

## 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:** Professor Matthew Hahn

I confirm the absence of any conflict of interest	<b>Signature:</b>  <b>Date:</b> 28-10-2020
---	---

*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

This is a very nice dissertation consisting of three substantial chapters. After a brief introduction, the first chapter is a review of the concepts and ideas studied in the dissertation, the second chapter is a study of parallel molecular evolution within the whole genomes of three clades, and the third chapter focuses on parallel molecular evolution within the mitochondrial genome of birds. The ideas being addressed are at the forefront of questions in evolutionary genomics right now, and the methods used to answer these questions are appropriate and up-to-date. There is a lot of excitement about convergence and parallelism, especially with the growing ability to identify the processes leading to them. The number of available whole genomes makes all of this possible, and the dissertation leverages such data to answer important questions.

All chapters are well-written and up to the standard expected at my own university and other universities around the world. The scientific results presented are of high quality, with chapter 3 recently having been published in a top journal in our field. I foresee little trouble publishing chapter 4, though if chapter 2 is to be published as a stand-alone review paper I think it would need more work. Right now, chapter 2 ends without much of an overall conclusion or roadmap for methods that do work, two things that would be expected in a good review paper.

Given all of this, I do have some scientific issues that I would like to see the candidate address in their formal defense, as well as some minor typographical and other errors that should be changed in the thesis.

1) One issue that is raised in the thesis is the idea that hemiplasy can be driven by adaptive natural selection. The example given in the text is the *Eda* gene in sticklebacks, where repeated selection on an ancestral variant has led to fixation independently in multiple lakes. I very much agree with this formulation, as it separates the mutational process (hemiplasy/homoplasy) from the selection process (neutral/adaptive). The line between these becomes a little less clear in cases of balancing selection, where selection is maintaining ancestral variation, but this possibility is not a focus of the thesis.

However, given this distinction, I wondered why in the Discussion of chapter 3 hemiplasy is dismissed so readily. Given the stickleback example above, it seems quite straightforward for selection to favor ancestral nonsynonymous variants over synonymous variants. Indeed, this is one of the findings of Pease et al. (2016, PLoS Biology). The conclusion here therefore requires much more justification.

2) It was unclear to me what a "robust" phylogeny of amphipods means, though it is likely referring to the bootstrap support on each branch. Much more important for inferences about hemiplasy, however, are concordance factor values on each branch. It is reported that the cichlid phylogeny has a lot of incongruence, but not information is given on the amphipod clade. A comparison of these values between phylogenies may help to reveal why patterns of parallelism are so different between them. As it stands, there is little biological or technical explanation as to the difference between cichlids and amphipods.

3) On a related note, I did not understand where the colonization history of Lake Baikal came from. Certain sub-clades are listed as having colonized first or second, but it was not clear how this was inferred. I do not think this information can be gleaned from the phylogeny shown in Figure 3.1. Some further explanation is therefore needed.

4) In section 3.3.2, results on parallelism for nonsynonymous and synonymous substitutions are given, with a higher rate for nonsynonymous than synonymous. But the text concludes "this can arise both from parallel adaptation and differences in strength of negative selection among sites," and goes on to introduce the P-test. I do not understand how a difference in slopes could be due to varying negative selection among sites, nor how the P-test would control for such an effect if present. Perhaps this can be explained more clearly.

5) More minor issues:

++The "GWAS" approach in chapter 4 seems a lot like the "phyloGWAS" approach laid out in Pease et al. (2016, PLoS Biology).

++In chapter 2 the text says that McGee et al. (2020) found that hybridization led to a shared predatory phenotype. But I do not believe that this paper showed any evidence for hybridization.

++Avice and Robinson (2008, Systematic Biology) created the term "hemiplasy," not Hahn and Nakhleh (2016)

++I do not understand the effect ascribed to bottlenecks on p. 19 of the thesis. Also, there is a typo in "Lefebure et al" at the same location.

++"Hahn" is misspelled on pages 20, 21, and 66 ("Hanh").

++"withinspecies" needs a space between the two words (p. 26).

++"Figure" is misspelled twice on p. 42 ("Figire")

++"McDonald" is misspelled on p. 44 ("MacDonald")

#### **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*