Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Kseniia Safina

PhD Program: Life Sciences

Title of Thesis: Molecular epidemiology of socially important infectious diseases

Supervisor: Professor Georgii Bazykin

Name of the Reviewer: Dmitri Pervouchine

I hereby confirm the absence of any conflict of interest	**Date: 14.11.2021**

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

The PhD thesis "Molecular epidemiology of socially important infectious deseases" by Ksenia Safina presents a study of two important human illnesses, which are both caused by viruses, the human immunodeficiency virus (HIV) leading to acute immunodeficiency syndrome (AIDS) if untreated and SARS-CoV-2 causing COVID-19. The thesis contains abstract, the list of author's publications, acknowledgments, list of abbreviations, introduction, literature review, two chapters describing results on the molecular epidemiology of HIV in Oryol Oblast and that of SARS-CoV-2 in Russia early in the pandemic, respectively, both containing their own introduction, methods, results, and discussion, followed by the common conclusion, bibliography and appendices. The list of author's publications indicated that she has at least one shared first author paper and other papers. Therefore, the prerequisites for PhD defense are met. The bibliography contains 430 (!) citations. Overall, the thesis is well-written and well-structured, all the display items are relevant, and references are carefully cited. The defendant arrives to several conclusions about HIV-1 epidemic in Oryol Oblast, including lower-bound estimation of the number imports, dominance of subtype A and CRF63, estimation of the

effective reproduction rate, and under-reporting the MSM transmission route. The conclusions on SARS-CoV-2 also include the estimation of the number imports, the analysis of transmission lineages and nosocomial outbreak in the Vreden hospital in Saint Petersburg. Overall, the style of the presentation is descriptive. The methods used in the thesis seems relevant, and the conclusions are justified. However, below are a few points to be clarified.

1. The analysis of HIV-1 epidemic is based on sequencing of only a part of the HIV genome that includes the pol gene. The defendant should spend some time to discuss the limitations of the analysis that are imposed by this, especially considering the existence of a circulating recombinant form.

2. The routes of transmission of HIV-1 and SARS-CoV-2 are drastically different: while HIV is known to be transmitted sexually and parenterally, SARS-CoV-2 is transmitted by exposure to respiratory fluids. This aspect deserves to be mentioned somewhere early in the thesis to highlight the differences between the two illnesses (unless I overlooked it).

3. The manuscript contains multiple abbreviations, not all of which are explained. For instance SIV (semian immunodeficiency virus), IDU (intravenous drug users), CDC on p. 38. Although the list of abbreviations exists, it would be convenient for the reader to have these terms explained as they first appear in the text.

Minor points:

p. 26. l.8: “As higher thermodynamic stability of the transmitted variants suggests” – it is not clear for the context how the thermodynamic stability is related to selection.

p.31. l.8: The author could make use of footnotes to take a moment and appreciate current progress in the field. Also, the use of non-English characters at the end of the sentence is not acceptable for a phD thesis.

p. 62 Figure 3.1. I propose converting this display item to a table because some of the low counts cannot be seen from it. For instance, the author later mentions on p. 75 that there were a few MSM samples, while from Figure 3.1B it looks that there were none.

p. 66. Figure 3.4. The dependence of the inferred number of singletons and transmission lineages on

the number of sequences – doesn’t it depend not only on the number of sequences, but also on *which* sequences were used in the analysis? I missed the point here.

Nevertheless most of my comments are cosmetic and do not detract from the value of the thesis, and its author, Ksenia Safina certainly deserves to be awarded a PhD degree.

Sincerely,

Dmitri D. Pervouchine

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*