Skoltech

Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology

GENETIC ASSOCIATION MAPPING OF AGRONOMICALLY IMPORTANT TRAITS IN RAPESEED AND SUNFLOWER

Doctoral Thesis

By

RIM GUBAEV

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Philipp Khaitovich

Moscow – 2022

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.

Candidate (Rim Gubaev)

Supervisor (Prof. Philipp Khaitovich)

Abstract

Oilseeds crops are the main source of vegetable oil production. Constant population growth, new industrial applications, as well as climatic change, require the selection of the high yield varieties that are adapted to changing populational requirements and unfavorable environmental conditions. Additionally, current breeding programs should meet the high-quality standards of the industry. One of the established approaches facilitating the breeding is marker-assisted and genomic selection, relying on modern genotyping techniques. Both approaches reduce the time and financial costs needed for plant evaluation in the field through the selection and monitoring of the preferred genotypes by means of genetic tests. One of the major steps that underlies both approaches is a genetic characterization of the mapping population followed by the identification of the genetic markers associated with the trait of interest. In Russia, the top three oilseed crops are sunflower, rapeseed, and soybean. Despite their substantial economic and food safety value, the application of genetic-information-based breeding approaches to these crops in Russia is currently lagging. This thesis focuses on the investigation of genetic markers associated with agronomically important traits in rapeseed and sunflower with the intent of enhancing marker-assisted rapeseed and sunflower breeding in Russia. Genetic markers were identified for glucosinolate content in rapeseed and for oil quality and seed morphology traits. Markers for glucosinolate content and seed morphology traits should be subjected to validation. While markers for oil quality for sunflower were validated and could be used directly for marker-assisted selection in sunflower.

Publications

Main publications:

Gubaev, R.; Boldyrev, S.; Martynova, E.; Chernova, A.; Kovalenko, T.; Peretyagina, T.; Goryunova, S.; Goryunov, D.; Mukhina, Z.;, Ben, C.; Gentzbittel, L.; Khaitovich P.; Demurin, Y. (2022). Genetic mapping of loci involved in oil tocopherol composition control in Russian sunflower (*Helianthus annuus* L.) lines. *G3: Genes, Genomes Genetics*, 12, 4, jkac036. https://doi.org/10.1093/g3journal/jkac036

Chernova, A.I.; **Gubaev, R.F.**; Singh, A.; Sherbina, K.; Goryunova, S.V.; Martynova, E.U.; Goryunov, D.V.; Boldyrev S.V.; Vanyushkina, A.A.; Anikanov, N.A.; Stekolshchikova E.A.; Yushina, E.A.; Demurin, Y.N.; Mukhina, Z.M.; Gavrilova, V.A.; Anisimova, I.N.; Karabitsina, Y.I.; Alpatieva, N.V.; Chang P.L.; Khaitovich P. E.; Mazin P. V.; Nuzhdin, S.V. (2021). Genotyping and lipid profiling of 601 cultivated sunflower lines reveals novel genetic determinants of oil fatty acid content. *BMC genomics*, 22(1), 1-15. https://doi.org/10.1186/s12864-021-07768-y

Gubaev, R.; Gorlova, L.; Boldyrev, S.; Goryunova, S.; Goryunov, D.; Mazin, P.; Chernova, A.; Martynova, E.; Demurin, Y.; Khaitovich, P. (2020) Genetic Characterization of Russian Rapeseed Collection and Association Mapping of Novel Loci Affecting Glucosinolate Content. *Genes*, 11, 926. https://doi.org/10.3390/genes11080926

Additional publications:

Goryunov, D.V.; Anisimova, I.N.; Gavrilova, V.A.; Chernova, A.I.; Sotnikova, E.A.; Martynova, E.U.; Boldyrev, S.V.; Ayupova, A.F.; **Gubaev, R.F.**; Mazin, P.V.; Gurchenko, E.A.; Shumskiy, A.A.; Petrova, D.A.; Garkusha, S.V.; Mukhina, Z.M.; Benko, N.I.; Demurin, Y.N.; Khaitovich, P.E.; Goryunova, S.V. (2019). Association Mapping of Fertility Restorer Gene for CMS PET1 in Sunflower. *Agronomy*, 9, 49. https://doi.org/10.3390/agronomy9020049

Chernova, A.; **Gubaev, R.**; Mazin, P.; Goryunova, S.; Demurin, Y.; Gorlova, L.; Vanushkina, A.; Mair, W.; Anikanov, N.; Martynova, E.; Goryunov, D.; Garkusha, S.; Mukhina, Z.; Khaitovich, P. (2019). UPLC–MS Triglyceride Profiling in Sunflower and Rapeseed Seeds. *Biomolecules*, 9, 9. https://doi.org/10.3390/biom9010009

Conference presentations:

Gubaev R.; Boldyrev S.; Chernova A.; Martynova E.; Kovalenko T.; Goryunova S.; Goryunov D.; Peretyagina T.; Demurin Y.; Khaitovich P. (2022). Mapping of loci associated with tocopherols content using genotyping by sequencing approach in sunflower. *20th International Sunflower Conference*, Novi Sad, Serbia - oral presentation

Gubaev R.; Boldyrev S.; Martynova E.; Chernova A.; Kovalenko T.; Peretyagina T.; Goryunova S.; Goryunov D.; Mukhina Z.; Ben C.; Gentzbittel L.; Khaitovich P.; Demurin Y. (2022). Quantitative and qualitative mapping of loci involved in tocopherol composition and oleic acid content control in Russian sunflower (*Helianthus annuus* L.) lines. *Population, Evolutionary, and Quantitative Genetics Conference (PEQG22)*, Pacific Grove, CA, United States of America - poster presentation (online)

Gubaev R.; Boldyrev S.; Chernova A.; Martynova E.; Kovalenko T.; Goryunova S.; Goryunov D.; Peretyagina T.; Demurin Y.; Khaitovich P. (2021). QTL mapping of oleic acid content and tocopherol composition in sunflower from VNIIMK collection. *Moscow Conference on Computational Molecular Biology (MCCMB21)*, Moscow, Russia - poster presentation

Gubaev R.; Boldyrev S.; Chernova A.; Martynova E.; Kovalenko T.; Goryunova S.; Goryunov D.; Peretyagina T.; Demurin Y.; Khaitovich P. (2021). Genetic map construction and QTL mapping of oil-related traits in sunflower from VNIIMK collection. *The Sixth International Scientific Conference "Plant Genetics, Genomics, Bioinformatics, and Biotechnology" (PlantGen2021)*, Novosibirsk, Russia - oral presentation

Gubaev R.; Goryunova S.; Goryunov D.; Mazin P., Boldyrev S.; Chernova A.; Ayupova A.; Martynova E.; Demurin Y.; Mukhina Z.; Khaitovich P. (2019). Mapping of genetic associations for agronomically important traits in Russian collection of rapeseeds (*Brassica napus*). *Moscow Conference on Computational Molecular Biology (MCCMB19)*, Moscow, Russia - poster presentation

Gubaev R.; Goryunova S.; Goryunov D.; Mazin P., Boldyrev S.; Chernova A.; Ayupova A.; Martynova E.; Demurin Y.; Mukhina Z.; Khaitovich P. (2019). Association mapping of agronomically important traits in Russian collection of rapeseed. *The Fifth International Scientific Conference "Plant Genetics, Genomics, Bioinformatics, and Biotechnology" (PlantGen2019)*, Novosibirsk, Russia - oral presentation

Acknowledgments

First of all I would like to express my gratitude to my supervisor Philipp Khaitovich for creating the productive environment which on the one hand facilitated my progress as an independent researcher and on the other hand forced me to collaborate with lab members and external scientists to do research. I sincerely thank our "OilGene" startup members Alina Chernova, Elena Martynova and Stepan Boldyrev for their constant support and initiation of the long process of implementation of our scientific results for oilseed breeding improvement. I thank the senior members of the Digital Agriculture Laboratory Laurent Gentzbittel and Cecile Ben for their support in mastering quantitative genetics approaches and for providing constructive critiques for our joint publications. I thank all the people who were involved in the realization of the project "Development of a technology for accelerated breeding of oil crops on the basis of high-performance methods of genotyping and molecular phenotyping to ensure food security in Russia" supported by the Ministry Of Science and Higher Education after which this thesis became possible, especially here I thank Svetlana Goryunova, Denis Goryunov, Pavel Mazin and our collaborators from VNIIMK Lyudmila Gorlova and Yakov Demurin. I thank my friend Elizaveta Grigoreva who supported me during this challenging period. I am also expressing my gratitude to my fellow graduate students Maria Osetrova and Alexandr Ilin for their moral support and positive attitude.

Table of contents

List of abbreviations

AFLP – Amplified fragment length polymorphism

ANOVA – Analysis of Variance

BC – Back cross

bp – base pair

CIM – Composite interval mapping

cMLM – Compressed mixed linear model

CMS – Cytoplasmic male sterility

DH – Doubled haploid

FA – Fatty acid

GBS – Genotyping by sequencing

GC – Gas chromatography

GS – Genomic selection

GWAS – Genome-wide association study

IMI – Imidazolinone

kb – kilobase

LAMP – Loop-mediated isothermal amplification

LD – Linkage disequilibrium

LOD – Logarithm of the odds

MAGIC – Multi-parent advanced generation inter-cross

MAS – Marker-assisted selection

Mb – Megabase

MLM – Mixed linear model

MPPs – Multi-parental populations

MS – Mass spectrometry

NAM – Nested association mapping

NGS – Next generation sequencing

PC – Principal component

PCA – Principal component analysis

PCR – Polymerase chain reaction

PVE – Phenotypic variation explained

QTL – Quantitative trait loci

R gene – resistance gene

RAD – Restriction-site associated DNA

RAPD – Random amplified polymorphic DNA

RFLP – Restriction fragment length polymorphism

RILs – Recombinant inbred lines

SIM – simple interval mapping

SLAF – Specific-locus amplified fragment sequencing

SNP – Single nucleotide polymorphism

SSR – Simple sequence repeat

SU – Sulfonylurea

TLC – Thin-layer chromatography

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

VNIIMK – Pustovoit All-Russia Research Institute of Oil Crops

Chapter 1. Introduction

Starting from the neolithic period people started to domesticate wild plants turning them into crops, thus securing a controlled food supply. From then on plant breeding relying on hybridization and selection continued fundamentally unchanged till the introduction of genetics. The genetics concepts entered the plant breeding field with the rediscovery of Mendel's Laws of Heredity at the end of the 19th century. Within the 20th century, increasing understanding of genetic mechanisms led to important innovations in plant breeding. These innovations included induced mutagenesis, relying on the application of radiation and chemicals, which facilitated the creation of plants with the traits lacking in the natural populations. In addition, directed mutagenesis based on the *Agrobacterium*-mediated transformation and gene gun technique was applied to transfer genes introducing novel agronomically important traits. The development of the molecular genetics approaches and DNA sequencing technologies led to another technological innovation called marker-assisted selection implementing genetic markers whose allelic states, in turn, may indicate the presence or absence of a specific trait. This technology allowed breeders to accelerate the breeding process by replacing lab or field phenotypic assessment with genetic testing and optimizing the breeding schemes, reducing time and financial expenditure. At the beginning of the 21st century, the other two technological breakthroughs emerged: genome editing techniques allowing the introduction of site-specific mutations in the genes of interest and high throughput genotyping. The high throughput genotyping of crops facilitated studies of genetic diversity and improved prediction of complex polygenic traits providing a basis for improved genomic selection.

All above mentioned approaches and technologies are nowadays applied in the broad spectrum of crops including oilseeds – the major source of vegetable oil, an important part of our diet. In addition to their dietary role, vegetable oil and oilseed by-products are used to produce biofuels, chemicals and serve as fodder for animals. The top five mostly produced oilseeds worldwide include soybean, rapeseed,

sunflower, peanut, and cottonseed. Due to the constant population growth, the demand for vegetable oil permanently increases. Additionally, increasing food quality standards, food and chemical industry requirements, and ecological and climate alterations require constant improvement of specific oilseed traits, such as productivity, composition and quality of the oil, and resistance to pathogens and unfavorable environmental conditions. All of these challenges could be met more timely by the implementation of genetics-based technological approaches.

In Russia, the most produced oilseed crops are sunflower, rapeseed, and soybean. Importantly, Russia is one of the world's leaders in sunflower production. Despite that, modern genetic technologies are not broadly used by scientific organizations and breeding companies. As a consequence, two problems arise. First, Russian oilseed germplasm is poorly characterized in terms of genetic diversity. Second, the implementation of genetic modification approaches and marker- and genome-assisted breeding is not a common trend. This impedes the creation of new competitive varieties meeting market demands. As a consequence, Russian oilseed producers often use imported seed material due to its higher quality, rooted in more efficient breeding strategies involving genetics-based technological solutions.

To improve this situation, several attempts are now made with an aim of supporting the introduction of novel genetic and biotechnological innovations in Russian scientific institutions and breeding companies. The present work is part of the bigger initiative supported by the Ministry of Science and Higher Education aimed at developing marker- and genomic-assisted approaches in oilseed crop breeding for providing food security in Russia. Within this initiative, the Skolkovo Institute of Science and Technology (Skoltech) collaborated with leading organizations involved in germplasm preservation and breeding of oilseeds, namely, the Vavilov Institute of Plant Industry (VIR), Pustovoit All-Russia Research Institute of Oil Crops (VNIIMK), and Agroplasma breeding company. The present thesis focuses on the association mapping of agronomically valuable traits in rapeseed and

sunflower. Additionally, a basic analysis of genetic diversity in rapeseed and sunflower germplasms from the above-mentioned organizations was performed.

The aim of the present study was to map agronomically important traits in rapeseed and sunflower. The following objectives were set within this aim:

1) Find the genetic associations for glucosinolate content by analyzing the rapeseed diversity panel from the VNIIMK.
2) Find genetic associations with the tocopherol composition and oleic acid content of sunflower oil by analyzing experimental crosses generated by the VNIIMK.
3) Find genetic associations with seed and husk size, seed to husk ratio using the diversity panel from VNIIMK, VIR and Agroplasma.
4) Characterize the genetic diversity of the rapeseed collection from VNIIMK.
5) Characterize the genetic diversity of the Russian sunflower collection from VNIIMK, VIR, and Agroplasma.

Chapter 2. Literature review

2.1 Marker-assisted and genomic selection as tools for plant breeding

Plant breeding is a set of methods for creating plant varieties and hybrids with the desired properties with an aim of increasing the yield and quality of crops. Starting from the neolithic revolution, people began to domesticate plants, select and reproduce ones with the desired properties. For a long time, plant breeding was performed empirically without any strong scientific background. Systematic plant breeding based on hybridization and selection began to be possible in the early 20th century after the rediscovery of Mendel's laws (Lee et al., 2015).

The idea of using markers, associated with certain traits that may help to facilitate the selection of prospective progeny appeared in the first half of the 20th century. The first markers were morphological traits, such as seed coat color, associated with the seed size in a common bean (Sax, 1923), or flower color associated with flowering time in pea (Rasmusson, 1935). In addition to morphological markers, cytological and biochemical ones were developed throughout the 20th century. To date, however, the majority of markers of traits are DNA-based (Nadeem et al., 2018).

The concept of marker-assisted selection (MAS) based on the DNA markers became popular in the 1980s, partially due to the development of laboratory techniques that allowed the identification of genetic polymorphisms among individuals in the population, some of which were related to the specific traits of interest (Ben-Ari & Lavi, 2012). The main idea was the following: not all genetic polymorphisms are associated with a trait of interest; however, some of them could be associated with monogenic or polygenic traits controlled by quantitative trait loci (QTL) and thus may substitute field testing with predictions based on the markers' states (Smith & Simpson, 1986; Soller & Beckmann, 1983). The marker-assisted selection is based on the concept of linkage disequilibrium (LD), a non-random

association between alleles at a certain region on a chromosome affected by different factors including genetic linkage, selection, mutation, non-random mating, population structure, genetic drift and recombination. Thus a genetic marker of a trait could represent a causal mutation in a gene that affects the phenotype (rare case) or it could be linked to the allele of the gene affecting the phenotype by LD (Mackay & Powell, 2007). MAS is commonly used to mark traits associated with the resistance to pathogens and R-gene pyramiding, resistance to herbicides, ability to restore fertility etc., It should be noted that most of these traits are mono-/oligogenic (Bulos, 2013; Melchinger, 1990; Ramalingam et al., 2020).

The modern marker-assisted selection is based on the implementation of genetic markers that could be of a different kind. To date, five more common groups of genetic markers could be highlighted: restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), Simple sequence repeats (SSR) and single nucleotide polymorphisms (SNPs). The first three marker systems (fragment-based polymorphisms) were the first ones to be applied widely in the 1980s and 1990s: RFLP and AFLP markers are based on the point mutations affecting the fragment restriction, while RAPD is based on point mutations that affect primer annealing (Ben-Ari & Lavi, 2012). All these markers do not require prior information about sequence but require gel electrophoresis for the detection, additionally, these markers are not so spread within a genome and to date are rarely used. Quite popular SSR markers represent short (2-6 bp long) repeated sequences that are widely distributed among plant genomes. The main advantage of these markers is a high level of polymorphism i.e. one marker can distinguish up to more than ten alleles (Kalia et al., 2011). The most commonly used markers nowadays are SNPs. The main advantage of the SNP markers is that these kinds of markers occur at higher rates through the genomes compared to the abovementioned markers which in turn makes it possible to find more genetic associations for traits of interest, especially polygenic ones. SNPs markers became widespread at the beginning of the 21st century partially due to the

rapid development of the high-throughput genotyping approaches, namely DNA microarrays and next-generation sequencing whose implementation towards the genetic diversity assessment is mostly concentrated on the detection of the SNP markers (Rasheed et al., 2017).

The ability to identify a high amount of genetic markers led to the development of the novel branch of marker-assisted selection called genomic selection (GS) at the beginning of the 21st century (Meuwissen et al., 2001). Unlike MAS which mostly concentrates on the one and/or several QTL associated with mono-/oligo genic traits with high heritability, GS relies on the hundreds or thousands of genetic markers, (usually SNPs or SSR) with small effects (Crossa et al., 2017). This allows the estimation of genome-estimated breeding values (GEBVs) of a certain individual and to select candidates for the breeding program (Nadeem et al., 2018). To date, several attempts have been made to facilitate the development of GS-based breeding programs, most of them were done on crops (wheat, maize, rice), and the key traits to improve included yield and resistance to unfavorable environmental conditions (Bhat et al., 2016).

The advantage of the MAS and GS results from the fact that the vast majority of the traits of interest are not easily assessed, and usually need additional laboratory testing (resistance to pathogens, metabolite content assessment, etc.) or field trials (yield, resistance to unfavorable environmental conditions, etc.). Thus the implementation of MAS and GS allows for reducing the time and financial costs of breeding (Ben-Ari & Lavi, 2012). It also should be noted that in recent years startups like NRGene, NeoGene, GeneCove e.t.c., are emerging and suggest a service in MAS and GS for breeding and seed companies with the aim of implementing genetic markers to facilitate new varieties' development.

Summarizing the information reviewed above it could be concluded that marker-assisted and genomic selection is now widely applied with the aim of obtaining new varieties at a higher rate with lower costs. It should be noted that the development of marker-assisted breeding approaches is based on the scanning for the

genetic markers of the agronomically important traits with subsequent development of genetic tests as well as marker panels.

2.2 Approaches for finding genetic markers and developing genotyping solutions

A basic procedure that underlies marker-assisted selection is scanning for the genetic markers for the traits of interest. Search for genetic markers consists of the following steps: 1) selection of the association panel, 2) phenotyping, 3) genotyping, 4) association mapping and 5) marker validation.

The scanning for the genetic markers usually starts with the creation and analysis of the population. In general, the two most common types of mapping populations are currently used to perform genetic mapping. The first one includes the diversity panels, which are usually formed by sampling the gene banks' accessions of phenotypically and genetically contrasting lines that may include wild, domesticated, landraces and modern ones (Brachi et al., 2011). The second type involves artificial populations obtained by means of experimental crosses. Artificial populations, in turn, may be classified in bi-parental including F2, double haploid (DH) backcross (BC), and recombinant inbred lines (RILs) populations and multi-parental populations (MPPs), including nested association-mapping populations (NAMs), multi-parent advanced generation inter-cross (MAGIC). Bi-parental populations are based on the crossing of two parents contrasting in the trait of interest followed by a subsequent rounds of selfing (F2, RILs), backcrossing (BC), or production of doubled haploids (DHs). Multi-parental populations include several parents: NAM consists of a series of biparental populations having one common founder as a parent. MAGIC has a simple funnel breeding design where several unrelated parental pairs (4-16) are crossed (Scott et al., 2020).

The advantage of the diversity panels is the higher resolution of association mapping. The analysis of these populations is based on the detection of associated regions generated by historical recombination events. Thus it is possible to detect loci

associated with traits in the narrow genomic regions (Mohammadi et al., 2020). The advantage of the bi-parental populations is the ability to study traits with low natural diversity, or traits artificially introduced by induced mutagenesis (Oladosu et al., 2016). Additionally, only two generations are required to produce a mapping population. However, the mapping resolution remains suboptimal due to a low number of recombination events (Arrones et al., 2020). Multi-parental populations allow for increasing mapping resolution compared to bi-parental populations. Further, it is possible to artificially increase the frequency of the rare alleles with an aim of detecting rare variants associated with the traits of interest. The disadvantage of MPPs is the substantial time needed to produce crosses and selfing to create a mapping population (Pascual et al., 2015).

The next step preceding the marker scanning is phenotyping, which is to some extent more important than the subsequent genotyping since the accuracy and replication of the phenotype data affect the prediction accuracy of the identified markers. Additionally, phenotyping costs are larger than the genotyping ones as genotypic data for a mapping population is obtained once and can be used to map several traits of interest. Thus, the collection of phenotype data might create a bottleneck in association studies due to the time and labor costs needed for its collection (Reynolds et al., 2019). Another important aspect of phenotyping is the collection of the phenotype under different climatic/year conditions and the usage of standardized protocols. Trials in several geographically distant fields and/or trials across different years facilitate the identification of a trait variance explained by the genotype x environment (GxE) and improve the accuracy of the prediction models for GS and markers identified by GWAS or QTL mapping (El-Soda et al., 2014). Standardized protocols, on the one hand, facilitate the collection and reusing of the phenotype data and on the other hand improve the exchange of the data among the plant breeding community (Pieruschka & Schurr, 2019).

An important procedure related to phenotyping is the collection of relevant environmental data, which mostly includes temperature, solar radiation, and

precipitation over a specific vegetation period. Recent trends in this direction lead to the development of so-called environmentomics which in addition to climatic data assumes the soil composition and structure. Microbiome data was also suggested to be included especially in case conventional and organic production is compared (Oyserman et al., 2021). All the above-mentioned information could be used to improve the accuracy of the yield and phenotype prediction models, especially for crops grown in contrast environmental conditions (Rogers et al., 2021).

The final step before the genetic association analysis is genotyping: detecting genetic polymorphisms in studied populations. The genotyping of the plants for scanning for the markers started in the 80s with low-throughput RFLP markers systems (Tanksley et al., 1989) and nowadays is mostly based on more occurring throughout the genome marker, namely SSRs and SNPs. Starting from the 2000s, following the deciphering of the plant genomes, DNA microarrays started to be used widely in order to genotype plant species including major crops like maize, rice and soybean (Rasheed et al., 2017).

The advantage of the DNA microarrays is related to their standard design and their ability to capture a significant amount of genetic diversity existing in specie compared to fragment-based polymorphisms analysis since the arrays are done within the consortium projects that usually analyze several germplasm collections of different origins (Morales et al., 2020; Unterseer et al., 2014). The power of the DNA-microarrays is restricted to the number of species for which such technology is available and the relatively high cost of the array's design and production for a small number of samples (Christiansen et al., 2021). In addition to DNA microarrays, high-throughput sequencing technologies are currently gaining popularity for plant genotyping due to their rapid development and constant reducing DNA sequencing cost (Egan et al., 2012). Although whole-genome sequencing data gives the opportunity to obtain complete information about causative and linked polymorphisms, it remains expensive to be obtained and computationally heavy to analyze. This, lead to the development of reduced genotyping approaches like

restriction-site associated DNA sequencing (RAD-seq), specific-locus amplified fragment sequencing (SLAF-seq), genotyping-by-sequencing (GBS) that are based on the sequencing of the restricted DNA-fragments (Baird et al., 2008; Elshire et al., 2011; X. Sun et al., 2013). These techniques allow for reducing the plant genome complexity overwhelmed with repetitive elements, as well as reducing genotyping costs and time for analysis.

Different types of examined population designs require different types of analysis to identify genetic markers. For bi-parental populations (BC, F2, RILs, DHs), the construction of the genetic maps based on the recombination fractions (RF) and LOD scores calculated between the pairs of the markers is required in order to reconstruct the true order of each marker in each linkage group (Zheng et al., 2019). Next, several approaches could be applied to scan for the associations, including single marker analysis which considers each marker as an independent observation, simple interval mapping (SIM), and more recently commonly used composite interval mapping (CIM) scans markers in the genetic intervals of linked markers and controls the effects of the QTL on other intervals or chromosomes on the tested QTL (Broman et al., 2003). For diversity panels, the genome-wide association studies approach (GWAS) is used (Brachi et al., 2011). Initially, GWAS was based on the single marker ANOVA that tests whether the marker is associated with the trait of interest (Bush & Moore, 2012). Such a naive approach does not take into account relatedness among individuals which may bias the results, since polymorphisms present at a higher frequency in a certain population could be falsely associated with the phenotype of interest of this population happened to be present at a higher frequency in a group expressing this phenotype (Tibbs Cortes et al., 2021). More recent GWAS methods are based on the linear models among which the mixed linear model (MLM) is widely used since it allows to account for the relatedness at two levels: kinship (*K*) and population structure (*Q*) (Yu et al., 2006). Kinship matrix (*K*) is used to control the random effects of genetic background, while the Q matrix explaining sample stratification, usually assessed by the principal component analysis or genetic

clustering algorithms implemented in software like ADMIXTURE or STRUCTURE, helps to control fixed effects of the population structure (Tibbs Cortes et al., 2021). In addition to the MLM, more than twenty additional approaches that are also based on the mixed models were developed in recent years mainly with the aim of reducing the computational demand and increasing the statistical power. Among them, SUPER (Settlement of MLM Under Progressively Exclusive Relationship), FaST-LMM (Factored Spectrally Transformed Linear Mixed Models) and FarmCPU (Fixed and random model Circulating Probability Unification) remain popular for trait mapping in plants (P. K. Gupta et al., 2019). Besides the family of linear model-based methods that often assume on and/or few markers at the time, Bayesian methods that use prior information on marker effects and phenotypes to estimate all marker effects simultaneously started to gain popularity. The main advantage of these methods is that the power of detecting associations is not inversely correlated to the number of polymorphisms as in the case of the application of multiple testing correction applied for linear model-based methods (Fernando & Garrick, 2013).

Regarding the MPPs several approaches could be highlighted. On the one hand, it is possible to apply linear models, as in the case of GWAS, for the diversity panels (Gangurde et al., 2020; Islam et al., 2016; Nice et al., 2016). On the other hand, since the parental genotypes are by default accessible for the NAM and MAGIC populations it is possible to construct a consensus genetic map for the studied population based on the information on recombination events (Zheng et al., 2019) and apply interval mapping approaches used for bi-parental populations that will increase the accuracy of the mapping (Bu et al., 2021).

After finding candidate genetic markers , they undergo validation based on the plant material used for breeding (Xu & Crouch, 2008). For this purpose, local genotyping tests, such as allele-specific PCR assays followed by the PCR product detection in an agarose gel, are commonly used. Then the predicted effects of these putative markers are compared to the test sample phenotype (L. Li et al., 2013; W.-C. Zhou et al., 2003). Additionally, such validation could be performed across different

environments and populations in order to assess the stability of the genetic markers (Knoll & Ejeta, 2008).

The next step after the genetic markers associated with the traits of interest have been verified based on the plant material used for the breeding and/or independent plant samples, is the development of a genotyping solution for routine material assessment. Here several technologies could be applied for the assessment of mono- and oligo-/poly-genic traits. Allele detection assays based on allele-specific PCR, for example SNP type (Fluidigm), KASPar (KBiosciences) and TaqMan (Life Technologies) could be used to predict the presence or absence of mono-/oligo-genic traits (Broccanello et al., 2018; Seeb et al., 2011). Notably, all the above-mentioned technologies require lab equipment. This obstacle is circumvented by the rapid development of isothermal PCR, like LAMP (loop-mediated isothermal amplification), which enables rapid phenotype prediction in the field (Prasannakumar et al., 2021; D. Zhou et al., 2016). For oligo-/poly-genic traits, alternative solutions like MALDI-based genotyping and low-density SNP-arrays are commonly applied (Pattemore et al., 2010; X.-L. Wu et al., 2016). The rapid development of the NGS further led to the emergence of the low-density NGS-based panels, for example, GT-seq (Genotyping-in-Thousands) technology, allowing multiplexing of sequence-specific regions carrying polymorphisms of interest (Campbell et al., 2015).

It could be concluded that scanning for the genetic markers requires a complex methodological pipeline including the choice of a suitable population design for the trait mapping, followed by a selection of optimal phenotyping and genotyping approaches. The collected data is then processed with an aim of scanning for candidate markers. Markers are then subjected to testing in different environmental and/or genetic backgrounds. All these steps lead to the creation of the marker panels, which are then used to reduce the financial and time costs for breeders in the context of marker-assisted breeding.

2.3 Sunflower and rapeseed: past and recent breeding trends

Rapeseed and sunflower are among the top three oilseed crops grown worldwide (Attia et al., 2021; Vinnichek et al., 2019). These crops are grown mainly with the purpose of vegetable oil production. Besides vegetable oil, rapeseed and sunflower are used to produce raw materials for biofuels and the chemical industry (Embaye et al., 2018). Additionally, byproducts of oil production, namely, rapeseed oil cakes and sunflower meals are used to feed the animals (de Oliveira Filho & Egea, 2021; Walker & Booth, 2001).

Rapeseed

Rapeseed (*Brassica napus*) as a plant species appeared due to a spontaneous hybridization between two other *Brassica* species – cabbage (*B. oleraceae*) and turnip (*B. rapa*). According to phylogenetic analysis this event happened from 7500 to 12,500 years ago (Chalhoub et al., 2014) resulting in amphidiploid *B. napus* genome (2n = 38) with the size of ~ 900 Mb that consists of A (2n = 20) and C (2n = 18) subgenomes of *B. rapa* and *B. oleraceae*, respectively. Domestication of rapeseed began in South and Southeast Asia around 6000 years ago (Snowdon et al., 2007). In Europe rapeseed was documented to be cultivated around 400 years ago (Gomez-Campo, 1999). Importantly, there are three ecotypes of rapeseeds: winter and semi-winter (lacking the frost hardiness) that are sown in the late autumn, and spring ecotype which is shown in early spring. Such plasticity has contributed to the widespread rapeseed cultivation across the world from Canada to Australia (D. Wei et al., 2017).

Significant changes in rapeseed breeding appeared throughout the second half of the 20th century. In the 1970s, the first source of cytoplasmic male sterility (CMS) was discovered, making hybrid breeding for this species possible. However, the stable CMS source was transferred later from radish (*Raphanus sativus*) to rapeseed (Rousselle & Bregeon, 1982). Heterosis was reported for rapeseed only in the late 1980s (Brandle & McVetty, 1990). Next, substantial progress was made by Canadian breeders in the late 1970s – their efforts led to the development of new cultivars with

low erucic acid and glucosinolates content in the oil (Mag, 1983). Lately, these traits were introduced in European germplasm (Przybylski et al., 2005). These modifications are important because high erucic content leads to myocardial lipidosis – increased triacylglycerol accumulation affecting the contractility of the heart (J. Alexander et al., 2016). In turn, glucosinolates negatively affect the liver and kidney in animals (Walker & Booth, 2001) and contribute to bitter taste across the *Brassicacea* family (Wieczorek et al., 2018). In addition to these traits, many efforts were made toward the selection of the high-oleic varieties of rapeseed in the 1990s, since oleic acid is considered optimal both for cooking and biofuel production (Spasibionek et al., 2020).

A more recent direction in rapeseed breeding is disease resistance. Three diseases mostly damaging rapeseed include stem rot (*Sclerotinia sclerotiorum*), blackleg disease (*Leptosphaeria maculans*) and clubroot disease (*Plasmodiopora brassicae*). For clubroot and blackleg there are qualitative disease resistance genes (R genes), thus the breeding for resistance mostly consists of scanning of the R gene in diversity panels and introduction to germplasm (Ton et al., 2020). For stem rot the disease resistance is quantitative and thus the breeding for resistance is more challenging (Z. Wang et al., 2019). Other important directions in rapeseed breeding related to the harvest preservation and effectiveness of harvesting include improvement of lodging resistance and pod shattering. Lodging resistance affects the uneven distribution of the light and the effectiveness of mechanical harvesting (C. N. Miller et al., 2018) while pod shattering is associated with yield loss during vegetation and harvesting as well (Kuai et al., 2016). As climate change significantly affects the sowing area for rapeseed (Jaime et al., 2018), several attempts and suggestions were proposed to diminish negative climatic effects. One of the breeding targets is the reduction of the flowering time as the rapeseed is sensitive to drought while transiting from the flowering stage to silique development (M. Zhu et al., 2016). Additionally, breeding for drought tolerance is also needed to reduce yield loss under non-optimal conditions (Khanzada et al., 2020).

Sunflower

Originally sunflower was domesticated by native North Americans around 4000 years ago (Crites, 1993). Cultivated sunflower is a diploid species (2n = 34) with a haploid genome size of 3-3.5 Gb (Badouin et al., 2017). As an oilseed crop, it was cultivated from the beginning of the nineteenth century in Russia. Significant progress in sunflower breeding was made by Soviet academician V. S. Pustovoit who developed open-pollinated varieties with increased oil content in the late 1950s (Terzić et al., 2020). After the discovery of a source for cytoplasmic male sterility (CMS) and the corresponding fertility restorer gene (Leclercq, 1971), sunflower hybrid breeding became widespread starting in the late 1970s (Dimitrijevic & Horn, 2018). Further, in the late 1990s, genes for imidazolinone and sulfonylurea herbicide tolerance were identified in the wild populations of sunflowers (Al-Khatib et al., 1998).

To date several trends in sunflower breeding could be highlighted. The first one is resistance to pathogens, including those that cause sunflower rust, downy mildew, sclerotinia white mold, etc. (Jocic´ et al., 2015; Rauf et al., 2020; Škorić, 2016). The high interest for Eastern Europe and Russian breeders is represented by breeding for resistance to the broomrape (*Orobanche cumana*) – a parasitic plant that evolves rapidly and significantly affects yield (Lukomets et al., 2021; Škorić, 2016). Another important direction of sunflower breeding that is specific to Eastern Europe is breeding for large-seeded varieties (Hladni & Miladinović, 2019; Lukomets et al., 2021). Such varieties are called confectionary and are used to make edible seeds for direct consumption or to produce snacks (Hladni & Miladinović, 2019). The third large direction of breeding focuses on the oil content, composition and quality. The total oil content was increased by V.S. Pustovoit from ~25% to ~50% of seed mass. Now, modern breeding programs, especially the ones relying on hybrid breeding, are aimed to maintain and improve this complex trait using genomic prediction models (Mangin et al., 2017). For frying oils the priority is given to the high oleic acid oils, since this monounsaturated fatty acid increases oil resistance to thermal oxidation. Another direction of oil content selection aims to create high stearic acid sunflower

for margarine and vegetable ghee manufacturing (Serrano-Vega et al., 2005). Additionally, the tocopherol composition of the seeds is of high interest for oil quality improvement, since tocopherols demonstrate vitamin E activity and increase the resistance of oil to oxidation, especially in combination with high oleic acid content (Rauf et al., 2017).

A novel trend in sunflower breeding is drought resistance, as current climate changes may significantly affect future sunflower productivity. Although sunflower is a low-to-medium drought sensitivity crop, increased temperatures reduce the total oil content, total yield, and protein level (Hussain et al., 2008; Rondanini et al., 2003). Thus, breeding efforts are currently made, including the consortium level such as SUNRISE (SUNflower Resources to Improve yield Stability in a changing Environment), with the main goal of making varieties that demonstrate high yield reproducibility across different environmental conditions (Debaeke et al., 2017).

Thus the breeding of both crops was aimed toward the improvement of the oil composition and yield quality for food and industry needs. As recent climate changes started to affect the yield, new breeding targets include resistance to unfavorable environmental conditions. Additionally, climatic changes accelerate pathogen evolution creating the need to select variants resistant to emerging pathogens.

2.4 Genetics of agronomically important sunflower traits

Being an industrially important species, agricultural and wild sunflower varieties are well characterized in terms of genetic diversity. The first genetic map of the sunflower was developed in the early '90s (Berry et al., 1995; Gentzbittel et al., 1995; Rieseberg et al., 1993) and its genome was sequenced using several approaches (Badouin et al., 2017). Accordingly, genotyping DNA microarrays were developed for this species. All of the above-mentioned features make a sunflower an appropriate object for marker scanning and subsequent application of marker-assisted selection approaches (Jocković et al., 2021). As a result, many important traits have already been dissected in terms of genetics for sunflowers.

To date, researchers uncovered more than 20 sources of cytoplasmic male sterility in *Helianthus* genus by experimental crosses (Terzić et al., 2020). However, only for two cytoplasms CMS-PET1 and CMS-PET2 derived from *H. petiolaris* the respective restorer genes have been identified (Z. Liu et al., 2012; Sajer et al., 2020, p. 2). For CMS-PET1 several germplasms were successfully scanned with an aim of finding the restorer genes (*Rf*) for CMS-PET1. It was demonstrated that loci associated with *Rf* genes mapped to chromosomes 2, 6, 12, 13, and 7 (Ma et al., 2021). Recently, *Rf-PET2* restoring CMS-PET2 was mapped on chromosome 13 (Sajer et al., 2020). For sunflower, as for other crops, these loci were demonstrated to carry genes encoding proteins carrying pentatricopeptide repeat (PPR) motifs (Goryunov et al., 2019; Horn et al., 2019; Ma et al., 2021; Sajer et al., 2020).

Weeds affect sunflower yield, thus the application of herbicides is economically important. For sunflower, two types of herbicides, namely sulfonylureas (SU) and imidazolinones (IMI) are applied. Both target acetohydroxyacid synthase key enzymes for branched amino acid biosynthesis encoded by *Ahsl* genes (Kaya et al., 2012). Herbicide-resistant plants were found among wild populations of the sunflower (Al-Khatib et al., 1998; White et al., 2002). Additionally, IMI-resistant plants were obtained based on induced mutagenesis (Sala et al., 2008). Using mapping populations three herbicide-resistance loci and their corresponding genetic markers (SSRs, CAPS and SNPs) were identified and validated: *AhasIl-1* (chromosome 9) determining moderate tolerance against both herbicides; *AhasIl-2* (chromosome 6) conferring tolerance to SU; and *AhasIl-3* (chromosome 2) determining resistance to IMI (Bulos, 2013; Kolkman et al., 2004).

There are more than 20 pathogens harming sunflower crops. Among them, the top three diseases leading to significant yield losses are caused by *Plasmopara halstedii* (downy mildew), *Orobanche cumana* (parasitic plant called broomrape), *Sclerotinia sclerotiorum* causing white mold (Rauf et al., 2020; Virányi, 2008). More than 30 resistance genes to *P. halstedii* damaging to sunflower leaves have been identified in sunflower and in its wild relatives (Qi et al., 2019). However, only

several of them were mapped genetically to loci on chromosomes 1, 2, 4, 8, 13 and the corresponding genetic markers were identified (Dimitrijevic & Horn, 2018). Broomrape is a parasitic plant that feeds on the sunflower's root system leading to the reduction of leaf and flower head size. Broomrape is the most harmful sunflower pathogen in Eastern Europe (Škorić et al., 2021). To date, eight races (A-H) of *O. cumana* have been identified. Several genes of vertical resistance to broomrape have been identified and mapped as well. The most recently identified resistance gene (and corresponding SNP marker) was localized to chromosome 4 for race G by analyzing an experimental cross of resistant and susceptible sunflower lines (Fernández-Aparicio et al., 2022; Martín-Sanz et al., 2020). Similarly, several studies reported lines carrying resistance to the race H. However, the genetic markers associated with this trait were not identified. Additionally, new potential broomrape races that overcome resistance to race H are emerging (Škorić et al., 2021). It should be noted that in addition to vertical resistance genes several QTLs that control sunflower resistance at different developmental stage of the broomrape were also identified (Louarn et al., 2016). Unlike the *P. halstedii* and *O. cumana,* resistance to *S. sclerotiorum*, a pathogen damaging all parts of the plants including the flower head, demonstrates mostly polygenic inheritance with narrow-sense heritability of up to 60% (Zubrzycki et al., 2017). Using F2 mapping populations over 15 potential QTLs associated with resistance to *S. sclerotiorum* were identified in sunflower (Micic et al., 2004). Of them, three SSR markers were validated using independent populations (Micic et al., 2005). In further studies, the utilization of the high-density SNP arrays could improve the accuracy of the *S. sclerotiorum* resistance predictions leading to a more rapid emergence of pathogen-resistant variants by genomic selection (Livaja et al., 2016; Talukder et al., 2021).

Sunflower total oil content ranges from 40 to 50% of dry seed mass in industrial varieties. Further, it was demonstrated that despite the complex genetic control of this trait, the estimated heritability was more than 80 percent (Bachlava et al., 2010). Several studies identified more than five QTLs controlling total oil content

explaining up to 51% of the trait variance (Bachlava et al., 2010; Merah et al., 2012). Application of the large-scale genotyping allowed to apply genomic selection approaches and to develop models for oil content prediction in hybrids, demonstrating high prediction accuracy of up to 78% (Mangin et al., 2017).

Genetic control of oleic acid content is one of the best-studied features, mainly due to the importance of this fatty acid to the food industry. Initially, it was assumed that the high oleic acid content is fully controlled by the dominant *Ol* gene located on chromosome 14 that encodes fatty acid desaturase FAD2 – an enzyme converting oleic acid to linoleic acid and explaining up to 86% of phenotypic variance (Lacombe & Bervillé, 2001). The first genetic markers associated with *Ol* were RAPD and RFLP were found by analyzing experimental crosses (BC and F2) (Dehmer & Friedt, 1998; Lacombe & Bervillé, 2001). Additional studies performed on the contrast F2 populations revealed additional minor effect loci associated with oleic acid content located on chromosomes 6, 8, and 9 and explaining up to 12% of the phenotypic variance (Premnath et al., 2016; F. Zhou et al., 2018). Besides oleic acid, several studies have been performed with an aim of mapping loci associated with other important fatty acids. For palmitic acid and stearic acids, which are of high importance for lubrication production and deep-fat frying, genetic markers were identified by experimental crosses analysis. QTL analysis revealed RFLP markers associated with the major effect locus controlling stearic acid content on chromosome 1 carrying a gene encoding stearoyl-ACP desaturase and explaining up to 80% of phenotypic variance (Pérez-Vich et al., 2002). For palmitic acid locus affecting its content was mapped on chromosome 9 by analyzing the F2 population derived from crossing conventional and high-palmitic lines. Markers of these loci were located close to gene encoding 3-ketoacyl-ACP synthase II associated with high levels of palmitic acid (Pérez-Vich et al., 2016). A recent GWAS study performed in our lab further helped to identify potential genetic markers (SNPs) associated with 12 fatty acids (Chernova et al., 2021).

Genes controlling tocopherol composition were mapped in sunflower variants by analyzing experimental crosses (BC, F2). It was demonstrated that this trait is controlled by two recessive/dominant genes *Tph1* and *Tph2*, located on chromosomes 1 and 8 encoding 2-methyl-6-phytylbenzoquinol methyltransferase and γ-tocopherol methyltransferase explaining up to 97% of phenotypic variance (Tang, Hass, et al., 2006; Vera-Ruiz et al., 2006). However, several studies indicate that this trait, besides the two major effect genes, might include additional minor non-allelic genes located on chromosomes 4, 9, 14, and 16 (García-Moreno et al., 2012a; Hass et al., 2006). Thus, despite the suggested mono-/dihybridism inheritance of oleic acid content and tocopherol composition, additional loci were reported for both traits' indicating the importance of the genetic background for traits being mainly under mono-/oligo genic control (García-Moreno et al., 2012a; Hass et al., 2006).

As climate change remains a serious challenge for plant breeding, more studies focus on QTLs of drought and high salinity resistance. Fourteen QTLs that determine high salinity resistance were identified by analyzing experimental crosses of wild sunflowers (Lexer et al., 2003). In cultivated sunflower variants, six candidate loci were linked with salt tolerance: the loci encoding genes involved in the mineral ion uptake (Lai et al., 2005). Several studies were performed with an aim of detecting QTLs associated with the agronomically important traits under different water regimes by analyzing experimental crosses (Poormohammad Kiani et al., 2007, 2008, 2009). As a result, a QTL associated with yield stability across water-stressed and standard conditions was identified (Poormohammad Kiani et al., 2009). Additionally, water use efficiency was assessed under drought conditions, and respective QTLs were identified on chromosomes 13, 15, and 9 cumulatively explaining up to 41% of trait variance (Adiredjo et al., 2014). A recent GWAS study identified five genetic markers associated with the thousand-kernel-weight traits, an important economic parameter, under drought response (Gosseau et al., 2019).

In conclusion, sunflower genomics includes numerous successful efforts leading to the dissection of genetic features underlying agronomically important

traits. These traits include herbicide and pathogen resistance, hybrid breeding oil quality and climate adaptations. Such progress not only improved our understanding of the genetic control mechanisms of various agronomic traits but also led to the identification of the corresponding genetic markers. This in turn has formed a substantial basis for the development of the marker-assisted selection approaches for sunflower improvement.

2.5 Genetics of agronomically important rapeseed traits

Similar to sunflower, rapeseed is well studied in terms of genetic diversity as it is the second-largest oilseed crop in the world. The first genetic maps for this crop were constructed in the mid-1990s based on the RFLP and RAPD markers (Ferreira et al., 1994; Parkin et al., 1995; Uzunova et al., 1995). However, the rapeseed genome was first sequenced and assembled relatively recently (Chalhoub et al., 2014). As in the case of sunflower, this knowledge facilitates genetic mapping and the development of markers for agronomically important traits in rapeseed.

There are six CMS systems identified in rapeseed, of which *ogu* and *pol* carrying lines are the most popular ones used to produce hybrid seeds (Anisimova & Dubovskaya, 2020). Thus, fertility restorer genes, as well as corresponding markers, were developed to facilitate hybrid breeding. For *ogu* fertility restorer gene *Rfo* introduced from radish (*Raphanus sativus*) RAPD markers were developed (Delourme et al., 1994). Further, a fine-mapping placed this locus to chromosome 19 and allowed to identify of the candidate gene encoding pentatricopeptide repeat (PPR) protein family member (X. Hu et al., 2008). For *pol* CMS derived from cultivar Polima, the monogenic dominant restorer gene Rfp was mapped on chromosome 18, by using RFLP-, RAPD-based genetic maps (Jean et al., 1997). A recent study focusing on fine mapping of *Rfp* by analyzing a large NIL population revealed a candidate gene encoding the PPR family protein (Z. Liu et al., 2012).

For oil and oilcake quality improvement, the genetic associations found two traits that significantly reduce the oil quality, namely glucosinolate content and erucic

acid. QTL mapping of these traits allowed to identify QTLs located on chromosomes 9 (A09), 12 (C02), 17 (C07), and 19 (C09) were considered as major effect loci controlling glucosinolate content (Howell et al., 2003; Uzunova et al., 1995; J. Zhao & Meng, 2003). Recent GWAS studies confirmed loci located on chromosomes 2 (A02), 9 (A09) and 12 (C02) explaining up to 40% of phenotypic variance (Harper et al., 2012; F. Li et al., 2014; D. Wei et al., 2019). The genes located in these regions included ones homologous to the *HAG1* of *Arabidopsis thaliana* encoding transcription factor involved in the glucosinolate biosynthesis regulation (Gigolashvili et al., 2007; F. Li et al., 2014). In addition to major effect loci, several minor effect loci carrying genes involved in glucosinolate metabolism have also been mapped (Gubaev et al., 2020; Kittipol et al., 2019; Y. Liu et al., 2020; D. Wei et al., 2019), thus making it possible the development of marker-assisted selection towards reduction of glucosinolate content. For erucic acid, QTL studies identified two loci controlling its content (Ecke et al., 1995; Jourdren et al., 1996; Thormann et al., n.d.). Recent GWAS studies demonstrated that *BnaA.FAE1* and *BnaC.FAE1* associated with the erucic acid content are located on chromosomes 8 (A08) and 13 (C03) and explain up to 25% of phenotypic variance (F. Li et al., 2014; C. Qu et al., 2017; D. Wei et al., 2019; Q. Zhu et al., 2019). These genes encode fatty acid elongases and were shown to be involved in the biosynthesis of erucic acid in rapeseed (N. Wang et al., 2008; G. Wu et al., 2008). Additional minor effect loci were also identified (Havlickova et al., 2018; F. Li et al., 2014; C. Qu et al., 2017; C.-M. Qu et al., 2015; D. Wei et al., 2019; Q. Zhu et al., 2019) indicating that such a complex trait should be controlled by the implementation of genomic selection approaches in the breeding program.

Several studies further focused on the mapping of other oil-related traits – fatty acid content and tocopherol composition. As in the case of the sunflower, the main direction of the selection effort is focused on reaching high content of oleic acid and improvement of tocopherol composition. QTL mapping revealed a major effect locus controlling oleic acid content in rapeseed on chromosome 5 (A5), explaining up to

76% of the phenotypic variance. This locus was shown to carry a gene encoding fatty acid desaturase (X. Hu et al., 2006; Schierholt et al., 2000). More recent GWAS studies confirmed the role of this fatty acid desaturase in oleic acid level control (Gacek et al., 2017) and allowed to identify additional loci carrying genes encoding acyl carrier protein and 3-ketoacyl-CoA synthase (Guan et al., 2019; Q. Zhu et al., 2019). For tocopherol composition mapping, one of the first studies identified two loci controlling tocopherol content explaining 43% of the phenotypic variance. Similarly, for tocopherol composition (α-/γ-tocopherol ratio), additional two loci were identified explaining 35% of phenotype variance (Marwede et al., 2005). Association mapping studies further identified candidate genes *VTE3* and *PDS1*, encoding MPBQ methyltransferase and hydroxyphenylpyruvate dioxygenase involved in the biosynthesis of tocopherols (Fritsche et al., 2012; X. Wang et al., 2012). Recent associative transcriptomic studies that facilitate identification of molecular markers linked to trait-controlling loci further supported the role of *VTE4* encoding γ‑tocopherol methyltransferase and controlling alpha-/-gamma-tocopherol ratio (Havlickova et al., 2018).

As the diseases significantly affect yield preservation and productivity, many attempts have been made towards resistance genes and loci associated with quantitative disease resistance (QDR) identification (Neik et al., 2017; Ton et al., 2020). For clubroot disease caused by soil-borne protist *Plasmodiopora brassicae* few R genes were identified in *B. napus* by means of analyzing mapping populations derived from crosses between susceptible and resistant lines (Diederichsen et al., 2006; Fredua-Agyeman & Rahman, 2016; Manzanares-Dauleux et al., 2000). Notably, it was demonstrated that the clubroot resistance controlled by R genes could be overcome by pathogens (LeBoldus et al., 2012) thus several studies focused on discovering QDR loci were made. A total of 19 QDR loci have been identified by means of QTL mapping in rapeseed, explaining up to 80 percent of phenotypic variance (Werner et al., 2007). Recent studies performed by means of association mapping in diversity panels allowed the identification of more than 30 additional

QDR loci (Hejna et al., 2019; L. Li et al., 2016). For parasitic fungus *Leptosphaeria maculans* causing blackleg disease at least seven R genes have been mapped on two chromosomes 7 (A7) and 10 (A10) by analyzing experimental crosses derived from R gene carriers (Ansan-Melayah et al., 1998; Balesdent et al., 2001; Rimmer, 2006). As the qualitative resistance is overcome by the pathogen several recent studies have been supporting the fact that the QDR loci increase the effect of the R genes (Brun et al., 2010; Delourme et al., 2014). Thus more recent studies are focused on the mapping of QDR loci. As a result, more than 20 QDR loci have been identified by analyzing genetically contrasting collections and contrasting DH lines of rapeseed (Fu et al., 2020; Huang et al., 2016; R. Raman et al., 2020). Unlike for the *L. maculans* and *P. brassicae,* the identification of R genes for parasitic fungus *Sclerotinia sclerotiorum* causing soft rot remains challenging due to the medium heritability of the *B. napus* resistance. As a result several attempts have been made to find QDR loci through assessment of genetic contrasting collections, DH mapping populations and integrative analysis based on previous studies (J. Li et al., 2015; L. Wei et al., 2016; J. Wu et al., 2013, 2016). Identified markers explained up to 32% of phenotypic variance (J. Wu et al., 2013) candidate genes were involved in oxidative stress response, secondary metabolites production, hormonal pathways functioning and pathogen recognition (J. Li et al., 2015; L. Wei et al., 2016; J. Wu et al., 2013, 2016). A very recent GWAS study identified 123 loci associated with rapeseed resistance to *S. sclerotiorum*. Additionally, genomic prediction model was developed with a predictive ability of 64% which in turn could be used for genomic selection for *S. sclerotiorum* resistance in rapeseed (Roy et al., 2021).

Crop preservation during the vegetation period for rapeseed is determined by pod shattering and resistance to lodging. For pod-shattering only recent studies based on the analysis of diversity panels and artificial mapping populations helped to identify candidate loci. Pod shattering demonstrates high broad-sense heritability (J. Liu et al., 2016). As a result, a major effect locus carrying *psr1* gene explaining 47% of the phenotypic variance was mapped on chromosome A09 (Z. Hu et al., 2012). An

additional GWAS study identified 12 additional QTL that jointly explained approximately 57% of the genotypic variation of pod shattering (H. Raman et al., 2014). Although lodging leads to a significant yield loss of up to 46% (Kendall et al., 2017), studies dissecting resistance to lodging appeared recently. By analyzing the diversity panel by means of an associative transcriptomic approach three genomic regions (explaining up to 30% of phenotypic variance) associated with the stem strength were identified. Candidate genes located in these regions were associated with plant cell wall biosynthesis and regulation of hormonal pathways (C. N. Miller et al., 2018). Another GWAS study revealed five loci associated with lodging, however, the proportion of explained variance ranged from 8 to 10 percent (L. Wei et al., 2017).

Rapeseed is significantly affected by the non-optimal water and temperature regimes, including drought, salinity and cold, therefore the breeding for abiotic stress tolerance remains crucial under a rapidly changing climate (Ahmar et al., 2019). Thus several recent studies aimed at identifying genetic markers for abiotic stress resistance have been made. A DH mapping population was used to map waterlogging and drought tolerance. As a result, 26 QTL and 31 QTL were identified explaining up to 23 % of phenotypic variance (Z. Li et al., 2014). Two studies based on GWAS identified novel loci and genes that determine drought and salt tolerance. Twenty-five QTLs explained up to 9.23% of salt tolerance and carried genes encoding aquaporins acting as water channels, transcription factors involved in abiotic stress response, proline biosynthesis enzymes (Wan et al., 2017). For drought stress tolerance two loci located on chromosomes 10 (A10) and 19 (C9) were identified; four genes encoding PPD5 regulating ROS levels and stomata closure as well as ribosomal protein (RP), multidrug and toxic compound extrusion (MATE) protein, and a histone superfamily protein previously shown to regulate abiotic stress responses were suggested as candidate ones controlling the trait (Shahzad et al., 2021).

Summarizing the information reviewed in the present chapter it should be concluded that the recent progress in genotyping and data analysis approaches

facilitated the dissection of the agronomically important traits in rapeseed. Significant progress was made both for monogenic traits, such as qualitative disease resistance and ability to restore fertility, and for polygenic traits including fatty acid composition, glucosinolate content and resistance to unfavorable environmental conditions. Identification of trait-associated genomic loci, in turn, facilitated the understanding of the genetic control of the traits including the identification of novel candidate genes. More practically, these studies revealed prospective genetic markers for marker-assisted and genomic selection in rapeseed.

2.6 Russian trends in rapeseed and sunflower breeding

Sunflower and rapeseed sowing areas rank first and third, respectively, among oilseed crops in Russia. It should be noted that for rapeseed and sunflower, up to 53 and 73 percent of planting material currently used in Russia is of foreign origin, which is explained by the insufficient development of local breeding companies and institutions. The leading Russian institute for sunflower breeding and the major governmental originator is VNIIMK (Pustovoit All-Russia Research Institute of Oil Crops). Besides VNIIMK institute that produces novel varieties of sunflower is the Altai Research Institute of Agriculture and the Federal Center of Agriculture Research of the Southeast Region. Additionally for sunflower, private breeding companies including "Agroplasma", "Galaktika" and "Triumph" made a substantial contribution to new varieties' development. For rapeseed, the leading variety breeder is the All-Russian research institute of rapeseed, in addition VNIIMK and Siberian Federal Scientific Center for Agrobiotechnologies. Unlike sunflower, rapeseed breeding in Russia is currently supported by a single private breeding company "Astra" (Ivanov et al., 2020).

For sunflower, there are several breeding trends currently addressed in Russia. One of the top problems that face sunflower breeders is related to the resistance to the broomrape. Active attempts are made for pyramidisation of the resistance from race A to recent race G (Demurin et al., 2020; Ryzhenko et al., 2021). In the area of hybrid

breeding that requires herbicide resistance and involvement of CMS/Rf systems significant progress was made in the identification and introduction of herbicide resistance and Rf genes (Anisimova et al., 2021; Chelyustnikova et al., 2017; Demurin, Borisenko, et al., 2006).

Successful attempts were made toward the development of hybrids with improved oil quality, basically the high-oleic sunflowers with altered tocopherol composition (Demurin, Borisenko, et al., 2006; Demurin et al., 2016). Another important direction in sunflower breeding is the creation of the confectionary sunflower which is generally between edible and oil forms. It should be noted that confectionary sunflower breeding is mostly based on the populational varieties (Bochkovoy et al., 2020).

For rapeseed, the main breeding targets in Russia are oil quality, resistance to herbicides, yield preservation, and productivity. For oil quality, the low erucic and low glucosinolate varieties are developing in the VNIIMK institute. Additionally, chemical mutagenesis was applied to introduce a high oleic trait in Russian germplasm (Gorlova et al., 2020). It also should be noted that most of the Russian rapeseed registered varieties are represented by inbred lines, while foreign variety producers mostly rely on hybrid breeding (Anisimova & Dubovskaya, 2020). Resistance to lodging and pod shattering also remains important traits in Russia and are evaluated for new varieties of high interest for breeding (Bochkaryova, Strelnikov, et al., 2021). Further, VNIIMK recently started to initiate hybrid rapeseed breeding since the productivity of hybrids is higher compared to inbred lines (Bochkaryova et al., 2020). The first hybrid based on the *ogu* CMS system "Debut" was registered in 2020. Notably, the productivity of the hybrid was equaled to the ones obtained in the EU and significantly higher than the "Loris" standard variety (Bochkaryova, Gorlova, et al., 2021).

Despite the huge work on the development of the varieties that meet the industry's requirements, marker-assisted breeding is poorly used in scientific institutions and private breeding companies in Russia. Only a few studies have been

made for mapping agronomically important traits in rapeseed and sunflower in Russia (Usatov et al., 2014). Notably, a significant part of these studies was made in collaboration with Skoltech (Chernova et al., 2021; Goryunov et al., 2019; Goryunova et al., 2019; Gubaev et al., 2020, 2022). The present work is a part of an effort of introducing marker-assisted selection in Russia for rapeseed and sunflower improvement. This work is focused on the mapping of the seed glucosinolate content in rapeseed, tocopherol composition and oleic acid content as well as seed-related traits in sunflower.

Chapter 3. Materials and methods

3.1 Plant material

3.1.1 Rapeseed diversity panel

Rapeseed diversity panel consisted of 90 inbred lines from the All-Russian Research Institute of Oil Crops (VNIIMK) and included 43 winter and 47 spring accessions (Gubaev et al., 2020). The phenotype data on glucosinolate content was collected during three vegetation periods (2015–2016, 2016–2017, 2018–2019) at VNIIMK experimental station, GPS coordinates were 45°10′51″ N and 39°02′42″ E. Seeds of plants used for genotyping were collected during 2018–2019 vegetation period. The detailed information on accession number, country and collection of origin, eco-type, glucosinolate content is provided as previously published supplementary information and could be accessed via web-link https://www.mdpi.com/2073-4425/11/8/926/s1 (Table S1). The accessions, as well as more detailed information on phenotypes, can be obtained upon request from VNIIMK collection (https://en.vniimk.ru/science/scientific-departments/otdel-selektsii-maslichnykh-kultur/laboratoriya-selektsii-rapsa/).

3.1.2 Experimental crosses for oil-related traits mapping in sunflower

The plant material used to map tocopherol composition and oleic acid content consisted of genetically contrasting mutant lines VK195 and VK876 (Demurin et al., 2016) as well as wild-type VK303 and VK101 (Trembak et al., 2018). According to hybridological analysis lines VK195 and VK876, carry mutant alleles *tph1*, *tph2* and *Ol* and thus express γ- and δ- tocopherol phenotype (~50% of γ- and ~50% δ-tocopherols) as well as demonstrate high oleic phenotype (>82% of oleic acid). Lines VK303 and VK101 carry wild type *Tph1*, *Tph2* and *ol* and demonstrate a high α- tocopherol phenotype (100% of α-tocopherol) as well as moderate content of oleic acid (30 - 45%). All lines were developed at the VNIIMK institute. Elite lines VK101

and VK303 are used to produce a simple interlinear middle-early sunflower hybrid "Typhoon" registered at VNIIMK (Trembak et al., 2018). While lines VK876 and VK195 are used to produce an "Oxy" hybrid that synthesizes oil with increased oxidative stability (Demurin et al., 2016).

Lines VK195 and VK876 were used as female lines after hand emasculation. Lines VK101 and VK303 were used as a source of pollen. The crosses were performed in the field (Krasnodar) in 2015. Individual F1 plants were selfed using isolators to produce the F2 progeny (F2 seeds) in the field (Krasnodar) in 2018. Random inflorescence was selected from a plot. For subsequent phenotyping and genotyping procedures F2 seeds were randomly sampled per F1 sunflower head. In total 142 F2 seeds for cross VK195xVK303 and 144 F2 seeds for the cross VK876xVK101 were collected. Seeds were harvested at the physiological maturity stage (R-9) according to the sunflower growth stages classification (Schneiter & Miller, 1981). For phenotyping, a half-seed technique was applied (Y. N. Demurin et al., 2004; Vera-Ruiz et al., 2005). In particular, a single seed was cut in half with a razor, and the first part containing the embryo was left for germination for subsequent genotyping, while the second part was used for phenotyping. Parental lines are available at the VNIIMK genetic collection upon request.

3.1.3 Sunflower diversity panel

A panel of 601 sunflower accession was provided by: N.I. Vavilov Institute of Plant Genetic Resources (VIR, St. Petersburg, Russia) - 255 accessions, V.S. Pustovoit All-Russia Research Institute of Oilseed Crops (VNIIMK) - 199 accessions, and LLC Agroplasma Seed and Breeding Company (Krasnodar, Russia) - 147 accessions. All plants were grown for genotyping procedures in the Krasnodar region in Russia in 2017. All accession were diploid, $2n = 34$. Plants were grown in fields of leached black earth soil type. The sunflower was sown following the preceding crop, winter wheat, at the seeding rate of 40,000 plants per hectare. Sowing was carried out according to the following sowing system: 70×35 cm, 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. Accessions catalog names are provided in supplementary information published previously (Chernova et al., 2021) that could be accessed via web link https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-021-07768-y/MediaObjects/12864_2021_7768_MOESM14_ESM.xlsx. Brief description of the three collections is also could be accessed via web link https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-021-07768-y/MediaObjects/12864_2021_7768_MOESM23_ESM.docx. The accessions, as well as detailed information on phenotypes, can be obtained upon request from VIR seed bank (https://www.elibrary.ru/item.asp?id=32542976; http://www.vir.nw.ru/en/), VNIIMK collection (https://vniimk.ru/science/geneticheskaya-kollektsiya-podsolnechnika/) and LLC Agroplasma (https://agroplazma.com/contacts).

3.2 Phenotyping procedures

3.2.1 Glucosinolate content measurement in rapeseed

For glucosinolate measurement, a modified colorimetric analysis was applied (Thies, 1982). For each measurement, 5 g of seeds were collected from three plants of each line from a single plot during each of the three years. In total 180 plants representing 90 lines were phenotyped. Seeds were homogenized using a mechanical grinder, then 0.2 g of the homogenate was used for extraction according to the previously published protocol (Gubaev et al., 2020). The obtained extract was subjected to measurement of absorbance at 420 nm on a spectrophotometer UNICO 2800. The concentration was calculated based on the calibration curve obtained using the standard samples with known concentrations.

3.2.2 Tocopherol composition and oleic acid content measurement in sunflower

A total of 142 F2 seeds for cross VK195xVK303 and 144 F2 seeds for cross VK876xVK101 were assessed for a single plant randomly selected from the plot.

Each parental line was phenotyped in at least 7 replicates. Phenotype data were collected within one vegetation period in 2018.

To determine tocopherol composition, a thin-layer chromatography (TLC) with subsequent densitometry was applied. Extraction of the tocopherols was made according to the previously described protocol (Demurin, Efimenko, et al., 2006; Y. N. Demurin et al., 2004). For the detection of tocopherols, the plate was sprayed with a freshly prepared mixture of 0.1% ferric chloride ($FeCl_3$) and 0.25% α, α'-dipyridyl in absolute ethanol (Emmery-Engel's reagent), taken by volume in a ratio of 1:1. The stained plates were scanned and the relative composition in tocopherol forms was quantified by measuring the relative intensity of the color of tocopherol forms (in %) using a Sorbfil videodensitometer (software version 1.5.0).

To determine the fatty acid composition of the oil and relative content of oleic acid, a half seed was ground in a coffee grinder, thoroughly mixed, and a 0.5 g sample was taken for extraction in 4 ml of hexane. Then 2 ml of miscella were taken for methylation according to the standard method (Demurin, Efimenko, et al., 2006; Y. N. Demurin et al., 2004). Chromatographic analysis was performed on a Khromatek-Kristall 5000 gas chromatograph with an automatic dispenser DAZH-2M on a SolGelWax 30 m × 0.25 mm × 0.5 μm capillary column in a flow of helium carrier gas, at a speed of 25 cm/s, with temperature programming in within 185-230 °C. The preparation of methyl esters and their chromatography and subsequent data analysis were performed in accordance with standard methods (Demurin, Efimenko, et al., 2006; Y. N. Demurin et al., 2004).

3.2.3 Seed morphology traits assessment in sunflower

A total of 601 inbred lines were used to perform seed morphology traits assessment. For each line, 10 seeds from a seed bank that were stored for 1-3 years were used. To estimate the seed and husk area, seeds were placed on the plate. For each line, 10 seeds were analyzed per plate. The next plate was scanned by means of PRDU-02 mobile X-ray diagnostic installation jointly developed by Agrophysical

Research Institute and St. Petersburg State Electrotechnical University LETI (Arkhipov et al., 2019). Image analysis was performed by SeedRentgen software (Brach et al., 2015). The mean values of husk and seed area were calculated for each of the plates (accession), given the are of one pixel of each photo corresponded to 100 μm x 100 μm (0.1 mm^2) of the plate area. Next, a ratio between seed area and husk area was calculated based on the rounded values for subsequent analysis. To assess the effect of the collection on the husk and seed area as well as seed to husk ratio analysis of variance was performed. To assess the correlation coefficient between husk and seed area Pearson correlation was estimated. Mann–Whitney U test was performed to assess the difference in seed area and husk area and seed to husk ratio between collections.

3.3 Genotyping procedures and SNP calling

3.3.1 Rapeseed genotyping and SNP calling

For DNA extraction 1–2 leaves of seedlings germinated within two weeks under room conditions were used. Three independent plants of each line were used for genotyping procedures. DNA was extracted according to the CTAB protocol using the NucleoSpin Plant II plant DNA extraction kit (Macherey-Nagel, Dylan, Germany). Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and gel electrophoresis were used to estimate the quality and concentration of the DNA. Extracted DNA was used to prepare the genotyping-by-sequencing (GBS) library as previously described (Chernova et al., 2021; Goryunov et al., 2019). PstI and MspI restriction enzymes were used to cut DNA. Sequencing was performed in the HiSeq 4000 (Illumina, San Diego, CA, USA). Raw sequence data are available on NCBI SRA under the project number PRJNA645178.

To find SNPs raw reads were first demultiplexed according to the barcodes which identify sample ID using Axe-demultiplexer software version 0.3.3 (Murray & Borevitz, 2018). Trimmomatic software version 0.36 was used to perform read filtering according to the following parameters: low-quality bases (Phred score < 20)

as well as sequences corresponding to Illumina adapters and reads of less than 40 bp were removed. Filtered reads were mapped on the most recent version of the reference genome of Brassica napus cultivar 'ZS11' (F. Sun et al., 2017) using Burrows-Wheeler aligner (H. Li & Durbin, 2009) version 0.7.17 with default parameters. For SNP calling, a modified version of the pipeline (https://github.com/RimGubaev/GATK_pipeline_customized) based on Genome Analysis Toolkit software version 4.1 (McKenna et al., 2010) was applied. The biallelic SNPs that passed the following filters were used for subsequent analysis: Minor allele frequency (MAF) > 0.03 and maximum missing data of 30%, sequencing depth was of at least 10 reads per position. Additionally, SNPs were filtered by maximum heterozygosity of 40%, as it was previously discussed fact of the high amount of heterozygous positions leads to read misalignment due to the amphidiploid nature of the rapeseed genome which results in the presence of highly homologous regions (Malmberg et al., 2018).

3.3.2 Sunflower genotyping and SNP calling

To perform genotyping of the sunflower plants, the half-seeds left after cotyledon separation were germinated in rolls of filter paper. DNA was extracted from the cotyledon leaves using the NucleoSpin Plant II kit (Macherey-Nagel) according to the manufacturer's protocol. Modified protocol (Elshire et al., 2011) was used to prepare a genotyping-by-sequencing (GBS) library as previously described (Chernova et al., 2021; Goryunov et al., 2019). Illumina HiSeq 4000 was used to sequence prepared libraries of single-end reads with a length of 150bp. For F2 crosses, parental plants were genotyped in at least seven biological replicates. Raw sequence data for F2 cross and parental lines are available on NCBI SRA under the project number PRJNA742188.

For SNP calling TASSEL-GBS pipeline v2 was used. Read mapping was performed using bowtie2 (Langmead & Salzberg, 2012) aligner with –very-sensitive parameter. The most recent assembly XRQ2.0

(https://www.heliagene.org/HanXRQr2.0-SUNRISE/) of the sunflower genome was used as a reference (Badouin et al., 2017) genome. Prior to mapping, reads derived from each parental line's replicate were merged. Next for genetic map construction only SNPs that were homozygous within each parent and polymorphic between the parents were selected. SNPs possessing less than 10% or more than 90% of heterozygosity level as well as ones that were genotyped in less than 20% of accessions, were discarded before the imputation procedure as previously recommended (Fragoso et al., 2016).

To perform SNP calling for diversity collection of the sunflower, a modified GATK pipeline was applied. Variants were filtered using hard filtering parameters: Map quality (MQ) > 36, QD > 24, and MQRankSum < 2, ensuring that reads were mapped to a unique place in the reference, 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). To select SNPs for population structure and association mapping analyses of seed-related traits missing calls rate < 0.3, variant depth (DP) > 4 and minor allele frequency (MAF) > 0.01 were applied, resulting in 15,068 SNPs.

3.4 Population structure analysis

To identify the number of subpopulations in sunflower and rapeseed datasets ADMIXTURE software version 1.3 (D. H. Alexander et al., 2009) was applied with the number of clusters ranging from 1 to 10. The PLINK software version 1.9 (Purcell et al., 2007) was used to perform principal component analysis with standard parameters, the number of principal components was set to 10, PCs were estimated using the variance-standardized genetic relationship matrix. To estimate linkage disequilibrium r^2 was calculated among SNP pairs located within a 1500 kb window using PLINK software version 1.9.

For diversity comparative analysis of rapeseed WGS data for 54 rapeseed accession from European, Chinese, Australian, Japanese, Moroccan, and Indian

collections (Malmberg et al., 2018). Lines were of different ecotypes (winter, semi-winter, and spring), SNP calling of raw WGS data was performed as described above (see chapter 3.3.1). To compare the genetic diversity of the sunflower, a raw vcf file included 1065 wild, 20 landraces, and 289 cultivated accessions (Hübner et al., 2019). Bcftools software version 1.9 (H. Li, 2011) was applied to find common SNPs among the datasets between the studied cohorts and the international accessions of rapeseed and sunflower.

3.4 Association mapping and candidate gene identification

3.4.1 Association mapping of glucosinolate content in rapeseed

Compressed mixed linear model (cMLM) implemented in TASSEL software version 5.2 (Bradbury et al., 2007) were used to find genetic associations to glucosinolate content. Kinship matrix calculated with centered IBS (Identity by state) method as well as first five principal components were included in the model to account for kin relationships and population structure, respectively. Phenotype data for each year were analyzed independently. To find the most reproducible SNPs across three years, the sum of the p-values was substituted by the corresponding densities of the Irwin-Hall distributions taking into account that the sum of p-values was formed by three uniform distributions. This was done using Unifed R package's Dirwin.hall function (Quijano Xacur, 2019). The adjusted p-value with Bonferroni correction was set to 0.0000041 (0.05/12,226, where 12,226 is the number of SNPs) to consider associations. An additional softer threshold was set to 0.0005. The scanning for potential candidate genes was performed using SnpEff software version 4.3 (Cingolani et al., 2012). A frame of 100 kb upstream and downstream of the significant SNPs was used, as such a threshold was previously used for scanning candidate genes in rapeseed (Guan et al., 2019; Q. Zhou et al., 2018; Q. Zhu et al., 2019).

To compare previously identified loci with ones found in the present study information on the exact physical location of the SNPs associated with glucosinolate

content was collected from previously published data (Jan et al., 2019; F. Li et al., 2014; S. Liu et al., 2020; C.-M. Qu et al., 2015; D. Wei et al., 2019). Next for each of the SNPs identified in the present study, corresponding positions were identified in the *Brassica napus's* Darmor bzh assembly (Chalhoub et al., 2014) as it was used in the previous studies. To do so the 50bp upstream and downstream flanking sequences of the SNPs were extracted and mapped to the Darmor bzh assembly version 4.1 using Blast software version 2.2.18 (Camacho et al., 2009). The following filters were used to identify the SNPs common for two assemblies: e-value < 1×10^{-40}. alignment length > 90. The frame of 100 kb was used to find the genetic regions in common with the previous studies.

3.4.2 Association mapping of the seed morphology traits in sunflower

To perform association mapping of the seed-related traits a compressed mixed linear model (cMLM) was implemented in TASSEL software version 5.2 (Bradbury et al., 2007). Kinship matrix calculated with centered IBS (Identity by state) method as well as first five principal components were included in the model to account for kin relationships and population structure, respectively. Additionally, the collection of origin was added to the model as a covariate as it was demonstrated that it is a factor significantly affecting seed and husk area as well as seed to husk ratio (ANOVA p-value << 0.05). The adjusted p-value with Bonferroni correction was set to 0.0000033 (0.05/15,058, where 15,058 is the number of SNPs) to consider associations. An additional softer threshold was set to 0.0001. The scanning for potential candidate genes was performed using SnpEff software version 4.3 (Cingolani et al., 2012). A frame of 100 kb upstream and downstream of the significant SNPs was used as on average SNP pairs located within this frame demonstrated a high level (r2 > 0.5) of linkage (for details see Chapter 6.3, Figure 6.3.1).

3.5 Genetic map construction

Before genetic map construction, missing genotypes were imputed with default parameters using the LB-impute pipeline (Fragoso et al., 2016). Imputed SNP genotypes were encoded as the A, B, and H, corresponding to maternal, paternal and heterozygous genotypes, respectively using a custom R script (https://github.com/RimGubaev/vcf_to_qtl). SNPs genotyped in less than 80% of individuals as well as ones with a significant segregation distortion (Bonferroni-corrected $\chi 2$ goodness-of-fit test P-value < 0.05) were discarded. Additionally, Individuals genotyped by less than 80% of SNPs were removed. As a result, 136 and 142 individuals were used to construct genetic maps for crosses VK195xVK303 and VK876xVK101, respectively. To construct a genetic map R/qtl software (Broman et al., 2003) was applied. Data on physical chromosomes available for SNPs was used to preorder the loci within chromosomes. Kosambi mapping function was implemented to order markers within chromosomes. Additional filtering for SNPs within each chromosome was performed using the droponemarker() function; markers with LOD scores equal to or above −20 were discarded. To estimate the concordance between physical and genetic marker order the Pearson correlation coefficient and the corresponding significance level were calculated using cor() and cor.test() functions, respectively, in the R base package between physical and genetic distances.

3.6 QTL mapping of tocopherol composition and oleic acid content

Quantitative and qualitative approaches were used to map genetic markers of tocopherol composition and oleic acid content. For quantitative mapping of the proportion of each of the tocopherol classes and oleic acid content were considered as independent observations. Analysis of variance (ANOVA) implemented in R's base statistical package was used to calculate the proportion of variance explained by genetic markers. For these traits, the interval mapping approach for non-normally distributed traits based on Haley–Knott regression was applied using the scanone()

function from r/qtl package. Additionally, composite interval mapping was used to map oleic acid content. For qualitative mapping of *Tph1* loci the phenotypes of plants that belong to the α (*Tph1*/_; *Tph2*/_) and γ (*Tph1*/_; *tph2*/*tph2*) classes were set to 1, while the α/β (*tph1*/*tph1*; *Tph2*/_) and γ/δ (*tph1*/*tph1*; *tph2*/*tph2*) classes were set to 0. For qualitative mapping of the *Tph2* the phenotypes of plants that belong to the α (*Tph*1/_; *Tph2*/_) and α/β (*tph1*/*tph1*; *Tph2*/_) classes were set to 1, while the β (*Tph1*/_; *tph2*/*tph2*) and γ/δ (*tph1*/*tph1*; *tph2*/*tph2*) classes were set to 0.

The phenotypes of plants that did not belong to any of the classes based on the tocopherol composition were set to NA. For qualitative mapping of the *Ol* gene phenotypes of plants that demonstrated oleic acid content equal to or higher than 83% were set to 1 while those that demonstrated less than 83% were set to 0. To scan for loci associated with *Tph1*, *Tph2* and *Ol,* the interval mapping approach for binary traits based on Haley–Knott regression implemented in the r/qtl's scanone() function was applied. Permutation analysis for the logarithm of odds (LOD) carried with 1,000 iterations was performed to identify a significance threshold corresponding to the 99 percentile of the permuted LOD values. To calculate 1.5-LOD confidence intervals for most significant markers lodint() function in r/qtl was applied.

3.7 Verification of the genetic markers for tocopherol composition and oleic acid content

To validate genotypes of polymorphic SNPs obtained with GBS, significant SNPs associated with *Tph1*- and *Tph2*-determined phenotypes were selected. For primer design, 500bp upstream and downstream sequences were extracted for SNPs within 1.5-LOD interval bedtools v2.27.1 (Quinlan & Hall, 2010). Primer3 plus (Untergasser et al., 2012) was used to design primers for amplification of the regions and subsequent Sanger sequencing. To test primers for dimer formation and self-complementarity OligoCalc tool was applied (Kibbe, 2007). PrimerBlast software (Ye et al., 2012) implemented in the PlantEnsemble genome browser was used to test the specificity of the primers within sunflower genome version XRQ2.0.

Two criteria were used to select markers, the first is the high LOD scores within a 1.5-LOD confidence interval, second is the ability to construct unique primer pairs matching unique regions. As a result, four pairs for amplification were constructed (Table 3.7.1).

Table 3.7.1 – Primers used to amplify loci carrying significant SNPs associated with *Tph1*- and *Tph2*-associated genotypes.

Primer name	Sequence 5'-3'	Tm used for amplification	Product Size, bp
S1_55196434_Tph1_VK195xVK303_F	ACATGGTTTCTTATCATTTGCAC	59.0 C	420
S1_55196434_Tph1_VK195xVK303_R	ACCGGATATTTGACAAAGTGC		
S8_30578572_Tph2_VK195xVK303_F	TGACTTACTTGGTCGAGCCG	60.5 C	416
S8_30578572_Tph2_VK195xVK303_R	GCACGTACCCGATTTCCTTG		
S1_71748138_Tph1_VK876xVK101_F	ATCCTCCACAACCCAACACG	60.0 C	459
S1_71748138_Tph1_VK876xVK101_R	GAAAGCATACTTTGGGCGACT		
S8_23941299_Tph2_VK876xVK101_F	TCTCGGATTACAGTGGTTCGA	59.5 C	287
S8_23941299_Tph2_VK876xVK101_R	GAAAACGATGGGGTTCTGG		

To validate markers 34 F2 plants from the mapping population eight plants for each of the following groups were selected: those carrying the paternal genotype, carrying the maternal genotype and plants carrying heterozygous genotype according to GBS data. Five parental plants were genotyped to validate paternal and maternal genotypes. Direct Sanger sequencing of the amplicons was performed, and Ugene software (Okonechnikov et al., 2012) was applied to analyze sequencing results. To test the concordance of the genotypes obtained by two different methods, we compared the GBS-derived genotypes of the markers with the ones obtained by Sanger sequencing and estimated the proportion of the matching genotypes.

In addition to the SNP variant validation, the ability to predict *Ol*-, *Tph1*- and *Tph2*-associated phenotypes was tested. To do so independent sample of 20 lines was genotyped. Among them, eight were high oleic (>83%), and 12 represented wild type, regarding tocopherol composition all four classes were represented by five lines. Each line was sequenced in five replicates. SNP calling was performed jointly with each of the crosses in order to capture the genetic polymorphisms common between the test sample and each of the crosses. SNP calling was performed as described in chapter 3.3.2. Next, 10 most significant markers were selected for each trait/cross combination based on the QTL-mapping results. For each marker, a proportion of the genotypes in the test sample was calculated as well as the proportion of the correctly predicted phenotypes associated with *Ol*, *Tph1* and *Tph2* was estimated.

Chapter 4. Genetic characterization of Russian rapeseed collection and association mapping of glucosinolate content

4.1 Introduction

In the present part of the study, a genotyping of the rapeseed collection from the VNIIMK was performed with the aim of genetic characterization of the collection and the mapping of the agronomically important trait – glucosinolate content. To do so 90 genetically and phenotypically contrasted rapeseed accessions consisting of 47 spring and 43 winter accession were analyzed.

Glucosinolates represent secondary metabolites that on the one hand act as part of the defense against plant pathogens and on the other hand negatively affect the higher animals when consumed as a part of the diet, additionally glucosinolates contribute to the bitter taste of Brassicas. Thus the reduction of the glucosinolates is one of the key breeding targets in rapeseed production (Walker & Booth, 2001; Wieczorek et al., 2018). As it was the first study that included the genotype data of the diversity panel of Russian rapeseed, a population structure was described for the studied cohort. Additionally, we compared the genetic diversity of the studied cohort with the available whole-genome sequencing data available at the end of 2018 (Malmberg et al., 2018). The phenotyping procedures were performed during the three vegetation periods (2015-2016, 2016-2017, 2018). The collected data was used for the association mapping and identification of loci responsible for the glucosinolate content as well as for scanning for the potential candidate genes. Additionally, we performed the comparison of the identified loci with previously identified ones; this comparison showed that the loci identified in the present study are new. The results presented in the following chapter were published in the Genes journal (Gubaev et al., 2020).

4.2 Genetic characterization of VNIIMK rapeseed collection

To describe the genetic diversity of the studied cohort, the GBS approach was applied to characterize 90 rapeseed accessions from the VNIIMK collection. Each accession was sequenced in three replicates. As a result, we identified 160257 raw biallelic SNPs. After the filtering procedures, 12226 SNPs were left for the subsequent analysis, which is approximately 7-8 times lower than it was previously made by using similar reduced representation sequencing approaches (Gacek et al., 2017; Lees et al., 2016). Which on the one hand is explained by the relatively strict filters used to subset the genotypes, and on the other could be the result of a relatively small sample size capturing a smaller amount of genetic diversity.

Collected genotype data was used to estimate the number of the potential subpopulations, to do so the ADMIXTURE algorithm was applied (Figure 4.2.1). A significant drop in the cross-validation error indicating the number of potential subpopulations was observed while dividing a studied cohort into two, three, five and seven clusters (K) (Figure 4.2.1 A).

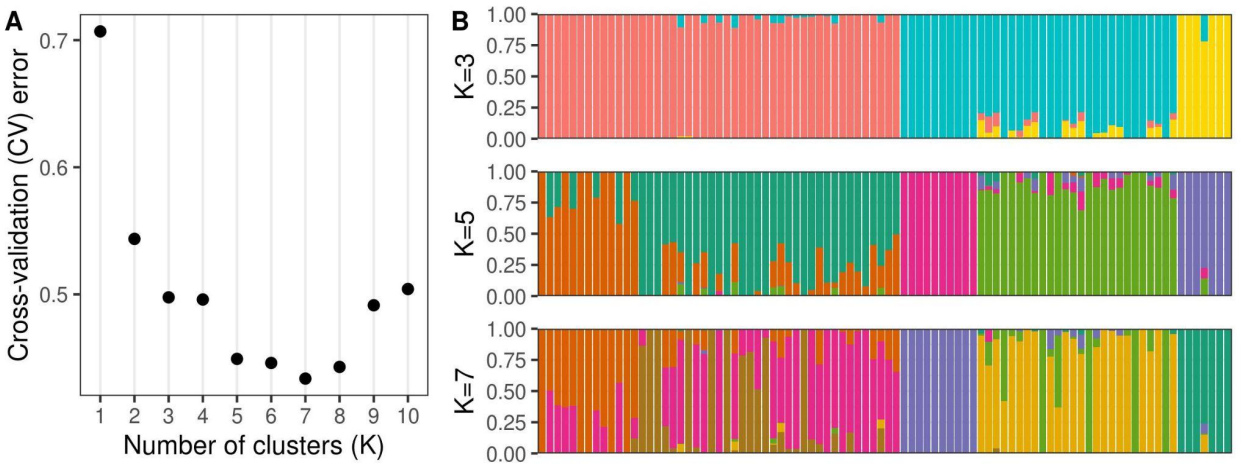

Figure 4.2.1 – Results of the population clustering assessment by the ADMIXTURE software. (A) Cross-validation error value for K ranging from 1 to 10. (B) ADMIXTURE bar plots correspond to subpopulations at K = 3, 5, 7, each bar corresponds to the rapeseed line; colors indicate the genetic admixtures for each of the studied lines. Red, blue and yellow bars reflect genetic admixture specific to spring, winter and yellow-seeded winter rapeseed accessions at K = 3.

At K = 2 the clusters were represented by the winter and spring ecotypes. A significant genetic difference between the spring and winter rapeseed accessions was expected and previously described as the major factor explaining the genetic structure since these two ecotypes demonstrate different growth habits as the winter rapeseed is sown in late autumn and requires vernalization while the spring ecotype is sown in the early spring (Gazave et al., 2016). Such segregation is explained by selection which led to the limiting of the rapeseed diversity and resulted in the separation of these ecotypes (Gazave et al., 2016; D. Wei et al., 2017; D. Wu et al., 2019). The principal component analysis confirmed this observation as the first principal component explaining 34.7% of genotype variance and also divided the studied sample into winter and spring ecotypes (Figure 4.2.2 A).

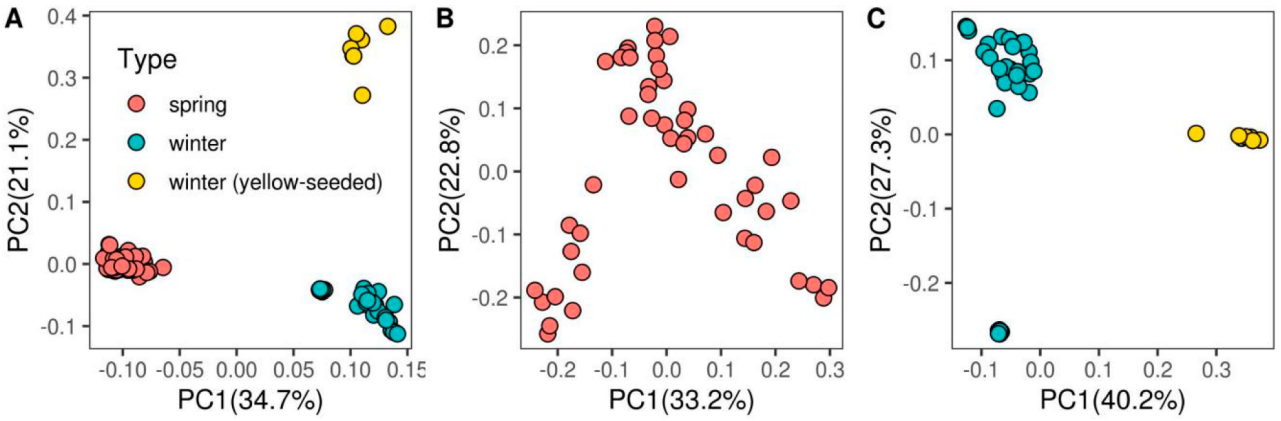

Figure 4.2.2 – Principal component analysis of the population structure of Russian rapeseed lines. PCA analysis was performed for the whole cohort (A) and separately for spring (B), and winter (C). Red, blue and yellow colors correspond to spring, winter and yellow-seeded winter rapeseed accessions, respectively.

A second significant drop in the cross-validation error was observed at K = 3; such segregation allowed additional distinguishing between yellow- and dark-seeded winter lines (Figure 4.2.2 C). The principal component analysis also distinguished these groups by the second principal component explaining 21.1% genotype variance. Such segregation by the seed color was quite a new observation, as the color of the seed coat is controlled only by several loci in *Brassicas* (HAWK, 1982; H. Zhao et

al., 2019). This presumably could be explained by the recent breeding history of yellow-seeded accessions in VNIIMK i.e yellow- and dark-seeded accessions were used to produce independent genetic lineages without mixing. A separate principal component analysis of the winter accession revealed three clusters. The first PC explaining 40.2% of genotype variance is expectedly distinguished between yellow- and dark-seeded accessions. While the second PC explaining 27.3% of genotype variance additionally divided dark-seeded winter rapeseeds into two subpopulations, there was no explanation for this according to the breeders' data. Although the substantial drop in CV was also observed at K = 5 and 7, such segregation did not correspond to any biochemical or phenotype data of VNIIMK breeders. The principal component analysis also did not reveal any clear clustering while analyzing the spring accessions separately (Figure 4.2.2 B).

In the next step, we estimated the linkage disequilibrium (LD) decay based on the squared correlation coefficient (r^2) calculated among SNPs located within 1500kb windows for the whole genome (Figure 4.2.3 A). The proportion of SNP pairs with the r^2 value greater than 0.25 within 30 kb bins was also estimated (Figure 4.2.3 B). Additionally, LD was calculated separately for the A subgenome derived from a turnip (*B. rapa*) and cabbage (*B.oleracea*) (Figure 4.2.3 C).

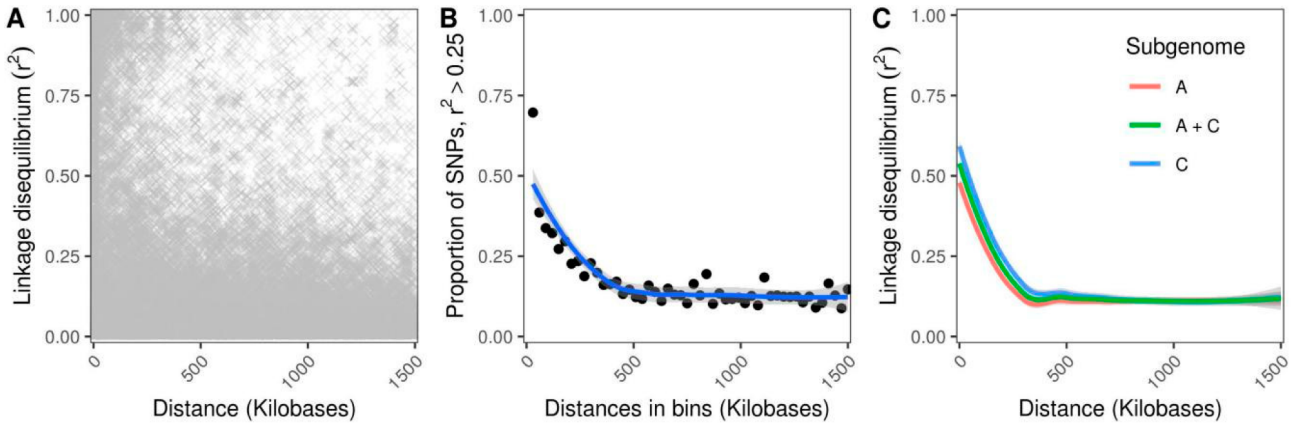

Figure 4.2.3 – Linkage disequilibrium (LD) decay in the studied rapeseed lines. (A) LD decay across genetic marker (SNP) pairs. Each cross corresponds to the r2 value between a pair of SNPs within 1500kb. (B) The proportion of SNP pairs with r*2* > 0.25 was calculated for the whole genome. Each dot indicates the proportion of SNP

pairs in the 30 kb bin. (C) LD decay for A and C subgenomes. Colored lines on panels B and C represent loess curves. Gray markers correspond to the 95% confidence interval.

For the entire genome at $r^2 = 0.25$ size of the LD blocks equaled 218.8 ± 14.5 kilobases (kb). Our findings demonstrated that at linkage cutoff of $r^2 = 0.25$ LD decay was longer for the C subgenome and equaled 265.1 ± 13.4 kb and shorter for the A subgenome equalling 177.1 ± 19.5 kb. Such difference in the LD decay between A and C subgenomes was repeatedly reported for the collections that did not include a substantial amount of accessions from Russian gene banks (Qian et al., 2014; D. Wei et al., 2017; Zheng et al., 2017). Such differences were explained by the longer chromosomes of the C subgenome compared to ones from the A subgenomes (F. Sun et al., 2017).

As the present study was the first to include a significant amount of the rapeseed accession from a Russian gene bank genotyped in a high-throughput manner we performed a comparison of the genetic diversity of studied accessions with international ones from other collections for which WGS data was available at the end of 2018 (Malmberg et al., 2018). As a result 54 geographically and ecotypically diverse rapeseed lines were selected for the comparison. For these 54 accessions WGS data was subjected to the same pipeline used for VNIIMK collection, resulting in 4,037,572 unfiltered SNPs which were joined with raw SNP data for VNIIMK collection. After filtering this joint dataset by the sequencing depth and missing values, common 20,848 SNPs were used to perform PCA analysis. The first and second principal components that jointly explained 38.7 % of genotype variance demonstrated that the international lines were clustered based on the collection of origin as well as the ecotype (Figure 4.2.4 A). This is in agreement with the previously published results where the collection of origin and ecotype served as the major factors affecting population structure in rapeseed (Bus et al., 2011; Gazave et al., 2016; D. Wei et al., 2017; D. Wu et al., 2019).

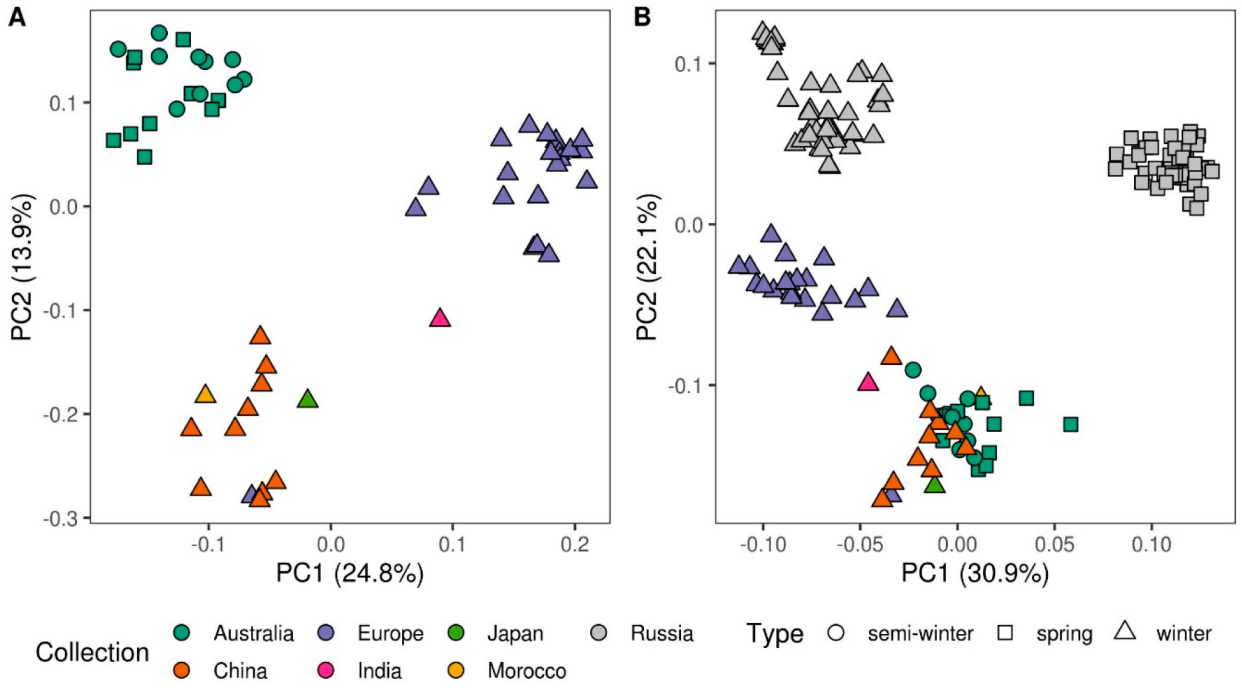

Figure 4.2.4 – Principal component analysis (PCA) plots reflecting the population structure of international and Russian collections. (A) The population structure of international lines is used in this study. (B) The comparison of the population structure of international and Russian lines. Shape corresponds to the ecotype, and colors represent the collection.

PCA analysis of the joint sample demonstrated the separation of the winter, spring and semi-winter accession by the first PC explaining 30.9% of genetic variance. The second PC explaining 22.1% genetic variance revealed clear differences between VNIIMK accessions and international ones (Figure 4.2.4 B). This in turn indicates a great potential for increasing the genetic diversity of collection for breeders and to find new genetic determinants for agronomically important traits. Additionally, this information could be used further for a deeper investigation of the rapeseed breeding history.

4.3 Measurement of the glucosinolate content

To perform the measurement of the glucosinolate, a modified chloropalade method followed by colorimetric analysis was applied at VNIIMK. The information on the glucosinolate content was collected for each of the 90 accessions across three years. Estimated glucosinolate content ranged from 11.0 to 35.8 micromoles per gram of fresh weight with an average value of 16.3 across three years. After that, a difference in glucosinolate content between winter and spring ecotypes was evaluated by ANOVA for three years as well as for mean values (Figure 4.3.1).

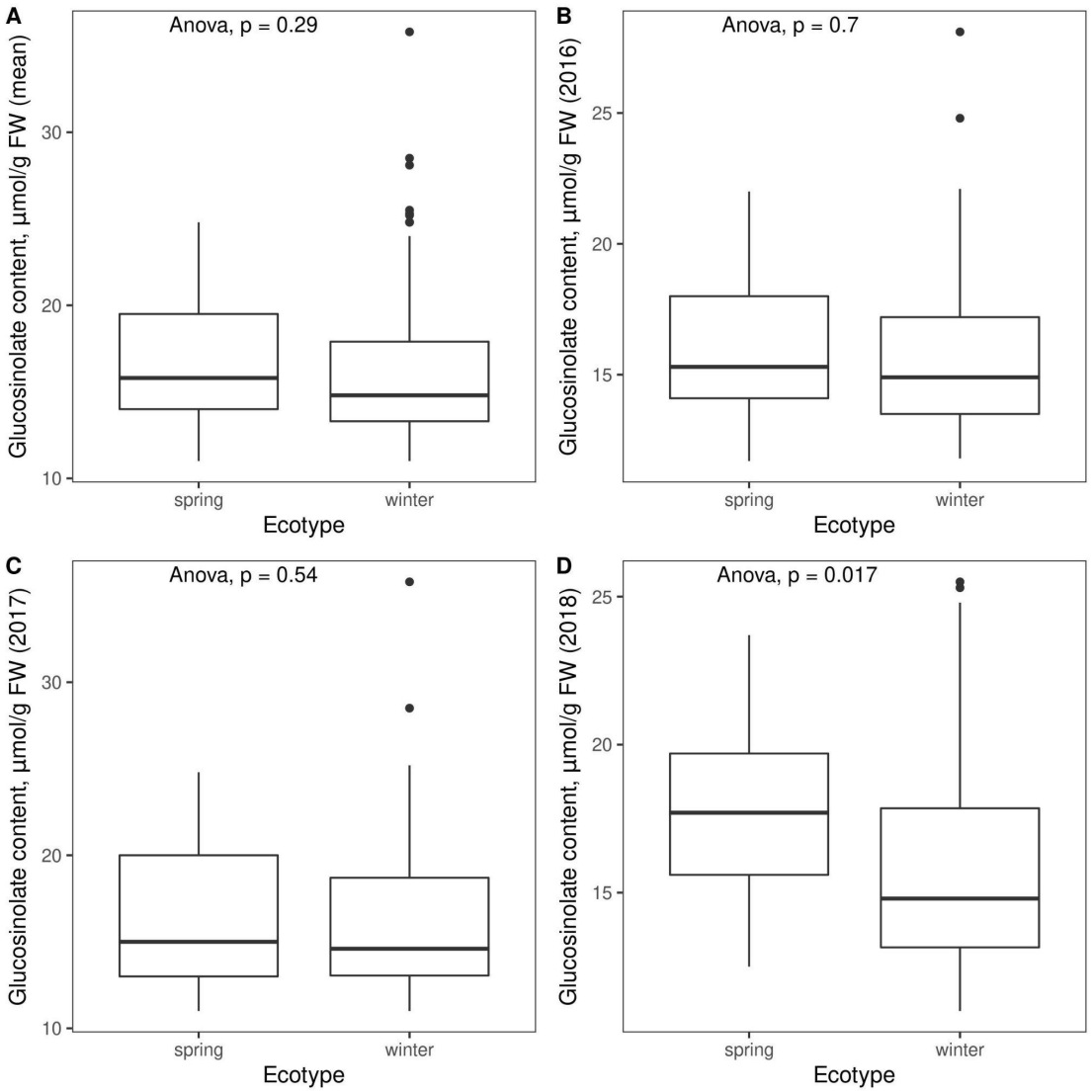

Figure 4.3.1 – Boxplots reflecting the distribution of the glucosinolate content across spring and winter accessions through 2016, 2017 and 2018.

This is due to the fact that it was previously shown that quantitative traits such as yield, oil content, and height significantly differ among rapeseed ecotypes (Assefa et al., 2018; Fridrihsone et al., 2020). It was demonstrated that no significant difference between the winter and spring phenotypes was identified at a 5% significance level for observations made in 2016 and 2017. Slight differences were only observed in 2018. Despite no strong climatic differences observed during the three vegetation seasons, this could be explained by the dryer conditions in November of 2017 (monthly cumulative precipitation = 49.9 mm) compared to the similar periods of 2015 (78.1 mm) and 2016 (93.8 mm). This probably could affect the glucosinolate accumulation in winter lines. Next, we tested the concordance of glucosinolate measurements across three years (Figure 4.3.2).

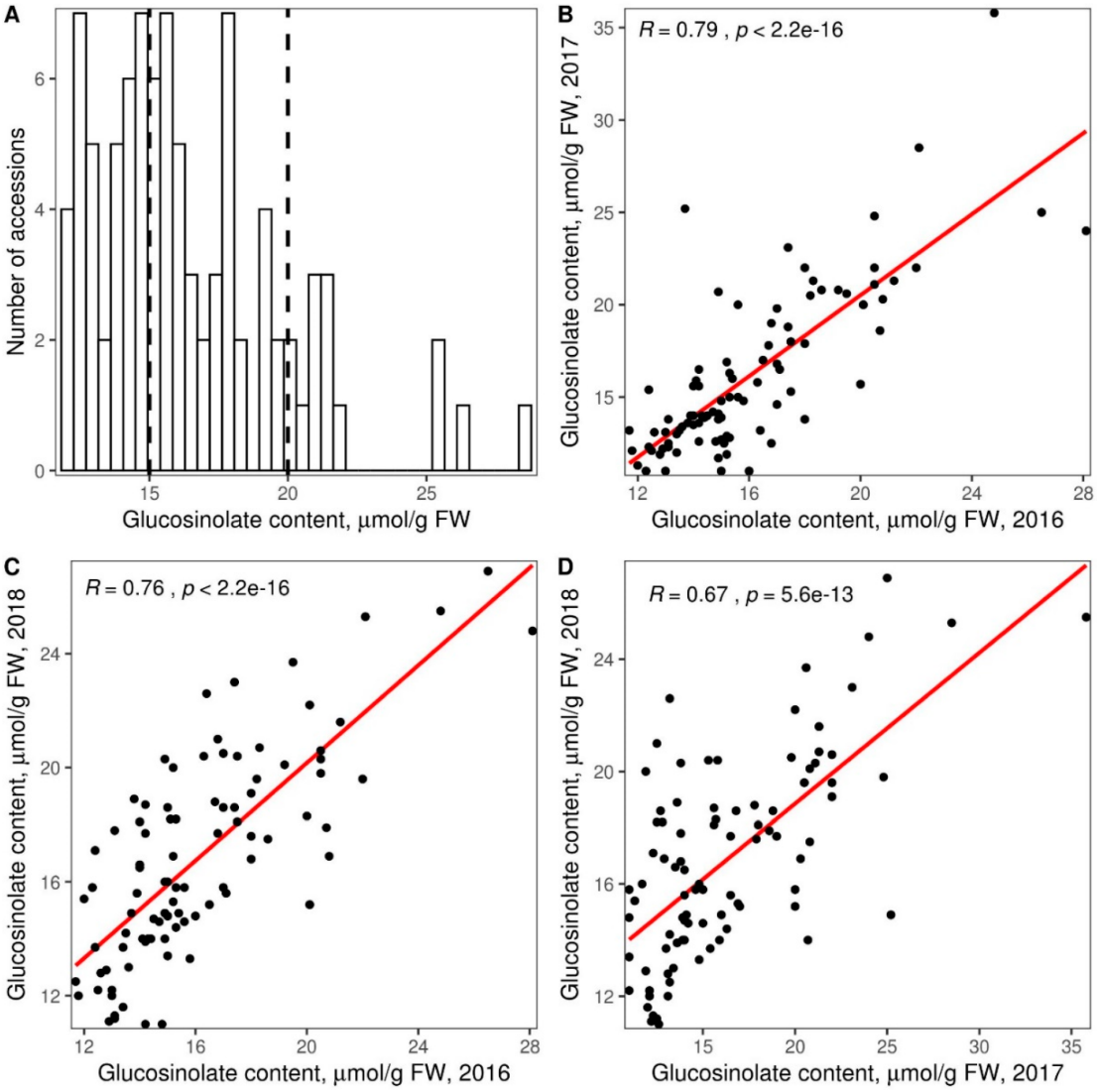

Figure 4.3.2 – Glucosinolate content distribution in the studied rapeseed cohort. Histogram depicting average glucosinolate content values for three years (A). Dashed lines divide low, middle, and high glucosinolate lines according to the (VNIIMK) classification. Correlation of glucosinolate content for three vegetational seasons (B–D). Each dot corresponds to a plant sample. Regression lines are shown in red.

The glucosinolate content correlated positively and significantly between years, with Pearson's correlation coefficient ranging from 0.67 to 0.79 (Pearson correlation, $p < 0.001$; Figure 4.3.2 A–C).

4.4 Association mapping and scanning for novel candidate genes

After the genotypes and phenotypes were collected, a scanning for genetic loci responsible for glucosinolate content was performed by using an MLM with additional PCs and kinship for fixed and random effects, respectively, to account for confounding derived from population structure and kin relationship. Since there were three years of observation, we applied an MLM for each year independently, after that a cumulative significance of genetic associations was calculated using Irwin-Hall distribution. After the application of the Bonferroni correction, two SNPs located on chromosome 7 (A7) remained above the threshold (Figure 4.4.1, Figure 4.4.2).

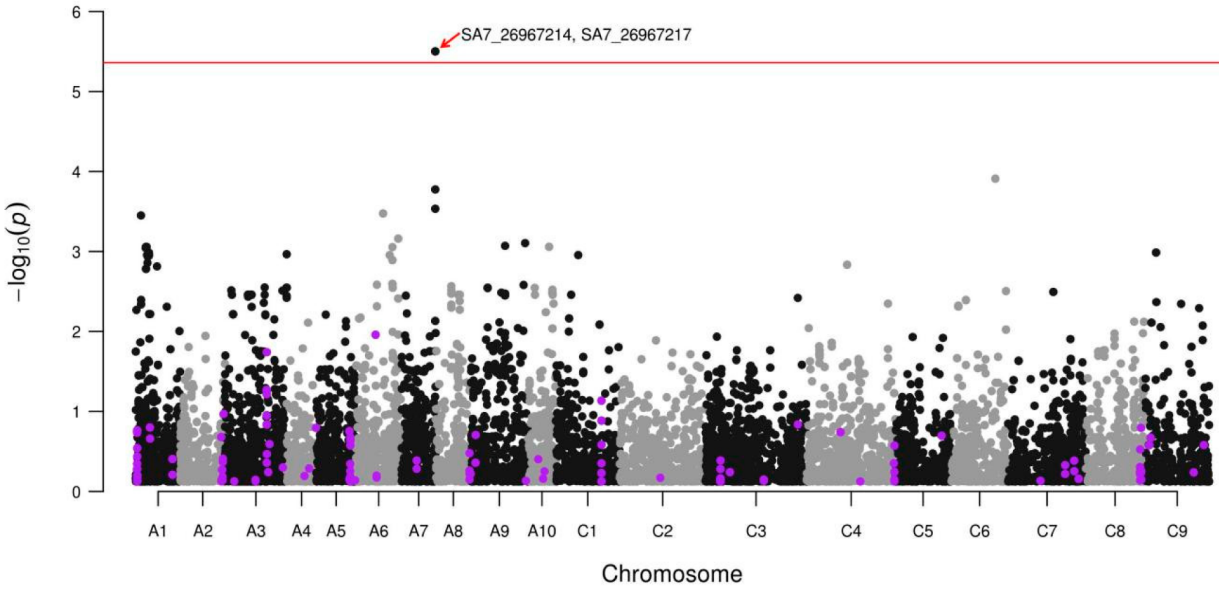

Figure 4.4.1 – Manhattan plot showing SNP markers associated with glucosinolate content. Each dot corresponds to a single SNP. Red line corresponds to the Bonferroni adjusted significance threshold. Purple dots correspond to SNPs demonstrated to be significantly associated with glucosinolate content in the previous studies.

These two SNPs (SA7_26967214 and SA7_26967217) explained from 13.8 to 20.4 % of phenotype variance across three-year observations (Table 4.4.1). Notably, both demonstrated stability across three years (ANOVA p-value for SNP:Year = 0.899).

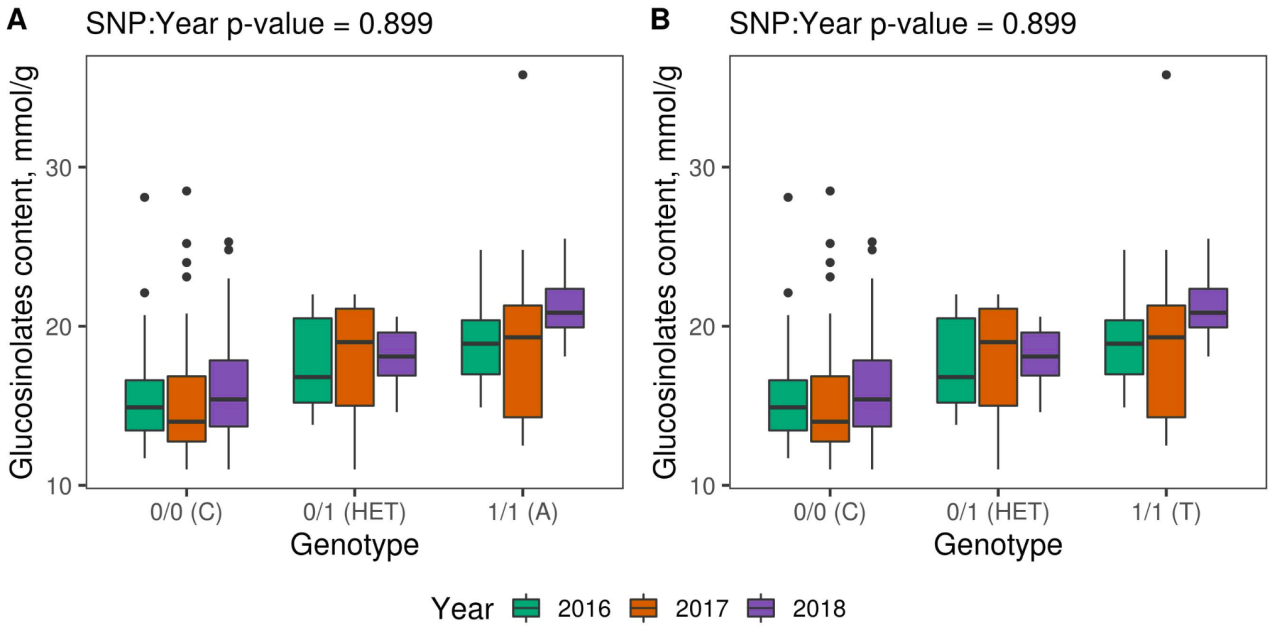

Figure 4.4.2 – Boxplots demonstrating the stability of the two significant SNPs across three years. Panel A corresponds to SA7_26967214, and panel B corresponds to SA7_26967217 Genotypes identified with 0/0, 0/1, 1/1 correspond to reference, heterozygous and alternative genotypes respectively, with corresponding nucleotides indicated in brackets. Colors correspond to the years.

A softer threshold of 0.0005 was also applied to find the additional genetic markers, as a result, additional five SNPs were detected that explained from 8.5 to 23.4% of glucosinolate content variation. In the previous studies it was demonstrated that the identified genetic markers explained 4 to 42% of phenotypic variance (Jan et al., 2019; F. Li et al., 2014; S. Liu et al., 2020; C.-M. Qu et al., 2015), this indicates that the effects identified in the present study are in the concordance with the previously detected range.

Table 4.4.1 – List of SNPs significantly associated with the glucosinolate content.

SNP	PVE (2016)	P-value (2016)	PVE (2017)	P-value (2017)	PVE (2018)	P-value (2018)	P-value (Irwin-Hall)
SA1_4407039	0.113	0.0136	0.117	0.0127	0.234	0.0001	0.0003
SA6_21541176	0.127	0.0079	0.112	0.0176	0.202	0.0002	0.0003
SA7_26967211	0.092	0.0168	0.181	0.0003	0.141	0.0011	0.0001
SA7_26967214	0.204	0.0002	0.138	0.002	0.172	0.0002	3.15E-06
SA7_26967217	0.204	0.0002	0.138	0.002	0.172	0.0002	3.15E-06
SA7_26967241	0.085	0.0229	0.191	0.0002	0.143	0.0009	0.0002
SC6_35450938	0.107	0.0035	0.083	0.0083	0.089	0.0038	0.0001

Since it was demonstrated that the genetic diversity of the VNIIMK lines significantly differs (Chapter 4.2), the comparison between the previously identified SNPs (Jan et al., 2019; F. Li et al., 2014; S. Liu et al., 2020; C.-M. Qu et al., 2015; D. Wei et al., 2019) were compared to SNPs from the present study. To do so, SNPs associated with glucosinolate content for which the physical location was known were located in the genomic assembly used in the present study within the frame of 100kb. As a result, no significant associations for SNPs from the previous studies were identified. This could be explained by the difference in genetic background identified by PCA. Additionally, a smaller amount of the significant SNPs identified may be due to the fact that previous studies included larger amounts of accessions demonstrating broader trait variance of glucosinolate content ranging from 8 to 146 micromoles per gram of fresh weight (Jan et al., 2019; F. Li et al., 2014; S. Liu et al., 2020). Another reason could be related to different environmental conditions in the present and previous studies, namely none of the previous studies were carried out in the environmental conditions of the Krasnodar region. In turn, the variability of the glucosinolate content is affected by environmental factors (Bohinc & Trdan, 2012), furthermore, QTLs controlling glucosinolate only under specific environmental

conditions were identified (He et al., 2018). As the SNPs identified in the present study were specific only for the VNIIMK accessions their application for other non-related germplasms could be limited. Thus they should be first tested before the development of marker-assisted selection approaches.

After the genetic associations were identified, scanning for potential candidate genes was performed within the 100kb region upstream and downstream of the significant SNPs. The such distance was selected as the r^2 for this region remained high according to the LD analysis performed for the studied cohort r^2 value equalling 0.25 or more was specific for more than 30% of SNP pairs located within the 100 kb frame (Figure 4.4.1). Additionally, the same distance was used in the previous LD mapping studies in rapeseed to identify candidate genes (Guan et al., 2019; Q. Zhou et al., 2018; Q. Zhu et al., 2019).

Two significant SNPs SA7_26967214 and SA7_26967217 and two marginally significant SA7_26967211 and SA7_26967241 were localized within the intergenic region of genes for the U-box domain-containing protein 35-like (NCBI GeneID: 106357364) and MTERF2 chloroplastic-like (NCBI GeneID: 106354661). Both genes were not previously discussed in terms of regulation of glucosinolate concentration. Further analysis of the adjacent 100kb region revealed gene (NCBI GeneID: 106354679) encoding BES1/BZR1 homolog protein 4-like located 75.3 kb downstream of these four SNPs (SA7_26967214, SA7_26967217, SA7_26967211, SA7_26967241). This BES1 of the BZR family proteins was previously shown to be involved in brassinosteroid-dependent signaling and regulating the glucosinolate biosynthesis pathway (Guo et al., 2013). Additionally, 38.1 kb downstream of the significant SNPs, a gene (NCBI GeneID: 106354656) encoding histone acetyltransferase HAC1 was identified. Genes encoding HAC1 histone acetyltransferase regulate senescence of the leaves, additionally it was demonstrated that the knockout mutants of *Arabidopsis thaliana* show down-regulation of genes encoding proteins involved in glucosinolate biosynthesis (Hinckley et al., 2019).

Of three other significant SNPs at $p = 0.0005$ the first one (SA6_21541176) represented a synonymous mutation within the gene encoding the derlin-2.1 protein (NCBI Gene ID: 106347898). The second one (SA1_4407039) was annotated as a missense mutation within the gene of uncharacterized protein BNAA01G06520D. Additionally, SA1_4407039 was located 31.8 kb downstream of the gene (NCBI Gene ID: 106392894) encoding the γ-glutamyl peptidase 1-like protein. It was previously shown that γ-glutamyl peptidases are involved in the biosynthesis of glucosinolates. Furthermore, *A. thaliana* plants with impaired genes encoding γ-glutamyl peptidases GGP1 and GGP3 demonstrated stained glucosinolate biosynthesis compared to the wild-type plants (Geu-Flores et al., 2009, 2011). The last SNP (SC6_35450938) was localized outside the gene regions.

4.5 Conclusions

In the present part of the work a genetic collection of rapeseed accessions from the VNIIMK collection has been characterized. To do so we applied a GBS which allowed us to collect information on polymorphisms for genetic diversity description and association mapping at a reasonable time and financial costs.

Both ADMIXTURE and PCA identified significant population structure in the studied cohort. A clear separation previously identified between the spring and winter ecotypes (Gazave et al., 2016; D. Wei et al., 2017; D. Wu et al., 2019) was confirmed for the VNIIMK rapeseed collection. Additionally, a strong difference was revealed between yellow and dark seeded winter rapeseeds that was not previously reported and in the present study could be explained by the divergence in recent breeding history. Linkage disequilibrium analysis supported the previous observation of the higher LD in the C subgenome compared to the A subgenome (Qian et al., 2014; D. Wei et al., 2017; Zheng et al., 2017) which is explained by the larger size of C subgenome chromosomes compared to ones from A subgenome (F. Sun et al., 2017). These observations suggest that the rapeseed peculiarities of population structure are common for international and VNIIMK lines.

The scanning for genetic loci revealed two genetic markers (SA7_26967214 and SA7_26967217) significantly associated with the glucosinolate as well as an additional five marginally significant SNPs (SA1_4407039, SA6_21541176, SA7_26967211, SA7_26967241, SC6_35450938). The proportion of the variance explained by these markers was comparable to the previously studied (Jan et al., 2019; F. Li et al., 2014; S. Liu et al., 2020; C.-M. Qu et al., 2015). Thus these genetic markers could be involved in the development of marker-assisted breeding in VNIIMK after verification.

A scanning for candidate genes allowed to identify novel genetic loci (chromosome A1:4307039-4507039; chromosome A6:21441176-21641176; chromosome A7:26867214-27067214; chromosome C6:35350938-35550938) associated with glucosinolate content. Notably, these ones were not previously identified by GWAS and QTL-based studies, which could be the result of studying the collections that differed from the VNIIMK collection at the genetic and phenotypic level. Previously genetic markers associated with glucosinolate content were identified close to genes encoding enzymes involved in the biosynthesis of glucosinolates. Specifically, MAM1 and MAM3 enzymes (Jan et al., 2019; Kittipol et al., 2019) are involved in glucosinolate side chain modification in *Arabidopsis* as well as AOP3 acting at the final stages of aliphatic glucosinolates biosynthesis (D. Wei et al., 2019). Additionally, genetic markers are located in the genes homologous to ones encoding enzymes for glucosinolate transporters GTR2 (S. Liu et al., 2020; C.-M. Qu et al., 2015). Several genetic markers strongly associated with glucosinolate content were located close to the genes encoding transcription factors of MYB28 and MYB34 families and transcription factor HAG1 controlling aliphatic glucosinolate biosynthesis (F. Li et al., 2014; D. Wei et al., 2019). In the present study SNPs marking the regions that were previously associated with glucosinolates were not detected. This could be the result of the genetic difference that was identified while comparing VNIIMK and international lines, as well as the effect of environmental factors. Additionally, the phenotypic diversity (glucosinolate concentration) was

lower compared to the cohorts previously used to perform GWAS for that trait (Jan et al., 2019; F. Li et al., 2014; S. Liu et al., 2020).

Notably, several SNP markers (SA7_26967214, SA7_26967217, SA7_26967211, SA7_2696724, SA1_4407039) of two loci (chromosome A1:4307039-4507039 and chromosome A7:26867214-27067214) that were identified in the present study by means of GWAS were previously reported to be involved in the control of glucosinolates. Namely, by GWAS we additionally indirectly supported the role of genes encoding γ-glutamyl peptidases (NCBI Gene ID: 106392894), HAC1 histone acetyltransferase (NCBI GeneID: 106354656), BES1 of the BZR family proteins (NCBI GeneID: 106354679) that were previously demonstrated to be involved in the regulation of the glucosinolate content control (Geu-Flores et al., 2009, 2011; Guo et al., 2013; Hinckley et al., 2019). Additionally, we identified novel genetic regions (chromosome A6:21441176-21641176 and chromosome C6:35350938-35550938), containing genes that were not previously described, related to the regulation of glucosinolate content regulation which could facilitate a better understanding of the genetic control of glucosinolate content in rapeseed.

Chapter 5. QTL-mapping of oil-quality traits in sunflower

5.1 Introduction

Two traits associated with the oil quality in sunflower include tocopherol composition and oleic acid. Tocopherols are natural antioxidants that serve as a source of vitamin E. There are four types of tocopherols: α, β, γ, and δ. Notably, antioxidant activity increases in the row from α to δ while vitamin E activity decreases in the row (J. M. Fernández-Martínez et al., 2007). It should be mentioned that the natural diversity of tocopherols in sunflower is low and normally sunflower produces only α form (Y. Demurin et al., 1996; Hass et al., 2006). Thus several attempts at screening large genetic collections have been made to identify the mutations leading to the altered tocopherol composition. Such mutations were identified in sunflower lines from the Vavilov Institute (VIR) in Russia (Y. Demurin, 1993) and the Institute of Sustainable Agriculture (CSIC) collection in Spain (Velasco, Domínguez, et al., 2004). The mutations identified in the VIR collections were transferred to the LG15 and LG17 lines that are now part of VNIIMK genetic collection (Y. Demurin, 1993; Y. Demurin et al., 1996; Y. N. Demurin et al., 2004).

It was demonstrated that the tocopherol composition is controlled by two major effect loci *Tph1* (*m*) and *Tph2* (*g*), which are independently inherited (Y. N. Demurin et al., 2004). Using the CSIC lines and QTL mapping approaches *Tph1* was mapped on linkage group (chromosome) 1 and *Tph2* was mapped on linkage group 8 (Velasco, Pérez-Vich, et al., 2004; Vera-Ruiz et al., 2005). Mutation in *Tph1* was shown to be associated with the gene encoding 2-methyl-6-phytyl-1,4-benzoquinone (MPBQ) methyltransferase (MPBQ-MT) which converts MPBQ to 3,2-dimethyl-6-phytyl-1,4-benzoquinone (DMPBQ) (Tang, Hass, et al., 2006). The accumulation of the DMPBQ or MPBQ leads to the accumulation of the γ- and δ-tocopherols, respectively (Figure 5.1.1).

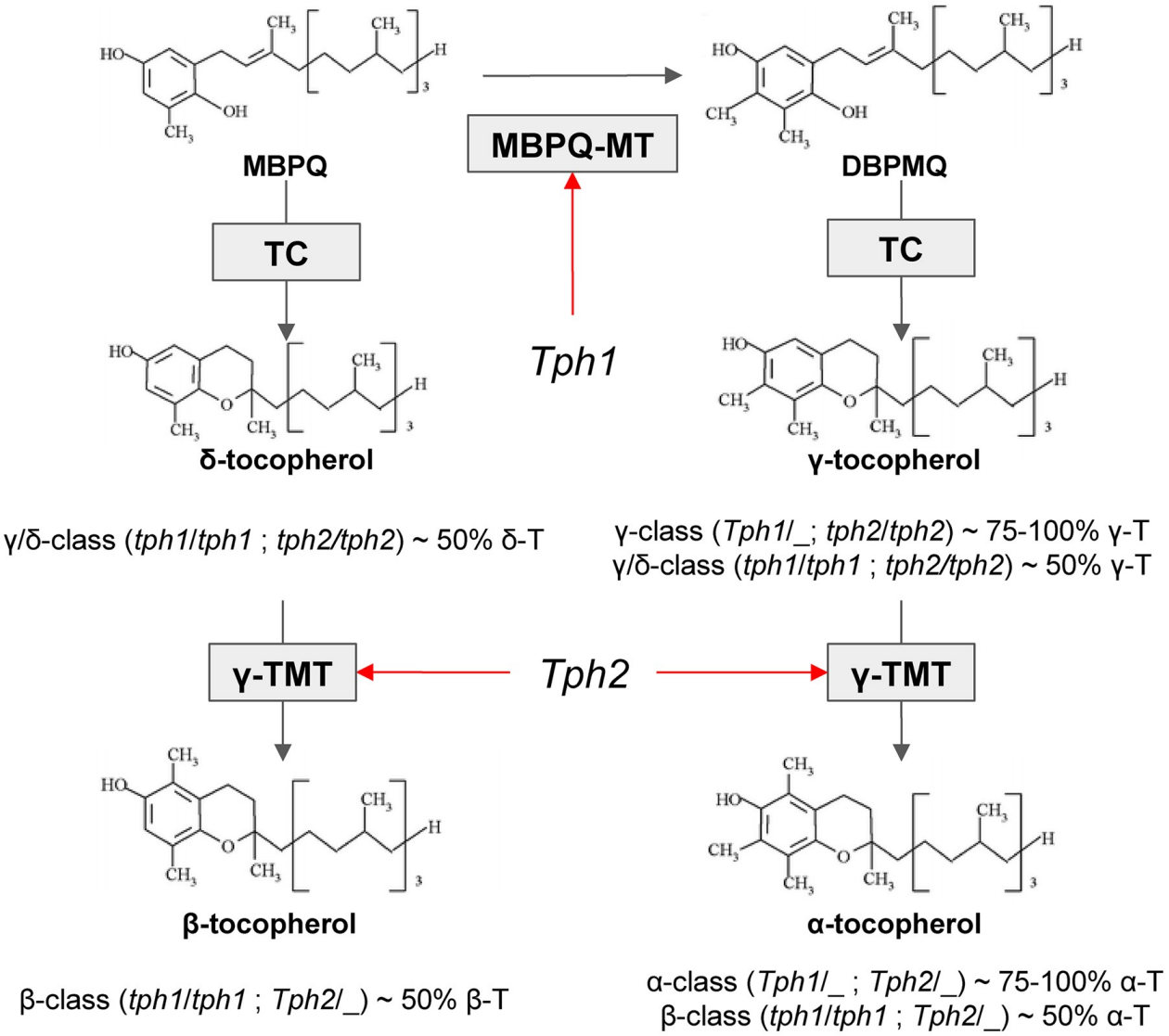

Figure 5.1.1 – Tocopherol biosynthesis scheme with corresponding phenotype classes. Names of the metabolites are indicated in bold: 2-methyl-6-phytyl-1,4-benzoquinone—MPBQ;3,2-dimethyl-6-phytyl-1,4-benzoquino ne — DMPBQ. Gray arrows correspond to the reactions catalyzed by the enzymes indicated in the gray squares: MPBQ methyltransferase—MPBQ-MT; tocopherol cyclase—TC; γ-tocopherol methyltransferase—γ-TMT. Red arrows indicate enzymes encoded by *Tph1* and *Tph2*. The biosynthesis pathway scheme was adapted from the previously published review (Lushchak & Semchuk, 2012).

In turn, the impaired *tph1* locus was associated with the increased accumulation of β-tocopherol. Mutation in *Tph2* was associated with the gene

encoding γ-tocopherol methyltransferase (γ-TMT), which converts γ- and δ-tocopherols to α- and β-tocopherols (Figure 5.1.1), respectively (Hass et al., 2006). Mutant *tph2* loci were associated with the accumulation of the γ-tocopherols. It also should be noted that the double mutant (*tph1*/*tph1*; *tph2*/*tph2*) plants accumulate γ- and δ-tocopherols in an almost equal ratio (Demurin, Borisenko, et al., 2006; Demurin et al., 2016; Gubaev et al., 2022).

Although it is assumed that the tocopherol composition is controlled by these two major effect loci, several studies identified additional and minor effect loci. Namely, a locus named *d* carrying gene coding for MPBQ-MT transferase (MT-2) paralogous to *m* (*Tph1*) was shown to interact epistatically with *m* (*Tph1*) and *g* (*Tph2*). This locus was mapped on chromosome 4 (Hass et al., 2006). Minor effect loci (modifying genes) affecting *tph2* mutation were mapped to chromosomes 1, 9, 14 and 16 (García-Moreno et al., 2012a). It was also assumed that the expressivity of *tph1* and *tph2* may depend on the genetic background of the lines (Y. Demurin et al., 1996). Thus the control of tocopherol composition in sunflowers remains under discussion. Additionally, currently, no genetic markers were identified for modern lines from the VNIIMK collection carrying *Tph1* and *Tph2* mutations.

Oleic acid is the monounsaturated fatty acid that on the one hand protects the lipids against thermooxidation during frying and on the other hand has positive health effects, namely lowering cholesterol and reducing inflammation (Sales-Campos et al., 2013). It should be noted that the antioxidant properties of oleic acid are increased in the oils with high content of γ- and/or δ-tocopherols (Skorić et al., 2008; Warner et al., 2008). Thus many breeding efforts in oilseeds including sunflower are made with the aim of increasing oleic acid content, especially in varieties used for frying oil production (Dehmer & Friedt, 1998; Monteros et al., 2008; Schierholt et al., 2000). As with tocopherols, there is low natural diversity for oleic acid in sunflowers (Cvejić et al., 2014). The first sunflower variety with high oleic acid content called Pervenets was obtained by chemical mutagenesis (Soldatov, 1976). Lately, this mutation was introduced within various collections with the aim of obtaining high-oleic varieties. It

was first assumed that the high oleic trait is controlled by a single dominant *Ol* gene (Urie, 1985). Studies based on hybridological analysis suggested more complex trait inheritance, in particular, it was assumed that this gene could be controlled by three dominant complementary genes (*Ol1*-*Ol3*) and one modificator gene (J. Fernández-Martínez et al., 1989; J. F. Miller et al., 1987). Development of the QTL mapping techniques allowed the identification of RAPD markers associated with the major *Ol* gene (Dehmer & Friedt, 1998). Lately, this gene explaining up to 56% of phenotype variance was mapped to chromosome 14 (Pérez-Vich et al., 2002). It was demonstrated that *Ol* is associated with the FAD2-1 gene encoding fatty acid desaturase that converts oleic acid to linoleic acid (Schuppert et al., 2006). A more recent study confirmed the role of *Ol* and identified additional minor effect loci located on chromosomes 8 and 9, explaining up to 10% of phenotypic variance (Premnath et al., 2016). A QTL study based on high-throughput genotyping additionally identified loci located in linkage groups 9 and 6, explaining 12 and 6% of phenotype variance (F. Zhou et al., 2018). These observations are in concordance with the hypothesis on the effect of genetic background on the *Ol* gene that has been previously put forward (Demurin & Škorić, 1996). Thus as in the case of tocopherols, the genetic control of oleic acid remains under discussion and similarly, there are no genetic markers that have been found with the implementation of lines from Russian genetic collections.

Thus the aim of this part of the research was to identify major and potential minor effect loci and perspective SNP markers associated with altered tocopherol composition and high oleic acid content. To do so we analyzed two experimental crosses derived from crossing contrast modern lines from VNIIMK. The part of the results related to the genetic map construction and mapping of the tocopherol composition presented in the following chapter was published in the G3-Genes Genomes Genetics journal (Gubaev et al., 2022).

5.2 Phenotype evaluation

To collect phenotype data two crosses derived from wild-type lines VK101 and VK303, used as parental lines of hybrid "Typhoon" as well as mutant lines VK876 and VK195 used as parental lines of hybrid "Oxy" producing oil with increased oxidative stability. Mutant VK876 and VK195 as maternal lines were crossed and selfed with wild-type paternal lines VK101 and VK303, respectively to obtain two F2 populations (VK195xVK303 and VK876xVK101). To characterize tocopherol composition a thin-layer chromatography (TLC) was applied. As a result 142 and 144 F2 seeds for crosses VK195xVK303 and VK876xVK101 were phenotyped, respectively. Parental lines were phenotyped in at least 7 replicates. Maternal lines VK195 and VK876 were attributed to the γ/δ phenotype as they demonstrated almost equal γ-/δ-tocopherol content. Paternal lines VK303 and VK101 demonstrated a wild-type α phenotype; the proportion of α-tocopherol equaled 100% (Table 5.2.1).

Table 5.2.1 – Tocopherol composition in parental lines.

Line name	Genotype	Tocopherol class	Mean proportion of α-tocopherol, %	Mean proportion of β-tocopherol, %	Mean proportion of γ-tocopherol, %	Mean proportion of δ-tocopherol, %
VK195	*tph1*/*tph1*;*tph2*/*tph2*	γ/δ	0	0	51.7 ± 5.41	48.3 ± 5.41
VK303	*Tph1*/*Tph1*; *Tph2*/*Tph2*	α	100	0	0	0
VK876	*tph1*/*tph1*;*tph2*/*tph2*	γ/δ	0	0	52.9 ± 8.71	47.1 ± 8.71
VK101	*Tph1*/*Tph1*; *Tph2*/*Tph2*	α	100	0	0	0

F2 progeny was classified into tocopherol phenotypic classes based on the tocopherol composition depending on the putative allelic states of two major genes *Tph1* and *Tph2*. Each of the plants was assigned to one of the four tocopherol classes. As a result, 142 and 144 phenotyped F2 seeds from the VK195xVK303 and

VK876xVK101, 137 (96.4%) and 143 (99.3%) were classified into the known phenotypic classes, respectively (Figure 5.2.1).

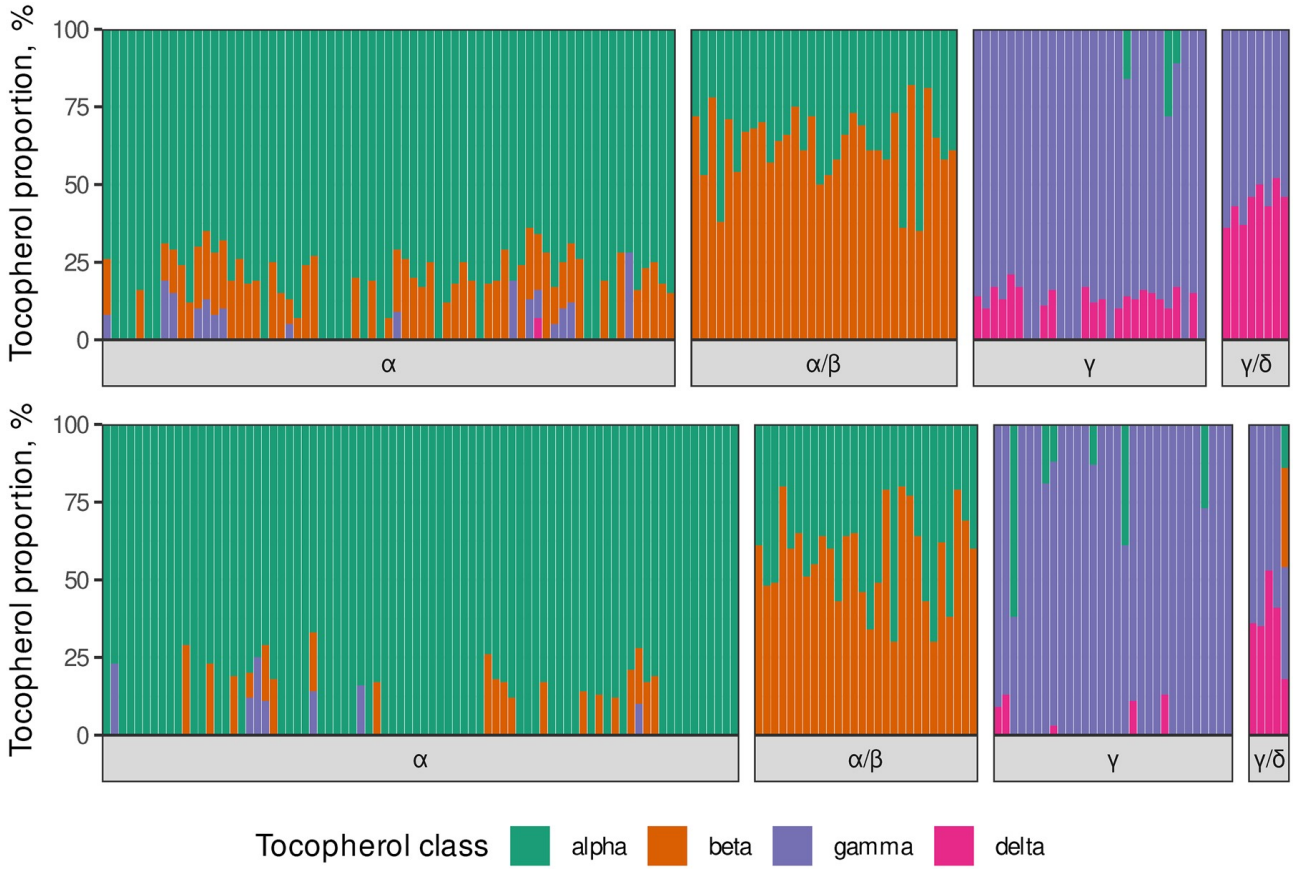

Figure 5.2.1 – Relative content of each of the tocopherol classes among the genotyped individuals in the F2 progeny. Each bar corresponds to a single F2 seed. Each colored bar shows the proportion of each of the four tocopherol classes. Facets show the distribution of phenotyped F2 seeds across the α-, α/β-, γ-, and γ/δ-tocopherol phenotypic classes. The upper panel corresponds to the VK195xVK303 cross. The lower panel corresponds to the VK876xVK101 cross.

Only one and five F2 seeds from VK195xVK303 and VK876xVK101 crosses, respectively, were not assigned to any specific phenotypic class since the tocopherol ratios for these accessions deviated from the specified classes. Phenotype classes of the classified F2 progeny matched the 9:3:3:1 distribution according to the $\chi 2$ goodness-of-fit test (Table 5.2.2) with high confidence (P-values of 0.4505 and

0.5382 for VK195xVK303 and VK876xVK101, respectively). These observations speak in favor of a two-gene model.

Table 5.2.2 – Number of plants assigned to one of four tocopherol classes for each of the crosses.

Cross	α-class	α/β-class	γ-class	γ/δ-class	P-values for χ2 goodness of fit test
VK195xVK303	69	32	28	8	0.4505
VK876xVK101	80	28	30	5	0.5382

To characterize relative oleic acid content, a gas chromatography followed by mass spectrometry was applied. As a result, the relative content of oleic acid was determined for parental lines in at least seven replicates and for 142 and 144 F2 seeds for crosses VK195xVK303 and VK876xVK101, respectively (Figure 5.2.2).

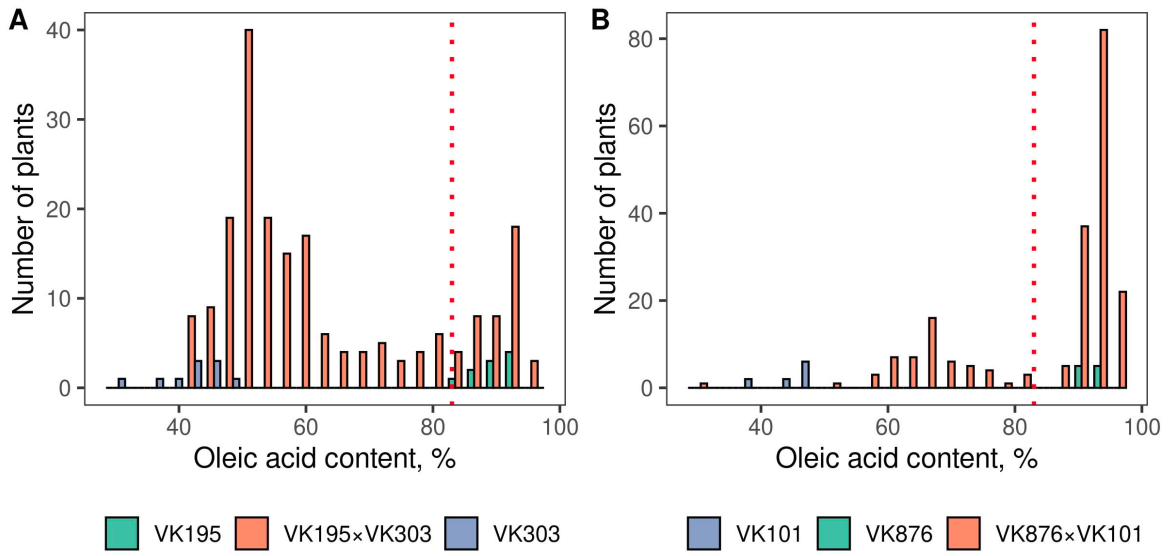

Figure 5.2.2 – A histogram of the distribution of the oleic acid content for parental lines and F2 progeny. Panel A reflects the distribution of oleic acid content for cross VK195xVK303, panel B reflects the distribution of oleic acid content for cross VK876xVK101. The green color of the bars corresponds to high-oleic mutant

parental line, the blue color corresponds to wild-type parental lines, and the orange color corresponds to F2 progeny. Dotted red lines correspond to oleic acid content specific for high-oleic plants.

As monohybridism due to the single Ol gene was assumed, plants were distributed to high-oleic and non-high-oleic classes. As a result 104 and 40 F2 seeds from cross VK876xVK101 were classified as high-oleic and non-high-oleic, respectively. For cross VK195xVK303 28 and 114 F2 seeds from the cross were classified as high-oleic and non-high-oleic, respectively. Next $\chi 2$ goodness of fit was applied to test for monohybridism and expected 3:1 phenotype distributions (Table 5.2.3).

Table 5.2.3 – The distribution of the F2 progeny into high-oleic and non-high oleic classes.

Cross	High-oleic	Non-high-oleic	P-values for $\chi 2$ goodness of fit test
VK195xVK303	40	104	<2.2e-16
VK876xVK101	114	28	0.44

For the population VK876xVK101, the $\chi 2$ goodness of fit test confirmed 3:1 segregation ratio, with the *Ol* allele behaving as dominant allele in that cross (Table 5.2.3). For the population VK195xVK303, the $\chi 2$ goodness of fit test also confirmed a 3:1 segregation ratio (three non-high-oleic and one high-oleic), but with the *Ol* mutant behaving as a recessive allele (P-value of $\chi 2$ goodness of fit test = 0.14).

5.3 Genotyping and genetic map construction

For genetic map construction we applied a genotyping-by-sequencing approach to identify genetic polymorphisms for studied crosses. As a result, 425213 and 499063 raw SNPs were collected for VK195xVK303 and VK876xVK101, respectively. Next SNPs were filtered according to the filters (see methods section) as

well as SNPs that were homozygous within and polymorphic between the parents were selected as this is a prerequisite for genetic map construction. As the low SNP coverage is a common problem for the GBS while being applied for F2 populations that demonstrate high amounts of heterozygous positions, an imputation based on the LB-Impute algorithm was applied (Fragoso et al., 2016). Imputation decreased the proportion of missing genotypes from 0.22 to 0.1 for the VK195xVK303 cross and from 0.2 to 0.09 for the VK876xVK101 cross. After imputation, 7028 SNPs for the VK195xVK303 cross and 5,876 SNPs for the VK876xVK101 cross were used for map construction (Figure 5.3.1) within the r/qtl package (Broman et al., 2003). During the genetic map construction genetic markers were additionally filtered by the expected genotype segregation pattern for F2 populations (1:2:1). Additionally, markers that significantly affected the map length were discarded (for details see methods section).

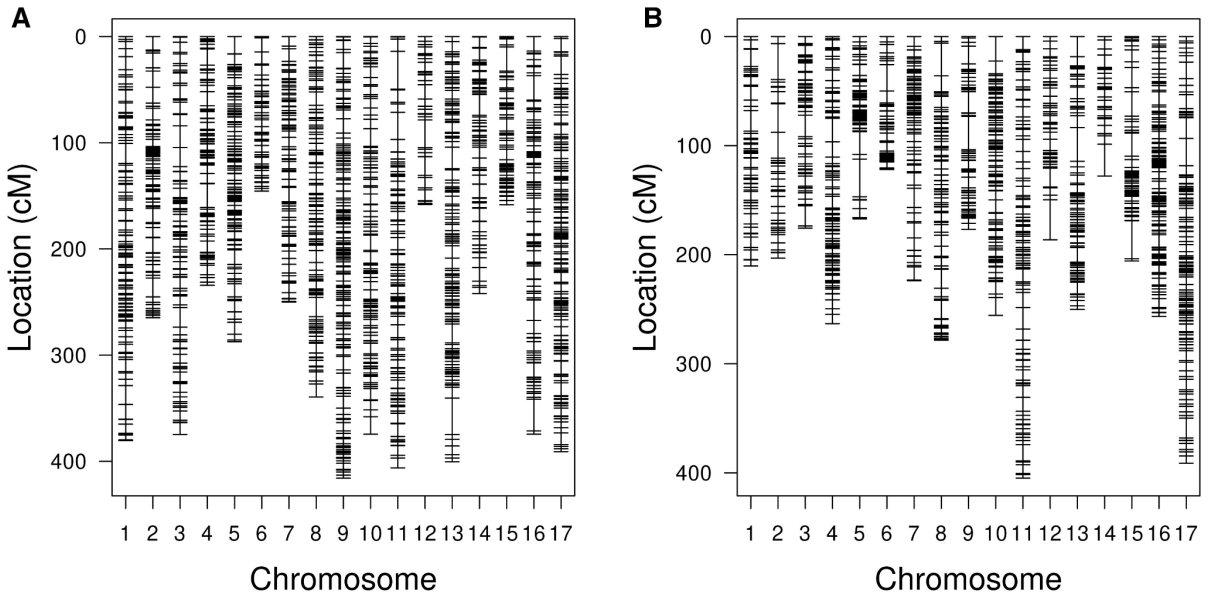

Figure 5.3.1 – SNP-based genetic linkage maps for the VK195xVK303 (panel A) and VK876xVK101 (panel B) crosses.

A genetic map for cross VK195xVK303 consisted of 3200 SNPs and spanned a total distance of 5,197.7 cM, with the maximum spacing of 35.5 cM and the average intermarker distance of 1.63 cM. The correlation between physical and

genetic map was significant (P-value for Pearson correlation coefficient < 1.223024e-19), Pearson correlation coefficient ranged from 0.62 to 0.97 across the chromosomes (Table 5.3.1).

Table 5.3.1 – Summary for the genetic maps constructed for VK195xVK303 and VK876xVK101 crosses.

Chromosome	Cross	Number of markers	Map length (cM)	Average spacing (cM)	Maximum spacing (cM)	Pearson correlation coefficient	P-value For Pearson correlation coefficient
1	VK195 x VK303	189	380.54	2.02	18.87	0.9	6.01E-68
2		188	264.79	1.42	17.82	0.72	3.1E-31
3		163	374.72	2.31	20.43	0.86	7.8E-49
4		136	234.29	1.74	22.41	0.85	1.18E-39
5		246	287.46	1.17	16.34	0.82	3.18E-62
6		95	145.51	1.55	13.32	0.88	1.52E-32
7		173	250.03	1.45	13.48	0.62	1.22E-19
8		238	339.45	1.43	18.24	0.94	1.42E-109
9		315	415.84	1.32	14.74	0.97	1.15E-188
10		240	374.47	1.57	16.69	0.95	1.42E-119
11		194	406.14	2.1	35.49	0.62	2.14E-22
12		57	158.27	2.83	26.02	0.91	2.02E-22
13		216	400.4	1.86	34.37	0.93	4.96E-98
14		126	242.1	1.94	19.66	0.9	3.36E-46
15		127	158.39	1.26	22.72	0.91	1.79E-48
16		187	374.41	2.01	29.86	0.79	1.8E-41
17		310	390.89	1.27	17.23	0.92	1.2E-126

overall		3200	5197.71	1.63	35.49	-	-
1	VK876 x VK101	98	210.29	2.17	18.77	0.9	1.7E-36
2		77	203.15	2.67	32.93	0.93	7.75E-34
3		118	175.66	1.5	19.68	0.9	8.98E-45
4		190	263.26	1.39	21.52	0.96	3.44E-106
5		138	167.13	1.22	34.49	0.89	2.48E-49
6		104	121.91	1.18	24.18	0.89	1.13E-36
7		215	223.92	1.05	17.5	0.48	5.21E-14
8		184	278.34	1.52	29.87	0.92	1.92E-75
9		104	176.82	1.72	30.57	0.95	5.43E-54
10		225	255.63	1.14	22.49	0.89	1.58E-79
11		137	404.86	2.98	19.73	0.9	1.25E-49
12		100	186.25	1.88	36.6	0.81	3.54E-24
13		165	250.15	1.53	30.8	0.96	9.5E-90
14		75	127.86	1.73	29.11	0.9	1.52E-27
15		105	205.8	1.98	34.89	0.94	7.57E-50
16		236	256.59	1.09	12.07	0.87	2.24E-73
17		300	391.19	1.31	24.98	0.89	2.8E-102
overall		2571	3898.81	1.53	36.6	-	-

For cross VK876xVK101 genetic map consisted of 2571 SNPs and spanned 3,898.8 cM. With the maximum spacing of 36.6 cM and average intermarker distance of 1.53 cM. The correlation between physical and genetic maps was significant (P-value for Pearson correlation coefficient < 5.205962e-14) and Pearson correlation coefficient ranged from 0.48 to 0.96 across the chromosomes (Table 5.3.1).

Previously the high-density SNP-based genetic maps based on the reduced representation sequencing (SLAF-, RAD-, GBS-sequencing) data for sunflower

included from 817 to 6,136 SNPs and was from 1444 to 2472 cM in length (Celik et al., 2016; Talukder et al., 2014; F. Zhou et al., 2018). There are two possible explanations that in the present study genetic maps were almost two times larger compared to the previously reported ones. First, relatively relaxed filters were applied for segregation distortion. Second, errors in genotyping were present due to the nature of the sunflower genome which contains a high amount of repetitive regions (Badouin et al., 2017). This in turn may lead to errors during read alignment (Kane et al., 2011; Treangen & Salzberg, 2012). It also should be noted that for other plant species large genetic maps and large chromosomes compared to the obtained ones were also reported. Large chromosomes that were longer than 300 cM were reported for maize (Su et al., 2017; J. Wang et al., 2018) and wheat (Yang et al., 2018); both maps were based on the reduced representation sequencing. A large RAD-seq based genetic map was also reported for sweetpotato (Shirasawa et al., 2017); the map was 7313.5 cM in length with the largest chromosome of 904.5 cM. The correlation between the genetic and physical distance reported in the present study was higher compared to the previously reported values of the Pearson correlation coefficient ranging from 0.2 to 0.7 (Celik et al., 2016). This speaks in favor that the constructed genetic maps are in concordance with the physical ones.

5.4 QTL-mapping of tocopherol composition

The next stage of the study was related to the association mapping of tocopherol composition. Here two approaches have been implemented; the first one was associated with mapping the proportion of each tocopherol class as independent observation (quantitative mapping). The second was related to the mapping of the tocopherol composition associated with the putative allelic states of *Tph1* and *Tph2* revealed by the belonging of the plant phenotypes to one of the four tocopherol classes (qualitative mapping).

The quantitative mapping approach revealed loci associated with α- β-, γ- and δ-tocopherols for both crosses (Figure 5.4.1). Additionally, a proportion of phenotypic variance explained by SNP tagged loci was calculated for markers 1.5-LOD interval estimated for the most significant marker (Supplementary Table 4 from https://academic.oup.com/g3journal/article/12/4/jkac036/6527637#supplementary-data). For cross VK195xVK303 loci associated with the proportion of α-tocopherol were located on chromosome 1, additionally, loci associated with a proportion α-, β-, γ- and δ-tocopherols were mapped on chromosome 8. The maximum proportion of variance explained by markers located on chromosome 1 explained 21.98% and 38.14% of α- and β-tocopherol variance. For cross VK876xVK101 loci associated with β-tocopherol were mapped on chromosome 1, additionally, loci associated with α-, γ-, and δ-tocopherols were mapped on chromosome 8. The maximum proportion of β-tocopherol variance explained by markers located on chromosome 1 was 44.44%. The markers located on chromosome 8 explained up to 71.09%, 85.74%, and 9.83% of α-, γ-, and δ-tocopherol content.

For both crosses associations for β-tocopherol were identified on chromosome 1 while associations for α-, γ-tocopherol were found on chromosome 8 which is in concordance with the previously published results (García-Moreno et al., 2012b; Tang, Hass, et al., 2006; Vera-Ruiz et al., 2006).

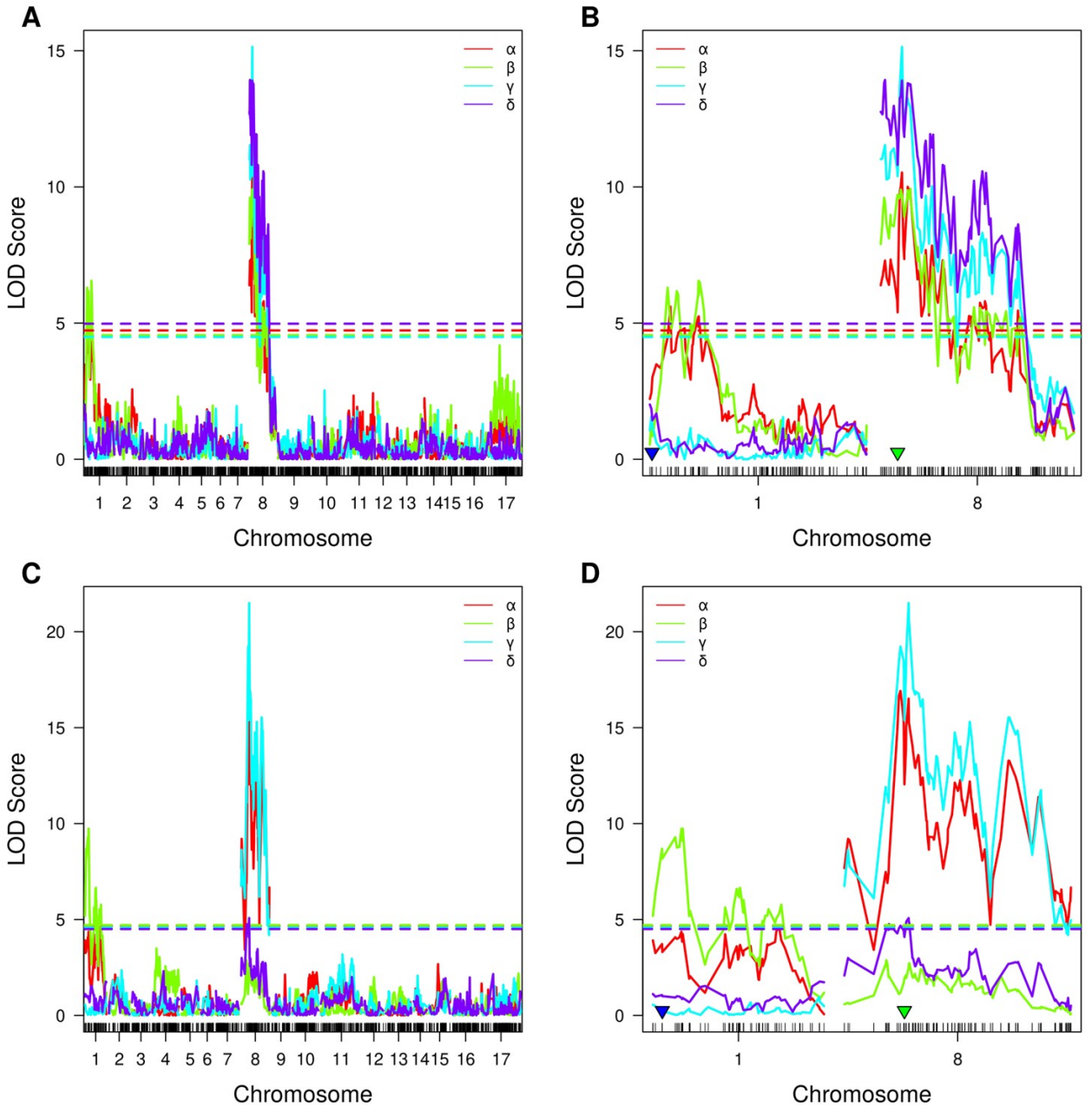

Figure 5.4.1 – Likelihood curve of the LOD score for α-, β-, γ-, and δ-tocopherol content for the VK195xVK303 a and b) and VK876xVK101 c and d) populations obtained by using the interval mapping approach. Dashed lines correspond to the permutation threshold. The color of lines corresponds to the tocopherol type. a and c) The mapping results for all chromosomes. b and d) The mapping results for chromosomes carrying significant markers. Blue and green triangles indicate markers that are most closely located to the *Tph1* and *Tph2* loci, respectively, based on the physical map data.

Additionally, the differences in mapping results were identified between the crosses which could be the result of different distributions of the relative content of tocopherol classes (Figure 5.2.1). Compared to the previous studies the proportion of variance explained by the markers was significantly higher, namely, for β- tocopherol genetic markers explained 90% of phenotypic variance (Vera-Ruiz et al., 2006) and for γ-tocopherols genetic markers explained 97% of the variance (García-Moreno et al., 2006). This could be explained by the fact that compared to the previous studies, double mutant lines were analyzed in which the tocopherol composition is controlled by two independent loci simultaneously. Furthermore, additional loci that were previously mapped on chromosomes 4, 9, 14, 16 were not identified (García-Moreno et al., 2012b; Hass et al., 2006). This could be explained by the relatively small amount of progeny used to find associations as well as with the strong effect of major effect loci located on chromosomes 1 and 8 which mask minor effect loci. This fact makes double mutant lines VK195 and VK876 carrying *tph1* and *tph2* mutant alleles with high expressivity convenient for breeding programs (Y. N. Demurin et al., 2004). Thus the next step of the present study was directly related to the mapping of major effect *Tph1* and *Tph2* loci.

To do so we applied qualitative mapping of *Tph1* and *Tph2* as the distribution of the classes followed a 9:3:3:1 distribution. For each phenotype of the plant the potential presence or absence of *Tph1* and *Tph2* were assigned as phenotypes. The quantitative approach mapped *Tph1* within a 1.5-LOD interval spanning from 77.77 to 93.35 cM on chromosome 1 with the maximum LOD score of 20.19 for VK195xVK303 cross (Figure 5.4.2 A, B). For cross, VK876xVK101 *Tph1* was mapped on chromosome 1 from 3.05 to 44.54 cM 1.5-LOD interval with the maximum LOD score of 15.6 (Figure 5.4.2 C, D). For cross VK195xVK303 *Tph2* was mapped on chromosome 8 from 32.12 to 64.69 cM with a maximum LOD score of 18.33 (Figure 5.4.2 A, B), for cross VK876xVK101 *Tph2* associated region was located from 75.98 to 83.57cM on chromosome 8 with maximum LOD value equal to 24.65 (Figure 5.4.2 C, D).

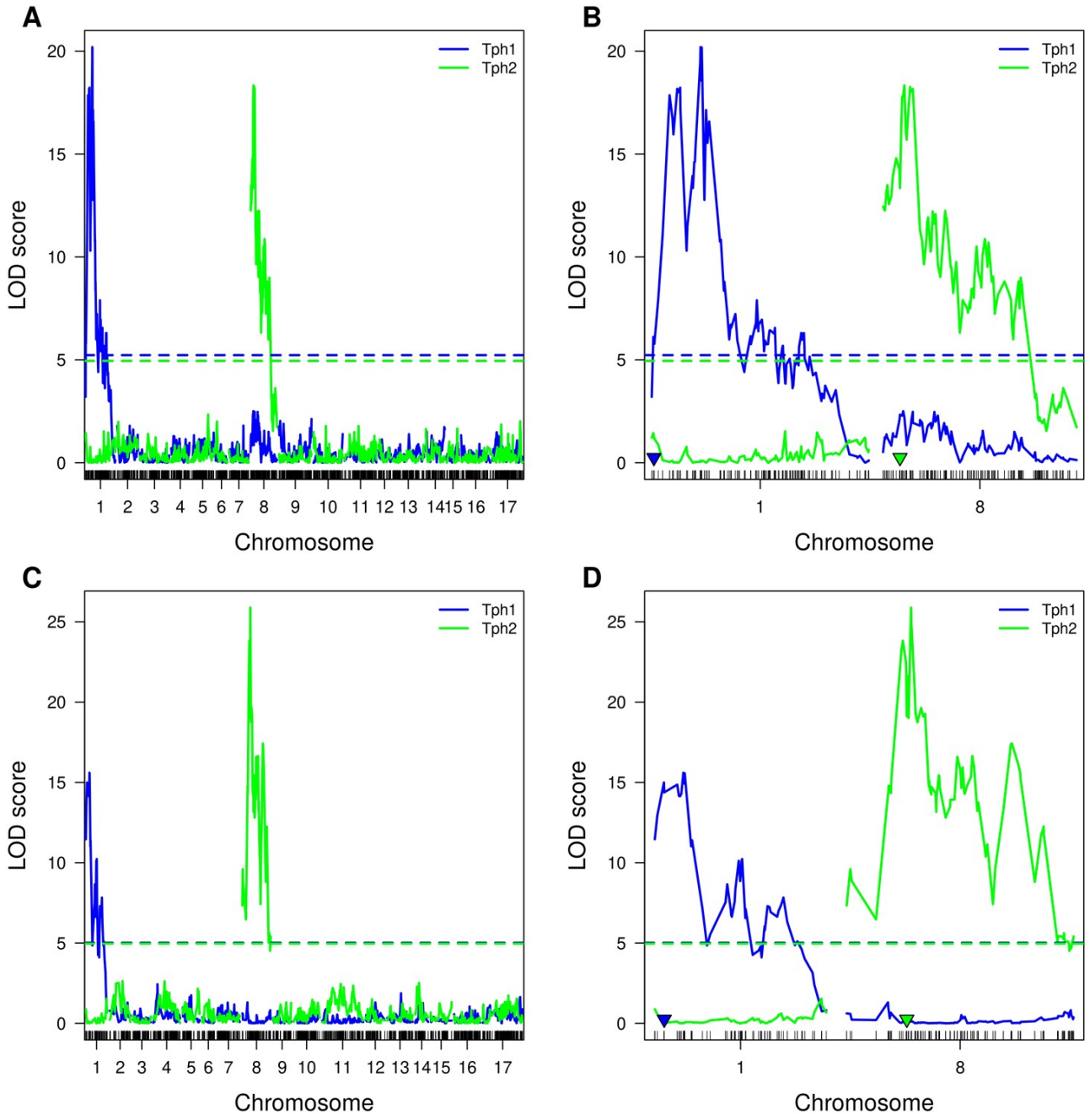

Figure 5.4.2 – Likelihood curve of the LOD score for *Tph1* and *Tph2* for the VK195xVK303 a and b) and VK876xVK101 c and d) populations. Dashed lines correspond to permutation results. Interval mapping results for *Tph1* and *Tph2* are indicated with the corresponding colors. a and c) The mapping results for all chromosomes. b and d) The mapping results for chromosomes carrying significant markers. Blue and green triangles indicate markers that are most closely located to the *Tph1* and *Tph2* loci, respectively, based on the physical map data.

Next, we scanned for the closest markers located to the causal genes associated with *Tph1* and *Tph2*. According to the information on the physical location of the *Tph1* gene encoding MPBQ-MT (NCBI Gene ID: 110937001) it was established that the closest significant markers were located 0.5 (marker S1_9228105) and 5.3 megabases (marker S1_4480460) for crosses VK195xVK303 and VK876xVK101, respectively. These markers demonstrated significant LOD scores of 6.11 and 14.37. For *Tph2* locus carrying genes encoding γ-TMT1 and γ-TMT2 (NCBI Gene IDs: 110872346 and 110872347) closest significant markers were located 0.34 (marker S8_21613153) and 0.42 megabases (marker S8_21536716), for crosses VK195xVK303 and VK876xVK101, respectively. The respective LOD scores were 13.34 and 19.10, respectively.

The results of mapping were in concordance with the published ones. Previously *Tph1* was also mapped on the upper end of chromosome 1 using F2 crosses (Tang, Hass, et al., 2006; Vera-Ruiz et al., 2006). The markers associated with *Tph1* were also physically located to the gene encoding MPBQ-MT. For *Tph2* it was also demonstrated that this locus is located on chromosome 8 for different mapping populations (García-Moreno et al., 2006; Hass et al., 2006). As in the case with *Tph1* markers, *Tph2*-associated ones were closely located to genes encoding γ-TMT1 and γ-TMT2 (Tang, Hass, et al., 2006). It should be noted that in the above-mentioned studies a quantitative approach was applied i.e. the mapping was performed for relative content of β- and γ-tocopherol proportions were analyzed (García-Moreno et al., 2006, 2006; Hass et al., 2006; Tang, Hass, et al., 2006). Notably, different statistical approaches were used to map these traits, including single marker ANOVA (García-Moreno et al. 2006; Hass et al. 2006; Vera-Ruiz et al. 2006) and interval mapping (García-Moreno et al. 2006), composite interval mapping (García-Moreno et al. 2012). Here, we have demonstrated that in our case, the qualitative approach based on interval mapping for traits with binary distribution was also suitable for mapping major effect loci and helped us to map *Tph1* and *Tph2* in both populations.

5.5 QTL mapping of oleic acid content

To map oleic acid content, three approaches have been used. First, the raw phenotypes were mapped, i.e the relative content of the oleic acid by two methods: interval mapping adapted for non-normally distributed traits and composite interval mapping as it was previously used to map this trait. Additionally, the transformation of the data into binary traits was performed following the mapping of high-oleic (>=83%) and non-high oleic traits (< 83%) as this threshold was previously used to divide high and non-high oleic phenotypes (Lacombe et al., 2009).

For cross VK195xVK303 a peak located on chromosome 14 was identified with all three methods (Figure 5.5.1). A 1.5-LOD confidence interval calculated for composite interval mapping results spanned from 199.76 to 208.4 cM. The markers that were located within the 1.5-LOD confidence interval explained up to 56.37% of phenotyping variance (Table 5.5.1, Figure 5.5.1). According to the physical map information, a genetic marker S14_137961667 was most closely (13.3 megabases away) located to the gene encoding FAD2-1 (NCBI GeneID 110904312) located from 151341021 to 151344179 bp and associated with high oleic acid content (Lacombe & Bervillé, 2001; Schuppert et al., 2006). Notably for the cross VK195xVK303 genetic markers for high content of oleic acid demonstrated the recessiveness, which was expected starting from the analysis of the phenotype distribution that was contrariwise compared to the expected 3:1 segregation ratio where 3 should correspond to high-oleic phenotypes and 1 part for non-high oleic phenotypes. The expected segregation was previously explained due to the dominant nature of the mutant *Ol* allele that carries two copies of gene encoding fatty acid desaturases, it was assumed that the one impaired copy FAD2-1 could reduce the expression of normal FAD2-1 via production of siRNA (Lacombe et al., 2009). The fact that the expected segregation ratio was not observed as well as the recessive nature of the genetic markers could be explained by the presence of the additional recessive *Ol* allele, and/or the effect of the genetic background of the wild-type line VK303 as the dominant state of the *Ol* was expected for the carrier line VK195.

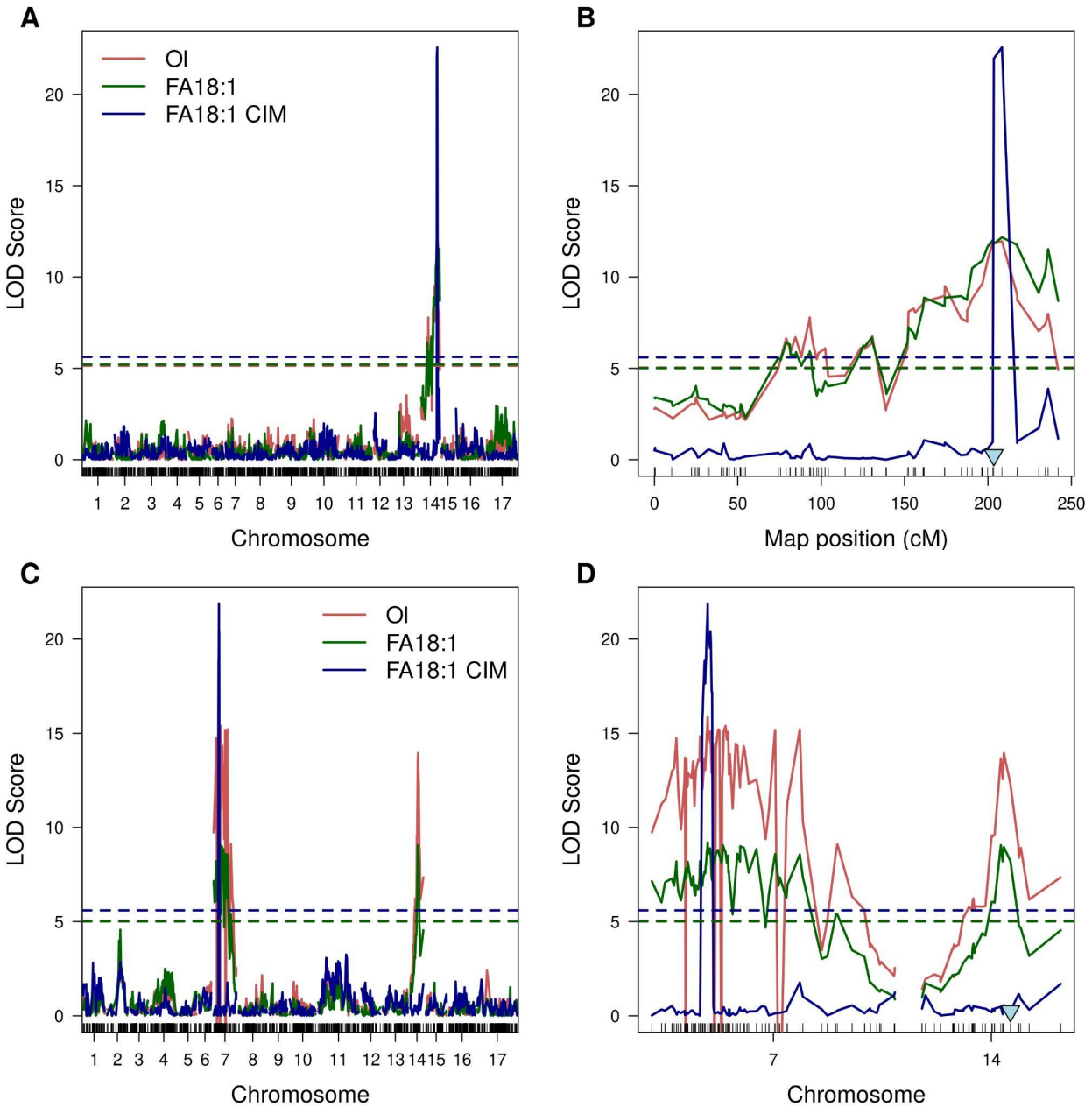

Figure 5.5.1 – Likelihood curves of the LOD score *Ol* gene and oleic acid content for VK195xVK303 (panels A and B) and VK876xVK101 (panels C and D) crosses. The red curve reflects the results of mapping the binary trait associated with the presence and/or absence of the *Ol* gene with the interval mapping approach. The green curve reflects the results of interval mapping adapted for non-normally distributed traits of oleic acid content. The blue curve reflects the results of composite interval mapping of oleic acid content. Light blue triangles indicate markers that are most closely located to the *Ol* gene encoding FAD2-1 based on the physical map data.

For cross VK876xVK101 interval mapping for both types of traits (binary and numeric) identified loci located on chromosomes 7 and 14 which contradicted the expectation of the presence of one major effect loci (Figure 5.5.1 C, D).

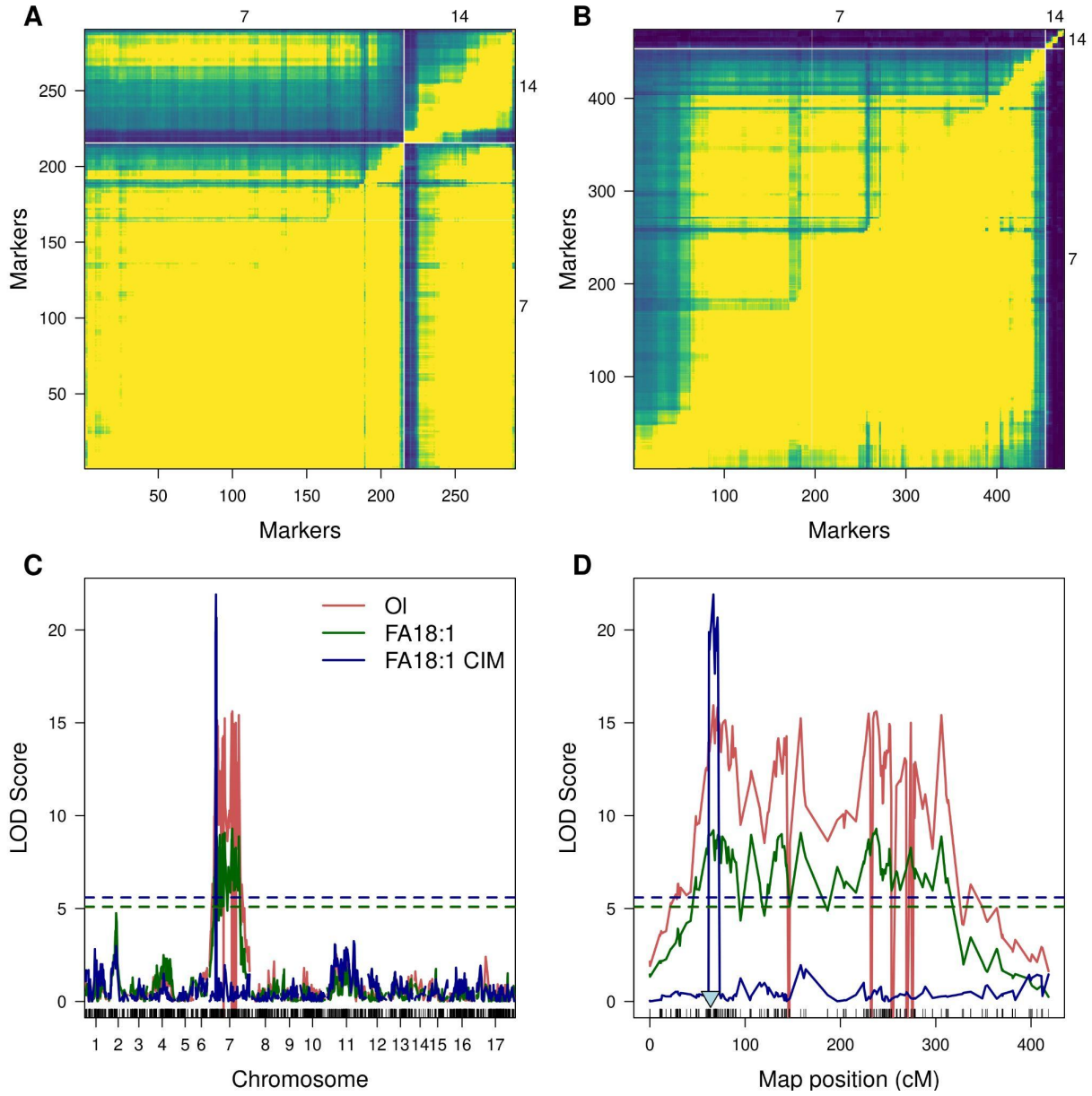

Figure 5.5.2 – Reassembling the chromosomes 7 and 14 for cross VK876xVK101 (Panels A and B) and updated mapping results (Panels C and D). The upper triangle of the matrices (panels A and B) corresponds to recombination fraction (RF), lower triangle corresponds to LOD scores. The blue color corresponds to high RF values and low LOD scores indicating poor linkage, while the yellow color corresponds to low RF values and high LOD indicating tight linkage between the pairs of markers.

Panel A reflects the RF and LOD scores for initial maps for chromosomes 7 and 14, panel B reflects RF and LOD scores for reassembled chromosomes 7 and 14. Red and green curves reflect the results of interval mapping of the binary *Ol*-associated trait and non-normally distributed oleic acid content, respectively. The blue curve reflects the results of composite interval mapping of oleic acid content. A light blue triangle indicates a marker located close to the *Ol* gene encoding FAD2-1.

In contrast, composite interval mapping detected only one locus located on chromosome 7 (Figure 5.5.2 D). Further analysis revealed the fact that the presence of loci located on chromosomes 7 and 14 could be explained by the tight linkage between markers from chromosome 7 and chromosome 14 (Figure 5.5.2 A, B). This could be the result of the potential translocation occurring between the markers from chromosome 7 and chromosome 14. Thus it was decided to reconstruct the genetic maps for markers from chromosomes 7 and 14 (Figure 5.5.2) without prior knowledge of the physical belonging of the markers to chromosomes. As a result, a significant amount of the genetic markers from chromosome 14 became a part of chromosome 7. Next, we performed the mapping with the re-assembled genetic map. As a result, we identified a single peak located on chromosome 7. Notably, the 1.5-LOD confidential interval obtained for significant markers identified by composite interval mapping from chromosome 7 spanning from 63.72 to 71.45 cM included one marker S14_78139380 which should be located on chromosome 14 based on the physical map (Table 5.5.1). Based on the physical map data S14_78139380is located 73 Mb away from the FAD2-1 gene. Thus it could be hypothesized that the FAD2-1 gene could be translocated from chromosome 14 to chromosome 7 and it is not a novel mutation since according to the breeding history the *Ol* mutation in the VK876 line originates from Prevents (Soldatov, 1976). Nevertheless, additional molecular cytogenetic analysis (for example fluorescence *in situ* hybridization (FISH)) is needed to test if this translocation actually occurred.

Table 5.5.1 – Summary of genetic markers located within 1.5-LOD confidence interval resulted from composite interval mapping.

Marker	Cross	Chromosomr	Position	Ol	FA18:1	FA18:1 CIM	PVE, %
S14_136819253	VK195xVK303	14	199.76	10.91	11.66	20.48	53.95
S14_137821336		14	202.86	11.91	12.01	22.27	56.37
S14_137795880		14	202.86	11.91	12.01	22.27	55.73
S14_137889873		14	203.49	11.85	11.82	21.96	56.07
S14_137951402		14	203.49	11.85	11.82	21.96	56.18
S14_137961667		14	203.49	11.85	11.82	21.96	56.41
S14_138978611		14	208.4	11.95	12.17	1.83	54.86
S14_78139380	VK876xVK101	7	63.72	13.95	8.96	19.88	46.99
S7_37821562		7	66.74	15.94	9.22	21.91	53.29
S7_39112905		7	66.74	15.94	9.22	21.91	50.46
S7_39776264		7	67.11	15.18	8.6	20.03	44.27
S7_39681220		7	67.11	15.18	8.6	20.03	47.39
S7_25411825		7	68.23	13.89	7.6	17.65	48.38
S7_25624783		7	68.23	13.89	7.6	17.65	46.14
S7_25717162		7	68.23	13.89	7.6	17.65	44.37
S7_25624784		7	68.23	13.89	7.6	17.65	46.14
S7_25664488		7	68.23	13.89	7.6	17.65	46.24
S7_25624756		7	68.23	13.89	7.6	17.65	46.14
S7_25421436		7	68.23	13.89	7.6	17.65	46.46
S7_25726978		7	68.23	13.89	7.6	17.65	44.37
S7_76273882		7	68.92	15.06	8.5	19.98	49.05
S7_25876036		7	68.92	15.06	8.5	19.98	48.62
S7_76319288		7	68.92	15.06	8.5	19.98	49.05
S7_25964564		7	68.92	15.06	8.5	19.98	48.83

S7_75968410		7	68.92	15.06	8.5	19.98	48.91
S7_76547907		7	68.92	15.06	8.5	19.98	49.11
S7_25961858		7	68.92	15.06	8.5	19.98	48.75
S7_25855919		7	68.92	15.06	8.5	19.98	48.62
S7_75140370		7	69.28	15.13	8.67	20.07	48.45
S7_72038884		7	69.28	15.13	8.67	20.07	51.02
S7_75826681		7	69.64	15.2	8.63	20.08	49.09
S7_75503792		7	69.83	14.59	8.23	18.88	46.9
S7_75503823		7	69.83	14.59	8.23	18.88	46.9
S7_75934390		7	70.02	15.2	8.63	20.08	49.01
S7_75892237		7	70.02	15.2	8.63	20.08	49.09
S7_75894251		7	70.02	15.2	8.63	20.08	49.09
S7_75535778		7	70.02	15.2	8.63	20.08	49.02
S7_76736113		7	71.1	15.83	8.78	20.66	50.33
S7_76758370		7	71.1	15.83	8.78	20.66	50.33
S7_76852814		7	71.45	15.2	8.96	19.73	48.72

The maximum proportion of variance explained by the genetic markers from the 1.5-LOD confidential interval was of 53.29% (Table 5.5.1, and Figure 5.5.3 B). Notably, unlike the significant markers from the VK195xVK303 the genetic markers from the VK876xVK101 demonstrated a dominant effect of the mutant *Ol* allele (Figure 5.5.3 B). This is in concordance with the identified segregation ratio of 3:1 reported for high-oleic and non-high oleic plants from cross VK876xVK101. The dominance of the genetic markers is also in agreement with the previously published results for the *Ol* gene (Lacombe et al., 2009; Lacombe & Bervillé, 2001; Schuppert et al., 2006).

It should be noted that the proportion of variance explained by the markers was similar compared to the previously published results that reported the range of 56 –

66% (Lacombe & Bervillé, 2001; Premnath et al., 2016). In addition, any additional loci that were previously described to affect the oleic acid content in addition to the major *Ol* gene (Premnath et al., 2016; F. Zhou et al., 2018) were not reported in the present study. This, first, could be related to the different genetic backgrounds of the lines from the present study and previously studied ones. Second, this could be due to the relatively low density of the genetic markers and sample size as well as the low resolution of F2 mapping populations.

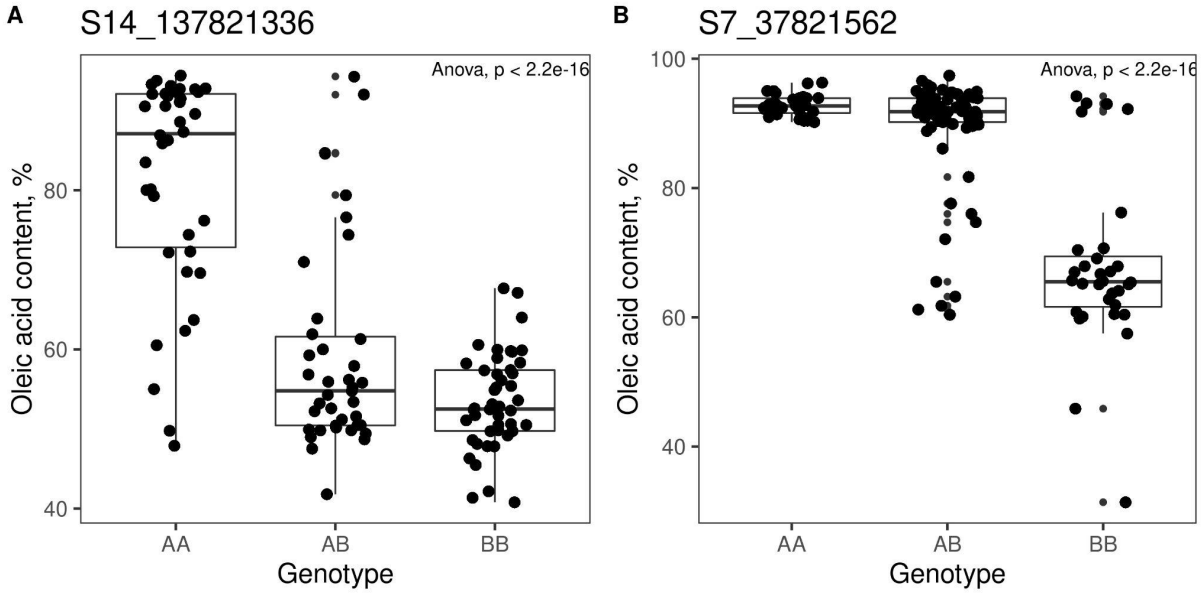

Figure 5.5.3 – Boxplots for genotype effects on oleic acid content. AA – genotype corresponds to maternal ones of a mutant line, BB – genotype wild ones derived from wild-type males. Panel A reflects the genotype effect of the most significant marker S14_137821336 for cross VK195xVK303. Panel B reflects the genotype effect of the most significant marker S7_37821562 for cross VK876xVK101.

The contradictory results of the unexpected effect of the recessive markers from cross VK195xVK303 laid the ground for deeper investigation of the *Ol* allele including sequencing of the targeted region and/or additional mapping experiments with the genetically contrasting parents. The potential translocation identified for cross VK876xVK101 could serve as a basis for the cytogenetic analysis of chromosomes 14 and 7 of lines VK876 and VK101. Additionally, a more precise

mapping of the major effect *Ol* gene is needed for this cross VK876xVK101. Although the potential genetic markers were identified for marker-assisted selection a deeper investigation of the loci potentially carrying the *Ol* gene is needed for both crosses.

5.6 Validation of the genetic markers

The genetic markers associated with the tocopherol composition were evaluated using Sanger sequencing. This was done to test the concordance between the alleles identified by GBS and Sanger sequencing being a "gold standard". To do so one marker for each of *Tph1* and *Tph2* and each of the two crosses (Table 5.6.1) was selected.

Table 5.6.1 – Summary of SNP verification.

Marker	Chromosome	Position (cM)	Population	Gene	LOD Score	Accuracy, %	Number matching genotypes
S1_55196434	1	88.71	VK195 X VK303	*Tph1*	20.18	94.12	32
S1_71748138	1	36.68	VK876 X VK101	*Tph1*	15.57	91.18	31
S8_23941299	8	79.34	VK876 X VK101	*Tph2*	24.66	85.29	29
S8_30578572	8	47.32	VK195 X VK303	*Tph2*	18.26	91.18	31

As a result, we tested four genetic markers. The main criteria for selection were the following: location of the marker within the 1.5-LOD confidence interval, unique sequence surrounding that makes it possible to construct primers for amplification of

a specific DNA region. Based on the above-mentioned criteria, the following markers were selected: S1_55196434 and S8_30578572 associated with *Tph1* and *Tph2* for the VK195xVK303 cross, respectively. For the VK876xVK101 cross, we selected S1_71748138 and S8_23941299 markers for *Tph1* and *Tph2*, respectively . As a result, four pairs of primers were constructed with the aim of amplification of four unique regions capturing the markers that underwent sequencing.

For verification of the markers, we selected 8 plants carrying the maternal (mutant) genotype, 8 plants carrying the paternal (wild-type) genotype and 8 ones carrying heterozygous genotypes based on the GBS data. Additionally, markers were tested on 5 female and 5 male parental plants. For this cohort, the allelic states of the markers were determined using Sanger sequencing and then compared to the allelic states obtained by GBS and calculated the proportion of matching genotypes. The genotypes obtained using the GBS approach that were used for genetic map construction and QTL mapping matched the genotypes obtained by Sanger sequencing with accuracy ranging from 85.29% to 94.12% (Table 5.6.1).

In addition to the verification of the allelic states with two different methods, an attempt was made to test whether the identified markers could be specific for non-relative lines. To test this, 20 genetically and phenotypically diverse lines were selected by collaborators from VNIIMK (Figure 5.6.1). These lines were sequenced in 5 replicates and phenotypically characterized in terms of the presence/absence of *Tph1*, *Tph2* (four phenotype classes) and *Ol* (high-oleic and non-high oleic) alleles of genes based on the tocopherol composition and oleic acid content previously reported for these lines.

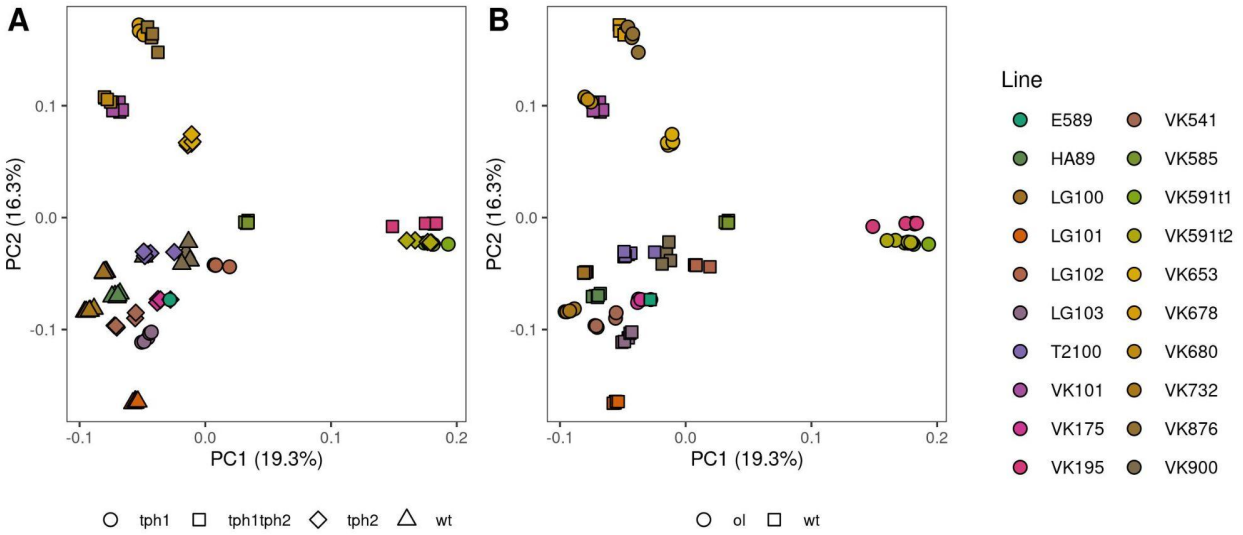

Figure 5.6.1 – Principal component analysis of the population structure of the verification sample. Color corresponds to the sunflower line. The shape of the figures reflects the phenotype associated with either tocopherol composition (Panel A) or oleic acid content (Panel B). The numbers in brackets correspond to the proportion of genotype variance explained by the principal component.

After the sequencing data was obtained, joint SNP calling with experimental cross samples was performed with the aim of finding genetic polymorphisms common in verification sample and experimental crosses. Next, 10 most significant markers were selected for each of the trait/population combinations. Next, *Ol*-, *Tph1*- and *Tph2*-associated phenotypes of the verification samples were predicted based on the allelic states of the significant markers identified either with VK195xVK303 or VK876xVK101 populations (Figure 5.6.2). It was demonstrated that the *Tph2*-associated phenotypes were most accurately predicted with the genetic markers identified with both populations. In particular genetic markers S8_30578572 mapped using VK195xVK303 cross and S8_24695806 mapped using VK876xVK101 cross demonstrated accuracy of 88.8% and 87.5%, respectively, being genotyped in 90% and 80% of lines from verification sample.

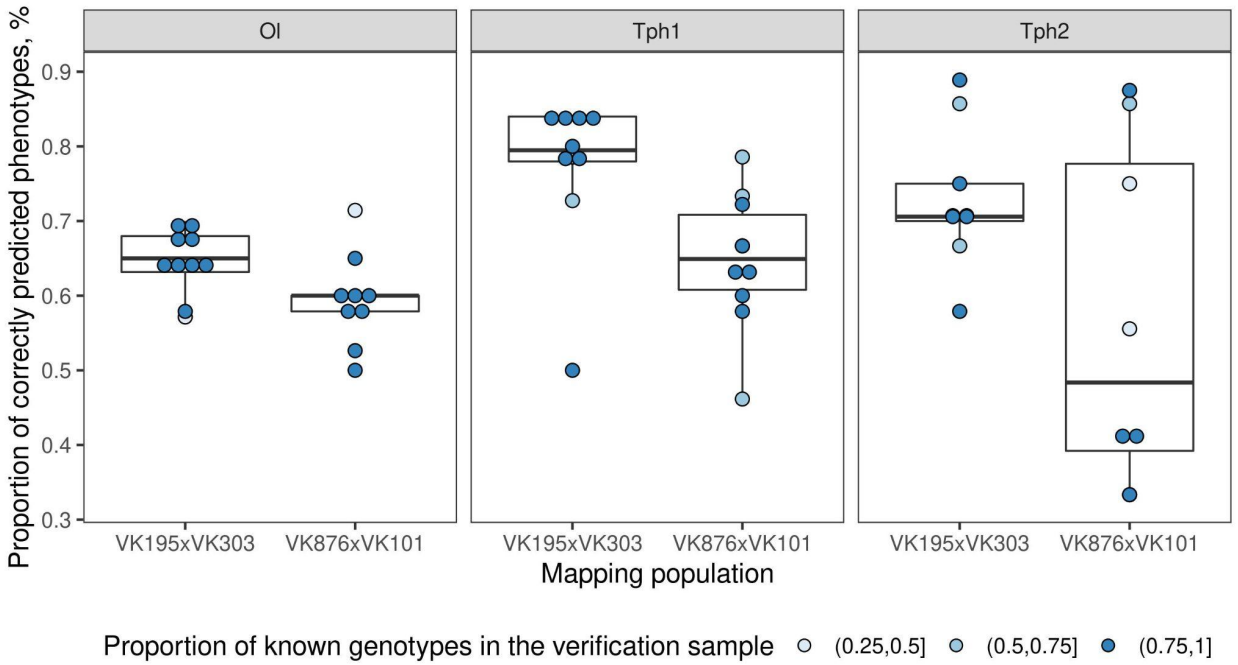

Figure 5.6.2 – Accuracy of the genetic markers to predict *Ol*, *Tph1* and *Tph2*-associated phenotypes (indicated by facets). Each dot corresponds to the marker. The color of the dot represents the proportion of known genotypes that were genotyped among individuals from the verification sample. The abscissa axis indicates the population for which the genetic marker was tested. The ordinate axis reflects the proportion of the correctly predicted phenotypes associated with *Ol*, *Tph1* and *Tph2*.

Such a high proportion of correctly predicted phenotypes and genotyping level could be explained by the physical proximity of these markers to the causal gene located on chromosome 8. For *Tph1* genetic markers S1_56474204, S1_55157943, S1_55196434 and S1_54474863 obtained by analyzing VK195xVK303 cross demonstrated the accuracy of 84% being identified in the more than 90% of verification lines. From cross VK876xVK101 only one genetic marker S1_71748138 predicted a *Tph1*-associated phenotype with an accuracy of 78.8%, the accuracy of the other markers was lower than 75%. For high oleic traits associated with *Ol*, the accuracy of the genetic markers was lower than 75%. In particular marker S7_76852814 from the cross, VK876xVK101 demonstrated a maximum accuracy of

71% and marker S14_136584727 from cross VK195xVK303 demonstrated a maximum accuracy of 70%. Lower predicting accuracy could be related to the fact that these markers were located far away from the causal gene FAD2-1, additionally for VK195xVK303 markers demonstrated unexpected recessive nature, while for population VK876xVK101 genetic markers were linked to chromosome 7, which was not previously reported to carry major effect gene encoding FAD2-1.

5.7 Conclusions

In the present part of the study the mapping of genetic markers associated with the oil quality was performed, namely tocopherol composition and oleic acid content. Two F2 populations derived from parental lines with contrast tocopherol composition and oleic acid content were analyzed. A GBS approach followed by SNP calling, imputation and genetic map construction was applied with the aim of mapping these traits. Here quantitative and qualitative mapping approaches were applied for both traits as it was assumed that in both crosses the two major effect genes *Tph1* and *Tph2* for tocopherol composition and *Ol* for oleic acid content control the trait expression. This assumption is on the one hand supported by the fact that the phenotypic segregation ratios of 9:3:3:1 and 3:1 for tocopherol composition and oleic acid content were observed, indicating a mono/di-genic mode of inheritance for these traits. On the other hand, quantitative mapping of these traits did not reveal any additional loci that control these traits.

In the previous studies aimed on finding markers associated with tocopherol composition markers associated with *Tph1* (Vera-Ruiz et al., 2006) and *Tph2* (García-Moreno et al., 2006) mapped on chromosomes 1 and 8 were identified. In several studies, additional loci controlling tocopherol composition were found (García-Moreno et al., 2012b; Hass et al., 2006). For oleic acid content major effect locus *Ol* was mapped on chromosome 14 (Schuppert et al., 2006). In addition, minor effect loci controlling this trait were discussed in addition to the major effect gene (Premnath et al., 2016; F. Zhou et al., 2018). The fact that here no additional loci

were identified could be explained by first that here we used a relatively small cohort, second by the different genetic backgrounds of the lines from the present and past studies, as well as by the high expressivity of *tph1* and *tph2* in the studied lines VK195 and VK876.

Due to the applied aim of this study associated with the scanning for genetic markers of oil quality, first a GBS approach was validated using the Sanger sequencing, to prove the concordance of the markers' alleles by two independent approaches. Second, genetic markers were tested for the ability to predict the phenotypes associated with major effect genes *Tph1*, *Tph2* and *Ol* in independent plant samples. These results demonstrated that *Tph1*- and *Tph2*-associated markers were identified in a large proportion of the test sample including unrelated lines (> 90%) and predicted associated phenotypes with high accuracy (> 84%). Thus, we selected a set of genetic markers for *Tph1* (S1_55196434 and S1_71748138) and *Tph2* (S8_23941299 and S8_30578572) that will be used to develop a marker-assisted selection approach for sunflower oil improvement in the context of tocopherol composition.

Regarding the oleic acid content, the genetic markers demonstrated lower parameters in the context of accuracy and presence in the test sample. This could be due to the fact that the genetic markers associated with oleic acid content for population VK195xVK303 were recessive, which on the one hand could be related to the genetic background and on the other a potential existence of an alternative allele of *Ol* behaving as recessive ones. For cross VK876xVK101 identified markers were dominant, however due to possible translocation between chromosomes 14 and 7, most of the genetic markers were attributed to chromosome 7 which was not previously reported to carry genes associated with oleic acid content and *Ol* gene in particular. It also should be noted that in both crosses the genetic markers associated with *Ol* were located more than 13 megabases away from the causal gene FAD2-1 based on the physical map information which in turn may indicate poor linkage with this gene. All these disconcordances with the initially set hypothesis could be set as

goals for future studies for a better understanding of *Ol* gene inheritance in the sunflower in lines from the VNIIMK collection.

Chapter 6. Association mapping of seed morphology traits in sunflower

6.1 Introduction

One of the key phenotypes of the sunflower that is of high interest to breeders and industry is seed-related traits – seed size (kernel size), husk size (hull size) and seed to husk ratio. Seed and husk size are of particular interest in the development of large-seeded confectionary sunflowers, additionally (Lukomets et al., 2021). One of the methods used to estimate seed size is a measurement of the weight of thousand seeds. This technique was widely applied to oilseed crops (Khan et al., 2019; Souza et al., 2016), including the sunflower (Radic et al., 2013) as the husk length and width significantly correlate with the thousand seed weight (Gjorgjieva et al., 2015). However, the main disadvantage of this method is that it does not provide information on the ratio of seed and husk area, being an important parameter of whether the photoassimilates are rationally utilized by plants in terms of increasing the yield output. One of the ways to estimate seed size, husk size and seed to husk ratio in a non-invasive manner is to perform x-ray radiography methods followed by image analysis (Arkhipov et al., 2019). Previously the X-ray computer methods were mostly applied to assess morphological aspects of the seeds (Rocha et al., 2014), quality including viable seeds, empty seeds and seeds damaged by the pathogen (Dumont et al., 2015), seed viability (Al-Turki & Baskin, 2017). In the present part of the study genetic mapping of seed-related traits collected by means of the X-ray computer methods by our collaborators from St. Petersburg Electrotechnical University LETI was performed. Traits that were collected included husk area, seed area as well as a seed-to-husk ratio. These characteristics were collected for the genetically and phenotypically diverse collection of the sunflower which included a total of 601 accessions from VIR and VNIIMK institutions and the Agroplasma breeding company. The part of the results related to the genetic diversity assessment

of the 601 lines in the following chapter was published in the BMC Genomics journal (Chernova et al., 2021).

6.2 Collection of phenotypes

To collect the information on the seed (kernel) size, husk (hull) size and seed to husk ratio the seeds were placed on the plate (Figure 6.2.1) and next scanned using PRDU-02 setup followed by analysis using SeedRentgen. As a result, mean values were collected for the seed area and husk area (Table 6.2.1).

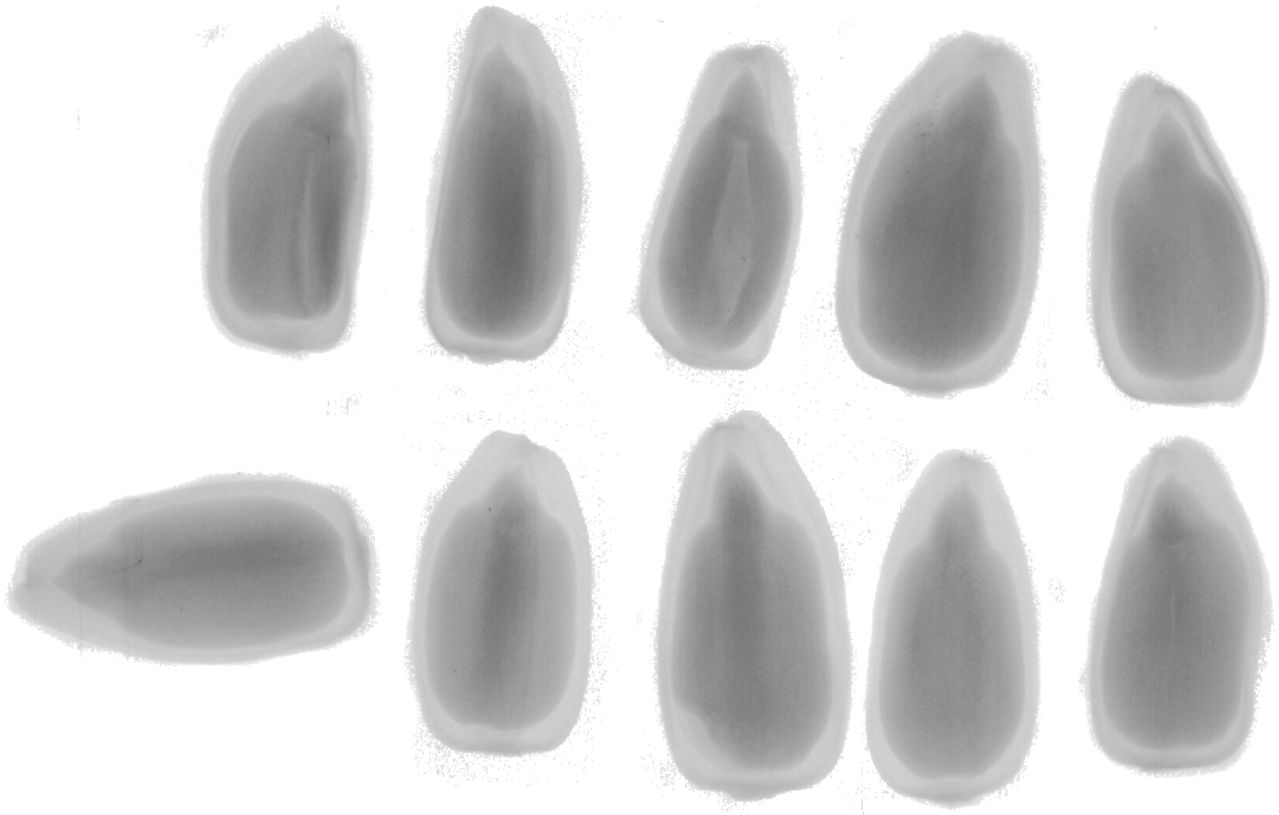

Figure 6.2.1 – Example of X-ray radiograph with inverted color. Each radiograph was made for 10 seeds of accession. The original pictures are with a black background and white seeds. The size of the pixel corresponds to 0.1 mm^2.

Table 6.2.1 – Summary table on seed-related traits. Mean values ± standard deviation are shown.

Collection	Number of accessions analyzed	Husk (hull) area, mm^2	Seed (kernel) area, mm^2	Husk to seed ratio
VIR	255	15.01 ± 4.53	7.51 ± 2.01	51.15 ± 7.01
VNIIMK	199	39.88 ± 8.11	19.28 ± 3.85	49.11 ± 8.34
AGROPLASMA	147	37.06 ± 11.68	20.56 ± 5.21	57.10 ± 7.75
Mean	-	28.54 ± 14.21	14.54±7.10	51.92 ± 8.23

Analysis of variance revealed a significant difference of seed and husk area between the collections (Figure 6.2.2). Namely, husk was on average 2.65 times smaller in VIR accessions when comparing VIR and VNIIMK collections (Mann–Whitney U test p-value < 2.2e-16) collections and on average 2.46 times smaller when comparing VIR and AGROPLASMA collections (Mann–Whitney U test p-value < 2.2e-16). The difference between AGROPLASMA and VNIIMK collection was far less prominent (1.07 times), but still significant (Mann–Whitney U test p-value = 0.012). The same pattern applied to the seed area: on average VIR accessions were 2.56 times smaller than VNIIMK accessions (Wilcoxon test p-value < 2.2e-16) and 2.73 times smaller than AGROPLASMA accessions. The difference between VNIIMK and AGROPLASMA was marginally significant (Mann–Whitney U test p-value = 0.041) and was 1.07 times.

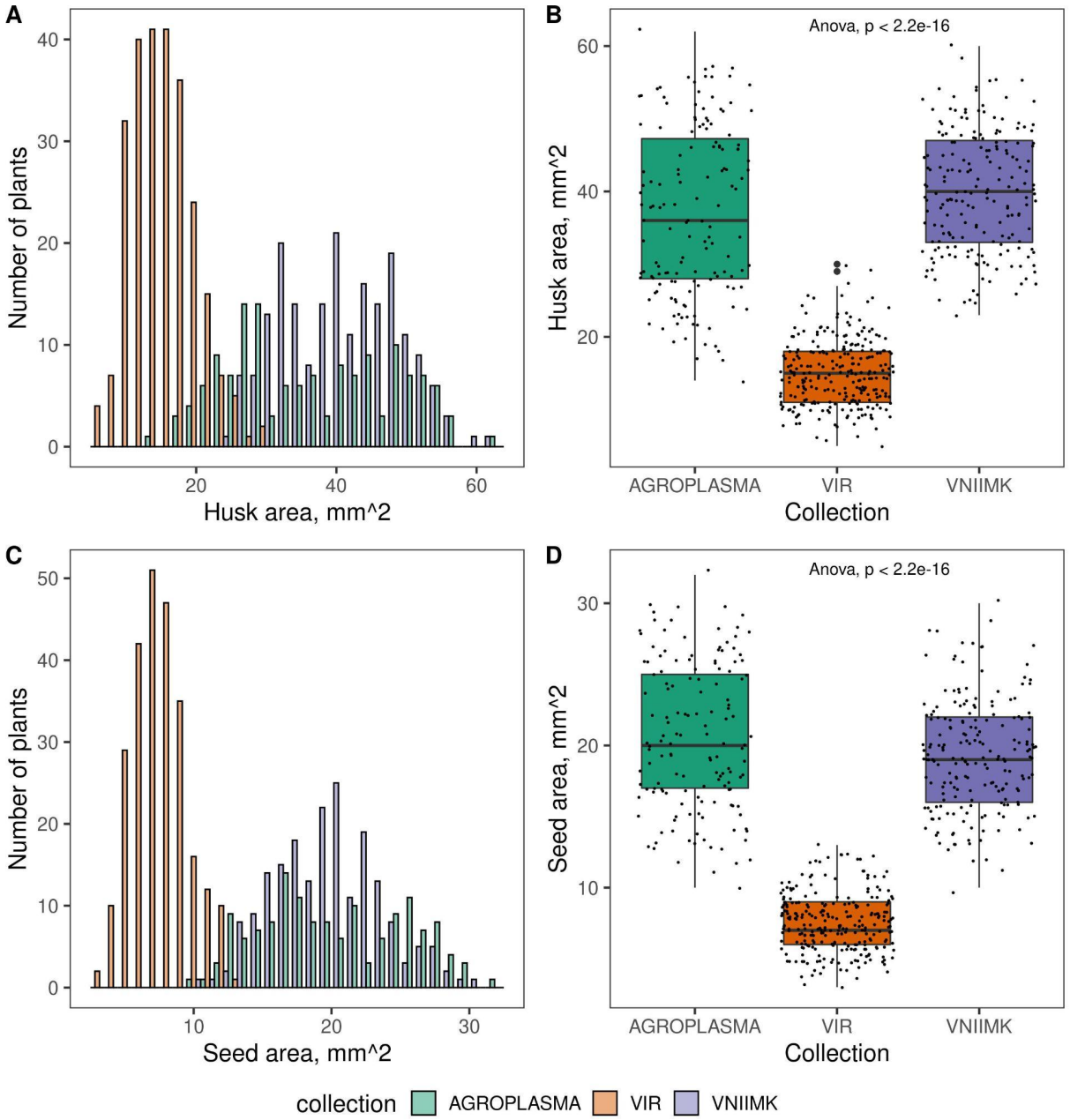

Figure 6.2.2 – Distribution of the husk (panel A and B) and seed (panels C and D) area between the collections. The color of the bars on histogram and box plots corresponds to the collection of the seed. Each dot corresponds to the mean value obtained for 10 seeds from the X-ray radiograph.

Next, a relationship between the husk and seed area was estimated by assessing the Pearson correlation coefficient. For the whole set of observations, the correlation between husk and seed area was significant (R = 0.93, p-value < 2.2e-16). However, it differed while being assessed separately for three collections (Figure 6.2.3)

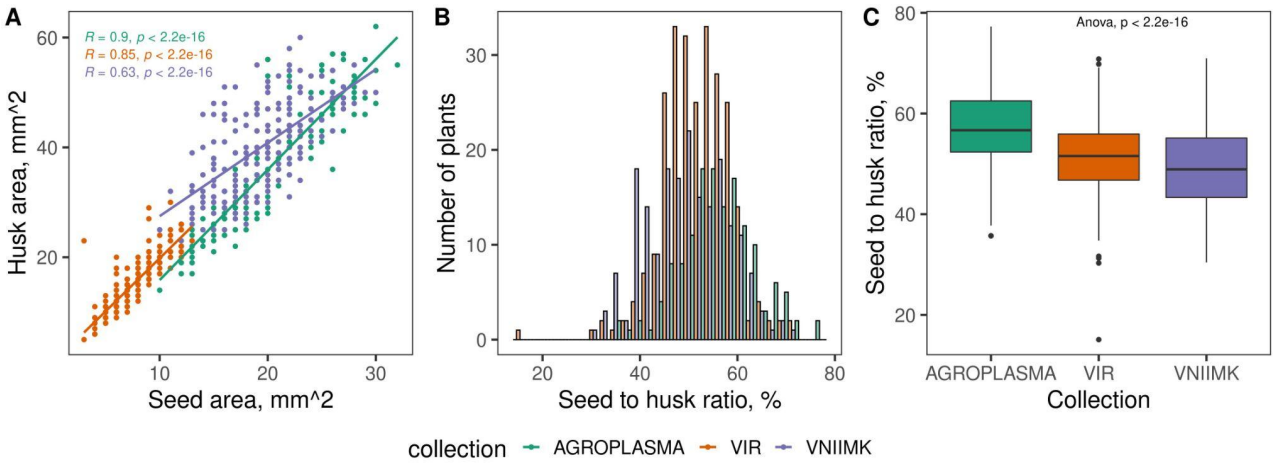

Figure 6.2.3 – Correlation between seed and husk area (panel A), distribution of the seed to husk ratio among collections. The color of regression lines, histogram bars and box plots correspond to the collection of the seed.

Additionally, the seed to husk ratio differed between collections (ANOVA p-value < 2.2e-16). Here an Agroplasma collection demonstrated the highest value of 57.10 which was significantly larger than the VIR sample by 1.11 times (Mann–Whitney U test p-value = 2.428e-13) and VNIIMK samples by 1.16 times (Mann–Whitney U test p-value = 4.651e-16).

Such differences in seed and husk sizes could be explained by the fact that the VIR collection mostly consists of historical samples, while VNIIMK and AGROPLASMA collections include economically valuable accessions used to produce commercial hybrids and lines for which the size of the seed is one of the key traits to be improved. Thus, the differences are explained by the artificial selection pressure different for the studied collections. The same could be applied to the difference in the seed to husk ratio the highest value was specific for the AGROPLASMA collection which also speaks in favor of the particular interest of increasing the seed area compared to the husk area so that the volume of the kernel was high. The performed analysis established that the collection of the accession is a

key factor affecting the seed-related traits. Thus, this factor is used as a covariate for subsequent association mapping.

6.3 Assessment of genetic diversity

To describe genetic diversity and perform mapping of the agronomically important traits, 15068 SNPs were obtained for 601 accessions from three collections by means of GBS followed by GATK pipeline analysis. As it was the first study that includes a substantial set of the genetically diverse sunflower accession collection, a characterization of the population structure was performed (Figure 6.3.1).

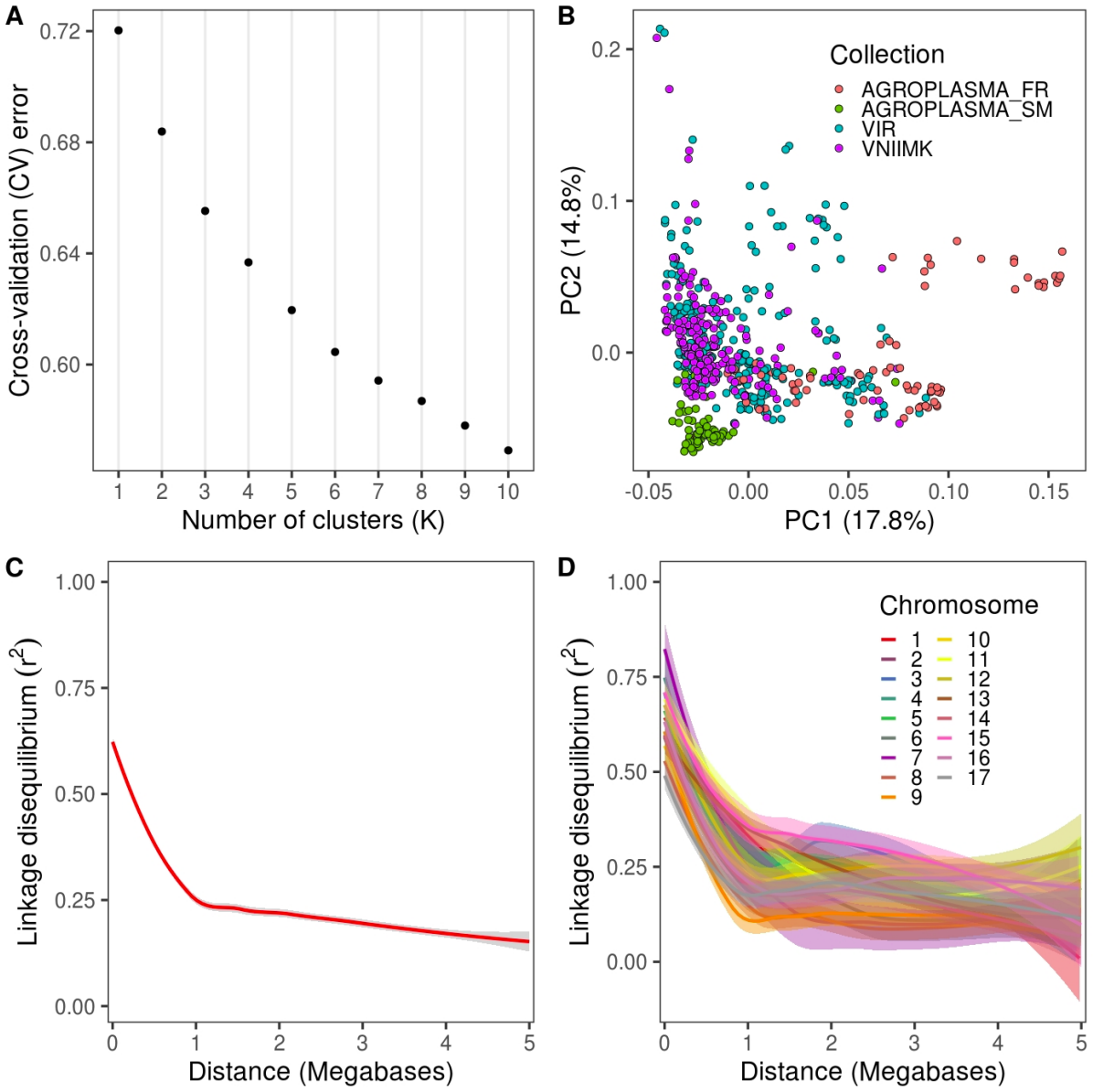

Figure 6.3.1 – Population structure of the 601 sunflower accession from VIR, VNIIMK and Agroplasma collections. Cross-validation error values for K ranging from 1 to 10 estimated by ADMIXTURE software (panel A). Visualization of first two principal components of PCA (panel B). Linkage disequilibrium decay across the whole genome (panel C) and separate chromosomes (panel D). FR and SM suffexes indicate fertility restorers and sterility maintainer lines, respectively in Agroplasma collection.

An ADMIXTURE and PCA analysis did not reveal clear population structure, neither between collections nor between phenotypically diverse lines. However, slight differences were identified for fertility restorer and fertility maintainer lines of the Agroplasma collection (Figure 6.3.1). Next, we assessed a Linkage disequilibrium decay for the whole genome as well as for the separate chromosomes. While the average genotype correlation (r^2) dropped to half of its maximum value at 0.7 Mb, linkage disequilibrium (LD) decay varied among the 17 chromosomes.

As the present study was the first one devoted to the genotyping of a significant amount of the accession from Russian collections the genetic differences with international collections that included wild representatives of the *Helianthus* genus, cultivated and landrace accessions were compared. To do so a principal component was performed on the 2345 SNPs shared between the collection of more than 1000 accessions (Hübner et al., 2019) for which an unfiltered VCF was provided by colleagues from the Biodiversity Research Centre of the University of British Columbia.

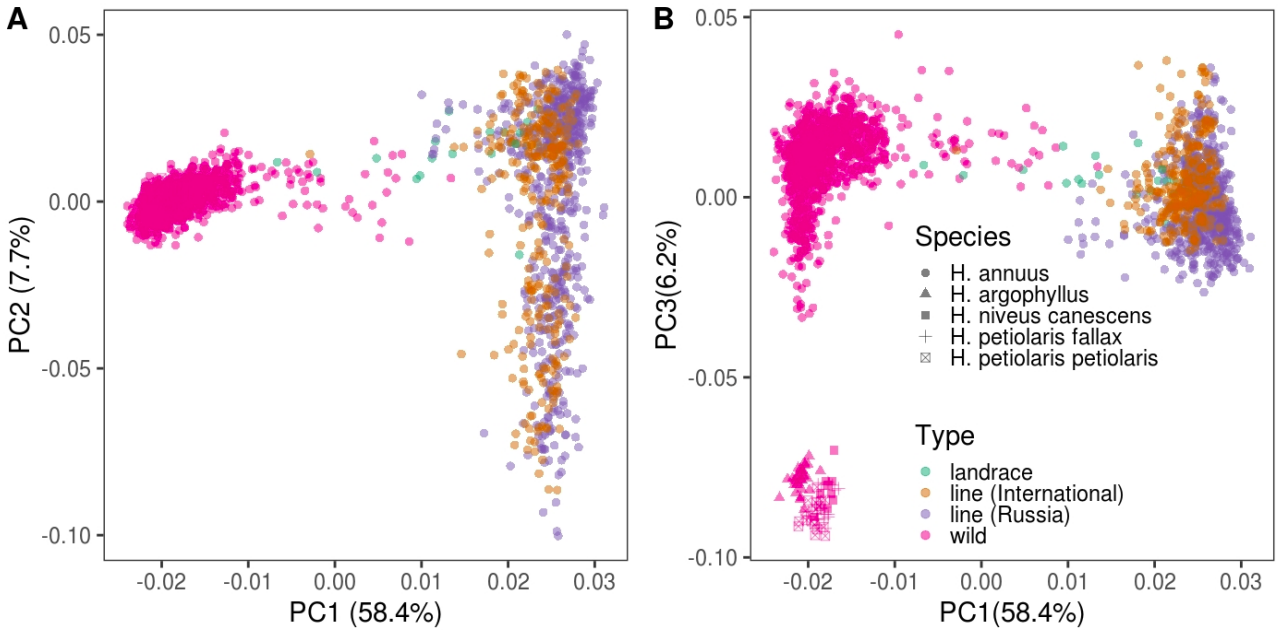

Figure 6.3.2 – Joint principal component analysis of sunflower accessions genotyped in this study and Hübner (2019) based on 2345 shared SNPs. The first and the second (A) or the first and the third (B) PCs are shown. Each dot corresponds to a plant accession. Shapes correspond to species. The origin (wild/line/landrace) is indicated by color.

The principal component analysis revealed a clear separation of wild and cultivated sunflowers by the first principal component explaining 58.4% of genotype variance. The first and third principal components revealed three clusters that consisted of wild *Helianthus annuus*, cultivated *Helianthus annuus*, and the third cluster was represented by other wild representatives of the *Helianthus* genus (Figure 6.3.2 B). The first three PCs demonstrated that the genetic diversities of the cultivated Russian and International sunflower accessions are matching.

6.4 Association mapping of seed-related traits and scanning for the candidate genes

As the information on the genotypes and phenotypes were collected, an association mapping of husk and seed areas as well as the seed-to-husk ratio was performed to find preliminary association mapping results. To do so, a compressed mixed linear model approach was applied within the TASSEL software. For each

trait, the first five principal components as well as the kinship matrix were added as confounding factors. Additionally, information on the collection was also added as a covariate as it significantly affected the distribution of all three traits.

As a result, only one SNP located on chromosome 10 associated with the husk area overcame the Bonferroni corrected threshold. Thus, a softer threshold of p=0.0001 was selected. As a result, we identified additional husk area-associated SNPs located on chromosomes 4, 9 and 17 (Figure 6.4.1). The proportion of the trait variance explained by these SNPs varied from 3.65 to 6.49 percent (Table 6.4.1).

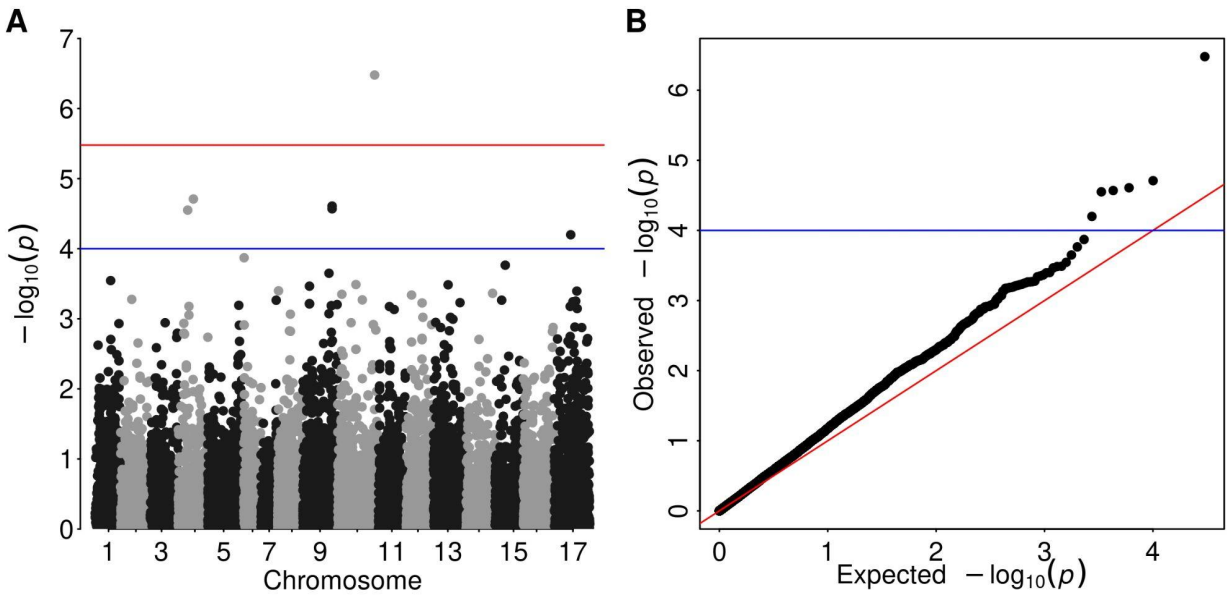

Figure 6.4.1 – Manhattan plot for association mapping of husk area results (panel A) and corresponding Q-Q plot (panel B). each dot corresponds to SNP. Black dots on panel A correspond to SNPs from odd chromosomes, gray dots correspond to SNPs from even chromosomes. The red horizontal line on panel A corresponds to the Bonferroni corrected p-value of 0.05. Blue horizontal lines on panels A and B correspond to a p-value threshold equal to 0.0001.

For the seed area, none of the SNPs overcame the Bonferroni multiple testing correction threshold. However, eight SNPs demonstrated a p-value below the softer

threshold of p=0.0001. These SNPs were located on chromosomes 2, 6, 9, 10, 11, 13 and 14 (Figure 6.4.2). The proportion of the trait variance explained by these SNPs varied from 3.82 to 5.52 percent.

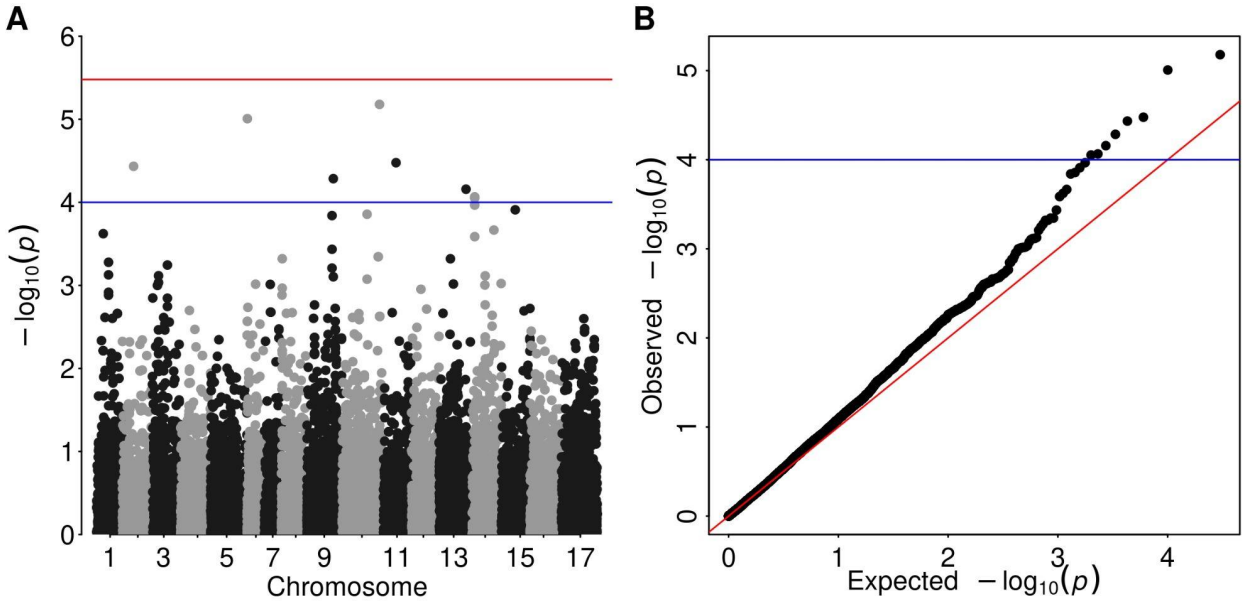

Figure 6.4.2 – Manhattan plot for association mapping of seed area results (panel A) and corresponding Q-Q plot (panel B). Each dot corresponds to the SNP. Black dots on panel A correspond to SNPs from odd chromosomes, gray dots correspond to SNPs from even chromosomes. The red horizontal line on panel A corresponds to the Bonferroni corrected p-value of 0.05. Blue horizontal lines on panels A and B correspond to a p-value threshold equal to 0.0001.

Finally, for the seed to husk ratio trait only one chromosome, chromosome 10, carried SNPs associated with the trait under the softer statistical threshold (Figure 6.4.3). The corresponding SNPs explained from 3.39 to 5.05 percent of trait phenotypic variation (Table 6.4.1).

Thus, the low amount of phenotypic variance explained by the markers was reported for all of the studied traits in the present study. The previous studies focusing on the association mapping of the seed size-related traits (hull length, hull width, hull area, kernel length, kernel width, seed weight) identified loci controlling these traits

that explained a higher proportion of phenotypic variance compared to the present study.

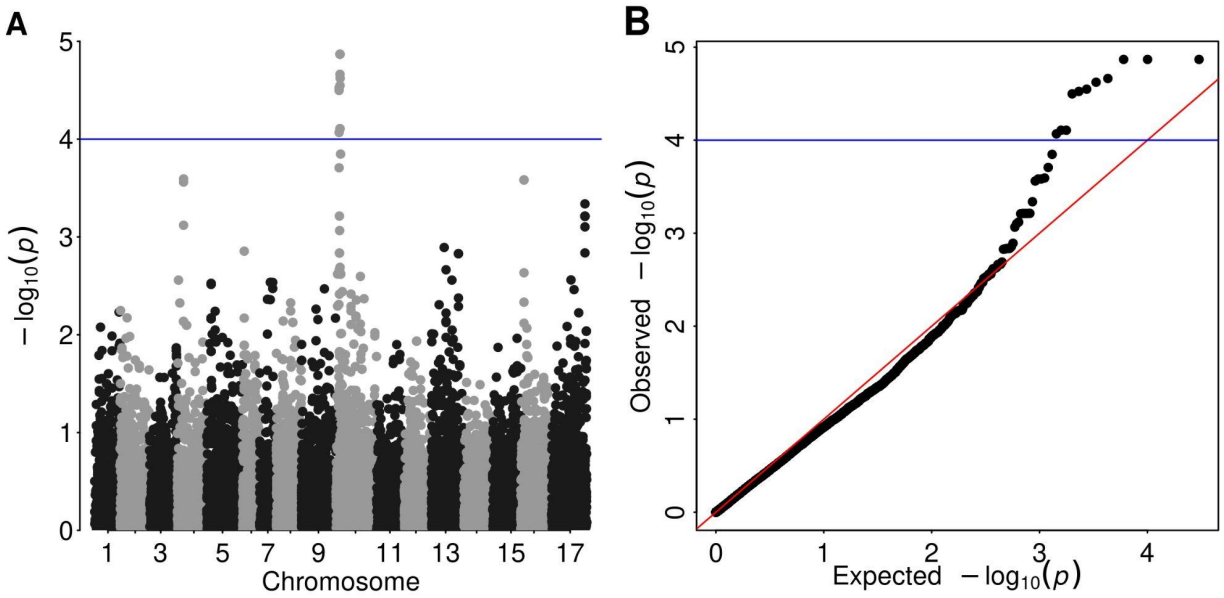

Figure 6.4.3 – Manhattan plot for association mapping of seed to husk ratio (panel A) and corresponding Q-Q plot (panel B). Each dot corresponds to the SNP. Black dots on panel A correspond to SNPs from odd chromosomes, gray dots correspond to SNPs from even chromosomes. Blue horizontal lines on panels A and B correspond to a p-value threshold equal to 0.0001.

A QTL mapping performed on the F2 mapping populations revealed seven genetic markers associated with 100-seed weight, seed (hull) length and seed (hull) width jointly explaining 56, 42 and 65 percent of phenotypic variance, respectively (Tang, Leon, et al., 2006). Another QTL study identified loci associated with the seed (hull) length, seed (hull) width, kernel length and kernel width explaining up to 14.8, 10.7, 13.5 and 23.2 percent of phenotypic variance, respectively. Associated loci were identified on chromosomes 1, 4, 7, 10, 13, 16 and 17 (Yue et al., 2009). A GWAS mapping of seed (hull) length, seed (hull) width and seed (hull) area based on sunflower diversity panels identified loci explaining up to 15, 85 and 47 percent of phenotypic variance, respectively. Loci associated with these traits were mapped on

chromosomes 10, 13 and 17 (Reinert et al., 2020). Additionally, it was previously demonstrated that the seed weight, which correlates well with the seed (hull) size, demonstrates a high level of heritability ranging from 79 to 89 percent in sunflower (Abdelsatar et al., 2020; Baraiya & Patel, 2018; A. Khan, 2001). Despite the fact that previously identified loci explained more variance, compared to the present study, they still explained less variation than expected considering the high rate of seed-related traits heritability. This discrepancy could be referred to as the missing heritability phenomenon discussed for plants previously (Brachi et al., 2011). Further, incomplete agreement between our current results and published studies, as well as among reported loci, indicates the need for further large-scale genetic studies of the seed-related phenotypes.

Table 6.4.1 – Summary of genetic markers associated with seed-related traits.

Marker	Trait	Chromosome	Position	p-value	PVE, %
S10_18982408	seed_husk_ratio	10	18982408	2.99E-05	4.33
S10_18982669	seed_husk_ratio	10	18982669	3.17E-05	4.21
S10_19187693	seed_husk_ratio	10	19187693	8.55E-05	3.39
S10_24409435	seed_husk_ratio	10	24409435	1.35E-05	4.85
S10_24409447	seed_husk_ratio	10	24409447	1.35E-05	4.85
S10_24409448	seed_husk_ratio	10	24409448	1.35E-05	4.85
S10_24409452	seed_husk_ratio	10	24409452	2.17E-05	5.05
S10_24409622	seed_husk_ratio	10	24409622	7.82E-05	4.32
S10_24409623	seed_husk_ratio	10	24409623	7.82E-05	4.32
S10_24409658	seed_husk_ratio	10	24409658	2.82E-05	4.57
S10_25468639	seed_husk_ratio	10	25468639	2.38E-05	4.56
S04_56847151	husk	4	56847151	2.81E-05	4.22

S04_91194852	husk	4	91194852	1.95E-05	3.65
S09_175848372	husk	9	175848372	2.46E-05	3.68
S09_175848377	husk	9	175848377	2.69E-05	3.66
S10_222897158	husk	10	222897158	3.32E-07	6.49
S17_99160416	husk	17	99160416	6.31E-05	3.78
S02_67897490	seed	2	67897490	3.68E-05	4.22
S06_1852907	seed	6	1852907	9.83E-06	4.95
S09_157078177	seed	9	157078177	5.18E-05	4.15
S10_222897158	seed	10	222897158	6.61E-06	5.52
S11_76005476	seed	11	76005476	3.33E-05	4.51
S13_159859510	seed	13	159859510	6.95E-05	4.09
S14_14459174	seed	14	14459174	8.86E-05	3.82
S14_14459481	seed	14	14459481	8.59E-05	3.84

Later we performed a scanning for the causal genes within the associated loci. To do so, we selected firstly genes located within 100kb intervals upstream and downstream the associated markers , as SNP pairs located within this frame demonstrated a high level ($r^2 > 0.5$) of linkage (Figure 6.3.1). As a result, more than 100 genes located within the selected frame were identified. Most of the genes encoded proteins of unknown functions, however some of the genes included meaningful annotation (Table 6.4.2) and encoded proteins that could be potentially involved in the seed size-related parameters.

Table 6.4.2 – Genes with annotation located within 100kb frame upstream and downstream significant SNPs.

Trait	SNP	GeneID	Product	Annotation	Distance
husk	S04_91194852	HannXRQ_Chr04g0109941	Putative cysteine-rich secretory protein allergen V5/Tpx-1-related	intergenic_region	0
husk	S09_175848377	ATMLH1	NA	downstream_gene_variant	23392
husk	S09_175848377	HannXRQ_Chr09g0266421	Putative UDP-glucuronosyl/UDP-glucosyltransferase	missense_variant	0
husk	S09_175848377	HannXRQ_Chr09g0266441	Putative cytochrome P450	downstream_gene_variant	99156
husk	S10_222897158	HannXRQ_Chr10g0312591	Putative cytochrome P450	upstream_gene_variant	22240
husk	S10_222897158	HannXRQ_Chr10g0312611	Putative transcription factor GRAS Scarecrow-like 29	downstream_gene_variant	37826
husk	S10_222897158	HannXRQ_Chr10g0312621	Putative germin	downstream_gene_variant	90730
husk	S17_99160416	HannXRQ_Chr17g0555021	Putative protein kinase-like domain, Leucine-rich repeat-containing N-terminal	downstream_gene_variant	84262
husk	S17_99160416	HannXRQ_Chr17g0555031	Probable S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	intergenic_region	0
husk	S17_99160416	HannXRQ_Chr17g0555031	Probable S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	upstream_gene_variant	14634
seed	S06_1852907	HannXRQ_Chr06g0164571	Probable translation elongation factor EF1B	downstream_gene_variant	47433
seed	S06_1852907	HannXRQ_Chr06g0164581	Probable ATP synthase D chain mitochondrial	downstream_gene_variant	40865
seed	S06_1852907	HannXRQ_Chr06g0164591	Putative tetraspanin/Peripherin	upstream_gene_variant	34142
seed	S06_1852907	HannXRQ_Chr06g0164601	Putative proton-dependent oligopeptide transporter family Major facilitator superfamily domain	downstream_gene_variant	8529
seed	S06_1852907	HannXRQ_Chr06g0164611	Putative proton-dependent oligopeptide transporter family Major facilitator superfamily domain	synonymous_variant	0
seed	S09_157078177	HannXRQ_Chr09g0261661	Putative tetratricopeptide-like helical domain	upstream_gene_variant	33065
seed	S09_157078177	HannXRQ_Chr09g0261671	Probable quinone reductase family protein	downstream_gene_variant	23147
seed	S09_157078177	HannXRQ_Chr09g0261681	Putative calcium/proton exchanger	upstream_gene_variant	19803
seed	S10_222	HannXRQ_Chr1	Putative dnaJ domain	downstream_gene	40006

	897158	0g0312581		_variant	
seed	S10_222897158	HannXRQ_Chr10g0312591	Putative cytochrome P450	upstream_gene_variant	22240
seed	S10_222897158	HannXRQ_Chr10g0312621	Putative germin	downstream_gene_variant	90730
seed	S11_76005476	HannXRQ_Chr11g0335701	Probable SAUR-like auxin-responsive protein family	upstream_gene_variant	3130
seed	S11_76005476	HannXRQ_Chr11g0335721	Putative meiotic nuclear division protein 1	downstream_gene_variant	53138
seed	S13_159859510	HannXRQ_Chr13g0416301	Putative cytochrome P450	downstream_gene_variant	52185
seed	S13_159859510	HannXRQ_Chr13g0416311	Probable peptidase M1 family protein	upstream_gene_variant	76515
seed	S14_14459481	BSK1	BR-signaling kinase 1	upstream_gene_variant	63202
seed_husk_ratio	S10_18982669	HannXRQ_Chr10g0281741	Putative nucleotide-binding alpha-beta plait domain	intergenic_region	0
seed_husk_ratio	S10_24409435	HannXRQ_Chr10g0282351	Putative ankyrin repeat-containing domain Gag-polypeptide of LTR copia-type	downstream_gene_variant	93932
seed_husk_ratio	S10_24409435	HannXRQ_Chr10g0282371	Putative ribosomal protein S11	downstream_gene_variant	6054
seed_husk_ratio	S10_24409447	HannXRQ_Chr10g0282361	Putative PGG domain	downstream_gene_variant	81280
seed_husk_ratio	S10_25468639	HannXRQ_Chr10g0282611	Putative protein kinase-like domain Leucine-rich repeat-containing N-terminal	upstream_gene_variant	33872

In particular, SNP S14_14459481 associated with the seed area was located 63kb away from gene BSK1 encoding BR-signaling kinase 1, a receptor that is involved in brassinisteroid signaling. Brassinosteroids are the plant hormones that regulate growth and development as well as seed formation (Clouse, 2011). S11_76005476 SNP associated with the seed area located 3kb upstream of the gene encoding SAUR-like auxin-responsive protein family SAURs were reported to act in auxin-responsive cell elongation (Stortenbeker & Bemer, 2019). Gene encoding probable Translation elongation factor EF1B was located 47kb downstream the S06_1852907 associated with seed area. Previously it was demonstrated that EF1B is involved in the control of cell wall biosynthesis in thale cress (Hossain et al., 2012). For the husk area, an SNP S10_222897158 was located close to the gene encoding

Putative transcription factor GRAS Scarecrow-like 29. GRAS Transcription Factors were shown to affect cell growth via gibberellin signaling (Cenci & Rouard, 2017). Unfortunately, previous studies aimed at scanning genetic markers associated with seed size-related traits were performed at a low mapping resolution not allowing for to identification of candidate genes. The only exception was the most recent study (Reinert et al., 2020), which identified 11 genes as strong candidates for trait control. Among them were genes encoding proteins involved in ubiquitination, glycosyltransferases, and sulfate anion transporters.

6.5 Conclusions

This part of the study lists only preliminary results of the association mapping seed-related traits as this project has been initiated, a new genome assembly of sunflower has been released and became publicly available. Also additional phenotype measurements during different years and/or locations are needed to estimate the genotype-environment interactions for such complex traits. Nevertheless, these preliminary analysis provided information on potential candidate genes associated with seed-related traits. To perform mapping, first, information on seed-related traits was collected by means of X-ray radiography followed by image analysis. Phenotype data analysis revealed significant differences in trait distribution between the studied collections, which could be explained by the different breeding directions applied to them. Second, genetic diversity analysis of 601 sunflower accessions from Russian genetic collections and comparison of the genetic diversity with the international lines were performed. It was established that the diversity of cultivated sunflower lines covers the diversity of the cultivated lines from other international collections. Additionally, a significant genetic difference between wild and cultivated sunflowers, independent of the source collections, was identified. Association mapping revealed 25 SNP loci associated with the seed size-related traits. Further, scanning the genetic regions carrying associated markers reveal several potential candidate genes that could be involved in the expression of the analyzed

traits. It should be noted that the identified genetic markers explained not more than seven percent of phenotypic variance, and thus could not be used for the development of marker-assisted approaches. Thus, since the studied traits apparently are under complex polygenic control, the genomic selection approaches could be more applicable to this trait than marker-assisted ones.

Chapter 7. Conclusions and future perspectives

This study presents an effort to identify genetic markers associated with valuable traits in rapeseed and sunflower crops. Further, as this study included, for the first time, full genome genotyping of a substantial number of diverse rapeseed and sunflower lines grown in Russia, I further performed a population structure analysis and diversity comparison of the Russian plan material with those used worldwide.

The main conclusions of the work are as follows:

1) Association mapping of the glucosinolate content in rapeseed revealed 7 SNP markers that explained up to 23% of the phenotypic variance. The identified SNP markers are located close to the new candidate genes that could be potentially involved in the control of glucosinolate composition.
2) The QTL mapping of oil quality traits in Russian plant material identified novel SNP markers in sunflower for previously reported *Tph1*, *Tph2,* and *Ol* loci associated with tocopherol composition and oleic acid content, respectively. The respective genetic markers were validated using independent plant samples.
3) Preliminary association mapping of seed, husk and seed to husk area traits conducted based on 601 Russian sunflower lines revealed 25 SNP markers associated with the seed-related traits. A candidate gene analysis revealed new candidate genes involved in the seed-related trait expression.
4) Analysis of the population structure of VNIIMK collection of the rapeseed revealed the strong population clustering, which is partially explained by the phenotypical and phenological features. Comparative analysis of the genetic diversity revealed genetic differences potentially underlying the separation between the international and VNIIMK accessions.
5) Analysis of the genetic diversity of the sunflower accessions from VNIIMK, VIR and AGROPLASMA collections did not reveal any significant population structure. Further, a joint population structure analysis of a large collection of

international and Russian accessions from VNIIMK, VIR and AGROPLASMA collections identified no significant genetic structure differences between cultivated accession from International and Russian accession. However, we observed significant genetic differences between the wild and cultivated sunflowers independent of the collection origin.

The genetic markers of glucosinolate content could be used to develop approaches for marker-assisted selection for oil and oilcake quality improvement right after the validation of these markers on the independent plant sample of rapeseeds to test if these markers are polymorphic and associated with glucosinolate content in other collections. Currently identified genetic markers for glucosinolate could be used to facilitate the transfer of low glucosinolate trait within the studied VNIIMK collection. Nevertheless additional validation by means of different methods, for example, allele-specific PCRs and/or Sanger sequencing is also needed.

Genetic markers associated with *Tph1* and *Tph2* were validated by two independent approaches and could be used to develop PCR-based systems for MAS with the aim of tocopherol composition prediction. Markers associated with the *Ol* should be additionally validated on the VNIIMK lines that will be used to obtain varieties with high oleic acid content. Thus, the markers for *Ol, Tph1* and *Tph2* could be used for example for the introduction of oil quality in the lines by means of marker-assisted backcrossing. To control oil quality in sunflower as well as glucosinolate content different approaches could be used including allele-specific PCR, for example, Real-Time PCR or LAMP-PCR. Additionally, solutions based on sequencing including sanger sequencing and high-throughput sequencing of the amplicons corresponding to the regions that carry associated polymorphisms could be applied. These methods could be used to trace the trait-associated markers for germplasm evaluation and/or trace the trait during its introduction by marker-assisted backcrossing.

Genetic markers associated with the seed-related traits could further provide a better understanding of these traits' genetic control. However, our study suggests that considering the low amount of the seed phenotypic variance explained by the markers, an alternative approach such as genomic selection should be applied to predict and select the proper candidate plants for seed-related trait breeding. Finally, all information on the genetic diversity obtained for rapeseed in sunflower lines could be used as a resource for selection scheme design and genetic diversity analysis by the breeders, as all genotyping data associated with the project is freely available.

Bibliography

Adiredjo, A. L., Navaud, O., Muños, S., Langlade, N. B., Lamaze, T., & Grieu, P. (2014). Genetic Control of Water Use Efficiency and Leaf Carbon Isotope Discrimination in Sunflower (Helianthus annuus L.) Subjected to Two Drought Scenarios. *PLOS ONE*, *9*(7), e101218. https://doi.org/10.1371/journal.pone.0101218

Ahmar, S., Liaqat, N., Hussain, M., Salim, M. A., Shabbir, M. A., Ali, M. Y., Noushahi, H. A., Bilal, M., Atta, B., & Rizwan, M. (2019). Effect of Abiotic Stresses on Brassica Species and Role of Transgenic Breeding for Adaptation. *Asian Journal of Research in Crop Science*, 1–10. https://doi.org/10.9734/ajrcs/2019/v3i130037

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Alexander, J., Barregård, L., Bignami, M., Brüschweiler, B., Ceccatelli, S., Dinovi, M., Edler, L., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L. (Ron), Nebbia, C. S., Oswald, I., Petersen, A., Rose, M., Roudot, A.-C., Schwerdtle, T., Vollmer, G., Wallace, H., Cottrill, B., … Vleminckx, C. (2016). Erucic acid in feed and food. *EFSA Journal*, *14*(11), e04593. https://doi.org/10.2903/j.efsa.2016.4593

Al-Khatib, K., Baumgartner, J. R., Peterson, D. E., & Currie, R. S. (1998). Imazethapyr resistance in common sunflower (Helianthus annuus). *Weed Science*, *46*(4), 403–407. https://doi.org/10.1017/S0043174500090809

Al-Turki, T. A., & Baskin, C. C. (2017). Determination of seed viability of eight wild Saudi Arabian species by germination and X-ray tests. *Saudi Journal of Biological Sciences*, *24*(4), 822–829. https://doi.org/10.1016/j.sjbs.2016.06.009

Anisimova, & Dubovskaya, A. G. (2020). CMS systems in rapeseed and their use in the breeding of domestic hybrids. *Proceedings on Applied Botany, Genetics and Breeding*, *181*(3), 171–180. https://doi.org/10.30901/2227-8834-2020-3-171-180

Anisimova, I., Карабицина, Ю., Alpatieva, N., Kusnetsova, E., Titov, N., Lyutko, A., & Gavrilova, V. (2021). Diagnostic value of Rf1 gene molecular markers in sunflower. *Plant Biotechnology and Breeding*, *4*. https://doi.org/10.30901/2658-6266-2021-2-o3

Ansan-Melayah, D., Balesdent, M. H., Delourme, R., Pilet, M. L., Tanguy, X., Renard, M., & Rouxel, T. (1998). Genes for race-specific resistance against blackleg disease in Brassica napus L. *Plant Breeding*, *117*(4), 373–378. https://doi.org/10.1111/j.1439-0523.1998.tb01956.x

Arkhipov, M. V., Priyatkin, N. S., Gusakova, L. P., Potrakhov, N. N., Gryaznov, A. Yu., Bessonov, V. B., Obodovskii, A. V., & Staroverov, N. E. (2019). X-Ray Computer Methods for Studying the Structural Integrity of Seeds and Their Importance in Modern Seed Science. *Technical Physics*, *64*(4), 582–592. https://doi.org/10.1134/S1063784219040030

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., Prohens, J., & Gramazio, P. (2020). The Dawn of the Age of Multi-Parent MAGIC Populations in Plant Breeding: Novel Powerful Next-Generation Resources for Genetic Analysis and Selection of Recombinant Elite Material. *Biology*, *9*(8), 8. https://doi.org/10.3390/biology9080229

Assefa, Y., Prasad, P. V. V., Foster, C., Wright, Y., Young, S., Bradley, P., Stamm, M., & Ciampitti,

I. A. (2018). Major Management Factors Determining Spring and Winter Canola Yield in North America. *Crop Science*, *58*(1), 1–16. https://doi.org/10.2135/cropsci2017.02.0079

Attia, Z., Pogoda, C. S., Reinert, S., Kane, N. C., & Hulke, B. S. (2021). Breeding for sustainable oilseed crop yield and quality in a changing climate. *Theoretical and Applied Genetics*, *134*(6), 1817–1827. https://doi.org/10.1007/s00122-021-03770-w

Bachlava, E., Tang, S., Pizarro, G., Schuppert, G. F., Brunick, R. K., Draeger, D., Leon, A., Hahn, V., & Knapp, S. J. (2010). Pleiotropy of the branching locus (B) masks linked and unlinked quantitative trait loci affecting seed traits in sunflower. *Theoretical and Applied Genetics*, *120*(4), 829–842. https://doi.org/10.1007/s00122-009-1212-1

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N. C., Bowers, J. E., Hubner, S., Bellec, A., Bérard, A., Bergès, H., Blanchet, N., … Langlade, N. B. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, *546*(7656), 7656. https://doi.org/10.1038/nature22380

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Balesdent, M. H., Attard, A., Ansan-Melayah, D., Delourme, R., Renard, M., & Rouxel, T. (2001). Genetic Control and Host Range of Avirulence Toward Brassica napus Cultivars Quinta and Jet Neuf in Leptosphaeria maculans. *Phytopathology®*, *91*(1), 70–76. https://doi.org/10.1094/PHYTO.2001.91.1.70

Ben-Ari, G., & Lavi, U. (2012). Marker-assisted selection in plant breeding. In *Plant Biotechnology and Agriculture* (pp. 163–184). Elsevier. https://doi.org/10.1016/B978-0-12-381466-1.00011-0

Berry, S. T., Leon, A. J., Hanfrey, C. C., Challis, P., Burkholz, A., Barnes, S. R., Rufener, G. K., Lee, M., & Caligari, P. D. (1995). Molecular marker analysis of Helianthus annuus L. 2. Construction of an RFLP linkage map for cultivated sunflower. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *91*(2), 195–199. https://doi.org/10.1007/BF00220877

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M., Jain, N., Singh, P. K., Singh, G. P., & Prabhu, K. V. (2016). Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. *Frontiers in Genetics*, *7*. https://www.frontiersin.org/article/10.3389/fgene.2016.00221

Bochkaryova, E. B., Gorlova, L. A., Strelkov, E. A., & Serdyk, V. V. (2021). First Russian hybrid of winter rapeseed Debyut. *Oilseed crops*, *2 (186)*. https://doi.org/10.25230/2412-608X-2021-2-186-98-100

Bochkaryova, Gorlova, L. A., Serdyuk, V. V., & Strelnikov, E. A. (2020). Breeding of winter rapeseed inbred lines for the development of parent forms for hybrids. *Proceedings on applied botany, genetics and breeding*, *180*(4), 4. https://doi.org/10.30901/2227-8834-2019-4-121-125

Bochkaryova, Strelnikov, Serdyk, B. B., & Gorlova, L. A. (2021). Breeding of winter rapeseed in VNIIMK: history and new results (review). *Oilseed crops*, *4 (188)*. https://doi.org/10.25230/2412-608X-2021-4-188-87-95

Bochkovoy, Khatnyansky, V. I., Kamardin, V. A., & Nazarov, D. A. (2020). Confectionary sunflower: Origin, history, struc- ture, breeding directions and features of cultiva- tion technology (review). *Oilseed crops*, *3 (183)*. https://doi.org/10.25230/2412-608X-2020-3-183-129-146

Bohinc, T., & Trdan, S. (2012). Environmental factors affecting the glucosinolate content in Brassicaceae. *Journal of Food Agriculture and Environment*, *10*, 357–360.

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biology*, *12*(10), 232. https://doi.org/10.1186/gb-2011-12-10-232

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brandle, J. E., & McVetty, P. B. E. (1990). Geographical diversity, parental selection and heterosis in oilseed rape. *Canadian Journal of Plant Science*, *70*(4), 935–940. https://doi.org/10.4141/cjps90-115

Broccanello, C., Chiodi, C., Funk, A., McGrath, J. M., Panella, L., & Stevanato, P. (2018). Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods*, *14*(1), 28. https://doi.org/10.1186/s13007-018-0295-6

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Brun, H., Chèvre, A.-M., Fitt, B. D. L., Powers, S., Besnard, A.-L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M., & Andrivon, D. (2010). Quantitative resistance increases the durability of qualitative resistance to Leptosphaeria maculans in Brassica napus. *The New Phytologist*, *185*(1), 285–299. https://doi.org/10.1111/j.1469-8137.2009.03049.x

Bu, S., Wu, W., & Zhang, Y.-M. (2021). A Multi-Locus Association Model Framework for Nested Association Mapping With Discriminating QTL Effects in Various Subpopulations. *Frontiers in Genetics*, *11*. https://www.frontiersin.org/article/10.3389/fgene.2020.590012

Bulos, M. (2013). Marker assisted selection for herbicide resistance in sunflower. *Helia*, *36*, 1–16. https://doi.org/10.2298/HEL1359001B

Bus, A., Körber, N., Snowdon, R. J., & Stich, B. (2011). Patterns of molecular variation in a species-wide germplasm set of Brassica napus. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *123*(8), 1413–1423. https://doi.org/10.1007/s00122-011-1676-7

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*, *8*(12), e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*(4), 855–867. https://doi.org/10.1111/1755-0998.12357

Celik, I., Bodur, S., Frary, A., & Doganlar, S. (2016). Genome-wide SNP discovery and genetic linkage map construction in sunflower (Helianthus annuus L.) using a genotyping by sequencing (GBS) approach. *Molecular Breeding*, *36*(9), 133.

https://doi.org/10.1007/s11032-016-0558-8

Cenci, A., & Rouard, M. (2017). Evolutionary Analyses of GRAS Transcription Factors in Angiosperms. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/article/10.3389/fpls.2017.00273

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Silva, C. D., Just, J., Falentin, C., Koh, C. S., Clainche, I. L., Bernard, M., Bento, P., Noel, B., … Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science*, *345*(6199), 950–953. https://doi.org/10.1126/science.1253435

Chelyustnikova, T. A., Guchetl, S. Z., & Antonova, T. S. (2017). Usage of molecular markers for identification of CMS-Rf system in parental lines of sunflower hy brids. *Oilseed crops*, *4 (172)*. https://elibrary.ru/item.asp?id=32398363

Chernova, A. I., Gubaev, R. F., Singh, A., Sherbina, K., Goryunova, S. V., Martynova, E. U., Goryunov, D. V., Boldyrev, S. V., Vanyushkina, A. A., Anikanov, N. A., Stekolshchikova, E. A., Yushina, E. A., Demurin, Y. N., Mukhina, Z. M., Gavrilova, V. A., Anisimova, I. N., Karabitsina, Y. I., Alpatieva, N. V., Chang, P. L., … Nuzhdin, S. V. (2021). Genotyping and lipid profiling of 601 cultivated sunflower lines reveals novel genetic determinants of oil fatty acid content. *BMC Genomics*, *22*(1), 505. https://doi.org/10.1186/s12864-021-07768-y

Christiansen, H., Heindler, F. M., Hellemans, B., Jossart, Q., Pasotti, F., Robert, H., Verheye, M., Danis, B., Kochzius, M., Leliaert, F., Moreau, C., Patel, T., Van de Putte, A. P., Vanreusel, A., Volckaert, F. A. M., & Schön, I. (2021). Facilitating population genomics of non-model organisms through optimized experimental design for reduced representation sequencing. *BMC Genomics*, *22*(1), 625. https://doi.org/10.1186/s12864-021-07917-3

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Clouse, S. D. (2011). Brassinosteroids. *The Arabidopsis Book / American Society of Plant Biologists*, *9*, e0151. https://doi.org/10.1199/tab.0151

Crites, G. D. (1993). Domesticated Sunflower in Fifth Millennium B.P. Temporal Context: New Evidence from Middle Tennessee. *American Antiquity*, *58*(1), 146–148. https://doi.org/10.2307/281459

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Cvejić, S., Miladinović, D., & Jocić, S. (2014). *Mutation breeding for changed oil quality in sunflower* (pp. 77–96). https://doi.org/10.3920/978-90-8686-796-7_4

de Oliveira Filho, J. G., & Egea, M. B. (2021). Sunflower seed byproduct and its fractions for food application: An attempt to improve the sustainability of the oil process. *Journal of Food Science*, *86*(5), 1497–1510. https://doi.org/10.1111/1750-3841.15719

Debaeke, P., Casadebaig, P., Flenet, F., & Langlade, N. (2017). Sunflower crop and climate change: Vulnerability, adaptation, and mitigation potential from case-studies in Europe. *OCL*, *24*(1), 1. https://doi.org/10.1051/ocl/2016052

Dehmer, K. J., & Friedt, W. (1998). Development of molecular markers for high oleic acid content in sunflower (Helianthus annuus L.). *Industrial Crops and Products*, *7*(2), 311–315. https://doi.org/10.1016/S0926-6690(97)00063-0

Delourme, R., Bouchereau, A., Hubert, N., Renard, M., & Landry, B. S. (1994). Identification of RAPD markers linked to a fertility restorer gene for theOgura radish cytoplasmic male sterility of rapeseed (Brassica napus L.). *Theoretical and Applied Genetics*, *88*(6), 741–748. https://doi.org/10.1007/BF01253979

Delourme, R., Bousset, L., Ermel, M., Duffé, P., Besnard, A. L., Marquer, B., Fudal, I., Linglin, J., Chadœuf, J., & Brun, H. (2014). Quantitative resistance affects the speed of frequency increase but not the diversity of the virulence alleles overcoming a major resistance gene to Leptosphaeria maculans in oilseed rape. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *27*, 490–499. https://doi.org/10.1016/j.meegid.2013.12.019

Demurin, Bochkarev, N., & Borisenko, O. (2016). High oleic sunflower hybrid Oxy with changed seed tocopherol content. *Proceedings of 19th International Sunflower Conference*, 3.

Demurin, Borisenko, O. M., Peretyagina, T. M., & Perstenyeva, A. A. (2006). Gene linkage test for imr with ol, tph1 and tph2 mutations in sunflower / Prueba de vinculación del gen imr con mutaciones ol, tph1 y tph2 en girasol / test de liaison du gène imr avec les mutations ol, tph1 et tph2 chez le tournesol. *HELIA*, *29*(44), 41–46. https://doi.org/10.2298/hel0644041d

Demurin, Efimenko, S. G., & Peretyagina, T. M. (2006). Expressivity of tocopherol mutations in sunflower. *Helia*, *29*(45), 55–62. https://doi.org/10.2298/HEL0645055D

Demurin, Savchenko, V. D., Borisenko, O. M., Levutskaya, A. N., Tolmachyova, N. N., Chebanova, Yu. V., Rubanova, O. A., Ryzhenko, E. N., Antonova, T., & Araslanova, N. (2020). Broomrape-resistant sunflower hybrid Taizar. *Oil Crops*, *184*, 87–90. https://doi.org/10.25230/2412-608X-2020-4-184-87-90

Demurin, & Skoric, D. (1996). Unstable expression of Ol gene for high oleic acid content in sunflower seeds. *Proc. 14th International Sunflower Conference, Beijing/Shenyang, 12-20 June 1996, China*, 145–150.

Demurin, Y. (1993). *Genetic variability of tocopherol composition insunflower seeds*. *16*, 59–62.

Demurin, Y. N., Efimenko, S. G., & Peretyagina, T. M. (2004). Genetic identification of tocopherol mutations in sunflower / Identificación genética de las mutaciones tocoferólicas en girasol / identification génétique de la mutation de tocophérol dans la culture de tournesol. *HELIA*, *27*(40), 113–116. https://doi.org/10.2298/hel0440113d

Demurin, Y., Skoric, D., & Karlovic, D. (1996). Genetic variability of tocopherol composition in sunflower seeds as a basis of breeding for improved oil quality. *Plant Breeding*, *115*(1), 33–36. https://doi.org/10.1111/j.1439-0523.1996.tb00867.x

Diederichsen, E., Beckmann, J., Schondelmeier, J., & Dreyer, F. (2006). Genetics of clubroot resistance in Brassica napus “Mendel”. *Acta Horticulturae*, *No.706*, 307–311.

Dimitrijevic, A., & Horn, R. (2018). Sunflower Hybrid Breeding: From Markers to Genomic Selection. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/article/10.3389/fpls.2017.02238

Dumont, J., Hirvonen, T., Heikkinen, V., Mistretta, M., Granlund, L., Himanen, K., Fauch, L., Porali, I., Hiltunen, J., Keski-Saari, S., Nygren, M., Oksanen, E., Hauta-Kasari, M., & Keinänen, M. (2015). Thermal and hyperspectral imaging for Norway spruce (Picea abies)

seeds screening. *Computers and Electronics in Agriculture*, *116*, 118–124. https://doi.org/10.1016/j.compag.2015.06.010

Ecke, W., Uzunova, M., & Weißleder, K. (1995). Mapping the genome of rapeseed (Brassica napus L.). II. Localization of genes controlling erucic acid synthesis and seed oil content. *Theoretical and Applied Genetics*, *91–91*(6–7), 972–977. https://doi.org/10.1007/BF00223908

Egan, A. N., Schlueter, J., & Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. *American Journal of Botany*, *99*(2), 175–185. https://doi.org/10.3732/ajb.1200020

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

El-Soda, M., Malosetti, M., Zwaan, B. J., Koornneef, M., & Aarts, M. G. M. (2014). Genotype×environment interaction QTL mapping in plants: Lessons from Arabidopsis. *Trends in Plant Science*, *19*(6), 390–398. https://doi.org/10.1016/j.tplants.2014.01.001

Embaye, W. T., Bergtold, J. S., Archer, D., Flora, C., Andrango, G. C., Odening, M., & Buysse, J. (2018). Examining farmers' willingness to grow and allocate land for oilseed crops for biofuel production. *Energy Economics*, *71*, 311–320. https://doi.org/10.1016/j.eneco.2018.03.005

Fernández-Aparicio, M., del Moral, L., Muños, S., Velasco, L., & Pérez-Vich, B. (2022). Genetic and physiological characterization of sunflower resistance provided by the wild-derived OrDeb2 gene against highly virulent races of Orobanche cumana Wallr. *Theoretical and Applied Genetics*, *135*(2), 501–525. https://doi.org/10.1007/s00122-021-03979-9

Fernández-Martínez, J., Jimenez, A., Dominguez, J., Garcia, J. M., Garces, R., & Mancha, M. (1989). Genetic analysis of the high oleic acid content in cultivated sunflower (Helianthus annuus L.). *Euphytica*, *41*(1), 39–51. https://doi.org/10.1007/BF00022409

Fernández-Martínez, J. M., Pérez-Vich, B., Velasco, L., & Domínguez, J. (2007). Breeding for specialty oil types in sunflower. *Helia*, *30*(46), 75–84.

Fernando, R. L., & Garrick, D. (2013). Bayesian Methods Applied to GWAS. In C. Gondro, J. van der Werf, & B. Hayes (Eds.), *Genome-Wide Association Studies and Genomic Prediction* (pp. 237–274). Humana Press. https://doi.org/10.1007/978-1-62703-447-0_10

Ferreira, M. E., Williams, P. H., & Osborn, T. C. (1994). RFLP mapping of Brassica napus using doubled haploid lines. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *89*(5), 615–621. https://doi.org/10.1007/BF00222456

Fragoso, C. A., Heffelfinger, C., Zhao, H., & Dellaporta, S. L. (2016). Imputing Genotypes in Biallelic Populations from Low-Coverage Sequence Data. *Genetics*, *202*(2), 487–495. https://doi.org/10.1534/genetics.115.182071

Fredua-Agyeman, R., & Rahman, H. (2016). Mapping of the clubroot disease resistance in spring Brassica napus canola introgressed from European winter canola cv. 'Mendel.' *Euphytica*, *211*(2), 201–213. https://doi.org/10.1007/s10681-016-1730-2

Fridrihsone, A., Romagnoli, F., & Cabulis, U. (2020). Environmental Life Cycle Assessment of Rapeseed and Rapeseed Oil Produced in Northern Europe: A Latvian Case Study. *Sustainability*, *12*(14), 14. https://doi.org/10.3390/su12145699

Fritsche, S., Wang, X., Li, J., Stich, B., Kopisch-Obuch, F. J., Endrigkeit, J., Leckband, G., Dreyer,

F., Friedt, W., Meng, J., & Jung, C. (2012). A Candidate Gene-Based Association Study of Tocopherol Content and Composition in Rapeseed (Brassica napus). *Frontiers in Plant Science*, *3*, 129. https://doi.org/10.3389/fpls.2012.00129

Fu, F., Zhang, X., Liu, F., Peng, G., Yu, F., & Fernando, D. (2020). Identification of resistance loci in Chinese and Canadian canola/rapeseed varieties against Leptosphaeria maculans based on genome-wide association studies. *BMC Genomics*, *21*(1), 501. https://doi.org/10.1186/s12864-020-06893-4

Gacek, K., Bayer, P. E., Bartkowiak-Broda, I., Szala, L., Bocianowski, J., Edwards, D., & Batley, J. (2017). Genome-Wide Association Study of Genetic Control of Seed Fatty Acid Biosynthesis in Brassica napus. *Frontiers in Plant Science*, *7*, 2062. https://doi.org/10.3389/fpls.2016.02062

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., Isleib, T., Holbrook, C. C., Xavier, A., Culbreath, A. K., Ozias-Akins, P., Varshney, R. K., & Guo, B. (2020). Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (Arachis hypogaea). *Plant Biotechnology Journal*, *18*(6), 1457–1471. https://doi.org/10.1111/pbi.13311

García-Moreno, M. J., Fernández-Martínez, J. M., Velasco, L., & Pérez-Vich, B. (2012a). Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds. *BMC Plant Biology*, *12*(1), 71. https://doi.org/10.1186/1471-2229-12-71

García-Moreno, M. J., Fernández-Martínez, J. M., Velasco, L., & Pérez-Vich, B. (2012b). Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds. *BMC Plant Biology*, *12*(1), 71. https://doi.org/10.1186/1471-2229-12-71

García-Moreno, M. J., Vera-Ruiz, E. M., Fernández-Martínez, J. M., Velasco, L., & Pérez-Vich, B. (2006). Genetic and Molecular Analysis of High Gamma-Tocopherol Content in Sunflower. *Crop Science*, *46*(5), 2015–2021. https://doi.org/10.2135/cropsci2005.10.0388

Gazave, E., Tassone, E. E., Ilut, D. C., Wingerson, M., Datema, E., Witsenboer, H. M. A., Davis, J. B., Grant, D., Dyer, J. M., Jenks, M. A., Brown, J., & Gore, M. A. (2016). Population Genomic Analysis Reveals Differential Evolutionary Histories and Patterns of Diversity across Subgenomes and Subpopulations of Brassica napus L. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00525

Gentzbittel, L., Vear, F., Zhang, Y. X., Bervillé, A., & Nicolas, P. (1995). Development of a consensus linkage RFLP map of cultivated sunflower (Helianthus annuus L.). *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *90*(7–8), 1079–1086. https://doi.org/10.1007/BF00222925

Geu-Flores, F., Møldrup, M. E., Böttcher, C., Olsen, C. E., Scheel, D., & Halkier, B. A. (2011). Cytosolic γ-Glutamyl Peptidases Process Glutathione Conjugates in the Biosynthesis of Glucosinolates and Camalexin in Arabidopsis. *The Plant Cell*, *23*(6), 2456–2469. https://doi.org/10.1105/tpc.111.083998

Geu-Flores, F., Nielsen, M. T., Nafisi, M., Møldrup, M. E., Olsen, C. E., Motawia, M. S., & Halkier, B. A. (2009). Glucosinolate engineering identifies a γ-glutamyl peptidase. *Nature Chemical Biology*, *5*(8), 8. https://doi.org/10.1038/nchembio.185

Gigolashvili, T., Yatusevich, R., Berger, B., Müller, C., & Flügge, U.-I. (2007). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology*,

51(2), 247–261. https://doi.org/10.1111/j.1365-313X.2007.03133.x

Gjorgjieva, B., Karov, I., Mitrev, S., Ruzdik, N. M., Kostadinovska, E., & Kovacevik, B. (2015). Correlation and Path Analysis in Sunflower (Helianthus annuus L.). *Helia*, *38*(63), 201–210. https://doi.org/10.1515/helia-2015-0008

Gomez-Campo, C. (1999). *Biology of Brassica Coenospecies*. Elsevier.

Gorlova, L. A., Bochkaryova, E. B., Strelnikov, E. A., & Serdyuk, V. V. (2020). The use of classical and modern methods in rapeseed (Brassica napus) breeding at VNIIMK. *Proceedings on applied botany, genetics and breeding*, *180*(4), 4. https://doi.org/10.30901/2227-8834-2019-4-126-131

Goryunov, D. V., Anisimova, I. N., Gavrilova, V. A., Chernova, A. I., Sotnikova, E. A., Martynova, E. U., Boldyrev, S. V., Ayupova, A. F., Gubaev, R. F., Mazin, P. V., Gurchenko, E. A., Shumskiy, A. A., Petrova, D. A., Garkusha, S. V., Mukhina, Z. M., Benko, N. I., Demurin, Y. N., Khaitovich, P. E., & Goryunova, S. V. (2019). Association Mapping of Fertility Restorer Gene for CMS PET1 in Sunflower. *Agronomy*, *9*(2), 2. https://doi.org/10.3390/agronomy9020049

Goryunova, S. V., Goryunov, D. V., Chernova, A. I., Martynova, E. U., Dmitriev, A. E., Boldyrev, S. V., Ayupova, A. F., Mazin, P. V., Gurchenko, E. A., Pavlova, A. S., Petrova, D. A., Chebanova, Y. V., Gorlova, L. A., Garkusha, S. V., Mukhina, Z. M., Savenko, E. G., & Demurin, Y. N. (2019). Genetic and Phenotypic Diversity of the Sunflower Collection of the Pustovoit All-Russia Research Institute of Oil Crops (VNIIMK). *Helia*, *42*(70), 45–60. https://doi.org/10.1515/helia-2018-0021

Gosseau, F., Blanchet, N., Varès, D., Burger, P., Campergue, D., Colombet, C., Gody, L., Liévin, J.-F., Mangin, B., Tison, G., Vincourt, P., Casadebaig, P., & Langlade, N. (2019). Heliaphen, an Outdoor High-Throughput Phenotyping Platform for Genetic Studies and Crop Modeling. *Frontiers in Plant Science*, *9*. https://www.frontiersin.org/article/10.3389/fpls.2018.01908

Guan, M., Huang, X., Xiao, Z., Jia, L., Wang, S., Zhu, M., Qiao, C., Wei, L., Xu, X., Liang, Y., Wang, R., Lu, K., Li, J., & Qu, C. (2019). Association Mapping Analysis of Fatty Acid Content in Different Ecotypic Rapeseed Using mrMLM. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.01872

Gubaev, R., Boldyrev, S., Martynova, E., Chernova, A., Kovalenko, T., Peretyagina, T., Goryunova, S., Goryunov, D., Mukhina, Z., Ben, C., Gentzbittel, L., Khaitovich, P., & Demurin, Y. (2022). Genetic mapping of loci involved in oil tocopherol composition control in Russian sunflower (Helianthus annuus L.) lines. *G3 Genes|Genomes|Genetics*, jkac036. https://doi.org/10.1093/g3journal/jkac036

Gubaev, R., Gorlova, L., Boldyrev, S., Goryunova, S., Goryunov, D., Mazin, P., Chernova, A., Martynova, E., Demurin, Y., & Khaitovich, P. (2020). Genetic Characterization of Russian Rapeseed Collection and Association Mapping of Novel Loci Affecting Glucosinolate Content. *Genes*, *11*(8), 8. https://doi.org/10.3390/genes11080926

Guo, R., Qian, H., Shen, W., Liu, L., Zhang, M., Cai, C., Zhao, Y., Qiao, J., & Wang, Q. (2013). BZR1 and BES1 participate in regulation of glucosinolate biosynthesis by brassinosteroids in Arabidopsis. *Journal of Experimental Botany*, *64*(8), 2401–2412. https://doi.org/10.1093/jxb/ert094

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2019). Chapter Two—Association mapping in plants in

the post-GWAS genomics era. In D. Kumar (Ed.), *Advances in Genetics* (Vol. 104, pp. 75–154). Academic Press. https://doi.org/10.1016/bs.adgen.2018.12.001

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., & Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species Brassica napus. *Nature Biotechnology*, *30*(8), 798–802. https://doi.org/10.1038/nbt.2302

Hass, C. G., Tang, S., Leonard, S., Traber, M. G., Miller, J. F., & Knapp, S. J. (2006). Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *113*(5), 767–782. https://doi.org/10.1007/s00122-006-0320-4

Havlickova, L., He, Z., Wang, L., Langer, S., Harper, A. L., Kaur, H., Broadley, M. R., Gegas, V., & Bancroft, I. (2018). Validation of an updated Associative Transcriptomics platform for the polyploid crop species Brassica napus by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds. *The Plant Journal: For Cell and Molecular Biology*, *93*(1), 181–192. https://doi.org/10.1111/tpj.13767

HAWK, J. A. (1982). Single gene control of seed color and hypocotyl color in turnip rape. *Canadian Journal of Plant Science*, *62*(2), 331–334. https://doi.org/10.4141/cjps82-051

He, Y., Fu, Y., Hu, D., Wei, D., & Qian, W. (2018). QTL Mapping of Seed Glucosinolate Content Responsible for Environment in Brassica napus. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00891

Hejna, O., Havlickova, L., He, Z., Bancroft, I., & Curn, V. (2019). Analysing the genetic architecture of clubroot resistance variation in Brassica napus by associative transcriptomics. *Molecular Breeding*, *39*(8), 112. https://doi.org/10.1007/s11032-019-1021-4

Hinckley, W. E., Keymanesh, K., Cordova, J. A., & Brusslan, J. A. (2019). The HAC1 histone acetyltransferase promotes leaf senescence and regulates the expression of ERF022. *Plant Direct*, *3*(8), e00159. https://doi.org/10.1002/pld3.159

Hladni, N., & Miladinović, D. (2019). Confectionery sunflower breeding and supply chain in Eastern Europe. *OCL*, *26*, 29. https://doi.org/10.1051/ocl/2019019

Horn, R., Radanovic, A., Fuhrmann, L., Sprycha, Y., Hamrit, S., Jockovic, M., Miladinovic, D., & Jansen, C. (2019). Development and Validation of Markers for the Fertility Restorer Gene Rf1 in Sunflower. *International Journal of Molecular Sciences*, *20*(6), 6. https://doi.org/10.3390/ijms20061260

Hossain, Z., Amyot, L., McGarvey, B., Gruber, M., Jung, J., & Hannoufa, A. (2012). The Translation Elongation Factor eEF-1Bβ1 Is Involved in Cell Wall Biosynthesis and Plant Development in Arabidopsis thaliana. *PLOS ONE*, *7*(1), e30425. https://doi.org/10.1371/journal.pone.0030425

Howell, P. M., Sharpe, A. G., & Lydiate, D. J. (2003). Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (Brassica napus). *Genome*, *46*(3), 454–460. https://doi.org/10.1139/g03-028

Hu, X., Sullivan-Gilbert, M., Gupta, M., & Thompson, S. A. (2006). Mapping of the loci controlling oleic and linolenic acid contents and development of fad2 and fad3 allele-specific markers in canola (Brassica napus L.). *Theoretical and Applied Genetics*, *113*(3), 497–507. https://doi.org/10.1007/s00122-006-0315-1

Hu, X., Sullivan-Gilbert, M., Kubik, T., Danielson, J., Hnatiuk, N., Marchione, W., Greene, T., & Thompson, S. A. (2008). Mapping of the Ogura fertility restorer gene Rfo and development

of Rfo allele-specific markers in canola (Brassica napus L.). *Molecular Breeding*, *22*(4), 663–674. https://doi.org/10.1007/s11032-008-9207-1

Hu, Z., Hua, W., Huang, S., Yang, H., Zhan, G., Wang, X., Liu, G., & Wang, H. (2012). Discovery of Pod Shatter-Resistant Associated SNPs by Deep Sequencing of a Representative Library Followed by Bulk Segregant Analysis in Rapeseed. *PLOS ONE*, *7*(4), e34253. https://doi.org/10.1371/journal.pone.0034253

Huang, Y. J., Jestin, C., Welham, S. J., King, G. J., Manzanares-Dauleux, M. J., Fitt, B. D. L., & Delourme, R. (2016). Identification of environmentally stable QTL for resistance against Leptosphaeria maculans in oilseed rape (Brassica napus). *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *129*, 169–180. https://doi.org/10.1007/s00122-015-2620-z

Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., Lee, J. S., Baute, G. J., Owens, G. L., Grassa, C. J., Ebert, D. P., Ostevik, K. L., Moyers, B. T., Yakimowski, S., Masalia, R. R., Gao, L., Ćalić, I., Bowers, J. E., Kane, N. C., … Rieseberg, L. H. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nature Plants*, *5*(1), 1. https://doi.org/10.1038/s41477-018-0329-0

Hussain, M., Malik, M. A., Farooq, M., Ashraf, M. Y., & Cheema, M. A. (2008). Improving Drought Tolerance by Exogenous Application of Glycinebetaine and Salicylic Acid in Sunflower. *Journal of Agronomy and Crop Science*, *194*(3), 193–199. https://doi.org/10.1111/j.1439-037X.2008.00305.x

Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., Deng, D. D., Hinchliffe, D. J., Jones, D. C., & Fang, D. D. (2016). A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. *BMC Genomics*, *17*(1), 903. https://doi.org/10.1186/s12864-016-3249-2

Ivanov, A. Y., Куликов, Р. С., Харченко, М. М., Волощенко, В. С., Гончаров, С. В., Резвый, Г. И., Баханова, Е. А., Гаврилова, О. А., Губаев, Р. Ф., Дискин, Е. И., Долматова, Н. А., Жегусов, В. О., Казбекова, В. О., Котова, Д. А., Марзиев, Т. Б., Николаенко, Н. Г., Позднякова, А. Э., & Ровнов, Ю. Е. (2020). *Селекция 2.0*. Издательский дом НИУ ВШЭ. https://doi.org/10.17323/978-5-7598-2349-0

Jaime, R., Alcántara, J. M., Manzaneda, A. J., & Rey, P. J. (2018). Climate change decreases suitable areas for rapeseed cultivation in Europe but provides new opportunities for white mustard as an alternative oilseed for biofuel production. *PLoS ONE*, *13*(11), e0207124. https://doi.org/10.1371/journal.pone.0207124

Jan, H. U., Guan, M., Yao, M., Liu, W., Wei, D., Abbadi, A., Zheng, M., He, X., Chen, H., Guan, C., Nichols, R. A., Snowdon, R. J., Hua, W., & Qian, L. (2019). Genome-wide haplotype analysis improves trait predictions in Brassica napus hybrids. *Plant Science*, *283*, 157–164. https://doi.org/10.1016/j.plantsci.2019.02.007

Jean, M., Brown, G. G., & Landry, B. S. (1997). Genetic mapping of nuclear fertility restorer genes for the 'Polima' cytoplasmic male sterility in canola (Brassica napus L.) using DNA markers. *Theoretical and Applied Genetics*, *95*(3), 321–328. https://doi.org/10.1007/s001220050566

Jocic´, S., Miladinovic´, D., & Kaya, Y. (2015). 1—Breeding and Genetics of Sunflower. In E. Martínez-Force, N. T. Dunford, & J. J. Salas (Eds.), *Sunflower* (pp. 1–25). AOCS Press.

https://doi.org/10.1016/B978-1-893997-94-3.50007-6

Jocković, M., Jocić, S., Cvejić, S., Marjanović-Jeromela, A., Jocković, J., Radanović, A., & Miladinović, D. (2021). Genetic Improvement in Sunflower Breeding—Integrated Omics Approach. *Plants*, *10*(6), 6. https://doi.org/10.3390/plants10061150

Jourdren, C., Barret, P., Horvais, R., Foisset, N., Delourme, R., & Renard, M. (1996). Identification of RAPD markers linked to the loci controlling erucic acid level in rapeseed. *Molecular Breeding*, *2*(1). https://doi.org/10.1007/BF00171352

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: An overview of the recent progress in plants. *Euphytica*, *177*(3), 309–334. https://doi.org/10.1007/s10681-010-0286-9

Kane, GillN, KingM.G, BowersJ.E, BergesH, GouzyJ, BachlavaE, LangladeN.B, LaiZ, StewartM, BurkeJ.M, VincourtP, KnappS.J, & RiesebergL.H. (2011). Progress towards a reference genome for sunflower. *Botany*. https://doi.org/10.1139/b11-032

Kaya, Y., Jocic, S., & Miladinovic, D. (2012). Sunflower. In S. K. Gupta (Ed.), *Technological Innovations in Major World Oil Crops, Volume 1: Breeding* (pp. 85–129). Springer. https://doi.org/10.1007/978-1-4614-0356-2_4

Kendall, S. L., Holmes, H., White, C. A., Clarke, S. M., & Berry, P. M. (2017). Quantifying lodging-induced yield losses in oilseed rape. *Field Crops Research*, *211*, 106–113. https://doi.org/10.1016/j.fcr.2017.06.013

Khan, S. U., Yangmiao, J., Liu, S., Zhang, K., Khan, M. H. U., Zhai, Y., Olalekan, A., Fan, C., & Zhou, Y. (2019). Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in Brassica napus L. *Industrial Crops and Products*, *142*, 111877. https://doi.org/10.1016/j.indcrop.2019.111877

Khanzada, H., Wassan, G. M., He, H., Mason, A. S., Keerio, A. A., Khanzada, S., Faheem, M., Solangi, A. M., Zhou, Q., Fu, D., Huang, Y., & Rasheed, A. (2020). Differentially evolved drought stress indices determine the genetic variation of Brassica napus at seedling traits by genome-wide association mapping. *Journal of Advanced Research*, *24*, 447–461. https://doi.org/10.1016/j.jare.2020.05.019

Kibbe, W. A. (2007). OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*(suppl_2), W43–W46. https://doi.org/10.1093/nar/gkm234

Kittipol, V., He, Z., Wang, L., Doheny-Adams, T., Langer, S., & Bancroft, I. (2019). Genetic architecture of glucosinolate variation in Brassica napus. *Journal of Plant Physiology*, *240*, 152988. https://doi.org/10.1016/j.jplph.2019.06.001

Knoll, J., & Ejeta, G. (2008). Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments. *Theoretical and Applied Genetics*, *116*(4), 541–553. https://doi.org/10.1007/s00122-007-0689-8

Kolkman, J. M., Slabaugh, M. B., Bruniard, J. M., Berry, S., Bushman, B. S., Olungu, C., Maes, N., Abratti, G., Zambelli, A., Miller, J. F., Leon, A., & Knapp, S. J. (2004). Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. *Theoretical and Applied Genetics*, *109*(6), 1147–1159. https://doi.org/10.1007/s00122-004-1716-7

Kuai, J., Sun, Y., Liu, T., Zhang, P., Zhou, M., Wu, J., & Zhou, G. (2016). Physiological Mechanisms behind Differences in Pod Shattering Resistance in Rapeseed (Brassica napus L.) Varieties. *PLoS ONE*, *11*(6), e0157341. https://doi.org/10.1371/journal.pone.0157341

Lacombe, S., & Bervillé, A. (2001). A dominant mutation for high oleic acid content in sunflower (Helianthus annuus L.) seed oil is genetically linked to a single oleate-desaturase RFLP locus. *Molecular Breeding*, *8*(2), 129–137. https://doi.org/10.1023/A:1013358711651

Lacombe, S., Souyris, I., & Bervillé, A. J. (2009). An insertion of oleate desaturase homologous sequence silences via siRNA the functional gene leading to high oleic acid content in sunflower seed oil. *Molecular Genetics and Genomics: MGG*, *281*(1), 43–54. https://doi.org/10.1007/s00438-008-0391-9

Lai, Z., Livingstone, K., Zou, Y., Church, S. A., Knapp, S. J., Andrews, J., & Rieseberg, L. H. (2005). Identification and mapping of SNPs from ESTs in sunflower. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *111*(8), 1532–1544. https://doi.org/10.1007/s00122-005-0082-4

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 4. https://doi.org/10.1038/nmeth.1923

LeBoldus, J. M., Manolii, V. P., Turkington, T. K., & Strelkov, S. E. (2012). Adaptation to Brassica Host Genotypes by a Single-Spore Isolate and Population of Plasmodiophora brassicae (Clubroot). *Plant Disease*, *96*(6), 833–838. https://doi.org/10.1094/PDIS-09-11-0807

Leclercq, P. (1971). La sterilite male cytoplasmique du tournesol. 1. Premieres etudes sur la restauration de la fertilite. *Annales de l'Amélioration des Plantes*, *21*(1), 45.

Lee, J., Chin, J. H., Ahn, S. N., & Koh, H.-J. (2015). Brief History and Perspectives on Plant Breeding. In H.-J. Koh, S.-Y. Kwon, & M. Thomson (Eds.), *Current Technologies in Plant Molecular Breeding* (pp. 1–14). Springer Netherlands. https://doi.org/10.1007/978-94-017-9996-6_1

Lees, C. J., Li, G., & Duncan, R. W. (2016). Characterization of Brassica napus L. genotypes utilizing sequence-related amplified polymorphism and genotyping by sequencing in association with cluster analysis. *Molecular Breeding*, *36*(11), 155. https://doi.org/10.1007/s11032-016-0576-6

Lexer, C., Welch, M. E., Durphy, J. L., & Rieseberg, L. H. (2003). Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of Helianthus paradoxus, a diploid hybrid species. *Molecular Ecology*, *12*(5), 1225–1235. https://doi.org/10.1046/j.1365-294X.2003.01803.x

Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., Harper, A. L., Trick, M., Liu, S., Gao, G., Wang, N., Yan, G., Qiao, J., Li, J., Li, H., Xiao, X., Zhang, T., & Wu, X. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (Brassica napus L.). *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *21*(4), 355–367. https://doi.org/10.1093/dnares/dsu002

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J., Zhao, Z., Hayward, A., Cheng, H., & Fu, D. (2015). Integration analysis of quantitative trait loci for resistance to Sclerotinia sclerotiorum in Brassica napus. *Euphytica*, *205*(2),

483–489. https://doi.org/10.1007/s10681-015-1417-0

Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., Huang, Q., Xiao, X., Zhang, T., Hu, J., Li, F., & Wu, X. (2016). A Genome-Wide Association Study Reveals New Loci for Resistance to Clubroot Disease in Brassica napus. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/article/10.3389/fpls.2016.01483

Li, L., Tacke, E., Hofferbert, H.-R., Lübeck, J., Strahwald, J., Draffehn, A. M., Walkemeier, B., & Gebhardt, C. (2013). Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *126*(4), 1039–1052. https://doi.org/10.1007/s00122-012-2035-z

Li, Z., Mei, S., Mei, Z., Liu, X., Fu, T., Zhou, G., & Tu, J. (2014). Mapping of QTL associated with waterlogging tolerance and drought resistance during the seedling stage in oilseed rape (Brassica napus). *Euphytica*, *197*(3), 341–353. https://doi.org/10.1007/s10681-014-1070-z

Liu, J., Wang, J., Wang, H., Wang, W., Zhou, R., Mei, D., Cheng, H., Yang, J., Raman, H., & Hu, Q. (2016). Multigenic Control of Pod Shattering Resistance in Chinese Rapeseed Germplasm Revealed by Genome-Wide Association and Linkage Analyses. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/article/10.3389/fpls.2016.01058

Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C., & Zhou, Y. (2020). Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of Brassica napus by genome‐wide association study. *Plant Biotechnology Journal*, *18*(6), 1472–1484. https://doi.org/10.1111/pbi.13314

Liu, Y., Zhou, X., Yan, M., Wang, P., Wang, H., Xin, Q., Yang, L., Hong, D., & Yang, G. (2020). Fine mapping and candidate gene analysis of a seed glucosinolate content QTL, qGSL-C2, in rapeseed (Brassica napus L.). *Theoretical and Applied Genetics*, *133*(2), 479–490. https://doi.org/10.1007/s00122-019-03479-x

Liu, Z., Mulpuri, S., Feng, J., Vick, B. A., & Jan, C.-C. (2012). Molecular mapping of the Rf3fertility restoration gene to facilitate its utilization in breeding confection sunflower. *Molecular Breeding*, *29*(2), 275–284. https://doi.org/10.1007/s11032-011-9563-0

Livaja, M., Unterseer, S., Erath, W., Lehermeier, C., Wieseke, R., Plieske, J., Polley, A., Luerßen, H., Wieckhorst, S., Mascher, M., Hahn, V., Ouzunova, M., Schön, C.-C., & Ganal, M. W. (2016). Diversity analysis and genomic prediction of Sclerotinia resistance in sunflower using a new 25 K SNP genotyping array. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *129*(2), 317–329. https://doi.org/10.1007/s00122-015-2629-3

Louarn, J., Boniface, M.-C., Pouilly, N., Velasco, L., Pérez-Vich, B., Vincourt, P., & Muños, S. (2016). Sunflower Resistance to Broomrape (Orobanche cumana) Is Controlled by Specific QTLs for Different Parasitism Stages. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/articles/10.3389/fpls.2016.00590

Lukomets, V. M., Trunova, M. V., & Demurin, Ya. N. (2021). Modern trends in breeding and genetic improvement of sunflower varieties and hybrids at VNIIMK. *Vavilov Journal of Genetics and Breeding*, *25*(4), 388–393. https://doi.org/10.18699/VJ21.042

Lushchak, V. I., & Semchuk, N. M. (2012). Tocopherol biosynthesis: Chemistry, regulation and effects of environmental factors. *Acta Physiologiae Plantarum*, *34*(5), 1607–1628. https://doi.org/10.1007/s11738-012-0988-9

Ma, G., Long, Y., Song, Q., Talukder, Z. I., Shamimuzzaman, M., & Qi, L. (2021). Map and sequence-based chromosome walking towards cloning of the male fertility restoration gene Rf5 linked to R11 in sunflower. *Scientific Reports*, *11*(1), 1. https://doi.org/10.1038/s41598-020-80659-6

Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, *12*(2), 57–63. https://doi.org/10.1016/j.tplants.2006.12.001

Mag, T. K. (1983). Canola oil processing in Canada. *Journal of the American Oil Chemists' Society*, *60*(2Part2), 380–384. https://doi.org/10.1007/BF02543522

Malmberg, M. M., Pembleton, L. W., Baillie, R. C., Drayton, M. C., Sudheesh, S., Kaur, S., Shinozuka, H., Verma, P., Spangenberg, G. C., Daetwyler, H. D., Forster, J. W., & Cogan, N. O. I. (2018). Genotyping‐by‐sequencing through transcriptomics: Implementation in a range of crop species with varying reproductive habits and ploidy levels. *Plant Biotechnology Journal*, *16*(4), 877–889. https://doi.org/10.1111/pbi.12835

Mangin, B., Bonnafous, F., Blanchet, N., Boniface, M.-C., Bret-Mestries, E., Carrère, S., Cottret, L., Legrand, L., Marage, G., Pegot-Espagnet, P., Munos, S., Pouilly, N., Vear, F., Vincourt, P., & Langlade, N. B. (2017). Genomic Prediction of Sunflower Hybrids Oil Content. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/article/10.3389/fpls.2017.01633

Manzanares-Dauleux, M. J., Delourme, R., Baron, F., & Thomas, G. (2000). Mapping of one major gene and of QTLs involved in resistance to clubroot in Brassica napus. *Theoretical and Applied Genetics*, *101*(5), 885–891. https://doi.org/10.1007/s001220051557

Martín-Sanz, A., Pérez-Vich, B., Rueda, S., Fernández-Martínez, J. M., & Velasco, L. (2020). Characterization of post-haustorial resistance to sunflower broomrape. *Crop Science*, *60*(3), 1188–1198. https://doi.org/10.1002/csc2.20002

Marwede, V., Gul, M. K., Becker, H. C., & Ecke, W. (2005). Mapping of QTL controlling tocopherol content in winter oilseed rape. *Plant Breeding*, *124*(1), 20–26. https://doi.org/10.1111/j.1439-0523.2004.01050.x

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Melchinger, A. E. (1990). Use of Molecular Markers in Breeding for Oligogenic Disease Resistance. *Plant Breeding*, *104*(1), 1–19. https://doi.org/10.1111/j.1439-0523.1990.tb00396.x

Merah, O., Langlade, N., Alignan, M., Roche, J., Pouilly, N., Lippi, Y., Vear, F., Cerny, M., Bouniols, A., Mouloungui, Z., & Vincourt, P. (2012). Genetic analysis of phytosterol content in sunflower seeds. *Theoretical and Applied Genetics*, *125*(8), 1589–1601. https://doi.org/10.1007/s00122-012-1937-0

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, *157*(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Micic, Z., Hahn, V., Bauer, E., Schön, C. C., Knapp, S. J., Tang, S., & Melchinger, A. E. (2004). QTL mapping of Sclerotinia midstalk-rot resistance in sunflower. *Theoretical and Applied Genetics*, *109*(7), 1474–1484. https://doi.org/10.1007/s00122-004-1764-z

Micic, Z., Hahn, V., Bauer, E., Schön, C. C., & Melchinger, A. E. (2005). QTL mapping of

resistance to Sclerotinia midstalk rot in RIL of sunflower population NDBLOSsel × CM625. *Theoretical and Applied Genetics*, *110*(8), 1490–1498. https://doi.org/10.1007/s00122-005-1984-x

Miller, C. N., Harper, A. L., Trick, M., Wellner, N., Werner, P., Waldron, K. W., & Bancroft, I. (2018). Dissecting the complex regulation of lodging resistance in Brassica napus. *Molecular Breeding*, *38*(3), 30. https://doi.org/10.1007/s11032-018-0781-6

Miller, J. F., Zimmerman, D. C., & Vick, B. A. (1987). Genetic Control of High Oleic Acid Content in Sunflower Oil1. *Crop Science*, *27*(5), cropsci1987.0011183X002700050019x. https://doi.org/10.2135/cropsci1987.0011183X002700050019x

Mohammadi, M., Xavier, A., Beckett, T., Beyer, S., Chen, L., Chikssa, H., Cross, V., Freitas Moreira, F., French, E., Gaire, R., Griebel, S., Lopez, M. A., Prather, S., Russell, B., & Wang, W. (2020). Identification, deployment, and transferability of quantitative trait loci from genome-wide association studies in plants. *Current Plant Biology*, *24*, 100145. https://doi.org/10.1016/j.cpb.2020.100145

Monteros, M. J., Burton, J. W., & Boerma, H. R. (2008). Molecular Mapping and Confirmation of QTLs Associated with Oleic Acid Content in N00-3350 Soybean. *Crop Science*, *48*(6), 2223–2234. https://doi.org/10.2135/cropsci2008.05.0287

Morales, K. Y., Singh, N., Perez, F. A., Ignacio, J. C., Thapa, R., Arbelaez, J. D., Tabien, R. E., Famoso, A., Wang, D. R., Septiningsih, E. M., Shi, Y., Kretzschmar, T., McCouch, S. R., & Thomson, M. J. (2020). An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS ONE*, *15*(5), e0232479. https://doi.org/10.1371/journal.pone.0232479

Murray, K. D., & Borevitz, J. O. (2018). Axe: Rapid, competitive sequence read demultiplexing using a trie. *Bioinformatics*, *34*(22), 3924–3925. https://doi.org/10.1093/bioinformatics/bty432

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, *32*(2), 261–285. https://doi.org/10.1080/13102818.2017.1400401

Neik, T. X., Barbetti, M. J., & Batley, J. (2017). Current Status and Challenges in Identifying Disease Resistance Genes in Brassica napus. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/article/10.3389/fpls.2017.01788

Nice, L. M., Steffenson, B. J., Brown-Guedira, G. L., Akhunov, E. D., Liu, C., Kono, T. J. Y., Morrell, P. L., Blake, T. K., Horsley, R. D., Smith, K. P., & Muehlbauer, G. J. (2016). Development and Genetic Characterization of an Advanced Backcross-Nested Association Mapping (AB-NAM) Population of Wild × Cultivated Barley. *Genetics*, *203*(3), 1453–1467. https://doi.org/10.1534/genetics.116.190736

Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics (Oxford, England)*, *28*(8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., Miah, G., & Usman, M. (2016). Principle and application of plant mutagenesis in crop improvement: A review. *Biotechnology & Biotechnological Equipment*, *30*(1), 1–16.

https://doi.org/10.1080/13102818.2015.1087333

Oyserman, B. O., Cordovez, V., Flores, S. S., Leite, M. F. A., Nijveen, H., Medema, M. H., & Raaijmakers, J. M. (2021). Extracting the GEMs: Genotype, Environment, and Microbiome Interactions Shaping Host Phenotypes. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/article/10.3389/fmicb.2020.574053

Parkin, I. A., Sharpe, A. G., Keith, D. J., & Lydiate, D. J. (1995). Identification of the A and C genomes of amphidiploid Brassica napus (oilseed rape). *Genome*, *38*(6), 1122–1131. https://doi.org/10.1139/g95-149

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J.-P., Le, Q. H., Chauchard, B., Verschave, P., & Causse, M. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnology Journal*, *13*(4), 565–577. https://doi.org/10.1111/pbi.12282

Pattemore, J. A., Rice, N., Marshall, D. F., Waugh, R., & Henry, R. J. (2010). Cereal variety identification using MALDI-TOF mass spectrometry SNP genotyping. *Journal of Cereal Science*, *52*(3), 356–361. https://doi.org/10.1016/j.jcs.2010.06.004

Pérez-Vich, B., del Moral, L., Velasco, L., Bushman, B. S., Knapp, S. J., Leon, A., Fernández-Martínez, J. M., & Berry, S. T. (2016). Molecular basis of the high-palmitic acid trait in sunflower seed oil. *Molecular Breeding*, *36*(4), 43. https://doi.org/10.1007/s11032-016-0462-2

Pérez-Vich, B., Fernández-Martínez, J. M., Grondona, M., Knapp, S. J., & Berry, S. T. (2002). Stearoyl-ACP and oleoyl-PC desaturase genes cosegregate with quantitative trait loci underlying high stearic and high oleic acid mutant phenotypes in sunflower. *Theoretical and Applied Genetics*, *104*(2), 338–349. https://doi.org/10.1007/s001220100712

Pieruschka, R., & Schurr, U. (2019). Plant Phenotyping: Past, Present, and Future. *Plant Phenomics*, *2019*. https://doi.org/10.34133/2019/7507131

Poormohammad Kiani, S., Grieu, P., Maury, P., Hewezi, T., Gentzbittel, L., & Sarrafi, A. (2007). Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (Helianthus annuus L.). *Theoretical and Applied Genetics*, *114*(2), 193–207. https://doi.org/10.1007/s00122-006-0419-7

Poormohammad Kiani, S., Maury, P., Nouri, L., Ykhlef, N., Grieu, P., & Sarrafi, A. (2009). QTL analysis of yield-related traits in sunflower under different water treatments. *Plant Breeding*, *128*(4), 363–373. https://doi.org/10.1111/j.1439-0523.2009.01628.x

Poormohammad Kiani, S., Maury, P., Sarrafi, A., & Grieu, P. (2008). QTL analysis of chlorophyll fluorescence parameters in sunflower (Helianthus annuus L.) under well-watered and water-stressed conditions. *Plant Science*, *175*(4), 565–573. https://doi.org/10.1016/j.plantsci.2008.06.002

Prasannakumar, M. K., Parivallal, B. P., Manjunatha, C., Pramesh, D., Narayan, K. S., Venkatesh, G., Banakar, S. N., Mahesh, H. B., Vemanna, R. S., & Rangaswamy, K. T. (2021). Rapid genotyping of bacterial leaf blight resistant genes of rice using loop-mediated isothermal amplification assay. *Molecular Biology Reports*, *48*(1), 467–474. https://doi.org/10.1007/s11033-020-06077-z

Premnath, A., Narayana, M., Ramakrishnan, C., Kuppusamy, S., & Chockalingam, V. (2016). Mapping quantitative trait loci controlling oil content, oleic acid and linoleic acid content in

sunflower (Helianthus annuus L.). *Molecular Breeding*, *36*(7), 106. https://doi.org/10.1007/s11032-016-0527-2

Przybylski, R., Mag, T., Eskin, N. a. m., & McDonald, B. e. (2005). Canola Oil. In *Bailey's Industrial Oil and Fat Products*. John Wiley & Sons, Ltd. https://doi.org/10.1002/047167849X.bio004

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Qi, L. L., Ma, G. J., Li, X. H., & Seiler, G. J. (2019). Diversification of the downy mildew resistance gene pool by introgression of a new gene, Pl35, from wild Helianthus argophyllus into oilseed and confection sunflowers (Helianthus annuus L.). *Theoretical and Applied Genetics*, *132*(9), 2553–2565. https://doi.org/10.1007/s00122-019-03370-9

Qian, L., Qian, W., & Snowdon, R. J. (2014). Sub-genomic selection patterns as a signature of breeding in the allopolyploid Brassica napus genome. *BMC Genomics*, *15*(1), 1170. https://doi.org/10.1186/1471-2164-15-1170

Qu, C., Jia, L., Fu, F., Zhao, H., Lu, K., Wei, L., Xu, X., Liang, Y., Li, S., Wang, R., & Li, J. (2017). Genome-wide association mapping and Identification of candidate genes for fatty acid composition in Brassica napus L. using SNP markers. *BMC Genomics*, *18*(1), 232. https://doi.org/10.1186/s12864-017-3607-8

Qu, C.-M., Li, S.-M., Duan, X.-J., Fan, J.-H., Jia, L.-D., Zhao, H.-Y., Lu, K., Li, J.-N., Xu, X.-F., & Wang, R. (2015). Identification of Candidate Genes for Seed Glucosinolate Content Using Association Mapping in Brassica napus L. *Genes*, *6*(4), 1215–1229. https://doi.org/10.3390/genes6041215

Quijano Xacur, O. A. (2019). The unifed distribution. *Journal of Statistical Distributions and Applications*, *6*(1), 13. https://doi.org/10.1186/s40488-019-0102-6

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Radic, V., Ovuka, J., Jockovic, M., Petar, C., Radanović, A., & Jocić, S. (2013). Sunflower 1000-seed weight as affected by year and genotype. *Ratarstvo i Povrtarstvo*, *50*, 1–7. https://doi.org/10.5937/ratpov50-3214

Ramalingam, J., Raveendra, C., Savitha, P., Vidya, V., Chaithra, T. L., Velprabakaran, S., Saraswathi, R., Ramanathan, A., Arumugam Pillai, M. P., Arumugachamy, S., & Vanniarajan, C. (2020). Gene Pyramiding for Achieving Enhanced Resistance to Bacterial Blight, Blast, and Sheath Blight Diseases in Rice. *Frontiers in Plant Science*, *11*. https://www.frontiersin.org/article/10.3389/fpls.2020.591457

Raman, H., Raman, R., Kilian, A., Detering, F., Carling, J., Coombes, N., Diffey, S., Kadkol, G., Edwards, D., McCully, M., Ruperao, P., Parkin, I. A. P., Batley, J., Luckett, D. J., & Wratten, N. (2014). Genome-Wide Delineation of Natural Variation for Pod Shatter Resistance in Brassica napus. *PLOS ONE*, *9*(7), e101673. https://doi.org/10.1371/journal.pone.0101673

Raman, R., Diffey, S., Barbulescu, D. M., Coombes, N., Luckett, D., Salisbury, P., Cowley, R., Marcroft, S., & Raman, H. (2020). Genetic and physical mapping of loci for resistance to blackleg disease in canola (Brassica napus L.). *Scientific Reports*, *10*(1), 1.

https://doi.org/10.1038/s41598-020-61211-y

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. (2017). Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. *Molecular Plant*, *10*(8), 1047–1064. https://doi.org/10.1016/j.molp.2017.06.008

Rasmusson, J. (1935). Studies on the Inheritance of Quantitative Characters in Pisum. *Hereditas*, *20*(1–2), 161–180. https://doi.org/10.1111/j.1601-5223.1935.tb03184.x

Rauf, S., Jamil, N., Tariq, S. A., Khan, M., Kausar, M., & Kaya, Y. (2017). Progress in modification of sunflower oil to expand its industrial value. *Journal of the Science of Food and Agriculture*, *97*(7), 1997–2006. https://doi.org/10.1002/jsfa.8214

Rauf, S., Warburton, M., Naeem, A., & Kainat, W. (2020). Validated markers for sunflower (Helianthus annuus L.) breeding. *OCL*, *27*, 47. https://doi.org/10.1051/ocl/2020042

Reinert, S., Gao, Q., Ferguson, B., Portlas, Z. M., Prasifka, Jarrad. R., & Hulke, B. S. (2020). Seed and floret size parameters of sunflower are determined by partially overlapping sets of quantitative trait loci with epistatic interactions. *Molecular Genetics and Genomics*, *295*(1), 143–154. https://doi.org/10.1007/s00438-019-01610-7

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., & Tardieu, F. (2019). What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Science: An International Journal of Experimental Plant Biology*, *282*, 14–22. https://doi.org/10.1016/j.plantsci.2018.06.015

Rieseberg, L. H., Choi, H., Chan, R., & Spore, C. (1993). Genomic map of a diploid hybrid species. *Heredity*, *70*(3), 3. https://doi.org/10.1038/hdy.1993.41

Rimmer, S. R. (2006). Resistance genes to Leptosphaeria maculans in Brassica napus. *Canadian Journal of Plant Pathology*, *28*(sup1), S288–S297. https://doi.org/10.1080/07060660609507386

Rocha, C. R. M. da, Silva, V. N., & Cicero, S. M. (2014). *Internal morphology and germination of sunflower seeds 1*. https://doi.org/10.1590/S2317-15372014000100006

Rogers, A. R., Dunne, J. C., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I. A., Edwards, J., Ertl, D., Flint-Garcia, S., Gore, M. A., Graham, C., Hirsch, C. N., Hood, E., Hooker, D. C., Knoll, J., Lee, E. C., Lorenz, A., Lynch, J. P., McKay, J., … Holland, J. B. (2021). The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. *G3 (Bethesda, Md.)*, *11*(2), jkaa050. https://doi.org/10.1093/g3journal/jkaa050

Rondanini, D., Savin, R., & Hall, A. J. (2003). Dynamics of fruit growth and oil quality of sunflower (Helianthus annuus L.) exposed to brief intervals of high temperature during grain filling. *Field Crops Research*, *83*(1), 79–90. https://doi.org/10.1016/S0378-4290(03)00064-9

Rousselle, P., & Bregeon, M. (1982). Premiers résultats d'un programme d'introduction de l'androstérilité « Ogura » du radis chez le colza. *Agronomie*, *2*(9), 859–864. https://doi.org/10.1051/agro:19820910

Roy, J., Shaikh, T. M., del Río Mendoza, L., Hosain, S., Chapara, V., & Rahman, M. (2021). Genome-wide association mapping and genomic prediction for adult stage sclerotinia stem rot resistance in Brassica napus (L) under field environments. *Scientific Reports*, *11*(1), 1. https://doi.org/10.1038/s41598-021-01272-9

Ryzhenko, E. N., Araslanova, N. M., & Goncharov, S. V. (2021). Breeding of sunflower lines

resistant to race G of Broomrape. *Agrarian science*, *0*(6), 6. https://doi.org/10.32634/0869-8155-2021-350-6-42-45

Sajer, O., Schirmak, U., Hamrit, S., & Horn, R. (2020). Mapping of the New Fertility Restorer Gene Rf-PET2 Close to Rf1 on Linkage Group 13 in Sunflower (Helianthus annuus L.). *Genes*, *11*(3), 269. https://doi.org/10.3390/genes11030269

Sala, C. A., Bulos, M., Echarte, M., Whitt, S. R., & Ascenzi, R. (2008). Molecular and biochemical characterization of an induced mutation conferring imidazolinone resistance in sunflower. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *118*(1), 105–112. https://doi.org/10.1007/s00122-008-0880-6

Sales-Campos, H., Souza, P. R. de, Peghini, B. C., da Silva, J. S., & Cardoso, C. R. (2013). An overview of the modulatory effects of oleic acid in health and disease. *Mini Reviews in Medicinal Chemistry*, *13*(2), 201–210.

Sax, K. (1923). The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS. *Genetics*, *8*(6), 552–560.

Schierholt, A., Becker, H. C., & Ecke, W. (2000). Mapping a high oleic acid mutation in winter oilseed rape (Brassica napus L.). *Theoretical and Applied Genetics*, *101*(5), 897–901. https://doi.org/10.1007/s001220051559

Schneiter, A. A., & Miller, J. F. (1981). Description of Sunflower Growth Stages1. *Crop Science*, *21*(6), cropsci1981.0011183X002100060024x. https://doi.org/10.2135/cropsci1981.0011183X002100060024x

Schuppert, G. F., Tang, S., Slabaugh, M. B., & Knapp, S. J. (2006). The Sunflower High-oleic Mutant Ol Carries Variable Tandem Repeats of FAD2-1, a Seed-specific Oleoyl-phosphatidyl Choline Desaturase. *Molecular Breeding*, *17*(3), 241–256. https://doi.org/10.1007/s11032-005-5680-y

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell'Acqua, M., Dixon, L. E., Filippi, C. V., Fradgley, N., Gardner, K. A., Mackay, I. J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R. K., Thudi, M., Varshney, R. K., … Mott, R. (2020). Multi-parent populations in crops: A toolbox integrating genomics and genetic mapping with breeding. *Heredity*, *125*(6), 396–416. https://doi.org/10.1038/s41437-020-0336-6

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, *11 Suppl 1*, 1–8. https://doi.org/10.1111/j.1755-0998.2010.02979.x

Serrano-Vega, M. J., Martínez-Force, E., & Garcés, R. (2005). Lipid characterization of seed oils from high-palmitic, low-palmitoleic, and very high-stearic acid sunflower lines. *Lipids*, *40*(4), 369–374. https://doi.org/10.1007/s11745-006-1396-y

Shahzad, A., Qian, M., Sun, B., Mahmood, U., Li, S., Fan, Y., Chang, W., Dai, L., Zhu, H., Li, J., Qu, C., & Lu, K. (2021). Genome-wide association study identifies novel loci and candidate genes for drought stress tolerance in rapeseed. *Oil Crop Science*, *6*(1), 12–22. https://doi.org/10.1016/j.ocsci.2021.01.002

Shirasawa, K., Tanaka, M., Takahata, Y., Ma, D., Cao, Q., Liu, Q., Zhai, H., Kwak, S.-S., Cheol Jeong, J., Yoon, U.-H., Lee, H.-U., Hirakawa, H., & Isobe, S. (2017). A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid

sweetpotato (Ipomoea batatas). *Scientific Reports*, *7*. https://doi.org/10.1038/srep44207

Škorić, D. (2016). Sunflower Breeding for Resistance to Abiotic and Biotic Stresses. In *Abiotic and Biotic Stress in Plants—Recent Advances and Future Perspectives*. IntechOpen. https://doi.org/10.5772/62159

Skorić, D., Jocić, S., Sakac, Z., & Lecić, N. (2008). Genetic possibilities for altering sunflower oil quality to obtain novel oils. *Canadian Journal of Physiology and Pharmacology*, *86*(4), 215–221. https://doi.org/10.1139/Y08-008

Škorić, D., Joiţa-Păcureanu, M., Gorbachenko, F., Gorbachenko, O., & Maširević, S. (2021). Dynamics of change in broomrape populations (Orobanche cumana Wallr.) in Romania and Russia (Black Sea area). *Helia*, *44*(74), 1–14. https://doi.org/10.1515/helia-2020-0025

Smith, C., & Simpson, S. P. (1986). The use of genetic polymorphisms in livestock improvement. *Journal of Animal Breeding and Genetics*, *103*(1–5), 205–217. https://doi.org/10.1111/j.1439-0388.1986.tb00083.x

Snowdon, R., Lühs, W., & Friedt, W. (2007). Oilseed Rape. In C. Kole (Ed.), *Oilseeds* (pp. 55–114). Springer. https://doi.org/10.1007/978-3-540-34388-2_2

Soldatov, K. (1976). Chemical mutagenesis in sunflower breeding. *Proceedings of the 7th International Sunflower Conference*, 352–357.

Soller, M., & Beckmann, J. S. (1983). Genetic polymorphism in varietal identification and genetic improvement. *Theoretical and Applied Genetics*, *67*(1), 25–33. https://doi.org/10.1007/BF00303917

Souza, R., Teixeira, I., Reis, E., & Silva, A. (2016). Soybean morphophysiology and yield response to seeding systems and plant populations. *Chilean Journal of Agricultural Research*, *76*(1), 3–8. https://doi.org/10.4067/S0718-58392016000100001

Spasibionek, S., Mikołajczyk, K., Ćwiek–Kupczyńska, H., Piętka, T., Krótka, K., Matuszczak, M., Nowakowska, J., Michalski, K., & Bartkowiak-Broda, I. (2020). Marker assisted selection of new high oleic and low linolenic winter oilseed rape (Brassica napus L.) inbred lines revealing good agricultural value. *PLOS ONE*, *15*(6), e0233959. https://doi.org/10.1371/journal.pone.0233959

Stortenbeker, N., & Bemer, M. (2019). The SAUR gene family: The plant's toolbox for adaptation of growth and development. *Journal of Experimental Botany*, *70*(1), 17–27. https://doi.org/10.1093/jxb/ery332

Su, C., Wang, W., Gong, S., Zuo, J., Li, S., & Xu, S. (2017). High Density Linkage Map Construction and Mapping of Yield Trait QTLs in Maize (Zea mays) Using the Genotyping-by-Sequencing (GBS) Technology. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00706

Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., Tong, C., Li, J., Du, D., Qi, C., Jiang, L., Liu, W., Huang, S., Chen, W., Yu, J., Mei, D., Meng, J., Zeng, P., Shi, J., Liu, K., … Wang, H. (2017). The high-quality genome of Brassica napus cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. *The Plant Journal*, *92*(3), 452–468. https://doi.org/10.1111/tpj.13669

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., Jiang, C., Guan, N., Ma, C., Zeng, H., Xu, C., Song, J., Huang, L., Wang, C., Shi, J., Wang, R., Zheng, X., Lu, C., Wang, X., & Zheng, H. (2013). SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *PLOS ONE*, *8*(3), e58700.

https://doi.org/10.1371/journal.pone.0058700

Talukder, Z. I., Gong, L., Hulke, B. S., Pegadaraju, V., Song, Q., Schultz, Q., & Qi, L. (2014). A High-Density SNP Map of Sunflower Derived from RAD-Sequencing Facilitating Fine-Mapping of the Rust Resistance Gene R12. *PLoS ONE*, *9*(7), e98628. https://doi.org/10.1371/journal.pone.0098628

Talukder, Z. I., Underwood, W., Misar, C. G., Seiler, G. J., Liu, Y., Li, X., Cai, X., & Qi, L. (2021). Unraveling the Sclerotinia Basal Stalk Rot Resistance Derived From Wild Helianthus argophyllus Using a High-Density Single Nucleotide Polymorphism Linkage Map. *Frontiers in Plant Science*, *11*. https://www.frontiersin.org/article/10.3389/fpls.2020.617920

Tang, S., Hass, C. G., & Knapp, S. J. (2006). Ty3/gypsy-like retrotransposon knockout of a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase is non-lethal, uncovers a cryptic paralogous mutation, and produces novel tocopherol (vitamin E) profiles in sunflower. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *113*(5), 783–799. https://doi.org/10.1007/s00122-006-0321-3

Tang, S., Leon, A., Bridges, W. C., & Knapp, S. J. (2006). Quantitative Trait Loci for Genetically Correlated Seed Traits are Tightly Linked to Branching and Pericarp Pigment Loci in Sunflower. *Crop Science*, *46*(2), 721–734. https://doi.org/10.2135/cropsci2005.0006-7

Tanksley, S. D., Young, N. D., Paterson, A. H., & Bonierbale, M. W. (1989). RFLP Mapping in Plant Breeding: New Tools for an Old Science. *Bio/Technology*, *7*(3), 257–264. https://doi.org/10.1038/nbt0389-257

Terzić, S., Boniface, M.-C., Marek, L., Alvarez, D., Baumann, K., Gavrilova, V., Joita-Pacureanu, M., Sujatha, M., Valkova, D., Velasco, L., Hulke, B. S., Jocić, S., Langlade, N., Muños, S., Rieseberg, L., Seiler, G., & Vear, F. (2020). Gene banks for wild and cultivated sunflower genetic resources. *OCL*, *27*, 9. https://doi.org/10.1051/ocl/2020004

Thies, W. (1982). Complex-Formation between Glucosinolates and Tetrachloropalladate(II) and its Utilization in Plant Breeding. *Fette, Seifen, Anstrichmittel*, *84*(9), 338–342. https://doi.org/10.1002/lipi.19820840903

Thormann, C. E., Romero, J., Mantet, J., & Osborn, T. C. (n.d.). *Mapping loci controlling the concentrations of erucic and linolenic acids in seed oil of Brassica napus L.* 5.

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, *14*(1), e20077. https://doi.org/10.1002/tpg2.20077

Ton, L. B., Neik, T. X., & Batley, J. (2020). The Use of Genetic and Gene Technologies in Shaping Modern Rapeseed Cultivars (Brassica napus L.). *Genes*, *11*(10), 10. https://doi.org/10.3390/genes11101161

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 1. https://doi.org/10.1038/nrg3117

Trembak, E. N., Savchenko, V. D., Kostevich, S. V., Ryzhenko, E. N., Goloschapova, N. N., Medvedeva, N. V., & Bochkaryov, B. N. (2018). Simple interlinear middle-early sunflower hybrid yphoon. *Oil Crops. Scientific and Technical Bulletin of All-Russian Research Institute of Oil Crops by the Name of Pustovoit V.S.*, *2*, 135–140. https://doi.org/10.25230/2412-608X-2018-2-174-135-140

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115.

https://doi.org/10.1093/nar/gks596

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T. M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K. F., & Schön, C.-C. (2014). A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics*, *15*(1), 823. https://doi.org/10.1186/1471-2164-15-823

Urie, A. L. (1985). Inheritance of High Oleic Acid in Sunflower1. *Crop Science*, *25*(6), cropsci1985.0011183X002500060021x. https://doi.org/10.2135/cropsci1985.0011183X002500060021x

Usatov, A. V., Klimenko, A. I., Azarin, K. V., Gorbachenko, O. F., Markin, N. V., Tikhobaeva, V. E., Kolosov, Y. A., Usatova, O. A., Bakoev, S. Y., Bibov, M. Y., & Getmantseva, L. V. (2014). DNA-markers of sunflower resistance to the downy mildew (Plasmopara halstedii). *American Journal of Biochemistry and Biotechnology*, *10*(2), 125–129.

Uzunova, M., Ecke, W., Weissleder, K., & Röbbelen, G. (1995). Mapping the genome of rapeseed (Brassica napus L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *90*(2), 194–204. https://doi.org/10.1007/BF00222202

Velasco, L., Domínguez, J., & Fernández-Martínez, J. M. (2004). Registration of T589 and T2100 Sunflower Germplasms with Modified Tocopherols. *Crop Science*, *44*(1), 362. https://doi.org/10.2135/cropsci2004.0362

Velasco, L., Pérez-Vich, B., & Fernández-Martínez, J. M. (2004). Novel variation for the tocopherol profile in a sunflower created by mutagenesis and recombination. *Plant Breeding*, *123*(5), 490–492. https://doi.org/10.1111/j.1439-0523.2004.01012.x

Vera-Ruiz, E. M., Pérez-Vich, B., Fernández-Martínez, J. M., & Velasco, L. (2005). COMPARATIVE GENETIC STUDY OF TWO SOURCES OF BETA-TOCOPHEROL IN SUNFLOWER / ESTUDIO GENÉTICO COMPARATIVO DE DOS FUENTES DE BETA-TOCOFEROL EN GIRASOL / ÉTUDE GÉNÉTIQUE COMPARATIVE DE DEUX SOURCES DE TOCOPHÉROL BÊTA DANS LE TOURNESOL. *HELIA*, *28*(42), 1–8. https://doi.org/10.2298/hel0542001v

Vera-Ruiz, E. M., Velasco, L., Leon, A. J., Fernández-Martínez, J. M., & Pérez-Vich, B. (2006). Genetic Mapping of the Tph1 Gene Controlling Beta-tocopherol Accumulation in Sunflower Seeds. *Molecular Breeding*, *17*(3), 291–296. https://doi.org/10.1007/s11032-005-5678-5

Vinnichek, L., Pogorelova, E., & Dergunov, A. (2019). Oilseed market: Global trends. *IOP Conference Series: Earth and Environmental Science*, *274*, 012030. https://doi.org/10.1088/1755-1315/274/1/012030

Virányi, F. (2008). *Research progress in sunflower diseases and their management*. 12.

Walker, K. C., & Booth, E. J. (2001). Agricultural aspects of rape and other Brassica products. *European Journal of Lipid Science and Technology*, *103*(7), 441–446. https://doi.org/10.1002/1438-9312(200107)103:7<441::AID-EJLT441>3.0.CO;2-D

Wan, H., Chen, L., Guo, J., Li, Q., Wen, J., Yi, B., Ma, C., Tu, J., Fu, T., & Shen, J. (2017). Genome-Wide Association Study Reveals the Genetic Architecture Underlying Salt Tolerance-Related Traits in Rapeseed (Brassica napus L.). *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00593

Wang, J., Zhang, X., & Lin, Z. (2018). QTL mapping in a maize F2 population using Genotyping-by-Sequencing and a modified fine-mapping strategy. *Plant Science: An International Journal of Experimental Plant Biology*, *276*, 171–180. https://doi.org/10.1016/j.plantsci.2018.08.019

Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., Shi, L., & Meng, J. (2008). A functional genomics resource for Brassica napus: Development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. *The New Phytologist*, *180*(4), 751–765. https://doi.org/10.1111/j.1469-8137.2008.02619.x

Wang, X., Zhang, C., Li, L., Fritsche, S., Endrigkeit, J., Zhang, W., Long, Y., Jung, C., & Meng, J. (2012). Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (Brassica napus L.). *PLOS ONE*, *7*(11), e50038. https://doi.org/10.1371/journal.pone.0050038

Wang, Z., Ma, L.-Y., Cao, J., Li, Y.-L., Ding, L.-N., Zhu, K.-M., Yang, Y.-H., & Tan, X.-L. (2019). Recent Advances in Mechanisms of Plant Defense to Sclerotinia sclerotiorum. *Frontiers in Plant Science*, *10*. https://www.frontiersin.org/article/10.3389/fpls.2019.01314

Warner, K., Miller, J., & Demurin, Y. (2008). Oxidative Stability of Crude Mid-Oleic Sunflower Oils from Seeds with High γ- and δ-Tocopherol Levels. *Journal of the American Oil Chemists' Society*, *85*(6), 529–533. https://doi.org/10.1007/s11746-008-1212-3

Wei, D., Cui, Y., He, Y., Xiong, Q., Qian, L., Tong, C., Lu, G., Ding, Y., Li, J., Jung, C., & Qian, W. (2017). A genome-wide survey with different rapeseed ecotypes uncovers footprints of domestication and breeding. *Journal of Experimental Botany*, *68*(17), 4791–4801. https://doi.org/10.1093/jxb/erx311

Wei, D., Cui, Y., Mei, J., Qian, L., Lu, K., Wang, Z.-M., Li, J., Tang, Q., & Qian, W. (2019). Genome-wide identification of loci affecting seed glucosinolate contents in Brassica napus L. *Journal of Integrative Plant Biology*, *61*(5), 611–623. https://doi.org/10.1111/jipb.12717

Wei, L., Jian, H., Lu, K., Filardo, F., Yin, N., Liu, L., Qu, C., Li, W., Du, H., & Li, J. (2016). Genome-wide association analysis and differential expression analysis of resistance to Sclerotinia stem rot in Brassica napus. *Plant Biotechnology Journal*, *14*(6), 1368–1380. https://doi.org/10.1111/pbi.12501

Wei, L., Jian, H., Lu, K., Yin, N., Wang, J., Duan, X., Li, W., Liu, L., Xu, X., Wang, R., Paterson, A. H., & Li, J. (2017). Genetic and transcriptomic analyses of lignin- and lodging-related traits in Brassica napus. *Theoretical and Applied Genetics*, *130*(9), 1961–1973. https://doi.org/10.1007/s00122-017-2937-x

Werner, S., Diederichsen, E., Frauen, M., Schondelmaier, J., & Jung, C. (2007). Genetic mapping of clubroot resistance genes in oilseed rape. *Theoretical and Applied Genetics*, *116*(3), 363. https://doi.org/10.1007/s00122-007-0674-2

White, A. D., Owen, M. D. K., Hartzler, R. G., & Cardina, J. (2002). Common sunflower resistance to acetolactate synthase–inhibiting herbicides. *Weed Science*, *50*(4), 432–437. https://doi.org/10.1614/0043-1745(2002)050[0432:CSRTAS]2.0.CO;2

Wieczorek, M. N., Walczak, M., Skrzypczak-Zielińska, M., & Jeleń, H. H. (2018). Bitter taste of Brassica vegetables: The role of genetic factors, receptors, isothiocyanates, glucosinolates, and flavor context. *Critical Reviews in Food Science and Nutrition*, *58*(18), 3130–3140. https://doi.org/10.1080/10408398.2017.1353478

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H.,

Chen, X., Wang, Q., Zhu, L., Maodzeka, A., Hussain, N., Li, Z., Li, X., Shamsi, I. H., Jilani, G., … Jiang, L. (2019). Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. *Molecular Plant*, *12*(1), 30–43. https://doi.org/10.1016/j.molp.2018.11.007

Wu, G., Wu, Y., Xiao, L., Li, X., & Lu, C. (2008). Zero erucic acid trait of rapeseed (Brassica napus L.) results from a deletion of four base pairs in the fatty acid elongase 1 gene. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *116*(4), 491–499. https://doi.org/10.1007/s00122-007-0685-z

Wu, J., Cai, G., Tu, J., Li, L., Liu, S., Luo, X., Zhou, L., Fan, C., & Zhou, Y. (2013). Identification of QTLs for Resistance to Sclerotinia Stem Rot and BnaC.IGMT5.a as a Candidate Gene of the Major Resistant QTL SRC6 in Brassica napus. *PLOS ONE*, *8*(7), e67740. https://doi.org/10.1371/journal.pone.0067740

Wu, J., Zhao, Q., Liu, S., Shahid, M., Lan, L., Cai, G., Zhang, C., Fan, C., Wang, Y., & Zhou, Y. (2016). Genome-wide Association Study Identifies New Loci for Resistance to Sclerotinia Stem Rot in Brassica napus. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/article/10.3389/fpls.2016.01418

Wu, X.-L., Xu, J., Feng, G., Wiggans, G. R., Taylor, J. F., He, J., Qian, C., Qiu, J., Simpson, B., Walker, J., & Bauck, S. (2016). Optimal Design of Low-Density SNP Arrays for Genomic Prediction: Algorithm and Applications. *PLOS ONE*, *11*(9), e0161719. https://doi.org/10.1371/journal.pone.0161719

Xu, Y., & Crouch, J. H. (2008). Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science*, *48*(2), 391–407. https://doi.org/10.2135/cropsci2007.04.0191

Yang, Q., Yang, Z., Tang, H., Yu, Y., Chen, Z., Wei, S., Sun, Q., & Peng, Z. (2018). High-density genetic map construction and mapping of the homologous transformation sterility gene (hts) in wheat using GBS markers. *BMC Plant Biology*, *18*(1), 301. https://doi.org/10.1186/s12870-018-1532-x

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1), 134. https://doi.org/10.1186/1471-2105-13-134

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 2. https://doi.org/10.1038/ng1702

Yue, B., Cai, X., Yuan, W., Vick, B., & Hu, J. (2009). Mapping quantitative trait loci (QTL) controlling seed morphology and disk diameter in sunflower (Helianthus annuus L.). *Helia*, *32*(50), 17–35.

Zhao, H., Basu, U., Kebede, B., Qu, C., Li, J., & Rahman, H. (2019). Fine mapping of the major QTL for seed coat color in Brassica rapa var. Yellow Sarson by use of NIL populations and transcriptome sequencing for identification of the candidate genes. *PloS One*, *14*(2), e0209982. https://doi.org/10.1371/journal.pone.0209982

Zhao, J., & Meng, J. (2003). Detection of loci controlling seed glucosinolate content and their association with Sclerotinia resistance in Brassica napus. *Plant Breeding*, *122*(1), 19–23. https://doi.org/10.1046/j.1439-0523.2003.00784.x

Zheng, Boer, M. P., & van Eeuwijk, F. A. (2019). Construction of Genetic Linkage Maps in

Multiparental Populations. *Genetics*, *212*(4), 1031–1044. https://doi.org/10.1534/genetics.119.302229

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., Liu, J., Sun, X., Wang, X., Xu, J., Hua, W., & Wang, H. (2017). Genome-Wide Association Study Reveals Candidate Genes for Control of Plant Height, Branch Initiation Height and Branch Number in Rapeseed (*Brassica napus* L.). *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01246

Zhou, D., Wang, C., Li, Z., Chen, Y., Gao, S., Guo, J., Lu, W., Su, Y., Xu, L., & Que, Y. (2016). Detection of Bar Transgenic Sugarcane with a Rapid and Visual Loop-Mediated Isothermal Amplification Assay. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/article/10.3389/fpls.2016.00279

Zhou, F., Liu, Y., Liang, C., Wang, W., Li, C., Guo, Y., Ma, J., Yu, Y., Fan, L., Yao, Y., Zhao, D., Liu, X., & Huang, X. (2018). Construction of a high-density genetic linkage map and QTL mapping of oleic acid content and three agronomic traits in sunflower (Helianthus annuus L.) using specific-locus amplified fragment sequencing (SLAF-seq). *Breeding Science*, *68*(5), 596–605. https://doi.org/10.1270/jsbbs.18051

Zhou, Q., Han, D., Mason, A. S., Zhou, C., Zheng, W., Li, Y., Wu, C., Fu, D., & Huang, Y. (2018). Earliness traits in rapeseed (Brassica napus): SNP loci and candidate genes identified by genome-wide association analysis. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *25*(3), 229–244. https://doi.org/10.1093/dnares/dsx052

Zhou, W.-C., Kolb, F. L., Bai, G.-H., Domier, L. L., Boze, L. K., & Smith, N. J. (2003). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breeding*, *122*(1), 40–46. https://doi.org/10.1046/j.1439-0523.2003.00802.x

Zhu, M., Monroe, J. G., Suhail, Y., Villiers, F., Mullen, J., Pater, D., Hauser, F., Jeon, B. W., Bader, J. S., Kwak, J. M., Schroeder, J. I., McKay, J. K., & Assmann, S. M. (2016). Molecular and systems approaches towards drought-tolerant canola crops. *New Phytologist*, *210*(4), 1169–1189. https://doi.org/10.1111/nph.13866

Zhu, Q., King, G. J., Liu, X., Shan, N., Borpatragohain, P., Baten, A., Wang, P., Luo, S., & Zhou, Q. (2019). Identification of SNP loci and candidate genes related to four important fatty acid composition in *Brassica napus* using genome wide association study. *PLOS ONE*, *14*(8), e0221578. https://doi.org/10.1371/journal.pone.0221578

Zubrzycki, J. E., Maringolo, C. A., Filippi, C. V., Quiróz, F. J., Nishinakamasu, V., Puebla, A. F., Rienzo, J. A. D., Escande, A., Lia, V. V., Heinz, R. A., Hopp, H. E., Cervigni, G. D. L., & Paniego, N. B. (2017). Main and epistatic QTL analyses for Sclerotinia Head Rot resistance in sunflower. *PLOS ONE*, *12*(12), e0189859. https://doi.org/10.1371/journal.pone.0189859