RESEARCH ARTICLE

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

Sergey Feranchuk\textsuperscript{1,2}, Natalia Belkova\textsuperscript{1,3}, Lubov Chernogor\textsuperscript{1}, Ulyana Potapova\textsuperscript{1}, Sergei Belikov\textsuperscript{1}

\textsuperscript{1}Limological institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, 664033, Russian Federation
\textsuperscript{2}Department of Informatics, National Research Technical University, Irkutsk, 664074, Russian Federation
\textsuperscript{3}Scientific Centre for Family Health and Human Reproduction Problems, Irkutsk, 664033, Russian Federation

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

Abstract

\textbf{Background}: 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.

\textbf{Methods}: The chloroplast genome fragment for the algae endosymbiont of Baikal sponge was assembled from metagenomic sequencing data. The distributions of polymorphic sites were obtained for the genome fragment, separately for samples from healthy sponge, diseased sponge and dead sponge tissues.

\textbf{Results}: The comparative analysis of chloroplast genome sequences suggests that the symbiotic algae from Baikal sponge is close to \textit{Choricystis} genus of unicellular algae. Also, the distributions of polymorphic sites allowed detection of the signs of extensive mutations in the chloroplasts isolated from the diseased sponge tissues.

\textbf{Conclusions}: The study demonstrate the particular case of evolution at the molecular level due to the conditions of a severe crisis of a whole ecosystem in Lake Baikal. The detection of adaptive mutations in the chloroplast genome is an important feature which could represent the behavior of an ecosystem in the event of a severe crisis.

\textbf{Keywords}
Chlorophyta, Lake Baikal, Chloroplas Genome, Genetic Polymorphism, Mutation Rate
Introduction

Lake Baikal, located in Southeastern Siberia, is the largest by volume and the oldest great lake on the planet, and several signs of ecological crisis in Lake Baikal have been observed since 2010–2011 [Bormotov, 2012; Khanaev et al., 2018; Kravtsova et al., 2014; Timoshkin et al., 2016]. One of these signs is severe disease and death of sponges which is now observed in almost all parts of the lake. The symptoms of the disease begin with the appearance of pink and brown spots on the surface of the sponge and terminate with complete destruction of the sponge tissues. The cause of the disease is still unclear.

Endemic freshwater Baikal sponges (Demospangiaceae, Labonim-Biskiiidae) dominate their biomass among the benthic organisms of the littoral at depths from 3 to 25 m covering 47% of the available surfaces [Pile et al., 1997]. In healthy condition sponges have a green color, mainly explained by the presence of a photosynthetic symbiont, an intracellular coccoid green algae. This algae belongs to the Chlororophyta division, and is close in taxonomy to the Chorecytis us genus [Chernogor et al., 2013]. It is natural to assume, that the photosynthetic symbiont is the source of feeding for sponge cells. And, the change of color of sponge tissue could indicate that the chloroplasts of the symbiont are damaged in the early stages of the disease. So, precise study of this algae symbiont could be of critical importance in investigating the cause and consequences of sponge disease.

The sequencing and comparative analysis of chloroplast DNA is a conventional method for detailed study of planctonic algae [Lemieux et al., 2014; Lemieux et al., 2015]. Normally, chloroplast DNA sequences are determined using cultivated algae and de novo assembly of genomic DNA reads [Twyford & Ness, 2017]. But for uncultured species, the chloroplast genome can be obtained using metagenome sequencing [Worden et al., 2012]. For symbiotic algae from Baikal sponges, this strategy is probably more efficient, and the comparative analysis of samples of healthy and diseased sponges can provide a deeper look into the features of chloroplast genome affected by sponge disease. The presence and properties of polymorphic sites on the chloroplast genome could be an effective way to investigate the variations of genome sequence depending on the disease state of the sponge. The study of the distribution of bacterial strains depending on geographic location [Truong et al., 2017] can be mentioned as a precedent, where gene-batteries typical for gut microbiome were compared using the distribution of polymorphic sites in genome sequences, using metagenome sequencing.

Methods

Sampling and sequencing

Three samples of freshwater sponge Lubominuska baicalensis were collected from Lake Baikal in the Bol’shiye Koty area (51° 90’ 69 N °, 105° 07’ 05 E’) at a depth of 10 m by scuba divers in June 2016. One sample was obtained from the sponge that was healthy in appearance (exhibiting a green colour), one sample was taken from diseased sponge and one from dead rotten sponge tissues. The collected samples were immediately placed in containers with Baikal water and ice and transported to the lab, maintaining a constant water temperature. For all three samples Illumina pair-end reads were obtained by DNA metagenome sequencing in Novogene Inc. (Illumina PE 150).

Identification of polymorphisms

In order to separate it from traces of sequencing errors, the selection of the polymorphic sites in the genome was implemented following the approach described in [Truong et al., 2017]. Each of the RNA and DNA samples represented as pair-end reads was separately aligned to the assembled fragment of the
Describing the algorithm, for each position \( s \) on the alignment of the reads against the \( N_s \) is defined as the total number of reads covering it, 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 \( N_s - T_s \) of reads coming from the non-dominant allele is \( < \alpha = 0.05 \). This is estimated using the probability mass function of a binomial distribution with \( N_s \) trials and the successful rate \( 1 - E \). The error rate was set to 0.01 for Illumina sequencing. The bases with quality below 30 were removed and the reads with an average identity to the reference below 99% were ignored before applying the statistical test. Failing to reject the null hypothesis reflects the absence of alternative alleles or inability of distinguishing between low-coverage potential alternative alleles and sequencing noise.

Thus, the number of polymorphic sites could 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 dominant allele \( (T_s < N_s - T_s) \). This property could detect mutations in the sample genotype and phenotype, for each gene.

**Phylogenetic analysis**

The chloroplast genome sequences of *Picocystis salinarum* (NC_024828), *Myrmecia israelensis* (KM462861), *Botryococcus braunii* (KM462884), *Coccomyxa subellipsoida* (NC_015084), Hydrodictyon reticulatum (NC_034655), Mychonastes jurisii (NC_028579) and Chlorella vulgaris (NC_001865) were used for a reconstruction of the phylogenetic trees for the 16S ribosomal RNA (rrs gene) and the ATP synthase subunit beta (atpB gene). The nucleotide sequences of the selected genes were aligned using Mafft 7.27 software [Katoh & Standley, 2013]. The trees were constructed using the FastMe 2.1.5.1 software [Lefort et al., 2015], with the distance-based neighbor-joining method to select tree topology and Jukes-Cantor measure to calculate the distances between genes.

Analysis was performed using custom scripts in Python 2.7 (see Data and software availability section).

**Results**

The chloroplast genome of the *Choricystis parasitica* algae is a circular DNA 94206 base pairs in length. The comparison of open reading frames of the candidate genome fragment from the metagenomic samples with annotated genes of *C. parasitica* support the statement that this genome fragment of length 55638 is a large part of the chloroplast genome of algae close to the *C. parasitica* species. Figure 1 illustrates the order of genes in these two related chloroplasts. The comparison of gene sequences shows them to be up to 98% identical in these two species.

The phylogenetic trees for the two selected genes, 16S ribosomal RNA and ATP synthase beta (Figure 2) in general confirm the conventional relations between Chlorophyta algae [Lemieux et al., 2014; Lemieux et al., 2015]. Figure 2 suggests that the symbiotic algae of *L. baikalensis* sponge is close in taxonomy to the *Choricystis* genus.

The bars on Figure 2 which show the proportion of polymorphic positions in the genes of symbiotic algae in metagenomic
samples is comparable in scale with the distances between genera. This observation needs discussion, because a timescale which separates the origins of the close genera in Figure 2 implies a much larger timescale than that which could characterize the separation of the chloroplast strains detected in the metagenome. Partially this can be explained by the RNA editing and similar modifications which lead to the accumulation of polymorphic positions.

A different view of the unexpectedly high proportion of polymorphic sites in the metagenomic samples is illustrated in Figure 3. Here, the proportion of polymorphic sites, and the proportion of polymorphic sites with a low abundance of dominant allele (“mutations”) is shown separately for each DNA and RNA sample. The results of Figure 3 are presented separately for each gene frame, and for a whole set of genes.

The proportion of polymorphic sites in the DNA and RNA metagenomes for the sample of healthy sponge tissues reflects the natural situation, where the quantity of matrix RNA in the chloroplast organelle is in general higher than the quantity of DNA. Here the polymorphic sites in the DNA sequences may arise due to natural heterogeneity of chloroplast genomes, but the dominant strain is clearly identified. The number of polymorphic sites in the RNA sequences is slightly higher than in DNA sequences due to RNA editing and other modifications.

In contrast, in the DNA and RNA samples of the diseased tissue, the quantity of chloroplast DNA is decreased, and the quantity of RNA is decreased even more, reflecting the fact of disease and the low level of chloroplast activity. Importantly, the number of polymorphic sites is sharply higher in the remaining DNA sequences, and the alternative alleles are presented in high proportion. And, for RNA sequences, the dominant allele is present much less than alternative alleles. For the case of dead tissue, where RNA couldn’t be extracted, the discussion about the observed number of DNA molecules and the proportion of polymorphic sites is beyond the scope of this study.

The natural assumption about the diseased but alive tissue is that living cells are desperately trying to survive. Adaptation to a changed environment is the one of the best ways to survive. The accumulation of mutations is a straightforward form of adaptation, and this could be confirmed by the results of Figure 3 for the sample of diseased sponge tissue. The rapid increase of mutations in the genome can be observed in the chart for chloroplast DNA, and as it may be suggested from the chart for chloroplast RNA, that the mutations which help survival are fixed in the cells which still continue to develop.

The observed signs of extensive mutations in response to severe stress are somewhat controversial when compared to the widely accepted concept of molecular clocks and a theory of neutral evolution [Kimura, 1968; Margoliash, 1963; Zuckerkandl & Pauling, 1962] where mutations are appear randomly and are independent from the environment. But in several studies the presence of adaptive mutations in response to stress has been detected in certain species, as reviewed in [Rosenberg, 2001; Wright, 2004]. So the present result cannot be treated as completely inadequate, in conditions of severe and unusual crisis of the whole ecosystem.

**Figure 2.** Phylogenetic tree for two chloroplast genes: 16S ribosomal RNA (rrs) and ATP synthase subunit beta (atpB). Bars located at the node for the studied chloroplast genome represent the relative number of polymorphic positions, in all 5 studied samples, at the 1:1 scale.
Discussion
The signs of the large-scale ecological crisis in Lake Baikal are confirmed from many sources, and ecological crises on such a large scale are rare in the documented history of water ecosystems. Sequencing technologies have appeared only in recent years, and, to the authors best knowledge, no cases of large and sharp changes in ecosystems have been documented using the tools of molecular biology.

The importance of Lake Baikal itself as an ecosystem with an unusual diversity of endemic species, and as a glorious source of pure drinking water, is a subject high above the economic and pragmatic reasons which are usually considered in molecular biology studies. However the present results suggest that the conventional approaches of molecular biology may be insufficient to adequately describe situations of ecological crisis. In particular, the observations of rapid accumulation of mutations in the chloroplast genomes in the diseased tissues could indicate that the concept of molecular clocks is inappropriate in rapidly changing ecosystems.

What's more, using the tools of molecular biology to study the Baikal ecosystem has another importance; it is a unique chance to accumulate observations about a rapidly changing environment. Great lakes are themselves simplified cases of large-scale marine ecosystems. The presented results, as a part of all Baikal ecosystem studies, could find application, not only in the challenge of minimizing the consequences of the crisis in Baikal, but also in the possible future global challenges caused by sudden changes in ecosystems of any scale.

In particular, the reconstructed genome of the symbiotic algae may improve knowledge about a cause of sponge disease, and indirectly narrow the possible strategies to prevent the spread of destruction in the Baikal ecosystem. The presented description of the genome may be helpful in the evolutionary studies of marine and freshwater Chlorophyta algae.

Data availability
The nucleotide sequence of the chloroplast genome fragment is deposited to Genbank under the accession number: MH591948

Nucleotide sequences have also been deposited with the European Nucleotide Archive (ENA) of the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) under study number: ERP110335.

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

Archived source code at time of publication available at: https://doi.org/10.5281/zenodo.1326765 [Feranchuk, 2018].

(license: CC BY 4.0).

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

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.

Acknowledgements

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

References

Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30(15): 2114–20.

Bormotov AE: What has happened to Baikal sponges? SCIENCE First Hand. 2012; 2: 32–35.

Cheunogor L, Denikina N, Kondratov I, et al.: Isolation and identification of the microalgal symbiont from primorphs of the endemic freshwater sponge Lubomirskia bicaelasis. (Lubomirskiaidae, Porifera). Eur J Phycol. 2013; 48(4): 497–508.

Chernogor L, Denikina N, Kondratov I, et al.: Isolation and identification of the microalgal symbiont from primorphs of the endemic freshwater sponge Lubomirskia bicaelasis. (Lubomirskiaidae, Porifera). Eur J Phycol. 2013; 48(4): 497–508.

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.

Katoh K, Standley DM: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30(4): 722–80.

Kimura M: Evolutionary rate at the molecular level. Nature. 1968; 217(5129): 624–626.

Kurtz S, Phillippy A, Delcher AL, et al.: Versatile and open software for comparing large genomes. Genome Biol. 2004; 5(2): R12.

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.

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

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.

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

Pile AJ, Patterson MR, Savasse 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.

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

Ticktin OA, Samsonov DP, Yamamura 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.

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

Twyford AD, Ness RW: Strategies for complete plastid genome sequencing. Mol Ecol Resour. 2017; 17(5): 658–668.

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

Zuckerand E, Pauling LB: Molecular disease, evolution, and genomic heterogeneity. In Kasha M. and Pullman B (eds). Horizons in Biochemistry. New York: Academic Press, 1962; 189–225.
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, 2007).

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 gramatic 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