![Skoltech logo](Skoltech - Skolkovo Institute of Science and Technology)

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Rim Gubaev

PhD Program: Life Sciences

Title of Thesis: Genetic association mapping for agronomically important traits in rapeseed and sunflower

Supervisor: Professor Philipp Khaitovich

Name of the Reviewer: Begoña Pérez-Vich

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 09-09-2022**

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

"Genetic association mapping for agronomically important traits in rapeseed and sunflower" dissertation is a comprehensive study focused on two of the most worldwide and Russian important oilseed crops, rapeseed and sunflower. Drawing on a wide amount of phenotypic, genetic and genomic data, and posing original and theoretically informed new questions regarding agronomic relevant traits, this dissertation explores and develops new breeding tools for genetic improvement of seed quality traits such as glucosinolate content in rapeseed and tocopherol and fatty acid profile in sunflower and seed morphology characteristics, and additionally, it stablishes new conclusions regarding their genetic basis and determines genetic diversity in Russian sunflower and rapeseed material. Therefore, it is considered that this dissertation presents an overall outstanding quality, with a high scientific relevance, since it fills gaps in existing scientific knowledge and it is an important contribution for breeding Russian material, and to understanding Russian germplasm diversity and the genetic basis of agronomically valuable traits.

The thesis adopts a well designed and thematic structure with three clear chapters (4, 5 and 6) focused of the results and discussion of the three independent studies carried out, and, in addition, a literature review (chapter 2), a materials and methods description (chapter 3), and final conclusions and future perspectives (chapter 7). It has also an introduction very useful to get into the subject of the dissertation (chapter 1). The dissertation presents therefore a clear and well-designed structure which allows the reader to follow and understand all the steps of the research studies carried out.

The dissertation's Literature review (second chapter) contains a helpful and well written overview of literature on marker-assisted breeding in plants, including new genomic approaches, and specifically in sunflower and rapeseed, and covers many aspects regarding the genetic basis of the seed traits that are subject of the research carried out. Finally, it provides the current context surrounding Russian trends in sunflower and rapeseed breeding. Some minor comments regarding this chapter are:
- Page 27, line 9. I would suggest to also add the reference Fernández-Aparicio et al. (2022) in the "most recently identified sunflower resistance gene localized to chromosome 4".
- Page 31, line 3. In the sentence "…QTLs located on chromosomes (…..) were considered as major effect loci controlling erucic acid content (Howell et al., 2003…" change "erucic acid content" by "glucosinolate content".

The dissertation's third chapter describes the plant materials and methods employed. The methodology used is relevant, the experiments are well designed and data analysis is described clearly and in detail. The use of genome-wide marker analysis together with well performed phenotypic measurements (please see below some comments relating to this) for association mapping of the agronomically important studied traits is accurately carried out and described in detail. Regarding phenotypic measurements, glucosinolate and tocopherol contents are based on colorimetric analysis, and on thin layer chromatography, respectively. For both traits, these techniques show in general a high correlation with those obtained with more sophisticated high-performance liquid chromatography (HPLC). These phenotypic measurements are not negatively considered since they provide rapid and cost-efficient methods for the measurement of these traits. However, it has to be taken into account that useful information regarding glucosinolate profile of the samples or a more accurate tocopherol profile provided by complementary HPLC analyses might contribute to increase the results accuracy and the already high scientific value of the research. Some other minor comments are indicated below:
- Page 46, line 8. Since individuals genotyped by less than 80% of SNPs were removed, please indicate the final number of individuals used to construct the genetic map.

The dissertation's fourth chapter contains the results and discussion of the study carried out for

determining (i) the genetic diversity and structure of a VNIIMK collection of 90 Russian rapeseed accessions and (ii) the genetic basis of glucosinolate content through an association mapping study. Using a genotyping by sequencing (GBS) approach and strict SNP filtering, a final set of 12226 SNPs was used. Collection structure characterization revealed, as expected, differentiation between spring and winter accessions, and, additionally, between winter yellow- and dark-seeded genotypes. Association mapping for glucosinolate content was based on three independent phenotypic measurements carried out across three years, which made results obtained robust and accurate. Notably, the SNPs detected on chromosome A7, which showed stability and significant effects across the three years, represent an outstanding result and the discussion regarding their underlying candidate genes an important advance and starting point to better understand glucosinolate biosynthesis. Please find below other detailed comments:

- Page 52, line 18. Please change Figure 4.2.2 B by Figure 4.2.2 C.
- Page 53, line 7. Please change Figure 4.2.2 by Figure 4.2.2 B
- Page 53, paragraph from line 4 to line 12. The first sentence of this parragraph (lines 4 to 7) is about the results of the analysis of the group of spring accessions, and the second sentence (lines 7 to 12) summarizes again the results of the analysis of the winter accessions. Please add this last sentence to the previous paragraph starting in page 52 (line 16) and ending in page 53 (line 3) where the discussion regarding the sub-structure of the winter accessions is carried out.
- Pages 54 and 55. I consider that the diversity analyses of the Russian collection compared to the international one should be carefully reviewed. It is stated in page 54 (lines 22-23) that selected SNPs for the analyses were polymorphic in "at least one data set". It is not clear if these SNPs were common between the international and the Russian sets, or that they were polymorphic within a set, and not present in the other. This is said because the clear differentiation between the international and the Russian collecions might be due to the existence of a very limited set of shared and common SNPs between the two collections. The number of SNPs shared between the Russian and the international collections to carry out the diversity analysis should be clearly stated in order to clarify this issue. Additionally, a brief summary of the breeding history and collection origin of the Russian rapeseed accessions, if available, will contribute to this discussion.

The dissertation's fifth chapter deals with the QTL mapping of tocopherol profile and oleic acid content in sunflower, using a GBS approach in experimental crosses. QTL-mapping of tocopherol composition clearly identified the *Thp1* and the *Thp2* loci on chromosomes 1 and 8, respectively, as the major factors underlying the modified profiles. In addition, the newly identified SNP markers associated to the different tocopherol classes were tested and validated using Sanger sequencing, making these SNPs prospective for marker-assisted selection. In relation to oleic acid content, a major QTL was found of chromosome 14 for the cross involving the high oleic parental line VK195. I have only minor comments which are detailed below:

- Pages 84- 85. In my opinion, the analysis of the oleic acid locus located on chromosome 7 should be reviewed, especially in relation to the chromosome 7 map construction. It is indicated in page 84 that the reassembling of chromosomes 7 and chromosome 14 resulted in a significant amount of the genetic markers from chromosome 14 becoming a part of chromosome 7. However, from data provided in Table 5.3.1, chromosome 7 in the VK876*VK101 map showed the lowest Pearson correlation coefficient (0.48), while that in chromosome 14 was high (0.9). Therefore it seems that initial chromosome 7 map was much less accurate than that on chromosome 14, and that the "true" oleic acid QTL was lying on chromosome 14 as the initial analyses indicated. It is unlikely that a new and not previously

described dominant mutation for oleic acid content lying on chromosome 7, different from that of the Soldatov mutant-derived material, exists in one single genotype (VK876). Breeding history of VK876 would be also useful in order to determine if the development of this line involved the Soldatov high oleic acid variety Pervenets or material derived from it.

The dissertation's sixth chapter describes an association mapping study in sunflower of seed morphology traits (seed size, husk size and seed to husk ratio) involving an important number of sunflower accessions (a total of 601), using also a GBS approach. Diversity analysis of the Russian collection revealed no clear population structure and a close similarity with the structure of international accessions. Phenotypic measurements of the seed morphology traits are clearly described. However, they were obtained in one single environment, which might be the cause of the low phenotypic variance that associated SNPs explained, together with the complex genetic basis of these traits. This is discussed by the author, which is highly appreciated. Some other minor comments are indicated below:

- Botanically speaking "sunflower seed" is an achene. Therefore, the term "sunflower seed" is actually a misnomer when applied to the seed (kernel) in its pericarp (hull). Please, clarify in the first mention in the introduction (page 94) if when you speak about "seed size" it is in reference to the aquene or the kernel size. Please clarify also this if possible in the cited references.
- Page 96. In Table 6.2.1 Add if possible the number of accessions analyzed within each collection.

Finally, I consider that global results derived from this thesis have great interest for the advance in the scientific knowledge of rapeseed and sunflower and have important implications for breeding Russian material using the newly developed and validated tools. The author has produced an excellent piece of research, methodologically sound and rich, with important findings. As a result, this dissertation represents an important contribution to oilseed crop breeding and knowledge. Additionally, indicate also that the quality of the publications is very high. All publications are in high impact factor journals (2021 SCIE I.F. from 3.54 in G3 to 4.56 in BMC Genomics) (2021 Journal Impact Factor, Journal Citation Reports, Clarivate 2022), being all included in the second quartile (Q2) of the SCIE category "Genetics and Heredity". It is highly appreciated that the thesis results included in each publication are clearly explained at the beginning of each chapter of the dissertation.

Cited references: Fernández-Aparicio M, Del Moral L, Muños S, Velasco L, Pérez-Vich B. Genetic and physiological characterization of sunflower resistance provided by the wild-derived Or_{Deb2} gene against highly virulent races of *Orobanche cumana* Wallr. Theor Appl Genet. 2022 Feb;135(2):501-525. doi: 10.1007/s00122-021-03979-9.

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*