



RESEARCH ARTICLE

REVISED The signs of adaptive mutations identified in the chloroplast genome of the algae endosymbiont of Baikal sponge. [version 2; peer review: 3 approved, 1 approved with reservations]

Sergey Feranchuk ^{1,2}, Natalia Belkova^{1,3}, Lubov Chernogor¹, Ulyana Potapova¹, Sergei Belikov¹

¹Limological institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, 664033, Russian Federation

²Department of Informatics, National Research Technical University, Irkutsk, 664074, Russian Federation

³Scientific Centre for Family Health and Human Reproduction Problems, Irkutsk, 664033, Russian Federation

v2 First published: 04 Sep 2018, 7:1405
<https://doi.org/10.12688/f1000research.15841.1>
 Latest published: 14 May 2020, 7:1405
<https://doi.org/10.12688/f1000research.15841.2>

Abstract

Background: Monitoring and investigating the ecosystem of the great lakes provide a thorough background when forecasting the ecosystem dynamics at a greater scale. Nowadays, changes in the Baikal lake biota require a deeper investigation of their molecular mechanisms. Understanding these mechanisms is especially important, as the endemic Baikal sponge disease may cause a degradation of the littoral ecosystem of the lake. Methods: The chloroplast genome fragment for the algae endosymbiont of the Baikal sponge was assembled from metagenomic sequencing data. The distributions of the polymorphic sites were obtained separately for the genome fragments from healthy, diseased and dead sponge tissues. Results: The distribution of polymorphic sites allows for the detection of the signs of extensive mutations in the chloroplasts isolated from the diseased sponge tissues. Additionally, the comparative analysis of chloroplast genome sequences suggests that the symbiotic algae from Baikal sponge is close to the Choricystis genus of unicellular algae. Conclusions: Mutations observed in the chloroplast genome could be interpreted as signs of rapid adaptation processes in the symbiotic algae. The development of sponge disease is still expanding in Baikal, but an optimistic prognoses regarding a development of the disease is nevertheless considered.

Keywords

Chlorophyta, Lake Baikal, Chloroplast Genome, Genetic Polymorphism, Mutation Rate

Open Peer Review

Reviewer Status

	Invited Reviewers			
	1	2	3	4
version 2 (revision) 14 May 2020		 report	 report	 report
version 1 04 Sep 2018	 report	 report		

1. **Michael G. Sadovsky** , Siberian Federal University, Krasnoyarsk, Russian Federation
2. **Roman Kondratov** , Cleveland State University, Cleveland, USA
3. **Ashraf Al-Ashhab** , Dead Sea and Arava Science Center, Masada, Israel
4. **OPhir Nave** , Jerusalem College of Technology, Jerusalem, Israel

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Sergey Feranchuk (feranchuk@gmail.com)

Author roles: **Feranchuk S:** Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Belkova N:** Conceptualization, Data Curation, Investigation, Resources, Writing – Review & Editing; **Chernogor L:** Investigation, Resources, Writing – Review & Editing; **Potapova U:** Validation, Visualization, Writing – Review & Editing; **Belikov S:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was supported by the Ministry of Education and Science of the Russian Federation by Government contract project no. 0345-2015-0002, “Molecular Ecology and Evolution of Living Systems of Central Asia in Terms of Fishes, Sponges, and the Microbial Flora Associated with Them” [VI.50.1.4], and the Russian Foundation for Basic Research [16-04-00065, 16-54-150007, 18-04-00224].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Feranchuk S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Feranchuk S, Belkova N, Chernogor L *et al.* **The signs of adaptive mutations identified in the chloroplast genome of the algae endosymbiont of Baikal sponge. [version 2; peer review: 3 approved, 1 approved with reservations]** F1000Research 2020, 7:1405 <https://doi.org/10.12688/f1000research.15841.2>

First published: 04 Sep 2018, 7:1405 <https://doi.org/10.12688/f1000research.15841.1>

REVISED Amendments from Version 1

The text of Introduction was rewritten, to remove specific terms and unnecessary references, and to make a narrative more ordered. The most of reference links were removed from the text to a separate table, to improve readability.

The text of Methods was modified, mostly in several parts:

- the description of sponge collection technology was inserted
- the description of bioinformatics pipeline was updated; the specifications of the software were removed from the text to a separate table.

The texts of Results and Discussion were rewritten almost completely.

- Subheadings were introduced, and the ordering of reported results was changed.
- First subsection include description of auxiliary tests, reservations about confidence of results and limitations of the approach. The annotation of chloroplast genome is mentioned in this section in brief, Figure 1 from first version was excluded in the new version.
- second subsection present distribution of polymorphic sites and hypothesis about adaptive mutations. Figure 3 (from first version) was improved in vector graphics editor and become Figure 1 in the new version. Figure 2 in the new version was introduced as a "vector graphics - styled" qualitative illustration in a support of the proposed hypothesis.
- third subsection describe the phylogenetic analysis and taxonomic assignment of the algae. Figure 2 (old version) become Figure 3 (in new version), one extra tree was added to the figure to clarify distribution of algae strains across Baikal.

The content of Discussion section was changed completely. Just the hypotheses about adaptive mutations is discussed in the new version, any auxiliary subjects were removed.

Any further responses from the reviewers can be found at the end of the article

Introduction

Lake Baikal in South-eastern Siberia is the source of the Angara River and quite close to Baikal – the source of the Lena River; both of these rivers flow far north to the Northern Ocean. The climate in the Baikal region is continental; in winter, the surface of the lake is covered with ice. Many of the plant species in the forests and steppes surrounding Lake Baikal no longer grow anywhere else in the world. Baikal seals, ‘nerpa’, some fish species,

and many other creatures that inhabit the lake, adapted to life in the lake, far separated from their closest relatives in the seas and in other lakes.

Baikal sponges live on the bottom of the lake in the coastal zone and have a greenish colour due to their ability to absorb sunlight. The adult sponge *Lubomirskia baicalensis*, after many decades of growth, becomes similar in shape and size to a low branchy tree. The green colour of sponges is caused by a photosynthetic symbiont, a unicellular green alga, adapted to live within the body of sponge cells, contributing in this way to its feeding.

Rapid changes in the ecosystem of Lake Baikal have been observed since 2010–2011. One of the central attributes of these changes is a severe disease and death of sponges, which is now observed in almost all parts of the lake. The symptoms of the disease start with the appearance of pink and brown spots on the surface of the sponge, which eventually lead to complete destruction of sponge tissues. The change in colour of sponge tissue indicates that the chloroplasts of the algal symbiont are damaged from the early stages of the disease.

Changes which happen in the microbial communities of sponges at the beginning of the disease have almost nothing in common, but the most probable cause of sponge disease is arranged and simultaneous attacks of heterotrophic microorganisms of different origin. Common to these pathogens is high changeability, noticed in the relatively high variation of genotypes and in a strategy to be ‘opportunistic’.

A series of research has been conducted on sponges in Baikal and their symbionts, overall signs of crisis in Baikal, and the reasons for sponge disease. Some of the appropriate publications are listed in Table 1, with a comment that the scope of the research is not too wide due to cases of misunderstanding, conflicts of interest and limited funding. But a detailed study of sponge disease provides an opportunity to observe the ways in which the unique ecosystem of Baikal has changed to pass the crisis. Since the advent of molecular biology, no crisis of so large a scale has happened, so the expected results may valuably expand the scope of knowledge about the evolution of alga species.

Table 1. References with the scope of research on sponges, their symbionts, ecology and evolution used in the recent study.

The scope of the research	References
Baikal and it's sponges	Chernogor <i>et al.</i> , 2013; Pile <i>et al.</i> , 1997
Crisis on Baikal	Bormotov, 2012; Khanaev <i>et al.</i> , 2018; Kravtsova <i>et al.</i> , 2014; Timoshkin <i>et al.</i> , 2016
Disease of sponges	Belikov <i>et al.</i> , 2019; Feranchuk <i>et al.</i> , 2018; Kulakova <i>et al.</i> , 2018
Evolution of Algae	Cai <i>et al.</i> , 2017; Fucíková <i>et al.</i> , 2014; Lemieux <i>et al.</i> , 2014; Lemieux <i>et al.</i> , 2015; Sun <i>et al.</i> , 2016
Adaptation and Evolution	Rosenberg, 2001; Wright, 2004

The algal symbiont of sponge is as conservative as other major constituents of the sponge hologenome in a healthy state. This alga is annotated by taxonomy as a Chlorophyta symbiont of *Lubomirskia* sp. (NCBI Tax. ID 752245); the closest of its relatives is the *Choricystis* genus of coccoid algae. Chloroplasts of the alga are abundant in sponge tissues and are a suitable object of focus for detailed investigations.

The taxonomy of the genus *Choricystis* and other similar unicellular algae from Chlorophyta division has been investigated thoroughly with the use of complete chloroplast genomes. But in a wider context of algal evolution, the credibility of the taxonomy may look insufficient, and advanced methods like taxon sampling, genome synteny analysis or estimates of site heterogeneity have been applied to improve it.

Matters concerning the evolution of a whole clade of algae are anyway 'enigmatic'. One of the ways to clarify it is to suppose that, in some periods of algal history, the rate of mutations was accelerated for some lineages. The possibility of an accelerated mutation rate, or 'stress-induced mutagenesis', or 'adaptive mutagenesis', or 'environment-associated increases in mutations' is discussed in the theory of evolution but is often treated as controversial. In the case of normal development, a need for accelerated adaptation is compensated by a need to keep the stability of a genotype, and any visible acceleration of mutations is almost excluded. But in the case of the sponge disease, the stress is deadly. And, due to the uniqueness of Baikal, with another look the phenomenon of adaptation may be discovered.

To obtain raw data for the analysis, three samples of sponge tissue were collected and processed using metagenome and metatranscriptome sequencing. This allowed the assembly of a fragment of a reference genome for the alga chloroplast, which covered at least half of the expected length. After this, a straightforward approach to detect a phenomenon of adaptation is to look at distributions of substitutions at polymorphic sites of the chloroplast genome in the three samples.

As one of the precedents of similar studies, the chloroplast genome of an uncultivated alga was obtained from metagenome sequencing data [Worden *et al.*, 2012]. And, in an extensive study of the population structure of microbial communities [Truong *et al.*, 2017], distributions of polymorphic sites were carefully investigated. The methods and ideas from those studies were suitable to be applied in the proposed analysis.

A precise taxonomic assignment for the alga species under study is also necessary for consistency of analysis. The composition of alga species across the whole lake is uneven, but the precision of amplicon sequencing only allows the selection of several closest entries in a reference database for the 16S rRNA gene. Classification of the alga species at the level of the 16S rRNA gene has been reported in previous publications about Baikal, and the results from these studies were used as a reference.

Materials and methods

Sampling and sequencing

The technology for collection of sponges by scuba divers, with the use of labelled transects, has developed since the detection of sponge disease and is described by Khanaev *et al.* [2018]. The samples collected by divers were immediately placed in containers with Baikal water over ice and transported to the lab, maintaining a constant water temperature.

The three samples of freshwater sponge *L. baicalensis* were collected in June 2016 in the Bol'shiye Koty area (51° 90' 69 N'', 105° 07' 05 E'') at a depth of 10 m within the same transect. One sample was obtained from a sponge that was healthy in appearance (exhibiting a green colour), one sample was taken from a diseased sponge and one from dead rotten sponge tissues.

Illumina pair-end reads were obtained by DNA metagenome sequencing at Novogene Inc. (Illumina PE 150), for all three samples; in addition, they were processed by a conventional bioinformatics pipeline, which included the filtering of sequencing errors and the assembly of contigs. The extraction and sequencing of RNA samples was also performed at Novogene Inc. This technique was possible only for healthy and diseased sponge tissues; not enough RNA was extracted from the rotten tissues.

Assembly and annotation of the chloroplast genome fragment

Chloroplast species are abundant enough in the genetic material available from metagenomic sequencing of sponge tissues. A template-based assembly is suitable for obtaining the sequence of a chloroplast genome or at least a substantial fragment of the genome. The sequence of the chloroplast genome of *Choricystis parasitica* (NC_025539) was used as a template for assembly and comparative analysis. This genome is a circular DNA with a length of 94206 base pairs.

To get a refined template from the scaffolds, obtained by de novo assembly of each separate metagenomic sample, scaffolds were selected which had a similarity to the template genome. Then, from cleaned reads, from both DNA and RNA sequencing, reads were selected which had a similarity to any of the selected scaffolds. The reads selected in all samples were merged to a single volume, and both reads in a pair were treated as unpaired. The 'lightweight' assembler Inchworm from the TrinityRnaSeq package was then applied to the volume of filtered reads, adjusted to the maximal size of k-mer (K = 31) and minimal sensitivity to sequencing errors. The contigs obtained after the second assembly were compared with the reference genome. This made it possible to select a single contig of 55638 bp in length, which was with confidence identified as a fragment of the chloroplast genome. The stages of the pipeline are listed in Table 2, to explain the data flow at each stage, and to provide the software specifications.

To check the obtained result, steps 3 and 5 were repeated with other settings: refined templates were obtained by running

Table 2. Stages of assembly and annotation of the chloroplast genome fragment.

	Stage	DNA	RNA	Type of output	Software / reference
1	cleaning	(Outsourced)	Trimmomatic 0.35	reads	Bolger et al., 2014
2	first assembly	(Outsourced) SoapDeNovo	-	scaffolds	Luo et al., 2012
3	template refining	blastn	-	secondary templates	Mature
4	filtering	bowtie 2.2.6		reads	Mature
5	second assembly 2	Inchworm 2.6.5		scaffolds	Grabherr et al., 2011
6	annotation	TrnaSCAN 1.4, Mummer 3.23		genes	Kurtz et al., 2004 ; Lowe & Eddy, 1997
7	phylogeny	Mafft 7.27, Fastme 2.1.5.1, PhyML 20160207		trees	Guindon & Gascuel, 2003 ; Kato & Standley, 2013 ; Lefort et al., 2015

both blastn with an evaluate of $1e-8$ instead of $1e-40$, and tblastx; the Inchworm assembly was run with default settings instead of the setup with the lowest sensitivity. The assembled contigs were not completely identical to the contig selected as a result, but the difference in any case was fewer than 10 substitutions.

Open reading frames were identified in the obtained contig and most of the identified proteins were annotated following the annotations of the reference genome. TrnaSCAN software was used to identify 13 transport RNA genes in the putative chloroplast sequence, and the locations of 18S rRNA and 23S rRNA were identified by direct alignment with reference rRNAs.

Identification of polymorphisms

In order to distinguish the polymorphic sites from traces of sequencing errors, their selection in the genome was implemented following a previously published approach [[Truong et al., 2017](#)]. Each RNA and DNA sample was represented as pair-end reads and separately aligned to the assembled fragment of the chloroplast genome using Bowtie2. The alignments were then processed using the Samtools 1.7 software pipeline with conventional settings, and the indexed archives of alignments were used as the input of the algorithm for the identification of polymorphic sites.

When describing the algorithm, s represents each position on the alignment of the reads, N_s is defined as the total number of reads and T_s is defined as the number of reads supporting the most abundant allele. Given the sequencing error rate E , the non-polymorphic null hypothesis was rejected if the probability that the number of reads equivalent to $N_s - T_s$ coming from the non-dominant allele was $< \alpha = 0.05$. This value was estimated using the probability mass function of a binomial distribution with N_s trials and a successful rate of $1 - E$. The error rate was set to 0.01 for Illumina sequencing. Bases with a quality below 30 were removed, and reads with an average identity to

the reference below 99% were ignored before applying the statistical test. Failing to reject the null hypothesis reflected the absence of alternative alleles or the inability to distinguish between low-coverage potential alternative alleles and sequencing noise.

Therefore, the number of polymorphic sites can be counted for each gene. Another property of each gene is the number of polymorphic sites where the count of alternative alleles is higher than the count of the dominant allele ($T_s < N_s - T_s$). This property could be used, in addition to the fraction of mutations in each gene, to detect the intensity of mutations in the DNA and RNA obtained in the samples.

Phylogenetic analysis

The chloroplast genome sequences of *Picocystis salinarum* (NC_024828), *Myrmecia israelensis* (KM462861), *Botryococcus braunii* (KM462884), *Coccomyxa subellipsoidea* (NC_015084), *Hydrodictyon reticulatum* (NC_034655), *Mychonastes jurisii* (NC_028579) and *Chlorella vulgaris* (NC_001865) were used to reconstruct the phylogenetic tree for the 16S ribosomal RNA (rrs gene) and ATP synthase subunit beta (atpB gene).

The 16S rRNA reference sequences from the Greengenes gg_13_7 database were used to clarify the relations between subspecies of the algae in Lake Baikal. These sequences were identified in the studies of [Belikov et al. \[2019\]](#) and [Feranchuk et al. \[2018\]](#) as the 16S rRNA templates which are extensively represented in the sponge microbiomes collected in different locations of the lake. Both templates, 4365343 (gb|JF495289.1) and 1118847 (gb|GU936925.1), were directed to the *C. parasitica* rRNA (SAG:17.98, gb|KM462878) as the closest annotated species, by a blast search at NCBI site (98% and 96%).

The nucleotide sequences of the selected genes were aligned using Mafft. A straightforward distance-based approach was used for phylogenetic analysis of the 16S rRNA and atpB genes of the chloroplasts of alga species from the Chlorophyta division. These

trees were constructed using Fastme, with a neighbour-joining method to select tree topology and the Jukes–Cantor measure to calculate the distances between genes. A maximum-likelihood algorithm was selected to be used for a tree which was based on 16S rRNA fragments of the *Choricystis* clade, to consider the contribution of the nucleotide substitution model to branch lengths in the constructed trees. A comparison of 16S rRNA genes for the selected strains was performed using PhyML with the default parameters (--model HKY85 -d nt).

Results

Remarks about the confidence of the assumptions

Indirect signs of adaptive mutations were detected in the integral distribution of nucleotides in polymorphic sites of the chloroplast genome. The polymorphisms detected in the samples could occur due to sequencing errors or to custom variations in the populations of alga substrains in closely located sponges, not only due to an adaptation to stress.

The samples were collected at the same place, and the alga strain was expected to be the same in all three samples. But micro-populations of algae can anyway be separated in any sponge, and their development can anyway change in response to stress. So, signs of an acceleration of mutations can be demonstrated only at a qualitative level. That is, assuming that there are several substrains of algae in the three samples, which are considered as a single environment, from the distribution of nucleotides at polymorphic sites one can estimate the relative abundance of dominant and alternative substrains in each sample. The observed difference in these abundances can be naturally explained, if the hypothesis about acceleration of mutations is assumed. It can also be explained by another reason, or just accepted as a random case, but to exclude the assumption about adaptive mutations would be a kind of ‘reduction’, an unnecessary simplification of the reality.

This approach should be treated not as a confident explanation of the observed event, but rather as a proof of the possibility. But even an approved possibility of extremely high mutation rates is of value, to clarify the ‘enigmas’ which are noticed in the evolution of algae, and to suggest the ways in which the community of sponges could save itself in a time of crisis.

The questions which should be answered to specify the degree of confidence in the presented results are: how adequate is the assembly of the chloroplast genome? To what extent is the estimated proportion of alternative alleles in polymorphic positions caused by an adaptation, compared with reasons like the natural diversity of genotypes, and with the biases introduced at the experimental and calculation stages?

The fragment of the genome obtained in the assembly is 55683 nucleotides in length, compared to 94206 nucleotides in the circular DNA of the *C. parasitica* chloroplast. An auxiliary run of the assembly pipeline with different settings showed that the fragment considered as the final result does not include at least 5000 nucleotides. But in all cases, the

differences were fewer than 10 nucleotides in the whole fragment.

The order of annotated genes in the reference genome corresponded to the order of annotated open reading frames in the assembled fragment, with a difference in a single event of reordering, such that the orientation of the segment from *rpoB* to *tufA* genes was reversed. The segment in the reference genome between the *rrs* and *tufA* genes was missing in the assembled fragment; instead, the *psaB* gene from the middle of that segment followed just after the *rrs* gene. The homology between nucleotide sequences of genes was from 80% up to 98%.

The average proportion of polymorphic positions in all genes was 98.4%, which is comparable with the previously reported 97.8% for bacterial genomes [Truong *et al.*, 2017]. These values are also consistent with the results reported by Feranchuk *et al.* [2018]. It was shown that on limiting the selection of 16S rRNA fragments of the chloroplast, sequenced by 454 technology, to 95% identity, variation was at most 2% between any of the fragments.

Coverage values for some coding frames were not sufficient to evaluate the distribution properties of alternative alleles. For all the annotated tRNA genes, coverage was at an adequate level, but no polymorphic positions were detected in any of these genes.

However, for 42 of the 56 annotated genes, the coverage was sufficient to detect polymorphisms with adequate significance. Within the purposes of the research, acceleration of mutations may be detected in the distribution of allele frequencies in polymorphisms only at a qualitative level. For qualitative analysis, for each gene and each sample, a coverage of reads was obtained where any of the alternative nucleotides was substituted in a polymorphic position, and those positions were counted for each gene where the coverage of reads with a dominant allele was less than half of the total coverage. These counts were then used to evaluate the abundance of a dominant strain, relative to minor substrains. The fractions of these numbers for each gene are comparable to the fractions for the whole genome, with acceptable dispersion.

The annotation of the alga species under study, with respect to close species of algae, including the species of sponge symbiotic algae in other parts of the lake, was also refined. This made it possible to compare the variations observed between strains in the three samples collected at the same location with those in strains of algae from sponge samples collected in other locations of the lake. The phylogenetic analysis explained above is reported in the last subsection of the results.

Distribution of polymorphic sites and its interpretation

The fraction of alternative alleles in polymorphic positions is shown in Figure 1, for a whole set of genes and for each separate gene. Within this fraction, the part is separated where a

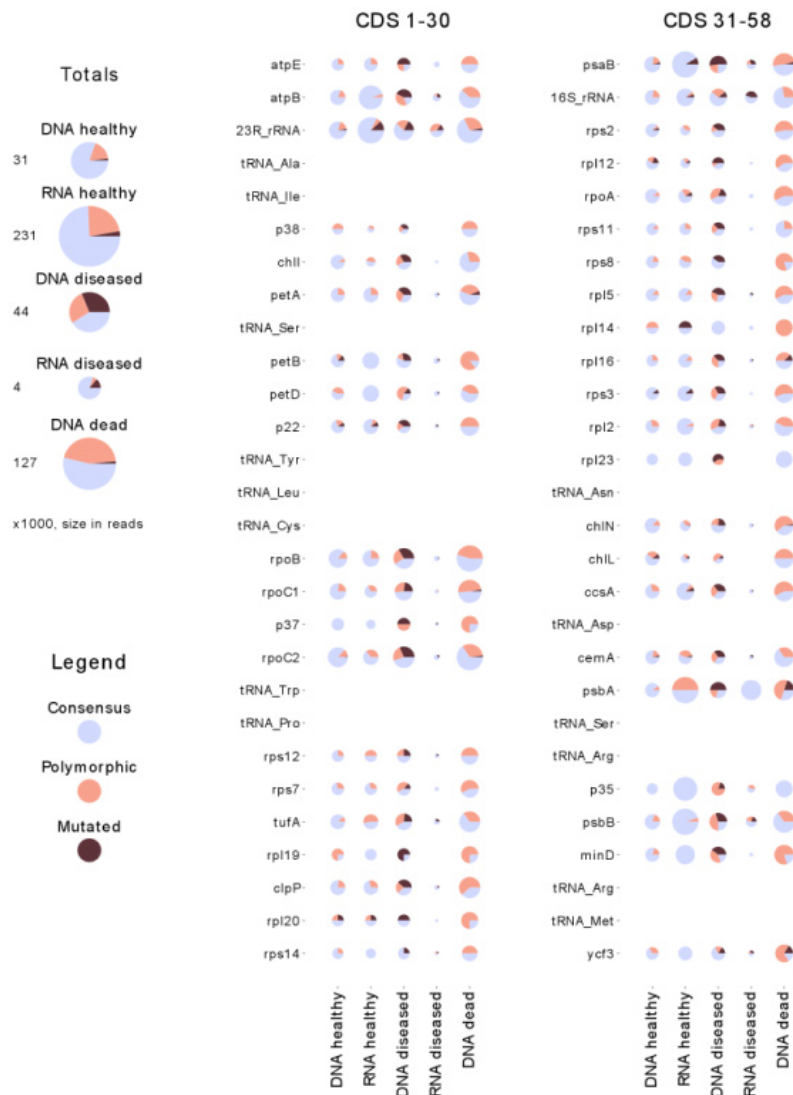


Figure 1. Relative proportion of polymorphic sites in the chloroplast genome of sponge symbionts, for the five samples studied. The results for each of the annotated genes are shown on the right. The left column presents the integrated results for each sample. The proportions of polymorphic sites and sites with high levels of alternative alleles ('mutations') are shown as pie charts, relative to the total number of polymorphic sites in all samples. The proportion of sites which are polymorphic in some other samples, but not in the given sample, are shown in light blue. The legend on the left shows the colour scheme used to represent three types of site. The circle radius represents the total number of sequencing reads aligned to the gene segment and used to identify polymorphic sites. The scale of the circle radii is transformed for better appearance, to compensate for the high variation in the number of aligned reads.

crucially low (< 50%) abundance of baseline nucleotides was detected in the polymorphic positions.

The separation into three fractions, which is shown in [Figure 1](#), allows one to guess, at a qualitative level, the abundance of the dominant strain, relative to the second most dominant strain and minor strains. Also, comparison of distributions for DNA and RNA allows one to guess how intensive is the development in each of the three separated groups, assuming that RNA synthesis is a sign of any development. This qualitative interpretation is shown in [Figure 2](#).

The precise quantitative estimates of these abundances, to our best knowledge, by no means can be obtained from the available volume of data. The abundance of substrains in the three sponges before the development of disease also cannot be reconstructed.

However, for the sample from the diseased sponge, the fraction of alternative alleles was much higher than in the other two samples. So, as a hypothesis which expands the neutral model, it can be suggested that in the alga cells of a diseased sponge, mutations become accelerated to a rate much higher

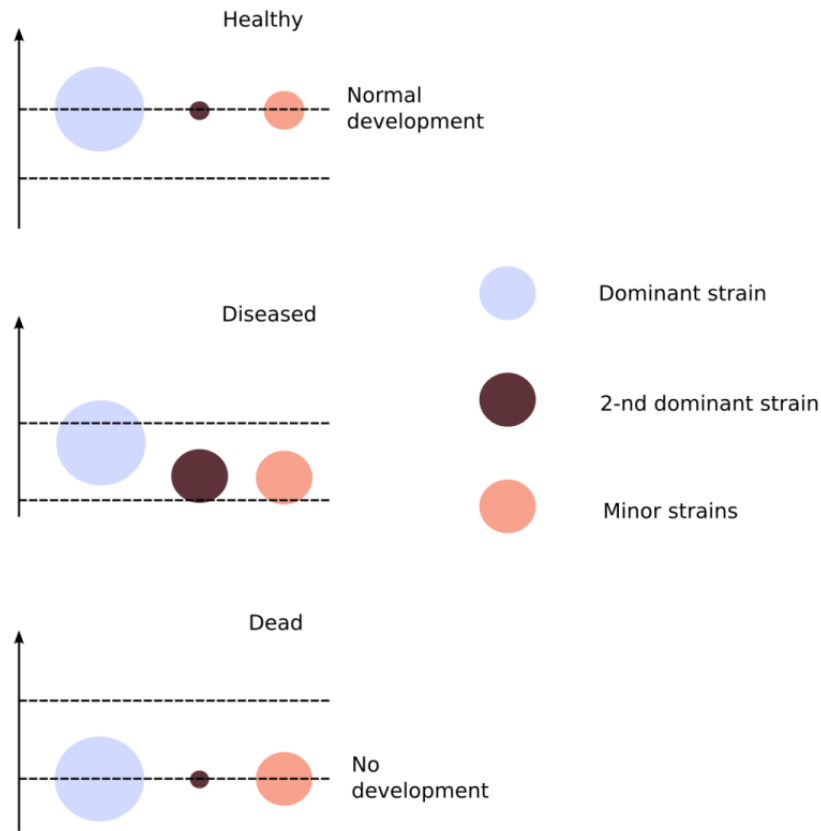


Figure 2. Qualitatively assessed abundance of chloroplast substrains in three samples of sponge. When a sponge is healthy, the major strain dominates, but mutations do not affect the rate of development. This corresponds to a minimal fraction of polymorphisms, and the dominant allele is anyway most frequent in any of the polymorphisms. When a sponge suffers from the disease, it causes, by assumption, an increase in mutation rates and, in turn, an increased abundance and re-distribution of minor strains. This corresponds to an increased fraction of alternative alleles in the genome and, to a lesser extent, in the transcriptome. When a sponge is completely destroyed, just the ‘white noise’ of mutations is found in the genetic material which is left from its chloroplasts.

than any rate compatible with the consistence of metabolic relations. In other words, the living cells in the diseased but alive tissue are desperately trying to survive. Subpopulations of mutated strains develop slowly and die earlier than a dominant lineage – a natural consequence of so many fast mutations in the genome.

But the proposed hypothesis means that the observed acceleration of mutations was a determined event, and some mechanism encoded in the genome triggers this event. But for what reason did this mechanism arise and was it kept conserved in evolution? It is possible just to guess that reason, and several suggestions for that guess are proposed the Discussion section.

Comparative analysis and taxonomic annotation of symbiotic algae

The phylogenetic trees for the two selected genes, 16S rRNA and ATP synthase beta (Figure 3 A, B), in general confirm

the conventional relationship between Chlorophyta algae [Lemieux *et al.*, 2014; Lemieux *et al.*, 2015]. The trees in Figure 3 support the assumption that the symbiotic algae of the sponge *L. baicalensis* are close in taxonomy to the genus *Choricystis*.

The bars in Figure 3 A and B, which show the proportion of polymorphic positions in the genes of symbiotic algae in metagenomic samples, are comparable in size with the scale of distances between genera. The timescale which separates the origins of the close genera in Figure 3 is much larger than the timescale which could adequately interpret the separation of the chloroplast strains detected in the metagenome. So, the observed relation between timescales may be interpreted as being caused by adaptation events, and in that case it can be used to estimate the order of magnitude of the acceleration rate.

Tree C in Figure 3 demonstrates in a more precise way the relation between the species and subspecies of symbiotic algae. The tree is composed an rrs gene for the reference species

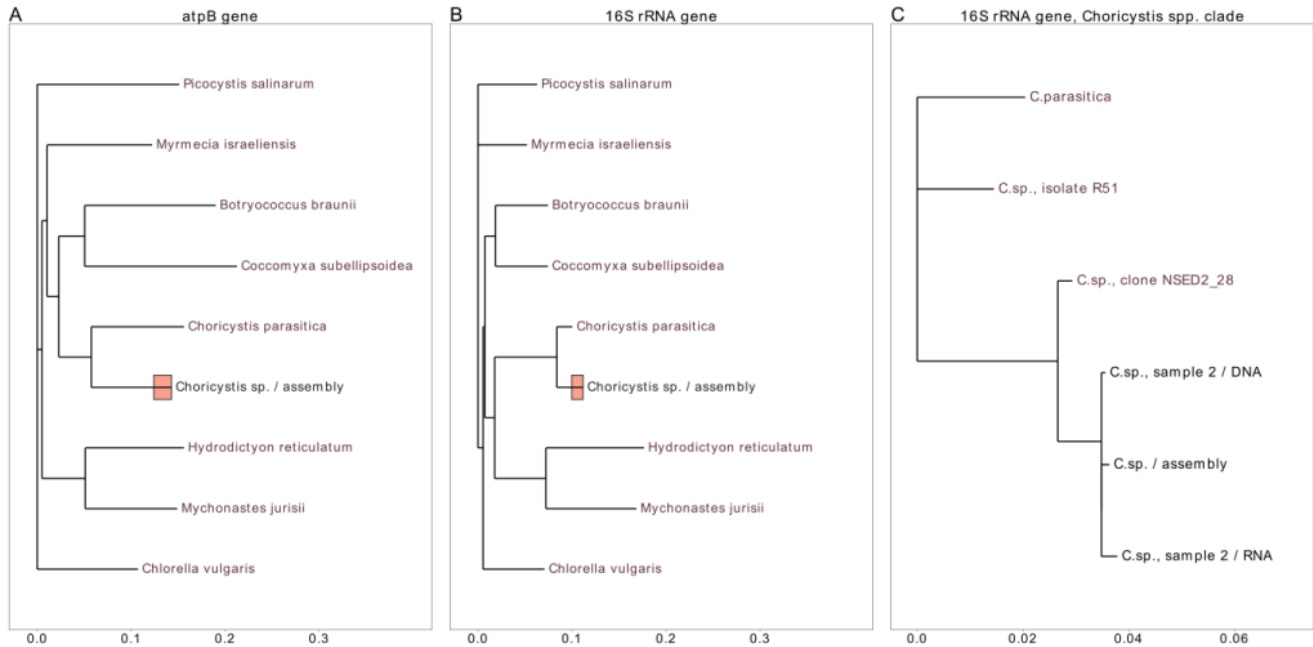


Figure 3. Phylogenetic trees for two chloroplast genes, ATP synthase subunit beta (**A**) and 16S rRNA (**B, C**). The trees in **A** and **B** are constructed for species within the Chlorophyta division; tree **C** is constructed for species within the genus *Choricystis*. Bars located at the node for the studied chloroplast genome in trees **A** and **B** represent the relative number of polymorphic positions in all five samples, at a 1:1 ratio.

of *C. parasitica*, the two abundant chloroplast rRNA templates for the sponges from other parts of the lake, and three consensus sequences for the three sponges under study. From these three sequences, one represents the 16S rRNA gene from the reference assembly, and two sequences represent specifically the sample of diseased sponge, as a DNA and as an RNA. The consensus sequences for the latter two datasets were created from the distribution of polymorphic positions.

The chart in **Figure 3C** illustrates in more detail the estimates of acceleration value provided below. The following assumptions can be introduced: the fraction of polymorphic positions in the whole chloroplast genome is 1.6%; the age of Baikal is about 20 million years, but let the separation of the alga species in Baikal from the rest of the *Choricystis* lineage be about 1 million years ago, and the genetic distance from the alga species under study to *C. parasitica* be about 20%; the fraction of mutations in the diseased sample is assumed to be 12.5% in polymorphic positions and 0.2% in the whole genome; the disease was developed in one season.

Under these assumptions, the relative increase of mutation rate in diseased sponge is estimated to be 10000. Such high acceleration is obviously incompatible with life but, as a comparable precedent, the mutations in cells of some tumours are accelerated 200 times [Bielas *et al.*, 2006] and even up to 10000 times [Berger *et al.*, 2012].

Discussion

The suggested proposal, that the acceleration of mutations in diseased sponge is a cause of the difference observed between distributions at polymorphic sites, is a too simplified model for the phenomenon observed. But, if a huge acceleration of mutations did anyway take place in a determined response to the disease, what could explain the presence of this determinism, if such an acceleration will inevitably lead to death?

A sponge is an organism which should be considered, to a greater extent than other groups, in the context of its symbionts, as a part of a hologenome. A response to attacks of pathogenic microbes in that meta-organism is, mainly, a load to the symbionts but not to the sponge itself. The strategy of most pathogens is to switch to a phase of aggression, after they adapt in a hologenome as a symbiont. And the decrease of biodiversity in healthy sponges, in comparison with times before the crisis [Belikov *et al.*, 2019], is evidence that the system of discrimination between friends and foes is malfunctioning so that the sponge rejects some of its allies which were constituents of its healthy conservative hologenome.

The question is, how can a sponge survive the crisis, even in a hypothetical case? The only direction is to strengthen the connections within the healthy part of the hologenome, excluding in this way the need for symbiosis with external microbes which may turn out to be opportunistic pathogens. But this

require a synchronous adaptation of all constituents of the healthy hologenome, with rates as high in the short period of crisis as when the sponges in Baikal were developed as species.

Unusually high rates of mutations, which were, by assumption, observed in the chloroplast of photosynthetic algae, would in most cases lead to the death of cells with a modified genome, and this will happen a bit earlier than the death of cells with an unmodified genome. And the cause of inhibition of these cells is mostly mutations in genes which are responsible for metabolic relations with other species. But, as was mentioned above, the stage of accelerated mutations cannot be avoided in a way to allow the survival of the whole hologenome and for the algae as its constituent. However, survival is possible if synchronous mutations result not in suffering but in strengthening of the metabolic relations between constituents of the hologenome.

The probability of the proposed scenario is extremely low. But, at least, the hypothetical possibility of this scenario provides a trade-off and balance to be considered in the developmental strategy, instead of inevitable death. The expectations of this scenario can explain the presence of determinism in the initiation of increased growth of the mutation rate. The chance of the species surviving in this scenario is extremely low. But the assumption that similar episodes happened anyway in previous stages of algal evolution would provide an additional degree of freedom, sufficient to explain many of the ‘enigmas’ in the history of algal development. Although the exact periods of these crises are unlikely to be reconstructed, the survival of ancestors of modern algae in these crises could

explain a need to keep in their genome a way to trigger again the accelerated adaptation.

Data and software availability

The nucleotide sequence of the chloroplast genome fragment is deposited to GenBank (ID: MH591948).

The reference project IDs for the nucleotide archives used in the study are PRJEB281624 (metagenomic and metatranscriptomic sequencing) and PRJNA369024 (16S rRNA gene sequencing).

The sequencing reads and source codes of scripts sufficient to reproduce the presented results are available at GitHub: https://github.com/sferanchuk/bsponge_chloroplast.

Archived source code at the time of publication is available at: <https://doi.org/10.5281/zenodo.1326765>. License: CC BY 4.0 [Feranchuk, 2018].

Custom scripts on Python (v 2.7) were used to run the pipeline and present the results. The Python libraries pysam (0.14.1), biopython (1.66) and matplotlib (2.2.2) are required to run the scripts.

Acknowledgements

We thank Dr Colin Brown for valuable help with the work presented in the manuscript; and we thank Dr Dmitry Kuzmin and Vadim Sharov from the Siberian Federal University for assistance with the data processing.

References

- Belikov S, Belkova N, Butina T, *et al.*: Diversity and shifts of the bacterial community associated with Baikal sponge mass mortalities. *PLoS One*. 2019; 14(3): e0213926.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Berger MF, Hodis E, Heffernan TP, *et al.*: Melanoma genome sequencing reveals frequent *PREX2* mutations. *Nature*. 2012; 485(7399): 502–506.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bielas JH, Loeb KR, Rubin BP, *et al.*: Human cancers express a mutator phenotype. *Proc Natl Acad Sci U S A*. 2006; 103(48): 18238–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15): 2114–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bormotov AE: What has happened to Baikal sponges? *SCIENCE First Hand*. 2012; 2: 32–35.
[Reference Source](#)
- Cai C, Wang L, Zhou L, *et al.*: Complete chloroplast genome of green tide algae *Ulva flexuosa* (Ulvophyceae, Chlorophyta) with comparative analysis. *PLoS One*. 2017; 12(9): e0184196.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chernogor L, Denikina N, Kondratov I, *et al.*: Isolation and identification of the microalgal symbiont from primmorphs of the endemic freshwater sponge *Lubomirskia baicalensis*. (Lubomirskiidae, Porifera). *Eur J Phycol*. 2013; 48(4): 497–508.
[Publisher Full Text](#)
- Feranchuk SI, Potapova UV, Chernogor LI, *et al.*: Microevolution processes in Baikal are detected in symbiotic microbiomes of Baikal sponges by the methods of fractal theory. *Limnology and Freshwater Biology*. 2018; 2: 122–134.
[Publisher Full Text](#)
- Feranchuk S: Supplement scripts and data files for the manuscript entitled “The signs of adaptive mutations identified in the chloroplast genome of the algae endosymbiont of Baikal sponge.” (Version august 2018). *Zenodo*. 2018.
<http://www.doi.org/10.5281/zenodo.1326765>
- Fučíková K, Leliaert F, Cooper ED, *et al.*: New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Front Ecol Evol*. 2014; 2: 63.
[Publisher Full Text](#)
- Grabherr MG, Haas BJ, Yassour M, *et al.*: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011; 29(7): 644–52.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003; 52(5): 696–704.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Katoh K, Standley DM: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4): 772–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Khanaev IV, Kravtsova LS, Maikova OO, *et al.*: Current state of the sponge fauna (Porifera: Lubomirskiidae) of Lake Baikal: Sponge disease and the problem of conservation of diversity. *J Great Lakes Res*. 2018; 44(1): 77–85.
[Publisher Full Text](#)
- Kravtsova L, Izhboldina LA, Khanaev IV, *et al.*: Nearshore benthic blooms of filamentous green algae in Lake Baikal. *J Great Lakes Res*. 2014; 40(2): 441–448.
[Publisher Full Text](#)
- Kulakova NV, Sakirko MV, Adelsin RV, *et al.*: Brown Rot Syndrome and Changes in the Bacterial Community of the Baikal Sponge *Lubomirskia baicalensis*. *Microb Ecol*. 2018; 75(4): 1024–1034.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kurtz S, Phillippy A, Delcher AL, *et al.*: Versatile and open software for

comparing large genomes. *Genome Biol.* 2004; 5(2): R12.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Lefort V, Desper R, Gascuel O: **FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program.** *Mol Biol Evol.* 2015; 32(10): 2798–800.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Lemieux C, Otis C, Turmel M: **Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae.** *BMC Evol Biol.* 2014; 14: 211.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Lemieux C, Vincent AT, Labarre A, *et al.*: **Chloroplast phylogenomic analysis of chlorophyte green algae identifies a novel lineage sister to the Sphaeropleales (Chlorophyceae).** *BMC Evol Biol.* 2015; 15: 264.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res.* 1997; 25(5): 955–64.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Luo R, Liu B, Xie Y, *et al.*: **SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler.** *GigaScience.* 2012; 1(1): 18.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Pile AJ, Patterson MR, Savarese M, *et al.*: **Trophic effects of sponge feeding within Lake Baikal's littoral zone. 2. Sponge abundance, diet, feeding**

efficiency, and carbon flux. *Limnol Oceanogr.* 1997; 42(1): 178–184.

[Publisher Full Text](#)

Rosenberg SM: **Evolving responsively: adaptive mutation.** *Nat Rev Genet.* 2001; 2(7): 504–515.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sun L, Fang L, Zhang Z, *et al.*: **Chloroplast Phylogenomic Inference of Green Algae Relationships.** *Sci Rep.* 2016; 6: 20528.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Timoshkin OA, Samsonov DP, Yamamuro M, *et al.*: **Rapid ecological change in the coastal zone of Lake Baikal (East Siberia): Is the site of the world's greatest freshwater biodiversity in danger?** *J Great Lakes Res.* 2016; 42(3): 487–497.

[Publisher Full Text](#)

Truong DT, Tett A, Pasolli E, *et al.*: **Microbial strain-level population structure and genetic diversity from metagenomes.** *Genome Res.* 2017; 27(4): 626–638.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Worden AZ, Janouskovec J, McRose D, *et al.*: **Global distribution of a wild alga revealed by targeted metagenomics.** *Curr Biol.* 2012; 22(17): R675–7.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wright BE: **Stress-directed adaptive mutations and evolution.** *Mol Microbiol.* 2004; 52(3): 643–50.

[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓ ✓ ✓

Version 2

Reviewer Report 23 November 2020

<https://doi.org/10.5256/f1000research.25657.r74617>

© 2020 Nave O. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ophir Nave 

Department of Mathematics and Computer Science, Jerusalem College of Technology, Jerusalem, Israel

The paper is very interesting and has an important application.

- The abstract summarized the main results of the paper.
- The authors revised the introduction according to the comments.
- **The introduction is not complete. The authors must add a relevant list of references to this section.**
- 5 paragraphs of the introduction are without even one reference.
- Also the section "Materials and methods" is without an extensive list of references. This section is very important, and the authors need to present the relevant research list.
- The results section needs to be revised. Since as it is present in the current version, it is not absolutely clear what the authors achieve from the study.
- The analysis of the Figures must be ended in detail.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mathematical model and application to engineering and medicine

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 16 November 2020

<https://doi.org/10.5256/f1000research.25657.r74181>

© 2020 Al-Ashhab A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ashraf Al-Ashhab 

Microbial Metagenomics Division, Dead Sea and Arava Science Center, Masada, Israel

The authors have based their study on chloroplast polymorphic sites in algae endosymbiont of Baikal sponge of diseased, healthy and dead sponge tissue.

- The main concern is that they based all their analysis and conclusions from one sample following each category. The authors also used 150 PE sequencing, while for better assembly and more accurate evaluation of polymorphic sites they could have used a longer amplicon sequencing.
- The authors have used PE reads to obtain while in the analysis they were treated and independent unpaired read, why did they do this? At which stage the sequencing errors were checked and quality control used is not clear.
- The authors have used a genome of *Choricystis parasitica* (NC_025539) as a template assembly, and not enough information of why they have chosen this and not others and more than one.
- Methods on RNA and DNA extraction, library preparation, and size selection are not listed in the manuscript.
- The authors also should indicate some information about contigs length distribution following assembly.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: microbiome

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 16 Nov 2020

Sergey Feranchuk, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russian Federation

Many thanks for the approving.

I agree and confirm that the study has been done with methodological inconsistencies which are mentioned in the concerns. A consistent research should be based on several replicates of sponges in each phase of disease; sequencing should be more fine-tuned, more deep and should make use of long amplicon extraction. The answers are in the reality of scientific business rather than in the science itself so I would like to omit the more detailed explanations.

Answering the more specific concerns:

- Reads were unpaired and mixed after filtering on a template of a similar chloroplast. After this stage, the chances of mis-assembly which are intended to be reduced by a filtering of mate-pairs, were of lesser importance than a need to obtain as long contig as possible, from a filtered pool of reads.

- A close similarity of the target algae species to *Choricystis* species was observed earlier in vitro [Chernogor 2013] and by 16S rRNA annotation to Silva database.

- In the primary Inchworm assembly, the targeted 55K contig was followed just by "noisy" contigs with lengths not more than 5K b.p., so the distribution of contig lengths was not explicitly indicated.

Competing Interests: No competing interests were disclosed.

Reviewer Report 28 May 2020

<https://doi.org/10.5256/f1000research.25657.r63440>

© 2020 Kondratov R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Roman Kondratov 

Center for Gene Regulation in Health and Disease (GRHD), Cleveland State University, Cleveland, OH, USA

The authors addressed my concerns. I am fine with the current version.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 23 January 2019

<https://doi.org/10.5256/f1000research.17291.r42769>

© 2019 Kondratov R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Roman Kondratov 

Center for Gene Regulation in Health and Disease (GRHD), Cleveland State University, Cleveland, OH, USA

The massive presence of sponges is known as playing a leading role in the process of biofiltration of Baikal's water. Recently the cases of sponge disease have been expanded rapidly. The significance of the current study is coming from the attempt to investigate molecular aspects of sponge disease using a systems biology approach. In the present submission the authors provide data on the genome and transcriptome of chloroplasts from symbiotic algae. Metagenomic and metatranscriptomic sequencing data was used for a template-based assembly of the chloroplast genome of algae, and a mapping of sequencing reads to the obtained genome sequence was a subject of interpretation, based on a detection of polymorphic sites. The advantage of the study is the exploration of natural samples, improvement of methods of bioinformatical analysis and provocative ideas. I think that the manuscript deserves to be indexed, but there are concerns which the authors need to address:

1. What was the number of analyzed biological replicas or independent samples? Do the authors expect the same spectra if another sample(s) of sponge tissues would be processed with the same pipeline? What would be the distribution of variations in polymorphic sites in another series of experiments?
2. Were the "healthy", "diseased" and "dead" samples collected from the same spot? How can the authors be sure that they are dealing with the same species? Could it be that what the authors proposed as "adaptive mutations" is in fact a difference between different subspecies?

I will recommend to tone down the conclusions and to discuss possible alternative interpretations. The statement on a presence of adaptive mutations in the chloroplast genome is too strong. It requires confirmation with more independently obtained samples. There is also no evidence that even if these mutations will be confirmed, that they are "true adaptive" mutations and provide the organism with any physiological advantage. Therefore the title is misleading in my opinion.

I suggest also to provide more details of the overall design of the study and on novelty in bioinformatics approaches.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Gene expression, biological rhythms, metabolism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 04 Apr 2020

Sergey Feranchuk, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russian Federation

Many thanks to Prof. Kondratov that he had found a possibility to read carefully the manuscript, so that he had found deep and precise issues in which where the proposed results can be doubted. I did consider his remarks in a most full extent.

I respect the experience of Prof. Kondratov who did suggest a presence of several subspecies of algae in a sponge host. In the revised version I explicitly told about possible presence of subspecies in the samples and add an extra chart which shows the distribution of algae subspecies around Baikal. And I did explicitly specify some of most important details of the study, and did remove some extra material from the second edition, trying to keep the format and to put more accents on the main of the declared results.

I did change the declared statement of the manuscript, to propose a proof of a possibility rather than a proof of an observation. In the revised version I tried to provide a more explicit "proof of possibility", that the adaptation can be just one of the many contributions to the observed distribution of polymorphic sites. And I add an additional argumentation why just a possibility of the adaptation is worth to be presented as a scientific result. In short, the situation around Baikal is complicated, and just an idea how this complication can

be resolved is worth to be said.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 January 2019

<https://doi.org/10.5256/f1000research.17291.r42738>

© 2019 Sadovsky M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Michael G. Sadovsky 

Siberian Federal University, Krasnoyarsk, Russian Federation

Well, on one hand I should say “Yes” to this submission. At least, it completely meets all the up-to-date customs and observances in genomics and molecular biology. On the other hand, the authors rely on a number of (quite complex and apparent) software packages, workbenches and pipelines, and this is the matter of my general scepticism. Actually, we all become the hostages of those software tools; one has to trust software designers and believe the output of the programs is correct, free of mistakes, stably working, etc. Somebody may say “You, physicists and mathematicians, do use Wolfram’s *Mathematica* and there is no collapse in math, nor in physics”. Reciprocally, I would like to draw attention to an Elsevier journal devoted to Microsoft Excel errors solely (McCullough and Heiser, 2007¹).

I am far from the idea to say that the study is wrongly arranged or badly accomplished; I just want to stress the point that there should be some special efforts done to ensure the results are at least stable. For example, what happens with assembled contigs, if we randomly remove a small part of reads? If a series of runs of an assembler yields (almost) the same contigs set, then the results could be used for further analysis. The problem arises, if one gets a number of sets of contigs with a sounding difference between them. The paper has no answer on that point; a comparative study (like that one presented by the authors) should have some proofs of the absence of artefacts affecting the comparison of fine differences between biological objects involved in the study. Meanwhile, I pretty well understand that such output testing falls beyond the customs and habits of NGS sequenced data treatment and I am in the smallest minority. So, from that point of view the paper completely meets all the custom data treatment procedures and in such capacity should be recommended for indexing.

Another important issue of the paper is that it presents an attempt to tie together ecological (environmental) processes, and some genetic background that may stand behind. Here the word ‘crisis’ used by the authors makes a point: regularly, *ecological crisis* is stipulated as a rather fast running process in a community resulting in serious (and inevitable) loss of the greater part of species from the community. Maybe, this word is too strong here: what if the observed infection intrusion is just a regular (while long ranged) periodic event in the community? Nonetheless, the scientific merit of the paper is obvious, the results and conclusions are sounding and up-to-date,

and paper should be indexed.

The paper needs major revisions in its English. The paper is written in a version I dare say is Runglish. There are too many lines in the manuscript that look like a literal translation from Russian of (quite boring) scientific Russian-style. I myself can decipher what the authors mean, since my mother language is also Russian. I am absolutely sure that the current version of the paper will fall out of comprehension for the greatest majority of readers who have no active Russian. To begin with, the title must be changed. No signs, at all. The correct version should be like *“Evidences of the adaptive mutations in chloroplast genomes of some algae endosymbionts of Baikal sponge”*.

Same in the Abstract (Background paragraph): instead of *“The study of ecosystems of the great lakes is important as observations can be extended to ecosystems of larger scale. The ecological crisis of Lake Baikal needs investigations to discover the molecular mechanisms involved in the crisis. The disease of Baikal sponges is one of the processes resulting in the degradation of the littoral zone of the lake”* there should be something like *“Monitoring and investigation of the great lakes ecosystem provides a sounding background to forecast the greater scale ecosystem dynamics. Changes in the Baikal lake biota observed nowadays demand deeper investigations of the molecular mechanisms standing behind these former. The endemic Baikal sponge disease may cause a degradation of littoral ecosystem of the lake”*. I am far from the idea that my version is the best, but the original one must be rewritten.

Unfortunately, there are many more similar problem lines in the manuscript, so very strong revisions in the English are absolutely necessary.

References

1. McCullough B, Heiser D: On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics & Data Analysis*. 2008; **52** (10): 4570-4578 [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 04 Apr 2020

Sergey Feranchuk, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russian Federation

I'm grateful to Prof. Sadovsky for his decision to review this manuscript. I did carefully consider his remarks and prepared a revised version with a respect to his position.

First of all, he was right that crises like the crisis on Baikal could anyway happen in the past. The need to survive in the times of severe crises can be encoded in genome. This idea was introduced to the revised version, as an additional support to the hypothesis about adaptive mutations.

To answer the remark about "closeness" and insufficient robustness of the software, I did several other runs of the assembly. I agree with Prof. Sadovsky about "closeness" and over-complication of some software, and this is why I did choose Inchworm assembler in the initial version of the pipeline, as the most lightweight and straightforward of the available assemblers. In additional runs I tried another assemblers. The correctness of the assembled chloroplast sequence was anyway confirmed, and the fact of verification was pointed out in the second revision.

To answer the remark about "Russian" style of language. This question is in part beyond the scope of the discussion. It is unlikely that me who is Russian will speak the same English as a man from England. But Prof. Sadosvky was right that the meaning of the text in the first edition was unclear in many parts. And in the revised version I put much more attention to a choice of words and grammatic constructions, to use only those words, for which I am certain in their meaning. The text can anyway look unusual to one who know in perfect the context of all words in English, but at least I do my best to make the meaning of the text the most clear.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research