sediminibacterium*. The *Rhodospirillaceae* and *Rhodobacter* phylotypes belonged to Alphaproteobacteria, their representatives described as phototrophic bacteria, can be expected to have an intermediate rate of adaptation flexibility (Baldani et al., 2014).

2.3 Estimates of the fractal dimension for the fluctuations of temperature in Baikal water

Temperature detectors were placed in four locations in Baikal; in Listvyanka, in Baykalsk town (South Baikal), near Uzur village (Olkhon island), and near Bolshye Koty village (direction to the north from Listvyanka). From 2010, the year when the crisis on Baikal was just beginning, detectors have recorded temperatures every two minutes. The collected data from the temperature detectors is certainly valuable in suggesting a clue as to the explanation of the processes in Baikal in time of crisis.

To interpret time series data, methods for time series analysis of various kinds have been developed for many specific applications. Estimates of fractal dimension can also be applied in time series analysis. There are several approaches which can be used to estimate fractal dimensions from time series data (Higuchi, 1988; Peng et al., 1994); the so-called 'Higuchi dimension' is the first and most straightforward of these approaches. It was applied to time series data from the temperature detectors in Baikal.

The method used to estimate fractal dimension proposed by Higuchi (Higuchi, 1988) is shown in Fig. 6, for time series data from Listvyanka. The total length of the lines which connect the data points has been changed when the time interval between the points is doubled. If a time series can be interpreted as a fractal object, the total length of the lines plotted against the interval between the points using logarithmic coordinates, should looks like a linear dependency, as in the classical case of coastlines described in the introduction section.

In the Fig 6, two fragments of temperature measurements are presented as test cases; June 2010 and September 2010. The higher variations of temperature in daytime in September lead to lower variations of difference in the total length of the lines constructed following the Higuchi methodology.

2.4 Specifications and availability of software

The heatmap in Fig. 4 with the results of the closed-reference pipeline was generated using the d3 javascript library (v.3 and v.4) within an interactive framework developed in-house for data visualization; the source scripts of the interactive system are available at https://github.com/sferanchuk/d3b_charts. The Scikit_bio python package (v. 0.4.2) was used to estimate biodiversity values for the samples which are presented in Table 1.
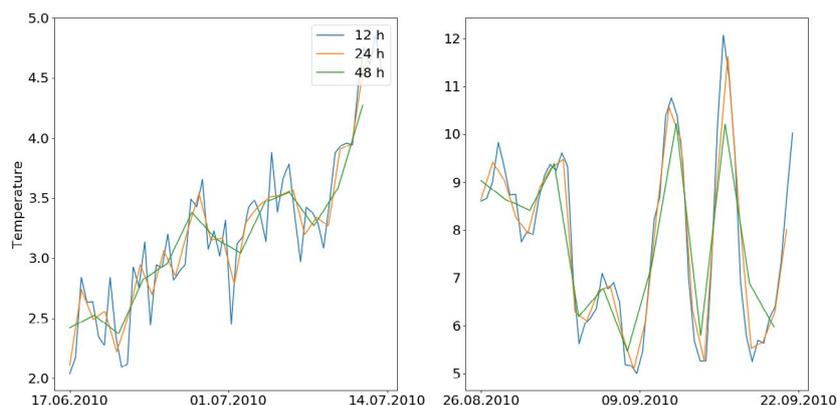
Fig. 6. Two test cases of fractal dimension estimates for fluctuations of temperature in Baikal water.
X axis - time of observations; Y axis - values of temperature, in degrees centigrade. The data from the temperature detector in Listvyanka are used. Three lines on each part of figure present the variations of temperature at different scales.

The libraries of raw sequencing reads identical to the archives available at NCBI (BioProject PRJNA369024) were preprocessed using the Mothur package v. 1.39.5 following conventional setup for filtering unreliable and short oligonucleotides; trim.seq function was applied to raw data files, with parameters ' maxambig = 0, maxhomop = 8, flip = T, bdiffs = 1, pdiffs = 2, qwindowaverage = 35, qwindowsize = 50').

The Usearch alogorithm (Edgar, 2010) implemented in the Uclust software (v. 10.0.240_i86), with the options '-id 0.95 -strand both -maxaccepts 0 -maxrejects 0 -top_hit_only', was used to align the sequencing reads to the database, constructed from the seven 16S rRNA reference sequences listed in Table 2. Paired alignments produced by the Usearch algorithm, in Sam format, were used to generate multiple alignments, by a custom software implemented in c + + programming language. The distance matrices were constructed from multiple alignments; for paired distances, the convention that a string of gaps is counted as a single gap was used, which is the default way to calculate paired distances between 16S rRNA gene fragments accepted in the Mothur package. The distributions of distances were calculated using a custom software in c + +. The source codes for the newly developed software utilities are available at https://github.com/sferanchuk/bsponge fractalmodels.

In the processing of data from temperature detectors, only the portions of the measurements which were continuous for at least 11 days were plotted. Measurements with shorter continuous periods were ignored. Statistical outliers (any value larger or smaller than the average temperature by more than 10SD, within a 16 hour interval) were changed to the value of the average temperature. The time interval used to calculate the Higuchi dimension was between two minutes and 16 hours. The values of the Higuchi dimension were obtained using a Higuchi Fractal Dimension python package; the package was modified to include a check for significance in the estimates of the dimension.

A functionality of the Matplotlib python package (v. 1.5.1) was used to present the results in figures 5, 6, 7 and 8, from the distributions of distances and the data files with records from temperature detectors. In Python, a procedure was developed to separate the two parts of the distance distributions (Fig. 5, 7), based on a linear regression function implemented in the Scipy package (v. 0.17.0) .

3. Results

3.1 The features of fractal dimension for selected microbial phylotypes evaluated separately in each sample

The main applied question, which the present research tries to answer, is to detect and interpret the separation of sponge samples by signs of disease and by geographic location. The separation of samples by disease state can easily be detected by conventional microbiological methods and is observed in the heatmap chart on Fig. 4. The microbiome of diseased sponges is more heterogeneous, and this is supported by the indicators of biodiversity like the Shannon index. But separation by location is difficult to detect, and, if it is detected by advanced statistical methods, it is difficult to interpret the observed separation.

The charts in Fig. 7 present the properties of the distributions, constructed from the alignments of sequencing reads in the same way as described above, but in the latter case the data from each of the samples was processed separately. And, in the same way, two regression lines were used to describe two parts of each of the distributions. As a result, one could use the slope coefficients of two regression lines to obtain two values which describe a sample, for all of the selected microbial phylotypes.

A comparison of the generic slope coefficients for the distributions in Table 2, with the values for separated samples in Fig. 4, suggests that combining several portions of the sequences into one alignment could change the properties of the distributions. The observed distributions are changed in multiple ways, and this problem could be the subject of a further
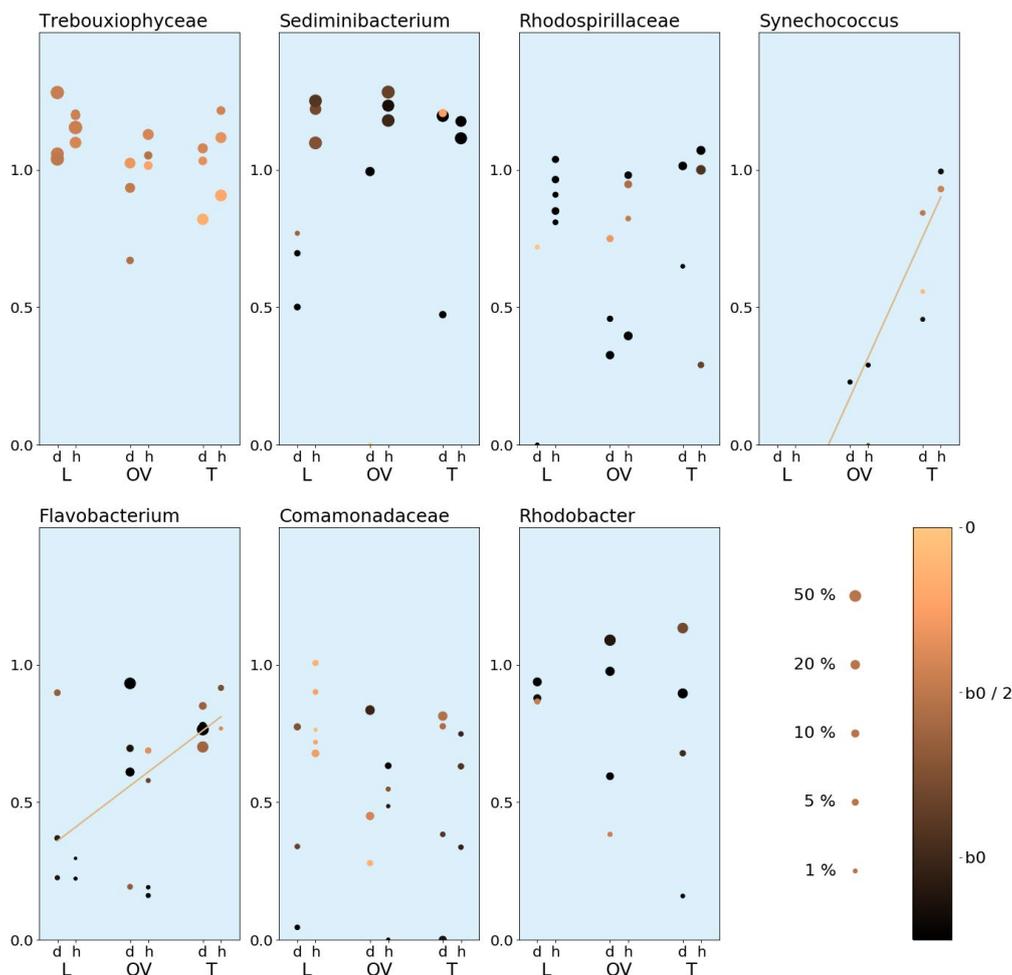
Fig. 7. Values of fractal dimension (Pareto coefficient) and a coefficient in the Boltzmann-Gibbs part of distance-based distributions for selected genera, estimated in separated samples. The separate samples are shown as circles in each part of the chart. The slope of the first part of the distribution (the fractal dimension part), defines the Y-coordinate of each sample. The size of the circle reflects the relative abundance of the selected genus in the microbial community, and color notations are used to show the slope of the regression line in the second part of the distributions (a coefficient for the Boltzmann-Gibbs model).

The X axis is used to separate the places of collection and the disease states of the samples. The scales for sample size and for the Boltzmann-Gibbs coefficient are shown at the bottom right of the chart. The values for the Boltzmann-Gibbs coefficient are scaled separately for each part of the chart, relatively to reference value b0 specific for each genera and listed in Table 2.

study. However, the generic coefficients are comparable by value with the coefficients for the separated samples, and the interpretation of the observed relative variation of coefficients is expected to be meaningful.

The area of Listvyanka at the south of Baikal is the region with high anthropogenic load; in contrast, Turali cape at the north of Baikal is the region without any kind of human activity; the Olkhon Gate area in middle Baikal is an intermediate region both by location and by anthropogenic load. For two phylotypes, *Flavobacterium* and *Synechococcus*, a significant (pv < 0.02) correlation is observed between the place of sample collection and the estimated fractal dimension, as is demonstrated by straight lines in two parts of Fig. 7.

The description of macroscopic systems using the Boltzmann-Gibbs distribution is developed less than a description with fractal dimension; and the slope coefficients for the a second part of the distribution, shown by color notations in Fig. 7, are harder to interpret. But anyway it should be noticed that for phylotypes belonged to chloroplast of class Trebouxiophyceae, a

clear North-South gradient (pv < 0.02) is observed in Fig. 7 for this parameter.

3.2 The fractal dimension for the fluctuations of temperature in Baikal water

The integral representation of temperature measurements in four locations, processed using the Higuchi approach, is shown in Fig. 8. The temperature time series definitely have a fractal structure; the presence of power-law dependency is confirmed with p-value below 10^{-20}. The variations in the estimated fractal dimensions are stressed using a curve constructed as moving average for 50 data points. These variations indicate the uncertainty in the values of the Higuchi dimension. The plots contain gaps and incomplete parts, because, to apply the estimates of fractal dimension in a uniform way, continuous measurements collected every two minutes during at least 11 days were required.

The fractal dimensions shown in Fig. 8 have lowest values in late summer and highest values
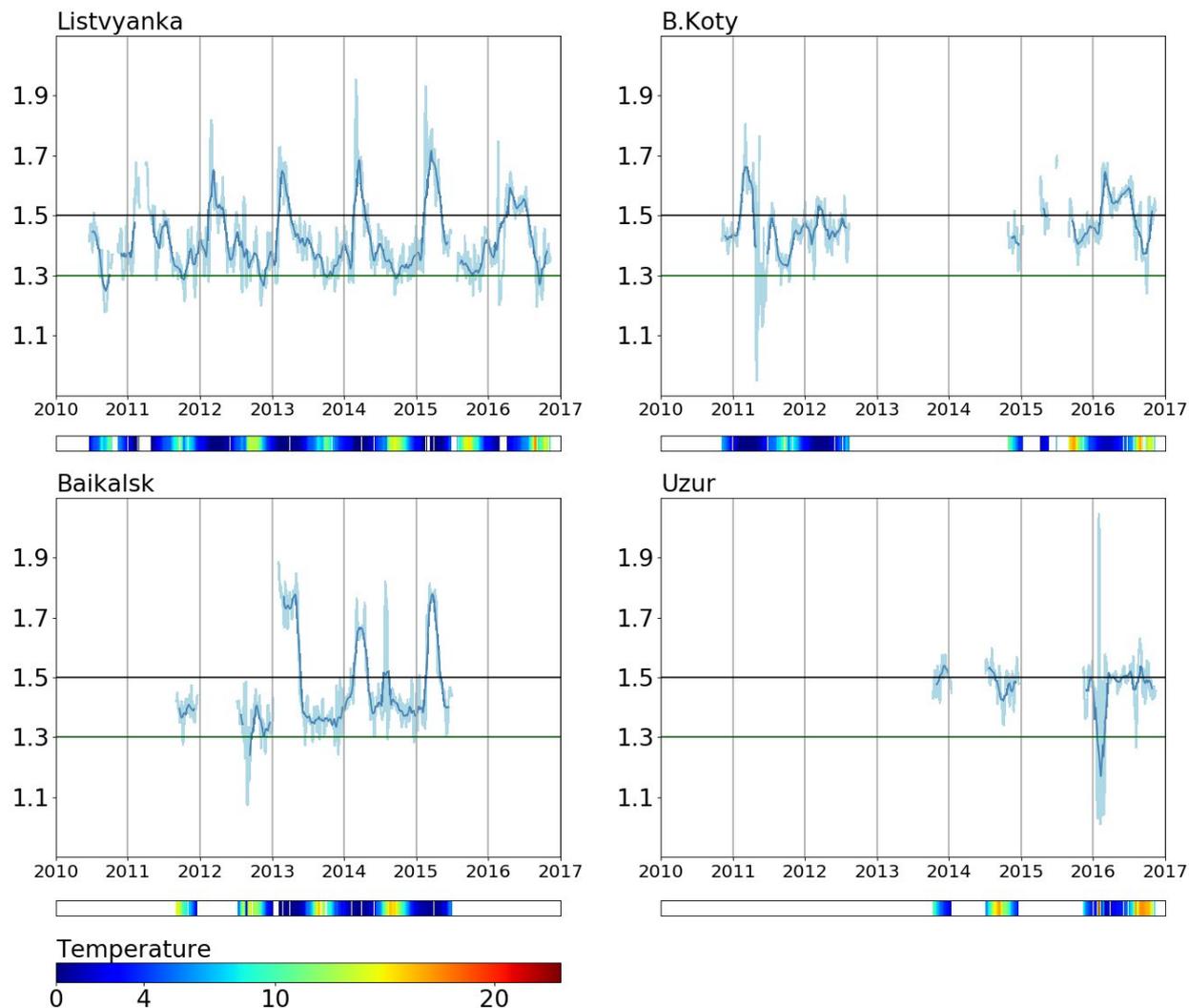
Fig. 8. The fractal dimension calculated for fluctuations of water temperature, in four locations in Lake Baikal.
The fractal dimension (Y axis) was calculated at 16-hour intervals, using the Higuchi approach.
The X axis shows the dates of measurement. The color bars at the bottom part of each subplot directly show the measured values of temperature. A light-blue line directly shows an estimated time dependency of the Higuchi dimension, and a dark blue line shows the same dependency smoothed using moving average with window size equal to 50 data points.

between February and April. Qualitative explanations can be suggested for these features; early spring in Baikal is the time when water is mixing in the near-shore zone beyond the ice, due to an increase in density caused by radiative heating. And, the high variations of temperature in late summer can lead to lower values of the Higuchi dimension, as was discussed above for the test cases in Fig. 6.

But, despite that, in this analysis it must be stressed that, in summer time in Listvyanka the values of the Higuchi dimension are consistently the lowest when compared to other locations and other seasons. And, the late summer in Listvyanka is the season and the place where the anthropogenic load is highest in all Baikal. Associations between the Higuchi dimension and the anthropogenic load can be extended; e.g. Baikalsk town is second after Listvyanka by anthropogenic load, and this can be observed in the distributions of Fig. 8. However, the limited completeness and precision of the measurements and the uncertainty of the estimated values, make these kinds of associations less sure.

4. Discussion

The methods applied to the sequencing data described above, are not intended to reconstruct specific metabolic processes or food chains in a sponge holo-genome. The whole of fractal theory is founded mostly on empirical observations, and therefore, so are the applied results presented in this study, since they are based of fractal theory. As discussed above in comments to the distributions of distances shown in Fig. 5, the differences observed for the selected microbial phylo-types can be interpreted in terms of evolution theory. Anyway, the microevolution processes of bacteria in Baikal sponges are too complex to reconstruct precisely at the molecular level from the available data. So, the presented methods are the most straightforward way which can be used to detect the events of microevolution on Baikal in times of crisis.

It can be concluded from the presented results that microevolution processes do certainly exist in

bacterial species in Baikal sponges. Most of the events in this microevolution remains uncertain, but several facts can be derived with a sufficient confidence. For the representatives of two bacterial phylotypes from the *Flavobacterium* and *Synechococcus*, a connection is detectable between the fractal dimension and the anthropogenic load at the place of collection. And, the variations of temperature at several locations in Baikal also provide a connection between values of the fractal dimension and the anthropogenic load.

In the terms of econophysics, the lower values of the fractal dimension mean that the inequality in the society, measured using the Gini index, would be lower, and the economy would be more robust in this society. Both bacterial phylotypes mentioned above have verified associations with sponge disease and with the contamination of freshwater ecosystems, so it can be interpreted that the development of opportunistic bacteria in the regions with high anthropogenic load is more extensive. Supporting this hypothesis, a slight (pv = 0.11) gradient in the opposite direction can be seen in Fig. 7 for algae chloroplasts, which are expected to suffer in the areas with high anthropogenic load.

The Boltzman-Gibbs part of distributions in econophysics should explain the basic and stable foundations of an economy. In the presented results (Fig. 7), the properties of the Boltzman-Gibbs part in the distributions for the microbial species are mostly difficult to interpret. But anyway, the observed gradient for this slope coefficient for chloroplasts of Trebouxiophyceae algae suggests that the symbiotic algae of sponges are adapting to the terrible disease of their hosts. And the presence of rapid adaptation processes in these algae is also confirmed by the preliminary results published in (Feranchuk et al., 2018b) where the presence of mutations in the algae chloroplast were detected using the sequencing of sponge metagenomes.

In order to put the obtained results into the global and important context of the reasons for the crisis, three considerations are listed below:

- The events of the crisis are manifested in the rapid restructuring of food chains and changes in sponge hologenomes.

- The adaptive changes in bacterial genomes in response to anthropogenic load have been detected for species which are all typical to Baikal, but which have flexible survival strategies.

- Microevolution in a microbial community normally implies interactions between several species, genera or even classes of bacteria, like the exchange of genomic islands or the restructuring of metabolic processes.

Three possible reasons should be considered to explain the beginning of the crisis on Baikal :

- contamination by toxic agents,
- contamination by alien microorganisms,
- events of macroscopic scale.

All three reasons have probably contributed to the crisis, but the above results stress the role of alien microorganisms in the processes as perhaps the most important.

5. Conclusions

The tools from fractal theory provide a level of abstraction which is high. Too high to be certain about any interpretation in molecular terms. This level of abstraction in natural sciences becomes close to that of subjects like history or social sciences. Such reasoning allowed Mandelbrot to entitle the basic conclusions of fractal theory as the "Noah effect" and the "Joseph effect", names borrowed from proverbs in the Bible. At this level of abstraction, one can say that the ecosystem of Baikal is strong and self-sustaining, but nevertheless any system of this kind is fragile and vulnerable. Once the processes of crisis have started, any kind of human intervention is unlikely to prevent the crisis continuing to develop.

But, one may hope that the ecosystem of Baikal, the greatest and most ancient lake on the planet, will remain self-sustaining at a deeper level. The observed signs of the adaptation of algae species to the severe situation of sponge disease give a hint as to how this resurrection of the ecosystem will take place in the near or distant future. Human intervention can be successful only if it is consistent with the processes of recovery which will anyway be continued naturally in the Lake.

And lastly, the methods used in this study are a case when a strategic question, the choice between global reasons for a crisis, can be partly resolved by the use of an experimental tools from molecular biology. But the qualitative analysis provided is obviously not precise enough to provide any suggestions about possible measures to minimize the scale and the damage of the crisis. It can only narrow the areas of interest and suggest some biological objects as a primary targets for a deeper and more extensive study at molecular level.

Acknowledgments

References

Baldani J.I., Videira S.S., dos Santos Teixeira K.R. et al. 2014. The family Rhodospirillaceae. In: Rosenberg E., DeLong E.F., Lory S. et al. (Eds.), The Prokaryotes: Alphaproteobacteria and Betaproteobacteria. Berlin: Springer-Verlag, pp. 533–618. DOI:10.1007/978-3-642-30197-1_300.

Banerjee A., YakovenkoV.M. 2010. Universal patterns of inequality. New Journal of Physics 12. DOI:10.1088/1367-2630/12/7/075032

Bernardet J.F., Bowman J.P. 2006. The genus *Flavobacterium*. In: Dworkin M., Falkow S., Rosenberg E. et al. (Eds.), The Prokaryotes. Volume 7: Proteobacteria, Delta, Epsilon, subclass. New York: Springer, pp. 481–531.

Berry D.L., Goleski J.A., Koch F. et al. 2015. Shifts in cyanobacterial strain dominance during the onset of harmful algal blooms in Florida Bay, USA. Microbial Ecology 70: 361–371. DOI: 10.1007/s00248-014-0564-5.

Bormotov A.E. 2011. What has happened to Baikal sponges? Science First Hand 32: 20–23. (In Russian).

Briée C., Moreira D., López-García P. 2007. Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. Research in Microbiology 158: 213–227. DOI: 10.1016/j.resmic.2006.12.012

Brown B.L., LePrell R.V., Franklin R.B. et al. 2015. Metagenomic analysis of planktonic microbial consortia from a non-tidal urban-impacted segment of James River. Standards in Genomic Sciences 10: 65. DOI: 10.1186/s40793-015-0062-5.

Caporaso J.G., Kuczynski J., Stombaugh J. et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335–336. DOI: 10.1038/nmeth.f.303.

Chernogor L., Denikina N., Kondratov I. et al. 2013. Isolation and identification of the microalgal symbiont from primmorphs of the endemic freshwater sponge *Lubomirskia baicalensis* (Lubomirskiidae, Porifera). European Journal of Phycology 48: 497–508. DOI: 10.1080/09670262.2013.862306.

Choudhury J.D., Pramanik A., Webster N.S. et al. 2015. The pathogen of the great barrier reef sponge *Rhopaloeides odorabile* is a new strain of *Pseudoalteromonas agarivorans* containing abundant and diverse virulence-related genes. Marine Biotechnology 17: 463–478. DOI: 10.1007/s10126-015-9627-y.

Denikina N.N., Dzyuba E.V., Belkova N.L. et al. 2016. The first case of disease of the sponge *Lubomirskia baicalensis*: investigation of its microbiome. Translated in Biology Bulletin 43: 263–270. DOI: 10.1134/S106235901603002X.

Edgar R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461. DOI: 10.1093/bioinformatics/btq461.

Feranchuk S., Belkova N., Potapova U. et al. 2018a. Evaluating the use of diversity indices to distinguish between microbial communities with different traits. Research in Microbiology 169: 254–261. DOI: 10.1016/j.resmic.2018.03.004.

Feranchuk S., Belkova N., Chernogor L. et al. 2018b. The signs of adaptive mutations identified in the chloroplast genome of the algae endosymbiont of Baikal sponge. [version 1; referees: awaiting peer review]. F1000Research 7. DOI:10.12688/f1000research.15841.1

Gladkikh A.S., Kaluyzhnaya O.V., Belykh O.I. et al. 2014. Analysis of bacterial communities of two Lake Baikal endemic sponge species. Mikrobiology 83: 787–797. DOI: 10.1134/S002626171406006X

Harte D. 2001. Multifractals: theory and applications. London: Chapman & Hall.

Higuchi T. 1988. Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena 31: 277–283.

Horňák K., Corno G. 2012. Every coin has a back side: invasion by *Limnohabitans planktonicus* promotes the maintenance of species diversity in bacterial communities. PLoS One 7. DOI: 10.1371/journal.pone.0051576.

Jelinek H.F., Fernandez E. 1998. Neurons and fractals: how reliable and useful are calculations of fractal dimensions? Journal of Neuroscience Methods 81: 9–18.

Jezberová J., Jezbera J., Znachor P. 2017. The *Limnohabitans* genus harbors generalistic and opportunistic subtypes: evidence from spatiotemporal succession in a canyon-shaped reservoir. Applied and Environmental Microbiology 83. DOI:10.1128/AEM.01530-17.

Kadnikov V.V., Mardanov A.V., Beletsky A.V. et al. 2012. Microbial community structure in methane hydrate-bearing sediments of freshwater Lake Baikal. FEMS Microbiology Ecology 79: 348–358. DOI: 10.1111/j.1574-6941.2011.01221.x.

Kaluzhnaya O.V., Itskovich V.B., McCormack G.P. 2011. Phylogenetic diversity of bacteria associated with the endemic freshwater sponge *Lubomirskia baicalensis*. World Journal of Microbiology and Biotechnology 27: 1955–1959. DOI: 10.1007/s11274-011-0654-1.

Karbauskaite R., Dzemuda G. 2016. Fractal-based methods as a technique for estimating the intrinsic dimensionality of high-dimensional data: a survey. Informatica 27: 257–281. DOI: 10.15388/Informatica.2016.84.

Khanaev I.V., Kravtsova L.S., Maikova O.O. et al. 2018. Current state of the sponge fauna (Porifera: Lubomirskiidae) of Lake Baikal: sponge disease and the problem of conservation of diversity. Journal of Great Lakes Research 44: 77–85. DOI:10.1016/j.jglr.2017.10.004.

Kulakova N.V., Sakirko M.V., Adelshin R.V. et al. 2018. Brown rot syndrome and changes in the bacterial community of the Baikal sponge *Lubomirskia baicalensis*. Microbial Ecology 75: 1024–1034. DOI: 10.1007/s00248-017-1097-5.

Luter H.M., Whalan S., Webster N.S. 2010. Exploring the role of microorganisms in the disease-like syndrome affecting the sponge *Ianthella basta*. Applied and Environmental Microbiology 76: 5736–5744. DOI: 10.1128/AEM.00653-10.

Mandelbrot B. 1960. The Pareto-Levy law and the distribution of income. International Economic Review 1: 79–106.

Mandelbrot B. 1967. How long is the coast of Britain? Statistical self-similarity and fractal dimension. Science 156: 636–638.

Manfredi R., Nanetti A., Ferri M. 1999. *Flavobacterium* spp. organisms as opportunistic bacterial pathogens during advanced HIV disease. Journal of Infection 39:146–52.

Marsac N.T., Houmard J. 1993. Adaptation of cyanobacteria to environmental stimuli: new steps towards molecular mechanisms. FEMS Microbiological Reviews 10: 119–189. DOI: 10.1016/0378-1097(93)90506-W.

O'Neil J.M., Davis T.W., Burford M.A., Gobler C.J. 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. Harmful Algae 14: 313–334. DOI:10.1016/j.hal.2011.10.027.

Peng C.K., Buldyrev S.V., Havlin S. et al. 1994. Mosaic organization of DNA nucleotides. Physical Review E 49: 1685–1689.

Prinzón J.H., Kamel B., Burge C.A. et al. 2015. Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. Open Science Journal 2. DOI: 10.1098/rsos.140214.

Pulkkinen K., Suomalainen L.R., Read A.F. et al. 2009. Intensive fish farming and the evolution of pathogen virulence: the case of columnaris disease in Finland. Proceedings of the Royal Society B: Biological Sciences 277: 593-600. DOI: 10.1098/rspb.2009.1659.

Richardson L.F. 1961. The problem of contiguity: an appendix to statistics of deadly quarrels. General systems: Yearbook of the Society for the Advancement of General Systems Theory 61: 139–187.

Saravia L.A. 2015. A new method to analyze species abundances in space using generalized dimensions. Methods in Ecology and Evolution 6: 1298–1310. DOI: 10.1111/2041-210X.12417.

Schloss P.D., Westcott S.L., Ryabin T. et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental

Microbiology 75: 7537–7541. DOI: 10.1128/AEM.01541-09.

Seuront L. 2015. On uses, misuses and potential abuses of fractal analysis in zooplankton behavioral studies: A review, a critique and a few recommendations. Physica A: Statistical Mechanics and its Applications 432: 410–434. DOI: 10.1016/j.physa.2015.03.007.

Stabili L., Cardone F., Alifano P. et al. 2012. Epidemic mortality of the sponge *Ircinia variabilis* (Schmidt, 1862) associated to proliferation of a *Vibrio* bacterium. Microbial Ecology 64: 802–813. DOI: 10.1007/s00248-012-0068-0.

Taylor M.W., Radax R., Steger D. et al. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. Microbiology and Molecular Biology Reviews 71: 295–347.

Timoshkin O.A., Samsonov D.P., Yamamuro M. et al. 2016. Rapid ecological change in the coastal zone of Lake Baikal (East Siberia): is the site of the world's greatest freshwater biodiversity in danger? Journal of Great Lakes Research 42: 487–497. DOI:10.1016/j.jglr.2016.02.011.

Våge S., Thingstad T.F. 2015. Fractal hypothesis of the pelagic microbial ecosystem can simple ecological principles lead to self-similar complexity in the pelagic microbial food web? Frontiers in Microbiology 6. DOI: 10.3389/fmicb.2015.01357.

Walsh K., Haggerty J.M., Doane M.P. 2017. Aura-biomes are present in the water layer above coral reef benthic macro-organisms. PeerJ 5. DOI: 10.7717/peerj.3666.

Wang H., Luo S., Luo T. 2017. Fractal characteristics of urban surface transit and road networks: Case study of Strasbourg, France. Advances in Mechanical Engineering 9: 1–12. DOI: 10.1177/1687814017692289.

Webster N.S., Negri A.P., Webb R.I. et al. 2002. A spongin-boring α-proteobacterium is the etiological agent of disease in the great barrier reef sponge *Rhopaloeides odorabile*. Marine Ecology Progress Series 232: 305–309. DOI: 10.3354/meps232305.

Webster N.S., Thomas T. 2016. The sponge hologenome. MBio 7. DOI: 10.1128/mBio.00135-16.

Willems A. 2014. The family Comamonadaceae. In: Rosenberg E., DeLong E.F., Lory S. et al. (Eds.), The prokaryotes: Alphaproteobacteria and Betaproteobacteria. 4th edition. Berlin: Springer-Verlag, pp. 777–851.

Xu Y., Wang Y., Tao X. et al. 2017. Evidence of Chinese income dynamics and its effects on income scaling law. Physica A: Statistical Mechanics and its Applications 147: 143–152. DOI: 10.1016/j.physa.2017.06.020.

Yakovenko V.M., Rosser J.B. 2009. Statistical mechanics of money, wealth, and income. Reviews of Modern Physics 81. DOI:10.1103/RevModPhys.81.1703