# Jury Member Report – Doctor of Philosophy thesis.

**Name of Candidate:** Sofia Medvedeva

**PhD Program:** Life Sciences

**Title of Thesis:** Natural diversity of CRISPR spacers

**Supervisor:** Prof. Konstantin Severinov

**Co-Supervisor:** Dr. Mart Krupovic

**Chair of PhD defense Jury:** Prof. Mikhail Gelfand **_Email_**: *m.gelfand@skoltech.ru*

**Co-Chair of PhD defense Jury:** Prof. Guennadi Sezonov

**_Email_**: *guennadi.sezonov@sorbonne-universite.fr*

**Date of Thesis Defense:** 3 June 2019

**Name of the Reviewer:**

| | |
|---|---|
| I confirm the absence of any conflict of interest<br>(Alternatively, Reviewer can formulate a possible conflict)<br>**NONE DECLARED** | **Signature:**<br>**Dmitri Pervouchine**<br>**Date: 06-05-2019** |

| |
|---|
| *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.*<br>*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:<br>• Brief evaluation of the thesis quality and overall structure of the dissertation.<br>• The relevance of the topic of dissertation work to its actual content<br>• The relevance of the methods used in the dissertation<br>• The scientific significance of the results obtained and their compliance with the international level and current state of the art<br>• The relevance of the obtained results to applications (if applicable)<br>• The quality of publications<br>The summary of issues to be addressed before/during the thesis defense |

The doctoral thesis by Mrs. Sofia Medvedeva, entitled “Natural diversity of CRISPR spacers”, is a systematic comparative study of the diversity of CRISPR spacers across different communities of prokaryotic organisms based on bacterial samples from a fossil mammoth intestine, from Antarctic snow, and from distant hot springs around the Earth. The manuscript contains the abstract, followed by the list of publications, on which the defendant is a co-author, the introduction, the formulation of the aims of the study, the results, which are subdivided into six chapters, each being represented by a peer-reviewed publication, the conclusions, and several supplementary sections, which include annexes and the list of citations. The dissertation is concluded with a Resume section written in French.

The abstract introduces the reader to the topic of the dissertation and briefly formulates the direction of the study. The introduction delivers the background by describing a variety of prokaryotic defense systems and proceeds to a detailed description of CRISPR systems as a particular subtype. The defendant did a great job: I am far from being an expert on CRISPR-Cas systems and I must say that for me it was an interesting and enlightening read. The introduction successfully describes the three functional modules of Cas proteins that are responsible for adaptation, expression, and interference, both from mechanistic and evolutionary perspectives. Then it discusses the distribution of CRISPR-Cas systems across bacterial and archeal domains and proceeds to a detailed description of CRISPR-Cas immunity in Sulfolobales. The introduction is concluded with applications of CRISPR-Cas in genomics. Overall, the introduction is very well-written, optimally structured, and contains all necessary background. I have only minor comments regarding the use of abbreviations. For example, at the bottom of page 12, the term “tracRNA” is used without definition. Similarly, the term “PAM sequence” first occurs on figure 3 on page 13, but it is introduced in the text only on page 14. All the acronyms need to be introduced before the first use, because otherwise it creates difficulties for a reader with little experience in CRISPR-Cas systems (like myself). It always helps reading when the author presents the table of acronyms beforehand. Also, I believe that the information on how CRISPR spacers are recycled is missing from the introduction.

The results are subdivided into six chapters, each represented by a peer-reviewed publication. The first chapter, called “Dynamics of Escherichia coli type I-E CRISPR spacers over 42 000 years”, is a study of type I-E CRISPR spacers of Escherichia coli from the intestine of a 42 000-year-old mammoth. This study revealed a large overlap between ancient and contemporary CRISPR spacers and also between their genomic organization (order) indicating that CRISPR arrays were not subject to intensive evolution over past time. I noticed that the contribution section on page 42 says that the defendant and another coauthor analyzed the data. However, since this is a qualification work, it would be necessary to know which exact part of the data was analyzed by the defendant, and which analyses were carried out by the other coauthor. The manuscript was prepared by the supervisor.

The second chapter, entitled “Metagenomic Analysis of Bacterial Communities of Antarctic Surface Snow”, is again represented by a paper under the same title. It is a study of flavobacterial type II-C CRISPR spacers in the samples of antarctic snow investigated by metagenome sequencing. The study reports a considerable diversity of bacterial species and their respective spacer arrays, and even more strikingly, a very different spectrum of CRISPR spacers compared to flavobacterial samples from the Northern hemisphere, indicating that bacterial strains in antarctic snow have evolved in a completely different environment compared to the rest of the world. The contribution section of the manuscript states that the defendant performed clustering and PCA analysis and prepared figures, while other steps including writing the paper were performed by other authors.

The third chapter, called “Natural diversity of CRISPR spacers of Thermus: evidence of local spacer acquisition and global spacer exchange”, investigates the diversity of CRISPR spacers of Thermus communities based on a diverse library of samples from Italy, Chile, and Russia. Unexpectedly, the study finds multiple common spacers in Thermus communities from different continents. It also finds examples of local acquisition of spacers from some phages that were isolated in the sampling sites, as well as distinct targeting patterns in different CRISPR-Cas systems. Overall, the study suggests a considerable amount of migration of thermophilic bacteria over long distances. As stated in the contribution sections, the defendant conducted bioinformatics analyses and prepared figures and tables, while the manuscript was written by the supervisor.

The chapter four references a manuscript called “Virus-borne mini-CRISPR arrays promote interviral conflicts and virus speciation”, in which the defendant is the first author. It is a study of all spacers contained in a hyperthermophilic archaea organism Sulfolobales. The study presents evidence of local adaptation, co-evolution of host/pathogen systems, and a special mechanisms of superinfection exclusion, a phenomenon in which a preexisting virus prevents a secondary infection with the same or a closely related virus. The defendant states that she performed CRISPRome data analysis, prepared display items and contributed to the text of the manuscript. However, the manuscript has no bibliographic reference and it would be good to know to which journal it is submitted. There is no reference in the list of publications on page 5 either. Perhaps the time that passed since the defendant prepared the manuscript allows to insert a valid bibliographic reference, or a letter of acceptance from a peer-reviewed journal.

The fifth chapter, called “Integrated Mobile Genetic Elements in Thaumarchaeota”, references a manuscript under the same title that is (presumably) accepted for publication in Environmental Microbiology. This work describes a family of integrated mobile genetic elements in archeal phylum Thaumarchaeota. The most interesting finding is that different members of that family use a shared gene network and thus may affect in an important way the fitness and adaptation of their hosts. The contribution of the defendant, as stated on page 112, was the identification of transposons in Thaumarchaeal genomes and the analysis of spacer diversity, which again represents a significant part of the analysis.

The last chapter, called “Avoidance of Trinucleotide Corresponding to Consensus Protospacer Adjacent Motif Controls the Efficiency of Prespacer Selection during Primed Adaptation.”, is represented by a published paper. In this study, the authors use a combination of bioinformatics and experimental approaches to identify factors that affect the efficiency of spacer acquisition in type I-E CRISPR-Cas system of E. coli during naive and primed adaptation. The results highlight differences in primed and naive adaptation manifested by the avoidance of AAG trinucleotide consensus PAM sequence, and support unidirectional spacer selection process during primed adaptation. A comment before this publication states that the defendant “obtained preliminary results of PAM avoidance in spacer sequences in primed adaptation experiments with different plasmids and lack of avoidance in CRISPRome mammoth data”. From this remark it is not clear whether the analysis of the preliminary data is different from the analysis of the published data, and whether the defendant participated in writing the manuscript.

The results are followed by conclusions, which as summarized in bullet points, and to a large extent coincide with the above mentioned conclusions about the evolution of CRISPR spacers and co-evolution of virus-host interactions. The conclusions are well-formulated and justified. The bullet points are followed by a discussion on similarities and differences between the studied systems, which brings together otherwise quite mosaic picture delivered by six different publications.

Overall, I find the dissertation of Mrs. Sofia Medvedeva presents a significant contribution to the current state of the art in research on the evolution of CRISPR-Cas systems; it is scientifically sound and uses the correct methodology. The defendant has demonstrated an ability to work with the literature, to perform bioinformatic analysis, to gather, analyze, and interpret the results. The validity of the results is supported by four peer-reviewed publications, on which the defendant is a co-author, one peer-reviewed publication, on which she is a main contributing co-author, and one under-review publication, on which she is the main contributing author. I recommend the following changes to be implemented in the manuscript:

1. A careful revision of the introduction in regard of acronym use, possibly adding a table of acronyms.

2. In order to highlight the contribution, I suggest a one-paragraph summary before each chapter in the Results section that would summarise the actual contribution of the defendant to that chapter, without references to display items therein, but rather explaining in plain language the analysis that was actually done.

With this amendments, I believe that the work by Mrs. Sofia Medvedeva would comply with all the PhD standards that are stated in Skoltech PhD policy, and that Mrs. Sofia Medvedeva deserves to be awarded with the PhD degree.

Sincerely,

Dmitri D. Pervouchine, PhD
Assistant Professor
Skolkovo Institute for Science and Technology
3 Nobel st., Skolkovo Innovation Centre,
Moscow, Russia 143025
+7 (495) 280.14.81 * 3925
+7 (916) 220.69.99

**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* |