The monitoring of Verkhniy Kaban lake by rbcL gene of freshwater organisms using next-generation sequencing

Anastasia Kharchenko, Anthony Elias Sverdrup, Liudmila L. Frolova

Department of Genetics, Institute of Fundamental Medicine and Biology, Kazan Federal University, 18 Kremlyovskaya St., 420008 Kazan, Republic of Tatarstan, Russia Kazan Federal University, Russian Federation

Abstract

Aim: The lake Verkhniy Kaban, along with the lakes Sredny Kaban and Nizhny Kaban, makes the part of the Kaban lake system. The lakes are located in the center of the large industrial city and experience anthropogenic load. According to the estimates of the Lake Kaban ecologists, they are related to polluted lakes. **Materials and Methods:** The sequencing of rbcL gene fragment sequences of the freshwater lake Verkhniy Kaban hydrobionts during autumn (2016) and summer (2017) sampling periods in FASTQ format is included in the international database on NCBI site with the following unique numbers: SRR7470846, SRR7459785, and SRR7463965. **Results:** The results of the analysis are given, and the water quality evaluation of the Verkhniy Kaban Lake (Kazan, Russia) is presented in the work on the basis of the rbcL gene of freshwater organisms by the method of new generation sequencing. The comparative analysis of metagenomic data shows that the majority of organisms of the Verkhniy Kaban Lake are grouped according to rbcL gene near b-mesosaprobic zone. The Verkhniy Kaban Lake can be characterized as contaminated with the water quality transitional to b-mesosaprobic one. **Conclusions:** The use of modern methods of molecular biology for the purpose of bioindication gives positive results and increases the effectiveness and reliability of water body ecological state evaluation. The obtained results are of great practical interest in the field of water body monitoring, in particular, and the environment as a whole.

Key words: Bioindication, rbcL gene, saprobity, the sequencing of new generation

INTRODUCTION

It is known that biological communities respond to numerous factors of the environment anthropogenic and natural changes. The identification of non-specific reaction of living organisms on the changes of their habitat conditions, for example, pollution or purification is used to assess the ecological state of ecosystems.[1] Bioindication methods have an advantage in comparison with other methods since the presence of indicator species of organisms allows a more accurate assessment of water quality in a reservoir.[2] Indicator species and the communities of organisms are able to live in certain optimal conditions with different contents of oxygen, organic matter, septic decomposition products, and mineralization products in water. There are different contamination zones: Very insignificant (oligosaprobic), moderate (beta-mesosaprobic), dirty (alpha-mesosaprobic), and very dirty (polysaprobic).[3]

Bioindication is based on the observation of the composition and the number of indicator species. Only highly qualified experts can determine an organism by the methods of microscopy with the accuracy to the species according to morphological features. Not every expert can identify organisms of the larval stage uniquely, many species are very small in size, some species have a strong gender dimorphism, or vice versa, there are twin species.[4]

With the development of sequencing methods, it becomes possible to identify a large number of organisms within a species. Thus, for example, DNA barcoding is used to identify

Address for correspondence:
Liudmila L. Frolova, Department of Genetics, Institute of Fundamental Medicine and Biology, Kazan Federal University, 18 Kremlyovskaya St., 420008 Kazan, Republic of Tatarstan, Russia Kazan Federal University, Russian Federation. E-mail: Lucie.Frolova@gmail.com

the organisms within a species.[5] Numerous DNA barcodes are accumulated in the international database of nucleotide sequences of GenBank by the kinds of organisms.[6] DNA barcoding method is based on the sequence of nucleotides of DNA barcode, which is the same among one type of species, for example, for animals, it is the variable fragment of CB1 gene with the length of 600–700 base pairs, and for plants and algae, the fragment of rbcL gene.[7,8] This technology was previously used by us to identify zooplankton organisms by the variable fragment of CO1 gene to assess the ecological state of freshwater reservoirs by bioindication method.[9]

The modern methods of sequencing can be used to assess the ecological state of the environment, including water. This article provides the assessment of Verkhniy Kaban Lake (Kazan, Russia) water quality using the marker gene rbcL of hydrobionts based on new generation sequencing method. The lake Verkhniy Kaban, along with the lakes Sredny Kaban and Nizhniy Kaban, makes the part of the Kaban lake system. The lakes are located in the center of a large industrial city and experience an anthropogenic load. According to the estimates of Lake Kaban ecologists, they refer to polluted lakes.

METHODS

The sampling from the Lake Verkhniy Kaban (Kazan) was conducted in September 2016 and July 2017 according to the standard hydrobiological methods.[10]

The isolation of DNA from the precipitate obtained by the centrifugation of 50 ml of the sample at the rate of 10,000 g for 15 min was carried out using the FAST DNA Kit (MP Biomedicals) according to the manufacturer's protocol. The amplification of the isolated DNA was performed by Phusion High-Fidelity DNA polymerase (Thermo Fisher) with the application of primers [Table 1].

The pair of primers rbcL_AB_FI (forward) and rbcL_AB_RI (reverse) was used to identify Cyanobacteria, Chlorophyta, and Proteobacteria; the pair of primers rbcL_D_FI (forward) and rbcL_D_RI (reverse) was used for the identification of Bacillariophyta, Pyrrophyta, Cryptophyta, and Haptophyta.

The purification of PCR products was carried out with Agencourt AMPure XP beads (Beckman Coulter) followed by a second PCR to index the samples (Nextera XT indices).

The resulting DNA libraries were cut off by Illumina MiSeq device (MiSeq Reagent kit v3). Metagenomic data were included in the International SRA Database on NCBI site.[6]

The resulting metagenomic data were aligned using the BLAST program to establish species diversity and subsequent analysis.

RESULTS AND DISCUSSION

The sequenced sequences of rbcL gene fragment from the freshwater hydrobionts of the lake Verkhniy Kaban during the autumn (2016) and summer (2017) sampling periods in FASTQ format are included in the International SRA Database on NCBI site with the following unique numbers: SRR7470846 (2016), SRR7459785 (2017), and SRR7463965 (2017). In 2016 and 2017, the identification of organisms such as Cyanobacteria, Chlorophyta and Proteobacteria was carried out by new generation sequencing. Furthermore, additional studies were carried out in 2017 to identify the organisms such as Bacillariophyta, Pyrrophyta, Cryptophyta, and Haptophyta. After filtering the quality reads, the trimming of the service sequences, and the removal of chimeric sequences, the resulting nucleotide sequences were aligned with the BLAST program to establish the taxonomic composition.

Based on the sequencing results of Cyanobacteria, Proteobacteria (Bacteria), and Chlorophyta (Viridiplantae) rbcL gene sequences of the Verkhniy Kaban freshwater lake during the autumn (2016) and summer (2017) periods, 63/43 species of bacteria and 89/91 species of viridiplantae were identified. The percentage ratio of taxonomic groups (Kingdom) of the Lake Verkhniy Kaban by species (autumn 2016/summer 2017) makes 41.45%/32.09% bacteria and 58.55%/67.91% viridiplantae [Figure 1].

The percentage ratio of the taxonomic groups (Kingdom) of the Lake Verkhniy Kaban makes the following by the reads (autumn 2016/summer 2017): 84.09%/82.72% bacteria and 15.91%/17.28% viridiplantae [Figure 2].

According to Figures 1 and 2, it is seen that Viridiplantae are prevalent as compared to bacteria, but bacteria dominate by reads significantly.

Table 1: Primers for PCR of rbcL gene of phytoplankton	
Primers	**Sequences**
rbcL_AB_FI (forward)	5'-tcgtcggcagcgtcagatgtgtataagagacagtcigciaaraactayggtcg-3'
rbcL_AB_RI (reverse)	5'- gtctcgtgggctcggagatgtgtataagagacagggcatrtgccaiacrtgrat-3'
rbcL_D_FI (forward)	5'-tcgtcggcagcgtcagatgtgtataagagacaggatgatgaraayattaactc-3'
rbcL_D_RI (reverse)	5'-gtctcgtgggctcggagatgtgtataagagacagatttgdccacagtgdatacca-3'

According to the variety of the Verkhniy Kaban Lake species, they identified 34/33 species of Cyanobacteria, 29/9 species of Proteobacteria, and 89/91 species of Chlorophyta in 2016/2017. During the autumn (2016) and the summer (2017) period, the percentage of the Lake Verkhniy Kaban species makes 22.37%/24.81% of Cyanobacteria species, 19.08%/6.77% of Proteobacteria species, and 58.55%/68.42% of Chlorophyta species [Figure 3].

By the number of organisms, they identified the following in 2016/2017: 5214/4201 of Cyanobacteria reads, 119/45 of the Proteobacteria reads, and 1009/887 of the Chlorophyta reads. The percentage of reads in rbcL gene of the freshwater lake of the Verkhniy Kaban during the autumn (2016) and summer (2017) period makes 82.21%/81.84% of cyanobacteria reads, 1.8%/0.88% of Proteobacteria reads, and 15.91%/17.28% of Chlorophyta reads [Figures 4 and 5].

According to Figures 3 and 4, it can be seen that Chlorophyta is almost 2.5 times larger in species diversity than cyanobacteria and 3–10 times larger than Proteobacteria, but Cynobacteria significantly dominates by reads. Its number is 4 times larger as compared with Chlorophyta and 8 times larger as compared with Proteobacteria. As is known, most

of the indicator species of cyanobacteria characterize water bodies as o- and b-mesosaprobic, proteobacteria - p-saprobic, and chlorophyta - b- and a-mesosaprobic.

Based on the results of additional sequencing of the rbcL gene fragment series of Bacillariophyta, rbcL Cryptophyta and rbcL Haptophyta, the freshwater lake Verkhniy Kaban was identified by the following species/reads during the summer period of 2017: 66/4400 of Bacillariophyta, 9/1029 of Cryptophyta, and 3/3573 of Haptophyta.

The percentage ratio of the taxonomic groups of the Verkhniy Kaban Lake by species and reads (summer 2017) makes the following: 11.54%/11.43% of Cryptophyta, 3.85%/39.69% of Haptophyta, and 84.62%/48.88% of Bacillariophyta [Figure 5]. As can be seen from Figure 5, the Bacillariophyta organisms significantly dominate over other microscopic algae by the diversity of species and the number of reads. In general, the indicator organisms of Bacillariophyta characterize the reservoir as b- and o-saprobic.

The results of series sequencing of rbcL gene fragment of Bacillariophyta, rbcL Cryptophyta, and rbcL Haptophyta of the freshwater lake Verkhniy Kaban for the diversity
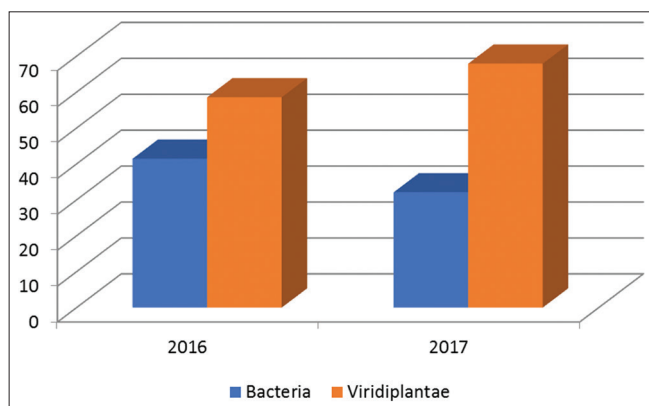

Figure 1: The percentage of taxon (Kingdom) by species diversity of Verkhniy Kaban Lake. (Autumn 2016; Summer 2017)
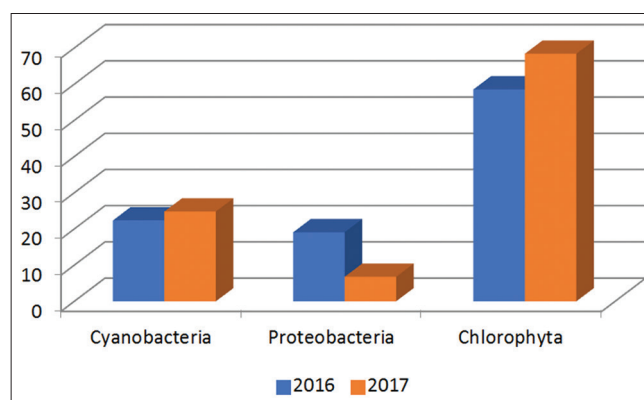

Figure 3: The percentage of species diversity of rbcL hydrobionts of Verkhniy Kaban Lake. (Autumn 2016; Summer 2017)
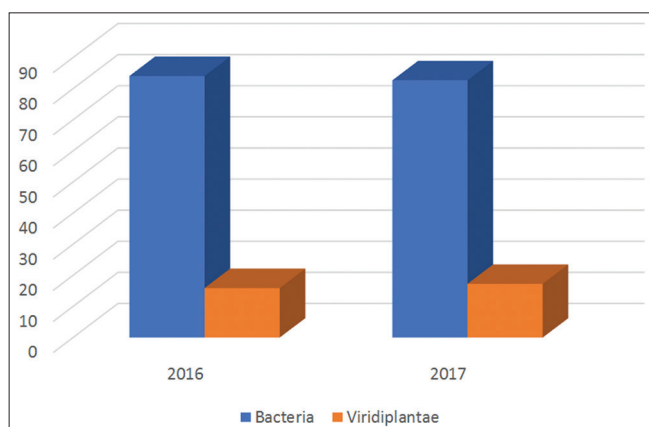

Figure 2: The percentage of taxon (Kingdom) by reads of Verkhniy Kaban Lake. (Autumn 2016; Summer 2017)
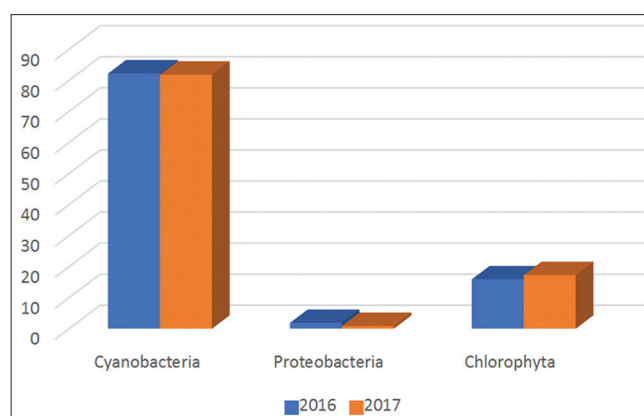

Figure 4: The percentage of rbcL hydrobionts reads of Verkhniy Kaban Lake. (Autumn 2016; Summer 2017)

of species and the abundance of organisms, as well as in percentage during the summer period (2017), are shown in Table 2 and Figure 6.
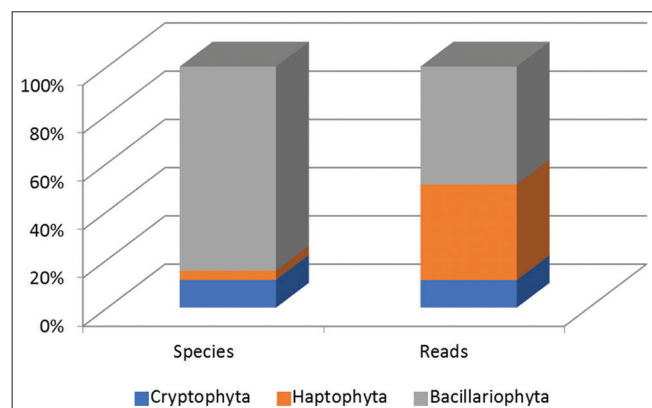

Figure 5: The percentage of taxon diversity and reads of hydrobionts of Verkhniy Kaban lake (Summer 2017)

It can be seen from Figure 6 that Bacillariophyceae (b-saprobity) and Coscinodiscophyceae (a-b-saprobity [Bacillariophyta]) dominate by species diversity, and Prymnesiales (Haptophyta) also dominates by reads.

From the analysis of the metagenomic data by the fragment of cyanobacteria rbcL gene, rbcL Chlorophyta, and rbcL Proteobacteria of the Verkhniy Kaban lake (autumn 2016), 13 species were identified as the indicator ones from the list of indicator organisms by Sladechek[11]:

1.Two types of cyanobacteria:
- *Microcystis aeruginosa* - b-mesosaprobity;
- *Oscillatoria tenuis* - a-mesosaprobity;

2.11 species of chlorophyta:
- *Actinastrum hantzschii, Micractinium pusillum, Phacotus lenticularis, Botryococcus braunii, Eudorina elegans, Gloeotila contorta,* and *Pediastrum*
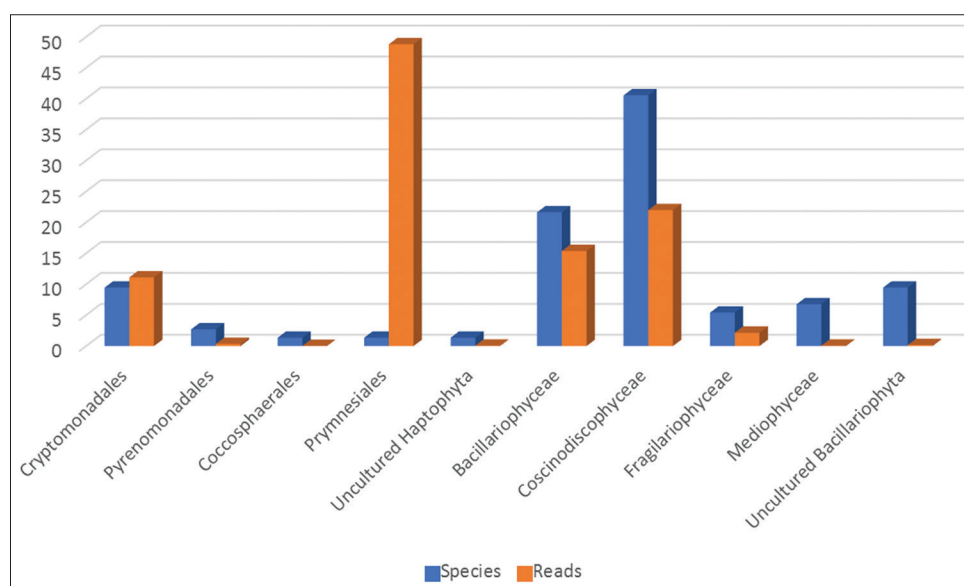

Figure 6: The percentage of species diversity and reads of hydrobionts of Verkhniy Kaban Lake (Summer 2017)

Phylum	Order	Species quantity	Reads quantity	Species%	Reads %
Table 2: The species diversity and quantity of freshwater organisms (the quantity of reads), also they are in the percentage of Verkhniy Kaban Lake (Summer 2017)					
Cryptophyta	Cryptomonadales	7	999	9.46	11.10
	Pyrenomonadales	2	30	2.70	0.33
Haptophyta	Coccosphaerales	1	1	1.35	0.01
	Prymnesiales	1	4393	1.35	48.80
	Uncultured	1	6	1.35	0.07
Bacillariophyta	Bacillariophyceae	16	1383	21.62	15.36
	Coscinodiscophyceae	30	1978	40.54	21.97
	Fragilariophyceae	4	193	5.41	2.14
	Mediophyceae	5	5	6.76	0.06
	Uncultured	7	14	9.46	0.16

duplex - b-mesosaprobity;
- *Ankistrodesmus falcatus* - b-a-mesosaprobity;
- *Chlamydomonas reinhardtii* and *Chlorella vulgaris* - a-mesosaprobity;
- *Chlorella pyrenoidosa* - p-saprobity.

Ten species were identified as the indicative ones from the list of indicator organisms by Sladechek according to the analysis of the metagenomic data by Cyanobacteria rbcL gene and rbcL Chlorophyta of the Lake Verkhniy Kaban (summer 2017)[11]:

3. One species of Cyanobacteria:
- *Anabaena flos-aquae* - b-mesosaprobity;

4. Nine species of Chlorophyta:
- *P. lenticularis*, *P. duplex*, *E. elegans*, and *Scenedesmus bijugus* - b- mesosaprobity;
- *Microthamnion kuetzingianum* and *A. falcatus* - b-a- mesosaprobity;
- *C. reinhardtii* and *C. vulgaris* - a-mesosaprobity;
- *C. pyrenoidosa* - p-aprobity.

The percentage of the species by rbcL gene of Cyanobacteria and rbcL of Chlorophyta according to saprobity (autumn 2016, summer 2017) is shown in Figure 7. Figure 7 shows a significant predominance of the species percentage by rbcL gene of cyanobacteria according to b-mesosaprobity both in 2016 and 2017.

The percentage of the number of reads by rbcL gene of Cyanobacteria and rbcL of Chlorophyta according to saprobity (autumn 2016, summer 2017) is shown in Figure 8.

Figure 8 shows a significant predominance of read percentage by rbcL gene of Cyanobacteria and rbcL gene of Chlorophyta according to a-mesosaprobity in 2016 and b-mesosaprobity in 2017.
Six species were identified as the indicative from the list of indicator organisms by Sladechek according to the analysis of additional metagenomic data on rbcL gene of bacillariophyta and rbcL gene of cryptophyta of the Lake Verkhniy Kaban (summer 2017)[11]:

5. Three species of Cryptophyta
- *Cryptomonas curvata* - b-mesosaprobity;
- *Cryptomonas marssonii* - o-b-mesosaprobity;
- *Cryptomonas ovata* - a-mesosaprobity.

6. Three species of Bacillariophyta:
- *Nitzschia fonticola* - o-b-mesosaprobity;
- *Cyclotella meneghiniana* - a-b-mesosaprobity;
- *Nitzschia palea* - b-mesosaprobity.

The percentage of the species by rbcL gene of Bacillariophyta and rbcL Cryptophyta according to saprobity (summer 2017) is shown in Figure 9. Figure 9 shows a significant predominance of the percentage ratio of hydrobiont species according to b- and o-b-mesosaprobity.
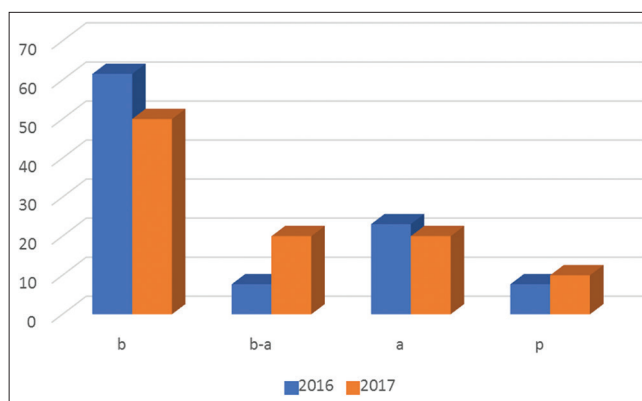

Figure 7: The percentage of hydrobionts saprobity by species of Verkhniy Kaban Lake. (Autumn 2016; Summer 2017)
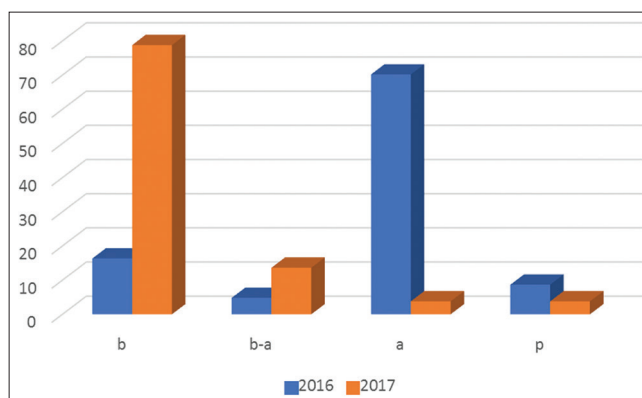

Figure 8: The percentage of hydrobionts saprobity by reads of Verkhniy Kaban Lake. (Autumn 2016; Summer 2017)
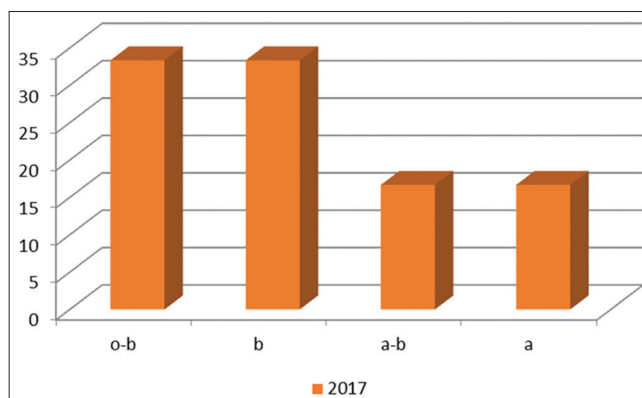

Figure 9: The percentage of hydrobiont saprobity by species of Verkhniy Kaban lake. (Summer 2017)

The percentage of reads by rbcL gene of Bacillariophyta and rbcL gene of Cryptophyta according to saprobity (summer 2017) is shown in Figure 10. Figure 10 shows the predominance of the percentage ratio concerning the number of hydrobiont reads according to b-mesosaprobity in 2017.

The obtained results for rbcL gene of Bacillariophyta, rbcL gene of Pyrrophyta, rbcL gene of Cryptophyta, and rbcL gene of Haptophyta supplement the results for the rbcL gene of Cyanobacteria, rbcL gene of Chlorophyta, and rbcL gene

of Proteobacteria, and the indicator species of which indicate predominantly the b-mesosaprobity of the lake Verkhniy Kaban.

The comparative analysis of metagenomic data for the rbcL hydrobiont gene in 2016 and 2017 shows that most of the indicator organisms of the Lake Verkhniy Kaban are grouped around the b-mesosaprobity. Figure 11 shows the Venn diagram, which demonstrates the number of indicator species of the Verkhniy Kaban Lake in 2016 and 2017.

Based on the results of the study, 13 indicator species were identified in 2016 and 10 indicator species in 2017, of which 7 species are identical for 2 years of research: *A. falcatus* - b-a-mesosaprobiont; *C. reinhardtii* and *C. vulgaris* - a-mesosaprobionts; *C. pyrenoidosa* - p-polysaprobiont; and *E. elegans*, *P. duplex*, and *P. lenticularis* - b-mesosaprobionts. Six indicator organisms were identified only in 2016: *A. hantzschii*, *B. braunii*, *G. contorta*, *M. pusillum*, *M. aeruginosa*, and *O. tenuis* - b-mesosaprobionts. Three indicator organisms were identified only in 2017: *A. flos-aquae* and *S. bijugus* - b-mesosaprobionts and *M. kuetzingianum* - b-a-mesosaprobiont.

In 2016, the diversity of Viridiplantae species is comparable to the diversity of *Bacteria* species, and in 2017, it exceeds it twice [Figure 12], whereas it is 4 times higher than Viridiplantae by the number of Bacteria. This is due to the intensive flowering of *Cyanobacteria* in the conditions of abnormal heat during this period.

CONCLUSIONS

Based on the results of the research, using the modern methods of a new generation sequencing, molecular and bioinformational analysis of water quality, the Verkhniy Kaban lake is transitional to b-mesosaprobic and can be characterized as contaminated.

SUMMARY

New generation sequencing methods can be successfully used to assess the ecological state of the environment, including water one. The use of modern molecular genetic methods improves the quality of the environment diagnostics, the accuracy, and the reliability of environmental monitoring results. The assessment of the environmental state of the environment and the environmental measures is carried out by the replacement of subjective visual studies for each organism with the results of objective instrumental research. Besides, an environmental assessment of the environment can be performed by less qualified experts. The use of modern methods of molecular biology for the purpose of bioindication gives positive results and increases the effectiveness of water body ecological state evaluation. The obtained results are of great practical interest in the field of water body monitoring, in particular, and the environment as a whole.
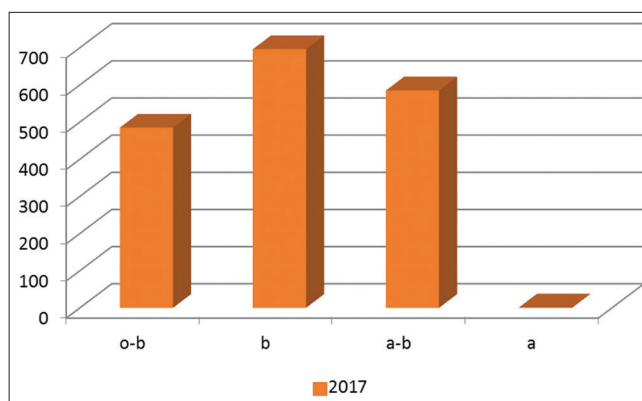

Figure 10: The percentage of hydrobiont saprobity by reads of Verkhniy Kaban Lake. (Summer 2017)
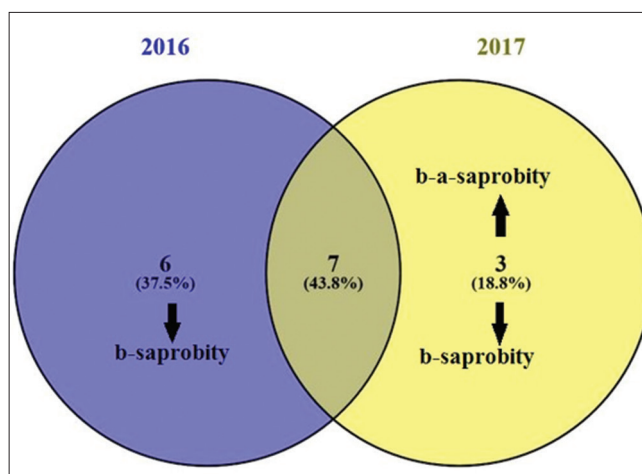

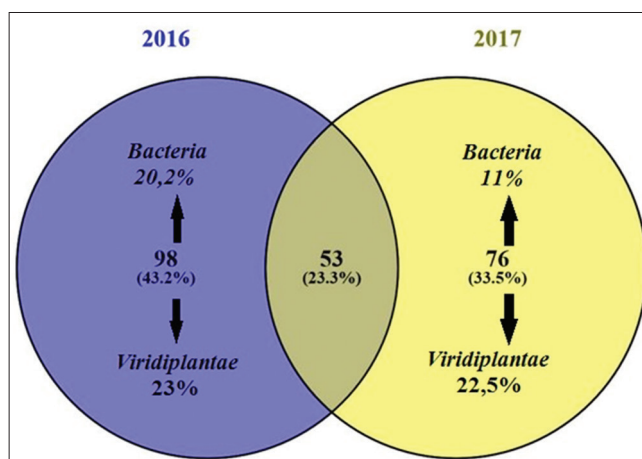**Figure 11:** Diagram of Vienna for indicator organisms of Verkhniy Kaban Lake. (Autumn 2016, Summer 2017)

Figure 12: Diagram of Vienna for species diversity of Verkhniy Kaban Lake. (Autumn 2016, Summer 2017)

ACKNOWLEDGMENTS

REFERENCES

1. Kurilenko VV, Zaitseva OV, Novikova EA, Osmolovskaya NG, Ufimtseva MD. Fundamentals of Ecology, Bioindication and Biotesting of Aquatic Ecosystems. Saint-Petersburg: SPBGU; 2003. p. 448.
2. Miseiko GN, Bezmaternyh DM, Tushkova GI. Biological Analysis of Fresh Water Quality. Barnaul: AltFU; 2001. p. 201.
3. Bick H. A review of central European methods for the biological estimation of water pollution levels. Bull World Health Organ 1963;29:401-3.
4. Gomez A. Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoans taxon. Proc B Royal Soc 2007;274;199-207.
5. Hebert P, Ratnasingham S, deWaard JR. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proc Roy Soc Lond B 2003;270:96-9.
6. National Center for Biotechnology Information – http://www.ncbi.nlm.nih.gov
7. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrate. Mol Mar Biol Biotechnol 1994;3:294-9.
8. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci U S A 2009;106:12794-7.
9. Frolova LL, Husainov AM. Identification of indicator species of zooplankton organisms by COI gene fragment for estimation of ecological state of a water body. Int J Pharm Technol 2016;8:24477-86.
10. Buillon VV. Methodological recommendations on the collection and the processing of materials during hydrobiological studies at freshwater reservoirs. Phytoplankton and its Products. Lake: State Scientific-Research Institute of Lake Fishery; 1981. p. 32.
11. Sladechek V. System of water quality from the biological point of view. Arch Hydrobiol Ergeb Limnol 1973;11:179-91.