Skoltech
Skolkovo Institute of Science and Technology

# Thesis Changes Log

**Name of Candidate:** Rim Gubaev

**PhD Program:** Life Sciences

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

**Supervisor:** Prof. Philipp Khaitovich

*The thesis document includes the following changes in answer to the external review process.*

I would like to thank all jury members for the careful review of my thesis and valuable comments. I have addressed all comments and questions and have made several changes reported in detail below.

**Professor Elena Potokina**

**1) 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.**

Thank you for your comment! Indeed there is a difference, however, we could not claim it as significant as the ADMIXTURE analysis did not reveal any clear clusterization.

**2) 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?**

Thank you for this question! The initial idea of using high-throughput genotyping and subsequent QTL analysis was because there were several studies discussing the presence of additional minor effect loci associated with the oleic acid content and tocopherol composition. So on the one hand we aimed to confirm the existing loci. On the other hand, we aimed to find additional minor effect loci, explaining, for example, variations in tocopherol composition in tocopherol phenotypic classes (corresponding changes were added to chapter 5.1). Unfortunately, we managed to find only major effect loci for oleic acid content and tocopherol composition since we used lines with high expressivity of these mutations. Nevertheless, we identified genetic markers for oleic acid that demonstrated recessive effects on phenotype these were one of the first reported cases.

**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?**

This was done since the distribution of the oleic acid content in the present study was non-normal it was bi-modal. So formally in that case a trait should be analyzed accounting for non-normality as it was done on the thesis, as the interval mapping function in r/qtl package has such an option. Notably, composite interval mapping implemented in r/qtl package is not adapted for the non-normally distributed traits. However, regardless of this composite interval mapping was applied to non-normally distributed traits several times including the analysis of tocopherol composition and oleic acid content.

**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.**

I agree with that comment, but unfortunately, the information on the years of entry as well as the origin was not provided to us. However, we see that the provided material demonstrated lower sizes of the seeds compared to VNIIMK and AGROPLASMA. The major reason to perform this analysis was to test if the collection of origin significantly affects the seed trait parameters. Our analysis revealed such differences, as a result, the additional confounding factor indicating a collection of origin was added to the model to correctly perform GWAS analysis.

**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?**

These markers explained from 13.8 to 20.4% in different years (across three years). This means that for different years these markers will explain the different proportions of variance explained due to the different phenotype distributions. The proportion of variance is the same for each of the years for these two SNPs.

**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.**

Thank you for this comment, indeed we identified this marginal difference which could be because spring/winter ecotypes could behave differently under stressful years. The corresponding discussion was added to the main text. According to the climatic data no strong changes within the vegetation period of 2018 compared to 2016 and 2017 except for the dryer in November of 2017 (monthly cumulative precipitation = 49.9 mm) compared to the similar periods of 2015 (78.1 mm) and 2016 (93.8 mm). This probably could affect the glucosinolate accumulation in winter lines. The corresponding discussion was added to section 4.3.

**6) 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.**

Fixed.

**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?**

Fixed.

**<u>Professor Yalcin Kaya</u>**

**1) In the abstract and conclusion parts, the author should mention some practical results of the conducted research, because there are no concrete results and suggestions for researchers who will work on same subjects in the sunflower and rapeseed as well as breeders how to utilize these findings and adapt their breeding programs.**

The information on practical results was added to the Abstract. The suggestions on how the validated oil-quality markers could be used were added to the "Conclusions and future perspectives" chapter.

**2) The author should mention especially for molecular results how to transfer these findings to formulate and facilitate the selection process for classical breeding program based on study results.**

Thank you for this comment! Different methods could be used to utilize molecular markers. The information on different tools i.e. allele-specific PCR, LAMP-PCR and ampli-seq were added to the "Conclusions and future perspectives" chapter.

**3) In some part of the thesis, some results especially molecular studies are not clearly mentioned for important traits. For instance, these used markers are polymorphic or not, it could be used for all genotypes and these are genotype dependent or not.**

The used markers are polymorphic in the studied populations, as by default SNP markers assume the presence of the polymorphisms which is required for marker identification. Regarding the genotypes, the markers for oil quality were tested across the different genotypes, while markers of the glucosinolate content and seed morphology should be subjected to additional testing. The corresponding information was added to the "Conclusions and future perspectives" chapter.

Minor corrections were done in accordance with comments in the attached pdf file.

**Dr. Dragana Miladinovic**

**1) Chapter 1 Introduction**
**The mutations mentioned ass innovations are not spontaneous but induced. Term used for NBT is not genetic editing, it should be replaced with genome editing.**

Thank you for this comment, agree with the incorrectness of using spontaneous instead of induced. This was fixed through the text. Agree with the fact that new breeding technologies (NBTs) assume genome editing, not "genetic editing". This was fixed through the text.

**2) 3.1.1 Rapeseed diversity panel**
**Please confirm if genotypes used are all inbred lines, or there are some varieties. The table with all genotypes used, some traits and origin would be of use.**

These are inbred lines, the corresponding notion was added to section 3.1.1. Detailed information on accessions on the sunflower diversity panel was previously provided as supplementary tables in Genes paper (Gubaev et al., 2020). Corresponding links to supplementary material were provided within section 3.1.1.

**3) 3.1.2 Experimental crosses for oil-related traits mapping in sunflower**
**The reference should be added related to info provided on traits of parental lines (VK101, VK303, VK876 and VK1959.**

References were added at the beginning of section 3.1.2 i.e the papers with detailed descriptions of VK101, VK303, VK876 and VK195 were cited.

**4) 3.1.3 Sunflower diversity panel**
**The table with all genotypes used, some traits and origin should be provided.**

The information on accessions on the sunflower diversity panel was previously provided as supplementary tables in our BMC Genomics paper (Chernova et al., 2021). Corresponding links, as well as addresses of the collection holders (for more detailed information), were provided within section 3.1.3.

**5) 3.2.1 Glucosinolate measurement**
**Add "in rapeseed" in the title. Please provide more information on the plant material used: how many plants/line, etc.**

The title was modified, the additional information was provided.

**6) 3.2.2 Tocopherol composition and oleic acid content measurement**
**Add "in sunflower" in the title. Please be more precise regarding plant material used: seed quantity, genotypes, plants/genotypes, was material collected only one year, or three? Provide the same info both for tocopherol and fatty acids.**

The title was modified, the additional information was provided.

**7) 3.2.3 Seed-related traits assessment**
**Add "in sunflower" in the title. Same as above, provide more detail on plant material analyzed.**

The title was modified, the additional information was provided.

**8) 3.4.1 Association mapping of glucosinolate content**
**Add "in rapeseed" in the title.**

Fixed.

**9) 4.4 Association mapping and scanning for novel candidate genes**
**No significant associations for SNPs from the previous studies were identified in the experiment. It has be done to a certain extent, but please elaborate on this further. Name other possible reasons, what does it mean from the aspect of the validity and applicability of your results.**

Thank you for this comment! As it is mentioned in the thesis the two possible reasons for not finding common regions were discussed, first the difference in the genetic background (including specific mutations associated with the identified SNPs) of the studied cohort. Second, the differences in the phenotypic variance between the studded and previously analyzed accessions. The third possible reason that was not mentioned in the text could be explained by the different environmental conditions, namely none of the previous studies were carried out in the Krasnodar region. At the same time, it was demonstrated that glucosinolate content is highly dependent on the climatic condition and there was a study identifying year-specific SNPs.

Regarding the applicability of the results, I could add that the identified SNPs should be used only for the studied accessions, as we identified these polymorphisms for a very limited collection of VNIIMK institute. The corresponding changes were added in the chapter in chapter 4.4.

**10) 4.5 Conclusions**
**Please name identified genes, regions and markers when mentioning them in the text.**

The name of the markers, genes and regions were added to chapters 4.4 and 4.5.

**11) 5.7 Conclusions**
**In the thesis, no additional loci controlling tocopherol composition were identified, whish is in contrast with some other studies. So, please elaborate further, explain possible causes, are these findings unique for this study, or some other authors also did not observe any additional loci. The same stands for discordances with the hypothesis initial set for oleic acid content.**

In fact, in several previous studies, the only single major effect locus was shown to be associated with for each of the Tph1- and Tph2-related phenotypes (as in the present study). While in other studies indicated some additional minor effect loci. The same is applied to the oleic acid content. According changes and discussions were added to the main text.

**12) When you mention a set of genetic markers for tocopherol content identified, please name them.**

Done.

**13) 6.5 Conclusions**
**In the thesis, only preliminary results of the association mapping of seed-related traits are provided. Hence, this should be pointed out both in the results, conclusions, final conclusions, etc. It is also claimed that these preliminary results provide some insights on seed-related traits. What are those insights? Are there relevant since this is only preliminary study without strong phenotyping data? Furthermore, It should be noted that the identified genetic markers explained not more than 7% of phenotypic variance.**
**Could they be considered genetic markers and of any breeding value at all? If yes, please explain why.**

I agree that the "insights" might be too promising I meant that we identified potential candidate genes for these traits and differences in seed-related traits among studied collections. This sentence was rewritten. I also agree that markers that explain the too low phenotypic variance, thus should not be considered for marker-assisted breeding. Corresponding changes were added to section 6.5.

**14) Chapter 7. Conclusions and future perspectives**
**Point 3) Add “Preliminary” before association. It is stated that 25 SNP markers associated with the seed-related traits are identified and, along with new candidate genes. Could you claim this, as it differs from what is said above in section 6.5. Should be rephrased as to reflect that the results obtained are preliminary.**

Point 3 was changed following the suggestion. In section 6.5 it was additionally stated that 25 SNPs are associated with the studied traits and that the results of chapter 6 are preliminary.

**Dr. Sreten Terzic**

**Minor corrections are needed according to the comments inserted in the thesis pdf document.**

The minor corrections highlighted in the attached manuscript were added to the text.

**Dr. Ilya Kirov**

**1) Chapter 4.2: the number of accessions used for genotyping should be mentioned in the beginning of this part.**

Fixed.

**2) The explanation of differences between yellow- and dark seeded winter lines revealed by PCA is unclear to me. What does "the recent breeding history" mean ?**

Thank you for this comment. It means yellow- and dark-seeded accessions were used to produce independent lineages without mixing which lead to the separation of these accessions. The corresponding changes were added to chapter "4.2 Genetic characterizations of VNIIMK rapeseed collection"

**3) Chapter 5.2: why these lines (VK101, VK303 etc.) were involved in crosses ? It may be better explained in the text. Also, the names of the lines differ between text and Table 5.2.1 ('B' should be replaced by 'V' letter).**

These lines represent modern perspective lines used in VNIIMK to produce a hybrid "Typhoon" (VK101 x VK303) and "Oxy" hybrid (VK876 x VK195) that synthesizes oil with increased oxidative stability. The corresponding changes were added to Chapter 5.2.

**Dr. Perez-Vich**

**1) 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".**

The citation was added, thank you for the suggestion.

**- 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".**

Changed, thank you for the suggestion.

**2) 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 phenotypicmeasurements, 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.**

Thank you for this comment! Indeed we did not use the sophisticated UPLC-based approaches to measure glucosinolates and tocopherol composition, as these methods were not available to our collaborators form VNIIMK. Thus cheaper and faster approaches were used instead. We could perform additional analysis by HPLC-based methods in the future to increase the 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 final number of genotypes used to construct genetic maps was 136 and 142 for crosses VK195xVK303 and VK876xVK101, respectively. The corresponding changes were added to chapter 3.5.

**3) 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.**

Done.

**- Page 53, line 7. Please change Figure 4.2.2 by Figure 4.2.2 B**

Done.

**- 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.**

Done.

**- 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.**

Thank you for this comment! These SNPs were polymorphic in each of the collections. So each collection was analyzed separately since the GBS and WGS require different analysis pipelines. Next, the unfiltered SNPs (160,257) for VNIIMK (i.e polymorphic for VNIIMK) were joined with the intersected SNPs (4,037,572) resulting in 20,848 common SNPs used to perform PCA analysis. I agree that the statement "polymorphic in at least one dataset" is confusing as these are 20,848 common SNPs, this statement was removed. Unfortunately, the detailed breeding history of the Russian collection was not available to us thus the discussion on that point is not possible.

**4) 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.**

Thank you for this comment the low correlation coefficient could be since the construction of the genetic map is based on stochastic algorithms that hardly could reconstruct the real order of the genetic markers. Additionally, this could be because this translocation indeed exists and, in turn, affects the quality of markers' order (i.e correlation between physical and genetic map). I agree that it is unlikely that the Soldatov mutation is located on chromosome 7, it is possible to speculate that the locus with the Soldatov mutation was translocated to chromosome 7, the fact that one SNP marker from the 14th chromosome translocated to the 7th chromosome was significantly associated with the oleic acid content. However, to test this hypothesis a detailed molecular cytogenetic analysis (for example fluorescence in situ hybridization (FISH)) is needed which was not the focus of the present study, however, could be applied in the future. According to the VK876 breeding history, the *Ol* was transferred from VNIIMK line LG18 which in turn obtained Ol from Pervenets. The corresponding discussion was added to the main text in chapter 5.5.

**5) 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.**

Kernel and hull terms were mentioned in the introduction section. When the term seed is used it corresponds to the kernel, while husk size corresponds actually to hull size. The corresponding changes were added while citing references.

**- Page 96. In Table 6.2.1 Add if possible the number of accessions analyzed within each collection.**

The information was added.

**Professor Viktor Korzun**

**1) Chapter 4 starts with the population structure analysis of the Russian rapeseed germplasm from the VNIIMK collection as well as its comparison with the international accessions. Here it should be noted that not so many international accessions were used to compare the genetic diversity.**

Thank you for this notion, this is related to the fact that at the end of 2018, very little sequencing data was publicly available for the rapeseed. Thus we selected 54 geographically and ecotypically diverse rapeseed lines for which the whole-genome sequencing data was available and used them for comparison.

**2) Chapter 5 is devoted to QTL mapping of oil-related traits in sunflower by analyzing F2 crosses. In the present chapter novel SNP markers were detected as well as validated using an independent plant sample. Here I would like to highlight several issues. First, the description of Figure 5.2.2 is confusing to some extent, looks like the mutant and wild-type lines are mixed up. Second, Table 5.2.3 as well as the subsequent paragraph is confusing due to chi-square values for 3:1 and 1:3 segregation ratios non correctly assigned for crosses.**

The caption for Figure 5.2.2 was changed. Table 5.2.3 as well as past and following paragraphs were rewritten.

**3) Chapter 6 describes the genetic diversity of the large sunflower diversity panel from two scientific institutions and one breeding company from Russia. A comparison of genetic diversity with international accessions was also demonstrated. Joint phenotype and genotype analysis revealed genetic markers explaining a low amount of phenotype variation (<10%) as well as candidate genes involved in seed parameters expression. Despite the novel results including new genetic markers and candidate genes identified for the studied cohort, these results are just preliminary ones as no information on GxE, as well as phenotype replications were provided. Thus, these results should be confirmed by the analysis of independent plant samples and/or additional data for different vegetation seasons and/or locations. This is also stated by the author.**

I agree with this notion, and indeed we are planning to do these experiments in the future since the result should be confirmed by analyzing GxE interaction for the part of the studied cohort. Or in better cases, the observations should be collected for the whole cohort with the aim of re-analyzing the whole mapping data. Thus the results currently are only preliminary.

**Professor Elena Salina**

**1) An introduction and literature review gives a clear problem statement as well as the current status of the research. This gives the opportunity for the reader to familiarize themselves with the relevant problems in sunflower and rapeseed breeding as well as with modern approaches used to solve them. At the same time, the information about the size of the genomes as well as its ploidy is missing and should be added to the text.**

Thank you for this comment, the information on the genome size and ploidy for rapeseed and sunflower was added to the section "2.3 Sunflower and rapeseed: past and recent breeding trends".

**2) The experimental part (methods) is clearly described. The described methods are considered modern and relevant to address the aim and the objectives of the thesis. The plant material includes plant diversity panels and experimental crosses which allow the application of different genetic mapping approaches. The author used a state-of-art genotyping approach based on high throughput sequencing to characterize plant accessions in the study. The phenotyping part of the presented methods section was done in collaboration with research institutions and private breeding companies. The modern bioinformatics and quantitative approaches applied were described in detail. The information on the sunflower and rapeseed lines is missing in the text so it should be added either by citing supplementary materials from published papers or added as a supplementary table to the thesis.**

The corresponding information was added to the text. In particular, web links to the previously published supplementary tables containing accession IDs as well as some descriptions of used lines were cited. The links were provided in sections 3.1.1 and 3.1.3.

**4) Unfortunately, there are certain inaccuracies in the description of the results presented in Figure 5.2.2 when mentioning mutant and wild-type lines. The data has not been entered in Table 5.2.1 very carefully (mix Russian and English styles) and there is a typo in the correspondence of the crosses and chi-square values in the description of Table 5.2.3**

The caption for Figure 5.2.2 was fixed. Table 5.2.3 as well as past and following paragraphs were rewritten. In Table 5.2.1 line names were switched to Latin.

**Professor Georgii Bazykin**

**The results of the work are published in five papers in renowned journals, on two of which Rim is the first author. These publications are sufficient for defense at Skoltech. I was unable to find information on presentation of these results at international meetings; this should be clarified at the defense.**

Thank you for this comment! The list of presentations at international conferences was added to the "Publications" section.