Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Alina Chernova

PhD Program: Life Sciences

Title of Thesis: Integrating high-throughput genotyping and lipidomic profiling for discovery of genetic determinants of cultivated sunflower seed oil content

Supervisor: Professor Philipp Khaitovich

Name of the Reviewer:

I confirm the absence of any conflict of interest Loren H. Rieseberg (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** Loren H. Rieseberg **Date:** 22-12-2020

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

This is an impressive dissertation, which reports on the generation and analysis of genotyping by sequencing (GBS) data for greater than 600 lines of the cultivated sunflower, along with morphological, developmental, and seed oil traits for the same set of lines. These data are then employed to assess the genetic and phenotypic diversity of Russian cultivated sunflower collections and compare them to international collections. In addition, genome-wide association (GWA) analyses are performed to assess the genetic basis of selected traits and identify candidate genes, focusing on agronomically important traits such as the restoration of cytoplasmic male sterility and various sunflower oil characteristics. The research reported not only represents a substantial effort – far beyond what is typically reported in a Ph.D. dissertation – but it also appears to be of very high quality.

The methods employed are appropriate and state of the art. However, I did have a couple of questions. Correcting for multiple tests is tricky with genomic data because there are many markers, which are not independent. Your approach of using LD blocks seems reasonable to me, but it was not clear to me how exactly you estimated them. That is, what thresholds did you use to call an LD block? Also, did you consider using the sunflower trait ontology tool to harmonize your phenotypic data (p. 39) with that generated by other groups (https://www.cropontology.org/terms/CO_359:ROOT/Sunflower%20traits)?

The results reported in this dissertation highly significant for several reasons. Most importantly (in my view), this represents the first comprehensive genomic analysis of Russian cultivated sunflower germplasm collections. Russian breeders were responsible for developing sunflower into an oil crop, so the germplasm held in Russian genebanks is valuable and unique – statements which are confirmed by the results reported here. Also important are the advances in high-throughput phenotyping of oil traits and subsequent GWA analyses, which are key to optimizing sunflower oil characteristics for different uses.

Although I am very impressed by this dissertation, I do have a few comments that might affect how the results are interpreted:

1) Unfortunately, the XRQ reference genome (which was employed for SNP calling) has a large number of mis-assemblies, stemming from the use of a faulty physical map during the assembly process. This can lead to false positives, especially if there are broad and strong GWA peaks. An example is the GWA peak for branching on chr. 10. (p. 54). The branching locus is known to be near the bottom of chr. 10, which is clearly seen in the GWA analysis. However, the additional hits scattered across the rest of the chromosomes are likely false positives. Todesco et al. (2020; Nature 584, 602-607) corrected for this by transferring SNPs to a new reference genome that is assembled correctly, so this is something you could consider as well (in the future!).

2) With respect to the restorer locus, you should mention that the locus is an introgression from *H. petiolaris*, which was characterized by Baute et al. (*New Phytologist* 2015; 206:830-838). Also, Owens et al. (Evolutionary Applications 2019; 12:54–65) reported on copy number variation in cultivated sunflower and found that one of your candidate genes (aldehyde dehydrogenase) shows a 10-fold increase in copy number in restorer relative to maintainer lines. Thus, it is our favored candidate for the restorer of fertility gene. The same paper reports on a PPR gene on chr. 8 that shows copy number differentiation between maintainer and restorer lines. Thus, it represents a candidate gene for the restorer locus on chr. 8.

The results from this research have important real-world applications. Russia is the largest grower of sunflowers, and the markers, QTLs, and phenotypic data reported here can be used to implement a modern molecular breeding program, as well as to more efficiently and precisely develop cultivars with an improved seed oil profile.

Lastly, I think the publications are of impressive and of high quality.

The issues I would like to see be addressed before/during the thesis defense include clarification of how LD blocks were estimated, discussion of the sunflower trait ontology in relation to your work, consideration of potential impacts of reference genome mis-assembly for interpretation of your results, and attention to the relevance of copy number variation for the candidate restorer loci.

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*