Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Aleksandra Bezmenova

PhD Program: Life Sciences

Title of Thesis: Evolutionary processes in hypervariable fungus Schizophyllum commune

Supervisor: Professor Georgii Bazykin

Co-supervisor: Professor Alexey Kondrashov, University of Michigan, USA

Name of the Reviewer:

I confirm the absence of any conflict of interest Confirmed (Alternatively, Reviewer can formulate a possible conflict)	**Date: 15-11-2021**

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 would like to congratulate Aleksandra Bezmanova on her contributions to science reported in this thesis. Here she examines mutation and recombination in the basidiomycete *Schizophyllum commune*, a fungus with extraordinarily high levels of DNA sequence variability. Her work combines mycological experiments with genomics to gain insights into rate of the fundamental properties of mutation and recombination rates in this species. In addition to a chapter providing some good introduction and background, the thesis consists of four research chapters. The first of these, which examines mutation rates in four strains under each of two growth conditions, is particularly strong. Each of the remaining chapters contains some intriguing results though I believe there are alternative interpretations that should be considered. Below I provide a few comments on each of these chapters.

Chapter 3

Estimates of mutation rates depend the number of mutations called (numerator) and the number of sites observed, i.e., callable sites (denominator). In this chapter as well as 4 and 5, the procedure for calculating the number of callable sites was less clear. Please clarify.

On p. 44 it says: "At these positions, we called variants that had the following properties: (i) at least in one sample, coverage in the 10-90% range and non-reference variant frequency >30%, or coverage in the 15-85% range and non-reference variant frequency >20% (13962 variants); (ii) not supported by any read in the reference sequence (289 variants). For these variants, we assessed their frequencies in all samples." It is unclear to me what the numbers in parentheses indicate (i.e., 13962 variants and 289 variants). It sounds like 13962 variants are identified at one stage but in the results there are only 289. Does this mean that 98% of the variants are excluded because of reads in the reference sequence? Have you considered how this issue will affect your estimate of the rate? In other words, if a TRUE mutation occurs at a random site, what is the chance you would then exclude that variant because that site would have a supporting read in the reference sequence. (By your procedure many sites may not be fully callable because there is an errant read supporting an alternative base.)

I would like to have seen an estimate of the "population size" (i.e., number of hyphal strands) in narrow and wide tubes.

How important are the number divisions when grown in liquid culture for sequencing?

Table 3.2 is very useful but I would also like to see the what proportion of callable sites are in each category so it would be easy to see if nonsynonymous mutations are underrepresented.

Chapter 4

Unfortunately, only 2 of 13 trunks examined, yielded the type of data desired (i.e., "siblings"). The mutation rate was reported as a function of the linear distance between fruiting bodies. It should be noted that this is a minimum distance as it is easy to imagine that hyphal pathways are very convoluted, much more so than when forced to grow in tubes as in Chapter 3.

The estimated mutation rate is described as similar to that observed in Chapter 3. However, the one Russian sample in Chapter 3 had a much higher mutation rate and the estimates of these Russian samples in Chapter 4. This should be acknowledged.

A major motivation of this thesis is understanding the high sequence diversity with this species. For this purpose, it is essential to know the mutation rate. In Chapter 3 you get quite a precise estimate of the mutation rate per meter (or per cell division). Chapter 4 also provides some information here too. However, you really want the mutation rate per generation. Here it seems you have great uncertainty in going from mutation rate per meter to mutation rate per generation. Much effort goes into measuring the former yet there is massive (and unknown) certainty in the cell divisions per generation. I would like to have seen more discussion of this issue. If "generation" is defined as sexual events, then you might be able to gain insight by relating mutation rate to recombination rate by comparing theta = 4 Ne*u to rho = 4 Ne*r.

Chapter 5

Table 5.1 reports the "Callable length" but the text below the table says it is "impossible to estimate the callable length of the genome", which seems to suggest that Table 5.1 has accomplished the impossible.

Can you simply use a single sample per fruiting body to estimate the mutation rates (averaging over the outcome of using different samples)? This should allow you to avoid the problem of "cluster" mutations.

This chapter reports a very striking result – a mutation rate 2-3X higher in homozygous than heterozygous regions. However, I was disappointed by the lack of discussion of this result. What are plausible explanations? Presumably, most of these mutations occur while it is growing as a dikaryon so the homologous chromosomes are isolated in different nuclei, correct?

Is there any bioinformatic (i.e., technical artifact) reason that could lead to this pattern. In particular, I wonder about this step of variant calling (p. 69):

> "ii) not a
> single read in both parental mappings supported non-reference nucleotide
> in case of 'both'
> genotype, or not a single read supported non-reference nucleotide in
> parent 1(2) in case of
> 'parent 1(2)' genotype."

Could that filtering step make it more likely to exclude variants from heterozygous regions than homozygous regions?

Chapter 6

This is a clever experiment. However, I have one serious concern. There is a very striking difference in recombination between your two crosses. Those crosses differ in heterozygosity of the focal region but

also in genotype at many parts of the genome. With these data alone, it is seems impossible to make a strong inference that the difference in recombination is due to heterozygosity or just genetic background.

If you think heterozygosity is directly affecting recombination, can you speculate on mechanism?

Summary

Overall, this is a good thesis that makes good contributions to evolutionary genetics.

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*