Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Anastasiia Stoliarova

PhD Program: Life Sciences

Title of Thesis: Genomic patterns of epistasis at macro- and microevolutionary scales

Supervisor: Professor Georgii Bazykin

Name of the Reviewer: Joshua B. Plotkin

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 17-11-2021**

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

Stolyarova's dissertation concerns the role of epistasis in molecular evolution. She offers the chapters of original research, well-positioned within the broader literature that she reviews and synthesizes at the start of the dissertation. The research is original, inciteful, and it spans an impressive range from mathematical models, to simulations, to new methods of inference from phylogenetic data.

The key idea in **Chapter 3** is to leverage data from an extremely polymorphic species (a fungus S. commune) to get a better view of the (still local) fitness landscape represented in natural variation than is possible in most other species.

The main method is look for short-range LD attributed to attraction of rare alleles (often in same gene) that are better in combination. The main result is that such LD elevation is stronger for non-synonymous variants than synonymous, suggesting a role of epistasis in protein structure driving patters of local LD.

To support this conclusion, Stolyarova analyzes individual proteins with known structure where the physical distance after folding is correlated with patterns of LD between non-synonymous variants.

More convincing yet is correlation of excess LD between divergent populations where mutations likely arose independently.

The work described in this chapter is impressive in its breadth: it includes sequencing and assembly within S. commune samples; identification of haploblocks, and computations with known PDB 3-d structures, inter alia.

Overall, the results in this chapter largely confirm previous inter-specific studies that identify a role physical proximity in protein structures on fitness interactions between sites in a protein. It's impressive to see this discovered within a single species, though.

Chapter 4 concerns extreme overdispersion of the molecular clock on protein coding regions. Previous work has shown, in general, that deleterious mutilations alone can lead to an over-dispersed clock. Whereas Stolyarova studies an extreme case of this: bursts of substitutions attributable to positive selection over very short timescales. Key to this work, again, is a good choice of study species with well-resolved phylogenies.

The work is largely descriptive, and although synergistic epistasis may play a role, independent selection on sites due a sudden environmental shift is not ruled out. The main novelty here is that, unlike studies of classical overdispersion, positive selection is almost certainly required to explain the extreme bursts discovered.

Chapter 5 is more intimately related to the thesis' central theme of epistasis, and it considers the how propensities for fixations in a protein-coding genet change over long timescales, due either to epistasis (typically reducing a propensity for further change) or environmental fluctuation (typically increasing a propensity for further change). This is a lovely dichotomy to setup and analyze, the this chapter is noteworthy for the combination of analytic models, simulations, and methods of inference from phylogenetic patterns.

Both increasing and decreasing fitness (entrenchment and senescence) of the current allele at a site are observed in the data analyzed (five metazoan genes harboring 40,000 or so alleles). The relative rates (approximately equal) of these two phenomena, and types of sites (positively or negative omega) that feature entrenchment or senescence adds empirical detail to an otherwise largely theoretical literature that has focused more often on entrenchment by epistasis. The work is important because it provides a phylogenetic method to identify both phenomena, and the results will force theoreticians to recognize the role of changing environments that play out, in this analysis, equally often as epistasis/entrenchment over these timescales.

Overall the dissertation shows a high degree of intellectual rigor and originality. It would easily qualify as a successful PhD dissertation at my own institution. I do not have suggestions for edits or changes to the document, especially considering that two-thirds of the primary research has already been published and the remainder is well-written and clearly ready for submission. I look forward to questioning Stolyarova on details of the work and her future directions during the thesis defense.

Signed,

Joshua Plotkin

University of Pennsylvania

Provisional Recommendation

☐ XX *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*