ARTICLE

# A NEW APPROACH FOR THE DETERMINATION OF SPECIES SAPROBITY FOR WATER QUALITY MONITORING BASED ON THE MOLECULAR PHYLOGENY

**Anthony E. Sverdrup*, Ludmila L. Frolova**

*Department of Genetics, Kazan Federal University, 18 Kremlevskaya Str., Kazan,, RUSSIA*

ISSN 0976-3104

## ABSTRACT

*The well-known method of bioindication is widely used for water quality monitoring. The number of indicator species is low in a compare with total number of species. For example, the indicator list of freshwater Bacillariophyceae made by V. Sladechek includes only 192 species of organisms from estimated 100000 species. The same situation happens with other freshwater organisms, which have not the status of indicator until nowadays. For solving this problem, we suggest a new approach for determination of indicator species for water quality monitoring based on the molecular phylogeny. Our choice is the rbcL gene and product of rbcL gene – protein of Bacillariophyceae, which is used as marker gene for plants. Phylogenetic analysis includes 66 sequences of rbcL gene and rbcL protein of Bacillariophyceae bioindicators accordingly from GenBank and GenPept Sequences Databases on NCBI website. As results, two phylogenetic trees on rbcL gene and two trees on rbcL protein of Bacillariophyceae were constructed using Neighbor-Joining (NJ) and Maximum Parsimony (MP) methods with bootstrap. The comparative analysis of phylogenetic trees shows stable clustering of the indicator species of different genera with the same or close saprobity with higher bootstrap by the rbcL gene than by the rbcL protein. Obtained results allow us to find new bioindicators faster than by using the traditional technology, using less time and resources. Thus, we conclude that our new technology can be used for the water quality monitoring and research results have the fundamental and practical value.*

**KEY WORDS**
*saprobity, determination, water quality, DNA/protein rbcL, Bacillariophyceae, bioindicators*

## INTRODUCTION

Water quality is a complex subject, because water is a complex medium intrinsically tied to the ecology of the Earth and it shows large geographical variation. The most common standards used to assess water quality relate to health of ecosystems, safety of human exposure, and drinking water quality. Water quality refers to the chemical, physical, and biological characteristics of water. Biological monitoring of water quality is based on bioindicators. A bioindicator is any species whose function, population, or status reveal the qualitative status of the environment. We used plankton, because it is geographically independent, it is convenient for detection and it quickly reacts to pollutants of any type. For example, members of plankton such as micro-algae.

We work with phytoplankton samples of diatoms or *Bacillariophyceae*, which are a major group of micro-algae, and the most common types of phytoplankton. Diatoms with other water inhabitants are used as saprobity indicator species in the monitoring of the water quality. Traditional technology for bioindicator identification include sample collection, identification of species by morphological characteristics using the microscope and construction of the bioindicators list. There are only 192 known indicator species of diatoms from list of Sladecek (1973) [1]. That is few as compared with more than 200 genera with an estimated 100000 species. The same situation is happening with other organisms, which have not the status of indicators until nowadays. This is explained by the long-time used in analysis and high cost of the experimental work with living organisms. To reduce costs and increase productivity we suggest the alternative approach of bioindicators determination, building on the existing database. It includes the sequences collection from free international database and analysis of molecular phylogeny [2]. Our choice is the gene and product of rbcL *Bacillariophyceae* which is used as marker gene for plants [3, 4]. We used the rbcL gene of fresh water plants for the monitoring of Kaban Lakes [5, 6]. Our aim is to development of more efficient methods for bioindication based on modern methods of bioinformatics and molecular genetics for water quality monitoring.

Received: 18 Aug 2019
Accepted: 12 Sept 2019
Published: 23 Sept 2019

## MATERIALS AND METHODS

Phylogeny analysis includes sequences of rbcL gene and rbcL protein of *Bacillariophyceae* – fresh water indicator species given by Sladechek's list and species from Kaban Lakes (Kazan, Russia). The rbcL sequences of 66 indicator species of *Bacillariophyceae* were performed in international databases GenBank Nucleotide and GenPept Protein Sequences on NCBI website (www.ncbi.nlm.nih.gov).

Multiple alignments of nucleotide and protein of rbcL sequences of indicator species were made with Clustal Omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/) [7]. Molecular phylogenetic trees by Neighbor-Joining (NJ) method [8-11] and Maximum Parsimony method (MP) [9, 11, 12] were constructed using the MEGA program (www.megasoftware.net) [13].

***Corresponding Author**
Email:
Anthony.Sverdrup8@gmail.com

IIOAB JOURNAL
ISSN: 0976-3104

[Table 1] shows the indicator species of *Bacillariophyceae* from the Sladechek's list and from Kaban Lakes with their saprobity and accession numbers from GenBank and GenPept databases. The new species name, if renamed, shown in parentheses.

**Table 1:** Accession numbers obtained from GenBank and GenPept databases of rbcL Bacillariophyceae species with the saprobity

| Species | Saprobity* | rbcL gene GenBank | rbcL protein GenPept | Species of Kaban lakes |
|---|---|---|---|---|
| *Achnanthidium coarctatum* | x | HQ912458 | AEB91214 | - |
| *Achnanthes minutissima (Achnanthidium minutissimum)* | o-b | KY863484 | ASK50010 | - |
| *Amphora ovalis* | b-o | KC954577 | AHK61273 | + |
| *Amphora pediculus* | - | HQ912403 | AEB39384 | + |
| *Anomoeoneis sphaerophora* | b-o | KJ011795 | AIY31906 | - |
| *Asterionella granulata (A. formosa)* | o-b | HQ912497 | AEB91253 | + |
| *Aulacoseira granulata* | b-o | KM999116 | AKS44355 | + |
| *Caloneis amphisbaena* | b-a | KM084980 | AIT92080 | - |
| *Caloneis silicula* | o-b | JN418663 | AER42044 | - |
| *Cocconeis pediculus* | b-o | KM084977 | AIT92077 | - |
| *Cocconeis placentula* | o-b | HQ912456 | AEB91212 | + |
| *Cocconeis placentula var. euglypta* | - | KT072907 | ALO24317 | + |
| *Cyclotella bodanica* | o | DQ514829 | ABF60387 | + |
| *Cymatopleura elliptica* | b | KX120659 | AQM56457 | - |
| *Cymatopleura solea* | b-a | KX120661 | AQM56459 | - |
| *Cymbella affinis* | b-o | KJ011796 | AIY31907 | - |
| *Cymbella aspera* | b | KJ011798 | AIY31909 | - |
| *Cymbella cistula* | b-o | KJ011802 | AIY31913 | - |
| *Cymbella granulata (C. helvetica)* | x-o | KJ011804 | AIY31915 | - |
| *Cymbella lanceolata (Brebissonia lanceolata)* | b | KJ011806 | AIY31917 | - |
| *Cymbella naviculiformis (Cymbopleura naviculiformis)* | b | KJ011815 | AIY31926 | - |
| *Didymosphenia granulata (D. geminata)* | x | KJ011819 | AIY31930 | - |
| *Encyonema ventricosum* | o-b | KU052340 | ALN50604 | + |
| *Epithemia sorex* | b | HQ912395 | AEB39376 | - |
| *Epithemia turgida* | b | KX120566 | AQM56364 | - |
| *Eunotia pectinalis* | x | HQ912500 | AEB91256 | - |
| *Fallacia pygmaea* | a | HQ912469 | AEB91225 | - |
| *Fragilaria capucina* | b-o | KC736594 | AGG86633 | + |
| *Fragilaria crotonensis* | o-b | KF959640 | AHE78119 | + |
| *Gomphonema acuminatum* | b-o | KJ011853 | AIY31964 | - |
| *Gomphonema angustatum* | o | KJ011835 | AIY31946 | - |
| *Gomphonema capitatum* | b | AY571751 | AAT78574 | - |
| *Gomphonema clevei* | x | JQ354682 | AFV95053 | - |
| *Gomphonema intricatum* | o-x | KJ011840 | AIY31951 | - |
| *Gomphonema intricatum v. pumilum (G. pumilum)* | o | KC736599 | AGG86638 | - |
| *Gomphonema parvulum* | b | JQ354693 | AFV95052 | - |
| *Gomphonema truncatum* | - | AM710509 | CAM97966 | + |
| *Gyrosigma acuminatum* | b | KM999078 | AKS44317 | - |
| *Melosira granulata var. angustissima (Aulacoseira granulata v.angustissima)* | b-o | FJ002130 | ACS92840 | - |
| *Melosira varians* | b-o | KM999081 | AKS44320 | - |
| *Navicula cryptocephala* | a | HQ912467 | AEB91223 | + |
| *Navicula tripunctata* | b-o | KM084935 | AIT92035 | - |

IIOAB JOURNAL
ISSN: 0976-3104

| Species | Saprobity* | rbcL gene GenBank | rbcL protein GenPept | Species of Kaban lakes |
|---|---|---|---|---|
| *Navicula gregaria* | b | KY320297 | ASC55339 | - |
| *Navicula radiosa* | b-o | KM084955 | AIT92055 | - |
| *Nitzschia acicularis* | a | KX889095 | ASF62417 | + |
| *Nitzschia dissipata* | o-b | KY320333 | ASC55375 | - |
| *Nitzschia fonticola* | o-b | HF675068 | CCQ77735 | - |
| *Nitzschia linearis* | o-b | KT072917 | ALO24327 | + |
| *Nitzschia palea* | a | FN557017 | CBH19895 | - |
| *Nitzschia sigmoidea* | b | FN557033 | CBH19911 | - |
| *Pinnularia borealis* | x-o | JN418640 | AER42021 | - |
| *Pinnularia viridis* | b | KM350021 | AKH66073 | - |
| *Rhoicosphenia abbreviata* | b-o | KJ011854 | AIY31965 | - |
| *Rhopalodia gibba* | o | KX120556 | AQM56354 | - |
| *Stauroneis acuta* | o | HQ912443 | AEB91199 | - |
| *Stauroneis anceps* | b | AM710475 | CAM97934 | - |
| *Stauroneis phoenicenteron* | b-o | KM084992 | AIT92092 | - |
| *Stephanodiscus hantzschii* | a | AB831882 | BAV19460 | + |
| *Surirella biseriata* | b | JX033009 | AGE34629 | - |
| *Surirella capronii* | b | JX033000 | AGE34620 | - |
| *Surirella splendida* | b-o | HQ912401 | AEB39382 | - |
| *Surirella spiralis* | o | JX032964 | AGE34584 | - |
| *Surirella tenera* | b | JX033012 | AGE34632 | - |
| *Synedra ulna (Ulnaria ulna)* | b | HQ912454 | AEB91210 | + |
| *Tabellaria flocculosa* | o-x | HQ912448 | AEB91204 | + |
| *Ulnaria delicatissima var. angustissima* | o | KT072900 | ALO24310 | - |

***** x- (pure), o-(clean), b- (polluted), a- (very polluted)

## RESULTS AND DISCUSSION

As known, the traditional technology needs of experimental work with living organisms. As a result, it takes a long time before the specie can be determined as a bioindicator. In a compare, our innovative approach uses modern methods of bioinformatics and molecular phylogenetics, which allow us to determine new bioindicators faster than by using the traditional technology, using less time and resources.

Thus, the new approach includes the selection of 66 primary sequences of rbcL gene and rbcL protein of *Bacillariophyceae* bioindicators from international databases; multiple alignment of all sequences; construction of phylogenetic trees by rbcL gene and rbcL protein of *Bacillariophyceae* bio indicators using computers.

As a result, two phylogenetic trees on rbcL gene [Fig. 1] and two phylogenetic trees on rbcL protein [Fig. 2] of 66 indicator species of *Bacillariophyceae* were constructed by NJ- and MP-methods with bootstrap. The percentage of bootstrap from 100 replicas for NJ/MP trees are shown accordingly next to the nodes.

### Phylogenetic analysis of rbcL gene of bacillariophyceae

As can be seen from [Fig. 1], there are 17 clusters on the phylogenetic tree with high bootstrap more than 50%, three of them include non-indicators species:

- cluster 1 includes species from the same genus – *Surirella* mainly of b-saprobity;
- cluster 2 includes species from the same genus – *Cymatopleura* mainly of b-saprobity;
- cluster 3 includes species from different genera – *Rhopalodia* and Epithemia mainly of b-saprobity;
- cluster 4 includes species from the same genus – *Aulacoseira* of b-o-saprobity;
- cluster 5 includes species from different genera – *Fallacia* and *Pinnularia* of different-saprobity;
- cluster 6 includes species from the same genus – *Stauroneis* mainly of b-o-saprobity;
- cluster 7 includes species from different genera – *Ulnaria*, *Synedra* and *Fragilaria* mainly of o-b-saprobity;
- cluster 8 includes species from the same genus – *Nitzschia* of a-saprobity;

IIOAB JOURNAL
ISSN: 0976-3104

- cluster 9 includes non-indicator species – *Amphora pediculus* and indicator species –*Amphora ovalis* of b-o saprobity;
- cluster 10 includes species from the same genus – *Navicula* of different-saprobity;
- cluster 11 includes non-indicator species – *Cocconeis placentula* var. *euglypta*, indicator species from different genera – *Cocconeis* and *Rhoicosphenia* of b-o-saprobity and species from different genera – *Cocconeis* and *Achnanthidium* of o-b- saprobity;
- cluster 12 includes species from the same genus – *Gomphonema* of different-saprobity;
- cluster 13 includes non-indicator species – *Gomphonema truncatum* and species from the same genus – *Gomphonema* mainly of b-o-saprobity;
- cluster 14 includes species from different genera – *Cymbella* and *Cymbopleura* of different-saprobity;
- cluster 15 includes species from different genera – *Asterionella* and *Tabellaria* mainly of o-saprobity;
- cluster 16 includes species from different genera – *Cyclotella* and *Stephanodiscus* of different-saprobity;
- cluster 17 includes species from the same genus – *Nitzschia* mainly of b-saprobity.

As we can see from [Fig. 1], all organisms are grouped in the clusters with the same and/or close saprobity with a high bootstrap. For example, cluster 8 includes the species *Nitzchia acicularis* and *Nitzchia palea* of a-saprobity with a high bootstrap more than 95%. This is a very good result.

Phylogenetic analysis of rbcL protein of *bacillariophyceae*

As can be seen from [Fig. 2], in a compare with phylogenetic tree on rbcL gene of *Bacillariophyceae* there are only 12 clusters with high bootstrap more than >50%, three of them include non-indicator species:

- cluster 1 includes species from different genera – *Navicula* and *Gyrosigma* mainly of b-o-saprobity;
- cluster 2 includes species from the same genus – *Nitzschia* mainly of a-saprobity;
- cluster 3 includes species from the same genus – *Nitzschia* mainly of o-b-saprobity and species from different genera – *Stephanodiscus* and *Cyclotella* of different-saprobity;
- cluster 4 includes species from different genera – *Ulnaria*, *Synedra* and *Fragilaria* mainly of o-b-saprobity;
- cluster 5 includes species from the same genus – *Aulacoseira* of b-o-saprobity;
- cluster 6 includes species from the same genus – *Surirella* mainly of b-saprobity;
- cluster 7 includes species from different genera – *Cymatopleura* and *Amphora* mainly of b-saprobity;
- cluster 8 includes species from the same genus – *Epithemia* of b-saprobity;
- cluster 9 includes non-indicator species – *Amphora pediculus* and species from different genera – *Rhopalodia* and *Surirella* of o-saprobity;
- cluster 10 includes species from the same genus – *Gomphonema* of different-saprobity;
- cluster 11 includes non-indicator species – *Cocconeis placentula* var. *euglypta*, and species from the same genus – *Cocconeis* mainly of b-o-saprobity;
- cluster 12 includes non-indicator species – *Gomphonema truncatum*, and species from the same genus – *Gomphonema* mainly of b-o-saprobity.

In a comparison of the rbcL protein tree [Fig. 2] with the rbcL gene tree [Fig. 1], we can see that less species grouped in clusters with high bootstrap. We may make the conclusion that the comparative analysis of phylogenetic trees shows stable clustering of indicator species with the same or close saprobity with higher bootstrap by the rbcL gene than by the rbcL protein.

**Checking assessment results**

We need to check the preliminary assessment results using data from natural lakes, for example the plankton species from the big Kaban Lakes in Kazan, which are situated in the center of the city. The Kaban Lakes are a system of lakes, which includes Nizhny Kaban, Verkhny Kaban, and Sredny Kaban. With a combined area of 1.86 square kilometers, they comprise the biggest lake in Tatarstan Republic (Russia). The ecologists appreciate the lakes as the transition from polluted (b-saprobity) to pure (o-saprobity).

IIOAB JOURNAL
ISSN: 0976-3104

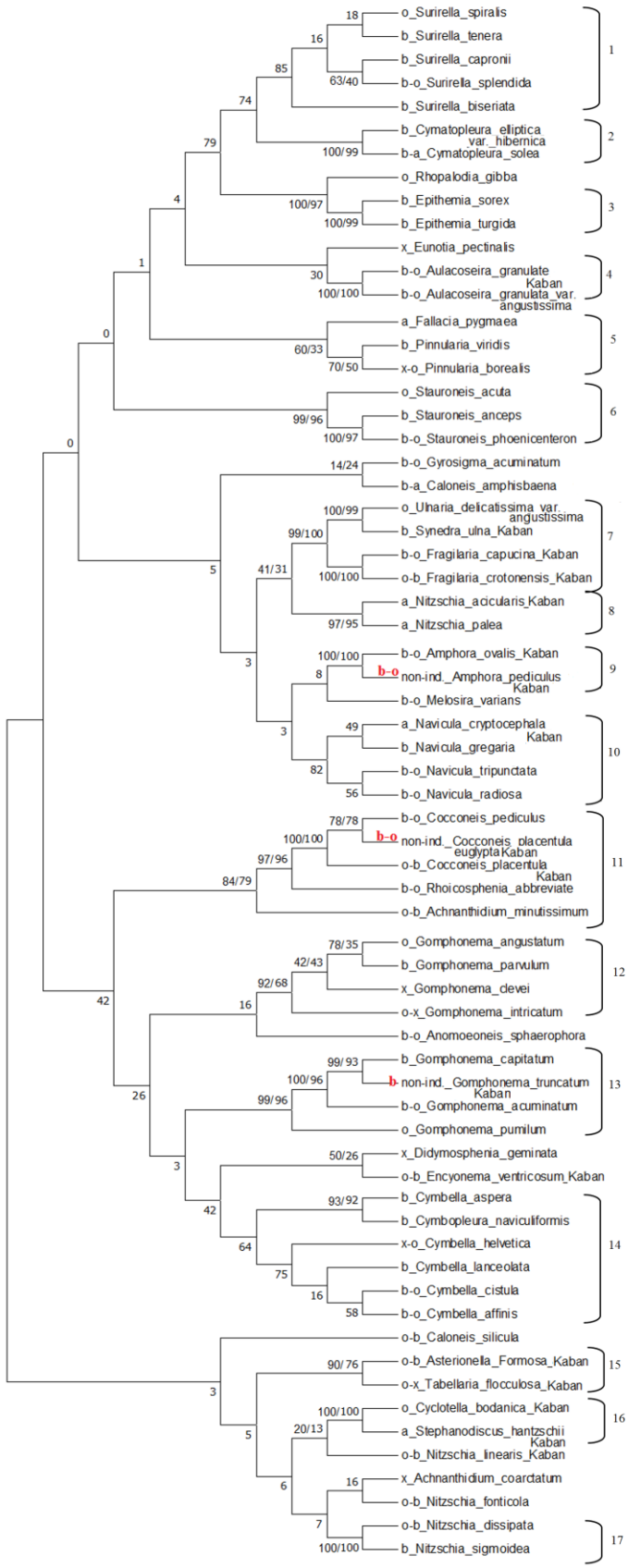


**Fig. 1:** Phylogenetic tree on gene of rbcL Bacillariophyceae (NJ/MP methods).

Molecular phylogenetic analysis includes indicator species and non-indicator species of *Bacillariophyceae* from Kaban Lakes. For non-indicator organisms, the saprobity can be determined based on phylogenetic analysis [Fig. 1]:

IIOAB JOURNAL
ISSN: 0976-3104

- cluster 9 with bootstrap equal 100% includes indicator species *Amphora ovalis* of o-saprobity and non-indicator species - *Amphora pediculus*, that means the last one should be the same o-saprobity [Fig. 3];

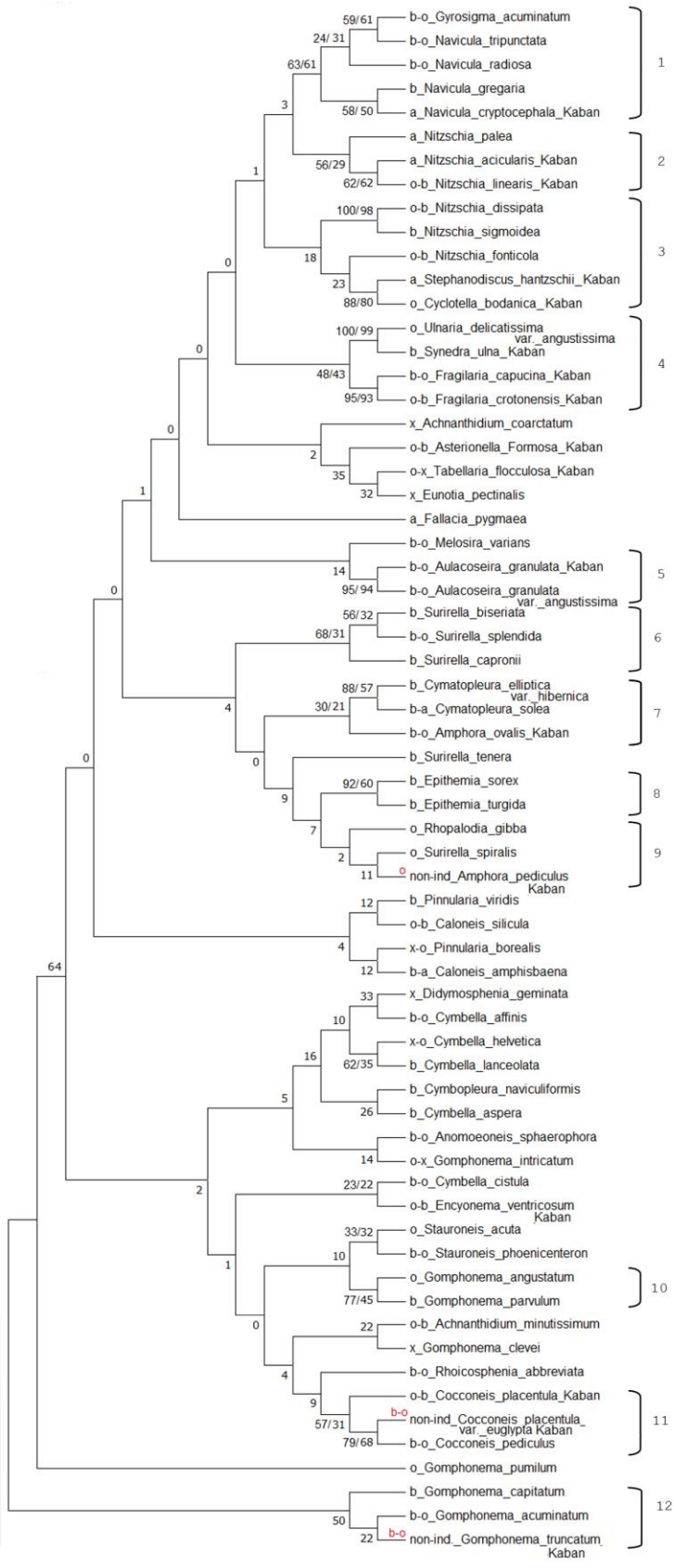


**Fig. 2:** Phylogenetic tree on protein of rbcL *Bacillariophyceae* (NJ/MP methods).

IIOAB JOURNAL
ISSN: 0976-3104

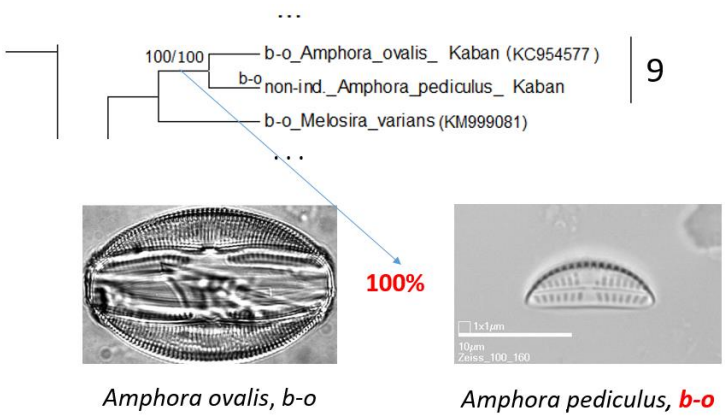


**Fig. 3:** The fragment on gene rbcL *Bacillariophyceae* with the cluster 9.

- cluster 11 with bootstrap equal 78% includes indicator species Cocconeis pediculus of b-o-saprobity and non-indicator species - Cocconeis placentula var. euglypta, that means the last one should be the same b-o-saprobity [Fig. 4];

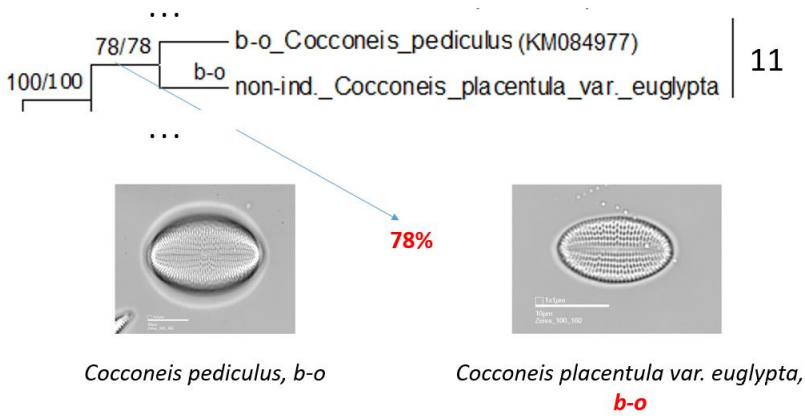


**Fig. 4:** The fragment on gene rbcL *Bacillariophyceae* with the cluster 11.

- cluster 13 with high bootstrap equal 99% includes indicator species Gomphonema capitatum of b-saprobity and non-indicator species - Gomphonema truncatum; that means the last one should be the same b-saprobity with high bootstrap [Fig. 5].

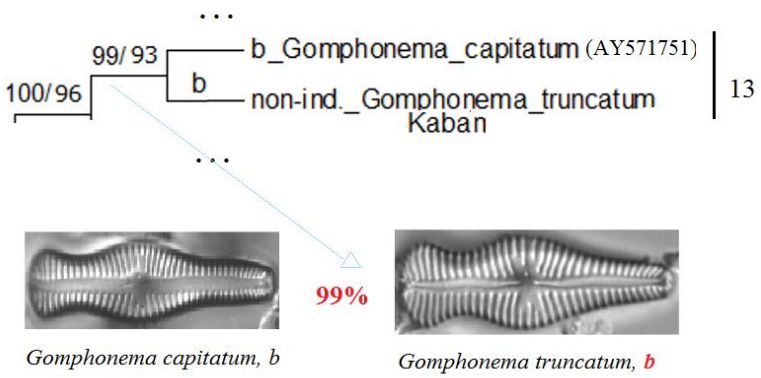


**Fig. 5:** The fragment on gene rbcL Bacillariophyceae with the cluster 13.

As a result of the phylogenetic analysis we have a new list of *Bacillariophyceae* bio indicators for the Kaban Lakes with additional bio indicators: *Amphora pediculus*, *Cocconeis placentula* var. *euglypta* and *Gomphonema truncatum* [Table 2].

IIOAB JOURNAL
ISSN: 0976-3104

**Table 2:** The list of *Bacillariophyceae* biondicators for the Kaban Lakes with additional bioindicators species

| Species | Saprobity | rbcL gene | rbcL protein |
|---|---|---|---|
| *Tabellaria flocculosa* | o-x | HQ912448 | AEB91204 |
| *Nitzschia linearis* | o-b | KT072917 | ALO24327 |
| *Cyclotella bodanica* | o | DQ514829 | ABF60387 |
| *Amphora ovalis* | b-o | KC954577 | AHK61273 |
| *Melosira granulata (Aulacoseira granulate)* | b-o | KM999116 | AKS44355 |
| *Synedra ulna (Ulnaria ulna)* | b | HQ912454 | AEB91210 |
| *Navicula cryptocephala* | a | HQ912467 | AEB91223 |
| *Nitzschia acicularis* | a | KX889095 | ASF62417 |
| *Stephanodiscus hantzschii* | a | AB831882 | BAV19460 |
| *Amphora pediculus* | b-o | HQ912403 | AEB39384 |
| *Cocconeis placentula v.euglypta* | b-o | KT072907 | ALO24317 |
| *Gomphonema truncatum* | b | AM710509 | CAM97966 |

In case of gene sequences absent in international database, it is easy to get the experimental sequences from water samples.

## CONCLUSIONS

Recent results allow us to conclude that a new technology will make water quality monitoring and assessment more efficient. Thus, the new method can be used for effective determination of bio indicators. The comparative analysis of the phylogenetic trees shows stable clustering of indicator species with the same and/or close saprobity with higher bootstrap by the rbcL gene than by the rbcL protein. The technology allows us to determine the saprobity for non-bio indicators based on phylogenetic analysis and add new bio indicators in the total list in an easier way than by experimental work.

### CONFLICT OF INTEREST

There is no conflict of interest.

### ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

### FINANCIAL DISCLOSURE

None.

## REFERENCES

[1] Sladecek V. [1973] System of water quality from the biological point of view. Archiv für Hydrobiologie Beiheft 7: 1–218.
[2] Lukashov VV. [2009] Molecular Evolution and Phylogenetic Analysis. (BINOM, Laboratory of Knowledge, Moscow). https://doi.org/10.1371/journal.pone.0118669.
[3] Hebert PD, Cywinska A, Ball SL, de Waard JR. [2003] Biological identifications through DNA barcodes. Proc Roy Soc Lond B. 64(2): 272-295.
[4] CBOL Plant Working Group [2009]. A DNA barcode for land plants. Proc Natl Acad Sci USA. 106 (31):12794-12797.
[5] Frolova LL, Sverdrup AE. [2018] The monitoring of Nizhniy Kaban Lake by rbcL gene of freshwater organisms using next-generation sequencing. Research journal of pharmaceutical, biological and chemical sciences (RJPBCS). 9(1):262-271.
[6] Kharchenko A, Sverdrup AE, Frolova LL. [2018] The monitoring of Verkhniy Kaban Lake by rbcL gene of freshwater organisms using next-generation sequencing. International Journal of Green Pharmacy. 12(03):756-762.
[7] Multiple Sequence Alignment - Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo).
[8] Saitou N, Nei M. [1987] The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4(4):406 – 425.
[9] Felsenstein J. [1985] Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39:783-791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.
[10] Tamura K, Nei M, Kumar S. [2004] Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences of the United States of America. 101 (30):11030-11035
[11] Kumar S, Stecher G, Tamura K. [2016] MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution. 33(7):1870-1874
[12] Nei M, Kumar S. [2000] Molecular Evolution and Phylogenetics. New York: Oxford University Press. 333. https://doi.org/10.1371/journal.pone.0118669.
[13] Molecular Evolutionary Genetics Analysis (MEGA) www.megasoftware.net.