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	**Signature:**  **Date: 30-12-2020**

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

Globally, sunflower is an important hybrid crop, grown around the world for oilseed. It is a major crop in Russia, where it was first grown as an oilseed crop, yet for hybrid cultivars comes from abroad. Thus, there is big potential to increase the breeding capacity for sunflower in Russia. This thesis provides new insight of the genetic and phenotypic variation for oil composition available in Russian seed banks and breeding programs. Knowledge gained from this thesis will assist in the development of locally bred and adapted cultivars with improved oil quality.

Below I provide my comment and summary of the thesis chapters. Attached to my review is a PDF of the thesis with minor corrections/suggestions.

Firstly, the paragraph structuring could be revised and improved. This is an issue throughout the thesis and makes it difficult for the reader to follow, particularly when many very short paragraphs occur (i.e. paragraphs containing 1 or 2 sentences only). Remember, each paragraph should be a single idea. Each paragraph should be between 4-8 sentences long, start with an introductory sentence and end with a concluding sentence.

Chapter 1 provides a nice introduction to the thesis, setting the scene for the research to be undertaken. I like the subheadings and the supporting figure!

Chapter 2 provides a literature review of the topic, including sunflower domestication, cultivation, genetic resources, breeding approaches and opportunities, sunflower oil properties, molecular markers for breeding applications, QTL mapping techniques, and genomics-assisted breeding strategies. I was fascinated to learn that sunflower was first grown as an oil seed crop in Russia in the late 1800s! Overall, the candidate has provided a throughout review. The literature review could benefit with a final section that highlights the research gaps or priorities for research. This would help guide and introduce the reader to the topics/research questions to be studied in this thesis.

Chapter 3 reports the materials and methods used for the research conducted as part of this thesis. I must say that I am not familiar with this type of structure for a PhD thesis (normally, in Australia the research chapters each comprises an Introduction, Methods, Results and Discussion section), but I think this is just a style thing and the arrangement works for this thesis, so this is fine with me! I think the "samples" section is quite confusing because there is a mix of a description of sunflower accessions, a description of how the seeds were produced for analyses and also some mention of rapeseed accession… please consider how this section could be presented more clearly and perhaps divided into 2 or 3 smaller sections to help guide the reader? Overall, the genotyping, phenotyping and GWAS were conducted using robust methodology.

Chapter 4 reports new insight of the phenotypic and genetic diversity of diverse sunflower accessions, in particular the Russian accessions that were genotyped for the first time. It is critical for researchers and breeders to understand the trait diversity if it is to be harnessed for crop improvement. Highlights from this research include GWAS that detect marker-trait associations for linoleic acid content on chromosomes 8, 9 and 17, and narrowing down a 7.7 Mb region on chromosome 13 associated with fertility restoration (specifically Rf1). A number of candidate genes were identified for the chromosome 13 region associated with Rf1. Based on the haplotype of this segment, and the passport information for the accessions you have genotyped, could you determine the likely origin of the Rf1 gene? Further, it would be very interesting to report whether any Rf1 haplotypes are associated with agronomic traits… Such as association could be possible via pleiotropic effects or through LD as a result of long term

selection in breeding programs. Lastly, I think it would be good to report and discuss the relationship among the traits studied, which is something plant breeders would be interested in.

Chapter 5 reports the implementation of UPLC-MS technologies for FAs and TAGs profiling in sunflower and rapeseed. The addition of rapeseed to this chapter serves as a nice way to demonstrate the methodology can approach can be extended to other oilseed crops. Two UPLC-MS-based approaches were compared to gas chromatography-flame ionization detection, which is a time-consuming process with limited ability to fully characterise the different fatty acids. The results are clear and the figures are high quality. Notably, the results presented in this chapter have been published in the form of two high quality articles in international journals. Interestingly, differences were observed between spring and winter rapeseed varieties. The candidate makes a nice discussion to propose the different lipid composition is possibly related to freezing/cold tolerance traits. The candidate links their results well with the literature and discusses some limitations, including the comparisons between methods.

Chapter 6 uses the phenotypic data from chapter 5 to perform a GWAS to reveal the underlying genetic controls of oil fatty acid content. This is an impressive study, with more than 600 accessions, the scale was sufficient to detect a large number of QTLs for fatty acid composition traits. This represents significant progress for the sunflower research and breeding community, as these genomic regions can be further studied or introgressed into new sunflower cultivars to improve oil quality.

Should the “Conclusions and future perspectives” section at the end of the thesis be a numbered chapter, like the Introduction (chapter 1)?

Overall, this is a high quality thesis and significant scientific contribution. Pending the requested minor changes and edits, in my view, this thesis satisfies the requirements for a PhD in plant genetics. I congratulate the candidate on very successful research outcomes that will assist the sunflower breeding objectives for Russia and around the world.

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*

Skolkovo Institute of Science and Technology

INTEGRATING HIGH-THROUGHPUT GENOTYPING AND LIPIDOMIC PROFILING FOR DISCOVERY OF GENETIC DETERMINANTS OF CULTIVATED SUNFLOWER SEED OIL CONTENT

Doctoral Thesis

By

ALINA CHERNOVA
DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Philipp Khaitovich

Moscow – 2020

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgment is made and has not been submitted for any other degree.

Alina Chernova

Prof. Philipp Khaitovich

Abstract

Since its domestication in North America approximately 4000 years ago, sunflower became one of the most important oilseed crops in the world and, in particular, in Russia. During the last century, oil content in sunflower has been subjected to strong selection. Further advancement of these selected cultivars with improved oil properties is one of the fundamental directions in modern oilseed crop breeding. The proportion of different fatty acids in oil is key to determining oil quality. Sunflower has excellent potential for oil improvement, and in the frame of climate change, it can outperform other oilseed crops like oil palm or soybean. Russia has a long history of success in sunflower cultivation and breeding. Sunflower development into the important oilseed crop took place in Russia. Despite the rich history of sunflower research and its substantial economic significance, currently, Russia lags behind in sunflower breeding and has to buy 80% of seeds abroad. To be competitive in the modern breeding process aimed at sunflower oil customization and oil production with different properties, it is crucial to search for genetic markers associated with desirable plant and seed oil phenotypes and use these markers to design optimized breeding selection schemes. In the present work, we combined high-throughput genotyping GBS (genotyping-by-sequencing) technology with high-throughput molecular phenotyping UPLC-MS (ultra-high-performance liquid chromatography coupled with mass-spectrometry) technology to study genome and lipidome variability of 601 diverse cultivated sunflower accessions from three Russian sunflower collections: Vavilov seed bank, VNIIMK Applied Agricultural Institute, and Agroplasma Breeding Company. Based on these data, we performed genome-wide association studies and identified several novel molecular markers and candidate genes that can be used to accelerate the breeding of sunflower plants with desired oil fatty acid composition. In addition to oil composition, we analyzed a set of classical phenotypic data for the genotyped accessions and identified genetic markers for several traits. We further compared the genetic variability present in Russian collections with that of international wild and cultivated accessions and suggested optimized methods for sunflower oil lipidome characterization to assess phenotypic variability of its fatty acid and lipid composition.

The presented work is the first large-scale study to use such an extensive sunflower germplasm collection, which adds to the existing knowledge of sunflower genetics and metabolism and provides new opportunities for marker-assisted (MAS) and genomic selection (GS) of oilseed crops.

Publications

.

1. Goryunova SV, Goryunov DV, **Chernova AI**, Martynova EU, Dmitriev AE, Boldyrev SV, et al. Genetic and Phenotypic Diversity of the Sunflower Collection of the Pustovoit All-Russia Research Institute of Oil Crops (VNIIMK). Helia. 2019 Jul 26;42(70):45–60.

2. Goryunov DV, Anisimova IN, Gavrilova VA, **Chernova AI**, Sotnikova EA, Martynova EU, et al. Association Mapping of Fertility Restorer Gene for CMS PET1 in Sunflower. Agronomy. 2019 Feb;9(2):49.

3. **Chernova AI**, Martynova EU. High-throughput technologies for sunflower oil improvement. In Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences 2019. p. 227–229.

4. **Chernova A**, Mazin P, Goryunova S, Goryunov D, Demurin Y, Gorlova L, et al. Ultra-performance liquid chromatography-mass spectrometry for precise fatty acid profiling of oilseed crops. PeerJ. 2019 Mar 6;7:e6547.

5. **Chernova A**, Gubaev R, Mazin P, Goryunova S, Demurin Y, Gorlova L, et al. UPLC⁻MS Triglyceride Profiling in Sunflower and Rapeseed Seeds. Biomolecules. 2018 27;9(1).

6. **Chernova A**, Gubaev R, et al. Genotyping and lipid profiling of 601 cultivated sunflower lines reveals novel genetic determinants of oil fatty acid content – under review

Patent: RU2717642.

Goryunova SV, Goryunov DV, Mazin PV, **Chernova AI,** et al. 2020. Snp-panel for genotyping and genomic selection of sunflower by content of fatty acids in seed oil.

Registered database RU 2019620084.

Pavlova AS, Gurchenko EA, Petrova DA, Khaitovich PE, Goryunova SV, **Chernova AI**, Boldyrev SV., 2019 Oilseed crops phenotypes database.

Acknowledgments

This work would have not been possible without the support of many people and organizations. First of all, I want to say the word of thanks to my advisor Philipp Khaitovich who granted me a lot of freedom and a high level of trust to develop this project and opportunities in collaborations and knowledge exchange in different places. Second, I want to extend my sincere thanks to my unofficial advisor Sergey Nuzhdin from the University of Southern California who shared with me plenty of exciting ideas and all the time was open for advice and support; he helped me to find my role in the project and the way it has to go; my committee members Georgii Bazikin and Petr Sergiev who always came with helpful comments during the review process every year; great members of my lab Pavel Mazin, Rim Gubaev, Elena Martynova, and Stepan Boldyrev, for taking a big part in the project, fruitful discussions, and real friend support. Svetlana Goryunova was an ideologist of the sunflower project in Skoltech. Waltraud Mair, Anna Vanushkina, Nikolay Anikanov, and Elena Stekolshikova helped a lot with the molecular phenotyping part; My collaborators from USC: Peter Chang, Katrina Sherbina, and Anupam Singh for their great help in data analysis and support; Wendy Wu for teaching me NGS library preparation procedures; Yulia Karabitsina from VIR made a significant contribution to the genotyping of VIR lines. The sunflower samples used in the study would never be available without three organizations and kindness of the people working there: VIR and Vera Gavrilova and Irina Anisimova who selflessly keep an eye on sunflower germplasm collection for many years, VNIIMK and Yakov Demurin, the holder of sunflower collection there who was always open to our ideas and initiatives, Agroplasma breeding company and Nikolay Benko and Ghanna Mukhina who we can name pioneers in implementing high technologies in sunflower breeding in Russia. I want to extend my gratitude to Liesbeth Veenhov and Anton Steen from the ERIBA center in Groningen (The Netherlands) my time in Skoltech started in Liesbeth's lab, and work with these people definitely convinced me to proceed with science and join a Ph.D. program. I want to acknowledge Prof. Loren Rieseberg and the members of his group Jean-Sébastien Légaré and Gregory Owens, who have kindly shared the data with us.
I want to mention Skoltech professors Laurent Gentzbittel and Dmitry Kulish, and all the people I met at the conferences: particularly Lee Hickey and Felipe Ferrao, discussions with these people were handy for my research and helped to reduce the level of uncertainty a lot.

Of course, I want to mention Skoltech and its Ph.D. program, which gave great opportunities to develop yourself as a scientist and to perform top-level research in Russia.

The support of my friends and family gave me confidence and a useful sense of calm. My parents, Nataliya and Igor, and my husband Sergey always support me and spend a lot of time with my son helping me to find time for work. My best friends Anastasia Latanova, Yuri Bykov (for him personal thank for the beautiful picture in my introduction section), Ekaterina Rozhavskaya, Ilya Fedorov, Mikhail Pogorely, Anastasia Troshina, and Yulia Muzaleva were near all the time and gave me unbelievable scientific and general support.

Table of content

List of Symbols, Abbreviations

GBS – Genotyping-by-sequencing

UPLC -MS – Ultra-performance liquid chromatography coupled with Mass-Spectrometry

VIR – N.I.Vavilov Research Institute of Plant Industry

VNIIMK – Pustovoit All-Russia Research Institute of Oil Crops

SNP – Single nucleotide polymorphism

MAS – Marker-assisted selection

GS – Genomic selection

NGS – Next generation sequencing

USC – University of Southern California

USDA – United States Department of Agriculture

CMS – Cytoplasmic male sterility

Rf – Fertility restoration

GEI – Genotype-by-environment interactions

TAG – Triacylglyceride

DAG – Diacylglyceride

FA – Fatty acid

PEP – Phosphoenolpyruvate

TP – triose phosphate

G3P – Glycerol-3-phosphate

G6P – Glycerol- 6-phosphate

ACP – Acyl-carrier-protein

FAS – Fatty acid synthase

ER – Endoplasmic reticulum

PDAT – Phospholipid diacylglycerol acyltransferase

VLCFA –Very long-chain fatty acids

LPA – Lysophosphatidic acid

PC – Phosphatidylcholine

PE– Phosphatidylethanolamine

PI – Phosphatidylinositol

PA – Phosphatidic acid

SPE- HPLC – Solid-phase extraction high-performance liquid chromatography

QTL – Quantitative trait locus

mQTL – Metabolomic QTL

PCR – Polymerase chain reaction

RFLP – Restriction fragment length polymorphisms

RAPD – Randomly amplified polymorphic DNA

AFPL – Amplified fragment length polymorphism

SSR – Simple sequence repeats

LD – Linkage disequilibrium

GWAS – Genome-wide association study

mGWAS – Metabolomic genome-wide association study

bp – base pair

Mb – Megabase

Gb - Gigabase

MLM – Mixed linear model

NMR – Nuclear magnetic resonance

GC – Gas chromatography

LC – Liquid chromatography

MS – Mass spectrometry

QTOF – Quadrupole time-of-flight

DF – Disk florets

RF – Ray florets

DTB – Planting-budding period

DTF – Planting-flowering period

DTM – Planting-physiological maturity period

HU – Heat units

PCA – Principal component analysis

PRP – Pentatricopeptide repeat

TRP – Tetratricopeptide repeat

"…Nothing, but large sunflowers"

Vincent Van Gogh

Chapter 1. Introduction

1.1 Significance of the work: Power of genomics help to face global challenges

In the twenty-first century, humanity faces significant challenges, requiring the improved means to address and overcome them. The human population is growing, with estimates of 10 billion people in 2050, which place global food security at risk. Even now, hunger is the reason number one of deaths in developing countries. More than 820 million people suffer from hunger worldwide (1). Climate change also a factor contributing to food availability. Currently, 40% of ice-free land is used for food production (2). Global warming makes new regions available for cultivation, while other areas may become unsuitable for planting crops. In light of these developments, it becomes crucially important to develop technologies for accelerated selection, which will help adapt existing cultivated species to changing environments. To achieve this goal, each species' existing genetic and phenotypic variability can be used as a source of genetic variation by plant breeders. Plant genomes became the principal value and currency, which can help to overcome food security issues and help to feed 10 billion people in the changing environment. This idea is going back to Nikolay Vavilov, the great scientist who was the first one who recognized the importance of studying and protecting genetic resources. Now we are on that step of technological progress where we can use all the power of genomics to improve crop species and increase the effectiveness of food production.

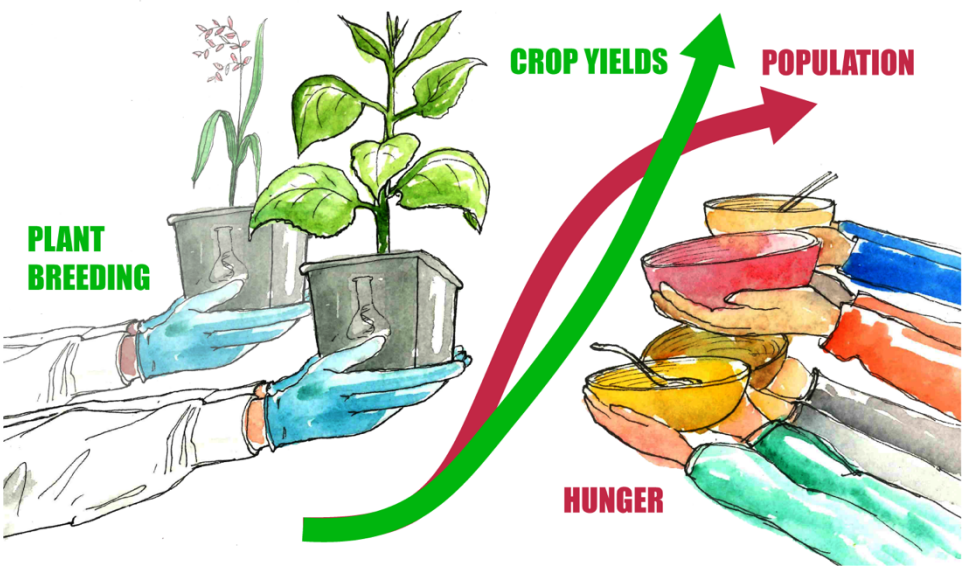


Figure designed by Yuri Bykov

1.2 Sunflower - a new “Arabidopsis” in the world of cultivated plants

The most well-studied plant is the model plant *Arabidopsis thaliana.* With a small, fully annotated genome, this plant helped to get significant insight into plant metabolism, genome organization, and physiology (3). Similarly, common sunflower (*Helianthus annuus*) serves as a model plant for domestication, evolution, and climate change studies (4). Evolution biologists have focused on the *Helianthus* species for a long time identifying the origins of domestication and tracing introgressions and genome reorganizations on the way for domestication (5–7). Wild sunflower species are interesting for understanding evolution processes, in general, exploring gene flows and adaptations (8, 9). In-depth knowledge of sunflower evolution owed to the work of Loren Rieseberg and his colleagues, sunflower genome sequence availability (10), and sunflower’s ability to grow under various conditions (11) stimulate sunflower use in modeling different environmental stresses and understanding the mechanisms underlying adaptations (12). Now, sunflower is a global oilseed crop, taking position four among the most cultivated oilseed crops, with 26.48 million hectares being occupied by it worldwide (13).

High genetic diversity captured in 52 species of *Helianthus* across North America through the wide variety of habitats has excellent potential for future crop improvement (14). It gives sunflower potential to become the oilseed crop of preference in future and a model crop for understanding adaptation to climate changes (11).

1.3 Plant oil – an essential component for human nutrition and industry

The role that is allocated to fatty acids has dramatically expanded during the last years (15). Plant oil is generally rich for mono- and polyunsaturated fatty acids (omega-3 and omega-6), which are considered good for the cardiovascular system and suggested as a healthy alternative for solid fats in the daily ration (16). Omega-3 fatty acids were also recognized as a nutrient that reduces the risk of type 2 diabetes (17) and as an essential element for mental health and fetal development (18,19). For omega-6 fatty acids, anti-cancer activity was reported (20). Improved plants can serve as an alternative source of polyunsaturated fatty acids, which can decrease the burden on aquaculture – the primary source of these acids at the present time (21). Due to the difference in saturation and carbon chain length, fatty acids exhibit different chemical and physical properties. These properties make fatty acids useful for the production of soaps, plastics, nylon, lubricants, paints, coatings, and adhesives (21). A significant vegetable oil application area is biofuel, where it can serve as an alternative to non-renewable energy sources (22). Advances in plant genetics and plant biotechnology will allow designing oilseed plants in order to target the previously impossible goals of both nutrition and industry (21). Novel experimental and data-analysis tools for detailed and systematic oil composition analyses were extensively developed during the past ten years, providing means to accurately detect and quantify major oil components and find minor components and get new insights into plant metabolic processes.

1.4 Aims of the study

My research focused on applying high-throughput genotyping and high-throughput oil lipidome phenotyping technologies to explore genotype-phenotype relationships in 601 accessions from three Russian sunflower germplasm collections (Vavilov seed bank, VNIIMK Applied Agricultural Institute, and Agroplasma Breeding Company). Specifically, in my work, I aimed to characterize germplasm genotypes at the whole-genome level, obtain lipid profiles of genotyped lines, and use them to search for genotype-phenotype associations in order to identify candidate genetic markers of phenotypic traits and candidate genes involved in phenotype expression. My project is a part of a larger initiative aimed at finding genetic determinants associated with agronomically important sunflower traits and developing the approaches for accelerated sunflower breeding, with a big goal to ensure Russian oilseed crop breeding independence and food security.

Chapter 2. Literature Review

2.1 History of sunflower domestication

2.1.1 From wild to cultivated

Sunflower is an important oilseed crop domesticated from the wild plant in North America approximately 4000 years ago. In the *Helianthus* genus, there are 53 species found in different ecological zones from the Atlantic coast to the Pacific coast and from Canada to Mexico. Like all cultivated plants, sunflower went through the domestication bottleneck in the course of domestication. However, its wild relatives still store vast genetic diversity, which could be used for crop improvement (23).

Cultivated sunflower originated from the wild-growing representatives of the annual species *Helianthus annuus* L. Sunflower seeds were used as food by Native Americans. According to archeological records, also in medicine and ritual ceremonies (24). Research suggests that the sunflower was domesticated in the 13th century by native tribes long before it was introduced to Europe (25, 26). According to P.M. Zhukovsky, sunflower seeds were brought to Spain from New Mexico in 1510 and placed in a botanic garden (27). In Europe, sunflower first becomes widespread as an ornamental plant. In the 18th century, the seeds started to be used for food and medicine—also, the first record of using sunflower seeds for oil is dated by 1716 (25).

2.1.2 Russia as the origin of oilseed sunflower

The sunflower was introduced in Russia by Peter the Great in the 18th century. The cultivation of sunflower as an oilseed crop started in Russia in the 19th century (28) after a peasant named Bokarev performed his "industrial oil extraction" in 1861 in the Voronezh province. The first sunflower varieties Maslyanka (k-619), Uspenka (k-768), and others contained about 28-33% of oil in the seeds (29). The success of sunflower in Russia may be connected with the academician V.S. Pustovoit because of his pioneering works. The VNIIMK institute in Krasnodar remained the main sunflower breeding center in the world till the end of 1960th. Pustovoit and his followers created the first high-oil varieties, with up to 51% of oil in the Peredovik variety (30).

The Soviet Union sunflower breeding program led by Pustovoit also reached excellent results in breeding for broomrape resistance and high oleic acid sunflower (31). Sunflower open-pollinated lines created in USSR were spread into other parts of Europe and formed the basis of

sunflower breeding worldwide (32). The discovery of cytoplasmic male sterility (CMS) (33) and later of the sources of fertility restoration (Rf) (34) was a huge breakthrough that predetermined sunflower cultivation as a hybrid crop, and it spread worldwide (35).

2.2 Sunflower cultivation and position on the global market

According to the USDA report, the total area of sunflower (*Helianthus annuus* L.) cultivation worldwide in 2020 is estimated to be 26.48 million hectares, with the entire seed production of about 55.25 million metric tons. Russia is the leading sunflower producer with 8.36 million hectares area occupied by sunflower (36). More than 27% of total sunflower production is provided by Russia (37). Sunflower is mostly grown for seed oil. Non-oil (confectionary) sunflower part varies between countries but globally accounts for about 10% of total sunflower production (38).

Based on oil production worldwide, sunflower ranks fourth after palm, soybean, and rapeseed with 20.95 million metric tons of oil produced, which account for about 10% of total plant oil production globally (37). According to the forecast, the oilseeds market is growing and will reach more than 255.2 million dollars by 2023 (39).

2.3 Sunflower genetic resources

Gene pool, which can be categorized into primary, secondary, and tertiary gene pools (40), serves the basis for selection and crop improvement (41). Conservation of genetic resources ensures global food security (42). For sunflower, the primary gene pool is presented by local landraces, open-pollinated cultivars, hybrids, and inbred lines. The secondary gene pool is most other *Helianthus* species, which make a significant contribution to diversity. Sunflower germplasm resources can also be categorized as in situ resources (wild populations and landraces) or ex-situ resources (accessions preserved in seed banks) (43).

The main advantage of in situ resources is their existence in natural conditions where they continue to evolve in response to the environment. But there is a risk of encroachment by human activities. Ex-situ conservation is essential because, in this case, genetic diversity is securely preserved and protected from environmental effects. Sunflower gene banks are maintained in different countries around the globe, and they all have different repertoires of accessions, and genetic variation is kept among them (23).

The world's biggest sunflower germplasm collection is the USDA sunflower gene bank (part of the US National Plant Germplasm System (NPGS)), which is extremely rich in both cultivated (2616) and wild accessions. Among wild accessions are 1057 representing *H. annuus* species,

and more than 1000 belong to other species in the *Helianthus* genus (https://npgsweb.ars-grin.gov/gringlobal/search.aspx).

The sunflower gene bank (Centre de Ressources Biologics [CRB]) at INRAE, Toulouse (https://www6.toulouse. inra.fr/lipm/Recherche/Genetique-et-Genomique-du-Tourne sol/CRB-Tournesol) holds 2300 cultivated sunflower accessions and around 800 wild ones. Big sunflower collection of cultivated and wild lines is maintained in Serbia in Novi Sad (https://www.ifvcns.rs/kolekcija-divljeg-suncokreta/wild.html). A lot of research is performed there to introduce wild sunflower into the programs of sunflower breeding and improvement of this crop performance. The big focus is on disease resistance (23).

The oldest sunflower bank is located in Russia (Vavilov Research Institute for Plant Genetic Resources (http://www.vir.nw.ru). The first entrees there date back to 1922. The sunflower collection at VIR totals 2730 accessions – 2288 of cultivated (*Helianthus annuus*) and 442 of wild sunflowers belonging to 24 species. Several more gene banks located in Bulgaria, India, Argentina, Spain, Germany, and Romania contribute to sunflower genetic diversity.

The second sizeable Russian sunflower collection is located in VNIIMK (44, 45), holding lines contrasting in oil content and composition.

All the mentioned above genetic resources provide sources of disease resistance, oil content, different maturities, and other agronomic characters for modern sunflower breeding programs (23).

2.4 Sunflower breeding

2.4.1 Peculiarity of sunflower breeding

After maize, sunflower is the second world crop in terms of hybrid breeding (46). It is possible to determine three stages in sunflower breeding history: mass selection, individual selection for various production traits, and hybrid breeding. The mass selection was performed in multiple locations and was the main force of sunflower domestication. Mass selection contributed to the development of local sunflower varieties (landraces) conserved in sunflower banks and used in current breeding programs. Some pest resistances are among the achievements of sunflower mass selection (47). Pustovoit developed the individual selection technique with seed reserves in 1920th (48). Individual selection made possible the development of sunflower high oil content varieties and made it one of the most important world oilseed crops globally cultivated for seed oil from the 1960th.

Until 1970th, open-pollinated sunflower varieties were cultivated, but then there was a switch to hybrid breeding, which gave more power to breeding through heterosis (32). The discovery of

CMS (49) and Rf genes (50) was another factor that stimulated the shift to hybrid breeding. It is essential to mention that the idea of sunflower hybrids dates back to the 1940th (51). The heterosis phenomenon helped increase the yield up to 60%. Still, the use of the CMS system made it possible to introduce it on a commercial basis (52), which led to the production of the first commercial sunflower hybrids. Hybrids' power is in heterosis, which is a result of high heterozygosity achieved when one crosses genetically divergent self-pollinated inbred lines. The studies on sunflower inbreeding were pioneered by Corden, who created the first inbreed sunflower line by self-pollinating (selfing) the Mammoth Russian variety (53). Now selfing of sunflower is a standard procedure to obtain inbred lines. For example, lines conserved in VIR seed bank undergo 8-25 generations of inbreeding (54).

2.4.2 Directions in sunflower breeding

Sunflower stays under the selection as an oilseed crop for more than 100 years. Since sunflower is cultivated worldwide, there are many characteristics that breeders control to get desirable phenotypes adapted to a wide range of climate conditions. Also, the directions in sunflower breeding have changed due to new breeding technologies described above. Currently, breeders mostly focus on increased oil and seed yield (amount of seeds), oil properties and oil composition, resistances to common sunflower pests and pathogens like broomrape or downy mildew, and tolerances to unfavorable environmental conditions and herbicides. Ornamental sunflower is also on-demand. (47). All traits require different approaches for their improvement since they all have different heritability. Some characters are inherited as a monogenic trait and have high heritability; these traits include CMS, Rf, pathogen resistance, and herbicide resistance.

Others are inherited as multigenic traits with environment affecting them largely, including quantitative traits like flowering time, oil yield, and oil composition (35).

Another significant factor is genotype-by-environment (GEI) interactions, which also have to be considered in breeding. This is a complex phenomenon that challenges breeding for quantitative traits (55). Oil content and composition are very dependent on the environment. GEI for such characteristics are actively studied for many crops like soybean (56), olive (57), peanut (58), rapeseed (59), and sunflower (60, 61). GEI and their cases should be studied individually on a cultivar-by-cultivar basis, as a deep understanding of GEI can be implemented in breeding to maximize genetic gains (62).

2.4.3 Future of sunflower breeding

In the future, more sunflower genomes will become available. Sequencing of wild species will help better understand the process of sunflower domestication; it will also give more insights into sunflower metabolic pathways (63).

A more detailed study of traits in wild and cultivated sunflowers will take place; this will give new opportunities to introgression and manipulating specific characteristics (64).

Marker-assisted selection (MAS), genomic selection (GS), and high-throughput technologies, which will be discussed later in this chapter, will extend their presence in breeding strategies (35).

Genome-editing tools like CRISPR-Cas technology can be used in breeding for monogenic traits like Broomrape resistance (65).

It is important to remember that proper understanding of germplasm genetic and phenotypic diversity gives a key for the successful implementation of new technologies into crop improvement.

2.5 Sunflower oil

Sunflower is mostly grown for seed oil (38). With 20.95 million metric tons of oil produced per year, sunflower takes position number four after palm, soybean, and rapeseed oil (13). Most of the oil is composed of triacylglycerides (TAGs) containing three fatty acid (FA) residues. The oil's physical and chemical characteristics mainly depend on the fatty acid residue composition of triacylglycerides and free FAs content, which in turn determine the final use of the oil. For example, the number of double bonds in fatty acids and fatty acid chain length affects thermo-oxidative oil stability. Nutritional properties, state (solid or liquid fats) is also dependent on the ratio between saturated and unsaturated fatty acids (free or captured in triacylglycerides) (66).

2.5.1 Oil chemistry and biosynthesis

The deposition of storage lipids in sunflower seeds is fully dependent on sucrose metabolism (67). Most of the carbon accumulated in seed oil comes from photosynthetic tissues during the period before flowering (68). The initiation of FA biosynthesis occurs in plastids, so metabolic exchanges between cytosolic and plastidial glycolytic pathways occur. Most of the carbon converted in the cytosol into phosphoenolpyruvate (PEP), is then transported to plastids, where it is converted into the pyruvate precursor of acetyl-CoA. Triose phosphate (TP) and glycerol-3-

phosphate (G6P) also can be provided by plastid metabolism (67). Acetyl-CoA is a key fatty acid precursor. There are three main pathways of how it can be produced: from citrate, from pyruvate directly, and from pyruvate through acetaldehyde and acetate (69). Then, in short, Acetyl-CoA is metabolized by Acetyl-CoA carboxylases to malonyl-CoA (70), which is exposed to malonyl-CoA: ACP transacylase. The product of this reaction, malonyl-ACP (acyl carrier protein), is used by FA synthases (FAS) for FA synthesis. FAS is a complex of six enzymes and ACPs responsible for the FA elongation cycle (71). Most of the sunflower FAS genes have been isolated and cloned (72, 73). The final product of FAS activity is generally palmitic (16:0) and stearic acids (18:0) (71). Sunflower oil is enriched by unsaturated FAs (16:1, 18:1, and 18:2), desaturase enzymes perform unsaturation. 16:1 and 18:1 fatty acids are produced directly in plastid stroma by Δ9 stearoyl-ACP desaturase. This enzyme is well characterized and sequenced and cloned for the many plant species, including sunflower (74).

Additionally, desaturase is present in other cell compartments. The fatty acid most present in traditional sunflower varieties is linoleic acid; it is produced in the ER in the reaction catalyzed by the oleate desaturase, FAD2 (71). In sunflower 3 FAD2 (FA 2-1, FAD2-3, FAD2-3) genes have been identified (75)

Acyl-ACP thioesterases release free FAs. There are two different enzymes in plants: *FatA* and *FatB*. Genes for both are identified in sunflower (76, 77). Long-chain acyl-CoA synthetases quickly convert free fatty acids in developing seeds to their corresponding acyl-CoAs (78). These acyl-CoAs are critical players in further lipid biosynthesis. Genes coding for long-chain acyl-CoA synthetases are described in sunflower (79).

Very long-chain fatty acids (VLCFAs) like 20:0, 22:0 are synthesized in ER using acyl-CoAs and malonyl-CoA by the FA elongase complex (80). In sunflower, VLCFAs are synthesized from 18:0-CoA by these membrane-bound fatty acid elongases. KCS-I and KCS-II enzymes of this type have been identified in sunflower (81).

The most abundant seed lipids, the triglycerides, are synthesized in ER using acyl-CoAs from plastids. (82). Multiple interconnected pathways for TAGs synthesis are described in plants (67). In the conventional, the so-called Kenedy pathway, glycerol-3-phosphate (G3P), serves as a carbon backbone. G3P is acetylated to form lysophosphatidic acid (LPA) through the action of sn-glycerol-3-phosphate acyltransferase. Then lysophosphatidic acid acyltransferase performs second acyl-CoA dependent acylation, and phosphatidic acid (PA) is formed, releasing phosphate to create DAG. DAGs are converted to TAGs by diacylglycerol acyltransferase (83,84).

Classical mechanisms of TAG formation in sunflower were widely studied in biochemical works (85, 86). The most important genes participating in these reactions are also identified.

An alternative mechanism of TAG formation is acyl-CoA independent. In this case, phospholipid diacylglycerol acyltransferase (PDAT) transfers an acyl group from the *sn*-2 position of phospholipids to *sn*-3 position of diacylglycerol, yielding TAG, and *sn*-1-lysophospholipid (87). It was shown that the first classical pathway is more common in sunflowers during seed development (88).

2.5.2 Oil Composition

The seed oil of standard cultivated sunflower is characterized by a high amount of linoleic acid (C18: 2) and a medium amount of oleic acid (C18:1). These two FAs account for about 88% of the total oil fatty acid content. Saturated fatty acids like palmitic (C16:0) and stearic (C18:0) comprise about 11%. Just about 1% of total fatty acids fraction is represented by FAs with fewer than 16 or more than 18 C atoms (89). Sunflower oil contains three main saturated VLCFAs: arachidic (20:0), behenic (22:0), and smaller amounts of lignoceric (24:0) (90).

Linoleic acid in classical concentration is very dependent on the environment (91), while in high oleic acid sunflower, the oleic acid concentration is independent of the environment (92). Soladatov first obtained the high oleic acid line; it was produced by treating Peredovic variety with a chemical mutagen and introducing the FAD2-1 gene's duplication leading to silencing and decrease of FAD2-1 expression (93). Later the same result was obtained by the changes in the FAD-1 sequence (94). High oleic acid sunflower contains up to 90% of oleic acid in its oil FA composition (89).

Fatty acids in sunflower oil mostly stay inside TAGs. The distribution across TAGs is not random. The most common saturated FAs like palmitic and stearic are mainly bound to positions sn-1 and sn-3 in TAGs (95). Unsaturated: oleic and primarily linoleic are distributed randomly between free positions, while linoleic prefer sn-2 (96). The majority of TAGs in sunflowers have unsaturated FAs in all three positions (97).

Among other possible components, it's essential to mention phospholipids, waxes, sterols, and tocopherols. Phospholipid fraction mostly consists of phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), and phosphatidic acid (PA), and account for less than 1.2% of the total oil (98). The quantitative assessment of phospholipids in sunflower oil was performed using SPE- HPLC technology (99). Waxes account for less than 1%. Sterols are minor components (100).

Other minor oil ingredients, such as tocopherols or vitamin E, are essential oil components that serve as antioxidants and contribute to oil stability (101). Tocopherols include alpha-, beta-, gamma- and delta-homologs. Tocopherol content varies between sunflower lines (102,103).

2.5.3 Applications

Sunflower oil applications lie on a scale between highly saturated and highly unsaturated oil. Talking about the food industry: standard sunflower oil with high polyunsaturated acid content serves as a healthy salad dresser. For frying and deep frying, the thermostable high oleic acid oil is preferable (66).

Non-food applications are mostly focused on biodiesel, biolubricants, cosmetics, and transformer oils (104). For these purposes, standard sunflower oil with the dominance of linoleic acid is not stable enough. High saturated acid content makes the oil performance on low temperatures poor, so high-oleic acid oil is the best choice. Non-food applications of sunflower oil could be extended to pesticide carrier and in the production of paints, soaps, and detergents, varnishes, agrichemicals, surfactants, adhesives, plastics, fabric softeners, and coatings (66)

2.5.4 Breeding for high oil yield and precise oil composition

Sunflower oil yield and oil content stay under the selection pressure since sunflower started to be cultivated as an oilseed crop. On average modern hybrids' seeds contain about 50% of oil; the maximum that can be achieved is around 60%, but this will affect other valuable compounds like proteins (32). Breeding for precise oil composition was mostly focused on oleic and linoleic fatty acid concentrations. The first sunflower variety with high oleic acid content was called Pervenets and was developed using chemical mutagenesis (105). This cultivar became the source of high oleic acid trait for breeding worldwide (35). Later different genetic markers and Quantitative trait loci (QTLs) associated with high oleic acid were identified in several studies (106–110). Now sunflower producers start thinking about different sunflower oil variations and ways to control other fatty acids like palmitic and stearic. High-stearic-high-oleic sunflower oil with low palmitic acid is a potential healthy alternative for palm oil (111).

Breeding affecting the concentrations of minor constituents using genetics can be beneficial in future oil improvement. For now, sufficient analysis was performed just for phytosterols (112). Lipid profiles and profiles of other minor oil components are very important for oil quality and its applications, and they will extend their presence in breeding programs (113).

2.6 Markers in sunflower breeding

Markers are one of the primary instruments in modern plant breeding since they can be used as "signs" to facilitate the choice of plants for a particular selection iteration. Markers can be divided into three groups: morphological, biochemical, and DNA markers (114). Morphological markers are the most intuitive and straightforward; the correlation between certain traits was noted by breeders long before DNA information started to be used in agriculture. By the way, just a few sunflower morphological markers were supported by the mapping of linked genes. For example, Chikkadevaiah and Nandini, 2003 reported the linkage between the nuclear male sterility gene and anthocyanin color gene (115). Under the biochemical markers, isozymes are usually considered – the enzymes that vary in amino acid sequence. The differences are generally detected by electrophoresis and special staining. Several studies were devoted to identifying polymorphic enzyme loci in sunflower (116–118). Morphological and biochemical markers are not the best for plant breeding since they are very limited in number and depend on environmental factors and developmental stages (119).

Many agriculturally important traits are complex and are regulated by multiple genes. DNA markers are a useful tool that helps to understand the genetic basis of agronomically important qualities and to learn how to control them. They are used to construct linkage maps and identify the genome regions responsible for the trait in the case of complex traits – QTLs (114). Since many agriculturally important traits like yield, plant height, oil, and protein content are complex and regulated by multiple genes, QTL mapping and estimation of the effect of each QTL on the phenotype is in high demand (120). QTL maps have been produced for many crop species, including sunflower. The next step for the implementation of QTL knowledge in agricultural practice is the identification of DNA markers for important QTLs and the use of them in breeding as tags for specific genotypes, which has to be selected to succeed in desirable phenotype. This approach is called marker-assisted selection (MAS) (121), and it allows us to simultaneously manipulate several QTLs. DNA markers are very abundant, and this is a huge advantage when we talk about multiple QTLs. Molecular or DNA markers capture the polymorphism within DNA sequences between different individuals. These differences may affect one nucleotide or longer fragments. It can be insertion, deletion, duplication, translocation, or just point mutations (122). DNA markers can be classified by mode of transmission (biparental nuclear inheritance, maternal nuclear inheritance, maternal organelle inheritance, or paternal organelle inheritance), way of gene action (dominant or codominant markers), and method of detection (hybridization-based or PCR-based markers) (123). Among the markers, it is essential to mention the most widely used markers in plants. First markers were detected using

hybridization; the most used among them are RFLP (restriction fragment length polymorphisms); they are based on restriction enzymes that produce different fragment patterns in different individuals (123). They began to be used for mapping in plants in 1980th (124). In sunflower, the first map based on RFLP markers was constructed in 1992 (125).

Among PCR-based markers, the most known are RAPD (randomly amplified polymorphic DNA), AFPL (Amplified fragment length polymorphism), SSR (simple sequence repeats), and SNPs (126). RAPD is the first genotyping technology based on PCR (127, 128). It was widely used for genotyping since it didn't require knowledge of DNA sequence, and the full procedure was fast and straightforward compared to RFLP, where radioactive isotopes were widely used (129). The biggest problem with RAPD technology was low reproducibility. To overcome it, AFPL markers were developed (130). This technology integrated PCR reaction using restriction enzymes to cut initial DNA and makes the genotyping more reproducible and robust in terms of detection of variation (122). This technique was successfully used for differentiation between closely related species and subspecies (131)

In sunflower, AFPL markers were used for germplasm fingerprinting (132) and QTL mapping, for example, for downy mildew resistance (133).

SSR or microsatellite markers are based on the amplification of repetitive sequences. In plants, they were first identified in soybean (134). They rapidly replaced RFPL markers in sunflower and became widely used for mapping (135–137).

All markers described above are available in a couple of dozens, but SNP markers can be produced in thousands and millions.

SNP is the smallest unit of mutation and inheritance, which became probably the most powerful tool used to capture the differences between the individuals and study the association between the genotype and phenotype. Compared to all other genetic markers, SNPs can provide the highest number of markers (126). In plants, there is about one SNP per 100-300 bp (138). Sunflower genome is highly polymorphic. One SNP per less than 20 bp has been reported (139). Fusari et al., 2008 reported 1 SNP per 69 bp in sunflower inbred lines (140).

It is essential to mention that markers in crops, particularly in sunflower, can be used for many different purposes. The most obvious are germplasm fingerprinting, resistance screening, gene mapping, mapping for other important traits, and evolutionary studies in wild species (141).

The abundance of SNPs in sunflower and the development of high-throughput NGS technologies, which allow the detection of millions of SNPs per individual, push them to become ideal markers in plant breeding that can be used for map construction, QTL-mapping, GWAS, MAS, and GS (43).

2.7 Sunflower genome and NGS in sunflower studies

2.7.1 Sunflower genome

Sunflower genome is large and comparable with the human genome (3.6 Gb, 17 chromosomes in the haploid set). The first genome sequence became available in 2017 (10). Like many other plant genomes, it is very rich in repetitive sequences (142), which slows down the assembly (143). The first sunflower sequenced was the XRQ inbred line. It was performed using the long-read PacBio technology. The second genome was obtained from the HA412 line using short reads and the software developed by NRGene (144). The third available genome is the fertility restorer line PSC8. Besides, the pan-genome based on 287 cultivated lines is available (145). NGS tools made the sequences of crop genomes available, but also had a high impact on genome functional annotation and linking phenotypes with genotypes with a view of further crop improvement (146).

2.7.2 NGS based genotyping technologies in sunflower studies

Apart from full sequences, single nucleotide polymorphisms, SNPs, are very useful for traits mapping, association studies, and germplasm characterization. NGS-based technologies are able to collect millions of SNPs within one sample (147).
Among multiple genotyping techniques for sunflower, the most widely used are microarray technologies and genotyping-by-sequencing techniques. Among the arrays, AXIOM chip (up to 62K) (148–150) and Illumina Infinium iSelect array (25K) (151) were used for sunflower genotyping. Compared to arrays that allow analyzing just the pre-selected set of genetic positions GBS approach is easily scaled to deal with genetic diversity, because it didn't rely on fixed nucleotide positions. GBS involves DNA processing by restriction enzymes to reduce genome complexity and get rid of repetitive sequences highly present in plants. (152). In sunflower studies, GBS helped to study wild sunflower genetic diversity and gene flow (153), introgressions from wild to cultivated sunflower (154), construct linkage maps (155), make genomic predictions (156), and perform QTL mapping (157).

2.8. Link the genotype and phenotype

There currently exist two approaches to understand the genetic basis of quantitative traits and, hence, to control these traits in breeding. The first approach is QTL mapping that relies on the so-called experimental crossings (F2, BC) with subsequent analysis of genotypes and phenotypes

of the progeny. The second approach is called association mapping that relies on mining the genetic and phenotype diversity in vast collections of breeding material (126).

2.8.1 *QTL mapping*

In QTL mapping, parents' contrast for the trait of interest are usually selected. The progeny from such parents is called a mapping population. The progeny is genotyped, and the genetic map is constructed based on these data. Phenotypes are used for QTL detection. Detected QTL has to be validated on other populations. If it is confirmed, the markers to detect this QTL are used in MAS (114).

For the sunflower, QTLs were found for all quantitative traits important for cultivation. For a flowering time, several QTLs in different chromosomes were identified using different populations (158–160). Multiple QTLs were identified for different diseases: downy mildew (133, 161, 162), *Diaporthe helianthi* (163), *Phoma macdonaldii* (133, 164–166), *Sclerotinia sclerotiorum* (163, 165, 167–170), *Orobanche cumana* (broomrape) (171, 172). For abiotic stress tolerance, QTLs were described for such traits as water status, osmotic adjustment, and yield under drought conditions (173,174).

Since sunflower is an oilseed crop, seed and oil yield-related traits are in priority. Seed yield is very environment-dependent and has low heritability, but for oil content, heritability is high, 65-70% (175).

Lately, QTLs for oil yield were mapped in several populations (158, 160, 168, 176–178). QTLs for oil composition are also described. Besides, oil content, oil composition also affects oil properties. A study (108) identified QTLs associated with stearic and oleic acid concentrations. Ebrahimi et al. 2008 identified QTLs, which affect palmitic, stearic, oleic, and linoleic acid content (179).

While there was a tremendous amount of QTL mapping studies on sunflower and other crops, and these studies contributed greatly to breeding, QTL analysis has several limitations. First, QTL analysis results are specific for each cross (180). There may be not enough resolution due to the small number of recombination events (181). Limitations mentioned above can be overcome with LD-based approaches, namely, association mapping (182). It has a higher resolution due to the higher number of meiotic recombination events between analyzed individuals (183).

2.8.2 LD concept in the context of association mapping

Linkage disequilibrium (LD) is a nonrandom association of different alleles at different loci. Jennings proposed the concept of LD in 1917, Lewontin made the first quantification of LD in 1964 (184). LD can occur between alleles in one chromosome or different chromosomes (185). LD usually is an outcome of recent selection events, gene migrations, mutations. Self-pollination, inbreeding, low recombination rate, genetic isolation, and epistasis extend the LD, while out-crossing, high recombination rate, and high mutation rate cause LD decay (186).

LD, usually estimated as r^2 (the squared value of correlation coefficient between allelic states) or with D, which describes the difference between the coupling gamete frequencies and repulsion gamete frequencies at two loci. LD is measured as normalized D' (187, 188). Compared to wild sunflower, where LD decays rapidly, falling to 0.1 and less after 200bp, cultivated sunflower demonstrates that higher levels of LD are maintained across greater distances. In highly inbred cultivars, it is extended up to 1 Mb (139). Therefore, for such inbred cultivars, just a moderate number of SNPs can provide sufficient resolution for association search.
Mapping based on LD allows to survey the entire gene pool, gives higher resolution, analyzes more allelic diversity, and doesn't require the production of a bi-parental mapping population. It further allows the mapping of many traits in one set of genotypes. And important is that this approach has potential in causal variant search (185).

2.8.3 Association mapping

Association mapping or genome-wide association study (GWAS) is usually performed on a big group of individuals, where the dataset is chosen in such a way as to capture maximum genetic diversity. It utilizes ancestral recombinations and natural genetic diversity based on the LD concept (189).
GWAS, based on genotyping data (not full genomes), does not look for causal variants in most cases. The main idea of such GWAS is that causal variant is located in a haplotype, and identified markers will be within the same LD block with a causal variant and, therefore, associated with a trait of interest. This makes the results unbiased and independent from in-depth trait knowledge on the genomic structure and makes the GWAS approach useful for mapping complex quantitative traits (190).

Usually, association mapping requires precise phenotyping, which is performed for various traits of interest, preferably in different locations and environments for many years. Extensive genotyping of selected individuals makes the analysis very powerful. Population structure and relatedness across individuals (kinship) are usually calculated and included in the analysis to avoid false-positive results. This is important since the difference in trait manifestation could be explained not only by a particular SNP or loci but also by the recent common ancestry of individuals or kin relationships between individuals (126).

First, GWAS studies were performed with general linear models (GLM), now mixed linear models are (MLM) more commonly in use due to the possibility to account for kinship (191). For the first time, the MLM approach was implemented on maize in 2006 (192). GWAS loses its power on rare alleles, so a minor allele frequency (MAF) filter is usually applied. Large datasets help deal with such problems by capturing more individuals carrying rare alleles thus ensuring their confident detection (193).

GWAS is best performed on inbred lines to have most of the polymorphisms in a homozygous state. That's why Arabidopsis was an ideal object for GWAS studies due to its self-pollination (194). GWAS for this model plant were performed on more than 100 phenotypes (78). For now, GWAS were performed for nearly on all crop species like maize (144, 195–197), soy (198–201), rapeseed (202–204), rice (205–207), wheat (208, 208–210), barley (210–212), and others. GWAS for sunflower has already been performed for many traits of interest, including flowering time (213), male fertility restoration (214), seedling growth (215), the plasticity of oil yield for combined abiotic stresses (216), basal and apical branching (217), flower morphological traits (218), and others.

Now, since next-generation sequencing (NGS) methods became widely available and economically feasible, successful GWAS are mostly limited by the availability of phenotyping data. Advances in phenotyping methods will push GWAS forward since, for now, it is the most powerful approach to look both for associations and traits' genetic architecture (219).

2.9 High-quality phenotypes as a key to successful crop improvement.

Genomic revolution and emerging gene technologies made sufficient progress in agriculture, but it should be carefully linked to the phenotype for the application of genomic information in crop sciences. Phenotyping remains the bottleneck in crop studies because it is technically challenging and very labor-consuming (220).

2.9.1 Advances in plant phenotyping technologies

To be productive in combination with genomics, phenotyping should utilize high-throughput approaches. Development and adoption of high-throughput and high-dimensional phenotyping formed a new discipline - phenomics (221). The trend is to collect as many phenotypes as possible using sensors, robotic platforms for image acquisition, machine learning, and automated high-throughput image processing (222). New high-throughput technologies are implemented both indoor and in the field (223). High-throughput phenotyping platforms are successful in the precise acquisition of information about the traits of interest (224). High-throughput phenotyping platforms were developed for many crop species, including sunflower (225). Field-based high-throughput phenotyping was also performed for sunflower (23). Phenomics works closely with envirotyping, the discipline which uses high-throughput technologies for environment conditions profiling (226).

Another page in crop phenotyping with the view of crop improvement – is molecular phenotyping, including metabolomics profiling (227). Metabolomics mostly relies on nuclear magnetic resonance (NMR) and mass spectrometry (MS) technologies coupled or not with chromatographic compound separation techniques (228). NMR and its modifications like isotope-labeled NMR (229) and micro-coil NMR (230) are non-disruptive, rapid, and highly quantitative techniques. Still, they lack sensitivity compared to MS. MS is usually coupled with gas chromatography (GC) for volatile compounds and liquid chromatography (LC), which can cover a large portion of metabolites. (227).

Phenotypes collected by metabolomic profiling techniques are widely used in genome-wide association studies (GWASs) assisted by metabolomics techniques (mGWAS) and metabolic quantitative trait loci (mQTLs)

Metabolomic GWAS were established in human studies to discover specific markers for diagnostic or targeted drugs (231). In plants, it was first implemented on Arabidopsis and later on crop species (232): maize (196), rice (233,234), tomato (235), and others.

2.9.2 Mass-spectrometry coupled with liquid chromatography as a powerful tool for lipidomic profiling

The metabolome is very complex and chemically diverse, so there could not be one single technique that would allow profiling all compounds in a single analysis (236).

To work with a certain fraction of metabolites, several extractions (237) and separation techniques (238) have been developed to reduce the complexity of the mixture, which will be

analyzed. Lipids are the most diverse group of biological molecules, so as is the case for all metabolites, no method can assess all lipids simultaneously (239).

The first implementation of MS techniques to the analysis of complex lipid mixtures dates back to 1994 (240, 241). Previously, lipids were analyzed by different chromatography techniques (242). During the past 20 years, various MS-based detection techniques combined with diverse chromatography separation methods for precise and specific lipidomic profiling in humans, animals, and plants have been developed. In contrast to animal material analysis, for plants, mostly targeted MS approached were implemented (243,244). In 2011, Hummel et al. proposed to couple ultra-performance liquid chromatography (UPLC)-based separation system with high-resolution mass spectrometry for plant lipid profiling (245). This allows accurate profiling of several hundreds of different lipid species extracted from a single plant sample. UPLC can be combined with orbitrap mass-spectrometry and QTOF to detect both nonfragmented ions and fragments in MS-MS mode (246). Later this approach was adapted for plant fatty acid profiling (247). Liquid chromatography is an alternative to the shotgun approach where lipids are injected in mass-spectrometer without pre-separation and, hence, low abundance lipid species may be missed (248). All parameters like solvents, gradient, time of separation, and ionization modes are set individually depending on the experiment (249).

There are many papers published on LC-MS lipidomics on the Arabidopsis model plant (228, 250–252). And there exist several studies where LC-MS were implemented to lipid profiling in crops like wheat (253), maize (254), soybean (255), rapeseed (256), and others.

2.10 MAS and GS as a future of sunflower improvement

2.10.1 Marker-assisted selection, from separate markers to full sequences

Marker-assisted selection (MAS) helped to overcome difficulties in the conventional selection and speed up breeding. Compared to phenotypes, genotypes are not affected by the environment and are detectable at all plant growth stages. The main goals that can be achieved by MAS can be summarized in three points: tracing favorable alleles, get rid of undesirable linkages, and find individuals beneficial for future crosses in segregating populations (257). MAS help breeders to perform phenotypic selection based on genetic markers (114). MAS has been successfully implemented to improve traits, both qualitative and quantitative, in many crops, including sunflower. Markers were discussed earlier in paragraph 2.7.

The marker-assisted selection has been moved to a new level by the advent of NGS technologies. In the post-genome era, the paradigm has shifted from separate markers to sequences, marker-assisted breeding to sequenced-based breeding (258).
While the main idea is still the same - to accelerate breeding, now it goes deeper into the understanding of the genetics of complex traits (259)

2.10.2 Genomic selection, towards precise predictions

Genomic selection (GS) determines the future of sunflower breeding. The genomic selection was developed for animal breeding (260). The availability of large whole-genome datasets made genomic selection attractive (26)
Genomic selection may be considered as a type of MAS or as a separate technique. The main difference from MAS consists of that desirable individuals are selected based on genomic estimated breeding values (GEBVs), that assess the overall performance of an individual based on its genetic composition (261).
GS establishes associations between markers and phenotypes using training populations. Mapping populations and QTL detection are skipped in GS. Based on the training population, the prediction model is constructed. The critical point here is the number of markers because even if the effect of each marker is minimal, a large amount of marker information covering the whole genome still has the potential to explain most of the phenotypic variation (262).
The accuracy of predictions is a big issue. It depends on many factors such as sample size, population structure, trait heritability, genotype-by-environment interaction (GEI), marker density, and LD size and strength. But usually, large datasets with well-characterized variable phenotypic data help overcome the accuracy problem (262). Analyses of hundreds and thousands of genomes became a standard (263).
Another way to improve genomic predictions is to integrate multi-omics approaches: genomics, transcriptomics, and metabolomics. This way allowed a significant improvement in predictions in maize (264, 265) and rice (266). There is a couple of studies in which genomic prediction approaches were applied to sunflower. A study (267) tested different prediction approaches to estimate hybrid performance based on AFPL genetic markers and such traits of interest as grain yield, oil yield, and oil content. Predictions were accurate for closely related plants, while for distant ones, it was challenging. Studies which identified SNPs facilitate the prediction studies in sunflower (151, 153, 155, 268). Mangin et al. tested the performance of several models for sunflower oil content prediction in hybrids based on SNP markers (150). Livaja et al., 2016 made predictions for sclerotinia resistance in sunflower based on the 25K SNP assay (135).

Overall, it is evident that more research is needed to successfully implement GWAS results in MAS and GS for the sunflower. A recent study shows that the implementation of results obtained on inbred lines can encounter some difficulties for predictions in hybrids (269). Genotyping and phenotyping of new accessions, implementing new techniques, and GWAS on different traits may help in a better description of germplasm. They will give new insights into the architecture of the traits of interest, which, in turn, will help to use genomic selection for successful crop improvement.

The future goes in the direction of combining different approaches. The 5G approach for plant breeding was recently proposed, suggesting integrating Genome assembly, Germplasm characterization, Gene function identification, Genomic breeding, and Gene editing (270).

Chapter 3. Materials and Methods

This work is combining several biochemistry, analytical chemistry, bioinformatics, and molecular biology approaches. The methodological part can be divided into genotyping, phenotyping and computational techniques.

3.1 Samples

The samples used in the study are presented by sunflower and rapeseed samples from 3 different seed banks listed below.

Two hundred ninety-two inbred lines from N.I. Vavilov Institute of Plant Genetic Resources (VIR, St. Petersburg, Russia). Mostly conventional lines in terms of fatty acid composition (18:2 range from 36 -79%). 3 middle – oleic (18:1 > 50%), 1 high-oleic line (18:1 > 80%). Inbred lines from VIR collection were obtained by repeated self-pollination of the interspecific and industrial hybrids obtained from different countries (8-25 generations of inbreeding). The majority of the lines made based on first heterosis hybrids Soldor (VIR704), Sunbred 265 from France (VIR630, 631) F1 (SW536xW635(France)) (VIR 636, 655,700, 734). Two lines–fertility restorer lines VIR 453 and VIR 658, obtained by continuous self-pollination of wild *H. annuus*. The study used lines with various morphological features: different height, branching, leaf shape, leaf and flower color, vegetation period, and resistance to downy mildew.

One hundred ninety-nine inbred lines from V.S. Pustovoit All-Russian Research Institute of Oilseed Crops (VNIIMK) (Krasnodar, Russia). Fatty acid composition is known for 99 lines: 2 lines high-oleic (18:1 > 80%), 7 middle – oleic (18:1 > 50%). Other lines with 18:2 range between 36 -70%). VNIIMK lines by origin close to VIR lines, but they are different in terms of morphological traits combination because an inbreeding was performed by different researchers in each generation. VNIIMK lines are contrasted by fatty acid composition.

One hundred forty-seven oil-producing sunflower lines were provided by Agroplasma Seed and Breeding Company (Krasnodar, Russia). The fatty acid composition is unknown. Agroplasma company collection is presented mostly by fertility restorer lines resistant to the broomrape (*Orobanche cumana*) and a smaller fraction of sterility maintaining lines. Lines were obtained by self-pollination of modern industrial hybrids.

All seeds were grown and collected in the Krasnodar region in Russia. For UPLC-MS analysis, the seeds themselves were used. For DNA extraction, seeds were germinated in the lab.

Soils of the leached black earth soil type. The sunflower was sowed following the preceding crop, fall wheat, at the seeding rate of 40,000 plants per hectare.

Sowing was carried out according to the following sowing system: 70×35cm, a single plant per planting pit. Farming techniques, as commonly used for sunflower.

Each line was grown on the plot with an area of 9.1 m^2.

For the addition, G×E interactions experiment six sunflower inbreed lines (one conventional and five high oleic) originating from the VNIIMK collection for three years with five biological replicates per year yielding a total of 89 accessions were used

For the testing and optimization of UPLC-MS analysis, 50 rapeseed lines from VNIIMK were used both spring-type and winter-type.

All the samples used in the study are presented in Annex for Chapter 3.

3.2 Reagents

NucleoSpin® Plant II plant DNA extraction kit (Macherey-Nagel), AMPureXP (Beckman Coulter A63881), Qubit dsDNA High-Sensitivity Assay Kit (Life Technologies Q32854), Agarose I (Amresco Amresco), SafeView (abm G108), EDTA (0.5M), Buffer EB (Qiagen 19086), polyethylene glycol 8000 solution (Sigma-Aldrich P5413), MgCl2, HindIII (20,000U/mL, NEB #R0104S), NlaIII(10,000U/mL, 2500U NEB #R0125), 10x NEB buffer 2.1 (NEB #B7202), 10x Cut Smart NEB (NEB #B7204), T4 ligase (NEB #A63881), Phusion® High-Fidelity DNA Polymerase (NEB M0530L),Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB M0531S), Ultrapure 1M Tris-HCL Buffer pH=7.5 (Life Technologies 15567-027),dNTP (10 mM, NEB N0447S), ETOH 95%, Agilent High Sensitivity DNA kit (Agilent), Methanol LC-MS (Scharlau, Spain), Methyl tert-butyl ether HPLC grade (Scharlau, Spain), Chloroform HPLC grade (Fisher Chemical, USA), Heptane LC/MS grade (Honeywell Fluka, USA), Water UHPLC-MS grade (Scharlau, Spain), Potassium hydroxide solution 45% in water (Sigma Aldrich, USA), NaCl USP grade (Helicon, Russia) HCl 37% (PanReac AppliChem, USA), Lipid standards (Oleic acid-$^{13}C_{18}$ (#490431 Sigma Aldrich), Palmitic acid-$^{13}C_{16}$ (#605573 Sigma Aldrich and Stearic acid-$^{13}C_{18}$ (#605581 Sigma Aldrich, 18:1-d7, LPC #791643 Avanti, and 15:0-18:1-d7 DG #791647 Avanti), Acetonitrile LC/MS grade (Fisher Chemical, USA), Isopropanol LC-MS grade (Honeywell Fluka, USA), Ammonium acetate (Honeywell Fluka,

USA), Formic acid 98%-100% LC-MS grade (LiChropur Merck Millipore, USA), Acetic acid Optima, LC-MS grade (Fisher Chemical, USA).

3.3 Genotyping

Genotyping was performed by NGS sequencing and was preceded by DNA extraction and NGS library preparation.

DNA extraction

DNA was extracted from chlorophyll-free sprouts after 1 week of germination without light. 100 mg of tissue for each sample was grounded to powder using FastPrep-96™ Automated Homogenizer (MP Biomedicals). Total DNA was extracted according to the CTAB protocol using the NucleoSpin® Plant II plant DNA extraction kit (Macherey-Nagel, Germany) and stored at -20 ˚C until needed. The purified DNA samples' quality and concentration were determined by gel electrophoresis and in the Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA).

GBS library preparation and sequencing

Illumina libraries were constructed using two restriction endonucleases – HindIII (rarely cutting enzyme, A/AGCTT) and NlaIII (frequently cutting, CATG/) according to the protocol described below. Before the procedure, DNA concentration for each sample was diluted to 10 ng/ µl in 96-well plates (1 well per sample), and working adapter stocks (0.5 mM) were prepared. 2 µl of "barcode" adaptor (5mM) was added into each well. To perform the first restriction digestion with HindIII, Master Mix containing 0.5 µl of HindIII (NEB, USA) with working concentration of 20U/ µl, 2 µl of CutSmart buffer (NEB, USA), and 5.5 µl of mQ was prepared. Then, 8 µl of Master Mix was added to each well containing 100 ng of DNA in 10 µl, and digested for 1 hour at 37°C in a thermal cycler (Thermo SimpliAmp). The digestion was terminated by incubation of the reaction mix at 65°C for 20 min. Right after that, the ligation reaction was carried out. The "barcode" adapter was ligated to the sticky ends produced by HindIII, allowing to pool the samples afterward. Master Mix including 1.5 µl of T4 ligase (NEB, USA) with a working concentration of 400U/µl, 5 µl of T4 buffer, and 23.4 µl mQ was prepared. Next, 30 µl of Master Mix was added to each well and incubated at 22°C for 2 hours. Inactivation of T4 ligase was performed by heating at 65°C for 10 min. Then, 10 µl was taken from each sample and pooled in one Eppendorf tube in order to simultaneously purify them using PEG8000 according to

instructions provided in a PEG purification protocol. To perform the second restriction digestion with NlaIII, Master Mix including 0.7 µl of NlaIII (NEB, USA) with a working concentration of 20U/ lL, 2 µL of CutSmart buffer (NEB, USA), and 16.1 µl of mQ+DNA mix was prepared. 2 µl of another adapter (5mM) called "common" 30 was added into each well to be ligated to the overhanging ends generated by NlaIII afterward. The digestion was carried out at 37°C for 15 min with the following inactivation at 65°C for 20 min. Then, the ligation reaction was performed. Master Mix consisting of 1.5 µl of T4 ligase (NEB, USA) with a working concentration of 400U/ µl, 5 µl of T4 buffer, and 23.4 µl mQ was prepared. After that, 30 µl of Master Mix was added to each well and incubated at 22°C for 2 hours. Inactivation of T4 ligase was performed by heating at 65°C for 10 min. The next step was to clean up the samples using AMPure beads (Agencourt AMPureXP kit) according to the instructions provided by the manufacturer. After purification, a PCR reaction was conducted to amplify the DNA fragments. The reaction mixture volume (50 µl) was distributed among 16 aliquots in order to reduce the risk of mistakes during PCR. Master Mix contained 28 µl of mQ, 5 µl of Primer 1 (5uM), 5 µl of Primer 2 (5 µM), and 10 µl of Phusion high-fidelity PCR Master Mix with HF buffer, dNTPs, and Phusion high-fidelity DNA polymerase (NEB, USA). PCR conditions were the following: 98°C for 30 sec, 14 cycles of 98°C for 10 sec, 65°C for 30 sec, 72°C for 15 sec, 72°C for 2 min, and then at 4°C. All 16 PCR reactions were separately purified with AMPure beads, and the double selection was conducted. The double selection allowed to remove small and large fragments, so the library mainly consisted of fragments about 400-500 bp. Library concentrations were first checked with Qubit 3.0 Fluorometer (Thermo fisher scientific), and the fragment length distribution was examined with 2100 BioAnalyser (Agilent) following Agilent High Sensitivity DNA Assay Protocol. Each 96-multiplexed library was sequenced across three lanes in Illumina HiSeq 4000 (San Diego, CA, USA) at the Skoltech Genomics Core Facility as either 150 bp or 75 bp paired-end reads. The sequencing dataset can be found in the NCBI repository: https://www.ncbi.nlm.nih.gov/bioproject/620114.

Computational analysis of sequencing data

For the results presented in Chapter 4:

Preprocessing of raw reads was performed using the Trimmomatic soft-ware (version 0.30) (271), after which 552,481,747 of good barcoded single-end reads were used as an input for the Tassel pipeline (version 5.0) (272) with k-mer length 65 bp. Additional filtering options were applied during variant calling: -mnMAF 0.01, -minMAPQ 10, and -minQS 20. Phylogeny reconstruction was carried out using the Mega software (version 7) (273) based on the SNP data

matrix. For SNP marker extraction bcftools (version 1.9) was used. To identify linkage disequilibrium (LD) blocks in the genome sequences of the analyzed sunflower samples, Plink (version 1.9) (274) was utilized.

For the results presented in Chapter 6:
Illumina reads were mapped onto the *Helianthus annuus* reference genome HanXRQr1.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_002127325.1/) using BWA MEM 0.7.9a-r786 (275) with consideration for uniquely mapped reads whose PE ends mapped within 1K of each other. Variants were called using the GATK pipeline (276), which considers indel realignment and base quality score recalibration and calls variants across all samples simultaneously through the HaplotypeCaller program in GATK. Variants were filtered using hard filtering parameters: MQ>36, QD>24, and MQRankSum<2, ensuring that the reads were mapped to a unique place in the reference with high quality (MQ), that the reads carrying both alleles were comparable in terms of mapping quality (MQRankSum), and that the actual variants were called with high quality (QD), filters that were not applied by default by GATK's HaplotypeCaller, resulting in the 2.3M SNP calls. To retain SNPs for population and GWAS analyses for oleic and linoleic acids missing calls rate < 0.3, DP > 4, MAF > 0.01 were applied, resulting in 15068 SNPs for GWAS for other fatty acids we used more strict MAF > 0.03 resulting in 12528 SNPs.

3.4 Phenotyping

3.4.1 Classical phenotyping

The following breeding goal traits were registered in the field in August:

Plant height (stem length from the soil level to the flower head);

The diameter of the flower head (distance between the flower head margins which passes through its center);

Stem branching (presence of lateral branches of the 1st and higher order);

Presence of pollen (male fertility of disk florets);

Presence of ray florets (pollinator attraction factor).

While the following traits were analyzed after harvesting:

Seed oil content

Seed husk content

100-seed weight

Seed number in the head

Phenotyping of each inbred sunflower line in the field was performed for 20 typical plants. Plants with the obvious damage caused by biotic (diseases, insects, or birds) and abiotic (flooding or anthropogenic traumas) stressors were not included in the study. Weather conditions during the vegetation period were characterized as generally favorable.

Plant height was measured with a ruler as a stem length from the soil level to the flower head at the end of the flowering period, which is considered the optimal developmental stage for this kind of measurement since stem growth has already finished, but the stem still retains its flexibility. Flower head diameter was also measured after the plant growth was completed as the distance between the head margins through the center. In the case of stem branching, the central head was measured.

Plants demonstrating all the multiple-head types described in sunflower, such as general, apical, and basal branching, including the "Neptune" morphotype with several lower lateral branches, were considered as branched plants. The presence of pollen was visually determined during the flowering period as male fertility of disk florets (DF) (normally androgyne). In the case of male sterility, anthers appeared to be reduced, light-colored and lacking bright yellow pollen. In the case of fertile genotypes, pollen appeared during the primary hours (between 8 and 12a.m.) when anthers emerge from the anther tube. The presence of ray (marginal, sterile) florets (RF), as one of the attraction factors for pollinators and a clear morphological marker character, was

registered during the flowering period by the presence of clearly-detectable petals of ray flowers around the head margin.

During phenological observations, the following indices were registered: planting-budding period (DTB), i. e. the number of days from the planting date to the development of primordial flower head ("star"); planting-flowering period (DTF), the number of days from planting date to 50% of plants in blossom; planting-physiological maturity period (DTM), i. e. the number of days from the planting date to 50 % of flower heads becoming yellow.

The data on climatic conditions during the year 2017 and the previous 5 years were obtained on the basis of the climatic records of the Kruglik meteostation (Krasnodar, Russia) readily available at https://rp5.ru. HU accumulations between the planting date and physiological maturity dates were calculated according to (277).

Seed oil content, seed husk content, 100-seed weight, and seed number in the head were assessed using seeds obtained by open pollination. Three flower heads were used for each line. Seed oil content was determined using the AMB- 1006 M NMR analyzer in the sample containing 20 g of seed in replication. Seed husk content was determined in a sample containing 10 g of seeds in replication. 100-seed weight was determined using a sampling containing 1000 seeds.

3.4.2 GC-FID

A total of 4–5 g of seeds were mixed together and homogenized, and 0.5 g was taken for fatty acid extraction with 4 ml of hexane. To obtain the methyl esters of fatty acids, 2–3 ml of the homogenized seed-hexane mixture was transferred into the new tube, and 0.1 ml of sodium methylate was added and mixed intensively for two minutes. The tube content was further transferred onto the paper filter with Na_2SO_4 on the bottom. The obtained filtrate was then placed into the DAG-2M automatic dispenser tube. GC-FID analysis was carried out using the <<Chromateck-Crystall 5000 >> GC chromatograph with the DAG-2M automatic dispenser. GC separation was performed in a SolGelWax column with the dimensions of 30 m × 0.25 mm × 0.5 µm; gas mobile phase - helium; speed - 25 cm/sec; temperature range - 185–230 °C.

FAs detection was performed based on retention times using FA methyl ester standards (Fluke). The percentage of each FA was calculated based on the peak area using GC software.

3.4.3 UPLC-MS

Lipid extraction

For lipid extraction, 10 mg (for each line) of sunflower seeds (1 sample-1 seed) with 400 µl of methanol/methyl tert-butyl ether mixture (1:, v:v) were homogenized in Precellys evolution

(Bertin corp. USA) (6800 rpm, 3* 20 sec, pause 30 sec) coupled with Cryolis filled with dry ice with 6 2.8 mm zirconium oxide beads (Bertin corp. USA) at the temperature not higher than 10 degrees. Then, extraction was performed using methanol/methyl tert-butyl ether mixture, according to (245) with minor modifications. After homogenization, 400 µl more of methanol/methyl tert-butyl ether mixture was added to each sample. Then each sample was vortexed. After sonication for 10 min in an ice-cooled sonic bath and incubation in 4°C for 30 min shaking after the sample was transferred in a new 1,5 ml eppendorff tube and 560 ml of water/methanol mixture (3:1, v:v) was added. This led to the formation of two phases: a lipophilic phase and a polar phase. After the addition of methanol-water mixture, the sample was vortexed for 10 min and centrifuged for 10 min at 4°C at 12700 rpm. The upper lipophilic phase was collected and vacuum dried (1,5 h, 30°C Concentrator plus, Eppendorf) and stored at - 80°C before measurement. For FA's quantification 5 isotopically labeled lipid internal standards were added to the extraction mixture (3 µg of each per sample).

For FAs analysis, the extracts obtained in the previous steps were hydrolyzed using the protocol adopted from Bromke, et al. 2015 (247). Lipid extracts were resuspended using 200 µl of a mixture of methanol and 6% KOH (4:1, v:v). The tubes were incubated for 2 h at 60°C with continuous shaking (1800 rpm). After cooling to room temperature, 100 µl of saturated NaCl solution was added. The reaction mixture was acidified by the addition of 50 µl of 29% HCl. Tubes were vortexed thoroughly and spun for 30 s at full speed using a table centrifuge. The FAs were extracted with 200 µl of chloroform–heptane mixture (1:4, v:v). After vortexing and 15 s of centrifugation, the organic phase was collected. The extraction with the chloroform/heptane mixture was repeated the second time, and the collected FA-containing organic phases were combined. The extract was washed by the addition of 200 µl of water followed by short vortexing on a table vortex and 15-sec centrifugation at 12700 rpm, which resulted in two phases. Finally, the organic upper phase was collected, dried in vacuum conditions (30min, 30°C Concentrator plus, Eppendorf) and stored at - 80°C before the measurements.

UPLC-MS profiling

Samples were processed using mass spectrometry (UPLC-MS) coupled with reversed-phase ultra-performance liquid chromatography (ACQUITY UPLC System; Waters, USA) in positive and negative ionization modes in Q-TOF Maxis Impact II, Bruker Daltonik, Germany. Settings: Ion Polarity: positive/negative, Scan mode: MS, Mass range: 50 -1200m/z, Spectra rate: 2Hz. UPLC separation was performed on the C8 Acquity Beh column (2.1 mm X 100 mm, 1.7-µm particle size; Waters) and the Acquity BEH C8 1.7 µm Vanguard precolumn (Waters) at 60°C.

For the FA profiling, spectrum acquisition in ESI- positive mode was performed. For UPLC separation gradient mobile phases consisted of two solvents.

For all FAs profiling excluding 18:1 and 18:2: Solvent A: 1%1 M NH4Ac and 0.1% formic acid in water; and solvent B, acetonitrile/isopropanol (7:3, 1% 1 M NH4Ac, 0.1% formic acid), with an injection volume of 3 µl. The following gradient profile was applied: 1 min, 55% B; 3 min, linear gradient from 55% B to 80% B; 8 min, linear gradient from 80% B to 85% B; 3 min, linear gradient from 85% B to 100% B. After washing the column for 4 min 50 sec with 100% B, the mixture was set back to 55% B, and the column was re-equilibrated for 4 min 10 sec (24.5 min total run time), with a flow rate of the mobile phase of 400 µl min. Final sample dilution: 1:5.

For the 18:1 and 18:2: Solvent A, 1%1 M NH4Ac and 0.1% acetic acid in water; and solvent B, acetonitrile/isopropanol (7:3, 1% 1 M NH4Ac, 0.1% acetic acid), with an injection volume of 3 µl. The following gradient profile was applied: 50 sec, 55% B; 1 min, linear gradient from 55% B to 75% B; 5 min, linear gradient from 75% B to 89% B; 1 min 10 sec, linear gradient from 89% B to 100% B. After washing the column for 2 min with 100% B, the mixture was set back to 55% B, and the column was re-equilibrated for 1 min 50 sec (11.5 min total run time), with a flow rate of the mobile phase of 400 µl min—final sample dilution 1:400.

For all other lipid profiling spectrum acquisition in ESI+ positive mode was performed. For the UPLC separation, the following solvent system was used: solvent A: 1%1 M NH4Ac and 0.1% formic acid in water; and solvent B, acetonitrile/isopropanol (7:3, 1% 1 M NH4Ac, 0.1% formic acid), with an injection volume of 3 µl. With the same gradient parameters that were used for FAs profiling. Final sample dilution: 1:25.

For the 6 lines which were collected from 3 different years, ESI - profiling with 1:10- dilution and ESI+ profiling in 1:25 and 1:3 dilutions were performed.

Lipidomic primary data analysis and annotation

For data processing, optimal parameters were generated using the Bioconductor IPO package. The subsequent peak peaking, chromatogram alignment, chemical noise subtraction, and intensity thresholding were performed using the XCMS 3.1 package (https://bioconductor.riken.jp/packages/3.1/bioc/html/xcms.html) (278). The output was a list of peaks, with retention time, m/z, and intensity for each sample. To exclude possible contaminants, the mean intensities of all sunflower peaks were compared to mean intensities in blank samples. (Figure 1). Only lipids with sample intensity at least two times higher than blank intensities were used in the analysis.

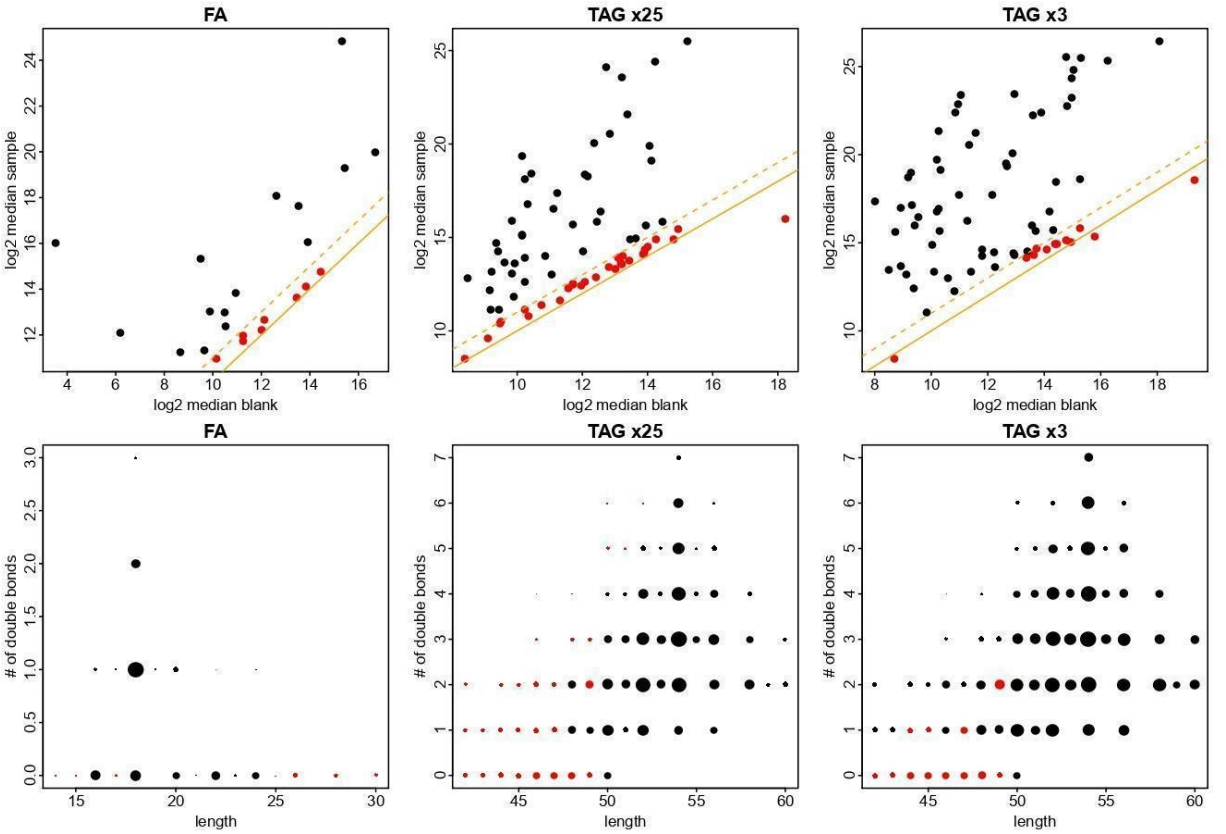


Figure 1. **Data clean up using blank samples**. The top panels show the dependence of average log2 seed sample intensity of FAs and TAGs (two dilutions) on average log2 intensity of the same lipids in blank samples (see Methods). Straight and dashed lines correspond to equal intensities in both types of samples and to two-fold higher concentration in seed samples compared to blanks, respectively. Bottom panels show the same lipids as top panels in coordinates of total FA chain length (x-axis) and a number of double bonds (y-axis). Point size is proportional to log2 average intensity in seed samples. Only lipids with log2(sample/blank) > 1 were used in the analysis. The remaining samples (shown in red) were filtered out.

To annotate FAs and TAGs, formulas for the possible lipids (irrespective to isomers) of these classes were generated. For FAs, chain lengths from C10:0 to C28:0 with not more than 6 double bonds were considered. For TAGs, the total chain length varied between 30 and 85 carbon atoms, and the number of double bonds varied from 0 to 12. Then, masses of generated lipids were compared to m/z of detected peaks. For FAs just one adduct ($-H^+$) was considered. For TAGs four adducts (H^+, Na^+, K^+, and NH_4^+) were considered. All peaks with ppm (ppm=abs(m1 − m2)/max(m1, m2) × 10^6), where m1 and m2 are masses of lipid and m/z of the peak, respectively) below 10 were considered as the possible lipids of the given class. Then, for each of the two lipid classes and for each adduct, the peaks were manually filtered based on the expectation that correct FAs and TAGs should form a net-like pattern on RT-m/z scatter plot. To annotate non-TAG lipids measured in the positive mode lipid Befdatabase (The LIPID MAPS®

Lipidomics Gateway, https://www.lipidmaps.org/) was used. First, all isomers were collapsed, then m/z of all non-TAG peaks were compared with the masses of all lipids from the lipidmap. The same adducts as were used for TAGs were considered. All lipid-peak pairs with ppm < 10 were considered as a valid annotation. All peaks annotated with lipids of just one category were assigned to this category. Peaks annotated with lipids of more than one category were considered as ambiguously annotated.

Reproducibility experiments with three years of replicates and the main dataset were measured and processed separately.

In the reproducibility experiment, only TAGs with NH_4^+ adduct and FAs were considered. For both dilutions of TAGs and for FAs, intensities of individual lipids were divided by the total intensity of all TAGs/FAs in the given sample and multiplied by 100. To assess the role of genetic and environmental factors ANOVA with the following model was used:

Lipid_concentration ~ line + year + line:year.

MDS analysis for two dimensions was performed based on one minus Spearman correlation coefficient distance.

3.5 Computational approaches for population structure estimation, LD analysis, and association search

Population structure

Genetic diversity among analyzed lines was estimated using PCA with the aid of PLINK (274) based on 15068 SNPs with minor allele frequency (MAF) > 0.01 called on all 17 chromosomes. Population structure was analyzed using ADMIXTURE v1.3.0 (279), with the number of clusters varying from 1 to 10.

Linkage disequilibrium

LD was estimated across the sunflower genome using VCFtools (280) to calculate frequency correlation (r^2) between 25431 biallelic SNPs with MAF > 0.03 whose genotypes were supported by at least 4 reads called in at least 60% of individuals.

Association analysis and annotation

For chapter 4:

Statistical analysis using the mixed linear models (MLMs) (281) implemented in the TASSEL 5 software (282) was performed for association mapping with PCA and kinship matrixes as

covariates. Multiallelic variants and those with the high missing call rates, MAF below 0.01 as well as the samples with many missing calls, were filtered out in PLINK 1.9 (274) before genome-wide association study (GWAS) analysis. Significant loci were identified based on Bonferroni and FDR adjusted q-values with 0.01 alpha significance level. GWAS results were visualized with the help of the qqman R package (version 0.1.4) (283).

For chapter 6:

For GWAS MLM (281) was implemented (MLM: Y = SNP + PCs + Kinship + e, where Y – phenotype, SNP and PCs – fixed effects, Kinship – random effect, e - error). In addition, internal standards intensity and LS-MS batch numbers were used as co-factors to account for the batch effect and sample weight in the model.

For GWAS, all the samples with 10% and more missing data (for phenotypes) were excluded from the analysis. GWAS was performed using TASSEL 5 (282). SNPs for the analysis were filtered out using the following criteria: missing calls rate < 0.3, DP < 4, and minor allele frequency (MAF) < 0.01 for such traits as oleic and linoleic acids and MAF< 0.03 for other traits. Filtering was performed using VCFtools. A mixed linear model was used where the SNP effect and population structure estimated by PCA were treated as fixed effects, and kinship was included in the model as a random effect. The genetic relatedness analysis was performed with the relative kinship coefficients (K-matrix) calculation using the TASSEL software (Centered IBS method). The collection and the batch number were also used as factors and sample weight and internal standards intensity as covariates. In order to estimate the mixed linear model performance, quantile-quantile plots (q-q plots) were used. Observed p-values were plotted against the expected probability of their distribution. For the representation of GWAS results, Manhattan plots were used where p-values were plotted for all sunflower linkage groups one by one.

All GWAS results were visualized with the help of the qqman R package (version 0.1.4) (283). To determine the significance of observed hits, 0.05/5000 p-value threshold was used. This is a Bonferroni correction based on the average number of LD blocks. The total number of SNPs used in GWAS was divided by 5000 - the number of LD blocks estimated from LD analysis. LD block analysis was performed using Haploview software (284). Gene annotation within each LD block was performed using the sunflower genome browser (https://sunflowergenome.org).

To estimate the variance in docosanoic acid concentration explained by identified 53 SNPs, the linear model with the same covariates as was used in GWAS analysis and with all 53 SNPs was used. Results show that 35.4% of the variance could be explained by these SNPs.

Chapter 4. Genetic and phenotypic diversity studies on Russian sunflower germplasm

4.1 Introduction and contributions

Germplasm collections are the key sources of new genes that can be used for crop improvement. Two publications support the results presented in this chapter. These studies' primary value is the genetic description and characterization of Russian lines that have never been sequenced before. Between the collections used in the two studies, there is a significant difference. In the first part, we used a collection from the VNIIMK research institute, which contains mostly lines contrast for oleic and linoleic acids ratio. The second part provides the data from Agroplasma breeding company collection, represented mainly by fertility restorer lines resistant to broomrape (Orobanche *cumana*) and a smaller fraction of sterility maintainer lines. Lines were obtained by self-pollination of modern industrial hybrids. In addition to the characterization of lines and their population structure, the pilot GWAS was performed, which yielded significant associations with linoleic acid content in chromosomes 8,9 and 17. In the second study, we isolated a region on chromosome 13 containing 21 candidate Rf1 fertility restorer genes. This chapter serves as a basis for the research presented in chapter 6 and gives excellent opportunities for the following studies on fertility restoration genetics.

This work is a part of a big project held in Skoltech for four years. In these studies, I have performed all experimental procedures in the genetic part. I have germinated all the study's accessions, extracted genomic DNA, and prepared GBS libraries. Phenotyping analysis was performed by collections holders, VNIIMK, and Agroplasma. I have performed phenotyping data organizing. Svetlana Goryunova and Denis Goryunov performed data analysis and taking care of results publishing.

4.2 Results: Genetic and Phenotypic Diversity of the Sunflower Collection of the Pustovoit All-Russia Research Institute of Oil Crops (VNIIMK).

4.2.1 VNIIMK lines genotyping

Sequencing of GBS-libraries and subsequent analysis have identified 65,553 variants, including SNPs and indels in 186 sunflower lines from the VNIIMK collection (Annex for chapter 4, Table S1). Overall transitions to transversions ratio was 1.73.

LD analysis revealed substantial variability across the genome. Mean, median, and the 1st and 3rd quartiles of LD block length distribution were 110.517 Kb, 0.053 Kb, and 0.021 Kb and 69.132 Kb, respectively. The most extended LD blocks (>5,000Kb) were found in linkage groups 1, 5, and 17, with the maximum LD block length of 6156.150 Kb in linkage group 5.

Phylogeny reconstruction was carried out using the Mega software (version 7) based on the SNPs data matrix. VNIIMK selection lines are present in all clades on the obtained tree (Figure 2)

Figure 2. **Tree of 186 VNIIMK sunflower lines constructed based on the SNP data.** The origin of lines: circles– VIR, stars- USDA, squares–NS, unmarked – VNIIMK.

It can also be observed that certain samples are similar to the samples from VIR and the United States and group with them. For example, sample s54 (line L 2586) from VNIIMK fall in the same clade with s477 (RHA 298) From USDA and s44 (VIR 391) from VIR. At the same time, some clades are formed solely by VNIIMK samples.

4.2.2 VNIIMK lines phenotyping

For phenotypic data collection, plants were grown in the field in the middle part of the Krasnodar Region during the year 2017.

Water scarcity is one of the critical limiting factors for sunflower growing in the Krasnodar Region. The areas on which VNIIMK experimental fields are located can be characterized as sub-humid, with the precipitation amount averaging to 738 mm in 2013–2016. During the sunflower growth period, the average rainfall is about 274 mm, and precipitation deficit can often be observed during the seed germination period. However, during the experimental period in May 2017, the rainfall was 104 % higher, although later precipitation amount did not exceed 78 % of the average for the last 5 years. In August, it decreased to a minimum of 35 % of the average (Figure 3).

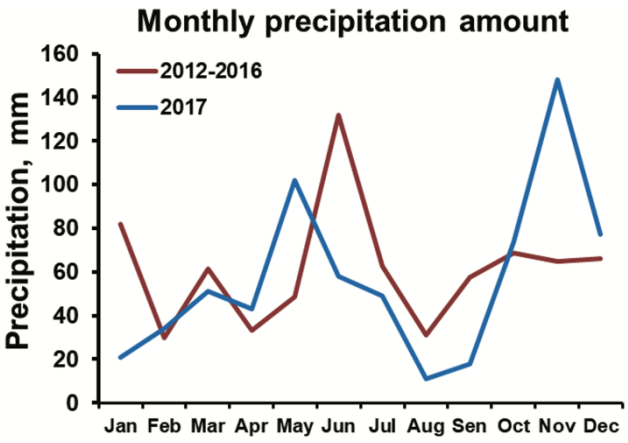


Figure 3. **Monthly precipitation amount.**

Temperature is another important factor affecting sunflower growth and development. The average annual temperature in 2012–2016 was+13.6°C, with the average daily minimum of 16.1°C being observed in January, and the maximum of + 30.8 °C, in July (Figure 4).

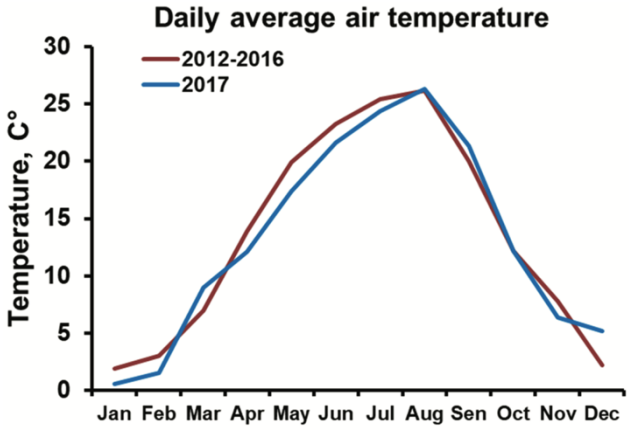


Figure 4. **Daily average air temperature.**

The summary positive air temperature in 2017 reached 4796 °C. The sum of the temperatures above 10°C equaled to 4076°C. Winters were snowless with frequent thaws. During the experimental period, the average air temperature varied within the 17.4–26.3°C range, which is 2.5–0.1°C below the climate normal.

Plant height ranged between 50.8 and 177.4cm (mean=119.2cm) in the analyzed inbred lines. Flower head diameter varied between 10.2 and 30.8cm (mean=18,8) (Annex for chapter 4, Table S1). Among all analyzed plants, 33 plants were branched, and 153 were non-branched with a single flower head. 135 sunflower lines were male fertile and 51 line, male sterile (CMS-PET1). The absence of ray flowers was observed only in 3 lines, appearing to be a rare event. The presence of ray florets in most samples may be accounted for by strong selection pressure since ray florets are considered a pollinator attraction factor. Seed oil content ranged from 23.4 % to 50.9 %, with the mean equal to 39.9 %. Seed husk content varied from 15.9 % to 47.1 %, and the mean value was equal to 29.3 %. 100-seed weight varied between 2.2 g and 12.5 g (mean = 5.3 g). Head seed numbers ranged from 210 to 1796 (mean = 839.7).

Length of DTB period varied from 36 to 52 days (mean 42), DTF period ranged from 54 to 82 days (mean 68), the period from planting date to PM (DTM) ranged from 79 to 114 days (mean 100) (Annex for chapter 4, Table S1).

Also, heat units (HU) accumulation between the planting date and physiological maturity dates was calculated (Figure 5).

A significant variation in HU accumulation by the physiological maturity date was observed among the analyzed sunflower lines (from 1129 to 1828 HU).

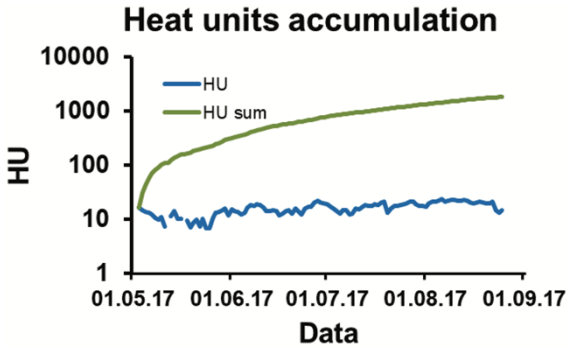


Figure 5. **Heat units accumulation**.

The histograms depicting quantitative morphological and phenological traits distribution among the sunflower lines from the VNIIMK collection are presented in Figure 6 (A–E).

A significant correlation was found between the planting date to the budding period and the planting date to the flowering period (Pearson coefficient = 0.851) and between the planting date to flowering and the planting date to physiological maturity periods (Pearson coefficient = 0.870). The correlation between the planting date to the budding period and the planting date to physiological maturity was not strong (Pearson coefficient = 0.695). It should be noted that the ratio between the analyzed periods varied between different lines. For example, for line no. 577,432, the period from the planting date to flowering was 59 days and the period from the planting date to physiological maturity was 94 days. In comparison, the line HA89 showed longer planting date to a flowering period (67 days) and shorter planting date to physiological maturity period than line no. 577,432 (89 days).

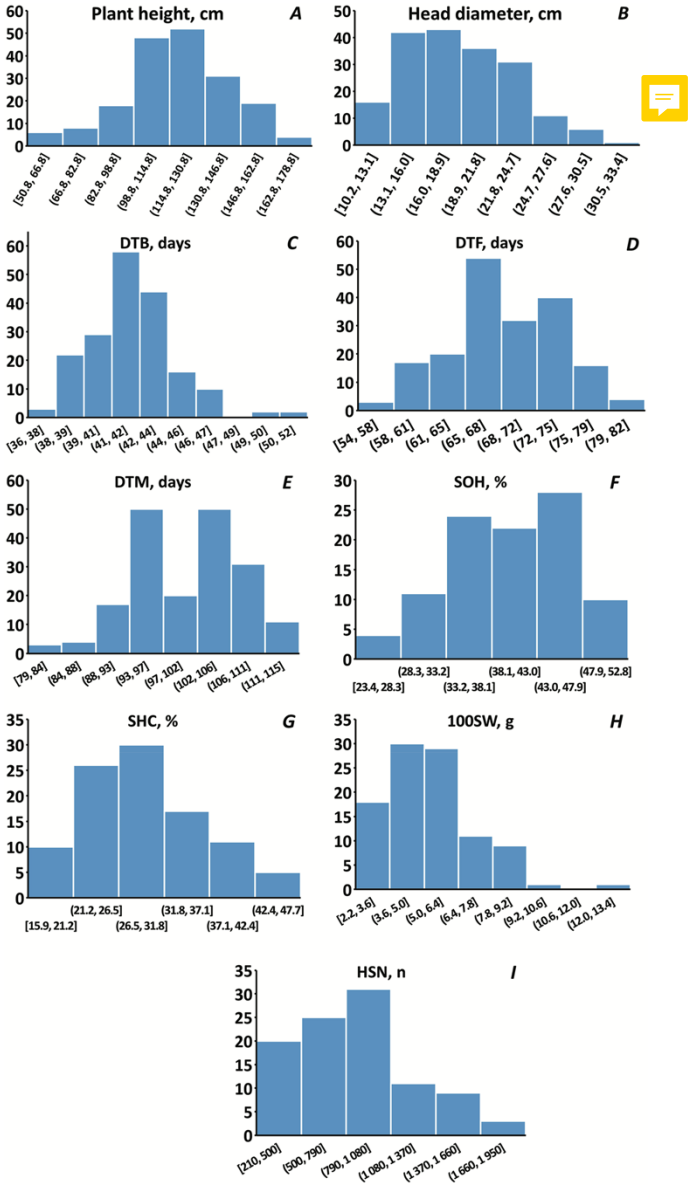


Figure 6. **Quantitative agronomically important traits in sunflower**. (A) Plant height, cm; (B) Head diameter, cm; (C) planting date to budding period, DTB, days; (D) length of the planting date to flowering, DTF, days; (E) planting-physiological maturity period, DTM, days; (F) seed oil content, SOH, %; (G) seed husk content, SHC, %; (H) 100-seed weight, 100 WS, g; (I) seed number in the head, HSN, %.

Fatty acid composition of seed oil from 99 sunflower line samples was analyzed using gas chromatography with flame ionization detection (GC-FID). As a result, it was detected 11 FAs: C14:1, C16:0, C16:1, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C22:0, and C24:0. The variability in sunflower oil fatty acid content observed among 99 VNIIMK sunflower lines is presented in Annex to chapter 4, Table S2. The most abundant FAs were C16:0, C18:0, C18:1, and C18:2, which constituted 97.5% of all FAs. The remaining 7 minor fatty acids constituted 2.51% of all FAs. The relative abundance of FAs was demonstrated to vary between samples.

High variation coefficients calculated for each fatty acid (from 14.9% to 84.5%) point to significant differences between genotypes. Oleic acid and such essential fatty acids as linoleic and linolenic are of major interest for sunflower breeding. Among the analyzed sunflower samples, ω-9 oleic acid content varied within a broad range (between 16.34% and 88.66%). Two samples, VK 464 and LG 26, were shown to be high oleic acid lines with an oleic acid content of 88.7 and 83.67 %, respectively. High variability was also demonstrated for ω-6 linoleic acid (from 3.13% to 68.55%). The variability range for ω-3 linolenic acid was between 0.05% and 0.88%. This acid is present in relatively low concentrations in the oil.

4.2.3 GWAS on VNIIMK lines

The genotypes and phenotypes of inbred lines obtained in this study were used in the GWAS analysis. Significant associations were found between the branching trait and the loci on the linkage groups (LGs) 10 and 1 (Figure 7). Novel significant associations with linolenic acid (18:3) content in the seeds were found on LGs 8, 9, and 17 (Figure 8).

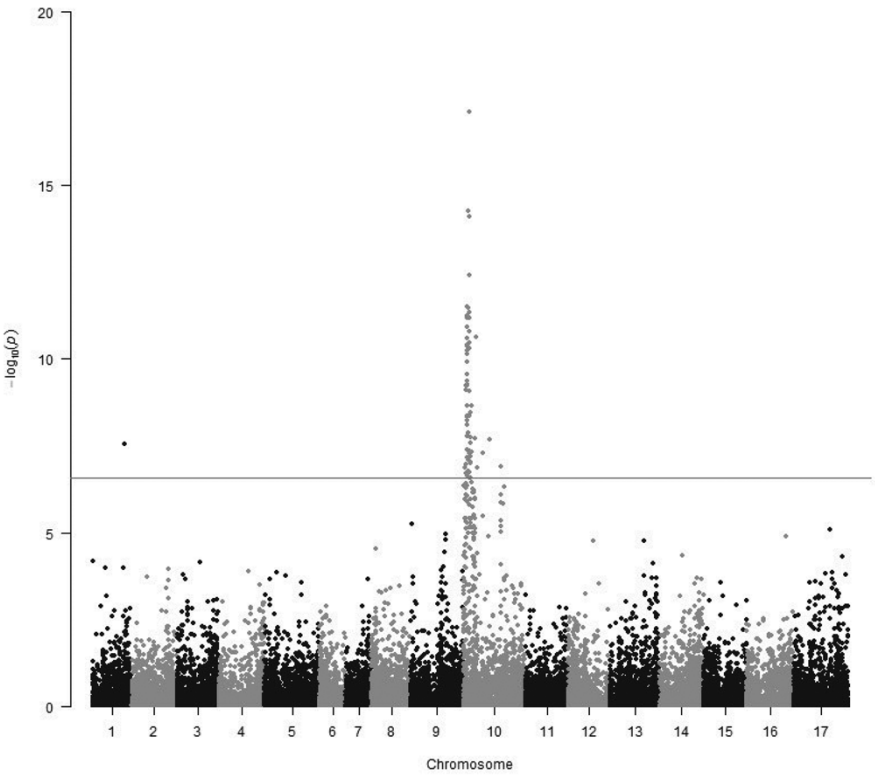


Figure 7**. Manhattan plot of branching associations**. The black line indicates the significance threshold based on the Bonferroni multiple testing correction (alpha = 0.01).

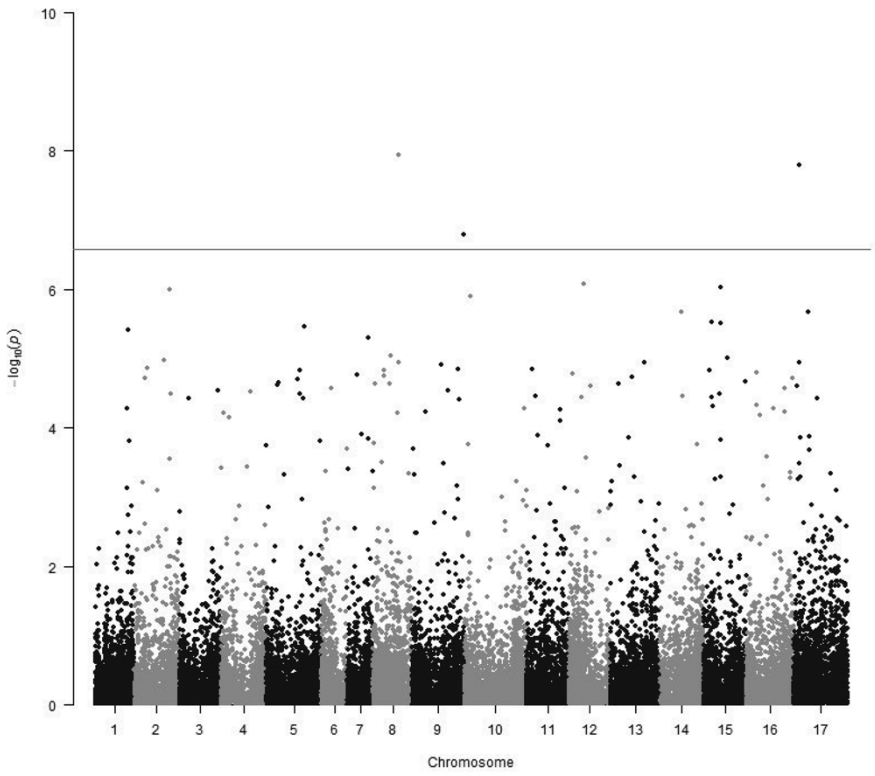


Figure 8. **Manhattan plot of linolenic acid content associations.** The black line indicates the significance threshold based on the Bonferroni multiple testing correction (alpha = 0.01).

4.3 Results: Association Mapping of Fertility Restorer Gene for CMS PET1 in Sunflower

4.3.1 Agroplasma lines genotyping

Sequencing of GBS-libraries and subsequent analysis has identified 28,153 SNP in 134 sunflower accessions (Annex for chapter 4, Table S3). Overall transitions to transversions ratio was 1.83.

PCA analysis revealed a significant population structure. Restorer lines and sterility maintainers form separate groups on the scatterplot (Figure 9).

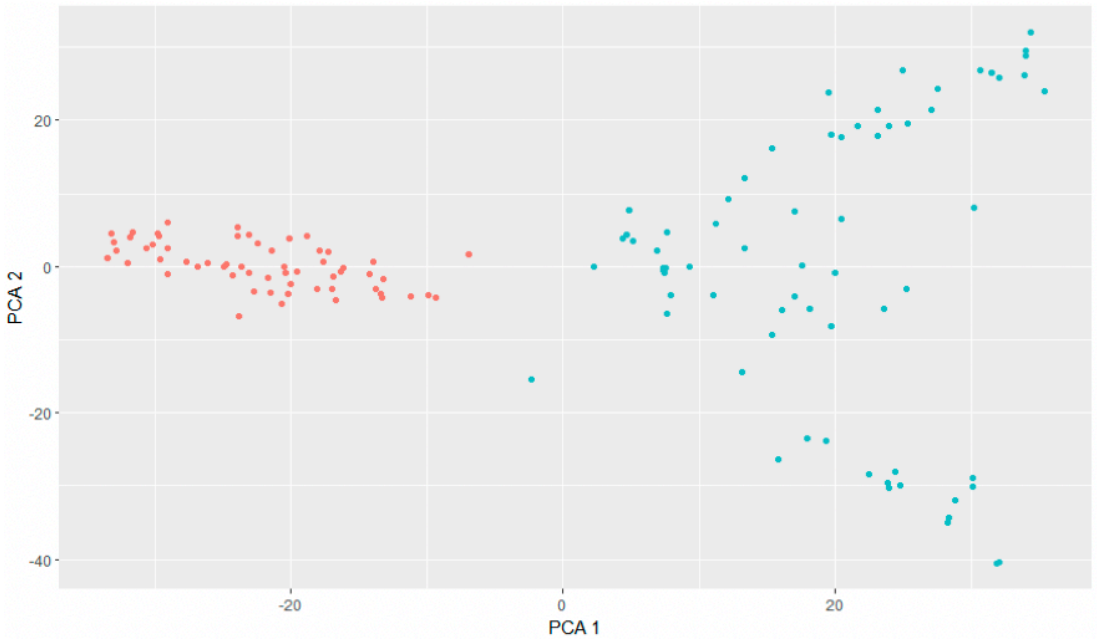


Figure 9**. Principal component analysis plot.** Pink dots–sterility maintainers, blue dots–restorer lines.

4.3.2 GWAS for fertility restoration

GWAS analysis revealed four loci associated with the ability to suppress CMS phenotype. A single significant marker was revealed at both 8 and 17 linkage groups. Most of the markers significantly associated with the trait under study and the markers with the highest *p*-values, were located at 10 and 13 LG (Figure 10, Annex for chapter 4, Table S4).

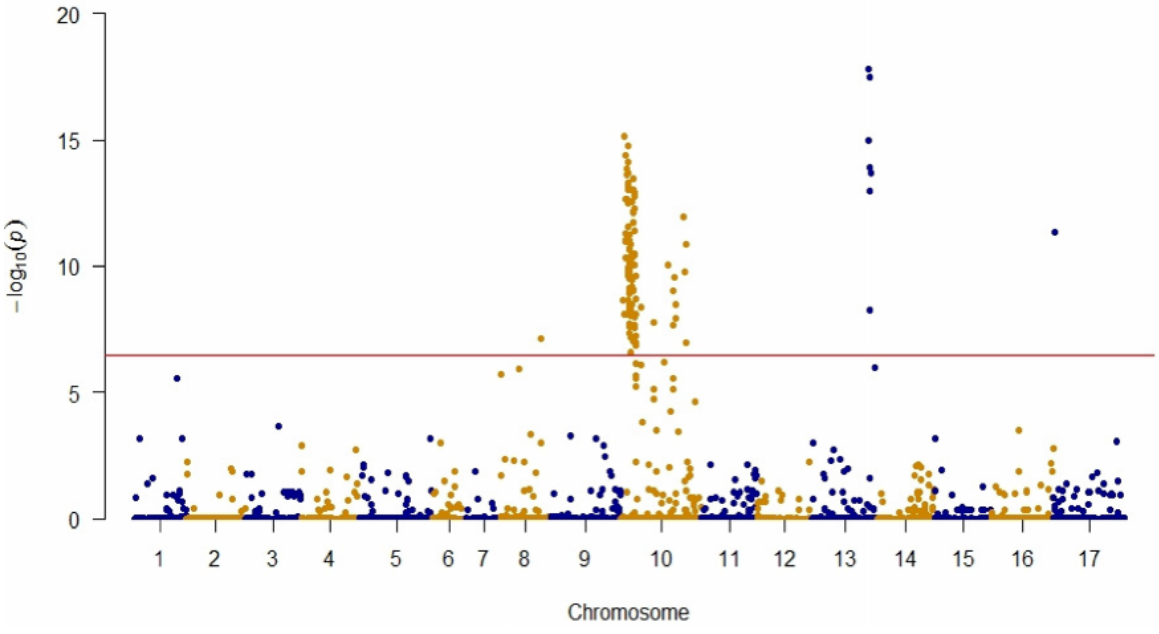


Figure 10. **Manhattan plot of associations with the ability to suppress CMS (cytoplasmic male sterility) phenotype**. Red line indicates the significance threshold based on the Bonferroni multiple testing correction (alpha = 0.01).

Within chromosome 13, based on the GWAS analysis results, a 7.72 Mb long section (coordinates170494693–178217103) can be distinguished, in which eight significant SNPs are located with *p*-values ranging from 5.69×10^{-9} to 1.53×10^{-18} (Table 1, Annex for chapter 4, Figure S1).

Table 1: List of Single nucleotide polymorphisms at linkage group 13, significantly associated with the ability to suppress CMS phenotype after Bonferroni correction.

Marker	Position	p-Value
S13_170494693	170494693	1.01×10^{-15}
S13_171053833	171053833	1.53×10^{-18}
S13_173268042	173268042	3.46×10^{-18}
S13_173832391	173832391	5.69×10^{-9}
S13_174474103	174474103	1.22×10^{-14}
S13_174474122	174474122	1.22×10^{-14}
S13_174809087	174809087	1.10×10^{-13}
S13_178217103	178217103	2.03×10^{-14}

To compare the localization of the 7.72 Mb region identified in this study with previously reported data, PCR primer sequences of the ORS511, ORS799, and ORS1030 markers were blasted against the reference genome. ORS511 and ORS1030 were mapped near each other on LG13, according to previously published data (136). Complete sequences of ORS1030 forward and reverse primers were mapped with 100% identity twice in the genome. Forward primer mapped to the positions 169535691–169535666 and 169655088–169655063 and reverse primer to the positions 169535262–169535287 and 169654659– 169654684 of LG 13. For ORS511 complete sequences of forward primer have no hits on the 100% identity threshold. The reverse primer of ORS511 was mapped at 169733686–169733704 of LG13. For the ORS799 marker, complete sequences of forward and reverse primers were uniquely mapped to the genome in position 186516272–186516291 and 186516418–186516399 of LG13 respectively.
These data suggest that the identified 7.72 Mb region (coordinates170494693–178217103) is located within segment of chromosome 13 flanked by SSR markers ORS799 and ORS1030 (coordinates 169535262–186516418).

4.3.3 Identification of Rf1 candidate genes

Within identified 7.72 Mb region in the HanXRQr1.0 reference genome sequence (10) 11 *PPR* genes are located, which are the most likely candidate genes for the fertility restorer gene *Rf1*. Besides, within this region, 11 genes of the *TPR* family are annotated. It is known that the sequences of PPR proteins are similar to the sequences of the TPR-family proteins, and it is assumed that the tetratricopeptide repeat (*TPR*)- family genes gave rise to *PPR* genes at the early stages of the evolution of eukaryotes (285).
Therefore, it was decided to include both the *PPR* and *TPR* families' gene sequence in further analysis. It should be noted that genome sections 7.72 Mb in length, flanking the region of the chromosome 13 mentioned above, did not contain any annotated sequence of the *PPR* family and only a single sequence belonging to the *TPR* family (HanXRQChr13g0421851).

The analysis of the translated amino acid sequences of genes of the *PPR (*pentatricopeptide repeats*)* and *TPR (*tentatricopeptide repeats*)* families located in the identified region and its flanking regions was conducted using ScanProsite tool of ExPASy SIB Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland). As a result, in all 11 amino acid sequences of the PPR family and in 10 of the 11 sequences of the TPR family, Pentatricopeptide (PPR) repeats were identified. Therefore, within the 7.72 Mb region and the flanking regions, 21 genes were detected, their protein products demonstrating the primary structure characteristic of the *PPR* family's sequences. Meanwhile, in addition to PPR repeats, the amino acid sequence of the protein product of one of the genes revealed a region of homology with UDP-glycosyltransferases. Therefore this gene was excluded from the list of possible candidate genes for *Rf1*.
In addition to *PPR* genes, a gene annotated as Probable aldehyde dehydrogenase 5F1 was detected in the 7.72 Mb region of chromosome 13. It was previously shown that *Rf2* gene of maize is the gene encoding aldehyde dehydrogenase (286). Therefore, this gene is also a possible candidate *Rf* gene. The list of identified candidate genes is shown in Table 2 and their arrangement within the 7.72 Mb region is shown in Figure 11.

Table 2. The list of candidate *Rf1* genes identified within the 7.72 Mb region.

Gene	Start	End	Strand	Product	Gene Bank Accession Number of Translated Protein	Hits for All PROSITE (Release 2018_11) Motifs
HanXRQChr13g0418841	170850155	170852002	+	Putative pentatricopeptide repeat	OTG02960	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0418861	170908019	170909110	+	Putative pentatricopeptide repeat	OTG02962	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0419621	173473487	173475525	−	Probable pentatricopeptide repeat-containing protein At2g41080	OTG03034	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0419631	173484455	173500401	+	Putative pentatricopeptide repeat	OTG03035	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0419931	174209661	174217234	−	Putative pentatricopeptide repeat	OTG03064	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420121	174799667	174801481	+	Probable pentatricopeptide repeat (PPR) superfamily protein	OTG03081	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420241	174944047	174945506	+	Putative pentatricopeptide repeat	OTG03093	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420261	174962084	174962512	+	Putative pentatricopeptide repeat	OTG03095	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420351	175219425	175219886	−	Putative pentatricopeptide repeat	OTG03099	PS51375 Pentatricopeptide (PPR) repeat

HanXRQChr13g0420811	176970038	176972308	+	Probable pentatricopeptide repeat (PPR) superfamily protein	OTG03141	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0421081	178216563	178219635	−	Probable putative pentatricopeptide repeat-containing protein At4g17915	OTG03166	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0418851	170877322	170879307	+	Putative tetratricopeptide-like helical domain	OTG02961	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0419881	174159006	174160682	−	Putative tetratricopeptide-like helical domain	OTG03060	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420271	175002640	175003793	+	Putative tetratricopeptide-like helical domain	OTG03096	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420281	175016437	175018065	+	Putative tetratricopeptide-like helical domain	OTG03097	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420301	175055952	175057826	+	Putative tetratricopeptide-like helical domain	OTG03098	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420371	175253986	175294219	−	Putative tetratricopeptide-like helical domain	OTG03101	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0420861	177597409	177599211	+	Putative tetratricopeptide-like helical domain	OTG03145	PS51375 Pentatricopeptide (PPR) repeat

HanXRQChr13g0420881	177609240	177611054	+	Putative tetratricopeptide-like helical domain	OTG03147	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0421271	178655189	178657150	−	Probable tetratricopeptide repeat (TPR)-like superfamily protein	OTG03183	PS51375 Pentatricopeptide (PPR) repeat
HanXRQChr13g0419821	174082899	174091500	−	Probable aldehyde dehydrogenase 5F1	OTG03054	NA

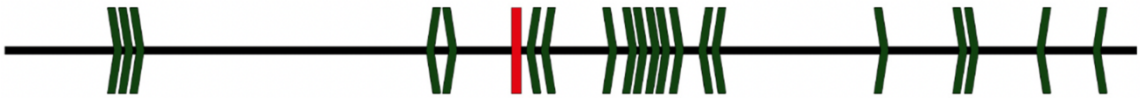


Figure 11. **Schematic localization of the candidate *Rf1* genes within the 7.72 Mb region**. Green arrows indicate the gene sequences of the *PPR* family. The direction of the arrow reflects the orientation of the sequence in the genome. The red box indicates the location of the Probable aldehyde dehydrogenase 5F1gene.

The number of PPR repeats in the sequence, and the length of the protein products of the candidate PPR family genes varied from 2 to 15 and from 110 to 756 amino acids, respectively. Genomic regions with increased LD could be recognized as strong selection pressure signatures on the traits encoded within these regions. The analysis results showed the presence of an extended section of elevated LD in 13 LG (Figure 12), of which the identified 7.72 Mb region forms part. This fact is an indirect proof of the localization of candidate genes in this region of the genome.

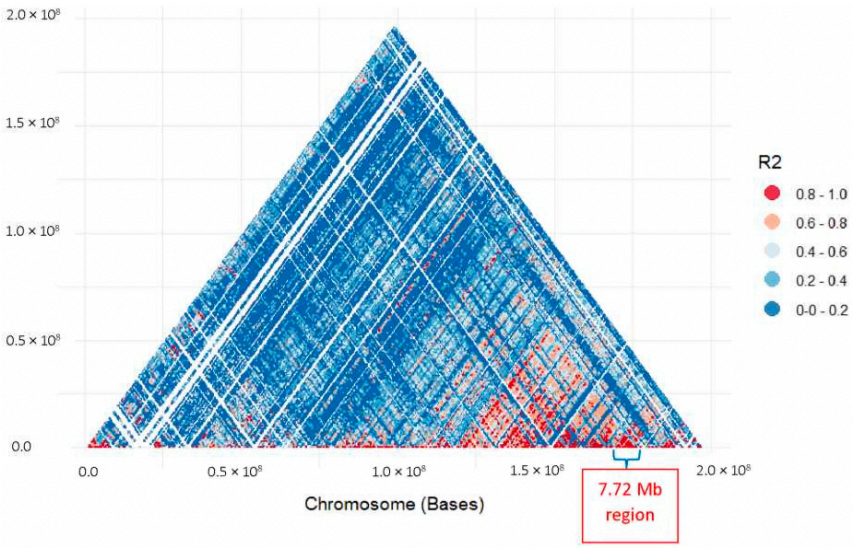


Figure 12. **Pairwise Linkage Disequilibrium (LD) Plot of the LG1** Individual data points reflect squared allele frequency correlations (R2) for all possible pairs of polymorphic SNP markers of LG13. The x- and y-axes correspond to the coordinates within 13 LG. Location of 7.72 Mb indicated by a curly bracket.

4.4 Discussion

In this chapter, two different sunflower collections were analyzed. The genetic material of the VNIIMK collection has made a significant contribution to the development of sunflower as the oilseed crop worldwide. However, notwithstanding its high value, the sunflower collection in VNIIMK remains poorly characterized in terms of using modern methods of genome-wide analysis. At the same time, a detailed study of genetic resources is essential for their practical use. In the present study, the genotypes of 186 VNIIMK lines were characterized by GBS sequencing for the first time. The data on their morphology, phenological, and biochemical characteristics were obtained. The analysis revealed a significant genetic and phenotypic diversity of the VNIIMK sunflower collection. New associations of genomic loci with linolenic acid content in seed oil have been revealed.

LD observations in VNIIMK collection stay in good agreement with the previous studies (287), where the authors accounted for their observations for by the proximity of the above-mentioned LD blocks to the genes or QTLs associated with the traits that served as selection targets during sunflower domestication and improvement.

Based on the obtained data, the major part of the analyzed lines in the VNIIMK collection may be described as the early-ripening or medium-ripening ones according to the existing

classification (288). This observation may reflect the high demand for the early- and medium-ripening sunflower and its prevalence among the sunflower hybrids in the Russian Federation. The use of early-ripening varieties and hybrids allows cultivating this crop in the regions with a short growing season, for example, in Siberia and Urals (289). Also, the use of early-ripening varieties and hybrids makes it possible to avoid the adverse effects of drought, frequent in some regions of sunflower cultivation in Russia.

The results obtained on morphological characteristics are consistent with the previous data obtained in the sunflower hybrid studies (277,290). Detected FA abundances are also in good agreement with data obtained in the previous works (291,291).

Our branching loci detection results stay in good correspondence with the data on branching loci localization to the upper half of the linkage group 10 (178,217). The data on the branching-associated loci on chromosome 1 (for basal type branching) was also obtained previously by Nambeesan et al. (217).

In chapter 4.4, sunflower lines, differing by the ability to suppress CMS phenotype, were analyzed. The GWAS results made it possible to isolate a segment 7.72 Mb in length on chromosome 13, in which 21 candidate *Rf1* fertility restorer genes were identified, including 20 *PPR*-family genes and one Probable aldehyde dehydrogenase gene.

Interestingly, in addition to the difference in the ability to restore pollen fertility in the crosses with sterile lines with PET1-type cytoplasm, the analyzed sunflower lines differed by the presence (restorer lines) or absence (sterility maintainer lines) of plant branching. This is because to obtain F1 hybrids, non-branched lines with a single large apical head are most often used as female parents, and lines with a recessive type of branching, with multiple small heads located on the lateral branches, are used as male parents. This approach allows an increase in the length of the flowering period of male parents due to the difference in the flowering times of the heads on the plant and getting F1 plants with a single large head. It is known from the literature that the branching locus is localized on chromosome 10 (178,217). Therefore, the associations identified on chromosome 10 seem to be linked to this trait.

At the same time, the associations identified on chromosome 13 correspond with the data obtained in the previous studies. For instance, Yu et al. combined RFLP, RFLP-SSR, and SSR maps and obtained data for localization of *Rf1* in LG13 [293). One year later, Kusterer et al. (294) map *Rf1* based on cosegregation with SSR markers ORS388 and ORS1030 belonging to LG 13 Tang et al. (136). Further, Kusterer et al. obtained a saturated map of the fertility restoration region *Rf1* (295). Mapping data have confirmed the location of *Rf1* on LG13 near marker ORS1030. According to Yue et al., *Rf1* is in the interval between markers ORS511 and

ORS799 of linkage group 13 (296). Based on this, the most likely location for the candidate *Rf1* genes appears to be chromosome 13, where a 7.72 Mb long region was observed. The results of the analysis showed the presence of an extended section of elevated LD in 13 LG. This fact is an indirect proof of the localization of candidate genes in this region of the genome.

During the annotation, the most attention was paid to PPR genes because almost all *Rf* genes in various plant species that have been identified so far belong to this family.
PPR genes are thought to be present in all eukaryotes, but they are most common in terrestrial plants' genomes, where they form one of the largest gene families (297). For example, in the genome of *Arabidopsis thaliana* L., there are about 450 genes of this family (298,299).
The total number of annotated *PPR* genes in the sunflower genome HanXRQr1.0 is 333. Therefore, the identified region of 7.72 Mb (comprising 0.214% of the genome length) contains 3.3% of all annotated *PPR* genes and is rich in *PPR* genes.

It should be noted that the reference genome used in the analysis was obtained by sequencing the XRQ line, which is a cytoplasmic male sterility maintainer (PET1 type) (10). At the same time, it is known that the *Rf* locus may undergo complex evolutionary events (300), and the structure of the identified site may differ in the genome of the fertility restorer lines. Therefore, to identify the *Rf1* gene, determine the sequence of the dominant alleles of the *Rf1* gene, and understand the evolution of the sunflower *Rf1* locus, the additional analysis of the structure of the 7.72 Mb region in the genome of fertility restorer lines is required.
The data obtained here will serve as the basis for further studies. They may contribute to further extensive utilization of the unique VNIIMK collection in oilseed sunflower breeding and give useful insights for further study of the genetic nature and molecular mechanisms for pollen fertility restoration in sunflower, as well as for the search of selection markers.

Chapter 5. Mass-spectrometry for sunflower lipid profiling

5.1 Introduction and contributions

In the literature review, the importance of implementing high-throughput technologies in crop studies was broadly discussed. New high-throughput phenotyping techniques can help speed up oilseed crop selection in oil production and oil quality. This chapter presents the data on the implementation of UPLC-MS technologies for FAs and TAGs profiling in sunflower and rapeseed seeds. These results serve as the basis for chapter 6, where the obtained lipidomic profiles were used in GWAS studies. We decided to analyze both sunflower and rapeseed seeds since our primary goal is to develop sunflower improvement strategies and make them expandable to other oilseeds. Two papers support the results presented in this chapter.

In chapter 5.2 published in PeerJ, we describe the UPLC-MS procedure for FA profiling in sunflower seeds and compare this methodology with the widely used in plants GC-FID technology. For oilseed crops, the knowledge on fatty acid composition, including minor fatty acids, is a critical factor of quality assessment on the one hand. On the other hand, the relative content of fatty acids determines the potential area of plant oil use (nutritional or technical). Therefore, precise and convenient methods for oilseed crop fatty acids profiling are in high demand. Commonly used techniques to measure FAs in plants, gas chromatography-mass spectrometry (GS-MS) or gas chromatography-flame ionization detection (GC-FID), are rather labor-consuming and don't allow to detect of the full spectrum of fatty acids.

Here, on a small cohort of samples from the full dataset (50 sunflower and 50 rapeseed accessions), we have optimized lipid extraction and MS profiling protocols and showed the advantages of the chosen procedures for phenotyping intending to use the obtained phenotypes in GWAS. Our study suggests using UPLC-MS technology for oilseed crops' fatty acid profiling as a highly sensitive, scalable, and suitable for individual seed analysis technique.

In chapter 5.3 published in Biomolecules, we tested the UPLC-MS methodology for TAG profiling in the same small cohort of samples. We have revealed significant differences in sunflower and rapeseed TAG composition and also found the differences between spring-type and winter-type rapeseed. We have demonstrated that winter-type rapeseed is enriched with the TAGs with a higher number of double bonds. In contrast, spring-type rapeseed was shown to contain TAGs predominantly with a high level of fatty acid chain saturation. We believe that our findings may give new insights into rapeseed freezing tolerance driving further research in this field, which is interesting in the frame of climate change.

In this chapter, I made the most significant contribution. I did the experimental design and carried out the experimental work from lipid extraction to the entire profiling procedure using mass-spectrometer. I have learned how to operate a mass-spectrometer, how to set all parameters, how to prepare extracts for analysis, and how to perform initial troubleshooting. I have optimized protocols for both extraction and profiling. I have participated in data analysis. By the way, Pavel Mazin and Rim Gubaev made a significant contribution to all calculations and figure preparation. In this chapter, I was fully responsible for paper writing and communication with the reviewers. Svetlana Goryunova and Denis Goryunov contributed to the draft revision. Anna Vanushkina and Waltraud Maier supervised the experimental part. All other people who coauthored these papers either contributed the experimental part or provided plant material.

5.2 Results: Ultra-performance liquid chromatography-mass spectrometry for precise fatty acid profiling of oilseed crops

In the present work, we compared the application of GC-FID coupled with hexane extraction and two UPLC-MS-based approaches: 1:5 dilution in the buffer system with formic acid (UPLC-5) and 1:400 dilution in the buffer system with acetic acid (UPLC-400) for FA profiling in sunflower and rapeseed (list of the samples provided in the annex for chapter 5, Table S1), both using MTBE extraction.

5.2.1 GC-FID data

Fifty lines of either of the two oilseed crops were analyzed using GC-FID. As a result, we have detected 11 and 13 FAs in sunflower and rapeseed, respectively (Figure 13C and 14C, the data can be upload from DOI: 10.7287/peerj.preprints.27501v1/supp-3). The lists of FAs seen in the two crops differed by minor FAs, while C16:0, C16:1, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C22:0, and C24:0, were common for both of them. FA C20:2, C22:1, and C24:1 were detected in rapeseed only, while C14:1 FAs we found in sunflower but were absent in rapeseed (Figure 13C and 14C)

In rapeseed oil, the most abundant FAs were C16:0, C18:0, C18:1, C18:2, C18:3, and C20:1, which together made for 98.6% of all FAs. In sunflower seed oil, the most abundant FAs were C16:0, C18:0, C18:1, C18:2, C22:0, which constituted 98.9% of all FAs. In both species, two FAs (18:2 and 18:1) constitute more than 80% of the total FA content. However, in rapeseed, 18:1 is three times more abundant than 18:2, while in sunflower, it is 18:2, which dominates (Figure 13C and 14C).

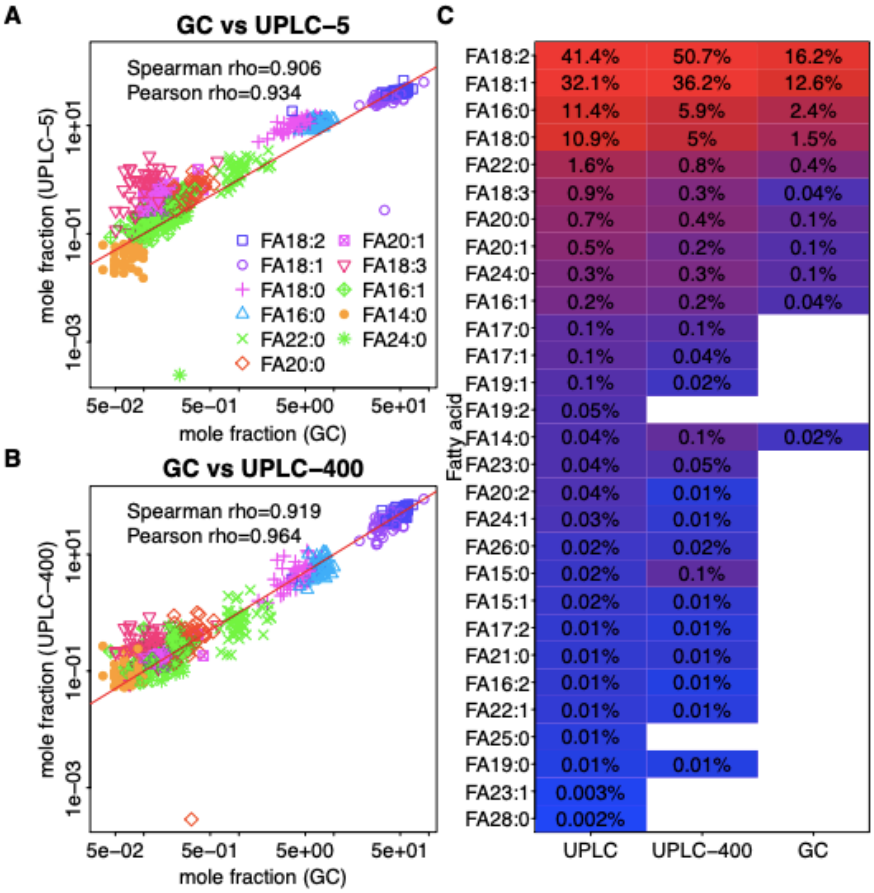


Figure 13. **Dependency between the mass fractions of FAs in sunflower (calculated for 11 FAs detected by GC-FID) estimated using different methods**. GC-FID vs. UPLC-5 and GC-FID vs. UPLC-400 are shown in (A) and (B), respectively. Each dot corresponds to a given FA (indicated by color) in a given sample (C) Mean mass fractions (relative to the total intensity of 11 FAs detected by GC-FID) for all FAs determined using the three techniques. White rectangles show FAs that were not detected by the indicated method. FAs are ordered according to their intensities, as obtained by UPLC-5.

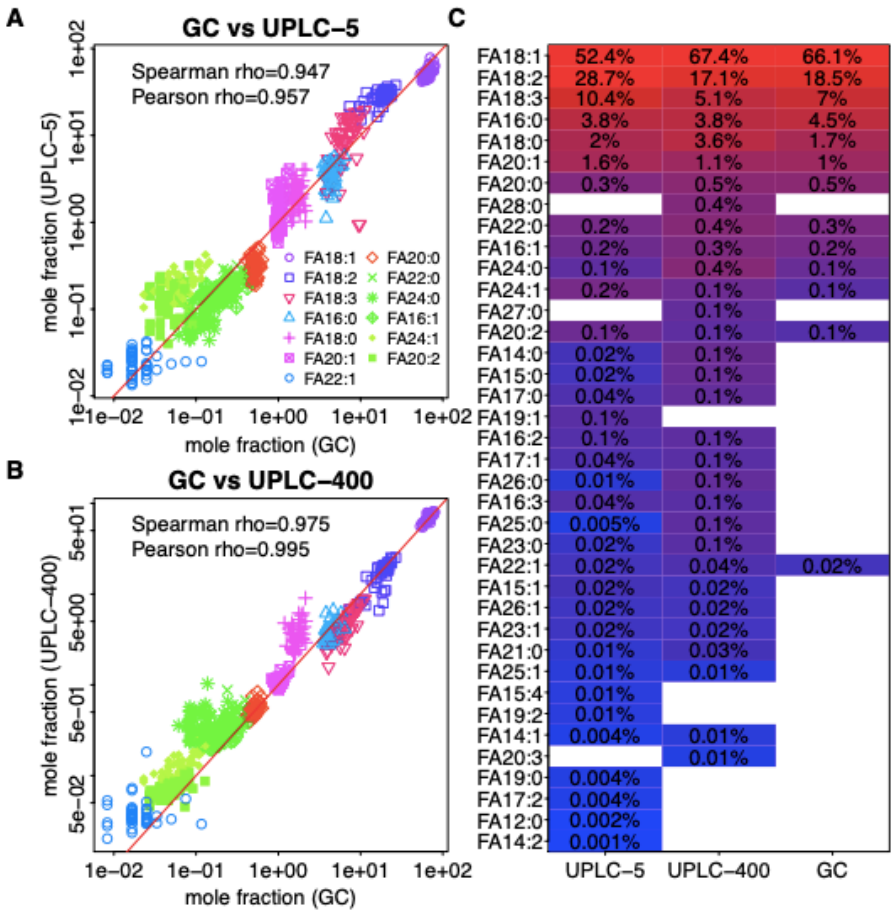


Figure 14. **Dependency between the mass fractions of FAs in rapeseed (calculated for 11 FAs detected by GC-FID) estimated using different methods**. GC-FID vs. UPLC-5 and GC-FID vs. UPLC-400 are shown in (A) and (B), respectively. Each dot corresponds to a given FA (indicated by color) in a given sample (C) Mean mass fractions (relative to the total intensity of 11 FAs detected by GC-FID) for all FAs determined using the three techniques. White rectangles show FAs that were not detected by the indicated method. FAs are ordered according to their intensities, as obtained by UPLC-5.

5.2.2 UPLC-MS data

For UPLC-MS fatty acids profiling, we used two different buffer systems: the UPLC buffer with formic acid and the buffer with acetic acid. The former buffer is the most common buffer used in UPLC-MS experiments. Ionization is more effective in the second buffer since acetic acid is a weaker acid, so a lower sample amount, and, therefore, higher sample dilution is required. For this reason, for the system with acetic acid, we used higher dilution (1:400), and for the system with formic acid, we used lower sample dilution (1:500).

Hydrolyzed extracts of all lines of both sunflower and rapeseed were analyzed using both UPLC-MS-based methods.

At low dilution (UPLC-5), we can detect 29 FAs for sunflower and 35 FAs for rapeseed. At high dilution (UPLC-400), 25 and 31 FAs were detected for sunflower and rapeseed, respectively (The data can be downloaded from DOI: 10.7287/peerj.preprints.27501v1/supp-2). Hence, lower dilution rates proved to be a more efficient approach that allowed to reveal the highest number of FAs. However, in this case, it becomes impossible to perform the relative quantification of major FAs: 18:1 and 18:2, because they cause detector saturation (Figure 15A and B). Apart from even-chain FAs, UPLC-MS also revealed odd-chain FAs (for example, 17:0, 17:1, and 17:2). FAs with an 18-carbon chain (stearic, oleic, and linoleic) were the most abundant. A net-like pattern on the M/z-RT plots can be observed (Figure 16A and B; 17A and B).

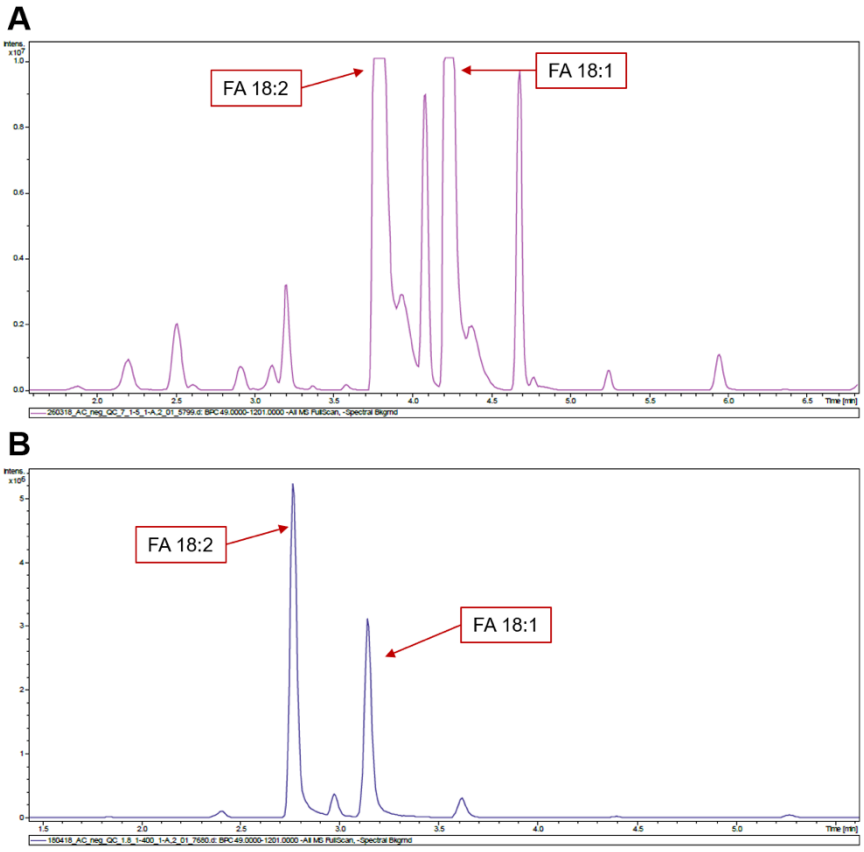


Figure 15**. LC chromatograms were obtained for sunflower.** (A) UPLC-5. (B). UPLC-400. Chromatographic peaks corresponding to 18:1 and 18:2 FAs stay out of dynamic range in (A). Higher dilution rates aid in resolving this problem (B).

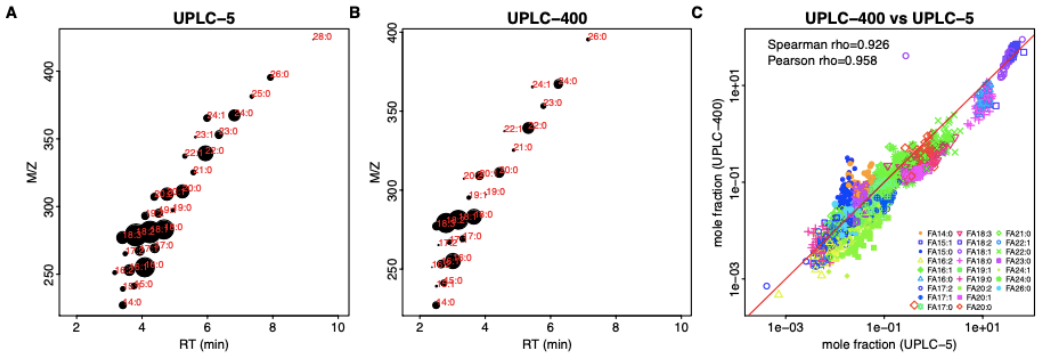


Figure 16. **Comparison of different techniques for quantitative assessment of FAs in sunflower**. (A, B) Retention time (x-axis, RT)–m/z (y-axis) scatter plots for UPLC-5 (A) and UPLC-400 (B). Each dot corresponds to the individual FA; the dot size indicates mean log-intensity. The red color text indicates chain length and number of double bonds. (C) Dependency between the mole fractions estimated using different methods, UPLC-5 vs. UPLC-400. Pearson correlation was calculated between logs of mole fraction.

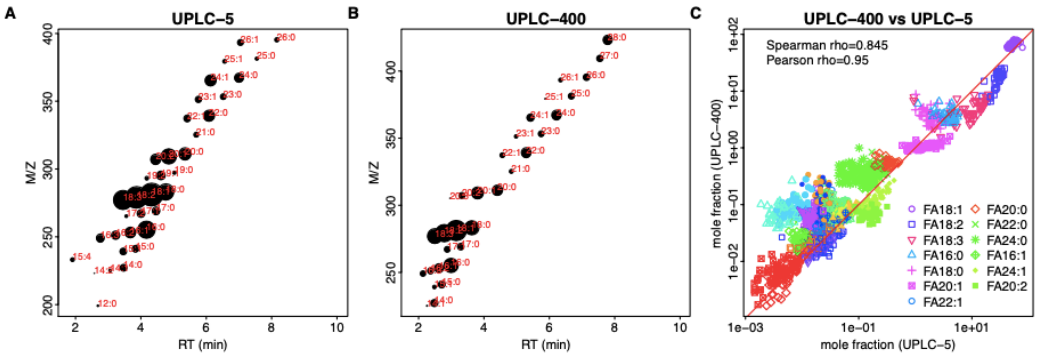


Figure 17**. Comparison of different techniques for FA quantification in rapeseed**. (A, B) Retention time (x-axis, RT)–m/z (y-axis) scatter plots for UPLC-5 (A) and UPLC-400 (B). Each dot corresponds to the individual FA; mean log-intensity is indicated by the dot size. The red color text indicates chain length and number of double bonds. (C) Dependency between the mass fractions estimated using different methods, UPLC-5 vs. UPLC-400. Each dot corresponds to the individual sample.

FA content shows significant variability between the lines (Annex for chapter 5, Figure S1). Results of both UPLC-MS methods show a strong correlation: Spearman rho = 0.927 and 0,854 for sunflower and rapeseed, respectively (Figure 16C and 17C)

5.2.3 Comparison of UPLC-MS with GC-FID

Results of UPLC-MS were compared with the results obtained using GC-FID. All 11 FAs detected in sunflower and 13 FAs detected in rapeseed were also identified using UPLC-MS. In both plants, UPLC-MS proved to be more sensitive and detected about 2.5 times higher number

of FAs than GC-FID (Figure 13C and 14C). GC-FID misses the majority of the minor FAs (mass fraction below 0.5%). The longest FA detected by GC-FID
was the FA with a 24-carbon chain, whereas UPLC-MS provided insight into the changes in longer FAs with the chains containing up to 28 carbon atoms.
As shown in Figures 13 A and B, 14 A and B, although the relative amounts of FAs measured by UPLC-MS and GC-FID are different, there is a significant correlation between them. Spearman rho = 0.908 and 0,918 for sunflower GC-FID and UPLC-5 and GC-FID and UPLC-400, respectively. Spearman rho = 0.979 and 0,947 for rapeseed GC-FID and UPLC-5 and GC-FID and UPLC-400, respectively.

5.3 Results: UPLC^{-}MS Triglyceride Profiling in Sunflower and Rapeseed Seeds

UPLC–MS measured triglycerides in 50 sunflowers and 48 rapeseed lines (Annex for chapter 5, Table S2). After the obtained measurement results were filtered, 34 TAGs in common between sunflower and rapeseed were identified (Annex for chapter 5, Table S3A and B, raw TAGs intensities available: https://www.mdpi.com/2218-273X/9/1/9/s1). The results are presented in Figures 18A and B. The total amount of double bonds per TAG varied from 1 to 7. All TAGs showed variability in their intensity. The most intense TAGs in sunflower were 54:3, 54:4, 54:5, 54:6, 54:7, 52:2, 52:3, and 52:4 and in rapeseed, 54:3, 54:4, 54:5, and 54:6. Based on their *m*/*z* and retention times, all TAGs may be represented as a net-like pattern. The net-like pattern obtained in this work corresponds with the results obtained previously by Hummel et al. (245). It can be seen in Figure 18 that TAGs with the same FA chain length lie on the same line, and all identified TAGs were taken together to form parallel lines on the graph. From bottom to top, these lines contain lipids with the increasing fatty acid chain length, while the number of double bonds decreases from left to right.

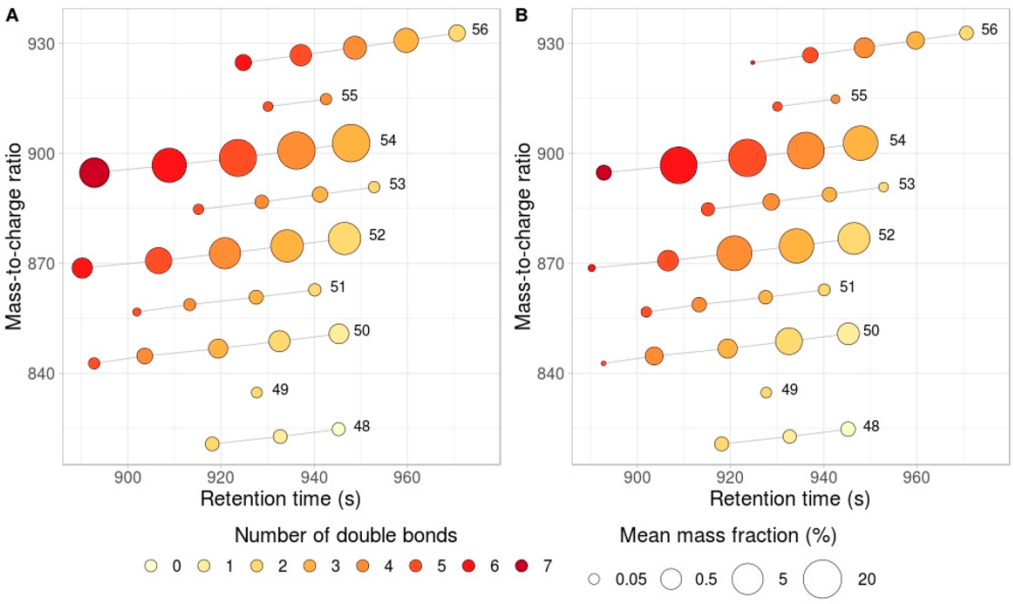


Figure 18. **Retention time –mass to charge plot for (A) rapeseed and (B) sunflower.** Single dot corresponds to a single triglyceride, retention time (seconds), and m/z are shown on the x- and y-axis, respectively. The color scale represents the number of double bonds. Circle size represents the mass fraction. Grey lines connect TAGs with the same fatty acids carbon chain length (indicated at the right end of each line).

5.3.1 Sunflower versus Rapeseed

The intensities of TAGs in sunflower were compared with those in rapeseed. Multidimensional scaling (MDS) analysis revealed apparent clustering based on all 34 TAGs common for the two plants (Figure 19A). Among them, 31 TAGs exhibited significant differences between sunflower and rapeseed (Annex for chapter 5, Table S3A; Figure 19B). Triglycerides 52:2, 52:5, 52:6, 54:3; 54:4, 54:7, 56:3, 56:4, and 56:5 showed the highest variability levels between sunflower and rapeseed with the higher presence in rapeseed. Triglycerides 50:2, 52:3, 52:4, 54:5, and 54:6 also showed substantially high variability but were the most abundant in sunflower.

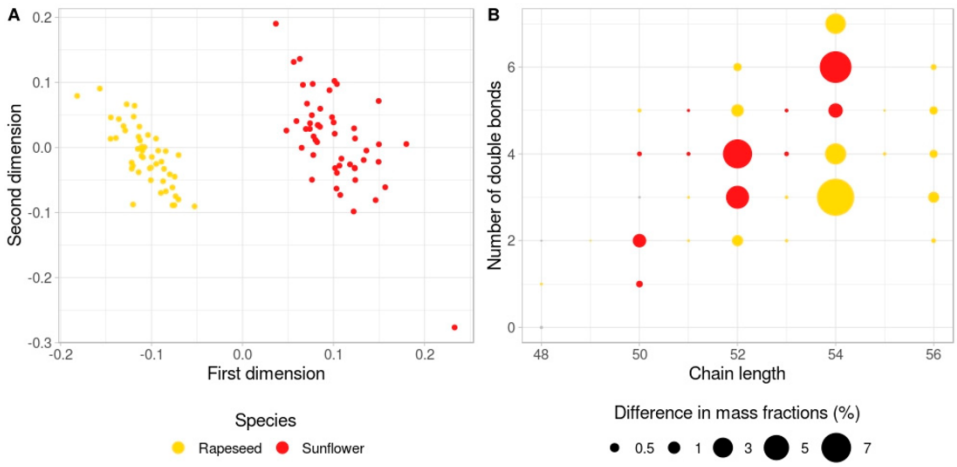


Figure 19. **Triglyceride comparison between sunflower and rapeseed**. (A) Multidimensional scaling plot (one minus Spearman correlation coefficient was used as the distance, two dimensions) of samples demonstrating clear segregation of rapeseed and sunflower samples. (B) The difference in the TAG mass fractions between rapeseed and sunflower. The color indicates which species possess a higher amount of a certain TAG. Circle size represents the absolute difference in mass fractions. Grey circles correspond to the TAGs that do not show statistically significant differences

Control samples were analyzed in the MS2 scan mode to reveal the precise FA content in TAGs. This allowed us to carry out a more detailed comparison between the sunflower and rapeseed TAGs. Triglycerides 54:3 is the most abundant in rapeseed. According to the MS2 results, this TAG contains three FAs 18:1 (oleic acid) (fragmentation pattern is depicted in Figure 20A). The most abundant in sunflower are TAGs 52:4 and 54:6, which according to the MS2 data, contain two linoleic (18:2) and one palmitic (16:0) acids and three linoleic acids, respectively (fragmentation pattern is depicted in Figure 20A and C. Linoleic acid is the most abundant fatty acid in sunflower on the whole.

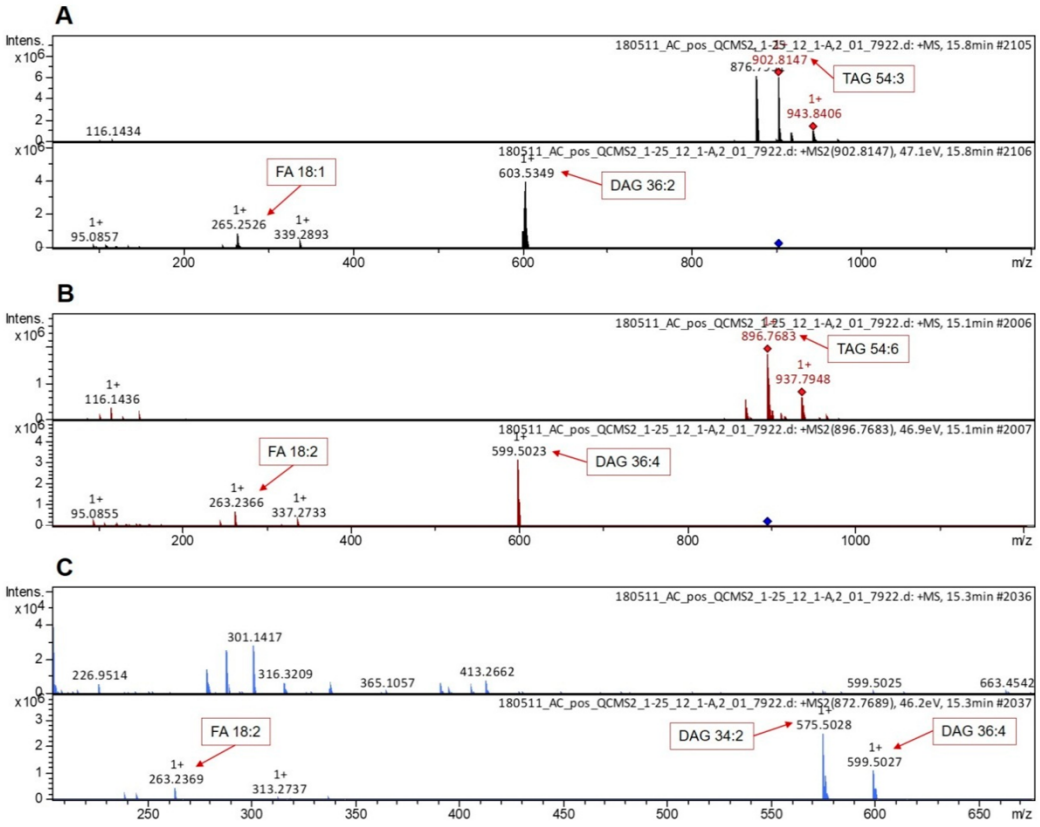


Figure 20**. MS2 fragmentation spectra with fragment annotation.** (A) Fragmentation spectrum for TAG 54:3. (B) Fragmentation spectrum for TAG 54:6. (C) Fragmentation spectrum for TAG 52:4. DAG: Diglycerides.

5.3.2 Spring-Type versus Winter-Type Rapeseed

Triglycerides intensities in spring-type and winter-type rapeseed lines were compared. Multidimensional scaling analysis revealed differences between these two groups (Figure 21A), with 17 TAGs demonstrating significant variability between the spring and winter lines (Annex for chapter 5, Table S3B, Figure 21A).

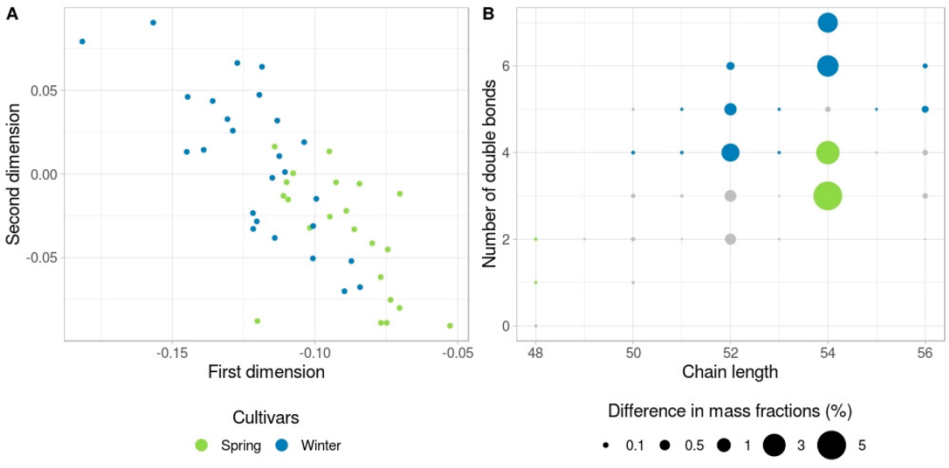


Figure 21**. TAG comparison between winter and spring rapeseed**. (A) Multidimensional scaling (MDS) plot of rapeseed samples (one minus Spearman correlation coefficient was used as the distance, two dimensions) demonstrating clear segregation of winter and spring rapeseed samples. (B) The difference in the TAG mass fractions between winter and spring rapeseed. The color indicates which cultivar possesses a higher amount of certain TAG. Circle size represents the absolute difference in mass fractions. Grey circles correspond to the TAGs that do not show statistically significant differences.

5.4 Discussion

FA composition of seed oil from 50 sunflower and 50 rapeseed lines was analyzed by GC-FID, which is traditionally used to measure FAs in plant oils, and two UPLC-MS-based approaches (1:5 dilution in the buffer system with formic acid (UPLC-5) and 1:400 dilution in the buffer system with acetic acid (UPLC-400)). GC-FID technique allowed to detect 11 and 13 FAs in sunflower and rapeseed, respectively, all of them representing even-chain FAs.

GC-FID based FAs abundances in sunflower and rapeseed are in good agreement with those obtained in previous investigations (44, 291, 292, 301, 302).

UPLC-MS is a more sensitive technique compared with GC-FID, so it was not surprising that we detected a considerable number of additional components in the FA profiles obtained using this method. This result corresponds well with the data obtained by (247) for Arabidopsis thaliana tissues.

In total, about 29 and 35 FAs were detected in sunflower and rapeseed samples, respectively, by UPLC-MS. It is worth noting that utilizing the UPLC-5 approach, we were able to identify significantly more FAs in both crops compared with the UPLC-400 method. However, with the 1:5 dilution implemented in this approach, it was impossible to perform the relative quantification of the two most abundant FAs, 18:1, and 18:2, due to detector saturation. Because of this, we suggest that the UPLC-400 approach is more suitable for FA profiling of oil crop samples.

Our study demonstrated that the UPLC-MS method is appropriate for the detection of long FAs both in sunflower and rapeseed. The longest FA detected by the GC-FID method was the FA with a 24-carbon chain, whereas UPLC-MS revealed FAs with tails up to 28 carbon atoms long. Generally, about half of the FAs identified by UPLC-MS belong to Very-long-chain fatty acids (VLCFA, fatty acids with the chain length of at least 20 carbon atoms). According to the previous investigations, VLCFAs are mainly located in the cuticular wax layer deposited at the surface of aerial plant organs; they form part of the triacylglycerides of seed oil and

sphingolipids and are essential for many aspects of plant development and play a role as signal molecules governing both biotic and abiotic stress (303,304).

Apart from even-chain FAs, odd-chain fatty acids were also detected by UPLC-MS. The latter are present in the oil extracted from the analyzed plants only in minor quantities, and selection for these FAs is not currently included in breeding programs. However, the beneficial effects of these compounds on human health have been recently demonstrated. For example, pentadecenoic and heptadecenoic acids contribute to reduced risks of developing multiple sclerosis and act as anti-inflammatory and edema-inhibiting agents (305,306). Additionally, odd-chain fatty acids inhibit developing certain plant pathogens and could be used as precursors for manufacturing agricultural and industrial chemicals (307,308,309,310). The chemical properties and potential biological activities of odd-chain fatty acids are continuously under investigation (311). Due to the importance of odd-chain fatty acids, efforts are being made to produce yeast strains with increased content of FAs of this kind (312).

Therefore, the analysis of minor FA content in oil may be important to make a complete assessment of the oil's functional and nutritional properties. Our results suggest that UPLC-MS has excellent potential as a precise tool for evaluation of the full FA profile in oilseed crops. It can be essential to create vegetable oils with increased nutrition value and new technical characteristics and provide additional markers for agronomically important traits in plants. Considering the advantages of UPLC-MS and its applicability for FA profiling in oilseed crops, we may observe that the overall result can be represented as a net-like pattern on the retention time M/z-RT plots (Figure 16A, 16B, and 17A, 17B). This makes the process of results interpretation and annotation easier compared to GC-FID results.

Another advantage of UPLC-MS is the possibility of analyzing thousands of samples per month and a small amount of plant material needed for the analysis. This technique requires only 5–10 mg of plant material per extraction, while GC-FID requires high amounts of material. For plants with big seeds like sunflowers, it is possible to take only a small part of the seed for the FA profiling analysis and germinate the rest of the seed and plant it, which allows exact assignment of phenotype to genotype in breeding programs.

It is also worth mentioning that the UPLC-MS technique involves no FA derivatization, which allows identifying more FAs compared to the conventional GC-FID approach. The present study results confirm this, and they are in good correspondence with the data obtained by Bromke et al. 2015 (247) for Arabidopsis.

It is essential to highlight that specific differences between the results obtained by GC and UPLC-MS may be connected with the lipid extraction procedure. We used MTBE extraction, which extracts both TAGs and phospholipids, which means that some detected FAs may come

from phospholipids, compared to GS, where extraction was performed with hexane, which typically extracts mostly TAGs.

Currently, gas chromatography with mass spectrometric or flame ionization detection is the gold standard for quantitative assessment of vegetable oils' FA composition. UPLC-MS has excellent potential to be used in the evaluation of the FA composition of oil crops as highly sensitive, scalable, and suitable for the individual seed analysis technique.

Taken together, our findings suggest that UPLC-MS provides an in-depth insight into the oil FA content and may be applied for precise identification of FA profiles of oilseed crops.

Then we implemented UPLC-MS for TAG profiles comparison in sunflower and rapeseed samples. The analysis revealed the similarity in the content of the most intense TAGs 54:3, 54:4, 54:5, and 54:6 between these two plants and also allowed to find significant differences in the TAG amounts between sunflower and rapeseed.

The most exciting results were obtained during the comparison of winter and spring rapeseed varieties. TAGs with a higher amount of double bonds (52:4, 52:5, 52:6, 54:6, and 54:7) are more abundant in the seeds of winter lines, while the TAGs with a higher level of FA chain saturation (48:1, 48:2, 54:2, and 54:3) are more abundant in spring lines. These results may most likely be accounted for by the chemical properties of FAs. The degree of saturation is highly essential for the FA crystallization processes. Triglycerides containing double bonds have significantly lower melting points than completely saturated TAGs (313).

Winter-type rapeseed seeds were shown to contain TAGs with a lower degree of saturation, which probably impacts winter-type rapeseed cold resistance. Freezing tolerance is one of the most crucial plant traits allowing them to survive in the low temperatures. Freezing tolerance correlates well with winter survival (314). The unsaturated fatty acid content of the plasma membrane is associated with cold resistance in plants. Plants with a higher content of unsaturated fatty acids in their membranes are more resistant to cold (315). Storage lipids, TAGs, are produced by the extension of the membrane-lipid biosynthetic pathway. That is why in the majority of plants, TAGs found in most seeds usually contain the same acyl groups as those found in membrane lipids. Simultaneously, while membrane lipid composition is highly conservative across plant species, the variability in the fatty acyl chains found in the seed oil is very high (316). A broad range of factors affects freezing tolerance in *Brassica* species with all classes of macromolecules being involved, lipids among them (317). The FA content patterns detected in the seeds of winter-type and spring-type rapeseed maybe just a footprint of the total membrane lipid content specific for the winter-type and spring-type plants, since during the germination period in oilseed plants, storage lipids, mainly TAGs, are catabolized with polar lipids, phospholipids, and galactolipids being synthesized de novo (318). But it cannot also be

excluded that TAGs play their specific role in the seeds of winter-type rapeseed. Since winter-type rapeseed lines start their germination late in the autumn and resume growth early in the spring, storage lipids with lower melting points may enable their development at relatively low temperatures and allow them to activate their metabolism early in the spring.
The observed differences in the saturation levels of TAGs in the winter-type and spring-type rapeseed may provide new insights into the cold tolerance mechanisms in plants, which is highly important in terms of global climate change.

Chapter 6. Genotyping and lipid profiling of 601 cultivated sunflower lines reveals novel genetic determinants of oil fatty acid content

6.1 Introduction and contributions

This chapter combined GBS sequencing data and the UPLC-MS profiling data to find SNPs associated with specific lipid phenotypes. Many oil properties, as well as its quality, are determined by seed lipid composition. That's why lipid profiling in oilseed crops and understanding genetic bases of specific lipidomic patterns are of particular interest. The data set in this study consists of 601 inbred lines, including highly inbred lines (8-25 generations of inbreeding) from the Vavilov collection. We were the first who have implemented UPLC-MS technology for lipid phenotyping of sunflower that allowed us to capture even very minor fatty acids in sunflower seeds and identify the genome regions, which have never been previously reported in the frame of association with these fatty acids. We have identified an extended region in chromosome 3, which is associated with a couple of minor fatty acids and contains many candidate genes for oil improvement. The results presented in this chapter mostly cover just FAs and TAGs. In the association studies, we used only FAs. But there is still much to analyze in future research.

.

This work was performed in strong collaboration with Prof. Nuzhdin's group at the University of Southern California. I led this project and was responsible for the whole spectrum of work, including seed germination, DNA extraction, NGS library preparation, lipid extraction, lipidomic profiling, data analysis, figure preparation, and paper writing and editing. However, many people contributed to this work. Members of Prof. Nuzhdin's group made considerable contributions to data analysis. Peter Chang did all work on primary sequencing data processing, including reads mapping, SNP calling, and quality control. Anupam Singh and Katrina Sherbina helped me with learning GWAS and metabolomic data processing. Skoltech members Pavel Mazin and Rim Gubaev helped with data analysis and figure preparation. Nikolay Anikanov helped a lot with lipid extraction. Elena Martynova and Stepan Boldyrev helped with DNA extraction, Svetlana Goryunova is one of the experimental design authors. Yulia Karabitsina from VIR participated in library preparation.

6.2 Results

6.2.1 GBS sequencing and SNP calling

We extracted DNA from inbred sunflower lines from the Vavilov seed bank, VNIIMK Applied Agricultural Institute, and Agroplasma Breeding Company collections (Annex for Chapter 3). Two to three technical replicates of each sample were sequenced on the Illumina HiSeq 4000 platform using a GBS protocol (see Chapter 3), resulting in 1490 genotypes. Reads were mapped onto the *Helianthus annuus* reference genome (HanXRQr1.0), with the mapping rates varying between 75 and 90%. Variant calling identified 2,360,111 single nucleotide polymorphisms (SNPs) spanning all 17 chromosomes. Homozygosity and Principal Component Analysis (PCA) showed no obvious bias with respect to plate batch or seed bank variables (Annex for chapter 6, Figure S1)

6.2.2 Population structure, relative kinship, and linkage disequilibrium

We assessed the populational structure of the genotypes used in GWAS analysis using the ADMIXTURE package. No visible clusters were observed for cases of K=1:10 (Figure 22A). However, visualization of genetic variation using the first two principal components of PCA revealed a distinct group of genotypes derived from the Agroplasma collection, which clustered separately (Figure 22B). While the average genotype correlation (r^2) dropped to half of its maximum value at 0.7 Mb, the linkage disequilibrium (LD) decay varied among the 17 chromosomes (Figure 22C, D; Annex for chapter 6, Figure S2). Notably, some chromosomes, such as chromosome 3, demonstrated extended LD within the 1-3 Mb interval (ANOVA, $p < 0.0001$).

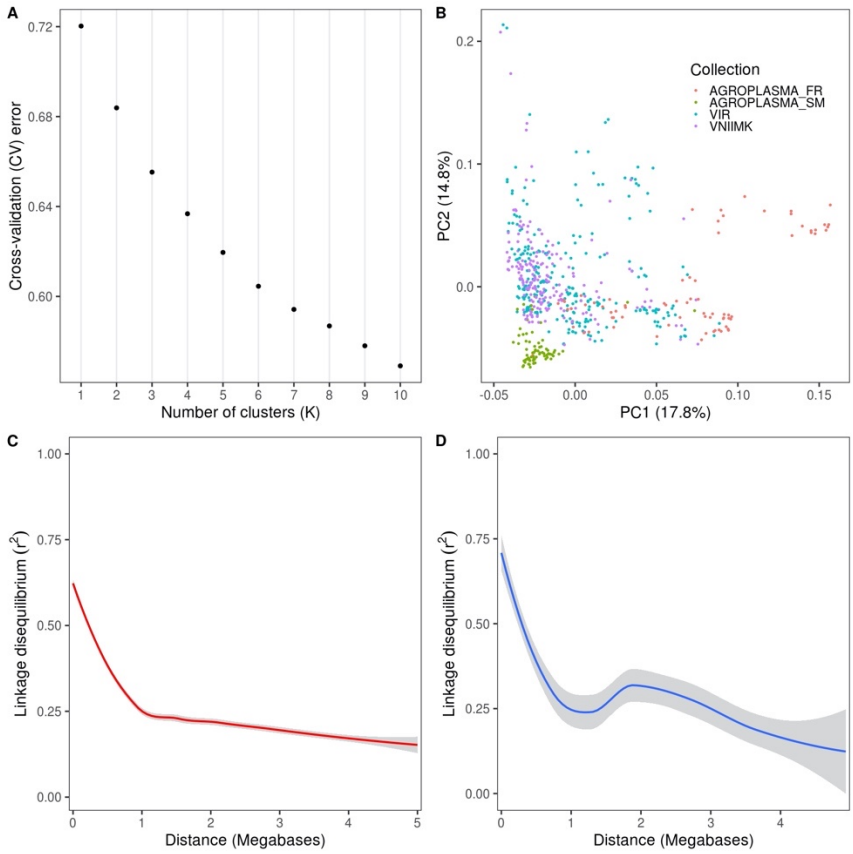


Figure 22. **Population structure of germplasm and linkage disequilibrium (LD) values.** (A) Estimated cross-validation error value for possible cluster numbers from 1 to 10. (B) Subpopulations were assessed using Principal Component Analysis. Each dot corresponds to a sunflower accession used in the study. Color corresponds to sunflower lines from different collections. Agroplasma_SM indicates sterility maintaining lines from Agroplasma; Agroplasma_FR indicates fertility restorer lines. (C-D) Genome-wide (C) and per-chromosome 3 (D) LD-decay. Lines correspond to loess curves.

6.2.3 Genotypes variability and relation to other sunflower germplasms

To place analyzed cultivars on a broader map of sunflower genotype variation, we compared our genotypes to previously sequenced 1065 wild sunflower varieties, 20 landraces, and 289 cultivated sunflower lines (145). Principal component analysis just on cultivated lines and landraces based on 2345 SNPs shared between the datasets showed that cultivated sunflower lines from the Russian dataset are clearly distinguishable from those collected worldwide by the third principal component (Figure 23A and B). The analysis further reaffirmed the broad genetic difference between cultivated and wild material (Annex for chapter 6, Figure S3A). However, it has to be noted that such analysis that is confined to the positions polymorphic in both datasets could, therefore, underestimate the differences between the datasets. The third principle component further separated some of the *Helianthus* species (Annex for chapter 6, Figure S3B).

Notably, the landraces present in the Hübner dataset mostly situated between the cultivated lines from foreign and Russian collections and the wild sunflower varieties (Annex for chapter 6, Figure S3A, S3B).

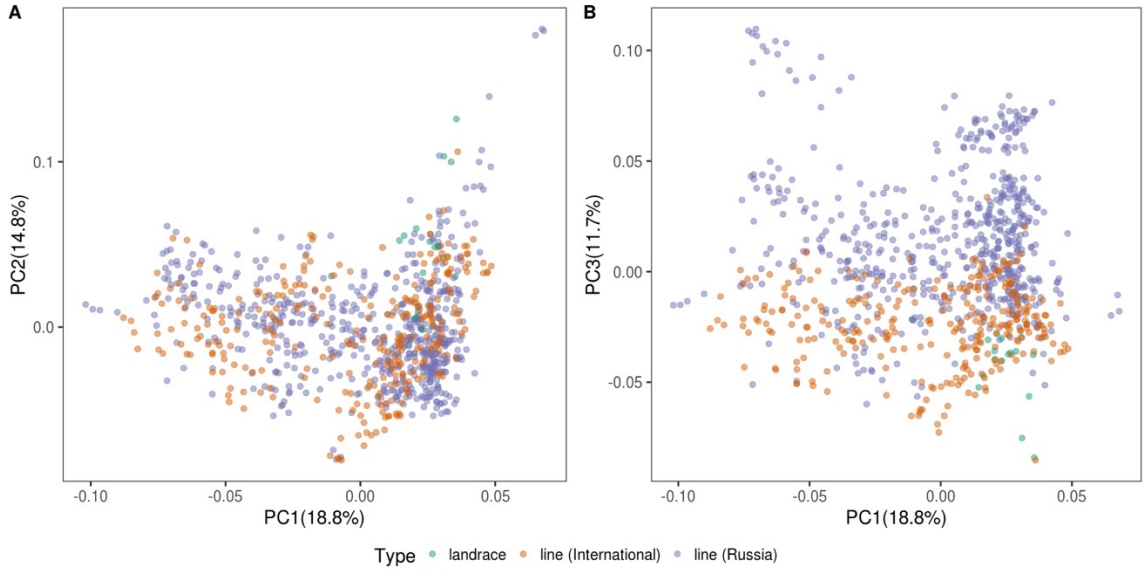


Figure 23**. The relationship between sunflower germplasm of different origins is estimated based on 2345 SNPs shared between this and the Hübner (2019) studies**. (A) The first and second components of the PCA. (B) The first and third components of the PCA. Each dot corresponds to a plant accession. Colors indicate the origin

6.2.4 Oil lipidome quantification

We extracted the total lipid fraction from sunflower seeds of same sunflower lines used in the genotype analysis. We then divided the lipid extracts into two fractions and analyzed them independently using UPLC-MS technology. The first fraction was kept intact, while the second was hydrolyzed before the analysis. The hydrolyzed fraction contained fatty acid residues of all oil lipids, as well as a minor fraction of free fatty acids present in intact samples before hydrolysis (FAs). Mass spectrometry analysis yielded 826 computationally annotated lipid peaks and 27 post-hydrolysis fatty acids. In our further analysis of intact lipids, we focused on a specific lipid class, the most important among sunflower oil lipids, the triacylglycerides (TAGs). To optimize the detection of both high and low abundancy FAs and TAGs, we conducted the UPLC-MS measurements at two extract dilutions (see Chapter 3).

6.2.5 Quantification of genetic and environmental effects on oil lipidome composition

To assess FA and TAG data environmental and biological reproducibility, we grew plants from six sunflower inbreed lines (1 conventional and 5 high oleic) originating from the VNIIMK collection for three years with five biological replicates per year, yielding a total of 89 accessions (Annex for Chapter 3). We conducted genotyping using the same GBS protocol and UPLC-MS measurements at extract dilutions to ensure quantitative coverage of the entire concentration range. We then tested the effects of the genotype-environment interaction using ANOVA with the following model: G + E + G:E (where G is genotype and E is environment, *i.e.,* year). All FAs and TAGs measured in both dilutions displayed significant differences between genotypes after BH-correction ($p < 0.05$, Figures 24A; Annex for chapter 6, Figures S4A, S5A, S6, S7, S8, Table S1). Further, the majority of FAs (11 out of 15) and TAGs (32 out of 42 and 43 out of 59 for 1:25 and 1:3 dilutions, respectively) also showed significant G:E interaction. However, the interaction effect, although statistically significant, had a much smaller amplitude than the impact of the genotype (Figure S9). Biological replicates of the same genotype collected in different years displayed more remarkable similarity than plants of different lines collected in the same year. The most substantial variation among genotypes was observed for oleic, linoleic, and palmitic acids, the major fatty acids in sunflower oil (Figure 24B-E), as well as for the following TAGs: 50:2, 51:3, 54:3, 54:4, 54:6 (Annex for chapter 6, Figure S4 B-E and S5 B-E).

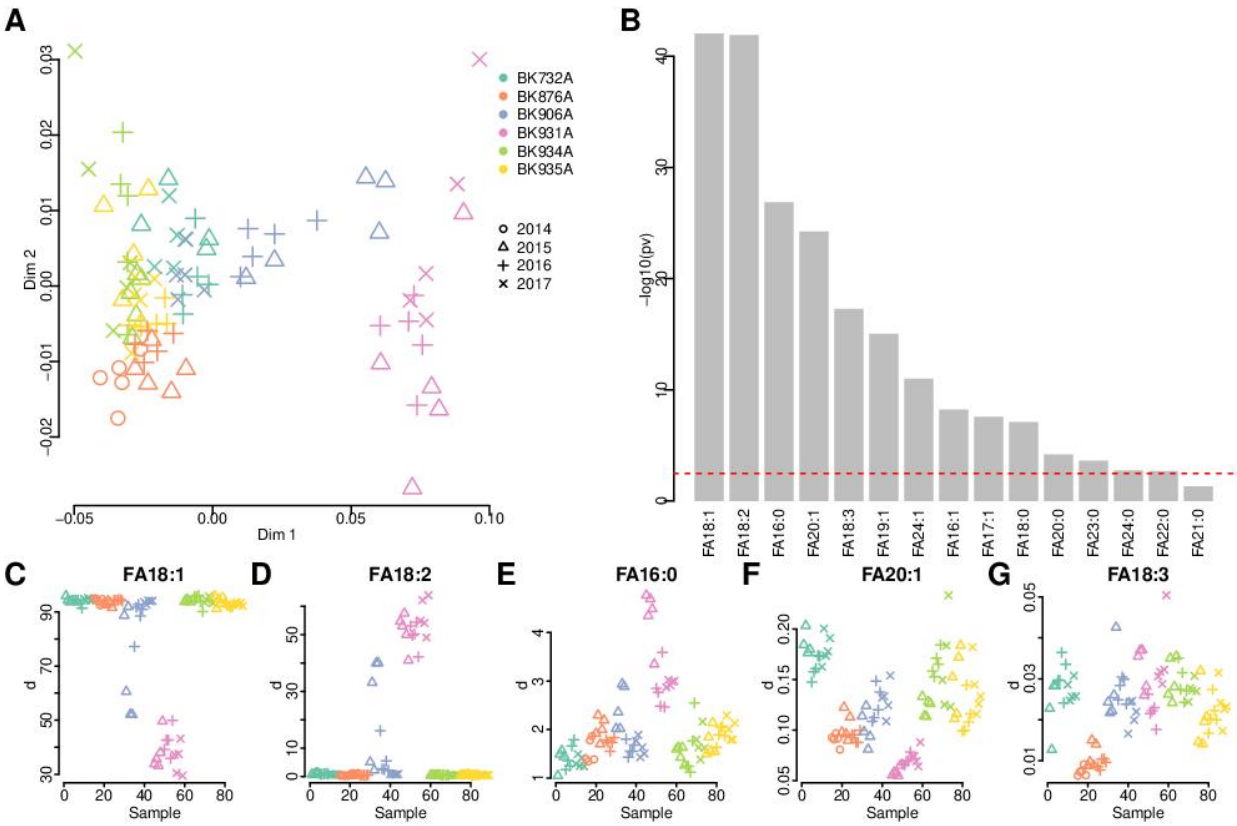


Figure 24**. FAs concentrations in replication experiments. (A) MDS plot (two dimensions, 1 - Spearman correlation coefficient between FAs concentrations was used as distance). One sample is shown by one point; different colors show lines, different years are shown by points of different shapes.** (B) Minus log10 p-values for the differences between lines (ANOVA) are shown. Bonferroni adjusted 0.05 significance level is shown by red line; (C) Linoleic acid (18:2); (D) Oleic acid (18:1); (E) Palmitic acid (16:0), (F) Eicosenoic acid (20:1), (G) Linolenic acid (18:3). Each point represents 1 sample, point shapes, and colors as in (A).

6.2.6 Oil lipidome variation analysis

Computational annotation of intact lipidome of the oil samples extracted from 601 sunflower lines yielded 687 lipids falling into seven lipid classes: glycerolipids (GL), glycophospholipids (GP), free fatty acids (FA), sterols (ST), prenols (PR), polyketides (PK), and saccharolipids (SP) (Figure 25A). A subclass of glycerolipids, TAGs, occupied 87% of lipid intensities of uniquely identified compounds (Figure 25B). The most present TAGs were 54:6, 54:5, 54:4, 54:3, 52:4, 52:3, and 52:2 (Figure 26A). Among computationally annotated 27 FAs, the highest abundance FAs were 18:1, 18:2, 16:0, and 18:0 (Figure 26B). The statistics on each fatty acid are presented in Table S2 (Annex for chapter 6).

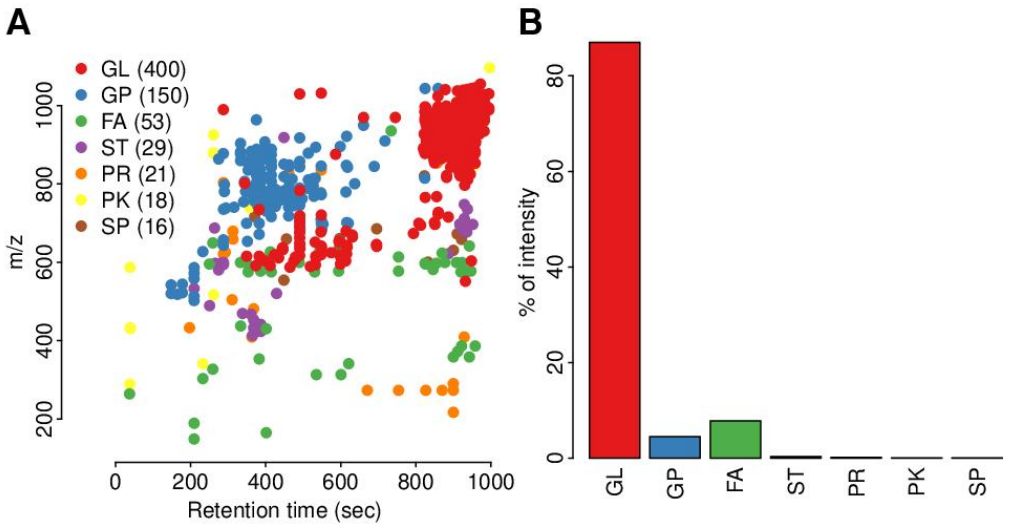


Figure 25. **Lipid annotation** (A) Mz/rt plot. One point represents one peak; different lipid categories are shown in different colors. Only peaks with sample intensities at least two times higher than blank intensities are shown. (B) Relative intensities of all lipid categories. The intensity of the given category was calculated as the sum of intensities of all lipids of the category. GL- glycerolipids, GP- glycophospholipids, FA -fatty acids, ST- sterols, PR- prenols, PK- polyketides, and SP-saccharolipids.

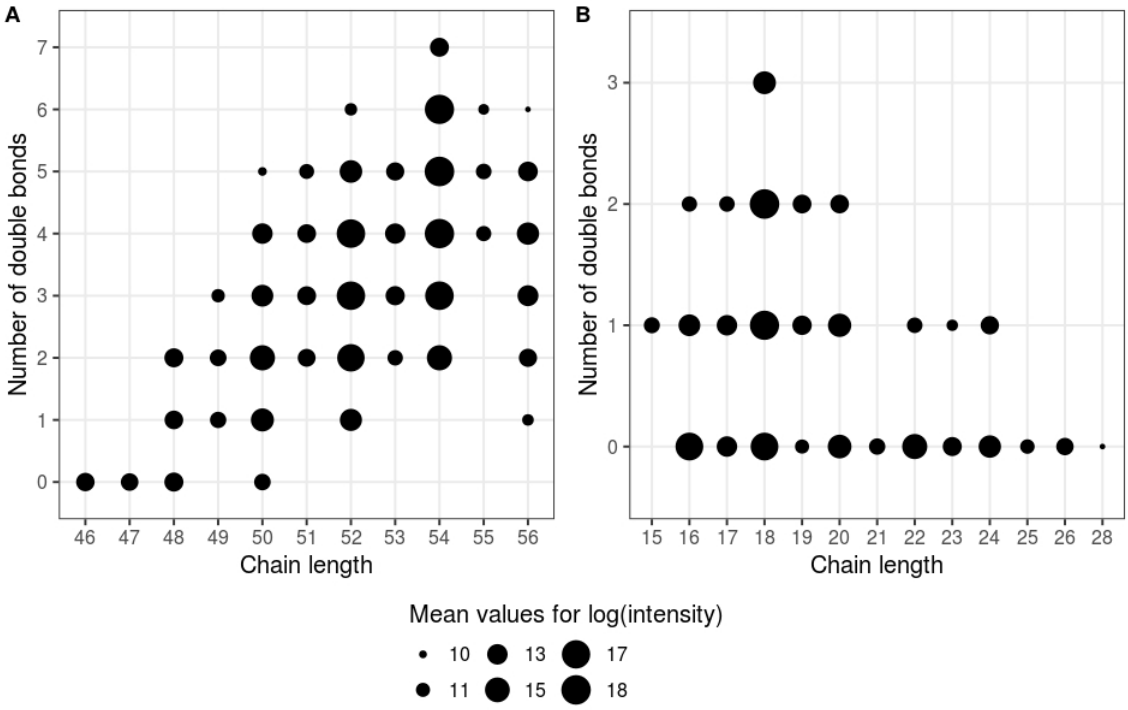


Figure 26. **Shematic representation of fatty acid properties (fatty acid chain length and degree of saturation) for detected lipids**. (A) Cumulative chain length and double bonds number of three fatty acid residues composing detected TAG molecules. (B) Chain length and double bonds number of fatty

acid (FAs) released after lipid hydrolysis. Each circle corresponds to a FA or a TAG. The circles' size corresponds to the mean relative amount of this molecule in a sample (log-transformed MS peak intensity).

6.2.7 Association analysis

Of the 601 sunflower lines taken into the study, we obtained both genotype and lipid intensity data for a total of 543 accessions. We conducted a GWAS analysis using the mixed linear model (MLM) approach to test for genetic determinants of FAs variation based on these data. The analysis included 15068 SNPs that passed the filtering criteria (missing calls rate < 0.3, DP > 4, MAF > 0.01) for oleic and linoleic acids and 12528 SNPs for other fatty acids (missing calls rate < 0.3, DP > 4, MAF > 0.03). From 27 detected FAs 23, satisfying the criteria for GWAS were selected. Of 23 analyzed FAs, we detected significant associations for eleven: stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2), nonadecanoic acid (19:0), eicosanoic acid (20:0), docosanoic acid (22:0), tetracosanoic acid (24:0), tetracosenoic acid (24:1), hexadecadienoic acid (16:2) and such rare fatty acids as 17:2 and 19:2 (MLM, Bonferroni-corrected $p < 0.00001$; Figure 27; Annex for chapter 6,, Figure S10A-F). We further performed GWAS for the oleic-linoleic acid ratio yielding six significant SNPs as response variables in MLM (Bonferroni-corrected $p < 0.00001$; Figure S10G-I). Altogether, we identified 140 trait-associated SNPs (MLM, Bonferroni-corrected $p < 0.$ 0.00001; Figure 27A). Among them, docosanoic acid (22:0) abundance variation showed the strongest association with 53 genotype variants located on chromosomes 3 and 14 (Annex for chapter 6, Table S3). These genetic variants cumulatively explain up to 35.4% of the quantitative variation of docosanoic acid abundance among sunflower lines.

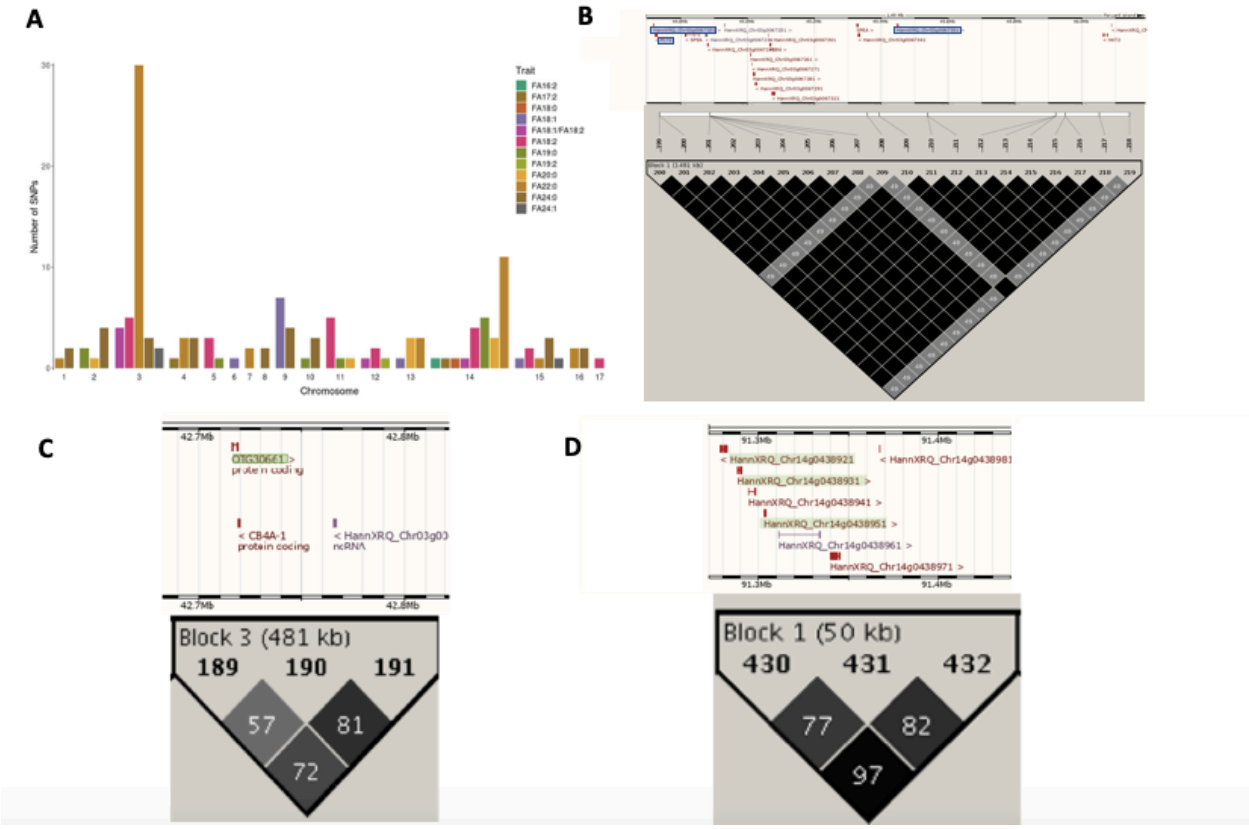


Figure 27. **GWAS results for FAs in Sunflower lines and candidate genes for docosanoic acid improvement**: (A) Cumulative plot representing the number of significant associations for each of all traits. Traits are represented by colors. Chromosome number and number of SNPs are presented on the X and Y-axes, respectively. (B) LD block in Chr3 (Location 44696624 - 46188263). (C) LD block in Chr3 (Location 42596595 - 43078214). (D) LD block in Chr14 (Location 91496885 -91547710). Candidate genes in blue associated with lipid metabolism, Candidate genes in green associated with lipid metabolism described by (Badouin *et al.*, 2017)(10).

6.2.8 SNP annotation and candidate gene identification

To annotate genes potentially linked to genetic variants significantly associated with FAs quantitative variation, we determined the boundaries of the corresponding LD blocks (Annex for chapter 6, Figure S11; Table 3). We then retrieved annotation of all genes located within these LD blocks. We checked their intersection with the genes annotated to be involved in sunflower oil metabolism (10) (Annex for chapter 6, Table S4). From 44144. sunflower genes Badouin *et al.,* 2017 (10) reported 429 genes involved in oil metabolism. Among 124 candidate genes located close to significant SNPs reported in the current study and stayed in LD with these SNPs, four genes coincide with genes from Badouin list, which is significantly more than expected by chance (Fisher exact test, p = 0.03, odds ratio = 3.4). According to the sunflower genome

annotation, these genes encode putative beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ, HotDog domain protein, probable phosphatidic acid phosphatase (PAP2) family protein, and putative MYB-CC type transcription factor, LHEQLE-containing domain and are located on Chr3 and Chr14 (Figure 6C, D).

Table 3: LD blocks with significant association

Phenotype	Chromosome	LD block Location		Length (kb)
		Start position	End position	
Oleic Acid (18:1)	6	64066219	64889534	823
	9	168736699	169306761	570
	13	116940760	117370881	430
	15	38043597	38078709	35
Linoleic Acid (18:2)	3	66733584	68666170	932
	5	37199838	37569381	396
	11	5004818	50619247	414
	11	95051157	92468132	416
Linolenic acid (18:3)	11	43846946	44328722	481
Oleic/Linoleic ratio	3	66733584	68666170	932
	12	121534492	121906701	372
Nonadecanoic acid (19:0)	2	179620148	179872251	252
	14	53394600	53480813	86
	14	59829070	60503626	664
Docosanoic acid (22:0)	3	32332262	32562669	230
	3	42596595	43078214	481
	3	44696624	46188263	1491
	3	48304030	49705352	1401
	3	53949047	54230339	281
	3	57635146	57714809	79
	14	91496885	91547710	50
	14	96632645	97927934	614
	16	176846705	176869659	22
Tetracosanoic acid (24:0)	2	56777868	56880436	102
	2	73398255	74229960	831
	3	102040303	102070280	29
Nervonic acid (24:1)	3	44696624	46188263	1491
	3	57635146	57714809	79

6.3 Discussion

Our results broaden the list of candidate genes and genetic variants associated with sunflower oil lipids' fatty acid composition. Our analysis, based on UPLC-MS mass-spectrometry, included the quantitative measurements of fatty acids present in sunflower oil in minor amounts, which were not previously assessed. Our results indicate that there are genetic loci with a substantial effect on the quantitative phenotype for at least some of these minor fatty acids, such as docosanoic acid (22:0). It requires further research, but there is a possibility of producing sunflower breeds with elevated levels of minor fatty acids in the future.

The reason for the lack of significant signals for 12 of 23 analyzed fatty acids, as well as relatively weak genetic signals for many of the eleven fatty acids with identified associations, could lay in the selection of the analyzed lines. The 543 lines used in the GWAS analysis, as well as all the 601 lines used in our study, were not preselected to contain contrasted phenotypes for fatty acid content, with the exception for oleic and linoleic acids, because until now, only these fatty acids together with stearic and palmitic acids were considered in breeding (63). Nonetheless, our approach involving a large number of diverse cultivated lines yielded enough variability for nine more sunflower fatty acids to produce significant genetic associations. Further work involving lines specifically selected to vary in terms of fatty acid content is required to determine the full scope of genetic associations underlying sunflower oil composition.

We have identified six large LD blocks containing SNPs significantly associated with FA content variation within chromosome 3 (Table 3). Furthermore, among the reported candidate genes predicted to affect oil quality, three genes associated with lipid metabolism localized within the large 1,491 kb LD block of the chromosome 3 (Annex for chapter 6, Figure S11, Table 3). These genes encode the putative phospholipase A2 (this protein releases FAs from the phospholipid), putative CRAL-TRIO lipid-binding domain-containing protein, and putative ethanolamine-phosphate cytidylyltransferase. Predicted functions of these genes, although not yet assessed experimentally in sunflower, single out this genomic region as one of the key regulators of sunflower oil FA composition (Figure 27B, Annex for chapter 6, Table S7). Further, among the genes located within the chromosome 14 region associated with FA 22:0 variation, two were annotated as membrane-bound proteins: putative membrane-bound transcription factor site-2 protease, and putative membrane-bound O-acyl transferase (MBOAT). This finding agrees with the fact that very-long-chain fatty acids in sunflower are synthesized by membrane-bound enzymes (81).

Among the genes essential for fatty acid metabolism according to (10) and located within the LD blocks linked to FA variation, one of the most interesting is the gene encoding a putative FabZ dehydratase, the protein responsible for FA elongation (Figure 27C). It has to be mentioned, however, that the genomic resolution of our study is limited to LD blocks, which typically include multiple genes. Thus, further work is needed to map associations to specific genes and causative genetic variants.

Genetic variants (SNPs) linked to the oleic-linoleic acid ratio also map to a chromosome 3 region (302 kb region; Annex for chapter 6, Figures S10I, S11). This LD block overlaps with the one carrying SNPs significantly associated with linoleic acid content (Annex for chapter 6, Figures S10H, S11). This finding supports the notion that genomic regions underlying linoleic acid content should also be involved in oleic-linoleic acid ratio determination. Unfortunately, no annotated genes were known to be directly related to fatty acid biosynthesis or modification in this region. Previous studies demonstrated that genes encoding desaturases, the major enzymes responsible for the oleic-linoleic acid ratio, are located on chromosomes 1, 12, 14 (75). Nonetheless, loci potentially associated with oleic and linoleic acid contents, were previously identified on chromosome 3 by means of QTL mapping (108,179), as well as by computational predictions (10). These loci, however, did not overlap with the locus obtained in the current study. Several reasons could cause the fact that the previously reported regions potentially related to the oleic-linoleic acid ratio were not identified by association mapping in the present study. First, the SNP coverage for these regions might not have been dense enough in our research. Second, previously identified associations might play lesser roles in determining linoleic and oleic-linoleic acid ratio under Russia's environmental conditions. Third, lack of overlap could be related to the specific genetic features of the studied cohort that was restricted to the lines from the Russian collections.

In addition to genetic variants linked to the oleic-linoleic acid ratio, we have identified nine and 22 SNPs significantly associated with individual oleic and linoleic acid content. These SNPs localized on chromosomes 9, 13, 15 for oleic acid and 3, 11, 12, 14, 15, and 17 for linoleic acid. Previously, a study reported QTLs identified by means of ORS markers for oleic acid content on chromosomes 8 and 9 and linoleic acid content on chromosomes 8 and 14 (109). We also identified significant associations and putative candidate genes on these chromosomes for linoleic acid and on chromosome 9 for oleic acid. However, our chromosome 9 LD block did not overlap with the QTL associated with oleate reported by Badouin *et al.* .2017 (10).

Interestingly, we have additionally identified a putative FAO1 gene on chromosome 9 as a candidate gene for docosanoic acid abundance. This is a long-chain fatty alcohol oxidase involved in the omega-oxidation pathway of lipid degradation (319). For minor FAs, we

identified a large LD block on chromosome 14 containing the associations with docosanoic and noncosanoic acids, in line with the computational predictions of Badouin *et al.* .2017 (10). Comparison of the Russian sunflower lines with the data on the cultivated and wild sunflower published by Hübner et al. 2019 (145) showed that Russian sunflower germplasm contains unique variation, which is not presented in international collections. This study contributes to the genetic characterization of Russian collections, and our findings can be implemented in future studies.

Due to climate change, sunflower can become the leading plant in oil production because of its ability to grow under different environmental conditions (11. In this view, sunflower varieties with oil properties customized for specific applications may become in-demand in the future. Our study makes a step in this direction by identifying the genetic associations both for major and for minor FAs represented in sunflower oil. Genetic markers for minor FAs, such as docosanoic and noncosanoic acids, have not been previously studied. We hope that future sunflower breeding programs will benefit from understanding the genetic bases governing the proportions of these oil components important for industrial applications.

Conclusions and future perspectives

This work represents how the implementation of high throughput genotyping and phenotyping technologies provides novel insights into the genetic variability of a sunflower germplasm collection in Russia and uncovers novel genetic variants associated with sunflower oil composition variability. I believe my results substantially contributed to the characterization of the genetic buildup of the Russian sunflower collection, thus opening up this genetic resource to a broader spectrum of practical applications.

Specifically, as part of the effort, me and my colleagues compared the genetic variation of Russian sunflower accessions with an international sunflower collection containing wild and cultivated accessions (1065 wild, 20 landraces, and 289 cultivated). Using phenotypic variation data accumulated by the collection holders and oil composition data collected in our laboratory, we performed genome-wide association studies leading us to discovery of genetic regions significantly linked with phenotypic variation of important agricultural traits.

To perform oil seed composition assessment in standardized and quantitative manner, we have optimized mass-spectrometry analysis of sunflower seed extracts and used this technique for systematic sunflower molecular phenotyping. Following the procedure optimization, the mass-spectrometry analysis of the sunflower oil composition yielded 826 computationally annotated lipid peaks and 27 post-hydrolysis fatty acids. The genome-wide association studies performed using genetic variation data collected for 15,483 SNPs yielded significant associations for 11 of the 27 fatty acids, including minor ones constituting less than 1% of the oil composition. For docosanoic acid (22:0), the identified SNPs located on chromosomes 3 and 14 explained up to 34.5% of this acid's total variation in sunflower oil.

In addition to the oil seed composition analysis, we identified genetic variants linked to classic phenotypic traits, such as new candidate genes for fertility restoration. Overall, the main value of the current work is perhaps in the systematic characterization of the genetic and oil composition variation contained in the Russian sunflower germplasm. This, in turn, may provide a valuable recourse for acceleration of sunflower selection and could serve as a base for in depth scientific investigations of the molecular mechanisms connecting genetic variation with the plant’s phenotype.

The main results of my scientific work described in this thesis could be listed as follows:

We identified a genetic locus contributing to pollen fertility restoration in sunflower on chromosome 13.

We identified genetic variants linked with linoleic acid content on chromosomes 8, 9 and 17 in VNIIMK sunflower lines.

Practical comparison of UPLC-MS technology with GC-FAD for sunflower oil fatty acid profiling demonstrated the significant advantages of the UPLC-MS technology in terms of sensitivity and scalability.

Russian sunflower germplasm collection contains unique genetic variants not presented in the international collections currently used in sunflower research.

We identified strong genetic determinants associated with minor fatty acids content in sunflower oil, particularly with docosanoic acid content, indicating that such fatty acids could also be selected as potential breeding targets for marker-assisted selection process.

Yet, we have pointed out new candidate genes as potential targets for oil content research and selection.

Further work for marker validation and estimation of their power in predictions have to be performed.

The environmental and economic situation on our planet is getting harder. Climate change and, consequently, the decree in the land available for cultivation in most countries is a growing problem. Due to world population growth, food demand is growing, and it becomes essential to produce more food on the same piece of land. To overcome these challenges, it is crucial to implement omics techniques in plant research and breeding. First, to get more insights into the genetic control of the traits we are interested in to ensure effective breeding, and second, to use all the power of genomics to improve crop varieties and adapt them for specific needs and conditions.

Each country has to ensure its food security to ensure it is independent of political concerns and foreign country economics. The Russian case is more severe and unique at the same time because of global warming. In contrast to most countries, Russia can benefit from climate change and obtain more land ready for cultivation, which was too cold for cultivation before. Russia has an opportunity to become one of the leading food producers, but this requires the fast implementation of new technologies. Due to the complicated legal situation around GMO and genome editing technologies, I would not bet on them. At least in the following 5-10 years, we would not use them widely for food production. MAS and GS are promising since they utilize natural variation existing within the population for further plant improvement. These technologies do not add anything new to the genome but just explore beneficial combinations of the existing genetic material.

Sunflower has all chances to become the plant oil source number one globally, and its oil can substitute palm and soybean oil on the global market. Sunflower is easier to grow; it is stable under different conditions, and new knowledge on genetics and high-throughput technologies will promote oil customization.

I believe that my work will have both social and scientific impact. From the scientific point of view: the first genotype data for 601 sunflower lines is a significant effort for future studies. These data may be used in numerous GWAS studies looking for associations with very different phenotypes; this data can also be used for germplasm purity checks. The second implementation of UPLC-MS for sunflower oil lipid profiling is a new step in plant phenotyping. I am sure that it will generate a lot of data successfully used in association studies with larger sets of individual plants in the future.

On the other hand, even now, we have a lot of data to analyze. Presently, we were most focused on fatty acids and TAGs. We have still captured much more primary annotated MS peaks corresponding to phospholipids, sterols, prenols, polyketides, and saccharolipids, which we plan to analyze using mathematical modeling approaches. And finally, LD blocks and candidate genes identified in association with several FAs have a value for further research. Probably new genes may be cloned and characterized in detail.

I am personally interested in behenic acid (22:0), a minor fatty acid in sunflower, but widely used in the cosmetic industry thanks to its soothing properties. The SNPs identified in association with it explain the substantial proportion of variation. I believe that based on these SNPs, it is possible to identify genes in wild-type sunflower populations that can be introgressed in the cultivated sunflower so that the high behenic sunflower oil would become a reality.

This project's social significance is that we pointed out that the implementation of modern omics techniques in Russian agriculture can lead to practical results. We showed that it is possible to build an excellent collaboration between research institutes, germplasm holders, and private breeding companies where all would benefit from the collaboration.

I was happy to develop this field during the four years of my Ph. D. I hope that I've managed to obtain results that will be useful for both research and practical application and that I have been promoted by my work this type of research in my country.

Bibliography

1. FAO, IFAD, UNICEF, WFP and WHO. 2019. The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns. Rome, FAO.

2. Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, et al. Solutions for a cultivated planet. Nature. 2011 Oct;478(7369):337–42.

3. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. Arabidopsis thaliana: A Model Plant for Genome Analysis. Science. 1998 Oct 23;282(5389):662–82.

4. Seiler GJ, Qi LL, Marek LF. Utilization of Sunflower Crop Wild Relatives for Cultivated Sunflower Improvement. Crop Sci. 2017;57(3):1083–101.

5. Blackman BK, Scascitelli M, Kane NC, Luton HH, Rasmussen DA, Bye RA, et al. Sunflower domestication alleles support single domestication center in eastern North America. Proc Natl Acad Sci. 2011 Aug 23;108(34):14360–5.

6. Burke JM, Knapp SJ, Rieseberg LH. Genetic Consequences of Selection During the Evolution of Cultivated Sunflower. Genetics. 2005 Dec 1;171(4):1933–40.

7. Burke JM, Tang S, Knapp SJ, Rieseberg LH. Genetic analysis of sunflower domestication. Genetics. 2002 Jul;161(3):1257–67.

8. Ostevik KL, Andrew RL, Otto SP, Rieseberg LH. Multiple reproductive barriers separate recently diverged sunflower ecotypes. Evolution. 2016;70(10):2322–35.

9. Rieseberg LH, Van Fossen C, Desrochers AM. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature. 1995 May;375(6529):313–6.

10. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature. 2017 May 22;546(7656):148–52.

11. Miladinović D, Hladni N, Radanović A, Jocić S, Cvejić S. Sunflower and Climate Change: Possibilities of Adaptation Through Breeding and Genomic Selection. In: Kole C. (Ed.), Genomic Designing of Climate-Smart Oilseed Crops. Cham: Springer International Publishing. 2019;173–238.

12. Debaeke P, Casadebaig P, Flenet F, Langlade N. Sunflower crop and climate change: vulnerability, adaptation, and mitigation potential from case-studies in Europe. OCL. 2017 Jan 1;24(1):D102.

13. Rauf S, Jamil N, Tariq SA, Khan M, Kausar M, Kaya Y. Progress in modification of sunflower oil to expand its industrial value. J Sci Food Agric. 2017 Feb 15;97(7):1997–2006.

14. Kane NC, Burke JM, Marek L, Seiler G, Vear F, Baute G, et al. Sunflower genetic, genomic and ecological resources. Mol Ecol Resour. 2013 Jan;13(1):10–20.

15. FAO. Fats and fatty acids in human nutrition Report of an expert consultation. 2010. 166 p.

16. U.S. Department of Health and Human Services and U.S. Department of Agriculture. *2015 – 2020 Dietary Guidelines for Americans*. 8th Edition. December 2015. Available at https://health.gov/our-work/food-and-nutrition/2015-2020-dietary-guidelines/.

17. Raheja BS, Sadikot SM, Phatak RB, Rao MB. Significance of the N-6/N-3 ratio for insulin action in diabetes. Ann N Y Acad Sci. 1993 Jun 14;683:258–71.

18. Molfino A, Gioia G, Fanelli FR, Muscaritoli M. The Role for Dietary Omega-3 Fatty Acids Supplementation in Older Adults. Nutrients. 2014 Oct 3;6(10):4058–72.

19. Swanson D, Block R, Mousa SA. Omega-3 fatty acids EPA and DHA: health benefits throughout life. Adv Nutr Bethesda Md. 2012 Jan;3(1):1–7.

20. Xu Y, Qian SY. Anti-cancer activities of ω-6 polyunsaturated fatty acids. Biomed J. 2014 Jun;37(3):112–9.

21. Kumar A, Sharma A, Upadhyaya KC. Vegetable Oil: Nutritional and Industrial Perspective. Curr Genomics. 2016 Jun;17(3):230–40.

22. Issariyakul T, Dalai AK. Biodiesel from vegetable oils. Renew Sustain Energy Rev. 2014 Mar 1;31:446–71.

23. Terzić S, Boniface M-C, Marek L, Alvarez D, Baumann K, Gavrilova V, et al. Gene banks for wild and cultivated sunflower genetic resources. OCL. 2020;27:9.

24. Heiser CB. The Origin and Development of the Cultivated Sunflower. Am Biol Teach. 1955;17(5):161–7.

25. Putt ED. Early History of Sunflower. In: A.A. Schneiter (Ed.), Agronomy Monographs. Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. 2015; 1–19.

26. Whiting AF. Ethnobotany of the Hopi. Northern Arizona Society of Science and Art. 1939. 136 p.

27. Zhukovsky PM. Cultivated plants and their relatives. Kolos. 1964. 791 p.

28. Zhukovsky PM. Main gene centres of cultivated plants and their wild relatives within the territory of the U.S.S.R. Euphytica. 1965 Jun 1;14(2):177–88.

29. Gavrilova VA, Anisimova IN. Genealogy of the Sunflower Lines Created on the Basis of Russian Varieties. Helia. 2017 Dec 20;40(67):133–46.

30. Pustovoit VS. Conclusions of work on the selection and seed production of sunflowers. 1964;5:662–97.

31. Pustovoit VS. Selection, seed culture and some agrotechnical problems of sunflower. Kolos. 1966. 368 p.

32. Vear F. Changes in sunflower breeding over the last fifty years. OCL. 2016 Mar 1;23(2):D202.

33. Leclercq P. Une stérilité mâle cytoplasmique chez le tournesol. Ann Amelior Pl. 1969;19:99–106.

34. Kinman M. New Developments in the USDA and State Experiment Station Sunflower Breeding Programs. Proc. 4th Int. Sunflower Conf., Memphis, Tenn, USA. 1970; p. 181–183.

35. Dimitrijevic A, Horn R. Sunflower Hybrid Breeding: From Markers to Genomic Selection. Front Plant Sci. 2018;8.

36. World Agricultural Production. United States Department of Agriculture; 2020 Apr. (Circular Series WAP 4-20). Available from: https://apps.fas.usda.gov/psdonline/circulars/production.pdf

37. Oilseeds: World Markets and Trade. United States Department of Agriculture; 2020 Apr. Available from: https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf

38. Hladni N, Miladinović D. Confectionery sunflower breeding and supply chain in Eastern Europe. OCL. 2019;26:29.

39. Ashmita Roy. Global Oilseeds Market Size and Demand | Industry Overview By 2023. Allied Market Research. Available from: https://www.alliedmarketresearch.com/oilseeds-market

40. Harlan JR, Wet JMJ de. Toward a Rational Classification of Cultivated Plants. TAXON. 1971;20(4):509–17.

41. Rajpal VR, Rao SR, Raina SN, editors. Gene Pool Diversity and Crop Improvement: Volume 1. Springer, Cham. 2016

42. Campbell BT, Saha S, Percy R, Frelichowski J, Jenkins JN, Park W, et al. Status of the Global Cotton Germplasm Resources. Crop Sci. 2010 Jul;50(4):1161–79.

43. Hu J, Seiler G, Kole C. Genetics, Genomics and Breeding of Sunflower. CRC Press. 2010; 374 p.

44. Demurin Y, Borisenko O. Genetic collection of oleic acid content in sunflower seed oil. Helia. 2011;34(55):69–74.

45. Gontcharov SV. Hybrid Sunflower Breeding in VNIIMK (Russia). In: Proc 18th International Sunflower Conference. 2012; 628–633.

46. Seiler GJ, Qi LL, Marek LF. Utilization of Sunflower Crop Wild Relatives for Cultivated Sunflower Improvement. Crop Science. 2017;57(3):1083–1101.

47. Jocic´ S, Miladinovic´ D, Kaya Y. Breeding and Genetics of Sunflower. In: Sunflower. Elsevier. 2015;1–25.

48. Pustovoit VS. Handbook of Selection and Seed Growing of Oil Plants. Israel Program for Scientific Translations. 1973; 328 p.

49. LECLERCQ P. Etude de divers cas de stérilité mâle cytoplasmique chez le tournesol. Agronomie. 1983;3(2):185–187.

50. Vrânceanu AV, Stoenescu FM. Genes for pollen fertility restoration in sunflowers. Euphytica. 1978 Jun 1;27(2):617–627.

51. Unrau J, Whites WJ. The Yield and Other Characters of Inbred Lines and Single Crosses of Sunflowers. Sci Agric. 1944 Jul 1;24(11):516–525.

52. Morozov VK. Sunflower Selection in USSR. Moscow, Russia. 1947; 1–272 p.

53. Corden PV. Sunflower Studies. Agron J. 1922;14:69–72.

54. Gavrilova VA, Rozhkova VT, Anisimova IN. Sunflower Genetic Collection at the Vavilov Institute of Plant Industry. Helia. 2014;37(60):1–16.

55. Farrall M. The Genetical Analysis of Quantitative Traits. J Med Genet. 1996 Nov;33(11):976.

56. Oliva ML. Genotype, environment and GE interaction effect on soybean oil composition [Thesis]. University of Missouri—Columbia. 2005. Available from: https://mospace.umsystem.edu/xmlui/handle/10355/4253

57. Mousavi S, de la Rosa R, Moukhli A, El Riachy M, Mariotti R, Torres M, et al. Plasticity of fruit and oil traits in olive among different environments. Sci Rep. 2019 Nov

18;9(1):16968.

58. Singkham N, Jogloy S, Kesmala T, Swatsitang P, Jaisil P, Puppala N. Genotypic variability and genotype by environment interactions in oil and fatty acids in high, intermediate, and low oleic acid peanut genotypes. J Agric Food Chem. 2010 May 26;58(10):6257–63.

59. Bocianowski J, Liersch A, Nowosad K. Genotype by environment interaction for alkenyl glucosinolates content in winter oilseed rape (Brassica napus L.) using additive main effects and multiplicative interaction model. Curr Plant Biol. 2020 Jan 1;21:100137.

60. Ma'ali S, Makgoga W, Erasmus J, Swanepoel S. Genotype-by-environment interaction and yield stability of sunflower hybrids across production environments in South Africa. South Afr J Plant Soil. 2019 Aug 8;36(4):271–278.

61. Regitano Neto A, Miguel AMR de O, Mourad AL, Henriques EA, Alves RMV, Regitano Neto A, et al. Environmental effect on sunflower oil quality. Crop Breed Appl Biotechnol. 2016 Sep;16(3):197–204.

62. Kang M.S., Prabhakaran V.T., Mehra R.B. Genotype-by-Environment Interaction in Crop Improvement. In: Jain H.K., Kharkwal M.C. (Eds.), Plant Breeding. Springer, Dordrecht. 2004.

63. Radanović A, Miladinović D, Cvejić S, Jocković M, Jocić S. Sunflower Genetics from Ancestors to Modern Hybrids—A Review. Genes. 2018 Nov;9(11):528.

64. Corbi J, Baack EJ, Dechaine JM, Seiler G, Burke JM. Genome-wide analysis of allele frequency change in sunflower crop-wild hybrid populations evolving under natural conditions. Mol Ecol. 2018;27(1):233–47.

65. Cvejić S, Radanović A, Dedić B, Jocković M, Jocić S, Miladinović D. Genetic and Genomic Tools in Sunflower Breeding for Broomrape Resistance. Genes. 2020 Feb;11(2):152.

66. Garcés R, Martínez-Force E, Salas JJ, Venegas-Calerón M. Current advances in sunflower oil and its applications. Lipid Technol. 2009 Apr 1;21(4):79–82.

67. Venegas-Calerón, M., Troncoso-Ponce, M. A., and Martínez-Force, E. Sunflower oil and lipids biosynthesis, In: Sunflower: Chemistry, Production, Processing, and Utilization,E. Martínez-Force, N. T. Dunford, and J. J. Salas (Eds.), Amsterdam: Elsevier. 2015;259–95.

68. Hall AJ, Connor DJ, Whitfield DM. Contribution of pre-anthesis assimilates to grain-filling in irrigated and water-stressed sunflower crops I. Estimates using labelled carbon. Field Crops Res. 1989 Mar 1;20(2):95–112.

69. Nikolau BJ, Oliver DJ, Schnable PS, Wurtele ES. Molecular biology of acetyl-

CoA metabolism. Biochem Soc Trans. 2000 Dec;28(6):591–3.

70. Sasaki Y, Nagano Y. Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. Biosci Biotechnol Biochem. 2004 Jun;68(6):1175–84.

71. Harwood JL. Recent advances in the biosynthesis of plant fatty acids. Biochim Biophys Acta BBA - Lipids Lipid Metab. 1996 May 31;1301(1):7–56.

72. González-Thuillier I, Venegas-Calerón M, Sánchez R, Garcés R, von Wettstein-Knowles P, Martínez-Force E. Sunflower (Helianthus annuus) fatty acid synthase complex: β-hydroxyacyl-[acyl carrier protein] dehydratase genes. Planta. 2016 Feb;243(2):397–410.

73. González-Thuillier I, Venegas-Calerón M, Garcés R, von Wettstein-Knowles P, Martínez-Force E. Sunflower (Helianthus annuus) fatty acid synthase complex: enoyl-[acyl carrier protein]-reductase genes. Planta. 2015 Jan;241(1):43–56.

74. Serrano-Vega MJ, Venegas-Calerón M, Garcés R, Martínez-Force E. Cloning and expression of fatty acids biosynthesis key enzymes from sunflower (Helianthus annuus L.) in Escherichia coli. J Chromatogr B Analyt Technol Biomed Life Sci. 2003 Mar 25;786(1–2):221–8.

75. Martínez-Rivas JM, Sperling P, Lühs W, Heinz E. Spatial and temporal regulation of three different microsomal oleate desaturase genes (FAD2) from normal-type and high-oleic varieties of sunflower (Helianthus annuus L.). Mol Breed. 2001 Sep 1;8(2):159–68.

76. Martínez-Force E, Cantisán S, Serrano-Vega MJ, Garcés R. Acyl-acyl carrier protein thioesterase activity from sunflower (Helianthus annuus L.) seeds. Planta. 2000 Oct;211(5):673–8.

77. Serrano-Vega MJ, Garcés R, Martínez-Force E. Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (Helianthus annuus L.). Planta. 2005 Aug;221(6):868–80.

78. Schnurr JA, Shockey JM, Boer G-J de, Browse JA. Fatty Acid Export from the Chloroplast. Molecular Characterization of a Major Plastidial Acyl-Coenzyme A Synthetase from Arabidopsis. Plant Physiol. 2002 Aug 1;129(4):1700–9.

79. Aznar-Moreno JA, Venegas Calerón M, Martínez-Force E, Garcés R, Mullen R, Gidda SK, et al. Sunflower (Helianthus annuus) long-chain acyl-coenzyme A synthetases expressed at high levels in developing seeds. Physiol Plant. 2014 Mar;150(3):363–73.

80. Cassagne C, Lessire R, Bessoule JJ, Moreau P, Creach A, Schneider F, et al. Biosynthesis of very long chain fatty acids in higher plants. Prog Lipid Res. 1994 Jan 1;33(1):55–

69.

81. Salas JJ, Martínez-Force E, Garcés R. Very Long Chain Fatty Acid Synthesis in Sunflower Kernels. J Agric Food Chem. 2005 Apr 1;53(7):2710–6.

82. Murphy DJ. Plant Storage Lipids. eLS. 2016 Feb 15;1–7.

83. Cagliari A, Margis R, Dos Santos Maraschin F, Turchetto-Zolet AC, Loss G, Margis-Pinheiro M. Biosynthesis of Triacylglycerols (TAGs) in plants and algae. Int J Plant Biol. 2011 Dec 22;2(1):10.

84. Kennedy EP. Biosynthesis of complex lipids. Fed Proc. 1961 Dec;20:934–40.

85. Stymne S, Stobart AK. The biosynthesis of triacylglycerols in microsomal preparations of developing cotyledons of sunflower (Helianthus annuus L.). Biochem J. 1984 Jun 1;220(2):481–8.

86. Triki S, Demandre C, Mazliak P. Biosynthesis of triacylglycerols by developing sunflower seed microsomes. Phytochemistry. 1999 Sep 1;52(1):55–62.

87. Banaś A, Dahlqvist A, Ståhl U, Lenman M, Stymne S. The involvement of phospholipid:diacylglycerol acyltransferases in triacylglycerol production. Biochem Soc Trans. 2000 Dec;28(6):703–5.

88. Banaś W, Sanchez Garcia A, Banaś A, Stymne S. Activities of acyl-CoA:diacylglycerol acyltransferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. Planta. 2013 Jun;237(6):1627–36.

89. Dorrell, D.G. and Vick, B.A. Properties and processing of oilseed sunflower. In: A.A. Schneiter (Ed.), Sunflower Production and Technology. Agronomy Monograph 35. ASA-CSSA-SSSA, Madison, WI, USA, 1997;709–746.

90. Gunstone FD. Fatty Acid and Lipid Chemistry. 1st ed. Springer US. 1996; XII: 252.

91. Unger PW, Thompson TE. Planting Date Effects on Sunflower Head and Seed Development1. Agron J. 1982;74(2):389–95.

92. Purdy RH. High oleic sunflower: Physical and chemical characteristics. J Am Oil Chem Soc. 1986 Aug 1;63(8):1062–6.

93. Hongtrakul V, Slabaugh MB, Knapp SJ. A Seed Specific Δ-12 Oleate Desaturase Gene Is Duplicated, Rearranged, and Weakly Expressed in High Oleic Acid Sunflower Lines. Crop

Sci. 1998;38(5):1245-49.

94. León, A.J., Zambelli, A.D., Reid, R.J, Morata, M.M., Kaspar, M., Martínez-Force, E., Garcés, R., Salas, J.J., Venegas-Caleron, M. Isolated Mutated Nucleotide Sequences That Encode a Modified Oleate Destaurase Sunflower Protein, Modified Protein, Methods and Uses. WIPO Patent WO/2013/004280, Jan 10, 2013.

95. Mattson FH, Lutton ES. The specific distribution of fatty acids in the glycerides of animal and vegetable fats. J Biol Chem. 1958 Oct;233(4):868–71.

96. Mattson FH, Volpenhein RA. The specific distribution of unsaturated fatty acids in the triglycerides of plants. J Lipid Res. 1963 Jan 10;4(4):392–6.

97. Kuksis A. Newer developments in determination of structure of glycerides and phosphoglycerides. Prog Chem Fats Other Lipids. 1972 Jan 1;12:1–163.

98. Borodulina AA, Popov PS, Harchenko LN. Biochemical characteristics of the present sunflower varieties and hybrids. In Pris, France: lnt. Sunflower Assoc. 1974; 239–24.

99. Carelli AA, Brevedan MIV, Crapiste GH. Quantitative determination of phospholipids in sunflower oil. J Am Oil Chem Soc. 1997 May 1;74(5):511–4.

100. Johansson A. The content and composition of sterols and sterol esters in sunflower and poppy seed oils. Lipids. 1979 Mar;14(3):285–91.

101. Krasil'nilov YN, Rzehin YP, Nedacina NA. Cerous substances in sunflower seeds- their composition and distribution on the morphological components of the seed-their distribution in various oil products. In: Proc 5th Int Sunflower Conf,. Paris, France: Int. Sunflower Assoc. 1972; 409–12.

102. Demurin Y, Skoric D, Karlovic D. Genetic variability of tocopherol composition in sunflower seeds as a basis of breeding for improved oil quality. Plant Breed. 1996;115(1):33–6.

103. Demurin YaN, Efimenko SG, Peretyagina TM. Expressivity of tocopherol mutations in sunflower. Helia. 2006;29(45):55–62.

104. Gupta MK. Sunflower oil and its applications. Lipid Technol. 2014 Dec 1;26(11–12):260–3.

105. Soldatov KI. Chemical mutagenesis in sunflower breeding. In: Proceedings of the 7th International Sunflower Conference. Krasnodar. 1996;352–357.

106. Dehmer KJ, Friedt W. Development of molecular markers for high oleic acid content in sunflower (Helianthus annuus L.). Ind Crops Prod. 1998 Jan 1;7(2):311–5.

107. Lacombe S, Souyris I, Bervillé AJ. An insertion of oleate desaturase homologous sequence silences via siRNA the functional gene leading to high oleic acid content in sunflower seed oil. Mol Genet Genomics MGG. 2009 Jan;281(1):43–54.

108. Pérez-Vich B, Fernández-Martínez JM, Grondona M, Knapp SJ, Berry ST. Stearoyl-ACP and oleoyl-PC desaturase genes cosegregate with quantitative trait loci underlying high stearic and high oleic acid mutant phenotypes in sunflower. Theor Appl Genet. 2002 Feb 1;104(2):338–49.

109. Premnath A, Narayana M, Ramakrishnan C, Kuppusamy S, Chockalingam V. Mapping quantitative trait loci controlling oil content, oleic acid and linoleic acid content in sunflower (Helianthus annuus L.). Mol Breed. 2016;36(7).

110. Schuppert GF, Tang S, Slabaugh MB, Knapp SJ. The Sunflower High-oleic Mutant Ol Carries Variable Tandem Repeats of FAD2-1, a Seed-specific Oleoyl-phosphatidyl Choline Desaturase. Mol Breed. 2006 Apr;17(3):241–56.

111. Anushree S, André M, Guillaume D, Frédéric F. Stearic sunflower oil as a sustainable and healthy alternative to palm oil. A review. Agron Sustain Dev. 2017 May 17;37(3):18.

112. Merah O, Langlade N, Alignan M, Roche J, Pouilly N, Lippi Y, et al. Genetic analysis of phytosterol content in sunflower seeds. TAG Theor Appl Genet Theor Angew Genet. 2012 Dec;125(8):1589–601.

113. Velasco L, Fernández-martínez JM. Breeding Oilseed Crops for Improved Oil Quality. J Crop Prod. 2002 Jan 1;5(1–2):309–44.

114. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica. 2005 Jan 1;142(1):169–96.

115. Chikkadevaiah C, Nandini R. Isozymes as markers for differentiating sunflower genotypes. Helia. 2003;26(39):51–8.

116. Kahler AL, Lay CL. Genetics of electrophoretic variants in the annual sunflower. J Hered. 1985 Sep 1;76(5):335–40.

117. Quillet MC, Vear F, Branlard G (Institut N de RA (INRA). The use of isozyme polymorphism for identification of sunflower (Helianthus annuus) inbred lines. J Genet Breed Italy. 1992;46(4):295–304.

118. Rieseberg LH, Choi H, Chan R, Spore C. Genomic map of a diploid hybrid

species. Heredity. 1993 Mar;70(3):285–93.

119. Winter P, Kahl G. Molecular marker technologies for plant improvement. World J Microbiol Biotechnol. 1995 Jul 1;11(4):438–48.

120. Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, et al. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed. 1997 Apr 1;3(2):87–103.

121. Ribaut J-M, Hoisington D. Marker-assisted selection: new tools and strategies. Trends Plant Sci. 1998 Jun 1;3(6):236–9.

122. Mondini L, Noorani A, Pagnotta MA. Assessing Plant Genetic Diversity by Molecular Tools. Diversity. 2009 Sep;1(1):19–35.

123. Semagn K, Bjørnstad Å, Ndjiondjop MN. An overview of molecular marker methods for plants. Afr J Biotechnol. 2006;5(25).

124. Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet. 1986 Sep 1;72(6):761–9.

125. Jan CC, Vick BA, Miller JF, Kahler AL. Progress in the development of a genomic RFLP map of cultivated sunflower (Helianthus *annuus*). In: PAG 1, San Diego, CA, USA. 1992;30.

126. Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, et al. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip. 2018 Mar 4;32(2):261–85.

127. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 1990 Dec 25;18(24):7213–8.

128. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 1990 Nov 25;18(22):6531–5.

129. Deragon JM, Landry BS. RAPD and other PCR-based analyses of plant genomes using DNA extracted from small leaf disks. Genome Res. 1992 Feb 1;1(3):175–80.

130. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995 Nov 11;23(21):4407–14.

131. Althoff DM, Gitzendanner MA, Segraves KA. The Utility of Amplified Fragment

Length Polymorphisms in Phylogenetics: A Comparison of Homology within and between Genomes. Syst Biol. 2007 Jun 1;56(3):477–84.

132. Hongtrakul V, Huestis GM, Knapp SJ. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theor Appl Genet. 1997 Aug 1;95(3):400–7.

133. Al-Chaarani GR, Roustaee A, Gentzbittel L, Mokrani L, Barrault G, Dechamp-Guillaume G, et al. A QTL analysis of sunflower partial resistance to downy mildew (Plasmopara halstedii) and black stem (Phoma macdonaldii) by the use of recombinant inbred lines (RILs). Theor Appl Genet. 2002 Feb 1;104(2):490–6.

134. Akkaya MS, Bhagwat AA, Cregan PB. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics. 1992 Dec;132(4):1131–9.

135. Tang S, Kishore VK, Knapp SJ. PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet. 2003 Jun 1;107(1):6–19.

136. Tang S, Yu J-K, Slabaugh B, Shintani K, Knapp J. Simple sequence repeat map of the sunflower genome. TAG Theor Appl Genet Theor Angew Genet. 2002 Dec;105(8):1124–36.

137. Zhang LS, Clerc VL, Li S, Zhang D. Establishment of an effective set of simple sequence repeat markers for sunflower variety identification and diversity assessment. Can J Bot. 2005 Jan 1;83(1):66–72.

138. Gupta PK, Roy JK, Prasad M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci. 2001;80(4):524–35.

139. Liu A, Burke JM. Patterns of Nucleotide Diversity in Wild and Cultivated Sunflower. Genetics. 2006 May 1;173(1):321–30.

140. Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB. Identification of Single Nucleotide Polymorphisms and analysis of Linkage Disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol. 2008 Jan 23;8:7.

141. Knapp SJ, Berry ST, Rieseberg LH. Genetic mapping in sunflowers. In: Phillips RL, Vasil IK (Eds.), DNA-Based Markers in Plants. Dordrecht: Springer Netherlands. 2001; 379–403.

142. Mehrotra S, Goyal V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. Genomics Proteomics Bioinformatics. 2014 Aug;12(4):164–

71.

143. Nagarajan N, Pop M. Sequence assembly demystified. Nat Rev Genet. 2013 Mar;14(3):157–67.

144. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, et al. High-resolution genetic mapping of maize pan-genome sequence anchors. Nat Commun. 2015 Apr 16;6(1):1–8.

145. Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Ziegler E, et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nat Plants. 2019 Jan;5(1):54–62.

146. Singh VK, Singh AK, Singh S, Singh BD. Next-Generation Sequencing (NGS) Tools and Impact in Plant Breeding. In: Al-Khayri JM, Jain SM, Johnson DV(Eds.), Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools. Cham: Springer International Publishing. 2015:563–612.

147. Kumar S, Banks TW, Cloutier S. SNP Discovery through Next-Generation Sequencing and Its Applications. Tuberosa R, editor. International Journal of Plant Genomics. 2012;2012:831460.

148. Duriez P, Vautrin S, Auriac M-C, Bazerque J, Boniface M-C, Callot C, et al. A receptor-like kinase enhances sunflower resistance to Orobanche cumana. Nat Plants. 2019 Dec;5(12):1211–5.

149. Louarn J, Boniface M-C, Pouilly N, Velasco L, Pérez-Vich B, Vincourt P, et al. Sunflower Resistance to Broomrape (Orobanche cumana) Is Controlled by Specific QTLs for Different Parasitism Stages. Frontiers in Plant Science. 2016;7:590.

150. Mangin B, Pouilly N, Boniface M-C, Langlade NB, Vincourt P, Vear F, et al. Molecular diversity of sunflower populations maintained as genetic resources is affected by multiplication processes and breeding for major traits. Theor Appl Genet. 2017 Jun 1;130(6):1099–112.

151. Livaja M, Unterseer S, Erath W, Lehermeier C, Wieseke R, Plieske J, et al. Diversity analysis and genomic prediction of Sclerotinia resistance in sunflower using a new 25 K SNP genotyping array. Theor Appl Genet. 2016 Feb 1;129(2):317–29.

152. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE. 2011 May 4;6(5):e19379.

153. Baute GJ, Owens GL, Bock DG, Rieseberg LH. Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. Am J Bot. 2016 Dec;103(12):2170–7.

154. Mondon A, Owens GL, Poverene M, Cantamutto M, Rieseberg LH. Gene flow in Argentinian sunflowers as revealed by genotyping-by-sequencing data. Evol Appl. 2018;11(2):193–204.

155. Celik I, Bodur S, Frary A, Doganlar S. Genome-wide SNP discovery and genetic linkage map construction in sunflower (Helianthus annuus L.) using a genotyping by sequencing (GBS) approach. Mol Breed. 2016 Sep 1;36(9):133.

156. Mangin B, Bonnafous F, Blanchet N, Boniface M-C, Bret-Mestries E, Carrère S, et al. Genomic Prediction of Sunflower Hybrids Oil Content. Frontiers in Plant Science. 2017;8:1633.

157. Talukder ZI, Seiler GJ, Song Q, Ma G, Qi L. SNP Discovery and QTL Mapping of Sclerotinia Basal Stalk Rot Resistance in Sunflower using Genotyping-by-Sequencing. Plant Genome. 2016;9(3).

158. Bert P-F, Jouan I, Labrouhe DT de, Serre F, Philippon J, Nicolas P, et al. Comparative genetic analysis of quantitative traits in sunflower (Helianthus annuus L.). 2. Characterisation of QTL involved in developmental and agronomic traits. Theor Appl Genet. 2003 Jun 1;107(1):181–9.

159. Leon AJ, Andrade FH, Lee M. Genetic Mapping of Factors Affecting Quantitative Variation for Flowering in Sunflower. Crop Science. 2000;40(2):404–7.

160. Mokrani L, Gentzbittel L, Azanza F, Fitamant L, Al-Chaarani G, Sarrafi A. Mapping and analysis of quantitative trait loci for grain oil content and agronomic traits using AFLP and SSR in sunflower (Helianthus annuus L.). TAG Theor Appl Genet Theor Angew Genet. 2002 Dec;106(1):149–56.

161. Dussle CM, Hahn V, Knapp SJ, Bauer E. PlArg from Helianthus argophyllus is unlinked to other known downy mildew resistance genes in sunflower. TAG Theor Appl Genet Theor Angew Genet. 2004 Sep;109(5):1083–6.

162. Mouzeyar S, Labrouhe DT de, Vear F. Effect of Host-race Combination on Resistance of Sunflower, Helianthus annuus L., to Downy Mildew Plasmopara balstedii. J Phytopathol. 1994;141(3):249–58.

163. Bert P-F, Jouan I, De Labrouhe DT, Serre F, Nicolas P, Vear F. Comparative genetic analysis of quantitative traits in sunflower (Helianthus annuus L.) 1. QTL involved in resistance to Sclerotinia sclerotiorum and Diaporthe helianthi. TAG Theor Appl Genet Theor Angew Genet. 2002 Nov;105(6–7):985–93.

164. Alfadil TA, Kiani SP, Dechamp-Guillaume G, Gentzbittel L, Sarrafi A. QTL mapping of partial resistance to Phoma basal stem and root necrosis in sunflower (Helianthus annuus L.). Plant Sci. 2007 Apr 1;172(4):815–23.

165. Bert P-F, Dechamp-Guillaume G, Serre F, Jouan I, de Labrouhe DT, Nicolas P, et al. Comparative genetic analysis of quantitative traits in sunflower (Helianthus annuus L.) 3. Characterisation of QTL involved in resistance to Sclerotinia sclerotiorum and Phoma macdonaldi. TAG Theor Appl Genet Theor Angew Genet. 2004 Aug;109(4):865–74.

166. Darvishzadeh R, Kiani SP, Dechamp-Guillaume G, Gentzbittel L, Sarrafi A. Quantitative trait loci associated with isolate specific and isolate nonspecific partial resistance to Phoma macdonaldii in sunflower. Plant Pathology. 2007;56(5):855–61.

167. Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Tourvieille De Labrouhe D, et al. Cloning of molecular markers for disease resistance in sunflower, Helianthus annuus L. TAG Theor Appl Genet Theor Angew Genet. 1998 Mar;96(3–4):519–25.

168. Mestries E, Gentzbittel L, Tourvieille de Labrouhe D, Nicolas P, Vear F. Analyses of quantitative trait loci associated with resistance to shape Sclerotinia sclerotiorum in sunflowers (shape Helianthus annuus L.) using molecular markers. Mol Breed. 1998 Jun 1;4(3):215–26.

169. Micic Z, Hahn V, Bauer E, Schön CC, Knapp SJ, Tang S, et al. QTL mapping of Sclerotinia midstalk-rot resistance in sunflower. TAG Theor Appl Genet Theor Angew Genet. 2004 Nov;109(7):1474–84.

170. Rönicke S, Hahn V, Vogler A, Friedt W. Quantitative Trait Loci Analysis of Resistance to Sclerotinia sclerotiorum in Sunflower. Phytopathology. 2005 Aug 1;95:834–9.

171. Lu YH, Gagne G, Grezes-Besset B, Blanchard P. Integration of a molecular linkage group containing the broomrape resistance gene Or5 into an RFLP map in sunflower. Genome. 1999;42(3):453–6.

172. Márquez-Lema A, Delavault P, Letousey P, Hu J, Pérez B. Candidate gene analysis and identification of TRAP and SSR markers linked to the Or5 gene, which confers sunflower resistance to race E of broomrape (Orobanche cumana Wallr.). In: Proceedings of 17th

International Sunflower Conference,Córdoba, Spain. 2008;661-666.

173. Kiani SP, Talia P, Maury P, Grieu P, Heinz R, Perrault A, et al. Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. Plant Sci. 2007; 173(4):773-87.

174. Kiani SP, Maury P, Nouri L, Ykhlef N, Grieu P, Sarrafi A. QTL analysis of yield-related traits in sunflower under different water treatments. Plant Breed. 2009;128(4):363–73.

175. Fick GN. Heritability of Oil Content in Sunflowers. Crop Sci. 1975;15(1):cropsci1975.0011183X001500010022x.

176. Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A. Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (Helianthus annuus L.). TAG Theor Appl Genet Theor Angew Genet. 2004 Nov;109(7):1353–60.

177. Leon AJ, Lee M, Rufener GK, Berry ST, Mowers RP. Use of RFLP Markers for Genetic Linkage Analysis of Oil Percentage in Sunflower Seed. Crop Sci. 1995;35(2):558-64

178. Tang S, Leon A, Bridges WC, Knapp SJ. Quantitative Trait Loci for Genetically Correlated Seed Traits are Tightly Linked to Branching and Pericarp Pigment Loci in Sunflower. Crop Sci. 2006 Mar 1;46(2):721–34.

179. Ebrahimi A, Maury P, Berger M, Kiani SP, Nabipour A, Shariati F, et al. QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. Genome. 2008 Aug;51(8):599–615.

180. Lübberstedt T, Chen Y, Brenner EA, Zein I. Development and application of functional markers in maize. Euphytica : international journal of plant breeding. 2013;146(1-2):101-8.

181. Jannink JL, Walsh B. Association mapping in plant populations. In: Kang MS (Ed.), Quantitative genetics, genomics and plant breeding. Wallingford: CABI. 2001;59–68.

182. Gupta PK, Rustgi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol Biol. 2005 Mar 1;57(4):461–85.

183. Kraakman ATW, Martínez F, Mussiraliev B, van Eeuwijk FA, Niks RE. Linkage Disequilibrium Mapping of Morphological, Resistance, and Other Agronomically Relevant Traits in Modern Spring Barley Cultivars. Mol Breed. 2006 Jan 1;17(1):41–58.

184. Abdurakhmonov IY, Abdukarimov A. Application of Association Mapping to Understanding the Genetic Diversity of Plant Germplasm Resources. Int J Plant Genomics. 2008; 574927.

185. Sahu PK, Sharma D. Linkage disequilibrium mapping as an advancement in crop breeding. Adv Plants Agric Res. 2017;6(2):40-41.

186. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol. 2003;54:357–74.

187. Hedrick PW. Gametic Disequilibrium Measures: Proceed with Caution. Genetics. 1987 Oct;117(2):331–41.

188. Hill WG, Robertson A. Linkage disequilibrium in finite populations. TAG Theor Appl Genet Theor Angew Genet. 1968 Jun;38(6):226–31.

189. Ersoz ES, Yu J, Buckler ES. Applications of Linkage Disequilibrium and Association Mapping in Crop Plants. In: Varshney RK, Tuberosa R (Eds.), Genomics-Assisted Crop Improvement: Genomics Approaches and Platforms. Dordrecht: Springer Netherlands. 2007;1:97–119.

190. Stranger BE, Stahl EA, Raj T. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. Genetics. 2011 Feb;187(2):367–83.

191. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet. 2012 Sep;44(9):1066–71.

192. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006 Feb;38(2):203–8.

193. Huang X, Han B. Natural Variations and Genome-Wide Association Studies in Crop Plants. Annu Rev Plant Biol. 2014;65(1):531–51.

194. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 2013 Jul 22;9(1):29.

195. Gao J, Wang S, Zhou Z, Wang S, Dong C, Mu C, et al. Linkage mapping and genome-wide association reveal candidate genes conferring thermotolerance of seed-set in maize. J Exp Bot. 2019 Sep 24;70(18):4849–64.

196. Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, et al. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci. 2012 Jun 5;109(23):8872–7.

197. Xiao Y, Liu H, Wu L, Warburton M, Yan J. Genome-wide Association Studies in Maize: Praise and Stargaze. Mol Plant. 2017 Mar;10(3):359–74.

198. Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. 2017 Aug 24;18(1):161.

199. Leamy LJ, Zhang H, Li C, Chen CY, Song B-H. A genome-wide association study of seed composition traits in wild soybean (Glycine soja). BMC Genomics. 2017 Dec;18(1):18.

200. Li D, Zhao X, Han Y, Li W, Xie F. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. Genomics. 2019 Jan 1;111(1):90–5.

201. Zatybekov A, Abugalieva S, Didorenko S, Gerasimova Y, Sidorik I, Anuarbek S, et al. GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. BMC Plant Biol. 2017 Nov 14;17(Suppl 1):179.

202. Gajardo HA, Wittkop B, Soto-Cerda B, Higgins EE, Parkin IAP, Snowdon RJ, et al. Association mapping of seed quality traits in Brassica napus L. using GWAS and candidate QTL approaches. Mol Breed. 2015 Jun 11;35(6):143.

203. Khan SU, Yangmiao J, Liu S, Zhang K, Khan MHU, Zhai Y, et al. Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in Brassica napus L. Ind Crops Prod. 2019 Dec 15;142:111877.

204. Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, et al. Genome-wide association mapping and Identification of candidate genes for fatty acid composition in Brassica napus L. using SNP markers. BMC Genomics. 2017 Mar 14;18(1):232.

205. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet. 2012 Jan;44(1):32–9.

206. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 2010 Nov;42(11):961–7.

207. Reig-Valiente JL, Marqués L, Talón M, Domingo C. Genome-wide association study of agronomic traits in rice cultivated in temperate regions. BMC Genomics. 2018 Sep 25;19(1):706.

208. Gahlaut V, Jaiswal V, Singh S, Balyan HS, Gupta PK. Multi-Locus Genome Wide Association Mapping for Yield and Its Contributing Traits in Hexaploid Wheat under Different Water Regimes. Sci Rep. 2019 Dec 20;9(1):1–15.

209. Li F, Wen W, Liu J, Zhang Y, Cao S, He Z, et al. Genetic architecture of grain yield in bread wheat based on genome-wide association studies. BMC Plant Biol. 2019 Apr 29;19(1):168.

210. Tsai H-Y, Janss LL, Andersen JR, Orabi J, Jensen JD, Jahoor A, et al. Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. Sci Rep. 2020 Feb 25;10(1):1–15.

211. Jabbari M, Fakheri BA, Aghnoum R, Nezhad NM, Ataei R. GWAS analysis in spring barley (Hordeum vulgare L.) for morphological traits exposed to drought. PLOS ONE. 2018 Sep 27;13(9):e0204952.

212. Novakazi, F, Afanasenko, O, Lashina, N, et al. Genome-wide association studies in a barley (Hordeum vulgare) diversity set reveal a limited number of loci for resistance to spot blotch (Bipolaris sorokiniana). Plant Breed. 2019; 00: 1– 15

213. Bonnafous F, Fievet G, Blanchet N, Boniface M-C, Carrère S, Gouzy J, et al. Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. TAG Theor Appl Genet Theor Angew Genet. 2018 Feb;131(2):319–32.

214. Goryunov DV, Anisimova IN, Gavrilova VA, Chernova AI, Sotnikova EA, Martynova EU, et al. Association Mapping of Fertility Restorer Gene for CMS PET1 in Sunflower. Agronomy. 2019 Feb;9(2):49.

215. Masalia RR, Temme AA, Torralba N de L, Burke JM. Multiple genomic regions influence root morphology and seedling growth in cultivated sunflower (Helianthus annuus L.) under well-watered and water-limited conditions. PloS One. 2018;13(9):e0204279.

216. Mangin B, Casadebaig P, Cadic E, Blanchet N, Boniface M-C, Carrère S, et al. Genetic control of plasticity of oil yield for combined abiotic stresses using a joint approach of crop modelling and genome-wide association. Plant Cell Environ. 2017 Oct;40(10):2276–91.

217. Nambeesan SU, Mandel JR, Bowers JE, Marek LF, Ebert D, Corbi J, et al. Association mapping in sunflower (Helianthus annuus L.) reveals independent control of apical vs. basal branching. BMC Plant Biol. 2015 Mar 11;15:84.

218. Dowell JA, Reynolds EC, Pliakas TP, Mandel JR, Burke JM, Donovan LA, et al. Genome-Wide Association Mapping of Floral Traits in Cultivated Sunflower (Helianthus annuus). J Hered. 2019 May 7;110(3):275–86.

219. Gupta PK, Kulwal PL, Jaiswal V. Association mapping in plants in the post-

GWAS genomics era. Adv Genet. 2019;104:75–154.

220. Furbank RT, Tester M. Phenomics – technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 2011 Dec 1;16(12):635–44.

221. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. Nat Rev Genet. 2010 Dec;11(12):855–66.

222. Fiorani F, Schurr U. Future Scenarios for Plant Phenotyping. Annu Rev Plant Biol. 2013;64(1):267–91.

223. Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, et al. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. Mol Plant. 2020 Feb 3;13(2):187–214.

224. Finkel E. Imaging. With "phenomics," plant scientists hope to shift breeding into overdrive. Science. 2009 Jul 24;325(5939):380–1.

225. Gosseau F, Blanchet N, Varès D, Burger P, Campergue D, Colombet C, et al. Heliaphen, an Outdoor High-Throughput Phenotyping Platform for Genetic Studies and Crop Modeling. Front Plant Sci. 2018;9:1908.

226. Xu Y. Envirotyping for deciphering environmental impacts on crop plants. TAG Theor Appl Genet Theor Angew Genet. 2016 Apr;129(4):653–73.

227. Razzaq A, Sadia B, Raza A, Khalid Hameed M, Saleem F. Metabolomics: A Way Forward for Crop Improvement. Metabolites. 2019; 9(12):30

228. Tian H, Lam SM, Shui G. Metabolomics, a Powerful Tool for Agricultural Research. Int J Mol Sci. 2016 Nov 17;17(11):1871

229. Pang Q, Zhang T, Wang Y, Kong W, Guan Q, Yan X, et al. Metabolomics of Early Stage Plant Cell-Microbe Interaction Using Stable Isotope Labeling. Front Plant Sci. 2018;9:760.

230. Lucas-Torres C, Wong A. Current Developments in μMAS NMR Analysis for Metabolomics. Metabolites. 2019 Feb 6;9(2):29.

231. Adamski J. Genome-wide association studies with metabolomics. Genome Med. 2012 Apr 30;4(4):34.

232. Chan EKF, Rowe HC, Kliebenstein DJ. Understanding the Evolution of Defense Metabolites in Arabidopsis thaliana Using Genome-wide Association Mapping. Genetics. 2010 Jul 1;185(3):991–1007.

233. Chen W, Wang W, Peng M, Gong L, Gao Y, Wan J, et al. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nat Commun. 2016 Oct 4;7(1):1–10.

234. Matsuda F, Nakabayashi R, Yang Z, Okazaki Y, Yonemaru J, Ebana K, et al. Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. Plant J Cell Mol Biol. 2015 Jan;81(1):13–23.

235. Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, et al. Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. Plant Physiol. 2014 Jul 1;165(3):1120–32.

236. Oldiges M, Lütz S, Pflug S, Schroer K, Stein N, Wiendahl C. Metabolomics: current state and evolving methodologies and tools. Appl Microbiol Biotechnol. 2007 Sep;76(3):495–511.

237. Mushtaq MY, Choi YH, Verpoorte R, Wilson EG. Extraction for metabolomics: access to the metabolome. Phytochem Anal PCA. 2014 Aug;25(4):291–306.

238. Iwasaki Y, Sawada T, Hatayama K, Ohyagi A, Tsukuda Y, Namekawa K, et al. Separation Technique for the Determination of Highly Polar Metabolites in Biological Samples. Metabolites. 2012 Aug 16;2(3):496–515.

239. Wenk MR. Lipidomics: new tools and applications. Cell. 2010 Dec 10;143(6):888–95.

240. Han X, Gross RW. Electrospray ionization mass spectroscopic analysis of human erythrocyte plasma membrane phospholipids. Proc Natl Acad Sci U S A. 1994 Oct 25;91(22):10635–9.

241. Kim H-Yong, Wang T-CL, Ma Y-Chung. Liquid Chromatography/Mass Spectrometry of Phospholipids using Electrospray Ionization. Anal Chem. 1994 Nov 15;66(22):3977–82.

242. Bausch JN. Lipid analysis. Curr Opin Biotechnol. 1993 Feb;4(1):57–62.

243. Chen M, Markham JE, Dietrich CR, Jaworski JG, Cahoon EB. Sphingolipid Long-Chain Base Hydroxylation Is Important for Growth and Regulation of Sphingolipid Content and Composition in Arabidopsis. Plant Cell. 2008 Jul;20(7):1862–78.

244. Lu W, Bennett BD, Rabinowitz JD. Analytical strategies for LC-MS-based targeted metabolomics. J Chromatogr B Analyt Technol Biomed Life Sci. 2008 Aug 15;871(2):236–42.

245. Hummel J, Segu S, Li Y, Irgang S, Jueppner J, Giavalisco P. Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. Front Plant Sci. 2011;2:54.

246. Ikeda K, Oike Y, Shimizu T, Taguchi R. Global analysis of triacylglycerols including oxidized molecular species by reverse-phase high resolution LC/ESI-QTOF MS/MS. J Chromatogr B Analyt Technol Biomed Life Sci. 2009 Sep 1;877(25):2639–47.

247. Bromke MA, Hochmuth A, Tohge T, Fernie AR, Giavalisco P, Burgos A, et al. Liquid chromatography high-resolution mass spectrometry for fatty acid profiling. Plant J. 2015 Feb;81(3):529–36.

248. Yang K, Han X. Accurate Quantification of Lipid Species by Electrospray Ionization Mass Spectrometry — Meets a Key Challenge in Lipidomics. Metabolites. 2011 Nov 11;1(1):21–40.

249. Knittelfelder OL, Weberhofer BP, Eichmann TO, Kohlwein SD, Rechberger GN. A versatile ultra-high performance LC-MS method for lipid profiling. J Chromatogr B Analyt Technol Biomed Life Sci. 2014 Mar 1;951–952(100):119–28.

250. Salem MA, Giavalisco P. Semi-targeted Lipidomics of Plant Acyl Lipids Using UPLC-HR-MS in Combination with a Data-Independent Acquisition Mode. In: António C (Ed.), Plant Metabolomics: Methods and Protocols. New York, NY: Springer. 2018;137–55.

251. Tarazona P, Feussner K, Feussner I. An enhanced plant lipidomics method based on multiplexed liquid chromatography–mass spectrometry reveals additional insights into cold- and drought-induced membrane remodeling. Plant J. 2015;84(3):621–33.

252. Zhou X-R, Callahan DL, Shrestha P, Liu Q, Petrie JR, Singh SP. Lipidomic analysis of Arabidopsis seed genetically engineered to contain DHA. Frontiers in Plant Science. 2014;5:419.

253. Riewe D, Wiebach J, Altmann T. Structure Annotation and Quantification of Wheat Seed Oxidized Lipids by High-Resolution LC-MS/MS. Plant Physiol. 2017 Oct;175(2):600–18.

254. Riedelsheimer C, Brotman Y, Méret M, Melchinger AE, Willmitzer L. The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. Sci Rep. 2013 Aug 21;3(1):1–7.

255. Li M, Butka E, Wang X. Comprehensive Quantification of Triacylglycerols in Soybean Seeds by Electrospray Ionization Mass Spectrometry with Multiple Neutral Loss Scans.

Sci Rep. 2014 Oct 10;4(1):1–11.

256. Kalo PJ, Ollilainen V, Rocha JM, Malcata FX. Identification of molecular species of simple lipids by normal phase liquid chromatography-positive electrospray tandem mass spectrometry, and application of developed methods in comprehensive analysis of low erucic acid rapeseed oil lipids. Int J Mass Spectrom. 2006 Jul 1;254:106–21.

257. Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, et al. Marker assisted selection in crop plants. Plant Cell Tissue Organ Cult. 2005 Sep 1;82(3):317–42.

258. Varshney RK, Pandey MK, Bohra A, Singh VK, Thudi M, Saxena RK. Toward the sequence-based breeding in legumes in the post-genome sequencing era. TAG Theor Appl Genet Theor Angew Genet. 2019;132(3):797–816.

259. Varshney RK, Tuberosa R (Eds.), Genomics-Assisted Crop Improvement: Vol 1: Genomics Approaches and Platforms. Springer Netherlands. 2007;386.

260. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci. 2009 Jan 1;92(1):16–24.

261. Nakaya A, Isobe SN. Will genomic selection be a practical method for plant breeding? Ann Bot. 2012 Nov;110(6):1303–16.

262. Wang X, Xu Y, Hu Z, Xu C. Genomic selection methods for crop improvement: Current status and prospects. Crop J. 2018 Aug 1;6(4):330–40.

263. Li J-Y, Wang J, Zeigler RS. The 3,000 rice genomes project: new opportunities and challenges for future rice research. GigaScience. 2014;3:8.

264. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet. 2012 Feb;44(2):217–20.

265. Xu Y, Xu C, Xu S. Prediction and association mapping of agronomic traits in maize using multiple omic data. Heredity. 2017 Sep;119(3):174–84.

266. Gärtner T, Steinfath M, Andorf S, Lisec J, Meyer RC, Altmann T, et al. Improved Heterosis Prediction by Combining Information on DNA- and Metabolic Markers. PLOS ONE. 2009 Apr 16;4(4):e5220.

267. Reif JC, Zhao Y, Würschum T, Gowda M, Hahn V. Genomic prediction of sunflower hybrid performance. Plant Breed. 2013;132(1):107–14.

268. Talukder ZI, Gong L, Hulke BS, Pegadaraju V, Song Q, Schultz Q, et al. A High-Density SNP Map of Sunflower Derived from RAD-Sequencing Facilitating Fine-Mapping of the Rust Resistance Gene R12. PLOS ONE. 2014 Jul 11;9(7):e98628.

269. Galli G, Alves FC, Morosini JS, Fritsche-Neto R. On the usefulness of parental lines GWAS for predicting low heritability traits in tropical maize hybrids. PLOS ONE. 2020 Feb 7;15(2):e0228724.

270. Varshney RK, Sinha P, Singh VK, Kumar A, Zhang Q, Bennetzen JL. 5Gs for crop genetic improvement. Curr Opin Plant Biol. 2020 Jan 28;S1369-5266(19):30119-0.

271 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

272. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLOS ONE. 2014 Feb 28;9(2):e90346.

273. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870–4.

274 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

275. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

276. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

277. Kaya Y, Baltensperger D. THE DETERMINING PHYSIOLOGICAL MATURITY TIME IN SUNFLOWER. 2004 Jan 1;

278. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006 Feb 1;78(3):779–87.

279.Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep;19(9):1655–64.

280. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.

281. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics. 2010 Apr;42(4):355–60.

282. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007 Oct 1;23(19):2633–5.

283. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014 May 14;005165.

284. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263–5.

285. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. Annual Review of Plant Biology. 2014;65(1):415–42.

286. Liu Z, Wang D, Feng J, Seiler GJ, Cai X, Jan C-C. Diversifying Sunflower Germplasm by Integration and Mapping of a Novel Male Fertility Restoration Gene. Genetics. 2013 Mar;193(3):727–37.

287. Mandel JR, Nambeesan S, Bowers JE, Marek LF, Ebert D, Rieseberg LH, et al. Association Mapping and the Genomic Consequences of Selection in Sunflower. PLOS Genetics. 2013 Mar 21;9(3

288 Fick GN, Miller JF. Sunflower Breeding. Sunflower Technology and Production. 1997 Jan;):395–439.

289 Puzikov AN, Suvorova YN. Large-seeded sunflower varieties in the southern forest steppe of Western Siberia. *Siberian Herald of Agricultural Science*. 2016;(3):23-29. (In Russ.)

290. Thompson, AP, Dougherty, D. A system to classify sunflower hybrids according to PM. Sunflower Research Workshop by National Sunflower Association. 15–16 Jan, 1998. Fargo, ND.

291 Alpaslan M, Gündüz H. The effects of growing conditions on oil content, fatty acid composition and tocopherol content of some sunflower varieties produced in Turkey. Nahrung/Food. 2000 Dec 1;44(6):434–7.

292 Pleite R, Martínez-Force E, Garcés R. Increase of the stearic acid content in high-oleic sunflower (Helianthus annuus) seeds. J Agric Food Chem. 2006 Dec 13;54(25):9383–8.

293 Yu J-K, Tang S, Slabaugh MB, Heesacker A, Cole G, Herring M, et al. Towards a Saturated Molecular Genetic Linkage Map for Cultivated Sunflower. Crop Science.

2003;43(1):367–87.

294 Kusterer B, Rozynek B, Brahm L, Prüfe M, Tzigos S, Horn R, et al. Construction of a genetic map and localization of major traits in sunflower (*Helianthus annuus* L.). Helia. 2004 Jul 1;27(40):15–24.

295 Kusterer B, Horn R, Friedt W. Molecular mapping of the fertility restoration locus Rf1 in sunflower and development of diagnostic markers for the restorer gene. Euphytica. 2005 Jan 1;143(1):35–42.

296 Yue B, Vick BA, Cai X, Hu J. Genetic mapping for the Rf1 (fertility restoration) gene in sunflower (Helianthus annuus L.) by SSR and TRAP markers. Plant Breeding. 2010;129(1):24–8.

297 Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in Plant Science. 2008 Dec 1;13(12):663–70.

298 Aubourg S, Boudet N, Kreis M, Lecharny A. In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants. Plant Mol Biol. 2000 Mar 1;42(4):603–13.

299 Small ID, Peeters N. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci. 2000 Feb;25(2):46–7.

300 Hernandez Mora JR, Rivals E, Mireau H, Budar F. Sequence analysis of two alleles reveals that intra-and intergenic recombination played a role in the evolution of the radish fertility restorer (Rfo). BMC Plant Biology. 2010 Feb 24;10(1):35.

301 Bocianowski J, Mikołajczyk K, Bartkowiak-Broda I. Determination of fatty acid composition in seed oil of rapeseed (Brassica napus L.) by mutated alleles of the FAD3 desaturase genes. J Appl Genet. 2012 Feb;53(1):27–30.

302 Sharafi Y, Majidi MM, Goli SAH, Rashidi F. Oil Content and Fatty Acids Composition in Brassica Species. International Journal of Food Properties. 2015 Oct 3;18(10):2145–54.

303 Roudier F, Gissot L, Beaudoin F, Haslam R, Michaelson L, Marion J, et al. Very-Long-Chain Fatty Acids Are Involved in Polar Auxin Transport and Developmental Patterning in Arabidopsis. The Plant Cell. 2010 Feb 1;22(2):364–75.

304 Bach L, Faure J-D. Role of very-long-chain fatty acids in plant development, when chain length does matter. C R Biol. 2010 Apr;333(4):361–70.

305 Degwert J, Jacob J, Steckel F. Use of Cis-9-heptadecenoic acid for treating psoriasis and allergies [Internet]. US5708028A, 1998. Available from:

https://patents.google.com/patent/US5708028/en/en17

306 Jenkins BJ, Seyssel K, Chiu S, Pan P-H, Lin S-Y, Stanley E, et al. Odd Chain Fatty Acids; New Insights of the Relationship Between the Gut Microbiota, Dietary Intake, Biosynthesis and Glucose Intolerance. Sci Rep. 2017 23;7:44845.

307 Fitton A, Goa KL. Azelaic Acid. Drugs. 1991 May 1;41(5):780–98.

308 Avis TJ, Boulanger RR, Bélanger RR. Synthesis and Biological Characterization of (Z)-9-Heptadecenoic and (Z)-6-Methyl-9-Heptadecenoic Acids: Fatty Acids with Antibiotic Activity Produced by Pseudozyma flocculosa. J Chem Ecol. 2000 Apr 1;26(4):987–1000.

309 Avis TJ, Bélanger RR. Specificity and Mode of Action of the Antifungal Fatty Acid cis-9-Heptadecenoic Acid Produced by Pseudozyma flocculosa. Appl Environ Microbiol. 2001 Feb;67(2):956–60.

310 Clausen CA, Coleman RD, Yang VW. Fatty acid-based formulations for wood protection against mold and sapstain. Forest products journal Vol 60, no 3 (May 2010): p 301-304. 2010;60(3):301–4.

311 Rezanka T, Sigler K. Odd-numbered very-long-chain fatty acids from the microbial, animal and plant kingdoms. Prog Lipid Res. 2009 Jul;48(3–4):206–38.

312 Park Y-K, Dulermo T, Ledesma-Amaro R, Nicaud J-M. Optimization of odd chain fatty acid production by Yarrowia lipolytica. Biotechnology for Biofuels. 2018 Jun 7;11(1):158.

313 Himawan C, Starov VM, Stapley AGF. Thermodynamic and kinetic aspects of fat crystallization. Adv Colloid Interface Sci. 2006 Sep 25;122(1–3):3–33.

314 Teutonico RA, Yandell B, Satagopan JM, Ferreira ME, Palta JP, Osborn TC. Genetic analysis and mapping of genes controlling freezing tolerance in oilseedBrassica. Molecular Breeding. 1995;1(4):329–339.

315 Roughan PG. Phosphatidylglycerol and Chilling Sensitivity in Plants. Plant Physiol. 1985 Mar;77(3):740–6.

316 Voelker TA, Kinney AJ. Variations in the Biosynthesis of Seed-storage Lipids. Annual review of plant physiology and plant molecular biology. 2001;52:335–61.

317 Huang Z, Zhang X, Jiang S, Qin M, Zhao N, Lang L, et al. Analysis of cold resistance and identification of SSR markers linked to cold resistance genes in Brassica rapa L. Breed Sci. 2017 Jun;67(3):213–20.

318 Fernández-Moya V, Martínez-Force E, Garcés R. Metabolism of Triacylglycerol

Species during Seed Germination in Fatty Acid Sunflower (Helianthus annuus) Mutants. J Agric Food Chem. 2000 Mar 1;48(3):770–4.

319 Vanhanen S, West M, Kroon JTM, Lindner N, Casey J, Cheng Q, et al. A Consensus Sequence for Long-chain Fatty-acid Alcohol Oxidases from Candida Identifies a Family of Genes Involved in Lipid ω-Oxidation in Yeast with Homologues in Plants and Bacteria. J Biol Chem. 2000 Nov 2;275(6):4445–52.

Annex

Annex for Chapter 3

Samples used in the study

Line_identificator	Collection	Illumina line in the run I1	barcode	Sequencing_id
A32/08 23-1121	AGROPLASMA	L007	TTCTGAG	s410
A32/08 308-111	AGROPLASMA	L001	CCTAGAT	s1
A40/08 36-1121	AGROPLASMA	L002	TTGGCAT	s97
A40/08 150-1111	AGROPLASMA	L002	TTCCACG	s98
A40/08 124(155-21111)	AGROPLASMA	L001	GATCTGG	s2
A40/08 155-22111	AGROPLASMA	L008	CGAGGTT	s506
A40/08 163-1111	AGROPLASMA	L008	AACAGAT	s507
A40/08 167-1111	AGROPLASMA	L008	GCCTCAC	s508
A40/08 197-1111	AGROPLASMA	L002	GGTTGAC	s99
A40/08 205-11111	AGROPLASMA	L007	CGTGGAG	s411
A40/08 254-1111	AGROPLASMA	L001	CTCTATG	s3
A40/08 280-1111	AGROPLASMA	L002	ATCTGTT	s100
A704/08 149-11111	AGROPLASMA	L001	CCGCATT	s4
A33/08 35-11111	AGROPLASMA	L007	ATTGCGT	s412
A33/08 51211	AGROPLASMA	L007	GTAACCT	s413
A33/08 14-1122	AGROPLASMA	L008	CCACGTC	s509
A33/08 35-2111	AGROPLASMA	L007	ACGTGAG	s414
A40/08 60-111	AGROPLASMA	L001	TTCCACG	s5
A40/08-150-2211	AGROPLASMA	L008	TGATCTC	s510
A40/08-162-111	AGROPLASMA	L008	CTAATGT	s511
A40/33-82-111	AGROPLASMA	L002	CGGAGGT	s101
BC2ЦМС AP83 x B32X36//BM08-41-12111	AGROPLASMA	L007	AACTCAG	s415
BC2ЦМС AP83 x B32X08//BF704XB08-14-1121	AGROPLASMA	L007	TTGTTAC	s416
BC2ЦМС AP83 x B32X08//BF704XBM08-21-112	AGROPLASMA	L002	GAATACC	s102
BC2ЦМС AP83 x BS33X08/BSH033XBM08-6-1112	AGROPLASMA	L007	CCGACTT	s417
BC2ЦМС AP83 x BS33X08/BSH033XBM08-19-111	AGROPLASMA	L001	CGGAGGT	s6

BC2ЦМС AP83 x BS33X08/BSH033XBM08-12111	AGROPLASMA	L001	TGCTACC	s7
BC2ЦМС AP83 x BS33X08/BSH033XBM08-31-11112	AGROPLASMA	L008	GGAATAC	s512
BC2ЦМС AP83 x B40X08/BF704XBM08-4-2112	AGROPLASMA	L008	TTCTAGT	s513
BC2ЦМС AP83 x B40X08/BF704XB08-32-112	AGROPLASMA	L007	ATCTGTT	s418
BC2ЦМС AP83 x BM40XBM08/BF704XBM08-50-1212	AGROPLASMA	L001	TGTTCAG	s8
BC2ЦМС AP83 x BM40XBS33/BM08-15-21121	AGROPLASMA	L007	CACACGT	s419
BC2ЦМС AP83 x BM32XBS33/BM36-7-1111	AGROPLASMA	L002	ACGTTGT	s103
BC2ЦМС AP83 x BM32XBS33/BM36-72-1111	AGROPLASMA	L007	ACTTCTG	s420
BC2ЦМС AP83 x BM32XBM08/BM08-14-12-111	AGROPLASMA	L008	TTCACTG	s514
BC2ЦМС AP83 x BM32XBM08/BM08-19-11111	AGROPLASMA	L002	ATGTTCG	s104
BC2ЦМС AP83 x BM32XBM08/BM08-24-111	AGROPLASMA	L001	CGAGGTT	s9
BC2ЦМС AP83 x BM32XBM08/BM08-29-1111	AGROPLASMA	L008	CCGTAAC	s515
BC2ЦМС AP83 x BM32XBM08/BM08-33-21111	AGROPLASMA	L008	GTTGAGC	s516
BC2ЦМС AP83 x BM32XBM08/BM08-83-22121	AGROPLASMA	L001	CTAATGT	s10
BC2ЦМС AP83 x B47/704-8	AGROPLASMA	L001	ATGCCGG	s11
BC2ЦМС AP83 x BHO33/704//47-2	AGROPLASMA	L008	GAATACC	s517
BC2ЦМС AP83 x BM40XBS33/BM08-25-211	AGROPLASMA	L001	CCTGCAC	s12
BC2ЦМС AP83 x C0SB-5	AGROPLASMA	L008	CCGCATT	s518
AP83X B08/P83-24211113	AGROPLASMA	L001	CGTGGAG	s13
AP83X B08/P83-24221113	AGROPLASMA	L007	TTGGCAT	s421
AP83X B08/83-31211211	AGROPLASMA	L008	TTAGCCG	s519
AP83X B08/P83-33421111	AGROPLASMA	L002	TTGTTAC	s105
AP83X BC1MB08/BP83-3322	AGROPLASMA	L002	GAGGCTG	s106
AP83X B08/P83-33421111/3322	AGROPLASMA	L007	GAGGCTG	s422
AP83X B40/P83-123111	AGROPLASMA	L002	GCCTCAC	s107
AP83X BC1-B40//P83-32112	AGROPLASMA	L001	GGACTCG	s14
AP83X BC1BM08//BP83-13211	AGROPLASMA	L001	ACCATAG	s15
AP83X BC1BM08//BP83-13211/1522	AGROPLASMA	L008	ATCTTGC	s520
AP83X BC1BM40//BP83-321111	AGROPLASMA	L002	CCTGCAC	s108
AP83X S5BC1-08//BP83-123	AGROPLASMA	L007	GATGGCT	s423
AP83X S5BC1-08//BP83-1821	AGROPLASMA	L002	CCGCATT	s109
AP83X S5BC1-08//BP83-1824	AGROPLASMA	L008	GTAACCT	s521
AP83X BM40///BP83-1	AGROPLASMA	L007	GATAGGC	s424
AP83X BM40///BP83-2	AGROPLASMA	L007	TGCGAGG	s425
AP83X BM40///BP83-3	AGROPLASMA	L008	GGACTCG	s522

AP83X BM08///BP83-1	AGROPLASMA	L001	CCACCGT	s16
AP83X B08/P83-24211113	AGROPLASMA	L008	GTTCGAT	s523
AP83X BC1-B40//P83-32112	AGROPLASMA	L008	CAGGTGG	s524
AP83X BC1BM40//BP83-321111	AGROPLASMA	L007	TAGGCGC	s426
RP38	AGROPLASMA	L001	TAACTGC	s17
RP38YL	AGROPLASMA	L002	ACTTCTG	s110
RP1	AGROPLASMA	L002	AGAAGTC	s111
RP14	AGROPLASMA	L002	GATCTGG	s112
RP24	AGROPLASMA	L007	CGATGCG	s427
Rlg3	AGROPLASMA	L001	AGAAGTC	s18
Rlg42	AGROPLASMA	L008	GGACATC	s525
R10	AGROPLASMA	L001	ATGTTCG	s19
RNM	AGROPLASMA	L008	TAAGAAC	s526
412pl	AGROPLASMA	L007	GCCAACT	s428
4120pl	AGROPLASMA	L008	AGCCGGT	s527
RR144	AGROPLASMA	L008	AACAATG	s528
ΦP81013	AGROPLASMA	L007	GAACAAT	s429
R10YL	AGROPLASMA	L008	TACTGTC	s529
RS//R10-132226	AGROPLASMA	L008	ACGTGAG	s530
RS//R10-132151	AGROPLASMA	L007	TAACTGC	s430
RS//R10-13/5B	AGROPLASMA	L008	TTCCACG	s531
R10/1231	AGROPLASMA	L007	AGCCGGT	s431
R10CLP-C1	AGROPLASMA	L007	CCTACCG	s432
Rdlf	AGROPLASMA	L007	GGACTCG	s433
4090pl	AGROPLASMA	L008	TGCCTAG	s532
Rd1	AGROPLASMA	L007	AAGGTCT	s434
4093pl	AGROPLASMA	L007	AAGCGTG	s435
RS15	AGROPLASMA	L002	ATCTTGC	s113
RS25	AGROPLASMA	L001	ACTCGCT	s20
4087pl	AGROPLASMA	L008	CCACCGT	s533
RS64	AGROPLASMA	L008	CCATTGC	s534
RC32(nord)	AGROPLASMA	L001	AGACCTT	s21
R6	AGROPLASMA	L001	AACGAGT	s22
RK35	AGROPLASMA	L002	GGACTCG	s114
4099pl	AGROPLASMA	L007	GTTACTC	s436
RR154	AGROPLASMA	L001	GTTACTC	s23
4117pl	AGROPLASMA	L008	TTGTTAC	s535
4099pl	AGROPLASMA	L007	ATGTTCG	s437

RD 164	AGROPLASMA	L007	CTCTATG	s438
RR114	AGROPLASMA	L008	TATGTTC	s536
4K738/R6-12111	AGROPLASMA	L008	ATTGCGT	s537
4K738/R6-4111	AGROPLASMA	L007	CCGGTAT	s439
RCM//R6-8211	AGROPLASMA	L008	CACATCG	s538
RCM///R6-31212	AGROPLASMA	L001	TGTAACT	s24
RCM///R6-3312	AGROPLASMA	L002	TTCTAGT	s115
RCM//R6-6312	AGROPLASMA	L007	GTTGAGC	s440
4096	AGROPLASMA	L007	GATCTGG	s441
4102	AGROPLASMA	L007	AGACCTT	s442
7678-2513	AGROPLASMA	L002	TAGAACG	s116
RF 12	AGROPLASMA	L001	GTAGGTC	s25
Rfob	AGROPLASMA	L001	GATAGGC	s26
R483	AGROPLASMA	L002	TGCTACC	s117
R60880	AGROPLASMA	L007	TTAGCCG	s443
R60875	AGROPLASMA	L008	CCTCAGC	s539
R60875CLP12	AGROPLASMA	L008	ATGTTCG	s540
RS3	AGROPLASMA	L001	CCACGTC	s27
R4	AGROPLASMA	L008	GATAGGC	s541
R4CLP	AGROPLASMA	L002	TAGGCGC	s118
R-fly	AGROPLASMA	L001	GAACAAT	s28
RBY	AGROPLASMA	L002	GGACATC	s119
RMO2	AGROPLASMA	L008	AGCTTCT	s542
RMO2CLP	AGROPLASMA	L008	GAACAAT	s543
R-ramb	AGROPLASMA	L007	AACGTAC	s444
RK325	AGROPLASMA	L007	CCACGTC	s445
RK806	AGROPLASMA	L007	CTAATGT	s446
RK-alz	AGROPLASMA	L007	TACTGTC	s447
Ralz	AGROPLASMA	L008	TGCGAGG	s544
RK-BLR	AGROPLASMA	L008	AGACCTT	s545
RL-mgs	AGROPLASMA	L001	AGCTTCT	s29
RL 65/35	AGROPLASMA	L001	GCCAACT	s30
R-eL	AGROPLASMA	L001	AGTAAGC	s31
R-exp	AGROPLASMA	L008	CCAATTG	s546
RT085	AGROPLASMA	L002	TGAACAT	s120
Rcrb	AGROPLASMA	L007	GCCTCAC	s448
Rs65HO	AGROPLASMA	L007	GAGTGCG	s449
9758R	AGROPLASMA	L007	CGAGGTT	s450

9802R	AGROPLASMA	L008	GGAGACT	s547
RC61	AGROPLASMA	L007	ATTCTTG	s451
RC61CLP41	AGROPLASMA	L001	TTGCATC	s32
RC8505	AGROPLASMA	L008	GTTACTC	s548
R67	AGROPLASMA	L001	GGATCGT	s33
AM32CLP	AGROPLASMA	L008	TTCTGAG	s549
AC42CLP	AGROPLASMA	L008	CCTGACT	s550
AM078CLP	AGROPLASMA	L002	GAACAAT	s121
AMO33	AGROPLASMA	L008	GTGGACG	s551
AMO8	AGROPLASMA	L001	GGACATC	s34
K 223	VNIMK	L002	GCCAACT	s122
K 225	VNIMK	L007	GGACATC	s452
K 370	VNIMK	L002	TTCTGAG	s123
K 562	VNIMK	L008	CGGAGGT	s552
K 581	VNIMK	L008	GTAGGTC	s553
K 651-3	VNIMK	L007	CCAATTG	s453
K 912	VNIMK	L002	CCTCAGC	s124
K 1391	VNIMK	L001	AACAGAT	s35
K 1459	VNIMK	L007	ACCATAG	s454
K 1505	VNIMK	L007	AACAATG	s455
K 1594	VNIMK	L007	CCTCTAG	s456
K 1687	VNIMK	L007	TTCCACG	s457
K 2068	VNIMK	L001	ACTTCTG	s36
K 2086	VNIMK	L001	GGAACTG	s37
K 2125	VNIMK	L002	AAGGTCT	s125
K 2235	VNIMK	L007	TGGCAGT	s458
K 2238	VNIMK	L008	CCGATCC	s554
K 2257	VNIMK	L002	ATTCTTG	s126
K 2462	VNIMK	L002	CCAATTG	s127
K 2479	VNIMK	L001	TACTGTC	s38
K 3035	VNIMK	L002	CCGACTT	s128
K 3059	VNIMK	L001	TTGTTAC	s39
K 3159	VNIMK	L008	TGGCAGT	s555
K 3350	VNIMK	L008	GGTTGAC	s556
K 3376	VNIMK	L008	AACGTAC	s557
Сл 1721	VNIMK	L008	TGCTACC	s558
Сл 1790	VNIMK	L007	AGAAGTC	s459
Сл 1813	VNIMK	L001	CCTACCG	s40

Сл 2039	VNIMK	L007	TTGAGGC	s460
Сл 2950	VNIMK	L001	TTAGCCG	s41
ВА 1	VNIMK	L008	GATCTGG	s559
ВА 4	VNIMK	L001	GAGGCTG	s42
ВИР 130-1	VNIMK	L002	CGTGGAG	s129
ВИР 172	VNIMK	L001	GAGTGCG	s43
ВИР 369	VNIMK	L007	GACCTCT	s461
ВИР 391	VNIMK	L001	GGAGACT	s44
ВК 15	VNIMK	L001	TAAGAAC	s45
ВК 30	VNIMK	L007	AGCTTCT	s462
ВК 102	VNIMK	L002	CCTAGAT	s130
ВК 206	VNIMK	L001	TTGGCAT	s46
ВК 268	VNIMK	L007	CCGCATT	s463
ВК 416	VNIMK	L008	GACCTCT	s560
ВК 428	VNIMK	L001	TTCACTG	s47
ВК 464	VNIMK	L001	GCCTCAC	s48
ВК 474	VNIMK	L001	TGATCTC	s49
ВК 475	VNIMK	L002	GGAACTG	s131
ВК 519	VNIMK	L001	ATTGCGT	s50
ЛГ 3	VNIMK	L007	CAGGTGG	s464
ЛГ 8-2	VNIMK	L007	GTTCGAT	s465
ЛГ 10	VNIMK	L002	AACAGAT	s132
ЛГ 26	VNIMK	L007	CCGATCC	s466
ЛГ 27	VNIMK	L001	AGTGGCC	s51
ЛГ 28	VNIMK	L007	TGAACAT	s467
КГ 7	VNIMK	L002	GGAATAC	s133
КГ 16	VNIMK	L002	AAGCGTG	s134
КГ 19	VNIMK	L008	TGTAACT	s561
КГ 21	VNIMK	L008	GAGGCTG	s562
КГ 32	VNIMK	L002	GTTGAGC	s135
КГ 48	VNIMK	L007	ATGCCGG	s468
КГ 104	VNIMK	L008	CTCTATG	s563
Л 1392	VNIMK	L008	CACACGT	s564
Л 2090	VNIMK	L008	TAGAACG	s565
л 2138	VNIMK	L001	TGGCAGT	s52
Л 2532	VNIMK	L008	AAGCGTG	s566
Л 2543	VNIMK	L002	ATTGCGT	s136
Л 2544	VNIMK	L007	ACGTTGT	s469

Л 2563	VNIMK	L001	AAGCGTG	s53
Л 2582	VNIMK	L007	CCGTAAC	s470
Л 2586	VNIMK	L001	TTGAGGC	s54
Л 2595	VNIMK	L001	CACATCG	s55
Л 3376	VNIMK	L007	AACGAGT	s471
Л 7247	VNIMK	L007	CCTCAGC	s472
М 1046	VNIMK	L008	TTGGCAT	s567
И7-235	VNIMK	L001	CGATGCG	s56
И7-246	VNIMK	L008	GAGTGCG	s568
1416-5	VNIMK	L007	GTGGACG	s473
355114	VNIMK	L007	GGAACTG	s474
МВГ-3	VNIMK	L007	GGAATAC	s475
МВГ-8	VNIMK	L007	TGTAACT	s476
Б 2073	VNIMK	L008	ATGCCGG	s569
№ 424924	VNIMK	L002	ATGCCGG	s137
№ 577083	VNIMK	L008	ATTCTTG	s570
№ 577432	VNIMK	L001	ACGTGAG	s57
№ 577433	VNIMK	L002	CCACCGT	s138
АН 512 Rf	VNIMK	L008	CGATGCG	s571
АН 70029 Rf	VNIMK	L002	CCTACCG	s139
НА 89	VNIMK	L008	ACTTCTG	s572
RHA 265-1	VNIMK	L008	AGAAGTC	s573
RHA 274-1	VNIMK	L001	CCGACTT	s58
RHA 297	VNIMK	L002	GGAGACT	s140
RHA 298	VNIMK	L007	ATGGCTC	s477
Sl 2966	VNIMK	L001	CCAATTG	s59
Z 231	VNIMK	L008	GCCAACT	s574
Z 1064	VNIMK	L007	CCTGCAC	s478
ZB x 231 AC	VNIMK	L008	TGTTCAG	s575
Черняка 66-2	VNIMK	L007	GGATCGT	s479
ЖС-17	VNIMK	L001	CCATTGC	s60
I4BC4 ANN 2188	VNIMK	L002	AGTGGCC	s141
I4BC4 ANN 2165	VNIMK			
I4BC4PET 2203	VNIMK	L002	TAACTGC	s142
ВК 680А	VNIMK	L007	ATCTTGC	s480
ВК 732А	VNIMK	L007	AGTAAGC	s481
ВК 905А	VNIMK	L002	CCTCTAG	s143
ВК 934А	VNIMK	L008	AACGAGT	s576

ВК 935А	VNIMK	L001	TGCGAGG	s61
ВК 932А	VNIMK	L001	AGCCGGT	s62
ВК 623А	VNIMK	L008	AGTGGCC	s577
ВК 901А	VNIMK	L001	CCTGACT	s63
ВК 902А	VNIMK	L007	TGCCTAG	s482
ВК 903А	VNIMK	L008	GGAACTG	s578
ВК 904А	VNIMK	L001	CCGTAAC	s64
ВК 906А	VNIMK	L001	TGAACAT	s65
ВК 900А	VNIMK	L001	AACGTAC	s66
ВК 927А	VNIMK	L007	AGTGGCC	s483
ВК 931А	VNIMK	L001	TAGGCGC	s67
ВК276А	VNIMK	L001	GGAATAC	s68
СЛ12-3732Б	VNIMK	L007	TCTCGAC	s484
СЛ12-3876Б	VNIMK	L007	AACAGAT	s485
СЛ13-2258Б	VNIMK	L008	CCTACCG	s579
СЛ14-4632Б	VNIMK	L001	CCTCTAG	s69
СЛ14-4658Б	VNIMK	L001	GACCTCT	s70
СЛ14-4662Б	VNIMK	L001	TGCCTAG	s71
СЛ14-4646Б	VNIMK	L001	CAGGTGG	s72
СЛ15-4488Б 2203	VNIMK	L008	GGATCGT	s580
СЛ15-4490Б 2205	VNIMK	L008	ACTCGCT	s581
СЛ15-4498Б 2209	VNIMK	L008	CCGACTT	s582
СЛ01-3839А 4383	VNIMK	L001	AACTCAG	s73
ВК499А 4355	VNIMK	L001	TCTCGAC	s74
СЛ15-4038А 4403	VNIMK	L007	TGCTACC	s486
СЛ01-3856А 4397	VNIMK	L001	GTTGAGC	s75
СЛ15-4508Б 2217	VNIMK	L007	GAATACC	s487
СЛ15-4464Б 2195	VNIMK	L002	AACAATG	s144
СЛ15-4480Б 2201	VNIMK	L001	ATGGCTC	s76
СЛ15-4476Б 2197	VNIMK	L002	TAAGAAC	s145
СЛ15-4472Б 2199	VNIMK	L008	ATCTGTT	s583
СЛ15-4520Б 2223	VNIMK	L007	CCATTGC	s488
СЛ15-4526Б 2229	VNIMK	L008	TCTCGAC	s584
СЛ15-4574Б 2275	VNIMK	L008	TAGGCGC	s585
СЛ15-4550Б 2273	VNIMK	L001	TTCTGAG	s77
СЛ15-4544Б 2271	VNIMK	L002	TGCCTAG	s146
СЛ15-4544А 2270	VNIMK	L007	ACTCGCT	s489
СЛ15-4514Б 2267	VNIMK	L001	ATCTGTT	s78

СЛ15-4528Б	2231	VNIMK	L002	CCTGACT	s147
СЛ15-4542Б	2241	VNIMK	L008	ATGGCTC	s586
СЛ15-4536Б	2263	VNIMK	L001	CACACGT	s79
СЛ15-4798Б	2259	VNIMK	L002	ATGGCTC	s148
ВК838А	2064	VNIMK	L007	TAGAACG	s490
СЛ13-2210А	2256	VNIMK	L007	GTAGGTC	s491
СЛ09-4750А	2970	VNIMK	L008	TAACTGC	s587
СЛ13-2186А	2076	VNIMK	L008	ACGTTGT	s588
СЛ13-2182А	2074	VNIMK	L008	TTGAGGC	s589
СЛ06-2545А	2038	VNIMK	L002	CGATGCG	s149
СЛ05-4154А	2036	VNIMK	L008	AAGGTCT	s590
СЛ13-2190А	2086	VNIMK	L001	GTTCGAT	s80
СЛ05-16А	2090	VNIMK	L001	CCTCAGC	s81
СЛ13-2756А	2254	VNIMK	L001	GTAACCT	s82
СЛ13-2260А	2252	VNIMK	L008	CCGGTAT	s591
СЛ13-2258А	2250	VNIMK	L008	ACCATAG	s592
СЛ13-2280А	2112	VNIMK	L007	TGTTCAG	s492
СЛ13-2224А	2110	VNIMK	L001	CCGGTAT	s83
СЛ13-2302А	2106	VNIMK			
СЛ13-2226А	2102	VNIMK	L008	AGTAAGC	s593
СЛ13-2218А	2098	VNIMK	L001	AAGGTCT	s84
СЛ13-2206А	2092	VNIMK	L008	TGAACAT	s594
СЛ13-2220А	2100	VNIMK	L007	TATGTTC	s493
СЛ15-4500Б	2265	VNIMK	L008	CCTCTAG	s595
СЛ13-2270Б	2121	VNIMK	L002	TACTGTC	s150
СЛ16-2278Б	2279	VNIMK	L008	TTGCATC	s596
СЛ16-2286Б	2287	VNIMK	L001	GATGGCT	s85
СЛ16-2282Б	2283	VNIMK	L008	CCTAGAT	s597
СЛ16-2290Б	2291	VNIMK	L001	GGTTGAC	s86
СЛ16-2284Б	2285	VNIMK	L008	GATGGCT	s598
СЛ16-2292Б	2293	VNIMK	L001	TATGTTC	s87
СЛ16-2290Б	2291	VNIMK	L007	CCTAGAT	s494
СЛ15-4472Б	2295	VNIMK	L001	ACGTTGT	s88
СЛ11-5244Б	2143	VNIMK	L001	GTGGACG	s89
СЛ05-4770А	2084	VNIMK	L002	CGAGGTT	s151
ВК276Б	4244	VNIMK	L008	CCTGCAC	s599
СЛ01-3828А	2021	VNIMK	L001	TAGAACG	s90
СЛ13-2210А	2256	VNIMK	L002	GACCTCT	s152

СЛ13-2232Б	2145	VNIMK	L002	CACACGT	s153
СЛ15-4512Б	4533	VNIMK	L007	TTGCATC	s495
СЛ16-2288Б	2289	VNIMK	L001	AACAATG	s91
СЛ11-5244Б	2143	VNIMK	L001	CCGATCC	s92
ВК678А	2007	VNIMK	L007	TTCACTG	s496
ВК639А	2005	VNIMK	L001	ATCTTGC	s93
ВК464А	2003	VNIMK	L007	TAAGAAC	s497
СЛ13-2256А	2114	VNIMK	L007	CCACCGT	s498
ВК861А	2116	VNIMK	L007	GGAGACT	s499
СЛ13-2270А	2120	VNIMK	L008	CGTGGAG	s600
СЛ13-2268А	2118	VNIMK	L007	CCTGACT	s500
СЛ13-2284А	2122	VNIMK	L008	AACTCAG	s601
СЛ13-2288А	2124	VNIMK	L007	TTCTAGT	s501
СЛ14-4638Б	2249	VNIMK	L007	TGATCTC	s502
СЛ13-2228А	2140	VNIMK	L001	GAATACC	s94
ВК653А	2136	VNIMK	L007	GGTTGAC	s503
СЛ15-2294А	2126	VNIMK	L001	ATTCTTG	s95
СЛ13-2246А		VNIMK	L007	CACATCG	s504
СЛ13-2258А		VNIMK	L007	CGGAGGT	s505
СЛ13-2260А		VNIMK	L002	TTGAGGC	s154
2261х 6232		VNIMK	L001	TTCTAGT	s96
ВИР 100А		VIR	L005	GATAGGC	s407
ВИР 100Б		VIR			
ВИР 101А		VIR	L004	CACATCG	s339
ВИР 101Б		VIR	L003	GAATACC	s240
ВИР 111А		VIR	L004	CCTACCG	s342
ВИР 114А		VIR	L004	ACGTTGT	s335
ВИР 114Б		VIR	L005	ATCTGTT	s401
ВИР 116А		VIR	L003	CCACCGT	s231
ВИР 116Б		VIR	L005	TTGGCAT	s402
ВИР 117А		VIR	L003	AACAATG	s234
ВИР 117Б		VIR	L005	AAGGTCT	s404
ВИР 125		VIR	L004	ATCTTGC	s325
ВИР 128		VIR	L003	CGAGGTT	s157
ВИР 129А		VIR	L005	CCGACTT	s403
ВИР 129Б		VIR	L003	CTAATGT	s233
ВИР 130Б		VIR	L004	TTGTTAC	s322
ВИР 136		VIR	L003	CTCTATG	s155

ВИР 137Б	VIR	L005	ATGTTCG	s405
ВИР 138А	VIR	L004	GGACATC	s336
ВИР 138Б	VIR	L003	TGTAACT	s232
ВИР 140	VIR	L003	CCAATTG	s215
ВИР 151А	VIR	L003	GGTTGAC	s237
ВИР 151RIGO	VIR			
ВИР 151Б	VIR	L004	TGTAACT	s338
ВИР 156	VIR	L003	CCGGTAT	s171
ВИР 162	VIR	L003	CGGAGGT	s179
ВИР 165	VIR	L004	CCTCTAG	s251
ВИР 171	VIR	L003	CCGCATT	s182
ВИР 172А	VIR	L003	AACGAGT	s228
ВИР 172Б	VIR	L004	ACCATAG	s332
ВИР 177	VIR	L004	CTAATGT	s273
ВИР 183	VIR	L004	AGCCGGT	s287
ВИР 184	VIR	L004	TTCACTG	s326
ВИР 190	VIR	L003	AACAGAT	s170
ВИР 195	VIR			
ВИР 196	VIR	L004	CCGATCC	s253
ВИР 197	VIR	L004	AACGAGT	s294
ВИР 200Б	VIR			
ВИР 205А	VIR	L003	TGCCTAG	s238
ВИР 210	VIR	L005	CCTCTAG	s356
ВИР 211	VIR	L004	GAACAAT	s295
ВИР 215А	VIR	L003	ATTGCGT	s236
ВИР 215Б	VIR	L004	CCACCGT	s337
ВИР 229А	VIR	L004	AGTGGCC	s296
ВИР 229Б	VIR	L004	AAGGTCT	s334
ВИР 230	VIR	L004	CCATTGC	s274
ВИР 234	VIR	L003	CACATCG	s158
ВИР 263	VIR	L005	ACGTTGT	s373
ВИР 278	VIR	L003	GATAGGC	s218
ВИР 319	VIR	L005	CGGAGGT	s358
ВИР 328	VIR	L004	GACCTCT	s318
ВИР 338	VIR	L005	TTCTGAG	s390
ВИР 339	VIR	L004	AACAATG	s297
ВИР 340Б	VIR			
ВИР 343	VIR	L004	GTTGAGC	s329

ВИР 349	VIR	L003	GTGGACG	s161
ВИР 362	VIR	L003	CCTGACT	s217
ВИР 364	VIR	L003	TTGGCAT	s188
ВИР 369	VIR	L005	CCTGCAC	s374
ВИР 370	VIR	L003	GACCTCT	s189
ВИР 371	VIR	L004	CCGTAAC	s298
ВИР 376	VIR	L005	CCAATTG	s348
ВИР 378	VIR	L003	GATGGCT	s190
ВИР 381	VIR	L005	TAGAACG	s349
ВИР 386	VIR	L003	TGCTACC	s191
ВИР 388	VIR	L005	TAAGAAC	s376
ВИР 395	VIR	L003	TAGAACG	s192
ВИР 421	VIR			
ВИР 435А	VIR			
ВИР 436Б	VIR			
ВИР 438	VIR	L004	AACTCAG	s299
ВИР 445	VIR	L004	TGATCTC	s331
ВИР 446	VIR	L003	TAGGCGC	s224
ВИР 448	VIR	L003	AACGTAC	s225
ВИР 450	VIR	L004	ATTCTTG	s256
ВИР 452	VIR	L004	GTTACTC	s300
ВИР 453	VIR	L005	AACAATG	s395
ВИР 456	VIR	L005	CCTACCG	s397
ВИР 471А	VIR	L003	TGGCAGT	s229
ВИР 471Б	VIR	L003	CCGACTT	s230
ВИР 479	VIR	L003	TTGCATC	s216
ВИР 480	VIR	L003	ATGTTCG	s193
ВИР 490	VIR	L005	GAACAAT	s377
ВИР 501	VIR	L005	GGACATC	s360
ВИР 581	VIR	L003	CGTGGAG	s194
ВИР 584	VIR	L004	TTCCACG	s301
ВИР 630	VIR	L005	CCTAGAT	s378
ВИР 631	VIR	L003	CCATTGC	s195
ВИР 632	VIR	L003	CCTGCAC	s162
ВИР 633	VIR	L004	CACACGT	s303
ВИР 634	VIR	L003	ATCTTGC	s196
ВИР 635	VIR	L003	GTTCGAT	s163
ВИР 636	VIR	L004	TGTTCAG	s304

ВИР 637	VIR	L004	GGACTCG	s258
ВИР 641	VIR	L004	GGAGACT	s259
ВИР 644	VIR	L004	CCACGTC	s305
ВИР 645	VIR	L004	TGGCAGT	s306
ВИР 646	VIR	L003	GAGGCTG	s221
ВИР 648	VIR	L004	GTAGGTC	s319
ВИР 649	VIR	L004	GTAACCT	s260
ВИР 651	VIR	L005	TTCCACG	s351
ВИР 652	VIR	L004	GCCTCAC	s307
ВИР 655	VIR	L004	ATTGCGT	s308
ВИР 656	VIR	L003	TTCACTG	s197
ВИР 658	VIR	L004	ATGTTCG	s261
ВИР 665	VIR	L003	CACACGT	s212
ВИР 679	VIR	L005	CCTGACT	s399
ВИР 682	VIR	L003	TAACTGC	s198
ВИР 684	VIR	L004	TGCTACC	s309
ВИР 692	VIR	L004	AGAAGTC	s275
ВИР 697	VIR	L004	AGCTTCT	s310
ВИР 700	VIR	L003	GCCAACT	s199
ВИР 702	VIR	L004	GAATACC	s311
ВИР 703	VIR	L005	TGCCTAG	s379
ВИР 704	VIR	L005	ATGCCGG	s380
ВИР 708	VIR	L005	GGAATAC	s391
ВИР 725	VIR	L003	TTCCACG	s174
ВИР 726	VIR	L003	GTAACCT	s200
ВИР 728	VIR	L003	TTCTGAG	s223
ВИР 730	VIR	L004	CCTCAGC	s327
ВИР 734	VIR	L005	AAGCGTG	s359
ВИР 739	VIR	L004	CCTGACT	s312
ВИР 740	VIR	L005	GGAACTG	s388
ВИР 743	VIR	L003	GAGTGCG	s201
ВИР 745	VIR	L005	ATTCTTG	s361
ВИР 747	VIR	L005	TTCTAGT	s396
ВИР 749	VIR	L005	ATGGCTC	s398
ВИР 751	VIR	L004	CCTGCAC	s314
ВИР 752	VIR	L004	TTGAGGC	s262
ВИР 753	VIR	L003	GGAACTG	s202
ВИР 755	VIR	L005	TACTGTC	s364

ВИР 757	VIR	L005	GGTTGAC	s381
ВИР 758	VIR	L004	CCTAGAT	s263
ВИР 759	VIR	L004	TGAACAT	s264
ВИР 762	VIR	L005	GGAGACT	s382
ВИР 763	VIR	L004	AGACCTT	s265
ВИР 764	VIR	L003	AAGCGTG	s164
ВИР 765	VIR	L003	ATGGCTC	s187
ВИР 767	VIR	L005	GCCAACT	s352
ВИР 768	VIR	L005	GAGGCTG	s353
ВИР 769	VIR	L005	ATCTTGC	s354
ВИР 770	VIR	L003	GCCTCAC	s165
ВИР 772	VIR	L003	ACGTTGT	s208
ВИР 776	VIR	L003	TATGTTC	s203
ВИР 777	VIR	L004	TTAGCCG	s266
ВИР 786	VIR	L005	TTAGCCG	s392
ВИР 787	VIR	L003	GGAATAC	s214
ВИР 796	VIR	L004	TTCTAGT	s267
ВИР 800	VIR	L003	TGCGAGG	s166
ВИР 817	VIR	L005	GCCTCAC	s383
ВИР 819	VIR	L005	TGCTACC	s371
ВИР 821	VIR	L003	AACTCAG	s175
ВИР 823	VIR	L003	AGTAAGC	s204
ВИР 825	VIR	L004	TATGTTC	s276
ВИР 830	VIR	L004	GAGTGCG	s289
ВИР 832	VIR	L004	CAGGTGG	s290
ВИР 833	VIR	L004	GGAATAC	s291
ВИР 836	VIR	L004	CCGCATT	s316
ВИР 839	VIR	L003	AAGGTCT	s185
ВИР 840	VIR	L004	CCGGTAT	s293
ВИР 841	VIR	L005	GACCTCT	s372
ВИР 843	VIR	L003	CCTAGAT	s205
ВИР 846	VIR	L003	ATGCCGG	s245
ВИР 847	VIR	L003	GGACTCG	s246
ВИР 848	VIR	L005	GAATACC	s408
ВИР 849	VIR	L004	CCAATTG	s343
ВИР 850	VIR	L003	CAGGTGG	s247
ВИР 855	VIR	L003	CGATGCG	s248
ВИР 858	VIR	L004	TAACTGC	s344

ВИР 859	VIR	L005	GTTGAGC	s409
ВИР 860	VIR	L003	TTGAGGC	s249
ВИР 900	VIR			
ВИР 902	VIR	L005	CCTCAGC	s386
ВИР 903	VIR	L004	CTCTATG	s324
ВИР АМ	VIR			
АД-66	VIR	L004	CGGAGGT	s271
L 2088	VIR	L004	CCGACTT	s269
ВД-690	VIR	L003	ACCATAG	s168
СМ 13	VIR	L003	CCACGTC	s180
СМ 27	VIR			
СМ 32	VIR	L004	ATCTGTT	s286
СМ 150	VIR	L004	ACTCGCT	s280
СМ 338	VIR			
СМ 359	VIR	L003	CCGATCC	s178
СМ 597	VIR	L005	CACACGT	s400
PI 331174	VIR			
ТА 716-18	VIR	L004	GGAACTG	s283
ТА 717	VIR			
ТА 728	VIR	L003	CCTACCG	s181
ТА 6463	VIR	L004	GGTTGAC	s272
RIL-80	VIR	L004	CGATGCG	s317
RIL-130	VIR	L004	TTGGCAT	s268
RIL-341	VIR			
RIL-353	VIR			
RIL-365	VIR			
ВБ-273	VIR	L004	TTGCATC	s279
ВК 571	VIR	L003	TTAGCCG	s183
ВК 17	VIR	L005	GATCTGG	s367
ВК 20	VIR	L005	TTGTTAC	s355
ВК 47	VIR	L003	AGCTTCT	s169
ВК 51	VIR	L004	ACGTGAG	s281
ВК 54	VIR			
ВК 580	VIR	L005	CCACCGT	s366
с.Зеленка	VIR			
с.ВНИИМК 8932	VIR			
с.Прогресс	VIR			
с.Армавирский 1813	VIR			

с.Армавирский 9345	VIR			
с.Спутник	VIR			
с.Чернянка 35	VIR			
HS 310	VIR			
У.З.Р.2-19	VIR			
xxx????k-1532	VIR			
Д-49 х 1069	VIR			
xxxx??? К-2238	VIR			
х 2006	VIR			
Guaycan YNTA	VIR			
х 782 Rf	VIR	L005	TAGGCGC	s365
ВИР 110А	VIR	L003	GGAGACT	s244
ВИР 110Б	VIR			
ВИР 137А	VIR	L003	ACTCGCT	s235
ВИР 143	VIR	L004	GATAGGC	s282
ВИР 158	VIR			
ВИР 159	VIR	L004	GCCAACT	s252
ВИР 220	VIR	L004	AAGCGTG	s254
ВИР 229А	VIR	L004	CGTGGAG	s333
ВИР 249	VIR	L003	AGACCTT	s159
ВИР 260	VIR	L004	TAGAACG	s255
ВИР 265	VIR	L003	TGTTCAG	s172
ВИР 283	VIR	L003	ACTTCTG	s173
ВИР 302	VIR	L003	TACTGTC	s160
ВИР 365	VIR	L004	AGTAAGC	s330
ВИР 387	VIR	L005	TAACTGC	s375
ВИР 449	VIR	L005	ATTGCGT	s350
ВИР 488	VIR	L004	GTTCGAT	s321
ВИР 580	VIR	L004	GAGGCTG	s257
ВИР 607	VIR	L003	TAAGAAC	s211
ВИР 674	VIR	L003	GGATCGT	s226
ВИР 788	VIR	L004	ATGCCGG	s328
ВИР 791	VIR	L003	CCGTAAC	s207
ВИР 795	VIR	L005	TTGAGGC	s369
ВИР 815	VIR	L003	GTTACTC	s210
ВИР 818	VIR	L005	CCGCATT	s370
ВИР 826	VIR	L004	TCTCGAC	s320
КГ-49	VIR	L003	GTTGAGC	s213

НА 89А	VIR	L003	AGCCGGT	s241
НА 89Б	VIR	L004	TAGGCGC	s340
СМ 142	VIR	L003	ATTCTTG	s177
RIL-258/1	VIR			
RIL 265	VIR			
Ждановский 6432	VIR			
ВНИИМК 8931	VIR	L003	AGAAGTC	s156
с. Одесский 19	VIR			
ВИР 106Б	VIR	L003	ATCTGTT	s243
ВИР 128	VIR	L004	GATCTGG	s270
ВИР 130	VIR	L003	GATCTGG	s227
ВИР 172	VIR	L004	TGCCTAG	s284
ВИР 175	VIR	L004	AACGTAC	s285
ВИР 205Б	VIR	L003	CCTCAGC	s239
ВИР 253	VIR	L005	AGCCGGT	s357
ВИР 340А	VIR	L003	ACGTGAG	s219
ВИР 372	VIR	L005	GGACTCG	s393
ВИР 397	VIR	L005	AGCTTCT	s389
ВИР 434	VIR	L005	TGAACAT	s387
ВИР 434А	VIR	L005	CGATGCG	s406
ВИР 434Б	VIR	L003	GGACATC	s242
ВИР 436А	VIR	L004	ATGGCTC	s341
ВИР 632	VIR	L004	GTGGACG	s302
ВИР 721	VIR	L003	TTGTTAC	s220
ВИР 743	VIR	L004	GGATCGT	s313
ВИР 744	VIR	L005	AGAAGTC	s363
ВИР 759	VIR	L003	TCTCGAC	s209
ВИР 766	VIR	L005	AGTGGCC	s368
ВИР 773	VIR	L004	AACAGAT	s278
ВИР 794	VIR	L004	TACTGTC	s288
ВИР 801	VIR	L003	CCTCTAG	s167
ВИР 814	VIR	L003	TGAACAT	s186
ВИР 820	VIR	L003	TTCTAGT	s184
ВИР 821	VIR	L005	CGAGGTT	s384
ВИР 823	VIR	L005	AACAGAT	s362
ВИР 825	VIR	L005	CCGTAAC	s385
ВИР 833	VIR	L004	TAAGAAC	s315
ВИР 834	VIR	L003	GAACAAT	s222

ВИР 835	VIR	L004	CGAGGTT	s292
ВИР 839	VIR	L004	TTCTGAG	s277
ВИР 845	VIR	L004	TGCGAGG	s323
ВИР 861	VIR	L004	GATGGCT	s345
ВНИИМК 8932	VIR	L005	CGTGGAG	s347
КГ 104	VIR	L003	AGTGGCC	s176
Китай с-03002	VIR	L004	ACTTCTG	s346
Китай Хейя-2	VIR	L003	GTAGGTC	s250
СЛ 2290	VIR	L005	ACTTCTG	s394
х 712	VIR	L003	TGATCTC	s206

Samples for Genotype by environment interactions test

sample_id	line_name	year	Collection
148/1	ВК732А	2015	VNIIMK
148/2	ВК732А	2015	VNIIMK
148/3	ВК732А	2015	VNIIMK
148/4	ВК732А	2015	VNIIMK
148/5	ВК732А	2015	VNIIMK
149/1	ВК732А	2016	VNIIMK
148/2	ВК732А	2016	VNIIMK
149/3	ВК732А	2016	VNIIMK
149/4	ВК732А	2016	VNIIMK
149/5	ВК732А	2016	VNIIMK
150/1	ВК732А	2017	VNIIMK
150/2	ВК732А	2017	VNIIMK
150/3	ВК732А	2017	VNIIMK
150/4	ВК732А	2017	VNIIMK
150/5	ВК732А	2017	VNIIMK
151/1	ВК876А	2014	VNIIMK
151/2	ВК876А	2014	VNIIMK
151/3	ВК876А	2014	VNIIMK
151/4	ВК876А	2014	VNIIMK
151/5	ВК876А	2014	VNIIMK
152/1	ВК876А	2015	VNIIMK
152/2	ВК876А	2015	VNIIMK
152/3	ВК876А	2015	VNIIMK

152/4	BK876A	2015	VNIIMK
152/5	BK876A	2015	VNIIMK
153/1	BK876A	2016	VNIIMK
153/2	BK876A	2016	VNIIMK
153/3	BK876A	2016	VNIIMK
153/4	BK876A	2016	VNIIMK
153/5	BK876A	2016	VNIIMK
154/1	BK906A	2015	VNIIMK
154/2	BK906A	2015	VNIIMK
154/3	BK906A	2015	VNIIMK
154/4	BK906A	2015	VNIIMK
154/5	BK906A	2015	VNIIMK
155/1	BK906A	2016	VNIIMK
155/2	BK906A	2016	VNIIMK
155/3	BK906A	2016	VNIIMK
155/4	BK906A	2016	VNIIMK
155/5	BK906A	2016	VNIIMK
156/1	BK906A	2017	VNIIMK
156/2	BK906A	2017	VNIIMK
156/3	BK906A	2017	VNIIMK
156/4	BK906A	2017	VNIIMK
156/5	BK906A	2017	VNIIMK
157/1	BK931A	2015	VNIIMK
157/2	BK931A	2015	VNIIMK
157/3	BK931A	2015	VNIIMK
157/4	BK931A	2015	VNIIMK
157/5	BK931A	2015	VNIIMK
158/1	BK931A	2016	VNIIMK
158/2	BK931A	2016	VNIIMK
158/3	BK931A	2016	VNIIMK
158/4	BK931A	2016	VNIIMK
158/5	BK931A	2016	VNIIMK
159/1	BK931A	2017	VNIIMK
159/2	BK931A	2017	VNIIMK
159/3	BK931A	2017	VNIIMK
159/4	BK931A	2017	VNIIMK
159/5	BK931A	2017	VNIIMK
160/1	BK934A	2015	VNIIMK

160/2	BK934A	2015	VNIIMK
160/3	BK934A	2015	VNIIMK
160/4	BK934A	2015	VNIIMK
160/5	BK934A	2015	VNIIMK
161/1	BK934A	2016	VNIIMK
161/2	BK934A	2016	VNIIMK
161/3	BK934A	2016	VNIIMK
161/4	BK934A	2016	VNIIMK
161/5	BK934A	2016	VNIIMK
162/1	BK934A	2017	VNIIMK
162/2	BK934A	2017	VNIIMK
162/3	BK934A	2017	VNIIMK
162/4	BK934A	2017	VNIIMK
162/5	BK934A	2017	VNIIMK
163/1	BK935A	2015	VNIIMK
163/2	BK935A	2015	VNIIMK
163/3	BK935A	2015	VNIIMK
163/4	BK935A	2015	VNIIMK
163/5	BK935A	2015	VNIIMK
164/1	BK935A	2016	VNIIMK
164/2	BK935A	2016	VNIIMK
164/3	BK935A	2016	VNIIMK
164/4	BK935A	2016	VNIIMK
164/5	BK935A	2016	VNIIMK
165/1	BK935A	2017	VNIIMK
165/2	BK935A	2017	VNIIMK
165/3	BK935A	2017	VNIIMK
165/4	BK935A	2017	VNIIMK
165/5	BK935A	2017	VNIIMK

Annex for chapter 4

Table S1

186 inbred *Helianthus annuus* L. lines

line	Origin	sample	H	D	Br	Fertility	Ray flo	DTB	DTF	DTM	HU
K 223	VIR accession	s122	175,6	15,2	1	1	1	52	81	111	1785,15
K 225	VIR accession	s452	127,8	13,2	1	1	1	41	68	97	1487,03
K 370	VIR accession	s123	116,6	13,6	1	1	1	38	58	89	1318,44
K 562	VIR accession	s552	103,8	21,6	0	1	1	43	71	96	1465,05
K 581	VIR accession	s553	106,6	14,8	1	1	1	40	62	91	1355
K 651-3	VIR accession	s453	125,8	15,6	1	1	1	41	63	91	1355
K 912	VIR accession	s124	92,8	28,8	0	1	1	42	66	96	1465,05
K 1391	VIR accession	s35	105,4	14	1	1	1	40	60	93	1397,66
K 1459	VIR accession	s454	147,2	18,2	1	1	1	44	70	102	1599,53
K 1505	VIR accession	s455	111	12,2	1	1	1	38	59	89	1318,44
K 1594	VIR accession	s456	113,6	18,6	0	1	1	38	59	89	1318,44
K 1687	VIR accession	s457	125,8	20,8	0	1	1	39	59	94	1419,99
K 2068	VIR accession	s36	110,8	11,4	1	1	1	37	57	89	1318,44
K 2086	VIR accession	s37	131,6	21,2	0	1	1	50	78	109	1744,09
K 2125	VIR accession	s125	129	19,8	0	1	1	42	66	96	1465,05
K 2235	VIR accession	s458	151,4	23,2	1	1	1	43	67	95	1443,77
K 2238	VIR accession	s554	99,6	18	0	1	1	41	65	96	1465,05
K 2257	VIR accession	s126	94,2	21	0	1	1	42	66	96	1465,05
K 2462	VIR accession	s127	128	18,6	0	1	1	43	72	102	1599,53
K 2479	VIR accession	s38	143	26,6	0	1	1	44	77	107	1703,58
K 3035	VIR accession	s128	101,4	20	0	1	1	40	70	96	1465,05
K 3059	VIR accession	s39	152,8	22,8	0	1	1	39	66	93	1397,66
K 3159	VIR accession	s555	93,8	16,4	1	1	1	38	62	94	1419,99
K 3350	VIR accession	s556	92,8	17,4	0	1	1	39	61	89	1318,44
K 3376	VIR accession	s557	109,4	28,8	0	1	1	40	70	103	1621,36
SL 1721	VNIIMK line	s558	143,8	22,4	0	1	1	41	73	106	1682,4
SL 1790	VNIIMK line	s459	143	25,4	0	1	1	40	66	96	1465,05
SL 1813	VNIIMK line	s40	105,4	19,2	0	1	1	38	61	89	1318,44
SL 2039	VNIIMK line	s460	141,6	14,8	1	1	1	40	70	96	1465,05
SL 2950	VNIIMK line	s41	108,8	14	1	1	1	38	61	88	1300,91
VA 1	VNIIMK line	s559	134,2	23,8	0	1	1	40	68	96	1465,05
VA 4	VNIIMK line	s42	144	29,2	0	1	0	41	66	94	1419,99

VIR 130-1	VIR line	s129	149	24	0	1	1	43	71	102	1599,53
VIR 172	VIR line	s43	103,8	13,8	1	1	1	42	67	96	1465,05
VIR 369	VIR line	s461	138,6	22,6	0	1	1	45	70	96	1465,05
VIR 391	VIR line	s44	120	16,2	1	1	1	43	67	96	1465,05
VK 15	VNIIMK line	s45	114,2	24,2	0	1	1	40	64	96	1465,05
VK 30	VNIIMK line	s462	81,6	27	0	1	1	40	61	94	1419,99
VK 102	VNIIMK line	s130	70,8	18,2	1	1	1	39	61	94	1419,99
VK 206	VNIIMK line	s46	120,8	25	0	1	1	41	66	95	1443,77
VK 268	VNIIMK line	s463	99,4	21	0	1	1	40	63	89	1318,44
VK 416	VNIIMK line	s560	87,4	18,8	0	1	1	41	65	96	1465,05
VK 428	VNIIMK line	s47	104,4	14	0	1	1	40	63	87	1283,33
VK 464	VNIIMK line	s48	120,8	21,8	0	1	0	47	76	107	1703,58
VK 474	VNIIMK line	s49	127,6	21,4	0	1	1	45	75	96	1465,05
VK 475	VNIIMK line	s131	157	23	0	1	1	45	76	106	1682,4
VK 519	VNIIMK line	s50	111,6	19,6	0	1	1	41	63	87	1283,33
LG 3	VNIIMK line	s464	151,6	15	1	1	1	45	75	96	1465,05
LG 8-2	VNIIMK line	s465	110	17,2	1	1	1	42	66	96	1465,05
LG 10	VNIIMK line	s132	82,2	13,4	0	1	1	38	58	81	1164,3
LG 26	VNIIMK line	s466	148	21,4	0	1	1	43	67	96	1465,05
LG 27	VNIIMK line	s51	146,2	24,4	0	1	1	45	74	105	1661,72
LG 28	VNIIMK line	s467	157	21,6	0	1	1	46	76	109	1744,09
KG 7	VNIIMK line	s133	98,2	22,4	0	1	1	40	61	96	1465,05
KG 16	VNIIMK line	s134	150,4	22,4	0	1	1	45	74	104	1641,74
KG 19	VNIIMK line	s561	50,8	12	0	1	1	37	54	80	1146,82
KG 21	VNIIMK line	s562	143,2	26,8	1	1	0	50	80	110	1764,02
KG 32	VNIIMK line	s135	108	14,2	1	1	1	42	65	87	1283,33
KG 48	VNIIMK line	s468	103	18,2	0	1	1	46	76	105	1661,72
KG 104	VNIIMK line	s563	96,2	23,8	0	1	1	43	67	95	1443,77
L 1392	VNIIMK line	s564	94	18,8	0	1	1	42	64	96	1465,05
L 2090	VNIIMK line	s565	119,4	24	0	1	1	46	74	105	1661,72
L 2138	VNIIMK line	s52	92	20,4	0	1	1	40	69	98	1509,91
L 2532	VNIIMK line	s566	145,2	23,6	0	1	1	42	72	103	1621,36
L 2543	VNIIMK line	s136	106,6	16,2	1	1	1	40	69	97	1487,03
L 2544	VNIIMK line	s469	89,2	14,4	1	1	1	38	60	89	1318,44

L 2563	VNIIMK line	s53	109,5	15,5	1	1	1	41	67	96	1465,05
L 2582	VNIIMK line	s470	96,4	14,4	1	1	1	43	73	104	1641,74
L 2586	VNIIMK line	s54	88,8	16,2	0	1	1	39	63	89	1318,44
L 2595	VNIIMK line	s55	103,6	21,6	0	1	1	40	65	94	1419,99
L 3376	VNIIMK line	s471	119,4	25,4	0	1	1	43	73	103	1621,36
L 7247	VNIIMK line	s472	123,4	28,6	0	1	1	43	71	100	1554,67
M 1046	VNIIMK line	s567	99,4	10,4	0	1	1	51	82	110	1764,02
I7-235	VNIIMK line	s56	62,6	24,4	0	1	1	43	71	101	1576,7
I7-246	VNIIMK line	s568	60,6	20,8	0	1	1	44	74	102	1599,53
№1416-5	VNIIMK line	s473	65	12,6	0	1	1	43	75	105	1661,72
№355114	VNIIMK line	s474	112	12,2	1	1	1	40	65	96	1465,05
MVG-3	VNIIMK line	s475	123,8	21	0	1	1	41	66	95	1443,77
MVG-8	VNIIMK line	s476	158	29,8	0	1	1	45	74	104	1641,74
B 2073	VNIIMK line	s569	159,6	27,2	0	1	1	46	76	106	1682,4
№ 424924	VIR accession	s137	96,6	11,4	1	1	1	39	66	96	1465,05
№ 577083	VIR accession	s570	61,2	14,2	0	1	1	36	55	79	1128,99
№ 577432	VIR accession	s57	99,8	18,8	0	1	1	40	59	94	1419,99
№ 577433	VIR accession	s138	95	17	0	1	1	43	62	91	1355
AH 512 Rf	VIR accession	s571	128,4	17,6	1	1	1	45	70	96	1465,05
AH 70029 Rf	VIR accession	s139	77,2	13,4	1	1	1	42	61	96	1465,05
HA 89	USDA line	s572	94,4	19	0	1	1	41	67	89	1318,44
RHA 265-1	USDA line	s573	130,2	23,4	0	1	1	43	72	102	1599,53
RHA 274-1	USDA line	s58	126,4	12,6	1	1	1	41	67	96	1465,05
RHA 297	USDA line	s140	126	14,2	1	1	1	40	64	89	1318,44
RHA 298	USDA line	s477	124	15,4	1	1	1	42	67	96	1465,05
Sl 2966	VNIIMK line	s59	107,4	16	0	1	1	40	63	96	1465,05
Z 231	VNIIMK line	s574	113,8	19,2	0	1	1	40	63	90	1335,37
Z 1064	VNIIMK line	s478	65,2	20,2	0	1	1	41	63	91	1355
ZB x 231 AC	VNIIMK line	s575	121,2	22,2	0	1	1	42	67	93	1397,66
Chernyanka 66-2	VNIIMK line	s479	70,4	16,2	1	1	1	38	59	89	1318,44
JS-17	VNIIMK line	s60	133,8	30,8	0	1	1	40	63	96	1465,05
I4BC4 ANN 2188	NS line	s141	136,4	23,6	0	1	1	45	70	100	1554,67
I4BC4PET 2203	NS line	s142	108	19,4	0	1	1	43	74	103	1621,36
VK 680A	VNIIMK line	s480	130	14,6	0	0	1	43	74	109	1744,09

VK 732A	VNIIMK line	s481	122	17	0	0	1	44	75	110	1764,02
VK 905A	VNIIMK line	s143	135	18	0	0	1	41	71	106	1682,4
VK 934A	VNIIMK line	s576	126	16,2	0	0	1	45	79	114	1828,24
VK 935A	VNIIMK line	s61	119	19	0	0	1	43	75	110	1764,02
VK 932A	VNIIMK line	s62	135	16,8	0	0	1	44	76	111	1785,15
VK 623A	VNIIMK line	s577	116	19,6	0	0	1	41	71	106	1682,4
VK 901A	VNIIMK line	s63	112	15,6	0	0	1	40	69	104	1641,74
VK 902A	VNIIMK line	s482	116	15,4	0	0	1	41	69	104	1641,74
VK 903A	VNIIMK line	s578	131	17,6	0	0	1	45	74	109	1744,09
VK 904A	VNIIMK line	s64	96	15,4	0	0	1	42	71	106	1682,4
VK 906A	VNIIMK line	s65	130	16,2	0	0	1	43	73	108	1723,91
VK 900A	VNIIMK line	s66	134	14,8	0	0	1	45	76	111	1785,15
VK 927A	VNIIMK line	s483	120	15,2	0	0	1	40	67	102	1599,53
VK 931A	VNIIMK line	s67	128	16,6	0	0	1	40	67	102	1599,53
VK276A	VNIIMK line	s68	69	15,4	0	0	1	38	62	97	1487,03
SL12-3732B	VNIIMK line	s484	135,6	26,4	0	1	1	41	67	102	1599,53
SL12-3876B	VNIIMK line	s485	140	16,6	0	1	1	43	72	107	1703,58
SL13-2258B	VNIIMK line	s579	123,4	13	0	1	1	42	65	100	1554,67
SL14-4632B	VNIIMK line	s69	118,4	16,4	0	1	1	45	70	105	1661,72
SL14-4658B	VNIIMK line	s70	112,8	21,2	0	1	1	44	69	104	1641,74
SL14-4662B	VNIIMK line	s71	111,6	15,4	0	1	1	46	74	109	1744,09
SL14-4646B	VNIIMK line	s72	113,4	16,2	0	1	1	41	67	102	1599,53
SL15-4488B	VNIIMK line	s580	157,2	22,4	0	1	1	43	69	104	1641,74
SL15-4490B	VNIIMK line	s581	117,2	13	0	1	1	42	71	106	1682,4
SL15-4498B	VNIIMK line	s582	144,8	18,2	0	1	1	43	71	106	1682,4
SL01-3839A	VNIIMK line	s73	101,6	14,8	0	0	1	40	66	101	1576,7
VK499A	VNIIMK line	s74	109,8	19,8	0	0	1	41	66	101	1576,7
SL15-4038A	VNIIMK line	s486	79,6	17,2	0	0	1	42	66	101	1576,7
SL01-3856A	VNIIMK line	s75	105,6	22,4	0	0	1	40	62	97	1487,03
SL15-4508B	VNIIMK line	s487	126,6	12,4	0	1	1	46	76	111	1785,15
SL15-4464B	VNIIMK line	s144	124,6	18,2	0	1	1	42	66	101	1576,7
SL15-4480B	VNIIMK line	s76	126	19,2	0	1	1	42	65	100	1554,67
SL15-4476B	VNIIMK line	s145	143	15,6	1	1	1	45	73	108	1723,91
SL15-4520B	VNIIMK line	s488	118,8	13,6	0	1	1	44	73	108	1723,91

SL15-4526B	VNIIMK line	s584	114,2	14,4	0	1	1	44	74	109	1744,09
SL15-4574B	VNIIMK line	s585	117,8	13,6	0	1	1	45	75	110	1764,02
SL15-4550B	VNIIMK line	s77	128	14,2	0	1	1	43	74	109	1744,09
SL15-4544B	VNIIMK line	s146	112,6	15,4	0	1	1	42	75	110	1764,02
SL15-4544A	VNIIMK line	s489	159,2	22,6	0	0	1	42	75	110	1764,02
SL15-4514B	VNIIMK line	s78	164,6	25	0	1	1	42	71	106	1682,4
SL15-4528B	VNIIMK line	s147	126,6	18,4	0	1	1	43	72	107	1703,58
SL15-4542B	VNIIMK line	s586	122,4	10,2	0	1	1	45	78	113	1813,46
SL15-4536B	VNIIMK line	s79	171,6	22,8	0	1	1	44	73	108	1723,91
SL15-4798B	VNIIMK line	s148	148,2	22	0	1	1	46	77	112	1800,18
VK838A	VNIIMK line	s490	119,4	18,2	0	0	1	43	74	109	1744,09
SL09-4750A	VNIIMK line	s587	126	16,8	0	0	1	40	65	100	1554,67
SL13-2186A	VNIIMK line	s588	128,6	20,4	0	0	1	41	67	102	1599,53
SL13-2182A	VNIIMK line	s589	115,4	22,8	0	0	1	39	64	99	1532,54
SL06-2545A	VNIIMK line	s149	152,6	27,2	0	0	1	41	67	102	1599,53
SL05-4154A	VNIIMK line	s590	146,4	19,6	0	0	1	42	70	105	1661,72
SL13-2190A	VNIIMK line	s80	132,2	15,2	0	0	1	42	71	106	1682,4
SL05-16A	VNIIMK line	s81	146	13,6	0	0	1	41	70	105	1661,72
SL13-2756A	VNIIMK line	s82	148,2	19,2	0	0	1	43	76	111	1785,15
SL13-2280A	VNIIMK line	s492	139,2	19,2	0	0	1	39	65	100	1554,67
SL13-2224A	VNIIMK line	s83	124,6	17,8	0	0	1	42	70	105	1661,72
SL13-2226A	VNIIMK line	s593	107,8	13,8	0	0	1	39	62	97	1487,03
SL13-2218A	VNIIMK line	s84	112,6	16,6	0	0	1	42	70	105	1661,72
SL13-2206A	VNIIMK line	s594	104,4	12,6	0	0	1	39	66	101	1576,7
SL13-2220A	VNIIMK line	s493	102	20,4	0	0	1	43	70	105	1661,72
SL15-4500B	VNIIMK line	s595	154	16,2	0	1	1	44	74	109	1744,09
SL13-2270B	VNIIMK line	s150	114,8	24,6	0	1	1	41	67	102	1599,53
SL16-2278B	VNIIMK line	s596	109,2	13,4	0	1	1	42	73	108	1723,91
SL16-2286B	VNIIMK line	s85	128,2	20	0	1	1	44	76	111	1785,15
SL16-2282B	VNIIMK line	s597	108,8	15	0	1	1	43	74	109	1744,09
SL16-2284B	VNIIMK line	s598	128	18	0	1	1	44	74	109	1744,09
SL16-2292B	VNIIMK line	s87	158,8	21,2	0	1	1	44	76	111	1785,15
SL05-4770A	VNIIMK line	s151	129,2	23,2	0	0	1	40	67	102	1599,53
VK276B	VNIIMK line	s599	72,2	15,6	0	1	1	38	62	97	1487,03

SL01-3828A	VNIIMK line	s90	107,8	20	0	0	1	41	67	102	1599,53
SL13-2232B	VNIIMK line	s153	133,4	18	0	1	1	38	61	96	1465,05
SL15-4512B	VNIIMK line	s495	177,4	30	0	1	1	43	75	110	1764,02
SL16-2288B	VNIIMK line	s91	143,8	16,8	0	1	1	42	72	107	1703,58
VK678A	VNIIMK line	s496	130,2	12,2	0	0	1	41	67	102	1599,53
VK639A	VNIIMK line	s93	105,8	11,6	0	0	1	41	66	101	1576,7
VK464A	VNIIMK line	s497	87	11,2	0	0	1	46	75	110	1764,02
SL13-2256A	VNIIMK line	s498	119,2	19	0	0	1	41	65	100	1554,67
VK861A	VNIIMK line	s499	146	17,8	0	0	1	47	77	112	1800,18
SL13-2270A	VNIIMK line	s600	114,8	24,6	0	0	1	42	67	102	1599,53
SL13-2268A	VNIIMK line	s500	89,2	18,6	0	0	1	41	65	100	1554,67
SL13-2284A	VNIIMK line	s601	131	23,2	0	0	1	43	74	109	1744,09
SL13-2288A	VNIIMK line	s501	115,6	18	0	0	1	42	66	101	1576,7
SL14-4638B	VNIIMK line	s502	137	21,4	0	1	1	43	70	105	1661,72
SL13-2228A	VNIIMK line	s94	145,4	23,8	0	0	1	42	67	102	1599,53
VK653A	VNIIMK line	s503	106	27	0	0	1	40	66	101	1576,7
SL15-2294A	VNIIMK line	s95	157,8	21,2	0	0	1	43	70	105	1661,72
SL13-2246A	VNIIMK line	s504	116,2	22	0	0	1	41	65	100	1554,67

VNIIMK (Pustovoit All-Russia Research Institute of Oil Crops), Russia
VIR (N. I. Vavilov Research Institute of Plant Industry), Russia
USDA (United States Department of Agriculture), USA
NS (Institute of Field and Vegetable Crops Novi Sad), Serbia

H- Plant height, cm
D - Diameter of the flower head, cm
Br - Stem branching (1 - presence of lateral branches; 0 - abscence of lateral branches)
Fertility - male fertility of disk florets (1- Presence of pollen; 0 - absence of pollen)
Ray florets (1- Presence; 0 - absence)
DTB - the number of days from the planting date to the development of primordial flower head
DTF - the number of days from planting date to 50% of plants in blossom
DTM - the number of days from the planting date to 50% of flower heads becoming yellow
HU - heat units accumulations between the planting date and physiological maturity date

Table S2

Biochemical and morphological seed traits of 99 sunflower lines

line	Fatty acid content											Seed oil content, %	Seed husk content, %	100-seed weight, g	Seed number in the head
	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1	C22:0	C24:0				
K 223	0,06	4,56	0,13	2,03	44,75	47,11	0,13	0,11	0,17	0,79	0,17	23,4	47,1	5,9	414
K 225	0,10	6,19	0,20	4,05	45,11	40,81	0,67	0,40	0,48	1,72	0,26	35,6	37,5	5,8	524
K 370	0,08	6,10	0,09	5,40	37,77	48,33	0,11	0,41	0,19	1,24	0,27	42,1	29,2	5,6	546
K 562	0,08	6,28	0,13	3,91	45,75	41,89	0,08	0,31	0,15	1,15	0,27	41,1	31,6	2,2	970
K 581	0,10	6,99	0,16	3,07	28,00	59,72	0,11	0,28	0,19	1,07	0,30	44,3	27,6	4,2	689
K 651-3	0,09	7,01	0,11	4,62	38,96	46,67	0,11	0,40	0,19	1,50	0,33	44,8	29,9	3,6	899
K 912	0,05	7,06	0,18	7,31	42,03	40,62	0,22	0,55	0,20	1,48	0,29	42,3	26,7	9,4	884
K 1391	0,07	5,94	0,07	4,32	26,48	61,26	0,16	0,29	0,19	1,02	0,19	35,9	37,3	4,5	446
K 1459	0,08	7,11	0,20	3,98	28,29	57,15	0,59	0,40	0,47	1,40	0,32	34,5	38,8	4,9	895
K 1505	0,05	5,59	0,07	5,55	33,74	52,44	0,13	0,45	0,17	1,46	0,35	37	32,3	3,1	517
K 1594	0,07	7,67	0,11	4,09	29,03	56,85	0,12	0,36	0,15	1,19	0,35	31,6	32	6	1012
K 1687	0,06	6,29	0,11	4,37	33,04	54,33	0,14	0,30	0,16	0,92	0,28	38,9	29	7,4	834
K 2068	0,05	5,24	0,04	4,07	34,36	54,01	0,11	0,28	0,19	1,38	0,27	43,2	29,8	4,6	412
K 2086	0,07	7,54	0,16	4,63	41,76	43,23	0,17	0,43	0,22	1,45	0,34	36,2	35,3	5,1	966
K 2125	0,04	5,87	0,18	3,35	28,43	60,29	0,15	0,28	0,17	0,93	0,29	36,5	34,7	6,6	829
K 2235	0,08	6,85	0,13	2,78	46,13	42,20	0,14	0,23	0,21	0,97	0,26	35,4	31,6	7,4	976
K 2238	0,05	6,44	0,07	5,13	38,20	47,74	0,09	0,41	0,18	1,39	0,28	37,6	30,1	6	416
K 2257	0,05	7,38	0,08	7,55	42,13	39,53	0,11	0,60	0,10	2,09	0,37	40,5	22,6	4,3	898
K 2462	0,06	8,73	0,31	2,83	25,37	61,14	0,07	0,23	0,14	0,91	0,20	47,1	23	4,5	1252
K 2479	0,04	4,79	0,09	3,50	43,65	45,72	0,14	0,32	0,19	1,22	0,34	24,2	43,7	4,8	340
K 3035	0,04	6,88	0,14	5,35	32,78	52,12	0,12	0,53	0,21	1,44	0,40	42,7	27,4	4	1380
K 3059	0,06	6,37	0,29	3,55	46,81	39,40	0,89	0,33	0,57	1,39	0,34	34,9	35,5	4,1	1488
K 3159	0,10	8,70	0,26	2,78	19,08	67,17	0,15	0,27	0,18	1,10	0,21	42,7	28,9	4,6	653
K 3350	0,08	8,19	0,18	4,44	27,16	57,23	0,33	0,37	0,19	1,51	0,32	38,3	25,4	6,1	343
K 3376	0,03	5,07	0,09	3,52	47,98	41,12	0,09	0,32	0,21	1,32	0,25	43,4	29,4	4,1	1427
SL 1721	0,05	5,37	0,06	6,51	51,77	33,40	0,11	0,48	0,15	1,69	0,39	48,5	19,4	4,2	1796
SL 1790	0,05	6,63	0,09	3,77	36,89	50,30	0,13	0,35	0,22	1,36	0,20	42,8	26,9	4,8	1314
SL 1813	0,07	7,41	0,10	4,01	31,01	55,50	0,12	0,32	0,15	1,00	0,30	47,4	15,9	8,5	579
SL 2039	0,07	6,72	0,11	5,24	31,08	54,69	0,23	0,35	0,17	0,99	0,34	44,3	30,3	2,9	1059
SL 2950	0,05	6,75	0,14	3,88	37,38	49,59	0,20	0,28	0,20	1,21	0,31	27,7	43,6	5,1	443
VA 1	0,07	6,72	0,11	4,85	43,46	42,55	0,11	0,40	0,15	1,26	0,32	48,1	19,5	4,9	1510
VA 4	0,05	6,75	0,14	3,88	37,38	49,59	0,20	0,28	0,20	1,21	0,31	44,8	19,9	3,7	1651
VIR 130-1	0,05	5,54	0,11	5,48	54,29	31,91	0,21	0,49	0,25	1,34	0,33	33,3	38,4	6,7	902
VIR 172	0,05	5,57	0,07	4,66	45,42	42,02	0,08	0,35	0,16	1,32	0,29	43	29,6	2,9	955
VIR 369	0,07	7,64	0,23	2,40	33,12	54,75	0,07	0,22	0,19	1,05	0,24	48,6	21,4	4,3	1720
VIR 391	0,03	5,99	0,09	6,43	42,70	42,58	0,06	0,46	0,11	1,24	0,31	50,2	22,2	2,8	1071
VK 15	0,06	6,69	0,14	2,56	29,75	58,87	0,10	0,23	0,18	1,14	0,26	43,7	23,5	3,9	1626
VK 30	0,07	7,65	0,15	5,37	32,87	52,17	0,08	0,36	0,11	0,97	0,19	33,9	28,1	7,1	654
VK 102	0,04	6,06	0,11	3,04	45,18	43,87	0,10	0,23	0,19	0,94	0,24	32,8	31,4	7,9	415
VK 206	0,07	7,12	0,11	5,67	27,41	57,65	0,14	0,39	0,11	1,06	0,25	43,2	22	6	1121
VK 268	0,05	5,71	0,09	6,99	40,69	42,98	0,12	0,49	0,14	2,39	0,34	42	20,2	4,1	1392
VK 416	0,04	6,30	0,10	5,61	50,67	35,16	0,05	0,39	0,13	1,26	0,28	44,9	25,2	8,5	721
VK 428	0,07	6,25	0,10	5,99	29,78	55,81	0,08	0,39	0,14	1,14	0,24	47,2	21,3	7,4	362
VK 464	0,04	3,98	0,24	1,62	88,70	3,63	0,07	0,20	0,41	0,76	0,34	35,6	33,7	5,7	639
VK 474	0,06	8,07	0,17	4,99	32,11	52,55	0,10	0,34	0,14	1,16	0,29	38,6	25,9	4,9	1085

VK 475	0,05	5,51	0,13	2,90	46,83	42,59	0,05	0,27	0,19	1,19	0,28	32,3	36,8	4,9	1055
VK 519	0,07	7,30	0,08	5,32	31,43	53,19	0,12	0,41	0,15	1,60	0,32	42,3	24,2	5,7	610
LG 3	0,08	7,18	0,12	3,59	32,49	54,52	0,13	0,27	0,16	1,22	0,22	32,7	40,7	3,2	716
LG 8-2	0,08	6,36	0,11	3,14	35,70	51,78	0,14	0,28	0,19	1,84	0,38	30,2	38,6	4,7	646
LG 10	0,05	6,00	0,09	4,47	25,70	61,61	0,14	0,30	0,16	1,23	0,25	41,8	20,2	3,6	210
LG 26	0,04	5,20	0,12	4,80	83,67	3,13	0,10	0,48	0,22	1,77	0,46	35,5	33,6	5,3	1018
LG 27	0,03	4,79	0,06	5,07	62,93	24,44	0,09	0,38	0,19	1,71	0,31	28,3	47,1	9	796
LG 28	0,06	8,62	0,26	1,84	35,51	51,85	0,10	0,20	0,19	1,01	0,34	34,2	41,3	5,1	1257
KG 7	0,05	5,95	0,10	4,85	35,31	51,69	0,08	0,31	0,13	1,23	0,29	36,5	30,6	5,8	501
KG 16	0,06	6,92	0,11	3,39	43,42	44,39	0,10	0,27	0,18	0,87	0,28	30,7	41,2	8,7	863
KG 19	0,06	6,88	0,08	4,23	16,74	70,21	0,14	0,30	0,13	0,99	0,23	36,4	24,2	5,3	243
KG 21	0,05	6,04	0,14	3,60	48,63	39,58	0,12	0,33	0,16	1,04	0,29	30,3	41,2	5,5	992
KG 32	0,05	6,60	0,09	6,57	31,51	52,60	0,14	0,49	0,11	1,51	0,32	50,4	24,2	2,5	858
KG 48	0,05	7,16	0,12	5,74	34,48	50,15	0,11	0,50	0,13	1,27	0,29	32,8	38,1	4,4	917
KG 104	0,06	6,56	0,11	6,29	45,53	39,23	0,12	0,52	0,12	1,12	0,34	30,1	38	6,8	515
L 1392	0,08	6,83	0,11	6,92	31,88	51,55	0,12	0,49	0,10	1,66	0,25	39,4	23,7	3,3	1046
L 2090	0,07	8,70	0,38	2,73	27,06	59,42	0,11	0,20	0,11	0,96	0,25	36,5	32,6	6,4	1230
L 2138	0,05	6,51	0,11	5,80	38,78	45,95	0,11	0,47	0,14	1,75	0,34	34,2	32,1	8	313
L 2532	0,05	6,63	0,09	4,01	34,51	51,86	0,12	0,32	0,15	1,98	0,27	43,2	30,1	5,7	1437
L 2543	0,05	6,10	0,11	3,62	32,39	54,97	0,15	0,23	0,19	2,00	0,20	44,3	28,3	6	210
L 2544	0,05	5,48	0,07	6,65	49,38	36,17	0,07	0,49	0,12	1,26	0,26	50,9	23,7	5,6	541
L 2563	0,06	6,55	0,13	4,97	47,01	39,31	0,06	0,37	0,12	1,11	0,30	46,2	26,9	3,1	929
L 2582	0,04	5,97	0,09	5,11	37,43	48,81	0,15	0,40	0,14	1,52	0,34	46,8	27,3	3,1	633
L 2586	0,07	6,71	0,09	4,90	28,87	57,36	0,12	0,35	0,13	1,04	0,34	46,9	22,6	2,5	935
L 2595	0,05	7,05	0,11	5,35	33,13	52,20	0,09	0,38	0,16	1,24	0,23	35,5	32,7	6,5	525
L 3376	0,03	5,96	0,17	4,39	40,10	47,42	0,09	0,38	0,16	1,11	0,18	41,7	27,2	3,9	1618
L 7247	0,06	6,01	0,10	4,70	44,69	42,37	0,07	0,38	0,14	1,10	0,37	34,5	32	6,9	1088
M 1046	0,06	6,69	0,14	2,56	29,75	58,87	0,10	0,23	0,18	1,14	0,26	50,6	23	3,9	450
I7-235	0,05	5,77	0,08	4,00	42,83	43,66	0,06	0,36	0,16	2,63	0,39	49,6	17,6	5,6	1154
I7-246	0,04	4,91	0,05	5,40	54,12	32,03	0,07	0,43	0,16	2,46	0,33	38,3	27	6	576
№1416-5	0,04	6,23	0,06	4,29	34,95	51,58	0,07	0,36	0,22	1,88	0,33	44,1	22,1	3	662
№355114	0,05	7,19	0,13	3,88	33,62	51,44	0,19	0,38	0,19	2,60	0,34	30	36,9	4	399
MVG-3	0,05	5,54	0,11	5,48	54,29	31,91	0,21	0,49	0,25	1,34	0,33	41,2	27	5,9	826
MVG-8	0,02	4,56	0,07	5,86	58,05	28,66	0,09	0,40	0,19	1,82	0,27	50,4	25,2	5,7	1706
B 2073	0,04	6,06	0,11	3,04	45,18	43,87	0,10	0,23	0,19	0,94	0,24	39	30,3	6,3	791
№ 424924	0,07	7,21	0,16	7,26	26,59	56,64	0,12	0,47	0,10	1,10	0,25	43,9	28,9	2,9	644
№ 577083	0,09	8,09	0,10	3,54	18,79	67,42	0,12	0,31	0,16	1,09	0,29	43,4	25,9	4,1	290
№ 577432	0,07	6,26	0,11	3,01	28,73	60,08	0,10	0,24	0,15	1,01	0,22	25,6	43,9	12,5	225
№ 577433	0,08	7,26	0,16	3,28	28,66	58,46	0,12	0,26	0,16	1,22	0,33	36,9	34,2	6,3	470
AH 512 Rf	0,05	5,57	0,07	4,66	45,42	42,02	0,08	0,35	0,16	1,32	0,29	46,3	24,2	5,1	649
AH 70029 Rf	0,07	7,12	0,11	5,67	27,41	57,65	0,14	0,39	0,11	1,06	0,25	38,2	33,1	3,9	339
HA 89	0,06	6,55	0,13	4,97	47,01	39,31	0,06	0,37	0,12	1,11	0,30	48,7	22,5	4,5	985
RHA 265-1	0,05	5,48	0,07	6,65	49,38	36,17	0,07	0,49	0,12	1,26	0,26	45,4	23,1	4,5	1108
RHA 274-1	0,07	7,41	0,10	4,01	31,01	55,50	0,12	0,32	0,15	1,00	0,30	46,2	29,2	2,8	846
RHA 297	0,05	6,63	0,09	3,77	36,89	50,30	0,13	0,35	0,22	1,36	0,20	43,8	30,9	3,3	950
RHA 298	0,07	6,72	0,11	5,24	31,08	54,69	0,23	0,35	0,17	0,99	0,34	44,7	23,8	3,7	473
Sl 2966	0,07	7,65	0,15	5,37	32,87	52,17	0,08	0,36	0,11	0,97	0,19	35	35,3	5,5	769
Z 1064	0,09	9,14	0,07	6,29	47,99	32,68	0,09	0,60	0,17	2,48	0,40	34,6	20,9	3,6	517
Z 231	0,03	5,99	0,09	6,43	42,70	42,58	0,06	0,46	0,11	1,24	0,31	47,8	17,7	7,9	853
ZB x 231 AC	0,08	6,28	0,13	3,91	45,75	41,89	0,08	0,31	0,15	1,15	0,27	45,2	17,4	9,2	807
Chernyanka 66-2	0,06	5,74	0,08	6,34	43,56	40,73	0,09	0,48	0,15	2,38	0,38	44,1	24,7	7,2	700
JS-17	0,04	5,59	0,09	4,04	42,03	45,04	0,11	0,33	0,23	2,19	0,31	39,9	27	8,8	1343

I4BC4 ANN 2188	0,04	6,26	0,08	5,05	26,21	59,98	0,12	0,37	0,15	1,44	0,29	41	24,7	5,2	764
I4BC4PET 2203	0,05	6,23	0,09	3,85	32,16	54,48	0,07	0,33	0,20	2,17	0,37	37,4	32,9	5,4	1110

Table S3

List of the sunflower lines selected for the study.

Id_Number	Line Name	Line Phenotype	Sequencing Number
1	A32/08 23-1121	sterility maintainer	s410
2	A32/08 308-111	sterility maintainer	s1
3	A40/08 36-1121	sterility maintainer	s97
4	A40/08 150-1111	sterility maintainer	s98
5	A40/08 124(155-21111)	sterility maintainer	s2
6	A40/08 155-22111	sterility maintainer	s506
7	A40/08 163-1111	sterility maintainer	s507
8	A40/08 167-1111	sterility maintainer	s508
9	A40/08 197-1111	sterility maintainer	s99
10	A40/08 205-11111	sterility maintainer	s411
11	A40/08 254-1111	sterility maintainer	s3
12	A40/08 280-1111	sterility maintainer	s100
13	A704/08 149-11111	sterility maintainer	s4
14	A33/08 35-11111	sterility maintainer	s412
15	A33/08 51211	sterility maintainer	s413
16	A33/08 14-1122	sterility maintainer	s509
17	A33/08 35-2111	sterility maintainer	s414
18	A40/08 60-111	sterility maintainer	s5
19	A40/08-150-2211	sterility maintainer	s510
20	A40/08-162-111	sterility maintainer	s511
21	A40/33-82-111	sterility maintainer	s101
22	BC2LJMC AP83 x B32X36//BM08-41-12111	sterility maintainer	s415
23	BC2LJMC AP83 x B32X08//BF704XB08-14-1121	sterility maintainer	s416
24	BC2LJMC AP83 x B32X08//BF704XBM08-21-112	sterility maintainer	s102
25	BC2LJMC AP83 x BS33X08/BSH033XBM08-19-111	sterility maintainer	s6
26	BC2LJMC AP83 x BS33X08/BSH033XBM08-12111	sterility maintainer	s7
27	BC2LJMC AP83 x BS33X08/BSH033XBM08-31-11112	sterility maintainer	s512
28	BC2LJMC AP83 x BM40XBM08/BF704XBM08-50-1212	sterility maintainer	s8
29	BC2LJMC AP83 x BM40XBS33/BM08-15-21121	sterility maintainer	s419
30	BC2LJMC AP83 x BM32XBS33/BM36-7-1111	sterility maintainer	s103
31	BC2LJMC AP83 x BM32XBS33/BM36-72-1111	sterility maintainer	s420
32	BC2LJMC AP83 x BM32XBM08/BM08-14-12-111	sterility maintainer	s514
33	BC2LJMC AP83 x BM32XBM08/BM08-19-11111	sterility maintainer	s104
34	BC2LJMC AP83 x BM32XBM08/BM08-24-111	sterility maintainer	s9
35	BC2LJMC AP83 x BM32XBM08/BM08-29-1111	sterility maintainer	s515
36	BC2LJMC AP83 x BM32XBM08/BM08-33-21111	sterility maintainer	s516
37	BC2LJMC AP83 x BM32XBM08/BM08-83-22121	sterility maintainer	s10
38	BC2LJMC AP83 x B47/704-8	sterility maintainer	s11
39	BC2LJMC AP83 x BHO33/704//47-2	sterility maintainer	s517
40	BC2LJMC AP83 x BM40XBS33/BM08-25-211	sterility maintainer	s12
41	BC2LJMC AP83 x C0SB-5	sterility maintainer	s518
42	AP83X B08/P83-24211113	sterility maintainer	s13
43	AP83X B08/P83-24221113	sterility maintainer	s421
44	AP83X B08/83-31211211	sterility maintainer	s519
45	AP83X B08/P83-33421111	sterility maintainer	s105
46	AP83X BC1MB08/BP83-3322	sterility maintainer	s106
47	AP83X B08/P83-33421111/3322	sterility maintainer	s422
48	AP83X B40/P83-123111	sterility maintainer	s107
49	AP83X BC1-B40//P83-32112	sterility maintainer	s14
50	AP83X BC1BM08//BP83-13211/1522	sterility maintainer	s520

51	AP83X BC1BM40//BP83-321111	sterility maintainer	s108
52	AP83X S5BC1-08//BP83-1821	sterility maintainer	s109
53	AP83X S5BC1-08//BP83-1824	sterility maintainer	s521
54	AP83X BM40///BP83-1	sterility maintainer	s424
55	AP83X BM40///BP83-2	sterility maintainer	s425
56	AP83X BM40///BP83-3	sterility maintainer	s522
57	AP83X BM08///BP83-1	sterility maintainer	s16
58	AP83X B08/P83-24211113	sterility maintainer	s523
59	AP83X BC1-B40//P83-32112	sterility maintainer	s524
60	AP83X BC1BM40//BP83-321111	sterility maintainer	s426
61	RP38	fertility restorer	s17
62	RP38YL	fertility restorer	s110
63	RP1	fertility restorer	s111
64	RP14	fertility restorer	s112
65	RP24	fertility restorer	s427
66	Rlg3	fertility restorer	s18
67	Rlg42	fertility restorer	s525
68	R10	fertility restorer	s19
69	RNM	fertility restorer	s526
70	4120pl	fertility restorer	s527
71	RR144	fertility restorer	s528
72	ΦP81013	fertility restorer	s429
73	R10YL	fertility restorer	s529
74	RS//R10-132226	fertility restorer	s530
75	RS//R10-132151	fertility restorer	s430
76	RS//R10-13/5B	fertility restorer	s531
77	R10/1231	fertility restorer	s431
78	R10CLP-C1	fertility restorer	s432
79	Rdlf	fertility restorer	s433
80	4090pl	fertility restorer	s532
81	Rd1	fertility restorer	s434
82	4093pl	fertility restorer	s435
83	RS15	fertility restorer	s113
84	RS25	fertility restorer	s20
85	4087pl	fertility restorer	s533
86	RS64	fertility restorer	s534
87	RC32(nord)	fertility restorer	s21
88	R6	fertility restorer	s22
89	RK35	fertility restorer	s114
90	4099pl	fertility restorer	s436
91	RR154	fertility restorer	s23
92	4117pl	fertility restorer	s535
93	4099pl	fertility restorer	s437
94	RD 164	fertility restorer	s438
95	RR114	fertility restorer	s536
96	4K738/R6-12111	fertility restorer	s537
97	4K738/R6-4111	fertility restorer	s439
98	RCM//R6-8211	fertility restorer	s538
99	RCM///R6-31212	fertility restorer	s24
100	RCM///R6-3312	fertility restorer	s115
101	RCM//R6-6312	fertility restorer	s440
102	4096	fertility restorer	s441
103	4102	fertility restorer	s442

104	7678-2513	fertility restorer	s116
105	RF 12	fertility restorer	s25
106	Rfob	fertility restorer	s26
107	R483	fertility restorer	s117
108	R60880	fertility restorer	s443
109	R60875	fertility restorer	s539
110	R60875CLP12	fertility restorer	s540
111	RS3	fertility restorer	s27
112	R-fly	fertility restorer	s28
113	RBY	fertility restorer	s119
114	RMO2	fertility restorer	s542
115	RMO2CLP	fertility restorer	s543
116	R-ramb	fertility restorer	s444
117	RK325	fertility restorer	s445
118	RK806	fertility restorer	s446
119	RK-alz	fertility restorer	s447
120	Ralz	fertility restorer	s544
121	RK-BLR	fertility restorer	s545
122	RL-mgs	fertility restorer	s29
123	RL 65/35	fertility restorer	s30
124	R-eL	fertility restorer	s31
125	R-exp	fertility restorer	s546
126	RT085	fertility restorer	s120
127	Rcrb	fertility restorer	s448
128	Rs65HO	fertility restorer	s449
129	9758R	fertility restorer	s450
130	9802R	fertility restorer	s547
131	RC61	fertility restorer	s451
132	RC61CLP41	fertility restorer	s32
133	RC8505	fertility restorer	s548
134	R67	fertility restorer	s33

Table S4

List of SNPs, significantly associated with the ability to suppress CMS phenotype after FDR correction.

Marker	LG	Position	*p*-Value
S01_128131042	1	128131042	2.864e-06
S08_126438	8	126438	1.846e-06
S08_53903259	8	53903259	1.2031e-06
S08_53903274	8	53903274	1.2031e-06
S08_123143887	8	123143887	7.3934e-08
S08_123143931	8	123143931	7.3934e-08
S10_8652678	10	8652678	2.2032e-09
S10_12216788	10	12216788	8.1699e-09
S10_12216789	10	12216789	8.1699e-09
S10_12216836	10	12216836	8.1699e-09
S10_13077995	10	13077995	6.9084e-16
S10_15015366	10	15015366	5.0982e-12
S10_15088058	10	15088058	2.2755e-13
S10_15088077	10	15088077	2.2755e-13
S10_15088083	10	15088083	2.2755e-13
S10_16502367	10	16502367	4.6206e-11
S10_16502383	10	16502383	4.6206e-11
S10_16502388	10	16502388	4.6206e-11
S10_16672099	10	16672099	4.3528e-15
S10_16672116	10	16672116	4.3528e-15
S10_16973978	10	16973978	8.9505e-12
S10_16978362	10	16978362	1.0841e-11
S10_16978398	10	16978398	1.0841e-11
S10_19215429	10	19215429	2.2929e-13
S10_19215436	10	19215436	2.2929e-13
S10_19215466	10	19215466	2.2929e-13
S10_20157601	10	20157601	1.4407e-14
S10_20731828	10	20731828	4.7903e-11
S10_20731859	10	20731859	4.7903e-11
S10_20950352	10	20950352	2.3488e-14
S10_22224869	10	22224869	4.8859e-14
S10_22868526	10	22868526	7.3097e-14
S10_22868559	10	22868559	7.3097e-14
S10_22888160	10	22888160	7.801e-15
S10_22888174	10	22888174	7.801e-15
S10_23593820	10	23593820	3.1714e-13
S10_23598288	10	23598288	1.7562e-15
S10_23979584	10	23979584	2.8835e-12
S10_23979613	10	23979613	2.8835e-12
S10_24103410	10	24103410	2.1836e-14
S10_24302532	10	24302532	2.2396e-09
S10_24631211	10	24631211	9.9572e-14
S10_24840720	10	24840720	2.2812e-10
S10_24844700	10	24844700	1.1134e-10
S10_24846679	10	24846679	8.1095e-09
S10_24846782	10	24846782	1.5141e-10
S10_25008166	10	25008166	2.2082e-09
S10_25009339	10	25009339	2.0194e-10
S10_25516085	10	25516085	1.1561e-09

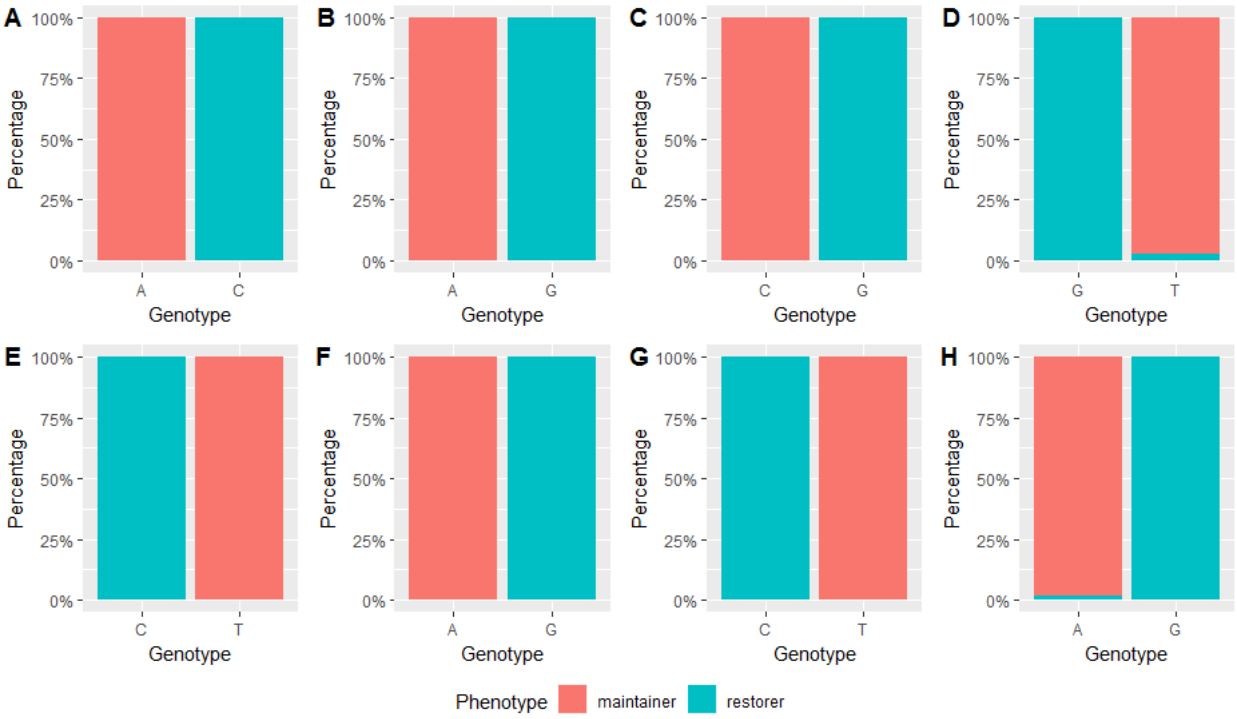
S10_25637813	10	25637813	2.5369e-10
S10_25639672	10	25639672	4.3426e-09
S10_26007668	10	26007668	5.4509e-12
S10_26122059	10	26122059	1.8139e-10
S10_26122060	10	26122060	1.8139e-10
S10_26261410	10	26261410	5.4698e-09
S10_26261413	10	26261413	5.4698e-09
S10_26261415	10	26261415	5.4698e-09
S10_26341283	10	26341283	8.3653e-10
S10_26373751	10	26373751	2.3673e-08
S10_26373770	10	26373770	2.3673e-08
S10_26818798	10	26818798	1.8809e-08
S10_26818815	10	26818815	1.8809e-08
S10_27005278	10	27005278	1.292e-10
S10_27012344	10	27012344	4.7488e-08
S10_27229641	10	27229641	1.3837e-10
S10_27229674	10	27229674	1.3837e-10
S10_27890499	10	27890499	3.6033e-09
S10_28082825	10	28082825	2.0364e-11
S10_28122177	10	28122177	9.6728e-12
S10_28336345	10	28336345	8.1738e-10
S10_28673860	10	28673860	1.305e-10
S10_28676538	10	28676538	2.9382e-10
S10_28751961	10	28751961	1.1888e-10
S10_28776947	10	28776947	2.3031e-11
S10_28776959	10	28776959	2.3031e-11
S10_28847299	10	28847299	7.7943e-10
S10_28847341	10	28847341	7.7943e-10
S10_29111412	10	29111412	6.7611e-11
S10_29616306	10	29616306	1.4473e-11
S10_32748646	10	32748646	9.3929e-09
S10_32917627	10	32917627	6.8689e-08
S10_32917642	10	32917642	6.8689e-08
S10_32922499	10	32922499	2.501e-07
S10_33104167	10	33104167	6.023e-08
S10_33109711	10	33109711	4.8895e-09
S10_33109725	10	33109725	8.1579e-11
S10_33109753	10	33109753	8.1579e-11
S10_33308922	10	33308922	3.5127e-09
S10_33789180	10	33789180	9.3829e-11
S10_35995267	10	35995267	3.814e-11
S10_35995268	10	35995268	3.814e-11
S10_36011406	10	36011406	4.7488e-11
S10_36109241	10	36109241	3.5516e-10
S10_36352875	10	36352875	6.5328e-10
S10_36493785	10	36493785	2.821e-13
S10_36493806	10	36493806	2.821e-13
S10_36682072	10	36682072	7.7958e-09
S10_37247856	10	37247856	9.8783e-10
S10_37247861	10	37247861	9.8783e-10
S10_37383929	10	37383929	9.3619e-14
S10_37383948	10	37383948	9.3619e-14
S10_37498670	10	37498670	2.6355e-10

S10_37842783	10	37842783	9.4891e-08
S10_38946665	10	38946665	8.2154e-10
S10_39033268	10	39033268	7.3084e-13
S10_39085019	10	39085019	1.8645e-12
S10_39085023	10	39085023	1.8645e-12
S10_39107249	10	39107249	1.066e-13
S10_39107267	10	39107267	1.066e-13
S10_39107275	10	39107275	1.066e-13
S10_40412799	10	40412799	3.6877e-14
S10_40412814	10	40412814	3.6877e-14
S10_40412836	10	40412836	3.6877e-14
S10_40854657	10	40854657	7.8152e-13
S10_40903825	10	40903825	3.516e-11
S10_41233859	10	41233859	1.2617e-13
S10_41233860	10	41233860	1.2617e-13
S10_41233867	10	41233867	1.2617e-13
S10_41262509	10	41262509	5.2572e-13
S10_41262539	10	41262539	5.2572e-13
S10_42386052	10	42386052	1.7433e-13
S10_42386104	10	42386104	1.7433e-13
S10_42395477	10	42395477	3.2433e-11
S10_42395516	10	42395516	3.2433e-11
S10_42506323	10	42506323	2.8167e-08
S10_42747724	10	42747724	1.0227e-08
S10_42783546	10	42783546	1.011e-07
S10_42783552	10	42783552	1.011e-07
S10_42798133	10	42798133	3.9265e-11
S10_43057361	10	43057361	9.3347e-11
S10_43392391	10	43392391	3.7861e-12
S10_43392398	10	43392398	3.7861e-12
S10_43593827	10	43593827	2.1151e-08
S10_45130435	10	45130435	2.3715e-10
S10_45169853	10	45169853	2.0718e-06
S10_45169874	10	45169874	2.0718e-06
S10_45169878	10	45169878	2.0718e-06
S10_46923521	10	46923521	1.9201e-09
S10_47145380	10	47145380	5.7909e-08
S10_47208285	10	47208285	1.4596e-07
S10_47208318	10	47208318	1.4596e-07
S10_47210455	10	47210455	1.1419e-07
S10_47280020	10	47280020	1.1606e-07
S10_48290966	10	48290966	8.4319e-09
S10_48291018	10	48291018	5.5934e-06
S10_48499825	10	48499825	2.6469e-06
S10_48504916	10	48504916	2.7204e-06
S10_48591005	10	48591005	7.3941e-07
S10_48591011	10	48591011	7.3941e-07
S10_62196502	10	62196502	4.0324e-09
S10_62553042	10	62553042	7.8219e-07
S10_101507311	10	101507311	1.8396e-05
S10_101507326	10	101507326	1.8396e-05
S10_101507335	10	101507335	1.8396e-05
S10_102520064	10	102520064	7.9563e-06

S10_102602662	10	102602662	1.7819e-08
S10_131614165	10	131614165	5.8924e-07
S10_131614189	10	131614189	5.8924e-07
S10_143378328	10	143378328	9.1335e-11
S10_150217716	10	150217716	5.2526e-05
S10_160313151	10	160313151	9.2075e-10
S10_160313155	10	160313155	9.2075e-10
S10_160316260	10	160316260	7.595e-06
S10_161195132	10	161195132	2.2124e-08
S10_161195163	10	161195163	2.2124e-08
S10_161207939	10	161207939	2.7968e-06
S10_161207951	10	161207951	2.7968e-06
S10_164607188	10	164607188	2.7492e-10
S10_166145792	10	166145792	3.277e-09
S10_167325919	10	167325919	1.2115e-08
S10_188891671	10	188891671	1.0917e-12
S10_195183448	10	195183448	1.7412e-10
S10_196674403	10	196674403	1.4375e-11
S10_196674438	10	196674438	1.4375e-11
S10_196674440	10	196674440	1.4375e-11
S10_198599962	10	198599962	1.032e-07
S10_198599968	10	198599968	1.032e-07
S10_198600004	10	198600004	1.032e-07
S10_225410637	10	225410637	2.1966e-05
S13_170494693	13	170494693	1.0112e-15
S13_171053833	13	171053833	1.5286e-18
S13_173268042	13	173268042	3.4582e-18
S13_173832391	13	173832391	5.6903e-09
S13_174474103	13	174474103	1.2222e-14
S13_174474122	13	174474122	1.2222e-14
S13_174809087	13	174809087	1.1002e-13
S13_178217103	13	178217103	2.0314e-14
S13_188440145	13	188440145	9.6827e-07
S13_188440166	13	188440166	9.6827e-07
S17_6085001	17	6085001	4.2694e-12

Figure S1

Distribution of the ability to suppress CMS phenotype across sunflower samples with different allelic states for 8 statistically significant markers: A)S13_170494693, B)S13_171053833, C)S13_173268042, D)S13_173832391, E)S13_174474103, F)S13_174474122, G)S13_174809087, H)S13_178217103



Annex for chapter 5

Table S1

List of the samples

Species	winter/spring	line	MS_ID
Brassica *napus*	spring	4005/16	R63
Brassica *napus*	winter	578/16	R109
Brassica *napus*	spring	3680/16	R43
Brassica *napus*	winter	824/16	R28
Brassica *napus*	winter	763/16	R1
Brassica *napus*	winter	675/16	R36
Brassica *napus*	spring	3295/16	R18
Brassica *napus*	spring	3298/16	R38
Brassica *napus*	spring	3530/16	R71
Brassica *napus*	spring	3316/16	R74
Brassica *napus*	winter	761/16	R35
Brassica *napus*	spring	3356/16	R90
Brassica *napus*	winter	1575/16	R11
Brassica *napus*	spring	3959/16	R56
Brassica *napus*	spring	3224/16	R29
Brassica *napus*	winter	529/16	R2
Brassica *napus*	spring	3980/16	R49
Brassica *napus*	winter	784/16	R81
Brassica *napus*	winter	1688/16	R13
Brassica *napus*	winter	734/16	R91
Brassica *napus*	spring	3285/16	R73
Brassica *napus*	winter	1540/16	R65
Brassica *napus*	spring	3957/16	R5
Brassica *napus*	spring	3300/16	R25
Brassica *napus*	winter	755/16	R54
Brassica *napus*	winter	1433/16	R62
Brassica *napus*	spring	3479/16	R93
Brassica *napus*	spring	3494/16	R70
Brassica *napus*	winter	842/16	R57
Brassica *napus*	spring	3674/16	R61
Brassica *napus*	spring	3527/16	R97
Brassica *napus*	winter	1607/16	R44
Brassica *napus*	winter	754/16	R22
Brassica *napus*	winter	549/16	R47
Brassica *napus*	winter	1548/16	R21
Brassica *napus*	winter	1594/16	R53
Brassica *napus*	winter	554/16	R52
Brassica *napus*	spring	4008/16	R55
Brassica *napus*	spring	3987/16	R76
Brassica *napus*	spring	3700/16	R33
Brassica *napus*	winter	641/16	R98
Brassica *napus*	winter	749/16	R108
Brassica *napus*	spring	3312/16	R75
Brassica *napus*	winter	798/16	R23

Brassica *napus*	spring	3353/16	R84
Brassica *napus*	winter	559/16	R32
Brassica *napus*	winter	706/16	R19
Brassica *napus*	spring	3269/16	R3
Brassica *napus*	winter	686/16	R37
Brassica *napus*	winter	583/16	R4
Helianthus *annuus*	-	M 1046	S2
Helianthus *annuus*	-	VK 30	S6
Helianthus *annuus*	-	VIR 130- 1	S7
Helianthus *annuus*	-	Z 231	S32
Helianthus *annuus*	-	RHA 297	S34
Helianthus *annuus*	-	I7-235	S35
Helianthus *annuus*	-	Z 1064	S63
Helianthus *annuus*	-	K 223	S89
Helianthus *annuus*	-	K 2086	S103
Helianthus *annuus*	-	VIR 391	S105
Helianthus *annuus*	-	GS 17	S134
Helianthus *annuus*	-	K 1459	S141
Helianthus *annuus*	-	№ 424924	S151
Helianthus *annuus*	-	RHA 298	S155
Helianthus *annuus*	-	№ 577083	S161
Helianthus *annuus*	-	VK 474	S173
Helianthus *annuus*	-	K 2479	S224
Helianthus *annuus*	-	VK 519	S229
Helianthus *annuus*	-	355114	S250
Helianthus *annuus*	-	RHA 274-1	S253
Helianthus *annuus*	-	I7-246	S254
Helianthus *annuus*	-	VA 1	S262
Helianthus *annuus*	-	K 2257	S274
Helianthus *annuus*	-	VIR 369	S299
Helianthus *annuus*	-	VK 428	S302
Helianthus *annuus*	-	K 2068	S311
Helianthus *annuus*	-	VA 4	S320
Helianthus *annuus*	-	B 2073	S321
Helianthus *annuus*	-	VIR 172	S329
Helianthus *annuus*	-	Sl 2966	S335
Helianthus *annuus*	-	AN 70029 Rf	S341
Helianthus *annuus*	-	VK 268	S342
Helianthus *annuus*	-	RHA 265-1	S390
Helianthus *annuus*	-	VK 464	S401
Helianthus *annuus*	-	K 2235	S417

Helianthus *annuus*	-	№ 577433	S434
Helianthus *annuus*	-	1416-5	S437
Helianthus *annuus*	-	K 1687	S465
Helianthus *annuus*	-	VK 15	S467
Helianthus *annuus*	-	ZB × 231 AC	S494
Helianthus *annuus*	-	VK 475	S497
Helianthus *annuus*	-	K 2125	S499
Helianthus *annuus*	-	K 2238	S520
Helianthus *annuus*	-	K 1594	S529
Helianthus *annuus*	-	VK 102	S541
Helianthus *annuus*	-	№ 577432	S553
Helianthus *annuus*	-	I4BC4ANN 2188	S558
Helianthus *annuus*	-	K 2462	S568
Helianthus *annuus*	-	VK 416	S569
Helianthus *annuus*	-	VK 206	S586

Figure S1

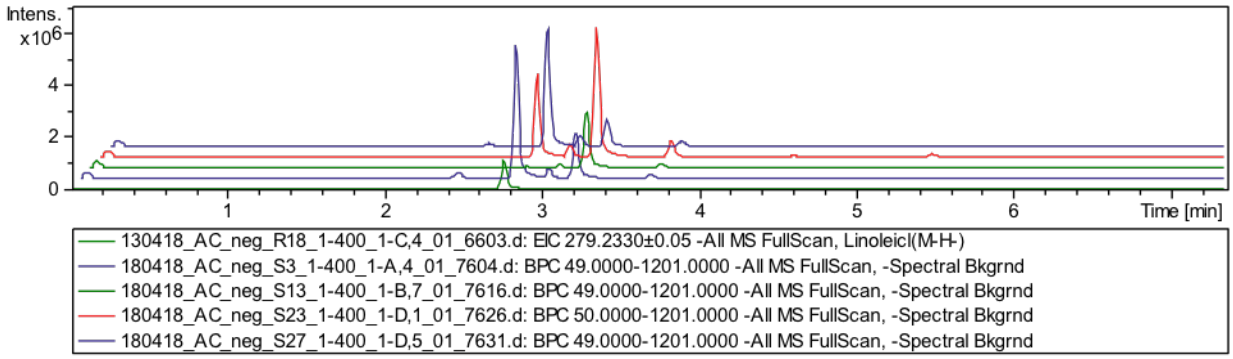

Intens.
x10^6
4
2
0
1
2
3
4
5
6
Time [min]
130418_AC_neg_R18_1-400_1-C,4_01_6603.d: EIC 279.2330±0.05 -All MS FullScan, Linoleicl(M-H-)
180418_AC_neg_S3_1-400_1-A,4_01_7604.d: BPC 49.0000-1201.0000 -All MS FullScan, -Spectral Bkgrnd
180418_AC_neg_S13_1-400_1-B,7_01_7616.d: BPC 49.0000-1201.0000 -All MS FullScan, -Spectral Bkgrnd
180418_AC_neg_S23_1-400_1-D,1_01_7626.d: BPC 50.0000-1201.0000 -All MS FullScan, -Spectral Bkgrnd
180418_AC_neg_S27_1-400_1-D,5_01_7631.d: BPC 49.0000-1201.0000 -All MS FullScan, -Spectral Bkgrnd

Table S2

Species	winter/spring	line	MS_ID
Brassica *napus*	spring	4005/16	R63
Brassica *napus*	winter	578/16	R109
Brassica *napus*	spring	3680/16	R43
Brassica *napus*	winter	824/16	R28
Brassica *napus*	winter	763/16	R1
Brassica *napus*	winter	675/16	R36
Brassica *napus*	spring	3295/16	R18
Brassica *napus*	spring	3298/16	R38
Brassica *napus*	spring	3316/16	R74
Brassica *napus*	winter	761/16	R35
Brassica *napus*	spring	3356/16	R90
Brassica *napus*	spring	3959/16	R56
Brassica *napus*	spring	3224/16	R29
Brassica *napus*	winter	529/16	R2
Brassica *napus*	spring	3980/16	R49
Brassica *napus*	winter	784/16	R81
Brassica *napus*	winter	1688/16	R13
Brassica *napus*	winter	734/16	R91
Brassica *napus*	spring	3285/16	R73
Brassica *napus*	winter	1540/16	R65
Brassica *napus*	spring	3957/16	R5
Brassica *napus*	spring	3300/16	R25
Brassica *napus*	winter	755/16	R54
Brassica *napus*	winter	1433/16	R62
Brassica *napus*	spring	3479/16	R93
Brassica *napus*	spring	3494/16	R70
Brassica *napus*	winter	842/16	R57
Brassica *napus*	spring	3674/16	R61
Brassica *napus*	spring	3527/16	R97
Brassica *napus*	winter	1607/16	R44
Brassica *napus*	winter	754/16	R22
Brassica *napus*	winter	549/16	R47
Brassica *napus*	winter	1548/16	R21
Brassica *napus*	winter	1594/16	R53
Brassica *napus*	winter	554/16	R52
Brassica *napus*	spring	4008/16	R55
Brassica *napus*	spring	3987/16	R76
Brassica *napus*	spring	3700/16	R33
Brassica *napus*	winter	641/16	R98
Brassica *napus*	winter	749/16	R108
Brassica *napus*	spring	3312/16	R75
Brassica *napus*	winter	798/16	R23
Brassica *napus*	spring	3353/16	R84
Brassica *napus*	winter	559/16	R32
Brassica *napus*	winter	706/16	R19

Brassica *napus*	spring	3269/16	R3
Brassica *napus*	winter	686/16	R37
Brassica *napus*	winter	583/16	R4
Helianthus *annuus*	-	M 1046	S2
Helianthus *annuus*	-	VK 30	S6
Helianthus *annuus*	-	VIR 130- 1	S7
Helianthus *annuus*	-	Z 231	S32
Helianthus *annuus*	-	RHA 297	S34
Helianthus *annuus*	-	I7-235	S35
Helianthus *annuus*	-	Z 1064	S63
Helianthus *annuus*	-	K 223	S89
Helianthus *annuus*	-	K 2086	S103
Helianthus *annuus*	-	VIR 391	S105
Helianthus *annuus*	-	GS 17	S134
Helianthus *annuus*	-	K 1459	S141
Helianthus *annuus*	-	№ 424924	S151
Helianthus *annuus*	-	RHA 298	S155
Helianthus *annuus*	-	№ 577083	S161
Helianthus *annuus*	-	VK 474	S173
Helianthus *annuus*	-	K 2479	S224
Helianthus *annuus*	-	VK 519	S229
Helianthus *annuus*	-	355114	S250
Helianthus *annuus*	-	RHA 274-1	S253
Helianthus *annuus*	-	I7-246	S254
Helianthus *annuus*	-	VA 1	S262
Helianthus *annuus*	-	K 2257	S274
Helianthus *annuus*	-	VIR 369	S299
Helianthus *annuus*	-	VK 428	S302
Helianthus *annuus*	-	K 2068	S311
Helianthus *annuus*	-	VA 4	S320
Helianthus *annuus*	-	B 2073	S321
Helianthus *annuus*	-	VIR 172	S329
Helianthus *annuus*	-	SI 2966	S335
Helianthus *annuus*	-	AN 70029 Rf	S341
Helianthus *annuus*	-	VK 268	S342
Helianthus *annuus*	-	RHA 265-1	S390
Helianthus *annuus*	-	VK 464	S401
Helianthus *annuus*	-	K 2235	S417
Helianthus *annuus*	-	№ 577433	S434
Helianthus *annuus*	-	1416-5	S437
Helianthus *annuus*	-	K 1687	S465
Helianthus *annuus*	-	VK 15	S467
Helianthus *annuus*	-	ZB × 231 AC	S494
Helianthus *annuus*	-	VK 475	S497
Helianthus *annuus*	-	K 2125	S499
Helianthus *annuus*	-	K 2238	S520

Helianthus	*annuus*	-	K 1594		S529
Helianthus	*annuus*	-	VK 102		S541
Helianthus	*annuus*	-		577432	S553
Helianthus	*annuus*	-	I4BC4ANN 2188		S558
Helianthus	*annuus*	-	K 2462		S568
Helianthus	*annuus*	-	VK 416		S569
Helianthus	*annuus*	-	VK 206		S586

Annex for chapter 6

Figures

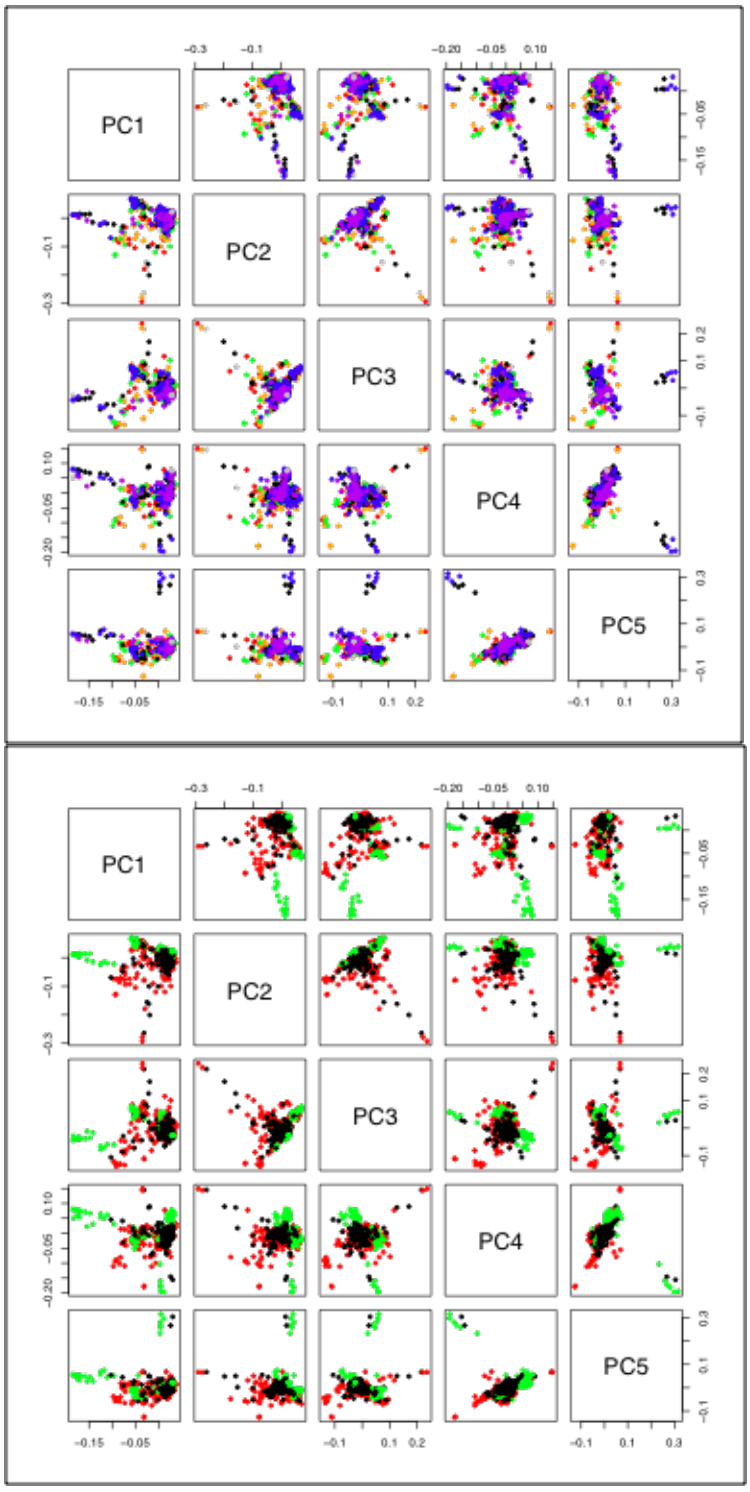


Figure S1: PCA plots reflecting the relationships between sunflower technical samples based on 15068 SNPs segregating in the Russian collection. Each dot corresponds to a sunflower technical sample used in the study. Color of dots corresponds to a) 96-well Plate ID or b) across the 3 different collections.

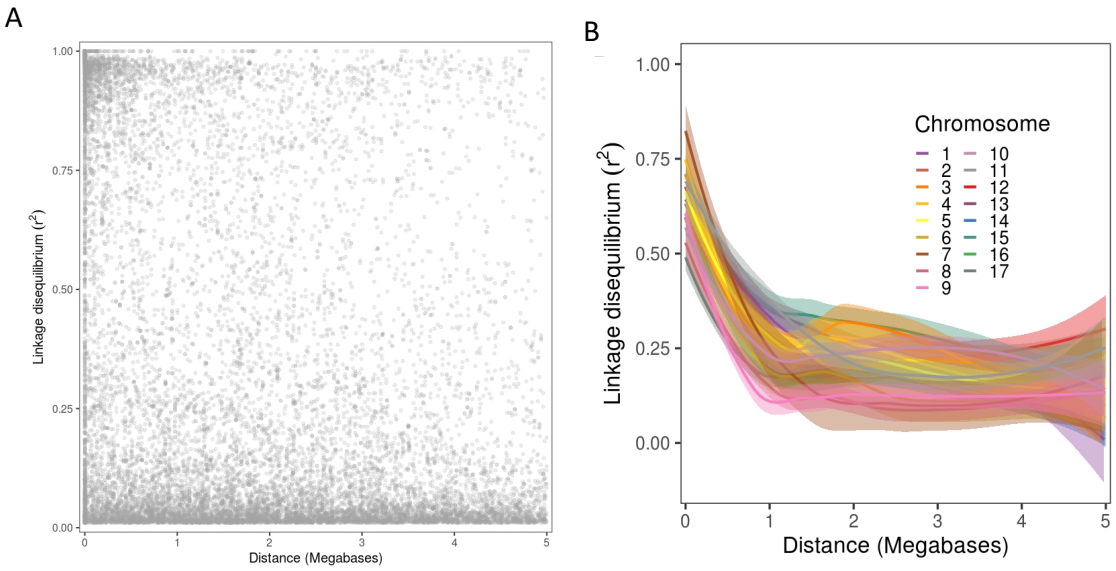


Figure S2: Linkage disequilibrium (LD) decay plot. (A) Genome-wide LD. Gray dots correspond to a SNP pair. (B) LD per each chromosome. Lines correspond to loess curves

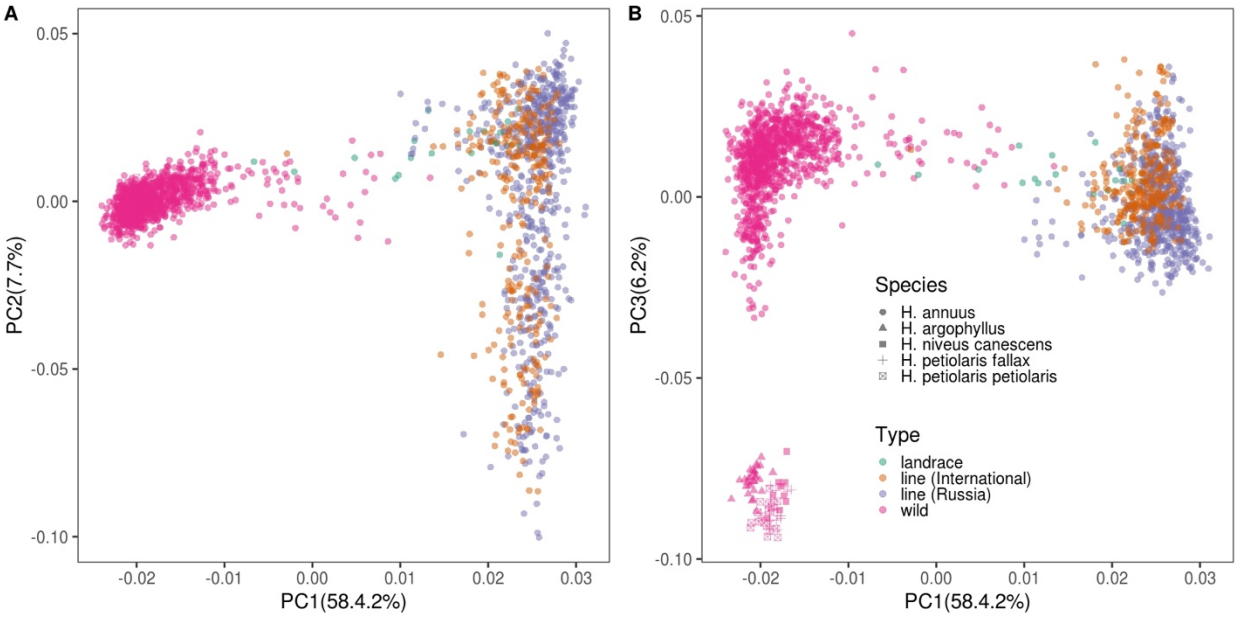


Figure S3: The relationship between sunflower germplasm of different origins estimated based on 2345 SNPs shared between this and the Hübner (2019) studies. (A) The first and the second components of the PCA. (B) The first and the third components of the PCA. Each dot corresponds to a plant accession. Colors indicate the origin (wild/line/landrace). Shapes indicate species.

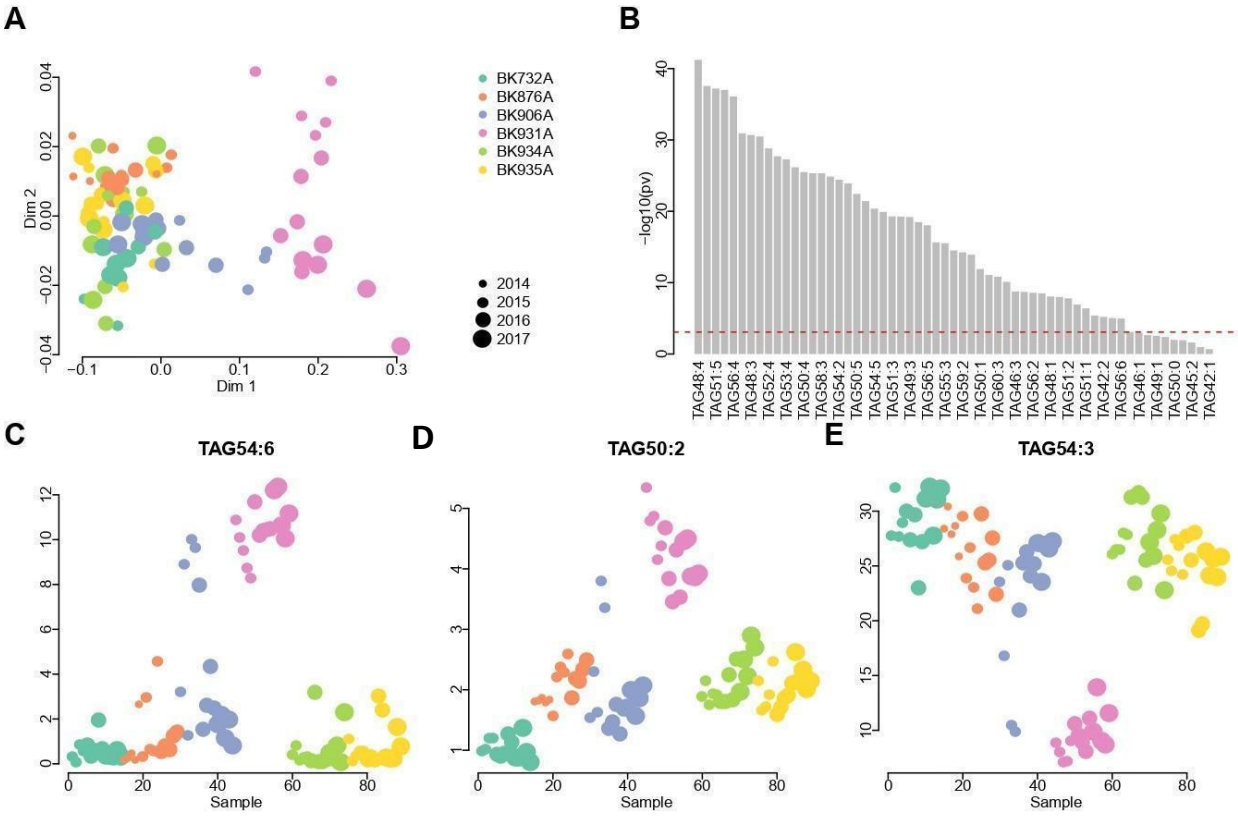


Figure S4: Replication experiment on 6 genotypes:TAG (dilution 1:3) amounts in different sunflower lines. (A) MDS plot based on all the data. (B) Significance level, (C) TAG 54:6; (D) TAG 50:2, (E) TAG 54:3. Each dot represents 1 sample, 1 color assigned to each line. Dot size represents the year of sample collection. In C.D.E, Y axis corresponds to the TAG fraction from the total amount of TAGs.

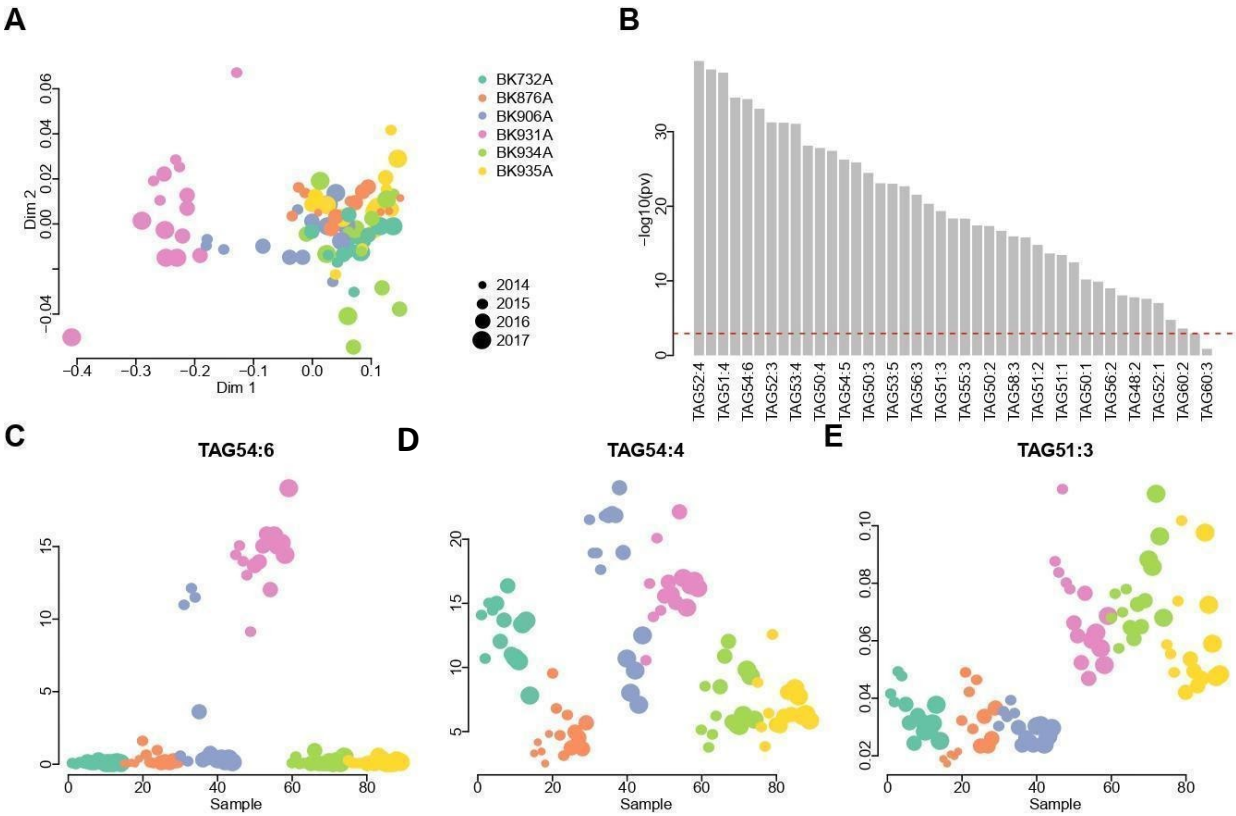


Figure S5: Replication experiment on 6 genotypes: TAG (dilution 1:25) amounts in different sunflower lines. (A) MDS plot based on all the data. (B) Significance level, (C) TAG 54:6; (D) TAG 54:4, (E) TAG 51:3. Each dot represents 1 sample, 1 color assigned to each line. Dot size represents the year of sample collection. In C.D.E Y-axis corresponds to the TAG fraction from the total amount of TAGs.

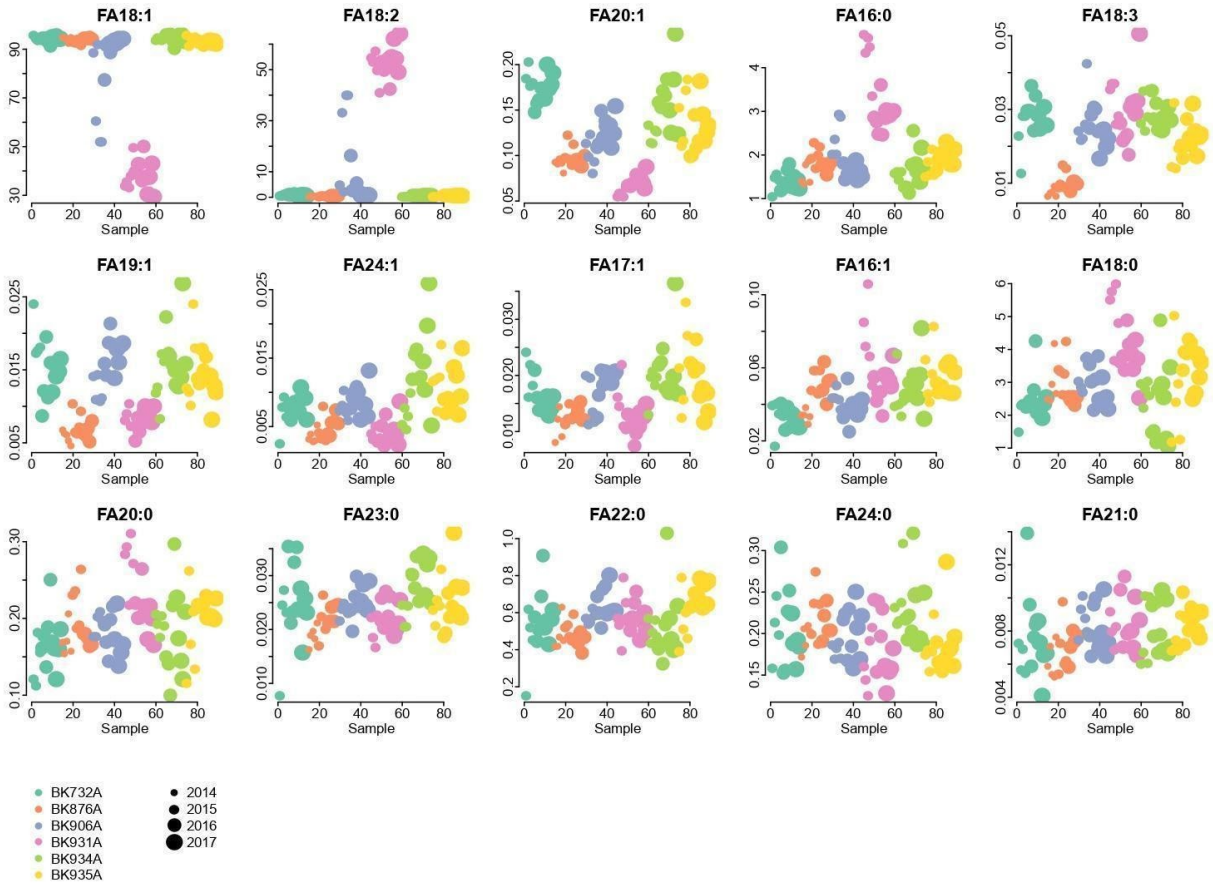
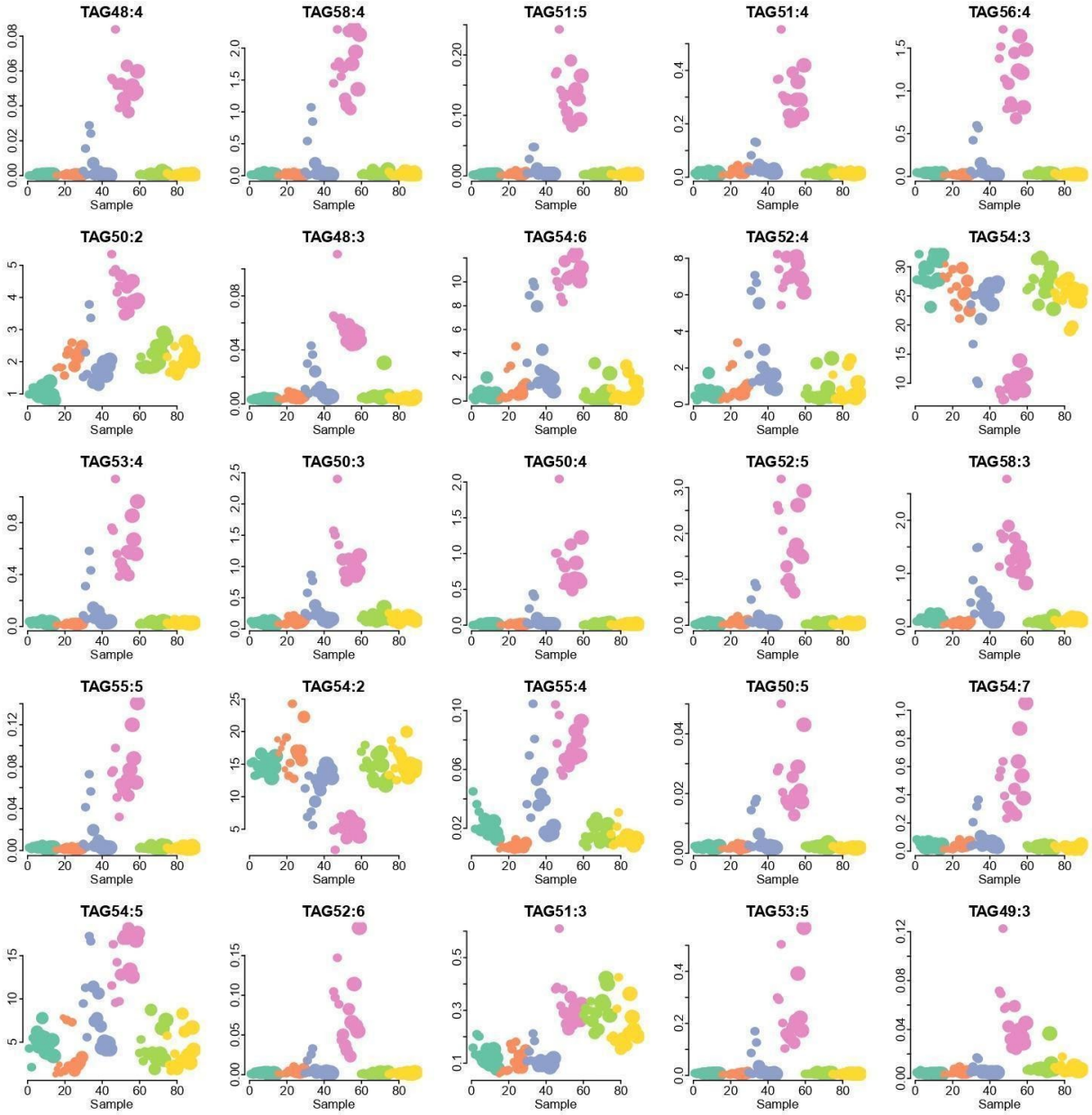
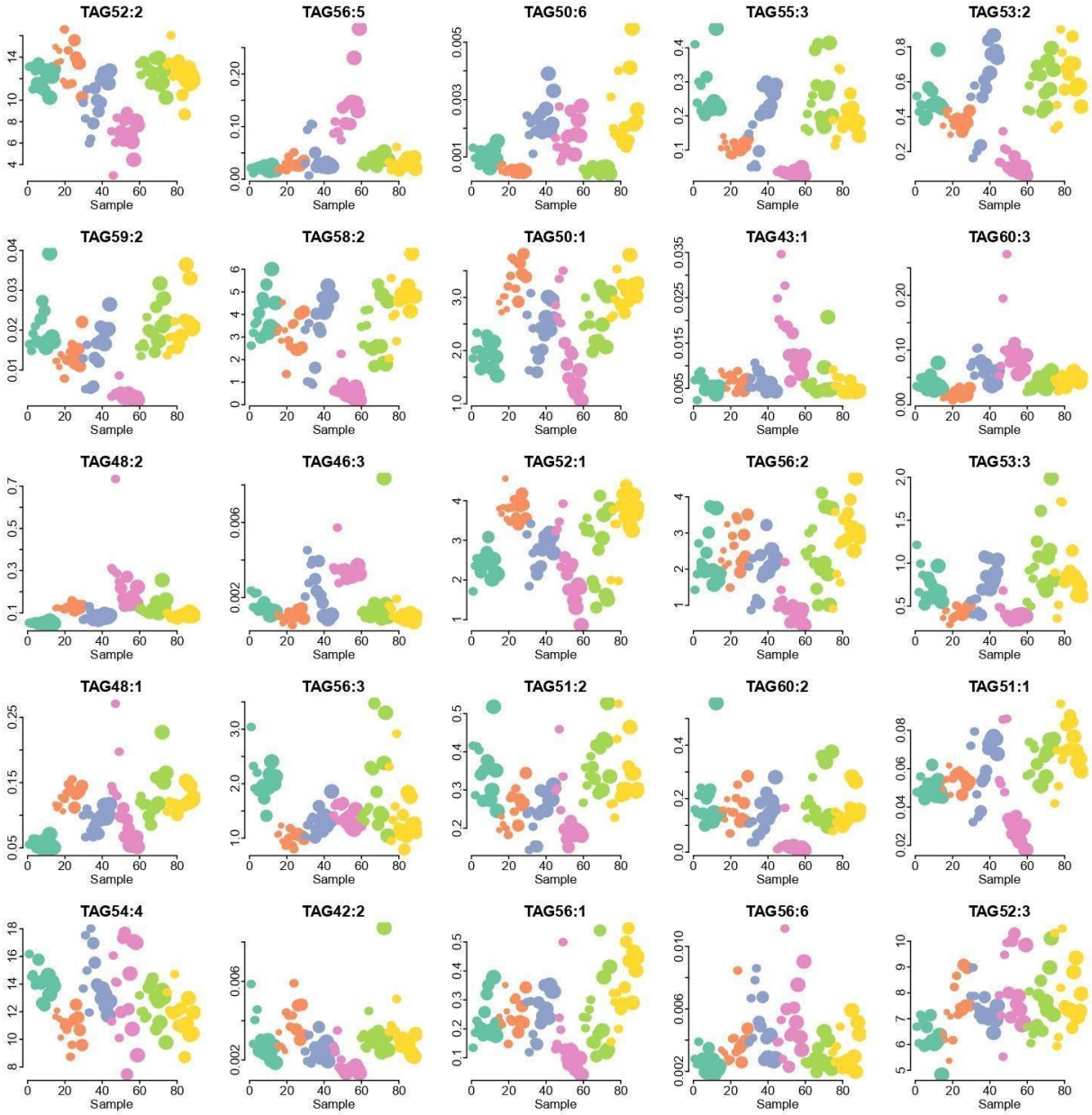


Figure S6: Replication experiment on 6 genotypes: Fractions off all the fatty acids. Each dot represents 1 sample, 1 color assigned to each line. Dot size represents the year of sample collection


TAG48:4
TAG58:4
TAG51:5
TAG51:4
TAG56:4
TAG50:2
TAG48:3
TAG54:6
TAG52:4
TAG54:3
TAG53:4
TAG50:3
TAG50:4
TAG52:5
TAG58:3
TAG55:5
TAG54:2
TAG55:4
TAG50:5
TAG54:7
TAG54:5
TAG52:6
TAG51:3
TAG53:5
TAG49:3
Sample


TAG52:2
TAG56:5
TAG50:6
TAG55:3
TAG53:2
TAG59:2
TAG58:2
TAG50:1
TAG43:1
TAG60:3
TAG48:2
TAG46:3
TAG52:1
TAG56:2
TAG53:3
TAG48:1
TAG56:3
TAG51:2
TAG60:2
TAG51:1
TAG54:4
TAG42:2
TAG56:1
TAG56:6
TAG52:3
Sample

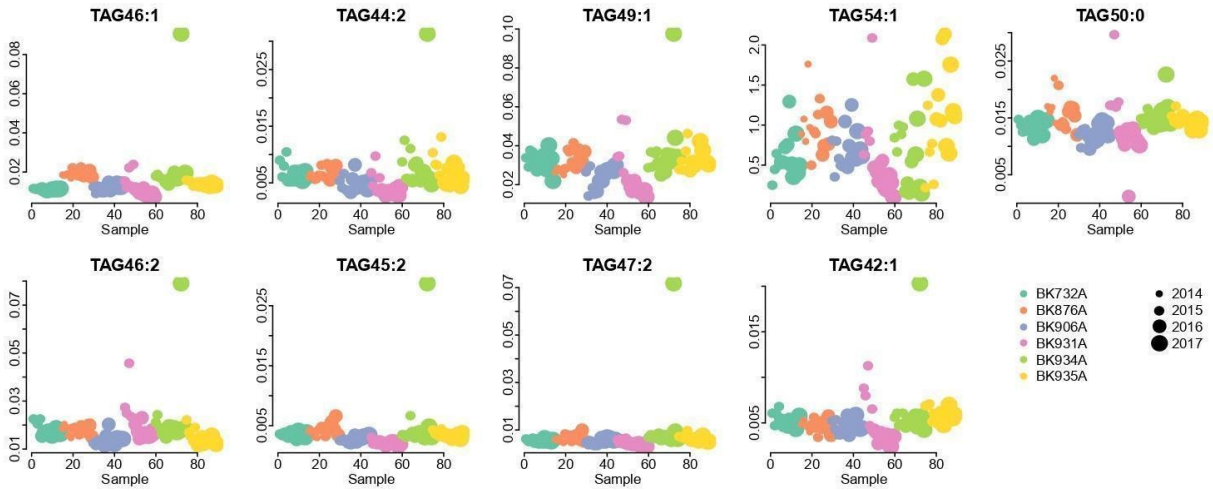
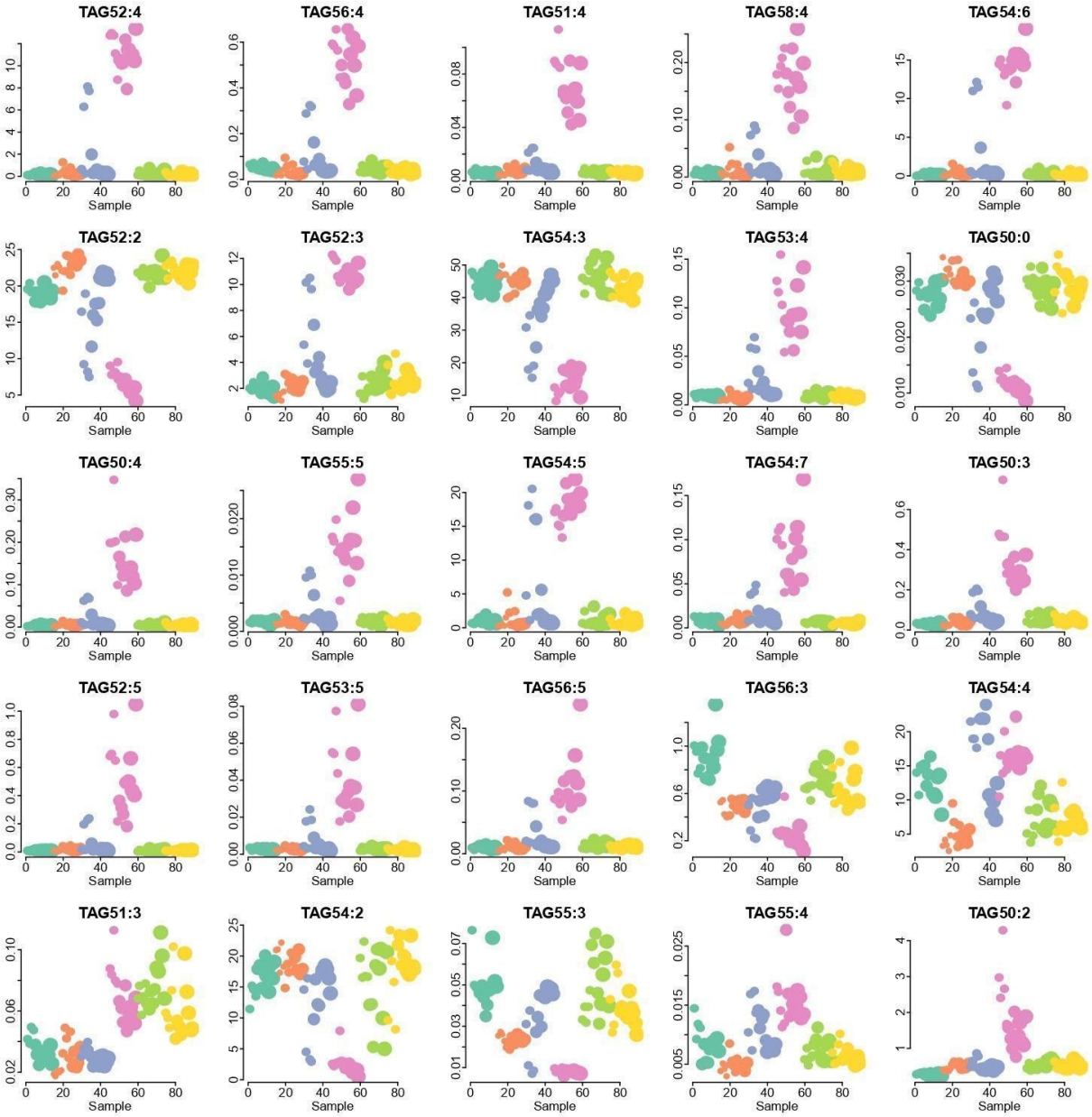


Figure S7: Replication experiment on 6 genotypes: Fractions of all TAGs (1:3). Each dot represents 1 sample, 1 color assigned to each line. Dot size represents the year of sample collection


TAG52:4
TAG56:4
TAG51:4
TAG58:4
TAG54:6
TAG52:2
TAG52:3
TAG54:3
TAG53:4
TAG50:0
TAG50:4
TAG55:5
TAG54:5
TAG54:7
TAG50:3
TAG52:5
TAG53:5
TAG56:5
TAG56:3
TAG54:4
TAG51:3
TAG54:2
TAG55:3
TAG55:4
TAG50:2
Sample

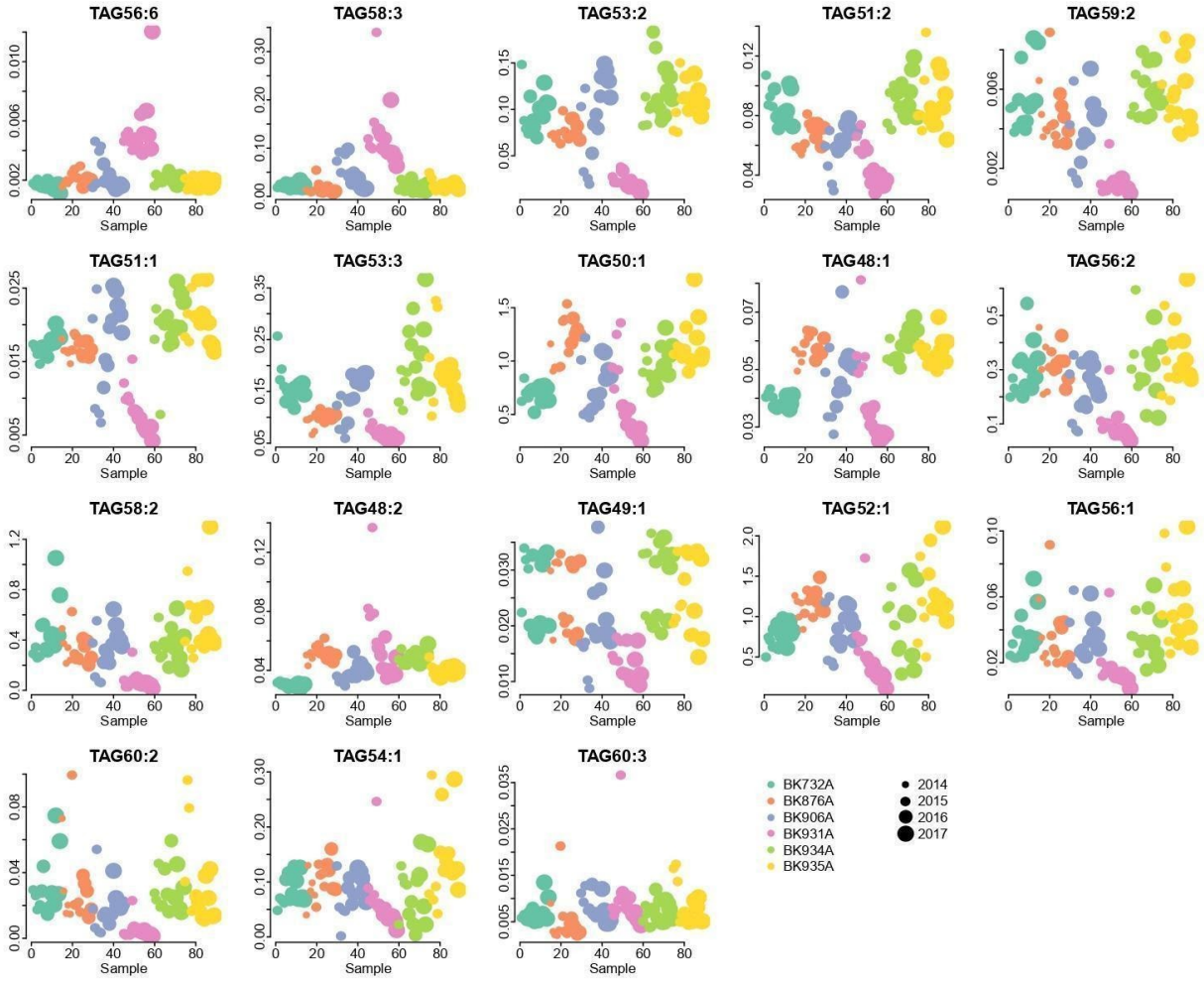


Figure S8: Replication experiment on 6 genotypes: Fractions of all TAGs (1:25). Each dot represents 1 sample, 1 color assigned to each line. Dot size represents the year of sample collection

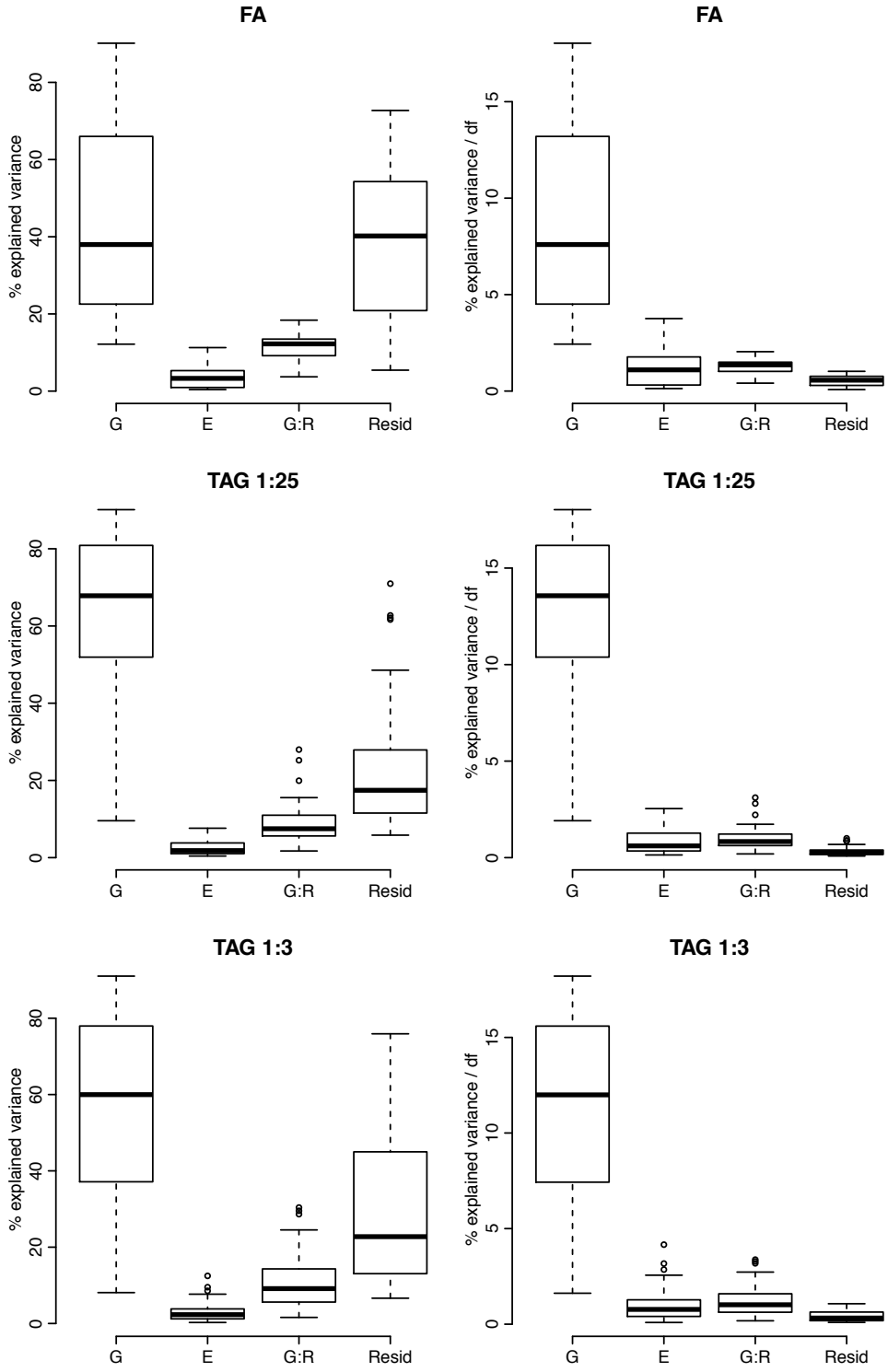
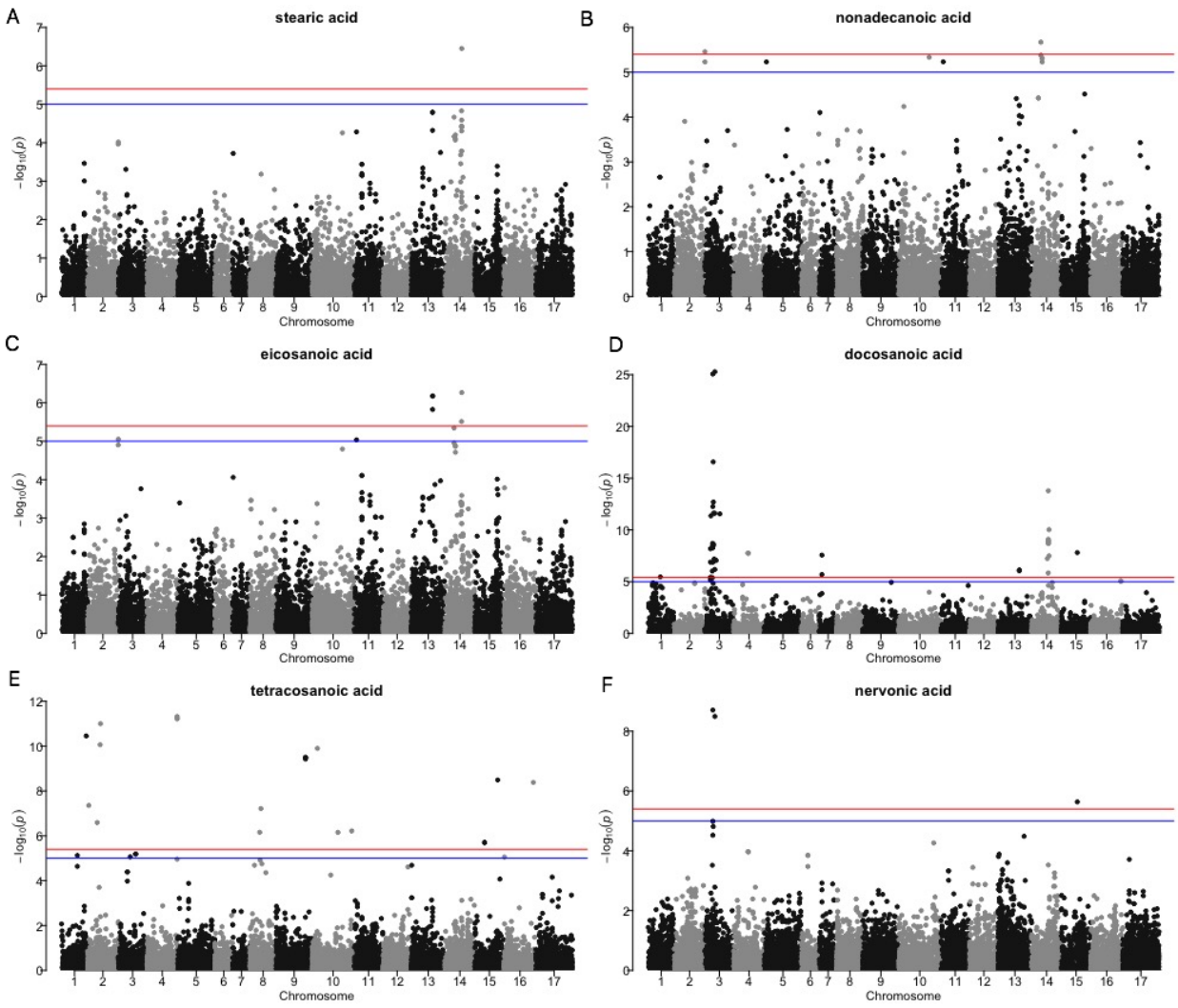


Figure S9: Replication experiment on 6 genotypes: Estimation of effects by genotype and environment on lipidomic profiles. Distributions (across lipids) of percentages of variance explained by each factor for FAs and TAGs in two dilutions are shown on the left, percentages of explained variance divided by number of degrees of freedom are shown on the right.


A
stearic acid
B
nonadecanoic acid
C
eicosanoic acid
D
docosanoic acid
E
tetracosanoic acid
F
nervonic acid
Chromosome

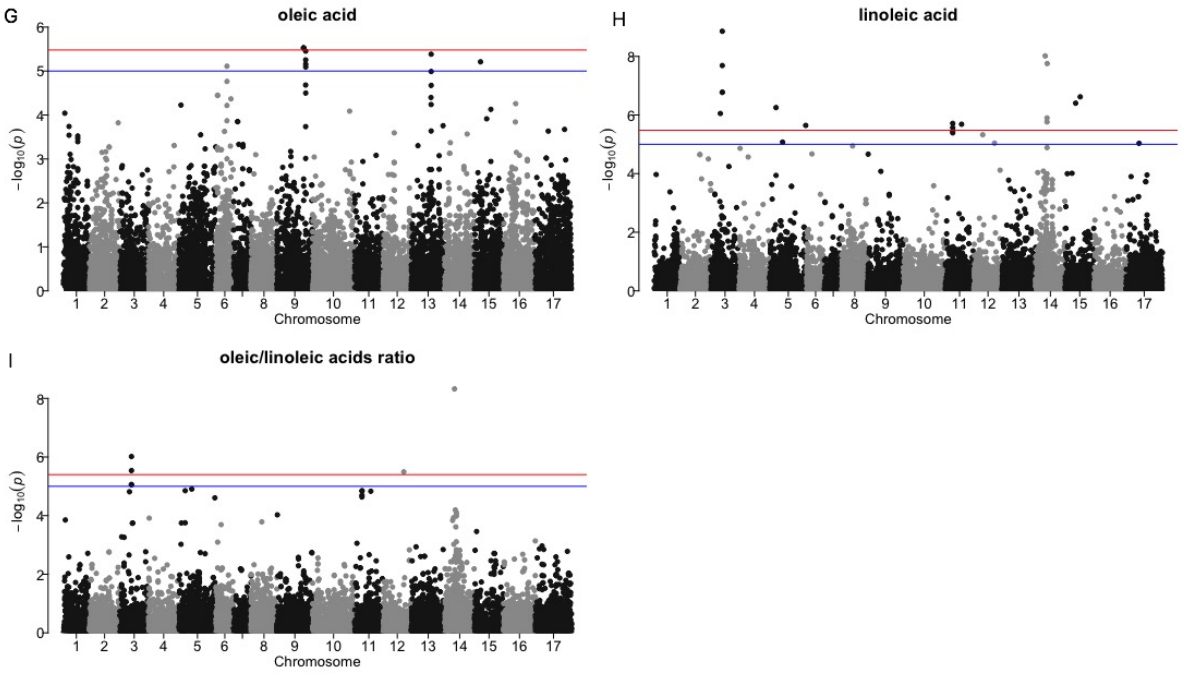
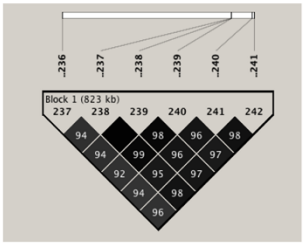
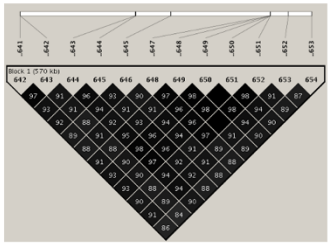


Figure S10: Manhattan plots representing significant associations (a) Stearic acid; (b) Nonadecanoic acid; (c) Eicosenoic acid; (d) Docosanoic acid; (e) Tetracosanoic acid; (f) Nervonic acid; (g) Oleic acid; (h) Linoleic acid; (i) Ratio between oleic and linoleic acids. FDR threshold - blue, Bonferroni threshold -red.

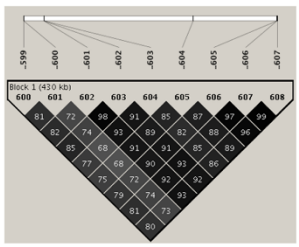
18:1 Chr6



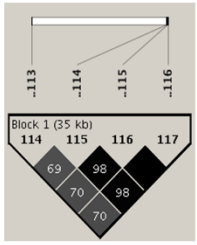
18:1 Chr9



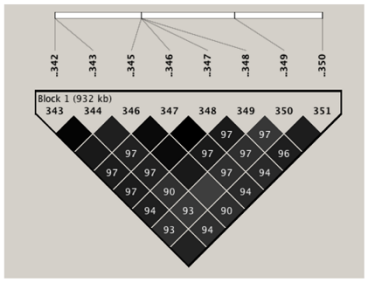
18:1 Chr13



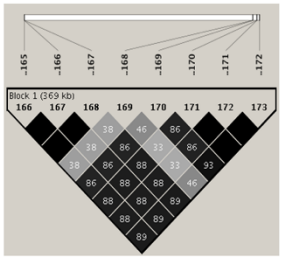
18:1 Chr15



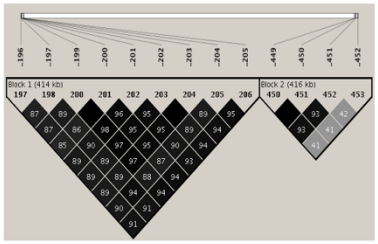
18:2 Chr3



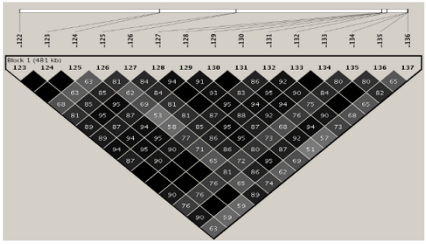
18:2 Chr5



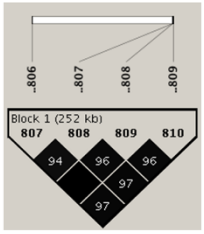
18:2 Chr11



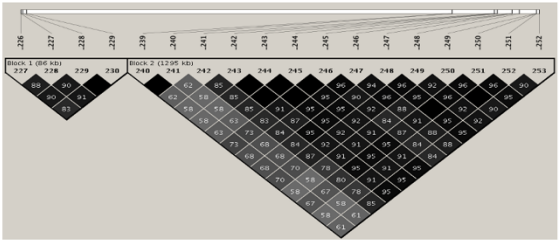
18:3 Chr11



19:0 Chr2



19:0 Chr14



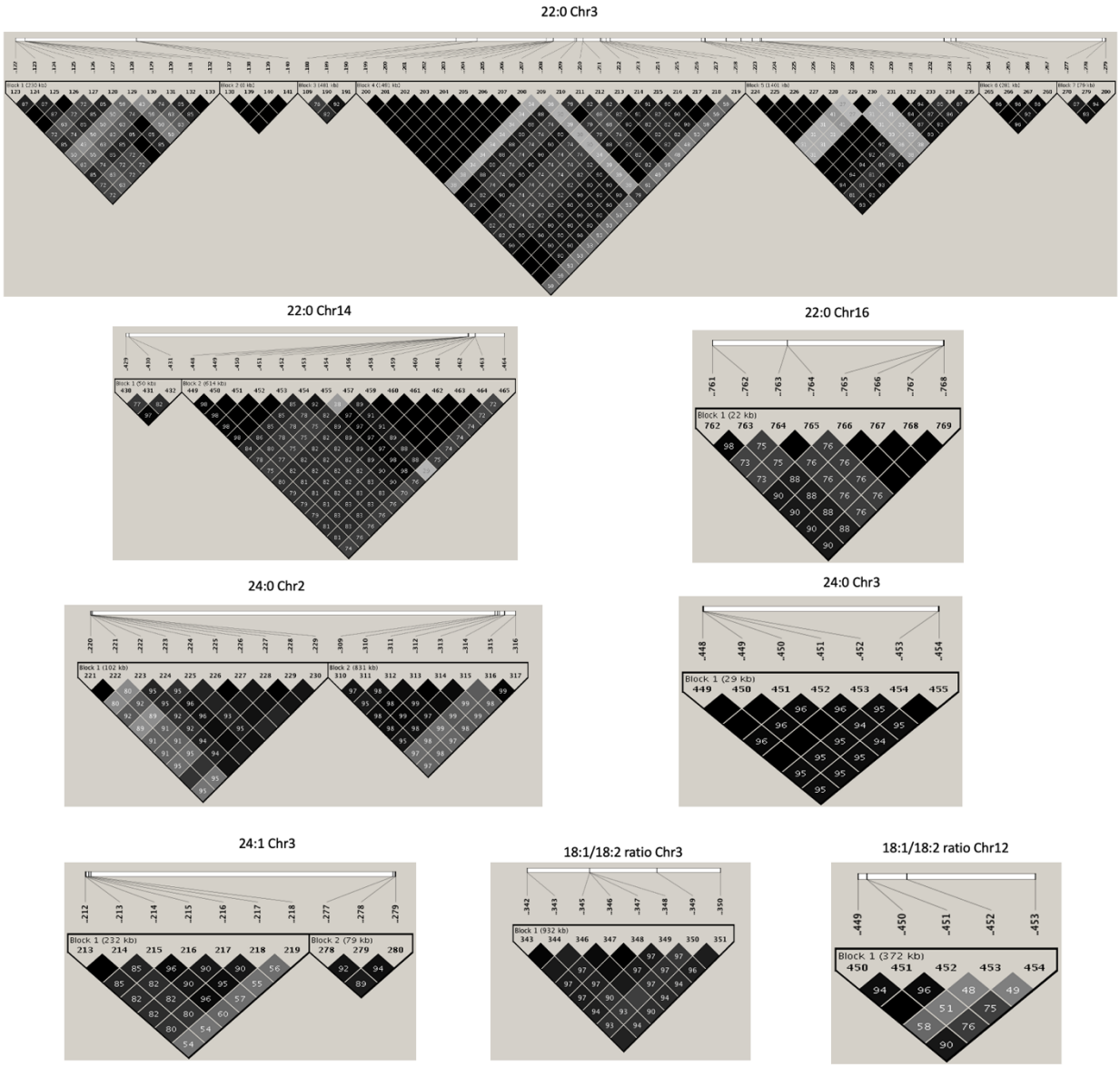


Figure S11: LD blocks containing significant SNPs.

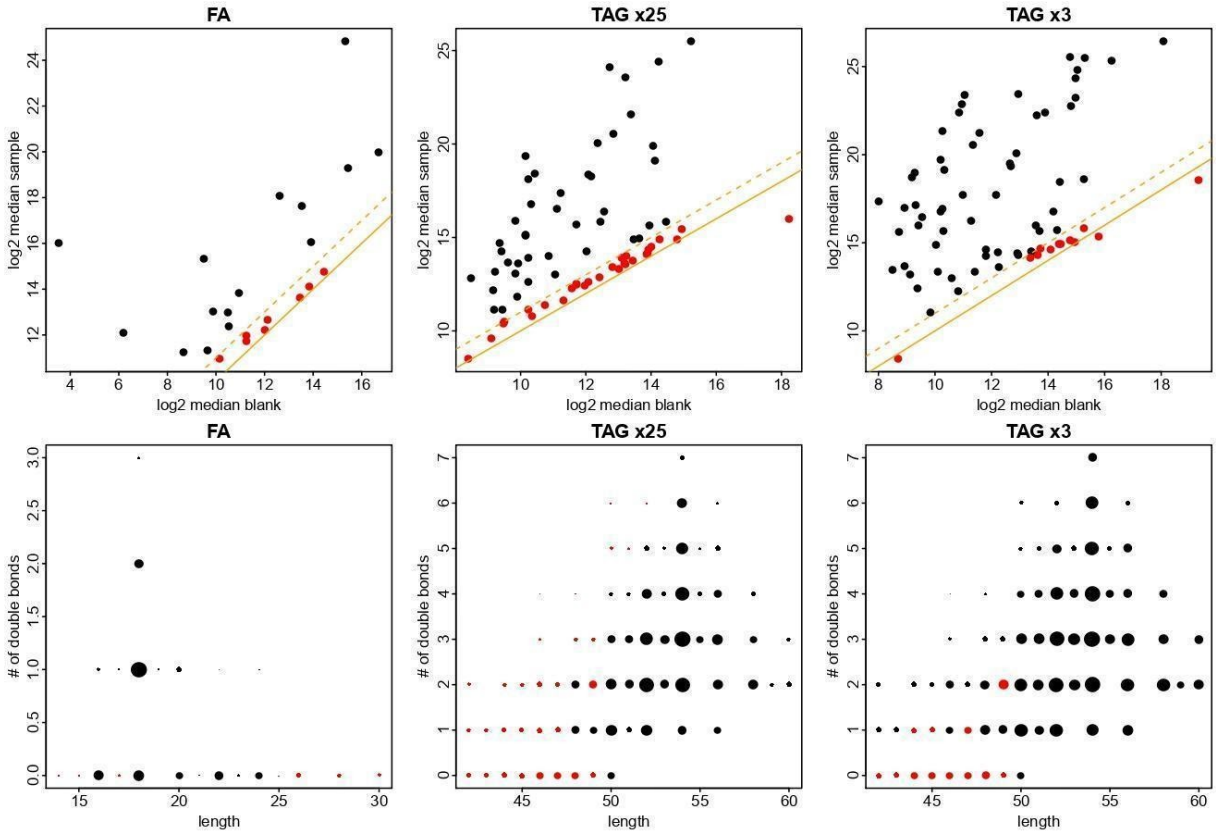


Figure S12: Data clean up using blank samples. Top panels show dependence of average log2 seed sample intensity of FAs and TAGs (two dilutions) on average log2 intensity of same lipids in blank samples (see Methods). Straight and dashed lines correspond to equal intensities in both types of sample and to two-fold higher concentration in seed samples compared to blanks, respectively. Bottom panels show the same lipids as top panels in coordinates of total FA chain length (x-axis), and number of double bounds (y-axis), point size is proportional to log2 average intensity in seed samples. Only lipids with log2(sample/blank) > 1 were used in analysis, remaining samples (shown in red) were filtered out.

Table S1

A

FAs

	line	year	line:year
FA_16:1	6,05E-09	0,000515	0,022557
FA_16:0	1,33E-27	8,2E-06	2,12E-08
FA_17:1	2,6E-08	0,618155	0,011788
FA_18:3	5,32E-18	0,616854	0,001526
FA_18:2	1,28E-42	0,066248	6,39E-06
FA_18:1	9,1E-43	0,029629	1,33E-05
FA_18:0	8,11E-08	0,12785	0,017013
FA_19:1	9,53E-16	0,812654	0,001865
FA_20:1	5,88E-25	0,042594	0,001144
FA_20:0	6,81E-05	0,086886	0,008437
FA_21:0	0,047331	0,106108	0,499518
FA_22:0	0,001976	0,32853	0,342691
FA_23:0	0,000242	0,012992	0,057457
FA_24:1	1,02E-11	0,029435	0,009706
FA_24:0	0,001781	0,366994	0,160552

B

TAGs 1:5

	line	year	line:year
TAG_50:1	8,05E-16	0,000312	1,01E-06
TAG_50:2	6,69E-25	0,000171	5,77E-08
TAG_50:3	3,21E-30	8,42E-05	2,22E-05
TAG_50:4	2,88E-28	0,031545	0,075544
TAG_50:0	4,23E-33	0,381703	4,64E-07
TAG_51:1	4,02E-16	0,245705	0,000614
TAG_51:2	2,16E-17	0,015842	0,00529
TAG_51:3	5,07E-25	2,41E-06	0,000519
TAG_51:4	1,43E-39	0,001468	0,00589
TAG_52:1	1,99E-08	0,885205	0,009418
TAG_52:2	1,18E-40	0,036495	1,61E-09
TAG_52:3	9,98E-39	0,000137	6,54E-08
TAG_52:4	9,82E-42	0,007636	0,000326
TAG_52:5	2,07E-25	0,031626	0,002868
TAG_53:2	2,11E-18	0,01146	0,006279
TAG_53:3	1,23E-14	0,735882	0,001178
TAG_53:4	1,44E-32	0,026345	0,004839
TAG_53:5	2,17E-24	0,072703	0,01504
TAG_54:1	0,000674	0,765934	0,05569
TAG_54:2	8,15E-20	0,261356	0,010622
TAG_54:3	5,01E-38	0,000619	1,43E-07
TAG_54:4	9,04E-29	3,64E-05	1,76E-08
TAG_54:5	7,63E-30	0,184061	2,65E-05
TAG_54:6	7,74E-38	0,173672	8,48E-06
TAG_54:7	6,25E-28	0,05255	0,002448
TAG_55:3	1,21E-21	0,063277	0,000245
TAG_55:4	7,1E-20	0,028964	0,004367
TAG_55:5	1,43E-28	0,373012	0,003603
TAG_56:1	3,03E-05	0,325417	0,686234
TAG_56:2	4,28E-09	0,890328	0,484695
TAG_56:3	5E-24	0,01937	0,003456
TAG_56:4	3,77E-39	0,006433	0,01154
TAG_56:5	1,49E-26	0,454232	1,12E-05
TAG_56:6	5,6E-18	0,487918	0,013054
TAG_58:2	1,04E-08	0,088185	0,511573
TAG_58:3	7,43E-17	0,027964	0,230569
TAG_58:4	1,24E-33	0,035936	0,220313
TAG_59:2	1,49E-14	0,076475	0,173341
TAG_60:2	0,000231	0,438548	0,250274
TAG_60:3	0,10278	0,1013	0,185267
TAG_48:1	4,54E-16	0,01253	1,32E-08
TAG_48:2	1,1E-13	0,00269	1,08E-08
TAG_49:1	1,37E-08	0,632649	0,058052

C

TAGs 1:25

	line	year	line:year
TAG_50:1	8,41E-19	0,000216	4,72E-08
TAG_50:2	2,52E-33	0,003206	0,002988
TAG_50:3	2,08E-32	4,55E-05	7,67E-06
TAG_50:4	2,22E-25	0,095961	0,165109
TAG_50:5	1,03E-23	0,038052	0,031506
TAG_50:6	1,85E-17	0,008699	0,059805
TAG_50:0	0,000497	0,003569	3,57E-05
TAG_51:1	1,85E-10	0,244829	2,41E-07
TAG_51:2	1,26E-09	0,173931	0,006812
TAG_51:3	2,13E-23	1,48E-05	0,011878
TAG_51:4	4,92E-37	0,021396	0,033834
TAG_51:5	7,54E-37	0,035811	0,046628
TAG_52:1	7,03E-12	0,348576	3,45E-05
TAG_52:2	5,12E-20	0,661162	0,00421
TAG_52:3	0,000505	0,150407	0,170343
TAG_52:4	2,19E-30	0,108563	0,002519
TAG_52:5	1,36E-29	0,005585	5,54E-05
TAG_52:6	6,46E-22	0,068887	0,005549
TAG_53:2	3,79E-17	0,086143	0,001347
TAG_53:3	1,46E-09	0,93337	0,031149
TAG_53:4	7,09E-30	0,029509	0,000782
TAG_53:5	8,31E-21	0,127424	0,012135
TAG_54:1	0,000803	0,383083	0,001156
TAG_54:2	1,37E-24	0,655035	0,031986
TAG_54:3	6,31E-30	0,006054	0,005379
TAG_54:4	7,7E-06	0,542438	0,539096
TAG_54:5	3,81E-23	0,38222	0,000244
TAG_54:6	2,6E-32	0,323992	0,000484
TAG_54:7	4,03E-26	0,026859	9,62E-06
TAG_55:3	3,26E-17	0,34487	0,00667
TAG_55:4	2,23E-28	0,002323	7,08E-05
TAG_55:5	2,18E-28	0,239709	3,15E-05
TAG_56:1	3,3E-06	0,871693	0,019817
TAG_56:2	2,1E-09	0,751115	0,082762
TAG_56:3	1,94E-09	0,818501	0,009233
TAG_56:4	4,43E-36	0,064389	0,020549
TAG_56:5	2,21E-23	0,052431	1,07E-06
TAG_56:6	5,04E-06	0,145618	0,125046
TAG_58:2	5E-16	0,026016	0,013654
TAG_58:3	6,2E-29	0,000846	0,001114
TAG_58:4	1,6E-37	0,145366	0,009169
TAG_59:2	1,61E-16	0,000513	0,133461
TAG_42:1	0,126788	0,114849	0,009367
TAG_42:2	2,98E-07	0,000544	0,056027
TAG_60:2	2,81E-07	0,242463	0,744573
TAG_60:3	2,21E-12	0,137774	0,024743
TAG_43:1	1,65E-17	2,58E-05	5,61E-07
TAG_44:2	0,0021	0,125248	0,513025
TAG_45:2	0,025696	0,389019	0,474039
TAG_46:1	0,000787	0,527775	0,121074
TAG_46:2	0,007215	0,303529	0,06673

TAG_46:3	8,76E-11	0,218231	0,003362
TAG_47:2	0,103934	0,452594	0,341957
TAG_48:1	2,18E-15	0,000691	1,46E-10
TAG_48:2	1,3E-13	0,041517	7,94E-05
TAG_48:3	1,43E-34	0,001245	0,000171
TAG_48:4	5E-40	0,054847	0,066761
TAG_49:1	0,000221	0,023901	0,00043
TAG_49:3	7,27E-26	0,00044	5,48E-07

Table S2

List of the detected FAs

Fatty acid	Mean	standart devi	min	max	coefficient of variation
FA15:1	10,2385326	0,66269095	8,47341715	12,089522	6,47251884
FA16:2	9,98312609	1,03723817	5,31536155	12,4301316	10,3899135
FA16:1	12,982438	0,69454512	11,5401268	15,2456843	5,34988204
FA16:0	17,1615131	0,33396737	15,7830296	17,8435919	1,94602523
FA17:2	10,0553376	1,03340444	5,19082053	11,99674	10,277173
FA17:1	12,2581979	0,44329156	10,7927607	13,4066194	3,6162865
FA17:0	12,2689301	0,40036596	11,2175041	13,3007369	3,26325077
FA18:3	13,6354804	0,9063083	11,4718123	16,0142258	6,64669139
FA18:0	17,1365557	0,38293585	15,9159791	17,7776049	2,23461385
FA19:2	11,4041371	0,77596992	7,47486597	13,1152879	6,80428434
FA19:1	11,7053152	0,52269008	10,0049514	13,4094555	4,46540797
FA19:0	9,5649866	0,50068281	8,00380705	11,211022	5,23453752
FA20:2	11,3602565	0,67666661	9,36214471	13,2459469	5,95643779
FA20:1	13,9355347	0,47950086	12,5458054	15,4195239	3,44085011
FA20:0	14,1922272	0,59263847	12,4931194	15,6927225	4,17579614
FA21:0	10,3631073	0,46605457	9,02627349	11,6598923	4,49724739
FA22:1	10,0614658	0,48043605	8,41033192	11,712953	4,77501048
FA22:0	15,1358061	0,54806147	10,0395463	16,4064662	3,62095991
FA23:1	8,87986311	0,76680272	6,44020727	11,2777903	8,63529893
FA23:0	11,5458096	0,47944754	10,2976928	12,9366987	4,15256756
FA24:1	11,2072016	0,76940463	6,4460317	13,4885266	6,8652698
FA24:0	13,3679796	0,64242517	6,77911957	15,0781053	4,80570127
FA25:0	9,70215812	0,54341724	8,45966774	11,6561766	5,60099346
FA26:0	10,7862711	0,54850002	9,16074706	12,7558468	5,08516815
FA28:0	8,28408566	0,85818449	5,07778678	11,2336295	10,3594352

These 2 FAs were mesured by a different technique so they can't be directly compared with Fas above

FA18:1	13,9291344	0,55564201	12,1500722	15,2230643	3,98906349
FA18:2	14,2540624	0,76122961	9,96342682	15,2377796	5,34043972

Table S3

List of significant associations

Trait	Marker	Chr	Pos	p	MarkerR2
FA16:2	S14_56559549	14	56559549	7,1681E-07	0,05446
FA17:2	S14_56559549	14	56559549	5,1558E-08	0,06577
FA18:0	S14_97927934	14	97927934	3,5466E-07	0,04895
FA18:1	S09_157078166	9	157078166	2,973E-06	0,03609
FA18:1	S09_157078199	9	157078199	2,973E-06	0,03609
FA18:1	S09_157078201	9	157078201	2,973E-06	0,03609
FA18:1	S09_169225964	9	169225964	8,1008E-06	0,03197
FA18:1	S09_169225972	9	169225972	7,0026E-06	0,03238
FA18:1	S09_169226015	9	169226015	5,5787E-06	0,03341
FA18:1	S09_169226211	9	169226211	3,5163E-06	0,03335
FA18:1	S13_116975962	13	116975962	4,1402E-06	0,03219
FA18:1	S15_36113913	15	36113913	6,1752E-06	0,03148
FA18:2	S03_56565662	3	56565662	8,8741E-07	0,05904
FA18:2	S03_68035606	3	68035606	1,3906E-09	0,0936
FA18:2	S03_68035691	3	68035691	1,6796E-07	0,06144
FA18:2	S03_68035696	3	68035696	1,6796E-07	0,06144
FA18:2	S03_68035765	3	68035765	2,0526E-08	0,08079
FA18:2	S05_37559606	3	37559606	5,5911E-07	0,08001
FA18:2	S05_76163363	3	76163363	8,411E-06	0,0571
FA18:2	S05_212061276	3	212061276	2,2686E-06	0,05684
FA18:2	S11_44296568	11	44296568	2,7182E-06	0,04957
FA18:2	S11_44296577	11	44296577	1,9341E-06	0,05092
FA18:2	S11_44296740	11	44296740	3,5515E-06	0,0578
FA18:2	S11_44296768	11	44296768	4,073E-06	0,05227
FA18:2	S11_96246686	11	96246686	2,0656E-06	0,0533
FA18:2	S12_52300121	12	52300121	4,7122E-06	0,06309
FA18:2	S12_121553510	12	121553510	9,1459E-06	0,05608
FA18:2	S14_56559549	14	56559549	9,6338E-09	0,08589
FA18:2	S14_67200094	14	67200094	1,2615E-06	0,06041
FA18:2	S14_67200134	14	67200134	1,723E-06	0,06071
FA18:2	S14_68135791	14	68135791	1,752E-08	0,08348
FA18:2	S15_61464794	15	61464794	3,9602E-07	0,07139
FA18:2	S15_88437752	15	88437752	2,3958E-07	0,07587
FA18:2	S17_77050404	17	77050404	9,2677E-06	0,05969
FA19:0	S02_179620148	2	179620148	5,8947E-06	0,039
FA19:0	S02_179872169	2	179872169	3,4931E-06	0,04129
FA19:0	S05_12696348	5	12696348	5,8838E-06	0,04244
FA19:0	S10_178263586	10	178263586	4,6462E-06	0,04506
FA19:0	S11_14356229	11	14356229	5,8534E-06	0,04781
FA19:0	S14_53423548	14	53423548	2,1321E-06	0,04415
FA19:0	S14_53480813	14	53480813	4,1691E-06	0,04448
FA19:0	S14_60721724	14	60721724	4,8479E-06	0,03748
FA19:0	S14_60721757	14	60721757	4,9559E-06	0,03741
FA19:0	S14_61085378	14	61085378	5,914E-06	0,05101
FA19:2	S12_121553510	12	121553510	3,0641E-06	0,04924

FA20:0	S11_14356229	11	14356229	9,2389E-06	0,04073
FA20:0	S13_124733368	13	124733368	6,6826E-07	0,04028
FA20:0	S13_124733371	13	124733371	6,6826E-07	0,04028
FA20:0	S13_124733445	13	124733445	1,4966E-06	0,04328
FA20:0	S14_53423548	14	53423548	4,5307E-06	0,03755
FA20:0	S14_96632645	14	96632645	3,0766E-06	0,0407
FA20:0	S14_97927934	14	97927934	5,4287E-07	0,04584
FA22:0	S01_71330751	1	71330751	3,4313E-06	0,03685
FA22:0	S03_32562406	3	32562406	5,8553E-06	0,03198
FA22:0	S03_32562452	3	32562452	1,2711E-07	0,04747
FA22:0	S03_32562479	3	32562479	6,5363E-09	0,05717
FA22:0	S03_32562596	3	32562596	3,8714E-06	0,03628
FA22:0	S03_32562603	3	32562603	4,2962E-06	0,03597
FA22:0	S03_32562660	3	32562660	6,8231E-06	0,03122
FA22:0	S03_32562669	3	32562669	6,8231E-06	0,03122
FA22:0	S03_32562679	3	32562679	6,8231E-06	0,03122
FA22:0	S03_35155696	3	35155696	4,2787E-12	0,06525
FA22:0	S03_43078172	3	43078172	4,0176E-06	0,03796
FA22:0	S03_43078214	3	43078214	9,4011E-07	0,04276
FA22:0	S03_44540519	3	44540519	5,4991E-09	0,05944
FA22:0	S03_44627483	3	44627483	3,4996E-09	0,04928
FA22:0	S03_44696807	3	44696807	1,9894E-09	0,05538
FA22:0	S03_45955683	3	45955683	2,6434E-09	0,0595
FA22:0	S03_45985345	3	45985345	5,5483E-13	0,07743
FA22:0	S03_45985383	3	45985383	8,6519E-26	0,17009
FA22:0	S03_46957204	3	46957204	1,2912E-07	0,04224
FA22:0	S03_48304030	3	48304030	2,973E-09	0,04534
FA22:0	S03_48304163	3	48304163	3,3405E-09	0,04585
FA22:0	S03_48384016	3	48384016	2,6733E-17	0,12247
FA22:0	S03_49228219	3	49228219	2,037E-13	0,094
FA22:0	S03_49488226	3	49488226	2,3357E-09	0,05974
FA22:0	S03_49897780	3	49897780	2,3967E-12	0,07708
FA22:0	S03_53949047	3	53949047	6,4443E-08	0,04636
FA22:0	S03_57635146	3	57635146	6,534E-07	0,04119
FA22:0	S03_57714775	3	57714775	2,4328E-12	0,07176
FA22:0	S03_57714809	3	57714809	5,2314E-26	0,16957
FA22:0	S03_65474677	3	65474677	8,8827E-08	0,05034
FA22:0	S03_87146728	3	87146728	2,7953E-12	0,08183
FA22:0	S04_84340196	4	84340196	1,754E-08	0,05275
FA22:0	S04_84340200	4	84340200	1,754E-08	0,05275
FA22:0	S04_84340205	4	84340205	1,754E-08	0,05275
FA22:0	S07_15277272	7	15277272	2,0219E-06	0,03504
FA22:0	S07_15568578	7	15568578	2,7107E-08	0,05017
FA22:0	S13_124733368	13	124733368	7,5873E-07	0,03614
FA22:0	S13_124733371	13	124733371	7,5873E-07	0,03614
FA22:0	S13_124733445	13	124733445	8,9324E-07	0,03936

FA22:0	S14_91496885	14	91496885	5,7702E-08	0,04837
FA22:0	S14_96627594	14	96627594	1,4405E-06	0,04103
FA22:0	S14_97315952	14	97315952	1,6471E-14	0,09657
FA22:0	S14_97431098	14	97431098	7,9917E-10	0,0615
FA22:0	S14_97431211	14	97431211	1,1224E-09	0,06211
FA22:0	S14_97431236	14	97431236	1,1455E-09	0,06201
FA22:0	S14_97431244	14	97431244	1,2403E-09	0,06163
FA22:0	S14_97431295	14	97431295	2,0298E-09	0,05494
FA22:0	S14_97431297	14	97431297	2,0298E-09	0,05494
FA22:0	S14_100895715	14	100895715	3,2002E-08	0,05043
FA22:0	S14_100895811	14	100895811	9,4062E-11	0,06146
FA22:0	S15_92320390	15	92320390	1,5299E-08	0,05427
FA22:0	S16_176846705	16	176846705	8,8098E-06	0,02944
FA22:0	S16_176846721	16	176846721	8,7231E-06	0,02947
FA24:0	S01_93308752	1	93308752	7,5359E-06	0,03865
FA24:0	S01_144787486	1	144787486	3,517E-11	0,07625
FA24:0	S02_7565103	2	7565103	4,4062E-08	0,05589
FA24:0	S02_56965554	2	56965554	2,5551E-07	0,04772
FA24:0	S02_73778898	2	73778898	8,7218E-11	0,07715
FA24:0	S02_75638693	2	75638693	9,9635E-12	0,08944
FA24:0	S03_69335993	3	69335993	8,6212E-06	0,03755
FA24:0	S03_102070278	3	102070278	6,5452E-06	0,03983
FA24:0	S03_102070280	3	102070280	6,5452E-06	0,03983
FA24:0	S04_176175991	4	176175991	5,7803E-12	0,07753
FA24:0	S04_176176013	4	176176013	6,0778E-12	0,077
FA24:0	S04_176176245	4	176176245	4,7921E-12	0,07774
FA24:0	S08_56199369	8	56199369	6,9325E-07	0,04616
FA24:0	S08_63370394	8	63370394	6,0819E-08	0,04957
FA24:0	S09_171231730	9	171231730	3,2034E-10	0,06224
FA24:0	S09_171231751	9	171231751	3,2034E-10	0,06224
FA24:0	S09_171231763	9	171231763	3,1924E-10	0,06226
FA24:0	S09_171231789	9	171231789	3,773E-10	0,06447
FA24:0	S10_31991298	10	31991298	1,2678E-10	0,07588
FA24:0	S10_151937534	10	151937534	6,9815E-07	0,04625
FA24:0	S10_232655851	10	232655851	6,0518E-07	0,04642
FA24:0	S15_57100996	15	57100996	1,9424E-06	0,0428
FA24:0	S15_57101285	15	57101285	2,0177E-06	0,04268
FA24:0	S15_134088807	15	134088807	3,2332E-09	0,06302
FA24:0	S16_2638460	16	2638460	8,7594E-06	0,0368
FA24:0	S16_172000404	16	172000404	4,1574E-09	0,05556
FA24:1	S03_45985383	3	45985383	1,9638E-09	0,08564
FA24:1	S03_57714809	3	57714809	3,2196E-09	0,07721
FA24:1	S15_92320390	15	92320390	2,3252E-06	0,05984
18;1/18:2ratio	S03_68035606	3	68035606	2,9022E-06	0,05995
18;1/18:2ratio	S03_68035691	3	68035691	8,7239E-06	0,04838
18;1/18:2ratio	S03_68035696	3	68035696	8,7239E-06	0,04838

18;1/18:2ratio	S03_68035765	3	68035765	9,6389E-07	0,06535
18;1/18:2ratio	S12_121553510	12	121553510	3,2185E-06	0,06479
18;1/18:2ratio	S14_56559549	14	56559549	4,7553E-09	0,09695

Table S4

List of candidate genes

Trait	Chromosome	Gene ID	Function
FA18:0	14	HanXRQChr14g0439881	Putative serine/threonine/dual specificity protein kinase, catalytic domain; Tyrosine-protein kinase, receptor ROR
FA18:1	6	HanXRQChr06g0180681	Probable 30S ribosomal protein S11, chloroplastic
FA18:1	9	HanXRQChr09g0261691	Uncharacterized protein, supported by expression data
FA18:1	9	HanXRQChr09g0261651	Uncharacterized protein, supported by expression data
FA18:1	9	HanXRQChr09g0261661	Putative tetratricopeptide-like helical domain
FA18:1	9	HanXRQChr09g0261671	Probable quinone reductase family protein
FA18:1	9	HanXRQChr09g0261681	Putative calcium/proton exchanger
FA18:1	9	HanXRQChr09g0264571	Probable long-chain-alcohol oxidase FAO1
FA18:1	9	HanXRQChr09g0264581	Uncharacterized protein, supported by expression data
FA18:1	9	HanXRQChr09g0264591	Putative ribosomal protein/NADH dehydrogenase domain; Thioredoxin-like fold
FA!8:2	3	HanXRQChr03g0070311	Putative uncharacterised protein family, basic secretory protein
FA!8:2	3	HanXRQChr03g0070321	Putative uncharacterised protein family, basic secretory protein
FA!8:2	3	HanXRQChr03g0070331	Probable plant basic secretory protein (BSP) family protein
FA!8:2	3	HanXRQChr03g0070351	Putative phosphoribosyltransferase-like
FA!8:2	3	HanXRQChr03g0070361	Uncharacterized protein, supported by expression data
FA!8:2	3	HanXRQChr03g0070371	Uncharacterized protein, supported by expression data
FA!8:2	3	HanXRQChr03g0070381	Uncharacterized protein, supported by expression data
FA!8:2	5	HanXRQChr05g0135651	Putative ATPase, AAA-type, core; Peptidase, FtsH; Peptidase M41, FtsH extracellular; P-loop containing nucleoside triphosphate hydrolase
FA!8:2	5	HanXRQChr05g0135661	Putative small GTPase superfamily; P-loop containing nucleoside triphosphate hydrolase
FA!8:2	5	HanXRQChr05g0135671	Putative sugar phosphate transporter domain
FA!8:2	12	HanXRQChr12g0368171	Uncharacterized protein
FA!8:2	12	HanXRQChr12g0368181	Uncharacterized protein
FA18:2	14	HannXRQChr14g0435061	lncRNA
FA18:3	11	HanXRQChr11g0330521	Probable 60S ribosomal protein L6
FA18:3	11	HanXRQChr11g0330531	Putative thioredoxin-like fold
FA18:3	11	HanXRQChr11g0330541	Putative bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain
FA19:0	2	HanXRQChr02g0059821	Probable transcription factor-related
FA19:0	2	HanXRQChr02g0059831	Probable histone deacetylase 8
FA19:0	2	HanXRQChr02g0059841	Putative seipin family
FA19:0	2	HanXRQChr02g0059851	Putative histone deacetylase complex subunit SAP30/SAP30-like
FA19:0	14	HanXRQChr14g0433231	Probable chitinase 5
FA19:0	14	HanXRQChr14g0433911	Putative tyrosine-protein kinase, neurotrophic receptor, type 3
FA19:0	14	HanXRQChr14g0433921	Putative mitogen-activated protein (MAP) kinase kinase kinase 10
FA19:0	14	HanXRQChr14g0433931	Putative protein kinase-like domain
FA19:0	14	HanXRQChr14g0433941	Putative ephrin receptor type-A /type-B
FA19:0	14	HanXRQChr14g0433951	Putative serine/threonine-protein kinase, active site; Ephrin receptor type-A /type-B
FA19:0	14	HanXRQChr14g0433961	Putative protein kinase-like domain
FA19:0	14	HanXRQChr14g0433971	Putative serine/threonine protein kinase, striated muscle-specific
FA19:0	14	HanXRQChr14g0433981	Uncharacterized protein, supported by expression data
FA19:0	14	HanXRQChr14g0433991	Putative serine/threonine-protein kinase, active site; Tyrosine-protein kinase, insulin-like receptor
FA19:0	14	HanXRQChr14g0434001	Putative tyrosine-protein kinase, Fes/Fps type
FA19:0	14	HanXRQChr14g0434011	Putative tyrosine-protein kinase, neurotrophic receptor, type 3
FA19:0	14	HanXRQChr14g0434021	Putative serine-threonine/tyrosine-protein kinase catalytic domain; Protein kinase-like domain
FA19:0	14	HanXRQChr14g0434031	Putative mitogen-activated protein (MAP) kinase kinase kinase, MLK1/MLK2/MLK4
FA19:0	14	HanXRQChr14g0434041	Uncharacterized protein, partly supported by expression data
FA19:0	14	HanXRQChr14g0434051	Putative tyrosine-protein kinase, neurotrophic receptor, type 3
FA19:0	14	HanXRQChr14g0434061	Putative ephrin receptor type-A /type-B
FA19:0	14	HanXRQChr14g0434081	Uncharacterized protein
FA19:0	14	HanXRQChr14g0434111	Putative serine/threonine-protein kinase Plk3
FA19:0	14	HanXRQChr14g0434121	Putative tyrosine-protein kinase, non-receptor Jak/Tyk2
FA19:0	14	HanXRQChr14g0434131	Putative mitogen-activated protein (MAP) kinase kinase kinase 7; Rho-associated protein kinase 1/2
FA19:0	14	HanXRQChr14g0434141	Putative mitogen-activated protein (MAP) kinase kinase kinase 10
FA19:0	14	HanXRQChr14g0434151	Putative tyrosine-protein kinase, neurotrophic receptor
FA19:0	14	HanXRQChr14g0434161	Putative protein kinase-like domain
FA19:0	14	HanXRQChr14g0434171	Putative tyrosine-protein kinase, CSF-1/PDGF receptor family
FA19:0	14	HanXRQChr14g0434181	Putative tyrosine-protein kinase, non-receptor Jak2
FA19:0	14	HanXRQChr14g0434191	Putative mitogen-activated protein (MAP) kinase kinase kinase 10
FA19:0	14	HanXRQChr14g0434071	Putative concanavalin A-like lectin/glucanase domain; Mitogen-activated protein (MAP) kinase kinase 10
FA19:0	14	HanXRQChr14g0434091	Putative protein kinase-like domain
FA19:0	14	HanXRQChr14g0434101	Putative serine/threonine/dual specificity protein kinase, catalytic domain; Fibroblast growth factor receptor family
FA22:0	3	HanXRQChr03g0065591	putative F-box family protein
FA22:0	3	HannXRQChr03g0065601	lncRNA
FA22:0	3	HannXRQChr03g0066951	putative beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ, HotDog domain protein
FA22:0	3	HanXRQChr03g0066961	putative chlorophyll a-b binding protein CP24 10A protein
FA22:0	3	HannXRQChr03g0066971	lncRNA

FA22:0	3	HanXRQChr03g0066981	mRNA-putative galactose oxidase/kelch repeat superfamily protein
FA22:0	3	HannXRQChr03g0066991	putative transducin/WD40 repeat-like superfamily protein
FA22:0	3	HanXRQChr03g0067001	putative mRNA export factor
FA22:0	3	HannXRQChr03g0067011	putative protein kinase-like domain-containing protein
FA22:0	3	HannXRQChr03g0067201	putative CRAL-TRIO lipid binding domain-containing protein
FA22:0	3	HanXRQChr03g0067211	putative ethanolamine-phosphate cytidylyltransferase
FA22:0	3	HanXRQChr03g0067221	putative arginine decarboxylase
FA22:0	3	HannXRQChr03g0067231	lncRNA
FA22:0	3	HannXRQChr03g0067351	putative phospholipase A2, Phospholipase A2 domain protein
FA22:0	3	HannXRQChr03g0067591	lncRNA
FA22:0	3	HannXRQChr03g0067601	mRNA-putative plant self-incompatibility S1 and OTG30718.1
FA22:0	3	HannXRQChr03g0067661	putative eamA domain, WAT1-related protein
FA22:0	3	HannXRQChr03g0068281	putative peptidase S10, serine carboxypeptidase, Alpha/Beta hydrolase fold protein
FA22:0	3	HannXRQChr03g0068271	putative peptidase S10, serine carboxypeptidase, Alpha/Beta hydrolase fold protein
FA22:0	3	HannXRQChr03g0068291	lncRNA
FA22:0	3	HannXRQChr03g0068301	putative ARID DNA-binding domain, Zinc finger, CCHC-type, MEKHLA
FA22:0	3	HannXRQChr03g0068901	putative ccc1 family
FA22:0	7	HanXRQChr07g0188431	Putative DNA-binding domain
FA22:0	7	HanXRQChr07g0188441	Uncharacterized protein, supported by expression data
FA22:0	7	HanXRQChr07g0188451	Putative BRCT domain; FCP1 homology domain; Armadillo-type fold; HAD-like domain
FA22:0	7	HanXRQChr07g0188461	Putative DNA-binding domain
FA22:0	14	HanXRQChr14g043979	putative phosphoglucomutase protein
FA22:0	14	HannXRQChr14g0439801	putative sas10/U3 ribonucleoprotein (Utp) family protein
FA22:0	14	HannXRQChr14g0439821	putative ribosomal protein L1p/L10e family
FA22:0	14	HanXRQChr14g0439811	putative CTC-interacting domain 12
FA22:0	14	HannXRQChr14g0439831	putative helicase, P-loop containing nucleoside triphosphate hydrolase
FA22:0	14	HanXRQChr14g0439851	putative membrane-bound transcription factor site-2 protease
FA22:0	14	HanXRQChr14g0439851	putative import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein
FA22:0	14	HanXRQChr14g0439841	putative related to ubiquitin 1
FA22:0	14	HanXRQChr14g0438921	Probable phosphatidic acid phosphatase (PAP2) family protein
FA22:0	14	HanXRQChr14g0438931	Putative homeodomain-like; MYB-CC type transcription factor, LHEQLE-containing domain
FA22:0	14	HanXRQChr14g0438941	Uncharacterized protein, supported by expression data
FA22:0	14	HanXRQChr14g0438951	Putative MYB-CC type transcription factor, LHEQLE-containing domain
FA22:0	14	HanXRQChr14g0438961	Uncharacterized protein
FA22:0	14	HanXRQChr14g0438971	Putative CCR4-Not complex component, Not N-terminal domain
FA22:0	14	HanXRQChr14g0438981	Putative ribosomal protein L2
FA22:0	14	HanXRQChr14g0438991	Putative protein kinase-like domain
FA22:0	14	HanXRQChr14g0439001	Putative mitogen-activated protein (MAP) kinase kinase kinase Ste11, Cryptococcus
FA22:0	14	HanXRQChr14g0439011	Putative zinc finger, RING/FYVE/PHD-type
FA22:0	14	HanXRQChr14g0439751	Probable VACUOLAR SORTING RECEPTOR 7
FA22:0	14	HanXRQChr14g0439761	Uncharacterized protein, supported by expression data
FA22:0	14	HanXRQChr14g0439781	Putative membrane bound O-acyl transferase, MBOAT
FA22:0	16	HanXRQChr16g0530601	Uncharacterized protein, supported by expression data
FA22:0	16	HanXRQChr16g0530611	Probable tubby-like F-box protein 5
FA24:0	2	HanXRQChr02g0040501	Uncharacterized protein, partly supported by expression data
FA24:0	2	HanXRQChr02g0040511	Probable protein kinase superfamily protein
FA24:0	2	HanXRQChr02g0041891	Probable myb domain protein 69
FA24:1	3	HanXRQChr03g0067361	Probable cation transporter HKT2
FA24:1	3	HanXRQChr03g0067371	Uncharacterized protein, partly conserved in plant genome(s)
FA24:1	3	HanXRQChr03g0068901	Putative ccc1 family
18:1/18:2ratio	3	HanXRQChr03g0070181	Probable PTEN 2
18:1/18:2ratio	3	HanXRQChr03g0070191	Putative C2 domain
18:1/18:2ratio	3	HanXRQChr03g0070201	Putative C2 domain
18:1/18:2ratio	3	HanXRQChr03g0070211	Putative C2 domain
18:1/18:2ratio	3	HanXRQChr03g0070221	Putative C2 domain
18:1/18:2ratio	3	HanXRQChr03g0070231	Putative DNA-directed RNA polymerase, alpha subunit
18:1/18:2ratio	3	HanXRQChr03g0070241	Putative leucine-rich repeat domain, L domain-like
18:1/18:2ratio	3	HanXRQChr03g0070251	Uncharacterized protein, supported by expression data
18:1/18:2ratio	3	HannXRQChr03g0070281	putative chlorophyll a/b binding protein domain-containing protein