Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Valentina Burskaia

PhD Program: Life Sciences

Title of Thesis: Positive selection in parallel evolution

Supervisor: Associate Professor Georgii Bazykin

Name of the Reviewer: Konstantin Popadin

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 22-11-2020**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

I was enjoying a lot by reading the PhD work by Valentina Burskaia: very deep and detail introduction, clear definitions of all terms, main problems and approaches; interesting results and detailed discussions. The dissertation is very well written with a straightforward logical flow.

The topic of the dissertation is relevant to its content, methods are adequate and fresh. Results of the work significantly improve the understanding of the scale and targets of parallel evolution. High quality of scientific research is supported by two first-author publications: one is in GBE (doi: 10.1093/gbe/evaa138) and one is uploaded to BioRxiv (**doi:** https://doi.org/10.1101/2020.09.15.298117).

Without any doubts, I recommend that the candidate should defend the thesis by means of a formal thesis defence.

I have two comments related to chapters 3 and 4 and would be glad to know the opinion of Valentina.

Chapter 3.
Intensive whole-genome parallel evolution in amphipods is very interesting. I would like to discuss the possibility of the parallel relaxation of selection in different species, which can at least partially contribute to the expected results. Taking into account that the main mode of selection is purifying and the diversification of species may be associated with a parallel decrease in Ne (as compared to the common ancestor) it is possible to assume, that pronounced decrease in Ne in some but not in other species of the quartets may drive the parallel molecular changes - accumulation of suboptimal slightly-deleterious substitutions. The main problem here is that it is difficult to approximate Ne and it is possible to use either some life-history traits, for example, the dimensionality of the environment (2D or 3D) or resource-poor/rich environments (see example here doi 10.1101/gr.212589.116) or genetic correlates of Ne - for example genome size (doi 10.1101/gr.212589.116: longer genomes on average mark less effective selection due to low Ne). As I see from chapter 3 authors used in analyzes many interesting traits, but none of them approximate Ne.

I would like to know the opinion of Valentina on the suggested scenario - which results can vote for or against the proposed scenario?

Chapter 4.

Mitochondrial genome of all vertebrates is extremely highly constrained and the NULL hypothesis is that the purifying selection is the only mode of selection, acting on mtDNA of vertebrates. The vast majority of studies, claiming positive selection in mtDNA of vertebrates is questionable. Additionally, birds are among the most constrained vertebrates because of high temperature and correspondingly high level of aerobic metabolism. From these points of view - the attempt to search for events of parallel adaptive evolution in mtDNA of birds is a low-success task from the very beginning. And, despite this, Valentina was able to find several candidate sites evolving in parallel. From my point of view, this is the amazing and unexpected result (I would expect zero sites evolving in bird mtDNA under parallel positive selection).

Another peculiarity of mtDNA and especially mtDNA of birds is very strong nonuniformity in mutagenesis, which, according to my opinion, may affect results/interpretation of the results. Two of the most common mtDNA substitutions, totally explaining ~ 80% of the whole mutational spectrum in mtDNA, are C>T and A>G substitutions on heavy chain (deamination of cytosine and deamination of adenine on the heavy chain, which spends a long time in single-stranded mode during mtDNA replication). Moreover, the fraction of A>G has been recently associated with oxidative damage in organisms with a high level of aerobic metabolism (doi: 10.1101/2020.07.25.221184). A>G substitution on heavy chain corresponds to T>C substitution on the light chain which is a reference strand in all public mtDNA databases. Thus, from mutagenesis point of view, we expect a strong mutational bias towards T>C (light chain notation) in species with high oxidative damage and no bias ("T>C" ~ "C>T" or even "T>C" < "C>T") in species with low oxidative damage. Assuming that environmental hypoxia (in high-altitude birds) may be associated with cellular level hypoxia we can expect that these species will have decreased frequency of mutations driven by oxidative damage: i.e. decreased fraction of T>C on the light chain, which automatically means an increased fraction of C>T. Altogether, we expect higher C>T fraction in high-altitude birds versus low-altitude birds.

One of the strongest results is based on four Histidin to Tyrosine substitutions in position 57 of the ND5 gene, which are going in parallel with switches from background to a high-altitude environment. The one-step trajectory from His to Tyr is possible due to C>T substitutions in the first codon position. This is exactly what we expect if the pressure of oxidative damage is decreasing in these species and the probability of C>T substitution is increasing.
So, one additional interpretation of the results from chapter 4 is that changes in the environment (high-altitude => hypoxia => decreased fraction of A>G on heavy chain == decreased fraction of T>C on light chain == increased fraction of C>T "which is opposite") may change rules of mutagenesis, allowing new categories of substitutions (C>T) becomes relatively more common. So, this is yet a parallel evolution, however, most likely due to neutral changes associated with mutagenesis.

It is not a very common situation when mutational spectrum drives amino-acid changes, however, it has been shown recently on SARS-CoV-2 evolution (doi: 10.1101/2020.05.01.072330) and also it has been recently shown for human mtDNA (unpublished data from my Kaliningrad laboratory).

I would greatly appreciate having comments of Valentina to the above-described scenario. More precisely the question might be rephrased like this: if TreeWAS takes into account different types of nucleotide substitutions during simulation of the “null” distribution?

Valentina, thank you for the interesting work!

Provisional Recommendation

☒ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate’s thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*