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:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	Signature: Date: 07-09-2022

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

Brief evaluation of the thesis quality and overall structure of the dissertation.

The doctoral thesis "GENETIC ASSOCIATION MAPPING OF AGRONOMICALLY IMPORTANT TRAITS IN RAPSE AND SUNFLOWER" is a solid, well-structured scientific report. This work will significantly contribute to the development of marker-assisted and genomic selection of oilseed crops in Russia.

The Introduction (Chapter 1) describes the problems associated with food quality, the requirements of the food industry, and climate change that can potentially be solved by introducing genetics-based technological approaches in agricultural practice. In particular, the problems of the production of oilseeds (rapeseed and sunflower) in Russia are well described. In accordance with the problems discussed, the goal of the study is formulated and specific tasks are set within the framework of this goal. It is noted that the dissertation is actually the results of a successful joint research project of the Skolkovo Institute of Science and Technology and leading organizations involved in plant germplasm conservation (VIR) and oilseed breeding (VNIIMK and Agroplasma breeding company).

The Literature review (Chapter 2) discusses the concept of marker-assisted selection (MAS), the main advantages of SNP markers for association mapping and genomic selection (GS). Various types of segregating populations, high-throughput genotyping technologies and approaches to search for associations between markers and trait variation are well discussed. A significant part of the literature review is devoted to the methods of genotyping technologies for the routine evaluation of plant material, which are in great demand in practical breeding. Past and current trends in sunflower and rapeseed breeding, as well as current knowledge about the genetics of agronomically important traits of these oilseeds, are highlighted. For rapeseed and sunflower, the main breeding targets in Russia are outlined.

The Material and Methods (Chapter 3) presents an overview of rapeseed and sunflower diversity panels used for the study, descriptions of experimental crosses performed, phenotyping procedures including glucosinolate, tocopherol composition and oleic acid content measurements, seed-related traits assessment as well as genotyping procedures and SNP calling. Genetic map construction and QTL mapping approaches are described.

Chapter 4 describes the significant results obtained during the genetic characterization of Russian rapeseed germplasm collection using 12,226 SNP markers. The principal component analysis expectedly revealed genetic distinction of spring and winter rapeseed, however, the clear genetic segregation of winter rapeseeds into two subpopulations of yellow- and dark seeded lines was reported for the first time. A remarkable conclusion was also made when comparing rapeseed genetic diversity from the international and Russian germplasm collections: there is a clear difference between the VNIIMK and international accessions which indicates that Russian rapeseed germplasm has great potential to increase the genetic diversity of rapeseed worldwide and search for new genetic determinants of agronomically important traits. As a result of the association mapping procedure, new genetic loci associated with glucosinolate content in rapeseed were discovered.

Chapter 5 is dedicated to genetic factors affecting oil-quality traits in sunflower – tocopherol composition and oleic acid. Two genes with major effect on tocopherol composition Tph1 and Tph2 were described earlier and mapped on chromosome 1 and 8 in sunflower. The corresponding causative genes encoding 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase and γ-tocopherol methyltransferase were also described. Here, the experiment was performed to identify loci and perspective SNP markers associated with altered tocopherol composition and high oleic acid content using two experimental crosses of modern lines from VNIIMK in order to search for possible new additional genetic determinants of the oil-quality traits. New high-density genetic maps for two crosses were developed with 3200 and 2571 SNP markers revealed with genotyping-by-sequencing approach. As a result of QTL mapping procedure, Tph1

and Tph2 loci were mapped in the chromosomes 1 and 8, as it was described before. SNP markers closely linked to the QTLs affecting tocopherol composition and oleic acid content were validated using Sanger sequencing, the accuracy varied from 85-94%. Thus, the SNP markers developed in the study are readily available for marker-assisted breeding applications.

In Chapter 6 a very diverse collection of 601 sunflower accessions from VIR and VNIIMK and the Agroplasma breeding company was evaluated for seed morphology traits (seed size, husk size and seed to husk ratio) in a non-invasive manner by X-ray radiography methods followed by image analysis. 15,068 SNPs were obtained for 601 accessions from three collections by means of GBS allowing to perform the genetic diversity analysis and reveal the population structure. I found a clear separation of fertility restorer and fertility maintainer lines of the Agroplasma collection from the other sunflower diversity in Figure 6.3.1, although the author classified that as "slight differences". I believe, this is another valuable finding of the dissertation. Association mapping of seed-related traits resulted in the discovery of highly significant SNPs on chromosome 10, as well as additional SNPs located on chromosomes 4, 9 and 17. Candidate genes for the mapped loci have been proposed.

Chapter 7 summarizes results, outlines conclusions and discusses future perspectives of applying the developed genetic markers for quality improvement of rapeseeds and sunflower. Seven main conclusions of the work are given, each conclusion is commented on in terms of the possibility of using it in breeding practice or the prospects for further in-depth study.

The relevance of the topic of dissertation work to its actual content
The dissertation reports an impressive set of SNP markers linked to the genetic loci underlying variation of agronomically important traits in rapeseed and sunflower. SNPs associated with glucosinolate content in rapeseed varieties from the VNIIMK were reported. SNPs linked to genes responsible for the tocopherol composition and oleic acid content of sunflower oil were identified. Genetic associations of SNPs with seed-related traits of sunflower (e.g. husk size, seed to husk ratio) were investigated. The topic of the doctoral thesis perfectly fits to its actual content.

The relevance of the methods used in the dissertation
Methods (both phenotyping and genotyping of plant material) used in the dissertation are adequate and impressively diverse. The analytical methods include a thin-layer chromatography (TLC) with subsequent densitometry to determine tocopherol composition in rapeseed, a gas chromatography followed by mass spectrometry to characterize relative oleic acid content, mobile X-ray diagnostic installation for seed-related traits assessment in sunflower. Genotyping-by-sequencing (GBS) approach was used to find SNPs, the original modified version of the pipeline for SNP calling was developed and published by the author. Statistical methods for linkage mapping, QTL mapping and detection of genetic associations are applied at a professional level.

The scientific significance of the results obtained and their compliance with the international level and current state of the art
For the first time a clear genetic difference between the Russian and International rapeseed germplasm collections has been demonstrated. The difference indicates the great potential of Russian rapeseed varieties as a source of new alleles for improving agronomically important traits.
For the first time, the genotypic profiles of sunflower accessions from Russian and international collections of *Helianthus* germplasm represented by wild, cultivated and local varieties were compared. This exhaustive genetic analysis was performed on 2345 SNPs shared by 601 accessions from the Russian sunflower germplasm collection and over 1000 accessions for which genotyping data was provided by the Biodiversity Research Centre of the University of British Columbia. A clear genetic separation between

wild and cultivated sunflowers has been shown, and it has also been well demonstrated that landraces create a genetic 'bridge' between wild and cultivated accessions.

The major achievement of the research is the discovery of genetic markers associated with agronomically important traits in rapeseed and sunflower which will promote marker-assisted breeding of these oilseed crops in Russia.

The relevance of the obtained results to applications (if applicable)
There are at least three achievements of this research work with clear practical applications: 1) 7 SNP markers associated with the glucosinolate content in rapeseed that can be used for oil and oilcake quality improvement with marker-assisted selection; 2) validated SNP markers of Tph1, Tph2, and Ol loci to assist breeding of sunflower with improved tocopherol composition and oleic acid content; 3) 25 SNP markers associated with the seed-related traits to assist sunflower breeding, e.g. the development of large-seeded confectionary sunflowers.

The quality of publications
The content of the Doctoral thesis is comprehensively presented in three publications in high-level scientific journals, including BMC genomics (Q1 in Genetics and Biotechnology), G3: Genes, Genomes Genetics (Q2 in Genetics) and Genes (Q2 in Genetics). Additional relevant papers have been published in Agronomy (Q1 in Agronomy and Crop Science) and Biomolecules (Q2 in Biochemistry and Molecular Biology).

There are some comments/questions to be addressed:

1. Page 110. "*The QTL mapping of oil quality traits in Russian plant material identified novel SNP markers for previously reported Tph1, Tph2, and Ol loci associated with tocopherol composition and oleic acid content, respectively*".

If *Tph1* and *Tph2* loci were previously identified and even causative genes (2-methyl-6-phytylbenzoquinol methyltransferase and γ-tocopherol methyltransferase) were suggested, what was the reason to search and validate the linked SNPs instead of simply re-sequencing the causative coding sequence to detect functional polymorphism in lines with contrast tocopherol composition?

2. Page 81. "*To map oleic acid content, three approaches have been used. First, the raw phenotypes were mapped, i.e the relative content of the oleic acid by two methods: interval mapping adapted for non-normally distributed traits and composite interval mapping as it was previously used to map this trait*".

Why is interval mapping supposed to be adapted for non-normally distributed traits?

3. Page 97. "*Analysis of variance revealed a significant difference of seed and husk area between the collections (Figure 6.2.2). Namely, husk was on average 2.65 times smaller in VIR accessions when comparing VIR and VNIIMK collections….. Such differences in seed and husk sizes could be explained by the fact that the VIR collection mostly consists of historical samples, while VNIIMK and AGROPLASMA collections include economically valuable accessions used to produce commercial hybrids and lines for which the size of the seed is one of the key traits to be improved. Thus, the differences are explained by the artificial selection pressure different for the studied collections*".

I am not sure that it is wise to consider the VIR collection as a one single slot in this analysis. It would be more reasonable to consider the stratification of the VIR collection according to the years of entry and origin of these 255 accessions. Indeed, there are many historical seed accessions in the VIR collection, but VIR, as a gene bank, was obliged to collect and keep modern varieties as well.

4. When mapping SNP markers that significantly associated with glucosinolate content in rapeseeds:

"… *two SNPs (SA7_26967214 and SA7_26967217) explained from 13.8 to 20.4 % of phenotype variance across three-year observations (Table 4.4.1).*", page 59.

Remarkably, these two SNPs are just 3 bp apart, while the percentage of phenotype variance they explain differs a lot. What could be the reason?

5. Page 57. "*It was demonstrated that no significant difference between the winter and spring phenotypes (in terms of the glucosinolate content) was identified at a 5% significance level for observations made in 2016 and 2017. Slight differences were only observed in 2018*".

This is an interesting observation. The gene systems that are responsible for plant fitness (e.g. spring/winter habit) may not affect the content of glucosinolates under stable environmental conditions, but turn on and play their role during stressful years. In this context, it will be useful to describe briefly the environmental conditions of the three-year field evaluation experiment.

Some minor spelling errors:

Page 57. "*This is due to the fact that it was previously demonstrated that quantitative traits such as yield, oil content, and height differences between the different rapeseed ecotypes (Assefa et al., 2018; Fridrihsone et al., 2020).*" The sentence is not finished.

Page 58. "After the application of the Bonferroni correction, two SNPs located 7 (A7) remained above the threshold (Figure 4.4.1, Figure 4.4.2)." Located on chromosome 7?

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*