Genome-wide association and genomic selection for oil and fatty acid profile in rapeseed (*Brassica napus* L.)

by

Jared Bento

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Plant Science

University of Manitoba

Winnipeg

ACKNOWLEDGEMENTS

On the first day of graduate studies, no one receives the skills or confidence required of a researcher. On the last day, I owe this distinction to the many remarkable people I've come to know along the way.

To Dr. Rob Duncan, my supervisor, who I first knew as the professor for PLNT 2520 where his enthusiasm in teaching plant science and genetics made a lasting impression on many including myself. When I first considered taking on MSc. research, Dr. Duncan shared his time to start us on the right path. His commitment to supporting students will impress upon my own peers in the future.

To the members of my committee, Dr. Harmeet Chawla, Dr. Curt McCartney, and Dr. Francis Zvomuya, for the meetings which answered many of my most difficult questions, but also sometimes filled my notebook with more questions than I arrived with. Your expertise helped to make the most of our time with this challenging research.

To many others who I've had the pleasure of learning from, you are living proof that no one should be afraid to ask for help. Namely, Dr. Marcos Cordeiro, Dr. Gabriel Dallago, Dr. Rob Guderian, Dr. Mike Domaratzki, Dr. Sven Weber, Dr. Subhadra Chakrabarty, Judith Nugent-Rigby, and especially Sakaria Liban – these people extended a helping hand from many disciplines. I consider you as positive influences forming the researcher I am today.

To my peers, who were always glad to share the trials and tribulations of grad research, I thank you. The PSGSA Halloween parties, Christmas potlucks, BBQs, tours, and conferences we organized and attended together were times that I'll remember most from plant science.

I couldn't have made it through this process without the unsung heroes of advisors and administrative staff: Ekta Badesha, Martha Blouw, Bev Godard, Alina Skorobohatova, Mariana Garcia Romero, and many others.

Lastly, I share my appreciation to NSERC, Nutrient Ag Solutions, Bunge Canada, and all other funders for supporting research across the Department of Plant Science.

DEDICATION

Dedicated to my parents, Maria and John, for their encouragement – practicing my proposal, seminar, and defense together in our living room was more important than you know.

To Avô, for being there every day after elementary school – you'd be happy to see where I am now.

TABLE OF CONTENTS

List of Tables

List of Figures

ABSTRACT

Overarching goals of rapeseed (*Brassica napus* L.) breeding efforts include the improvement of yield- and seed-quality-related traits to meet the evolving demands of a growing population. Genome-wide association studies (GWAS) and genomic selection (GS) are methods that provide the potential to improve selection efficiency, facilitating responses to agronomic and quality challenges while enhancing the sustainability of plant breeding programs. We applied these biotechnologies to *B. napus* L. in pursuance of three main objectives: 1) GWAS to identify quantitative trait loci (QTL) for five seed quality traits (overall oil content, erucic, oleic, linoleic, and linolenic acids); 2) evaluating GS prediction accuracy for hybrid seed quality traits; and 3) evaluating the GWAS-guided GS method proposed to improve GS prediction accuracy. We analyzed 454 *B. napus* L. genotypes (92 parents, 362 hybrids) across 48 site-years in the three Canadian prairie provinces. FarmCPU GWAS analyses identified 89 peak QTL, including 14 QTL for oil content. Several QTL coincide with candidate genes identified in previous studies, while novel QTL warrant further candidate gene investigation. GS prediction accuracies were compared across 135 unique GS analyses for each trait, evaluating responses to factors including model choice (nine regression models), population (five training/validation population designs), and marker density (three marker sets containing low, intermediate, and high densities). Prediction accuracies (represented by Pearson's correlation coefficient (r) for the relationship between predicted and actual phenotypes) ranged from negative values (oil content) to 0.89 (linoleic acid content); however all five traits could be predicted with $r > 0.70$ depending on the combination of aforementioned factors. Prediction accuracies exhibited negative responses to increasing trait complexity, positive responses to increasing training population size and degree of training/validation population relatedness, and no significant differences among marker densities or parametric models. Machine learning models performed either equivalent to or worse than common parametric models. GWAS-guided GS exhibited slight numeric improvements relative to conventional GS accuracies for the same traits. Although improvements were not statistically significant, the consistency of extremely low-density marker sets is conducive to reducing genotyping density while maintaining or improving genetic gains. The promising accuracy of GS techniques in this study supports their potential implementation in future *Brassica* breeding programs.

FOREWORD

This thesis has been written in the grouped manuscript style and follows the format guidelines outlined by the Faculty of Graduate Studies at the University of Manitoba. The thesis includes a general introduction, a literature review, three manuscripts, a general discussion and future research recommendations. Each of the three manuscripts will be submitted to Molecular Breeding or another suitable journal.

CONTRIBUTIONS OF AUTHORS

The *Brassica napus* L. population was sourced from University of Manitoba High Erucic Acid Rapeseed (HEAR) breeding program, managed by Dr. Rob Duncan. Plot design and replicates designed by Dr. Duncan. Population was previously analyzed in an earlier study (Sun 2021). DNA extraction by Dr. Jia Sun. Genotyping by Dr. Isobel Parkin. Seed quality phenotyping by Debbie Witko, Kelsey Dickson and Judith Nugent-Rigby, as part of Dr. Duncan's HEAR breeding program. Study conceived by Dr. Duncan. Experiments 1, 2, and 3 originally designed by Dr. Sun. Experiments 1, 2, and 3 redesigned, proposed, and conducted by Jared Bento. Sakaria Liban and Dr. Mike Domaratzki provided assistance with machine learning model refinement in experiment 2. Manuscripts for experiments 1, 2, and 3 written by Jared Bento under supervision and review by co-advisors Dr. Rob Duncan Dr. Harmeet Chawla, reviewed by committee members Dr. Curt McCartney and Dr. Francis Zvoyuma.

1. GENERAL INTRODUCTION

The first canola-quality *Brassica napus* L. cultivar, Tower, was released by the University of Manitoba in 1974, setting in motion fifty years of plant breeding achievements which have established *B. napus* as the most widely sown species on Canadian acreage (Daun 2011; Statistics Canada 2024; Stefansson and Kondra 1975). The value of *B. napus* is derived by separating its high-quality oil from nutritional seed meal. In 2023, Canada exported over 6.6 million metric tonnes of seed and 2.8 million metric tonnes of oil after canola farmers harvested 18.3 million metric tonnes (Canadian Oilseed Processors Association 2023; Canola Council of Canada 2024b; Statistics Canada 2024). Canada leads the international *B. napus* market – between 2004 to 2016, approximately 70 % of total international canola exports were Canadian (Barthet 2016). The rising demand for *B. napus* can be attributed to its diverse range of applications including healthy cooking oils, livestock feed, biofuels, biodegradable plastics, and household products such as cleaners and cosmetics (Aoki et al. 2015; Aukema et al. 2011; Leonard 1992; Wang et al. 2022; Wolinetz 2017).

Using conventional breeding methods, *B. napus* was transformed from an inedible oilseed suitable only for industrial uses to a cornerstone of Canadian food, feed, fibre, and renewable energy industries. For example, Tower and its respective low-erucic-acid and low-glucosinolate progenitors, Oro and Bronowski, were all produced through conventional pedigree selection methods (McVetty et al. 2009; Stefansson et al. 1961; Stefansson and Downey 1995; Stefansson and Kondra 1975). Alongside canola, the improvement of rapeseed and high erucic acid rapeseed (HEAR) (low glucosinolates with erucic acid contents over 50 %) focuses on increasing yield, erucic acid content, and oil content, while reducing glucosinolate content, days to maturity, and plant height (McVetty et al. 2009). Several HEAR cultivars have been developed using conventional breeding methods, including the “Millenium” series which incorporated blackleg resistance (McVetty et al. 2009).

Further improvements to canola yield and pest resistance were achieved through biotechnological advancement in the 1990s and 2000s. The development of mutagenesis, transgenics, and cytoplasmic male sterility (CMS) systems gave rise to high-yielding herbicide-tolerant hybrid cultivars (Morrison 2016). Similarly, as blackleg and clubroot became a rising threat to canola farmers, resistant cultivars were widely adopted to protect yield (Canola Council of Canada 2024c).

More recent contributors to *B. napus* improvement include marker-assisted selection (MAS) and high-throughput genotyping techniques.

The development of hybrid canola cultivars, which have constituted more than 95 % of canola acreage in the Canadian prairies since their introduction, is facilitated by MAS. Hybrid breeding utilizes MAS to select for sterility and restorer genes in CMS systems (Hansen et al. 1997; Liu et al. 2005, Huang et al. 2012). Relative to conventional (phenotypic) selection, MAS enhances breeding efficiency by facilitating selections based on linkage between causative genes and sequence polymorphisms (e.g. markers linked to sterility genes) (Collard et al. 2005). The broad applications of molecular markers stimulated their adoption for *B. napus* improvement and other important crop species (Moose and Mumm 2008; Nadeem et al. 2017). In addition to CMS systems, MAS has also been used to improve agronomic and pest resistance traits in *B. napus* including overall oil content, fatty acid profile (oleic, erucic, linoleic acid), seed coat colour, lodging resistance, and resistance to clubroot, blackleg, and white rust. (Cheung et al. 1998; Hirani et al. 2016; Jourdren et al. 1996a; Jourdren et al. 1996b; Kawasaki et al. 2021; Liu et al. 2005; Li and Du 2023; Rahman et al. 2014; Rakow et al. 1999; Somers et al. 1999; Spasibionek et al. 2010). When combined with high-throughput genotyping data such as the popular and cost-effective *Brassica* 60K Illumina Infinium™ SNP array, recent *B. napus* genome assemblies also provide reliable alignment and gene annotation for downstream research including MAS (Mason et al. 2017; Rousseau-Gueutin et al. 2020).

The increasing availability and quality of genome-wide marker data realizes the potential of genome-wide association studies (GWAS) (Wainschtein et al. 2019; Young 2019). The main objective of GWAS is the identification of markers which are significantly associated to traits of interest, known as quantitative trait loci (QTL) (Uffelmann et al. 2021). With large and diverse populations, GWAS can provide high resolution for QTL identification (Lander and Schork 1994; Risch and Merikangas 1996). Numerous studies have utilized GWAS to investigate *B. napus* traits including yield, flowering time, height, oil content, fatty acid profile, germination, clubroot resistance, and blackleg resistance (Cantila et al. 2023; Dakouri et al. 2021; Gacek et al. 2017; Hatzig et al. 2015; Helal et al. 2021; Huang et al. 2021; Korber et al. 2016; Li et al. 2016; Liu et al. 2022; Luo et al. 2021; Pal et al. 2021; Qin et al. 2022; Qu et al. 2017; Raman et al. 2019; Schiessl et al. 2015; Tang et al. 2021; Xiang et al. 2023; Xue et al. 2018; Yang et al. 2023; Zhang

et al. 2023). Results from GWAS are often implemented to MAS, but are also used in linkage mapping, investigating candidate gene function, genomic selection, transgenics, and other applications (Deng et al. 2017; Liu et al. 2019; Bian and Holland 2017; Fiedler et al. 2017; Zhu et al. 2011; Tsai et al. 2020; Wang et al. 2016).

Plant breeders have turned their attention to other methods of marker-based selection to drive genetic gain towards future populations' demand. Genomic selection (GS) has become highly attractive to breeders by capturing extensive genotypic variation, detecting small-effect genes for the prediction of complex traits (Heffner et al. 2010; Jubair and Domaratzki 2023; Krishnappa et al. 2021; Varshney et al. 2017). By combining phenotypic and genotypic data from a training population (e.g. parental genotypes) with only the genotypic data from a breeder's testing population (e.g. hybrid lines), GS models produce phenotypic predictions known as genomic estimated breeding values (GEBVs) (Habier et al. 2007; Jannink et al. 2010). These GEBVs, based on genotypic data collected at the seedling stage, are valuable criteria to inform selections within a breeder's population (represented by a validation population (VP) in cross-validation methods evaluating GS performance). Informing the removal of undesirable germplasm with GEBVs (which would otherwise require multiple generations of phenotypic data to inform) is a significant benefit to plant breeders, facilitating the allocation of resources to desirable germplasm more likely to contribute to the final cultivar. This lessens the volume of required phenotyping, improving genetic gain per unit time while reducing costs per breeding cycle and (potentially) the overall number of breeding cycles (Hickey et al. 2014).

Identifying an "optimal" GS method is highly subjective – currently, no methods are known to be effective in all circumstances (Werner et al. 2020; Zhang et al. 2023). The subjectivity of GS accuracy can be attributed to factors which inherently differ among breeding programs. Factors such as trait architecture, marker density, population design, and choice of GS model produce wide variations in prediction accuracy. This subjectivity contributes to the notion regarding GS methodology: there is no "one size fits all" (Lorenz et al. 2011).

Studies on *B. napus* have utilized GS to predict yield traits, seed-quality traits, morphology, and disease resistance (Derbyshire et al. 2021; Fikere et al. 2020; Hu et al. 2021a; Koscielny et al. 2020; Luo et al. 2017; Roy et al. 2021; Zou et al. 2016). These studies demonstrate the predictive power achievable with GS, but also its highly subjective nature among traits, population designs,

and models. One technique, known as GWAS-guided GS, has been proposed to improve GS accuracy by focusing model training on informative markers from GWAS results (Spindel et al. 2016; Werner et al. 2018; Zhang et al. 2023). GWAS-guided GS remains largely unexplored, having only been applied to *B. napus* by Werner et al. (2017), Werner et al. (2018), and Sun (2021).

This following research included three experiments to evaluate GWAS, GS, and GWAS-guided GS in application to five seed quality traits in *B. napus*: overall oil content (OIL), erucic acid content (ERU), linoleic acid content (LLE), linolenic acid content (LLN), and oleic acid content (OLE). We analyzed a breeding population of 454 high-erucic acid rapeseed (HEAR) genotypes phenotyped across 48 site years in the Canadian prairie provinces. Each accession was genotyped using the *Brassica* 60K SNP array. The main objective of the first study (Chapter 3) was to identify seed quality QTL for potential implementation to HEAR breeding. The first study also aimed to demonstrate how population structure, marker set density, and choice of model affect QTL discovery. The second study (Chapter 4) evaluated GS in *B. napus* by observing an extensive range of potential seed quality breeding scenarios. For each trait, 135 unique analyses were conducted, demonstrating prediction accuracy's response to GS model (nine regression models), population (five training/validation population designs), and marker density (three marker sets containing low, intermediate, and high densities). The objectives of the final study (Chapter 5) were to predict seed quality traits with GWAS-guided GS, utilizing a similar array of analyses to compare accuracy to conventional GS from Chapter 4.

The overall objective of this research was to demonstrate the effectiveness of GWAS and GS in application to seed quality improvement in *B. napus*. Plant breeders may also observe the responses in QTL discovery and GS prediction accuracy to inform their decisions on factors (population design, genotyping density, choice of model, etc.) according to their goals and available resources.

2. LITERATURE REVIEW

2.1 History of *Brassica napus* L.

Brassica napus L. (rapeseed) is an economically important crop worldwide. Behind only soybeans (*Glycine max* L.), *B. napus* is the second-largest oilseed commodity worldwide, producing 84.34 million metric tons in 2022 (USDA 2023). Alongside *B. napus*, the *Brassicaceae* family contains 338 genera and over 3,700 other species, many of which are also important agricultural species (Petruzello 2023). Crops such as broccoli (*Brassica oleracea v. italica* L.), cabbage (*Brassica oleracea v. capitata* L.), bok choy (*Brassica rapa, v. chinensis* L.), radish (*Raphanus sativus* L.), kale (*Brassica oleracea v. acephala* L.), and more comprise economically significant members of the *Brassicaceae* family (Petruzello 2023). The *Brassica* genus hosts the most closely related crops to the model species *Arabidopsis thaliana* L. (Yang et al. 1999).

On an evolutionary scale, *B. napus* crops are relatively "young" considering the domestication of other important crops before it such as wheat (*Triticum aestivum* L.) (~10,000 years ago) and soybean (~5,000 years ago) (Eckardt 2010; Jeong et al. 2019). Debate surrounds the origin of *B. napus*, however one leading theory suggests that approximately 7,500 years ago, a hybridization event (or series of hybridization events) occurred between *B. rapa* (genome AA, 2n = 20) and *B. oleracea* (genome CC, 2n = 18), forming the allopolyploid *B. napus* (genome AACC, 2n = 38) (Chalhoub et al. 2014). The *Brassica* "Triangle of U" (Figure 2.1), which includes several important *Brassica* crops, illustrates the origination of allotetraploid *Brassica* species through the hybridization of diploid *Brassicas* (Rakow 2004). The centre of origin for *B. napus* is also debated to have occurred in Mediterranean Europe or Asia (Downey and Rimmer 1993; Friedt and Snowdon 2009; McNaughton 1976). The earliest record of *Brassica* cultivation (probably *B. rapa*) is in India as early as 4000 BC, before its spread through Europe from 2000 BC to approximately the 13th century (Friedt and Snowdon 2009). European domestication of *B. napus* is thought to have occurred in 16th century Netherlands (Daun 2011; Friedt and Snowdon 2009; Gómez-Campo 1999; McNaughton 1976; OECD 1997). In the following period, rapeseed oil was used primarily as lamp oil or, in some cultures, as an edible oil (Freidt and Snowdon 2009; OECD 1997). Cultivation spread eastward and northward into Scandinavia, Russia, and Poland by the 19th century (Bell 1982; Kroll 1994; Kimber and McGregor 1995; Friedt and Snowdon 2009).

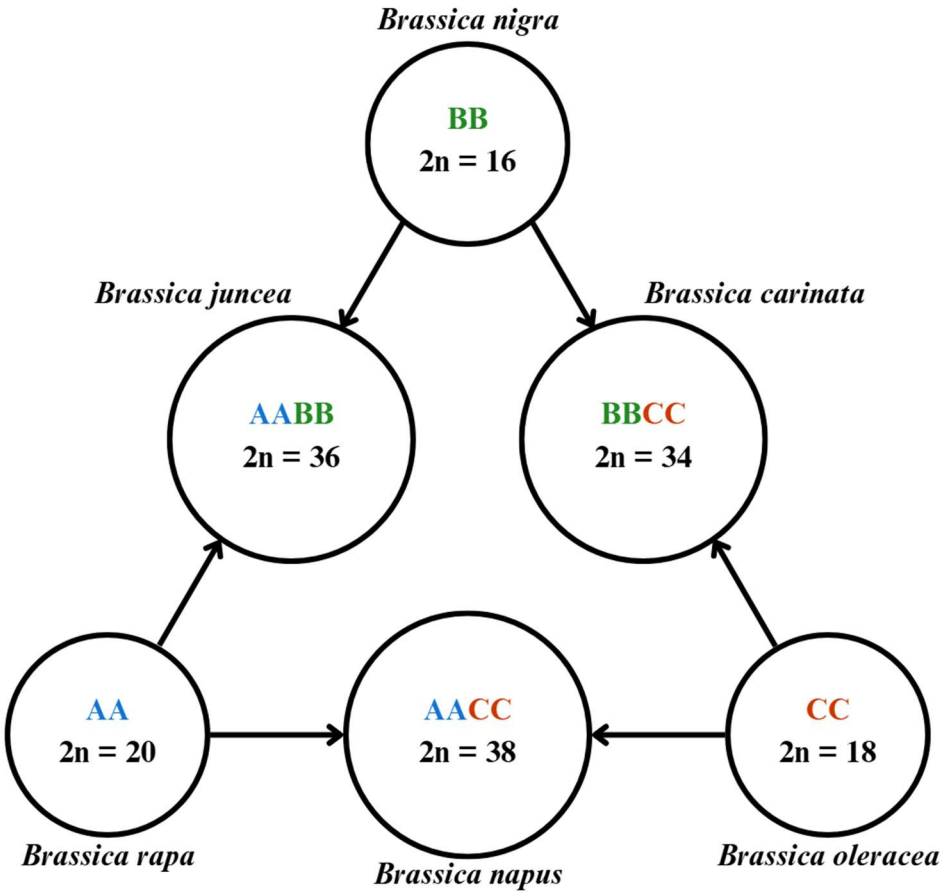


Figure 2.1 Triangle of U, representing the relationship between three diploid *Brassica* species and their hybridizations giving rise to three allopolyploid *Brassica* species. Adapted from U, 1935.

Canadian soils first grew rapeseed in 1936 when Fred Solvonik, an immigrant farming in Shellbrook, Saskatchewan, received *B. rapa* seeds from his native Poland (Bell 1982). Rapeseed oil production increased significantly during World War II for use as a lubricant in naval and merchant ship engines (Daun 2011; Canola Council of Canada 2024c). To supplement the production of rapeseed oil while seed imports from Europe could not be made during World War II, the government of Canada imported *B. napus* from Argentina (Daun 2011). These original sources of *B. rapa* and *B. napus* gave rise to the common names of the two species in Canada as Polish and Argentine rapeseed, respectively (Daun 2011). *Brassica* oilseed varieties developed from the original *B. rapa* and *B. napus* sources showed excellent adaptation to the Canadian prairies and became established as significant crops in the 20 years following World War II, meeting demand as a domestic commodity as well as export commodity, especially to Japan (Daun 2011). Up to this period, rapeseed oil contained high levels of a fatty acid known as erucic acid while its seed meal contained compounds known as glucosinolates. Following reports of erucic acid's negative effects to heart health and reports of livestock feed containing glucosinolates exhibiting toxicity, breeders turned their efforts to producing a *Brassica* whose oilseeds could fit this industrial niche (Daun 2011). "Canola" was developed through artificial selection of rapeseed wherein the low-erucic acid rapeseed (LEAR) cultivar "Liho" and low-glucosinolate cultivar "Bronowski" were crossed and subsequently backcrossed with the popular cultivar "Turret," creating a cultivar whose oil contains low amounts of both erucic acid and glucosinolates (AgMRC 2022; Stefansson and Kondra 1975). At the University of Manitoba, Dr. Baldur Stefansson released the first "canola-quality" *B. napus* cultivar named "Tower" in 1974 (Daun 2011). The term "canola" describes seeds of the genus *Brassica* (*B. napus*, *B. rapa*, or *B. juncea*) of which the oil is "double-low," more specifically, "contains less than 2 % erucic acid in its fatty acid profile and the solid component shall contain less than 30 micromoles of any one or any mixture of 3-butenyl glucosinolate, 4-pentenyl glucosinolate, 2-hydroxy-3 butenyl glucosinolate, and 2-hydroxy- 4-pentenyl glucosinolate per gram of air-dry, oil-free solid" (Canola Council of Canada 2024c).

2.2 Market for *Brassica napus* L.

The significance of canola oil in food and feed industries can be attributed to a combination of desirable qualities. Canola oil contains no trans fats, while its saturated fatty acid content is also very low relative to other vegetable oils (CanolaInfo n.d.). Canola oil also contains a source of

omega-6 and omega-3 fatty acids, constituted by its polyunsaturated fat content (Canadian Grains Commission 2022). Finally, monounsaturated fatty acid composes the largest proportion of the canola oil profile (Canadian Grains Commission 2022). These qualities represent an edible oil which aids in cardiovascular health, immune system, diabetes, and cancer (Stricker et al. 2008; Iggman et al. 2011; Risérus et al. 2009; Ren et al. 2010; Granado-Casas and Mauricio 2019; Gnoni et al. 2010; Raboanatahiry et al. 2021). Three main fatty acids constitute canola oil: oleic acid (18:1), linoleic acid (18:2) and linolenic acid (18:3), along with a small proportion of saturated fatty acids and other trace monounsaturated and polyunsaturated fatty acids (Canadian Grains Commission 2022; Przyblyski and Eskin 2011). Oleic acid (OLE) is an omega-9 fatty acid with an 18:1 backbone (18 carbons, monounsaturated) (Gnoni et al. 2010). Importantly, OLE is a precursor in the biosynthesis of linoleic, linolenic, and erucic acids (Figure 2.2) (Kaur et al. 2019). Linoleic acid (LLE) is an omega-6 polyunsaturated fatty acid with an 18:2 backbone that serves as an essential component in the human diet – it is regarded as the most consumed polyunsaturated fatty acid worldwide (Whelan and Fritsche 2013). In the human diet, LLE is responsible for several bodily functions including sources of energy, precursors to membrane lipids, and families of signaling molecules (Smith and Borgeat 1985; Whelan and Fritsche 2013). Linolenic acid (LLN) is an omega-3 polyunsaturated fatty acid with an 18:3 backbone derived from LLE (Whelan and Fritsche 2013). Along with LLE, LLN is also considered an essential dietary fatty acid, however, the number of double bonds in LLN impair frying quality (lower thermal stability) and shorten shelf life relative to other monounsaturated oils (Browse et al. 1998; Tanhuanpää and Schulman 2002; Fitzpatrick and Scarth 1998). Significant plant breeding efforts have been applied to "specialty oil profiles" – oilseed cultivars whose oils contain proportions of fatty acids conducive to certain goals. One of the earliest objectives for canola specialty oil profiles was the reduction of LLN. The world's first low-LLN cultivar "Stellar," developed at the University of Manitoba, contained less than 3 % LLN (Scarth et al. 1987). Since then, several competitively yielding low-LLN canola cultivars have been released to fulfill this niche market (McVetty et al. 2009). In 2022–2023, canola oil consumption reached 29.8 million metric tonnes worldwide (Hameed 2023). Canadian farmers harvested 18.3 million metric tonnes of canola seed in 2023, of which, 7.9 million metric tonnes were crushed in Canada to produce 3.3 million metric tonnes of oil (Statistics Canada 2024; Canadian Oilseed Processors Association 2023). A large proportion of harvested canola seed is exported internationally – over 6.6 million metric tonnes of seed and 2.8

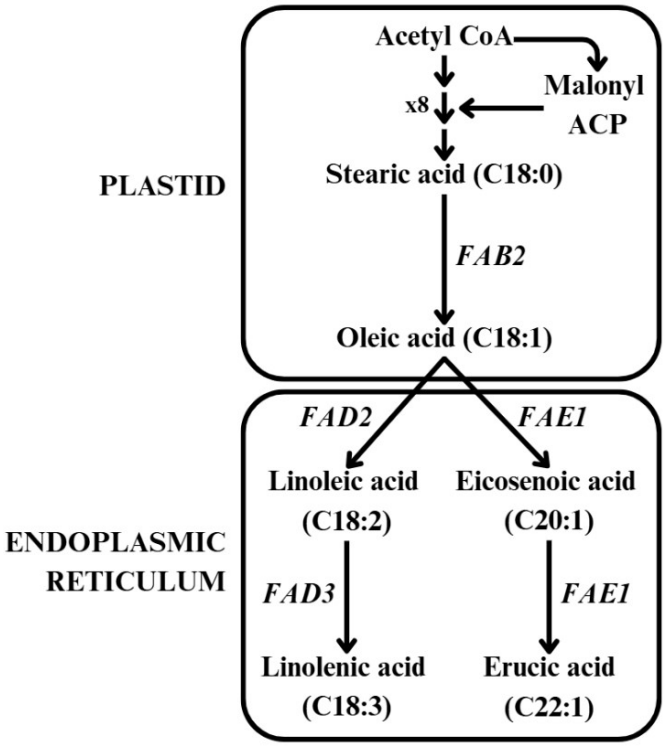


Figure 2.2 Simplified flowchart of fatty acid biosynthesis in *A. thaliana*. Adapted from Kaur et al. (2019).

million metric tonnes of oil were exported in 2023 (Canola Council of Canada 2024b). Between 2004 to 2016, approximately 70 % of total international canola exports were Canadian (Barthet 2016). Overall, the canola industry generated an average of $43.7 billion to the Canadian economy between 2020 and 2023, employing approximately 206,000 Canadians over that time (Canola Council of Canada 2024a; GlobalData 2024).

While canola is distinguished as *B. napus* genotypes whose seeds are "double-low," high erucic acid rapeseed (HEAR) is distinguished as *B. napus* genotypes whose seeds contain the same standard of low glucosinolates but typically 50 % or more erucic acid (ERU) in the fatty acid profile (Sanyal et al. 2015). The HEAR fatty acid profile is very similar to canola's in that it also contains OLE, LLE, and LLN, however ERU constitutes a significant proportion of the total oil profile, accompanied by a reduction in other fatty acids relative to canola. ERU is a monounsaturated omega-9 fatty acid with a 22:1 backbone (U.S. National Library of Medicine 2004). Through various processes, ERU can be converted to several forms with broad industrial applications. Several methods have been implemented for the production of biodiesel using *B. napus* oil, a renewable fuel capable of reducing millions of tonnes of greenhouse gas emissions relative to fossil fuels. (Wang et al. 2022; Wolinetz 2017). Through hydrogenation, ERU can produce behenic acid: a compound useful in lubricants, plasticizers, and stabilizers in plastic, pharmaceutical, and food industries (Bährle-Rapp 2007; Wang et al. 2022). Oxidation reactions convert ERU to tridecanedioic acid or nonanoic acid used in the production of nylon, fragrances, and low-temperature plasticizers (Wang et al. 2022). Other forms of ERU can act as emulsifiers or emollients used in cosmetic products, silver behenate used in photography, and pharmaceutical compounds such as behenyl alcohol, the antiviral agent in Abreva® (Aukema et al. 2011; Leonard 1992; Aoki et al. 2015). Alongside canola, approximately 56,000 metric tonnes of HEAR oil are consumed per year (USDA 1996). The United States is the largest consumer of HEAR oil, importing most of the 18,000 metric tonnes consumed domestically per year (Bhardwaj and Hamama 2000). Europe and Japan are among the other major importers of HEAR oil (Bhardwaj and Hamama 2000). Of its wide range of industrial applications, HEAR oil is used mainly in the production of lubricants, slip agents, and biofuels. *B. napus* oil constituted approximately 45 % of EU biodiesel consumption in 2016 (Dusser 2019).

2.3 Plant breeding progress in *Brassica napus* L.

Since the release of Tower, fifty years of advancement have propelled *B. napus* to the single most widely cultivated species on Canadian acreage (Statistics Canada 2024). Herbicide tolerance was first introduced to the industry in 1984 with the release of "OAC Triton", a triazine-tolerant *B. napus* cultivar produced through classical breeding with *B. campestris* (Beversdorf and Kott 1987). The canola industry of the 1990s was marked by a massive spike in herbicide-tolerant cultivar development. In 1995, the first imidazolinone-tolerant *B. napus* cultivar "45A71" was registered, whose tolerance was imparted through a mutagenesis-induced trait (Swanson et al. 1989; Government of Canada 2016). In the same year, a major milestone was achieved with the registration of two herbicide resistant cultivars produced through transgenics: "Quest," resistant to glyphosate herbicides, and "Innovator," resistant to glufosinate herbicides (Canola Council of Canada 2024c; Harker et al. 2000). The University of Manitoba also developed multiple bromoxynil-resistant cultivars beginning in the 1990s (McVetty et al. 2001). Between 1995 to 2001, over 100 herbicide tolerant *Brassica* cultivars received registration recommendation (Canola Council of Canada 2023).

Also of significance was the implementation and popularization of hybrids. The discovery of a *B. napus* cytoplasmic male sterile (CMS) system brought the production of the first canola hybrid "Hyola 40" in 1989 (Canola Council of Canada 2024c; Canadian Food Inspection Agency 1989). The "Ogura" CMS system has been widely implemented by *Brassica* breeders due to its practical transferability and full pollen abortion for use in hybrid cultivar production (Ren et al. 2022). Since their introduction, more than 95 % of seeded canola acreage in the Canadian Prairies has been constituted by herbicide tolerant hybrid canola cultivars (Morrison 2016). It is estimated that, from 2000-2013, the adoption of herbicide tolerant hybrid cultivars contributed to approximately 32 of the 54 $kg\,ha^{-1}\ year^{-1}$ yield gain observed throughout that period (Morrison 2016). The popularization of herbicide tolerant hybrids could be attributed to their high yields, excellent weed control achieved with lesser herbicide applications, and the resultant increase in profits compared to traditional *B. napus* (O'Donovan et al. 2006).

Improvements in *B. napus* cultivation have also adapted to disease pests, most notably blackleg and clubroot, each capable of decimating crop yields once established in soils. Since its discovery in Saskatchewan in 1975, the rising threat of blackleg disease in Canada saw yield losses ranging

from 30-50 %, prompting Canadian breeders to develop blackleg-resistant varieties by the early 1990s (McGee and Petrie 1978; Wang et al. 2023; Hall et al. 1993; Barbetti and Khangura 2001; Kutcher et al. 2010). Similarly, after its discovery in Canada in 1997, clubroot quickly became a serious concern for Canadian farmers, contributing to yield losses as great as 91 % in susceptible varieties (Pageau et al. 2006). The first clubroot-resistant canola cultivar for the Canadian market was registered in 2009 (Hwang et al. 2017). Nearly all cultivars registered now are resistant or moderately resistant to both blackleg and clubroot, however, the emergence of new virulent isolates has caused significant erosion of resistance in previously resistant cultivars (Canola Council of Canada 2024c; Hwang et al. 2016). The breakdown of disease resistance highlights the importance of continued investigation into new and improved modes of resistance.

2.4 *Brassica napus* L. genome

The *B. napus* genome is an allotetraploid, consisting of two copies of the "A" genome (*Brassica rapa,* 10 chromosomes) and two copies of the "C" genome (*Brassica oleracea,* 9 chromosomes), otherwise denoted AACC, or 2n = 4x = 38 (Hu et al. 2021a). As an allopolyploid whose ancestors themselves are polyploids, many *B. napus* genes exist as copies across its respective *Brassica* genomes (Mason et al. 2017).

Several *B. napus* genome assemblies have been created by cataloguing sequence data. The first of these genome assemblies as was called "*Darmor-bzh,*" based on the European winter rapeseed cultivar with the same name (Chalhoub et al. 2014). The *Darmor-bzh* genome assembly is extremely significant for *B. napus* development, with wide applications to plant breeding and genetics research.

At the time of writing, there are three versions of *Darmor-bzh*: the original by Chalhoub et al. (2014), an updated version by Bayer et al. (2017), and the most recent assembly by Rousseau-Gueutin et al. created using long-read sequencing technology (2020). Relative to the original *Darmor-bzh* assembly, Rousseau-Gueutin et al. achieved significant improvements in several important metrics of genome assembly quality. Most notably, long-read sequencing technology produced an extremely contiguous assembly with N50 = 11,486,274 bases (i.e. 50 % of the assembly's contigs were sequences longer than the N50), relative to the original whose N50 = 37,644 bases (Rousseau-Gueutin et al. 2020). Also compared to the original *Darmor-bzh* assembly

which was overall 850 Mb, with 645 Mb confidently anchored, and only 80 % of 101,040 genes located on specific chromosomes, the recent long-read assembly is 924 Mb in length, containing 867 Mb anchored, and 98.8 % of 108,190 genes located on chromosomes (Rousseau-Gueutin et al. 2020).

High-quality *B. napus* genome assemblies allow the investigation of polyploidy's effect on structural and the functional activity of duplicated genes and their connection to important agronomic traits. These resources also allow insights into many of the phenomena instilled by *B. napus'* polyploidy genome, including changes in gene and protein expression, transposon activation, epigenetics, and alternative splicing (Rousseau-Gueutin et al. 2020).

2.5 Conventional breeding

"Relative fitness", a factor in population genetics which describes the probability of certain genotypes to survive, reproduce, and pass their alleles to the next generation, is the basis of evolution through natural selection (Orr 2009). For example, genotypes exhibiting the greatest seed yield constitute larger portions of subsequent generations. Similarly, genotypes with certain qualities such as flavour, milling ability, or oil content were selected and sown by farmers, contributing their alleles to next generations. In practice, the relative fitness of these genotypes is imposed by a farmer's selection – "favourable" alleles in the gene pool are influenced not by natural selection, but instead "artificial" selection – this is the basis of plant improvement and conventional plant breeding.

The domestication of *B. napus* is an example of conventional breeding. Preliminary goals of Canadian *B. napus* breeding were the improvement of yield, seed quality, and environmental adaptation, which included earlier maturity, plant height, and uniformity (Stefansson and Downey 1995). Conventional breeding methods carried *B. napus* from an un-adapted crop species with inedible oil suited only for industrial use to an established cornerstone of Canadian food, feed, and fuel industries. The first Canadian-bred *B. napus* cultivar, "Golden", was registered in 1954 (Canola Council of Canada 2024d). Breeders at Agriculture and Agri-Food Canada produced the first ever low ERU *B. napus* cultivar, Oro, in 1968, utilizing pedigree selection of progeny from Liho (low ERU *B. napus* accession) and Nugget (high ERU *B. napus* cultivar released in 1961) (McVetty et al. 2009; Stefansson et al. 1961). Similarly, using pedigree selection methods, "Tower"

(first canola-quality *B. napus* cultivar) was registered in 1974 (McVetty et al. 2009; Stefansson and Kondra 1975). The oil content of "Tower" was approximately 41 % and yielded equally to its check variety in cooperative tests with 784 kg/ac (Stefansson and Kondra 1975). The oil content of "Hyola 401", the highly popular successor to the first canola hybrid "Hyola 40," contained 43.7 % oil content, toting an average yield of 962 kg/ac in a variety trial (Minnesota Canola Council 1998).

Before canola-quality *B. napus* was developed, the ERU content of oil generally ranged from 40 to 50 % (McVetty et al. 2009; Wang et al. 2022). Concurrent to the canola breeding program, a HEAR breeding program was established at the University of Manitoba whose objectives focused on increasing ERU content, oil content, and yield while reducing glucosinolates, days to maturity, and plant height (McVetty et al. 2009). Canada's first high ERU, low glucosinolate cultivar, Reston, contained 40-45 % ERU and was widely successful among Canadian farmers as it competed strongly with the foremost canola cultivar of the time, exhibiting 3 % higher oil content and 1 % higher yield (Stefansson and Downey 1995). Following Reston, several HEAR cultivars were developed using conventional breeding methods, including the "Millenium" series of cultivars which focused on blackleg resistance (McVetty et al. 2009). Throughout the many cultivars from Reston to Mill03, significant improvements were achieved including yield increase from approximately 930 kg/ac to nearly 1214 kg/ac, oil content increase from approximately 46 % to over 48 %, and ERU content increase from 40-50 % to 50-55 % (McVetty et al. 2009).

HEAR breeders have investigated the development of super-high-erucic acid, low-glucosinolate rapeseed (SHEAR) which surpasses the theoretical maximum ERU concentration of 66 % (McVetty et al. 2009). Pedigree selection methods have been applied to resynthesized *B. napus* genotypes to develop SHEAR, resulting in a maximum erucic acid content as high as 64 % (McVetty et al. 2009). The accomplishments in conventional HEAR breeding have been significant, however transgenic molecular approaches have achieved greater ERU levels as high as 72 % (Wang et al. 2022; Nath et al. 2008).

Phenotypic improvement through plant breeding has often been summarized through the breeder's equation (Equation 1), wherein ΔG represents genetic improvement regarding a trait of interest (Fehr 1987). The genetic improvement (ΔG) is a function of the phenotypic variability of the population (σ_P), the heritability of the trait (h^2), the selection intensity/accuracy (i), and the time required for one selection cycle (L) (Moose and Mumm 2008). Essentially, by increasing any of

the phenotypic variability, trait heritability, or selection intensity/accuracy, or reducing the time per breeding cycle, a breeder can expect greater genetic improvements within their germplasm. Functioning within these principles, conventional breeding has been effective for thousands of years, however certain limitations have been highlighted in light of future agricultural demands. Conventional plant breeding relies on "phenotypic selection" – choosing parents based on observed phenotypic variation within generations (Moose and Mumm 2008; Ahmar et al. 2020). Phenotypic selection is a major limiter regarding selection accuracy, and thus genetic improvement. For many important agronomic traits, accurately measuring phenotypes can be difficult, time consuming, and expensive. Significant inaccuracies and inefficiencies may also be introduced through environmental effect, genotype by environment interaction, data collection dependent on certain environmental conditions and stages of development, and measurements requiring the destruction of breeding material (Moose and Mumm 2008). Overall, phenotypic selection is effective in selecting major-effect alleles, however, polymorphism in minor-effect alleles is relatively poorly selected for (Prohens 2011). Underlying minor-effect alleles are not actively selected, and many are likely lost during early-stage selections (Raboanatahiry et al. 2021). All of these factors contribute to overall reduction in selection accuracy, limiting potential genetic gain, and increasing the number of breeding cycles required to achieve breeding goals.

These limitations have emphasized the inadequacy of conventional breeding methods in their ability to meet the future demands of agriculture. Mounting evidence shows that modern breeding methods are not only an option to approach the future demand, but a necessity (Collard and Mackill 2008; Voss-Fels et al. 2019).

$$\Delta G = \frac{h^2 \times \sigma_P \times i}{\mathrm{L}} \qquad [1] \text{ (adapted from Moose and Mumm 2008)}$$

Where: ΔG = genetic gain in a breeder's population

h^2 = trait heritability

σ_P = phenotypic variability in the population

i = selection intensity (proportion of population selected for production of next generation)

L = length of cycle interval (usually 1 generation)

2.6 Marker-assisted selection

Several forms of molecular markers have been developed for use in a wide array of plant breeding applications. In general, markers are used as "flags", performing on the principle of genetic linkage to flag the presence or absence of a closely linked gene of interest (Collard et al. 2005). During the rise of plant biotechnology in the 1980s, molecular marker systems were developed to facilitate the introgression of desired traits and reduction of linkage drag (Moose and Mumm 2008). Important molecular markers utilized in modern plant breeding include restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), retrotransposons, and single nucleotide polymorphisms (SNP) (Nadeem et al. 2017).

Details regarding development and application of molecular markers and marker-assisted selection (MAS) are available in reviews (Collard et al. 2005; Ribaut and Hoisington 1998; Francia et al. 2005; Xu and Crouch 2008). In essence, molecular markers provide researchers with tools to more efficiently and accurately analyze and manipulate the genetic variation among individuals (Collard et al. 2005). These variations occur as genetic differences (sequence polymorphisms) in genetic regions linked to genes of interest – not polymorphism in genes of interest themselves (Collard et al. 2005).

The use of MAS in plant breeding requires heavy preliminary research to ensure the reliability of markers. Marker development begins with the creation of a population which possesses a significant degree of variation (allelic segregation) in one or several traits of interest (Collard et al. 2005; Young 1994). From the members of this segregating population, DNA is extracted, and mapping software analyzes markers for polymorphism, producing a linkage map of the markers based on their segregation relative to one another (Collard et al. 2005). By extracting both phenotypic and genotypic data from the original population, polymorphisms in molecular markers can be associated with variation in traits of interest, allowing identification of Quantitative Trait Loci (QTL) (Nadeem et al. 2017; Bernardo et al. 2015). Once QTL are identified in the original mapping population, their locations must be validated to assure reliability in further applications. Marker validation is achieved through additional QTL mapping procedures upon germplasm genetically unique from the original mapping population (Collard and Mackill 2008; Melchinger et al. 1998; Langridge et al. 2001). Markers which are validated as strongly linked to QTL and

capable of reliably predicting phenotypes are then deemed viable for use as markers in MAS (Collard and Mackill 2008).

The broad applications of molecular markers stimulated their widespread adoption and rapid development. A shortlist of the significant uses for molecular markers in plant breeding includes phylogenetics, QTL mapping, assessing genetic diversity, investigating heterosis, and MAS (Nadeem et al. 2017). MAS has been adopted as standard practice for cultivar development in many crops (Moose and Mumm 2008).

The SNP is the most commonly utilized molecular marker in plant genetics, largely due to their prevalence on chromosomes of all plant species, presence in coding (genic) and non-coding (intergenic) regions, high stability, and reproducibility (Nadeem et al. 2018; Xu 2010; Syvänen 2001). As high-throughput, long-read sequencing technologies have become more cost-effective, researchers are able to easily produce high-density linkage maps based on SNPs (Bernardo et al. 2015; Ahmar et al. 2020). Long-read sequencing is also complimentary to SNP marker use, thus SNPs are currently regarded as the ideal molecular marker for plant genetics studies (Song et al. 2013; Bernardo et al. 2015; Ahmar et al. 2020; Tian et al. 2022).

2.6.1 Advantages and limitations of marker-assisted selection

The incorporation of MAS in plant breeding strategies may offer significant advantages over phenotypic (conventional) selection. These advantages to genetic gain are generally through improved breeding efficiency. For instance, differentiation between heterozygous and homozygous individuals facilitated MAS backcrossing methods for rapid trait introgression and reduction of linkage drag by selecting individuals which contain introgression targets (e.g. transgenes) and greater proportions of favorable alleles from the recurrent parent (Collard and Mackill 2008; Moose and Mumm 2008). MAS is highly advantageous for traits which are expensive or laborious to phenotype or traits which occur in later developmental stages since selection can be imposed at the seedling stage (Collard and Mackill 2008). Individual plants may be screened based on their genotype, as compared to conventional breeding where single-plant selection can be unreliable, requiring the growth of entire families for trait screening (Collard and Mackill 2008). Relative to phenotypic selection, MAS provides improved selection accuracy and greater selection pressure, improving genetic gains by more efficiently selecting against

undesirable germplasm, lessening phenotyping costs, and appropriately allocating plant growth resources.

Utilizing MAS has proven effective for qualitative traits controlled by one or very few major genes, however the major limitation of MAS is its ineffectiveness in selecting for complex traits controlled by several minor genes (Bhat et al. 2016). Traditional phenotypic selection may even provide superior genetic gains for complex quantitative traits relative to MAS (Bhat et al. 2016; Zhao et al. 2014). In broad terms, QTL used for MAS are only as effective as the genes to which they are linked to. Genetic gain through MAS is restricted to the amount of genetic variance that QTL can explain, which can often be limited or unreliable (Desta and Ortiz 2014). This relates to another limitation: QTL validity is not certain between distant germplasm and one's own germplasm. Breeders must often perform validation procedures to re-estimate the effect of QTL in their own germplasm, ensuring that traits truly exhibit association to loci that facilitate MAS (Heffner et al. 2009). The above limitations are inherent to MAS' design as an indirect selection method.

2.6.2 Marker-assisted selection in *Brassica napus* L.

The discovery and use of genetic markers for *Brassica* began with the development of RFLP linkage maps in *B. napus*, *B. oleracea*, and *B. rapa* (Landry et al. 1991; Slocum et al. 1990; Song et al. 1991). Since then, numerous genetic mapping studies have investigated important traits in *Brassicas*, identifying various forms of markers with the potential for use in MAS. Despite the rapid discovery of MAS-capable markers in *Brassicas*, there is a disparity in the incorporation of such markers into actual MAS breeding strategies.

Some examples of MAS in *Brassica* breeding include molecular markers identified by Rahman et al. (2008) for the two major erucic acid-controlling genes. These markers were implemented to the University of Manitoba HEAR breeding program to improve the efficiency of backcrossing schemes which included canola-quality rapeseed (McVetty et al. 2009). Breeding for other seed quality traits, such as oleic acid, linoleic acid, overall oil content, and yellow seed coat, have also incorporated MAS (Cheung et al. 1998; Jourdren et al. 1996a; Jourdren et al. 1996b; Rakow et al. 1999; Somers et al. 1999; Spasibionek et al. 2010; Liu et al. 2005). Hybrid breeding has implemented MAS to select for sterility and restorer genes in cytoplasmic male sterility systems

(Hansen et al. 1997; Liu et al. 2005, Huang et al. 2012). Markers for determinant inflorescence were used to select genotypes with superior lodging resistance (Li and Du 2023). Resistance traits for diseases such as clubroot, blackleg, and white rust have also been selected for using MAS (Kawasaki et al. 2021; Cheung et al. 1998; Somers et al. 1999; Hirani et al. 2016; Rahman et al. 2014).

The application of MAS to generally less complex traits in *Brassicas* is demonstrative of its limited ability to select for traits besides those controlled by one or very few genes.

2.7 High-throughput genotyping

Several SNP arrays are produced by Illumina Infinium™ (CA, USA) for a wide range of plant species including wheat, maize, and *Brassicas,* as well as animals including bovines, canines, humans, and more (Illumina n.d.). Of current high-throughput genotyping methods, the *Brassica* 60K and 90K Illumina Infinium SNP arrays are among the most extensively utilized in *Brassica* breeding and genetics research (Scheben et al. 2018). On any specific array, for example the *Brassica* 60K, there are 52,157 SNP-specific primers bound to the microarray flowcell (Clarke et al. 2016). DNA samples are amplified and fragmented, fragments hybridize to the SNP-specific primers on the microarray, followed by a single-base extension via labelled bases whose nucleotide-specific fluorescence represents SNP reads (Mason et al. 2017).

Genotyping with the *Brassica* 60K Illumina Infinium™ SNP array provides several important benefits. With a relatively simple DNA preparation protocol that is cost-effective for large populations, researchers receive easily interpretable and highly informational data with genome-wide distribution which includes physical position, indel variation, presence-absence variation, and more (Mason et al. 2017). Furthermore, SNP array data is highly reproducible, wherein fixed subsets of SNPs allow for direct comparison amongst populations (Mason et al. 2017). Relative to GBS, SNP arrays can be considered more reproducible and efficient for procedures such as comparative QTL mapping, however “reproducibility” (via an array’s subset of specific SNP markers) is a two-sided coin, such that the analysis of genetic diversity is restricted only to the SNPs present in the array (Mason et al. 2017). The choice of genotyping platform relies on a researcher’s goals, however, the *Brassica* 60K array remains a highly flexible tool for application to a wide range of applications in plant breeding and genetics research.

The *Brassica* 60K array has been used for mapping traits such as seed weight and quality (Li et al. 2014), seed number per silique (Xin et al. 2023), seed glucosinolate content (Qu et al. 2015), fatty acid content (Liu and Li 2014; Guan et al. 2019), flowering time (Xu et al. 2016), *Sclerotinia* resistance (Wei et al. 2016), root system architecture (Xu et al. 2022; Fletcher et al. 2014), and drought tolerance (Zhang et al. 2015). Additionally, it has been applied to dissecting meiotic recombination patterns (Yan et al. 2023) and homoeologous exchanges (Higgins et al. 2018; Stein et al. 2017) as well as input to genomic selection (Werner et al. 2017; Werner et al. 2020; Jan et al. 2016).

2.8 Genome-wide association study

The term "genome-wide association study" (GWAS) describes a wide range of methodologies in genetics research which use computer modelling to identify genetic regions that are associated with phenotypic variation in traits of interest (Cortes et al. 2021). Most GWAS assess genetic variation in the form of SNPs (Uffelmann et al. 2021). The main objective of GWAS is the identification of genome-wide QTL, and therefore is considered an important technique for understanding complex traits and implementation to plant improvement strategies (Uffelmann et al. 2021). The function of GWAS relies on genetic linkage to identify markers that are significantly associated with candidate genes, rather than directly identifying genic regions themselves. The first application of GWAS was in 2002 to identify genetic associations to human heart attacks (Ozaki et al. 2002). The first non-human application of GWAS was in *Arabidopsis* to investigate flowering time and pathogen resistance (Aranzana et al. 2005).

Results from GWAS are often implemented in subsequent MAS strategies; however, they are also used in validation of loci identified in other studies, linkage mapping, investigating candidate gene function, genomic selection, implementation to transgenics, and other uses (Deng et al. 2017; Liu et al. 2019; Bian and Holland 2017; Fiedler et al. 2017; Zhu et al. 2011; Tsai et al. 2020; Wang et al. 2016).

The general workflow for GWAS begins with assembling a variable population. Researchers can utilize several different population types for GWAS, including preexisting breeding populations, bi-parental mapping populations, nested association mapping (NAM) populations, multi-parental advanced generation intercross (MAGIC) populations, and others (Gupta et al. 2014; Cavanagh et

al. 2008). The composition of any assembled population must be considered as it inherently introduces unequal relationships among individuals, in turn introducing false-positive associations between predominant phenotypes and highly frequent SNPs within subpopulations (Cortes et al. 2021). The earliest GWAS models are considered "naïve" as they did not consider relatedness among individuals. Several models and statistical methods have since been developed to account for false positives and false negatives (Cortes et al. 2021).

DNA extraction and genotyping are performed upon individuals in the assembled population, followed by genotypic data quality control. Data quality filters are applied with thresholds upon sample call rate (proportion of successful genotype calls within individuals' DNA samples) and marker call rate (proportion of successful genotype calls for each marker across all individuals) (Uffelmann et al. 2021; Yao et al. 2020; Qu et al. 2017; Zou et al. 2016). Additionally, most GWAS studies filter markers whose minor allele frequency (MAF) is below 5 % due to supposed low statistical power of such variants (Uffelmann et al. 2021; Cortes et al. 2021; Turner et al. 2011).

2.8.1 Advantages and limitations of genome-wide association study

The advantages of GWAS are relative to earlier linkage mapping methods which are considered the precursors to GWAS (Cortes et al. 2021). Linkage mapping typically analyzes cosegregation of QTL using biparental cross progeny. Such progeny limits the diversity of segregation, exhibiting large linkage blocks and therefore lower resolution regions for subsequent candidate gene investigation (Cortes et al. 2021). Using large and highly diverse populations, GWAS detects cosegregation in small linkage blocks, thereby providing high resolution and the power to detect small-effect genes for complex traits (Lander and Schork 1994; Risch and Merikangas 1996). Improvements in high-throughput genotyping have facilitated cost-effective, high-density, pan-genome markers which expand the potential of GWAS to access wider sources of heritability in complex traits (Wainschtein et al. 2019; Young 2019).

A major limitation of GWAS is the inherent risk of false positives. The shortcomings of naïve models led to their replacement by methods which effectively control false positives. Mixed linear models have become commonplace for GWAS as they account for relatedness within the population by considering both a fixed effect of population structure (Q) and random effect by kinship (K) (Yu et al. 2005). Multi-locus models, such as FarmCPU and MLMM, improve on other

MLM models by providing high statistical power and computation speed while maintaining control over population relatedness factors (Cortes et al. 2021; Liu et al. 2016; Segura et al. 2012).

Multiple linear regression introduces another factor towards the risk of false positives in GWAS, stemming from the fact that each analyzed SNP is an individual test (Brzyski et al. 2017). With the conventional significance threshold set at 0.05, and considering the number of markers assessed in GWAS (ranging from hundreds to millions) the likelihood of encountering false positives increases due to the sheer volume of individual tests (Brzyski et al. 2017). Two correction methods for multiple linear regression have been widely adopted to further control the occurrence of false positives: corrected false discovery rate (FDR) or using a Bonferroni correction (Cortes et al. 2021).

The practice of omitting "rare variants" with low MAF (<5 %) may also be considered a limitation in GWAS analyses. Although GWAS studies have discovered many QTL for complex traits using common variants, it is considered that rare variants could significantly contribute to proportions of heritability which are not yet explained (Manolio et al. 2009). Improvements in sequencing methods have facilitated more comprehensive sequencing of rare variants, and statistical designs have emerged to more effectively analyze the effect of these variants (Lasky-Su 2017; Lee et al. 2014). These designs allow analysis upon rare variants often by aggregating and evaluating their combined significance (Lee et al. 2014; Cortes et al. 2021).

2.8.2 Use of GWAS in *Brassica napus* L.

As with many crop species, GWAS has become a common tool in genetic research and improvement of *B. napus.* A wide range of QTL and candidate genes have been identified for important traits affecting yield, seed quality, physiology, disease resistance, abiotic stress resistance, and more.

The genetic control of *B. napus* yield has been investigated using GWAS (Pal et al. 2021; Zhang et al. 2023; Liu et al. 2022; Xiang et al. 2023; Schiessl et al. 2015). Other yield-related traits have been investigated using GWAS including flowering time (Helal et al. 2021; Zhang et al. 2023; Huang et al. 2021; Raman et al. 2019; Korber et al. 2016), pod and seed characteristics such as seeds per pod, pods per plant, pod length, and thousand kernel weight (Pal et al. 2021; Zhang et al. 2023; Yang et al. 2023; Xiang et al. 2023), plant height and architecture (Dong et al. 2022; Zhang

et al. 2023; Liu et al. 2021; Xiang et al. 2023), harvest index (Qin et al. 2022; Lu et al. 2016), and lodging coefficient (Li et al. 2018).

Seed quality traits such as oil content (Zhao et al. 2022; Tang et al. 2021; Wang et al. 2021; Pal et al. 2021; Yao et al. 2021; Li et al. 2014; Xiao et al. 2019), erucic acid content (Xiang et al. 2023; Korber et al. 2016; Wang et al. 2018a; Li et al. 2014), oleic acid content (Xiang et al. 2023; Zhao et al. 2019), fatty acid composition (Gacek et al. 2017; Qu et al. 2017; Xue et al. 2018), glucosinolate content (Tan et al. 2021; Tang et al. 2023; Bhinder et al. 2022; Xiang et al. 2023), fibre content (Gacek et al. 2021), and seed coat colour (Wang et al. 2017a) have also been investigated using GWAS.

Researchers have also identified QTL and candidate genes associated with physiological traits such as seed germination (Luo et al. 2021; Hatzig et al. 2015), root architecture (Ibrahim et al. 2021; Ahmad et al. 2022 ; Duan et al. 2021; Wang et al. 2017b), calcium accumulation (Alcock et al. 2017; Chen et al. 2018), and magnesium accumulation (Alcock et al. 2017).

Traits for which QTL have been identified relating to biotic resistance include clubroot resistance (Dakouri et al. 2021; Li et al. 2016), blackleg resistance (Cantila et al. 2023; Raman et al. 2016), and sclerotinia stem rot (Roy et al. 2021; Wei et al. 2016; Wu et al. 2016). Abiotic stress resistance has been investigated with GWAS for freezing tolerance (Chao et al. 2021), salt tolerance (Wassan et al. 2021; Zhang et al. 2022), and aluminum tolerance (Zhou et al. 2022; Du et al. 2022).

2.9 Genomic selection

Genomic Selection (GS) is a technique implemented to enhance genetic gains in breeding programs. Originally designed for livestock, GS has been widely studied for its effectiveness in plant breeding (Budhlakot et al. 2022; Meuwissen et al. 2001). A multitude of GS methods have been developed and applied to a wide range of crops, traits, environments, population types, and more (Budhlakot et al. 2022). In general, GS employs trained regression models to produce genomic estimated breeding values (GEBVs) – phenotypic predictions which inform a breeder's selections and lessen reliance on repeated phenotyping over multiple generations (Combs and Bernardo 2013; Habier et al. 2007). Approaches used by GS methods are considered advantageous relative to the very limited genetic variation captured by MAS, which utilizes only major-effect QTL requiring prior identification, mapping, and diverse germplasm validation (Collard and

Mackill 2008; Kaler et al. 2022; Krishnappa et al. 2021). In combination with next generation sequencing and high-throughput phenotyping methods, GS has become highly attractive to breeders by capturing extensive genotypic and phenotypic variation, reducing costs through optimized resource allocation, and shortening the overall breeding cycle (Heffner et al. 2010; Jubair and Domaratzki 2023; Krishnappa et al. 2021; Varshney et al. 2017).

To evaluate the accuracy of GS, researchers may use "cross validation" procedures, which imitate the workflow of GS in practical applications (Figure 2.3). The practical GS workflow includes two populations: a training population (TP) and a test population (represented by a validation population (VP) in cross validation procedures). The TP (e.g. parental genotypes, preliminary hybrids) is sown, phenotypic and genotypic data are collected, and these data are used to "train" a GS model. As opposed to MAS, which informs selections based on a small number of major-effect markers, the training of GS models estimates the additive effects of all markers in a genome-wide marker set (Bernardo 2008; Budhlakot et al. 2022). The trained GS model is then capable of producing GEBVs. The test population (e.g. hybrids derived from TP parental genotypes, siblings of preliminary hybrids) is sown and DNA samples are collected. The test population's genotypic data are input to the trained GS model, which produces GEBVs through whole-genome regression (WGR) (summing effects of all genome-wide markers) (Budhlakot et al. 2022). Selection can be conducted including each genotype's GEBV as a criterion (Xu et al. 2012). To evaluate GS prediction accuracy using cross validation, prediction accuracy is calculated as Pearson's correlation between the VP's observed phenotypes and model-generated GEBVs (Werner et al. 2020). Using reliable GS models, breeders may minimize the amount of germplasm required for phenotyping and continue selection in off-season nurseries or other non-target environments, effectively reducing phenotyping costs and the time required to achieve breeding goals (Lorenz and Nice 2017; Krishnappa et al. 2021).

Several strategies are available for plant breeders looking to incorporate GS within their program. These include parametric and non-parametric models, machine learning and deep learning models, uni-trait and multi-trait approaches, multi-trait multi-environment approaches, GWAS-guided GS, and more (Budhlakoti et al. 2022; Emmert-Streib et al. 2020; Jubair and Domratzki 2023). To

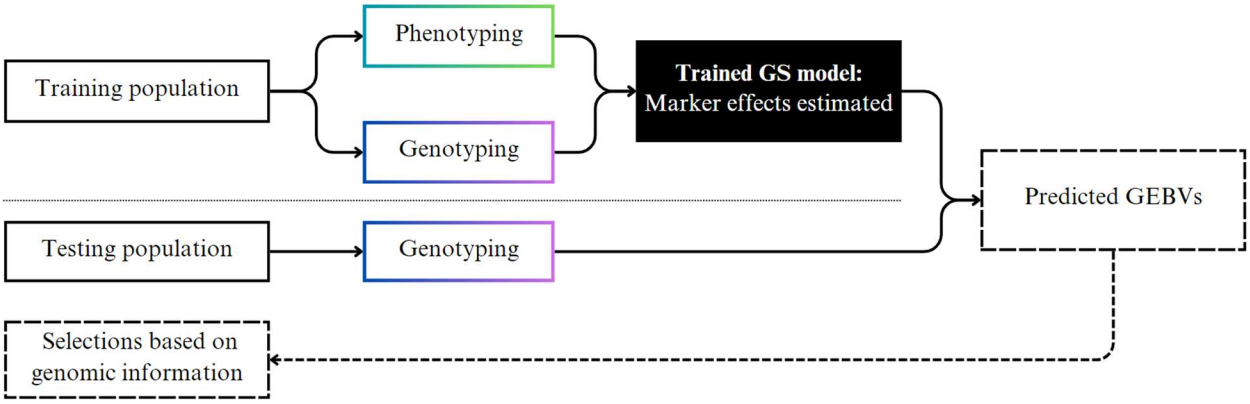


Figure 2.3 General workflow of genomic selection for crop improvement. Adapted from Bhat et al. (2016). Abbreviations: GEBV: Genomic estimated breeding value; GS: Genomic selection.

minimize TP phenotyping costs and maximize prediction accuracy for the VP, breeders can also consider a wide array of population designs depending on the objective(s) of their program. Population designs are available for recurrent crosses and selection, selection within a biparental family, selection within diverse accession germplasm, and other breeding populations (Lorenz and Nice 2017; Windhausen et al. 2012). Depending on breeding program structure, GS can be used to effectively address components in the breeder's equation. Selection intensity, selection accuracy, and especially breeding cycle time can be directly addressed to enhance genetic gain per unit of time (Hickey et al. 2017; Werner et al. 2020).

2.9.1 Factors affecting accuracy of genomic selection

Currently, there are no GS methods that provide high prediction accuracies in all circumstances, therefore, identifying an "optimal method" is a highly subjective process which requires comparisons using the data at hand (Werner et al. 2020; Zhang et al. 2023). Many factors affect the accuracy and overall effectiveness of GS. These factors are inherent in all breeding programs and vary widely among species, and even among traits within the same species (Figure 2.4). Given its subjectivity, the general notion regarding GS methodology is that there is no "one size fits all" (Lorenz et al. 2011).

The effectiveness of GS does not only depend on the availability of software, but also an understanding of GS' role within the context of multi-stage selection breeding programs (Figure 2.5) (Werner et al. 2020). The goal of increased genetic gain per unit time without increased costs is much easier said than done (Crossa et al. 2017). Effective implementation of GS requires the breeder to quantify the potential benefits and limitations associated with GS when compared to alternative methods for long-term breeding programs (Heslot et al. 2015).

2.9.1.1 Training population and validation population design

The accuracy of marker effect estimates by GS models is a function of the allelic variation provided within their TP. The TP should include wide phenotypic and genotypic variation to extensively capture haplotype combinations and their effects on traits of interest (Calus 2010; Werner et al. 2020). The "quality" of TP design is subjective to the VP (or test population in practical applications) upon which the trained models predict GEBVs. Specifically, GEBV accuracy is subjective to the genetic relationship between the TP and the VP (Calus 2010; Lorenz and Nice

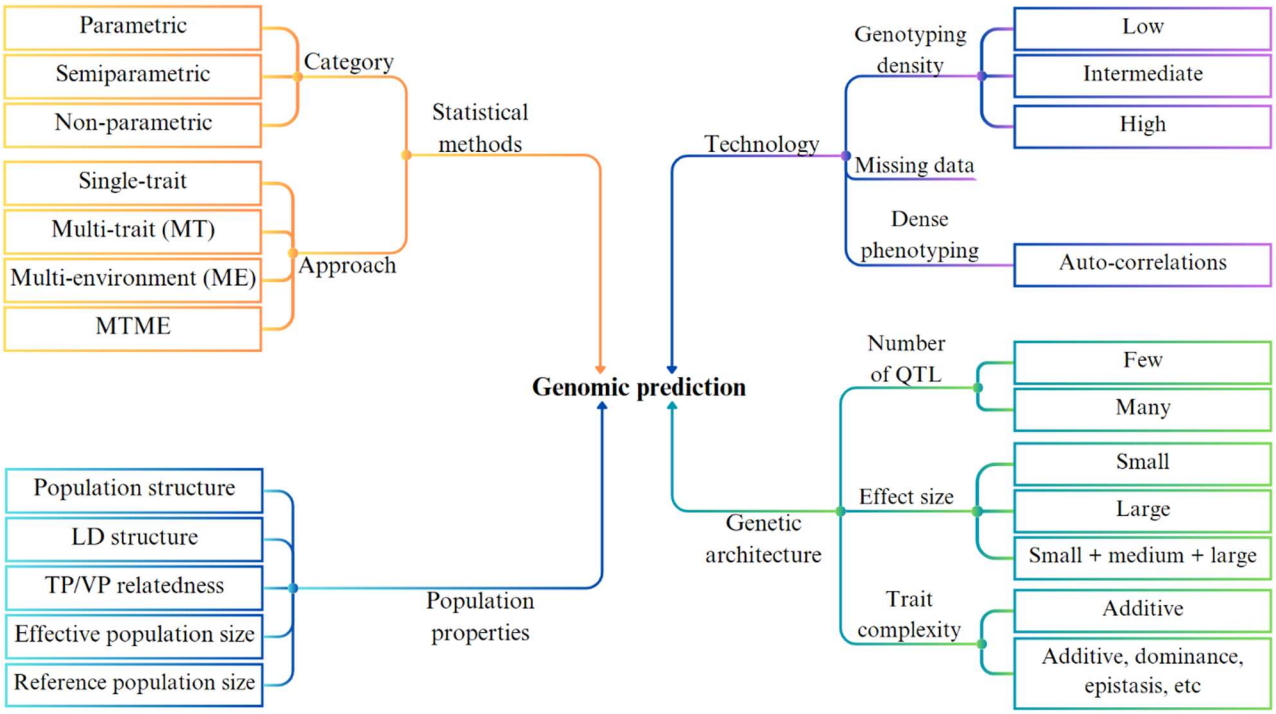


Figure 2.4 Factors affecting genomic prediction accuracy. Adapted from Farooq et al. (2023). Abbreviations: LD: Linkage disequilibrium; QTL: Quantitative trait loci; TP: Training population; VP: Validation population

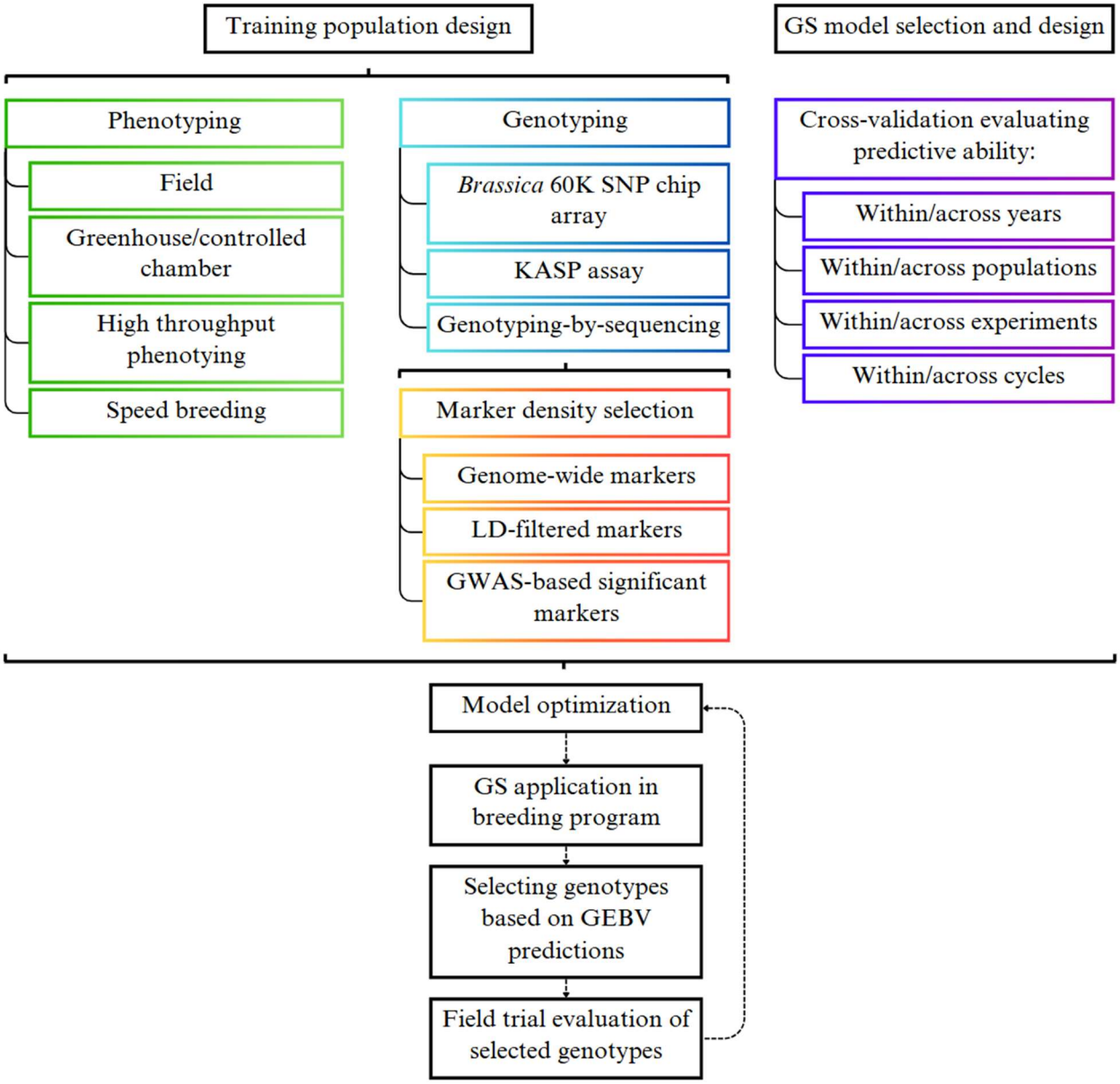


Figure 2.5 Overview of genomic selection (GS) in application to plant breeding. Adapted from Yadav et al. (2023). Abbreviations: GEBV: Genomic estimated breeding value; GWAS: Genome-wide association study; KASP: Kompetitive allele-specific PCR; LD: Linkage disequilibrium; SNP: Single nucleotide polymorphism.

2017). Within a single population, linkage disequilibrium (LD) patterns between markers and QTLs exhibit homogeneity which is known to be disrupted when considering stratified populations (de Roos et al. 2009; Rio et al. 2022). These disruptions in LD patterns may lead to differences in estimated markers effects, impacting accuracy when one population is used to predict another (Hayes et al. 2009; de Roos et al. 2009; Rio et al. 2022). Designs which include high genetic relatedness between TP and VP enhance GS accuracy by ensuring linkage phase between markers and QTL is similar for model training and model application (Werner et al. 2020). By including similar population structure in TP and VP, population structure has been demonstrated as a significant contributor to GS prediction accuracy (Clark et al. 2011; Makowsky et al. 2011; Bastiaansen et al. 2012; Werner et al. 2020). In addition to genetic relatedness, TP size is a significant factor when evaluating genomic heritability and genomic prediction in real and simulated breeding populations (Guo et al. 2014; Neyhart et al. 2017). Increasing TP size more extensively captures genotypic and phenotypic variation for GS model training, increasing prediction accuracy (Daetwyler et al. 2008; Desta and Ortiz 2014; Edwards et al. 2019; Hickey et al. 2014; Zhang et al. 2017). Although it would be ideal for plant breeders to capture as much variation as possible within the TP (e.g. large collections of highly exotic or resynthesized germplasm), it is not practical (Calus 2010). Plant breeders are prompted to achieve balance between minimizing TP phenotyping/genotyping costs, while maximizing genetic relatedness to targeted VPs to which GS models will be applied (Lorenz and Nice 2017). An important consideration regarding cross validation procedures is that accuracies measured within a specific TP/VP design are not necessarily representative of GS effectiveness in any application (Werner et al. 2020).

2.9.1.2 Marker densities

Similar to TP design, the choice of optimal marker density for GS requires a consideration of balance. Providing a GS model with low marker density can be considered underfitting (Montesinos-López et al. 2022b). An underfit model is one which is insufficiently adapted to understand marker effects due to low marker coverage, suffering a similar disadvantage as MAS wherein many important minor/major effect markers are not considered (Montesinos-López et al. 2022b). Alternatively, while high density marker sets provide a GS model with extensive genic and non-genic coverage, they may be vulnerable to overfitting, wherein marker effects are adapted

from non-genetic effects (e.g. environment) and noise (random marker variation) from the TP such that it decreases predictive power when applied to VPs which exhibit independent non-genetic effects (Hickey et al. 2014; Montesinos-López 2022b). The phenomenon known as the "large p small n problem" arises when the number of markers (p) greatly exceeds the number of samples (n), causing model overfitting (Schmidt et al. 2016; Crossa et al. 2017). For practicality, a breeder may not want to maximize marker density for the sake of genotyping cost and computational resources, while also considering that the predictive ability of GS models tends to reach a plateau when increasing marker density (Combs and Bernardo 2013; Onogi et al. 2015). It may be beneficial, however, to have such dense genotypic markers available for long-term breeding cycles as they may be utilized to buffer LD disruption caused by recombination over cycles (DoVale et al. 2022).

Optimal marker density is considered a function of LD decay distance (Desta and Ortiz 2014; Kaler et al. 2022). As an indirect selection method, GS accuracy relies on marker linkage to causal genes. As such, it is not simply "marker density" but instead the maximum distance/distribution of markers to causal genes. For example, in a potato breeding population whose LD block size is estimated 1 Mb, the presence of markers within a maximum 500 kb span on either side of causal genes can be considered a reliable marker density (Aalborg et al. 2024; Vos et al. 2017). In a study comparing GS accuracy among maize, rice, and soybean using LD-filtered marker sets, maize exhibited the shortest LD decay, thus requiring the highest-density marker set, followed by rice and soybean (Kaler et al. 2022). The plateau effect on GS prediction accuracy was demonstrated by Kaler et al. (2022) whose LD-filtered marker set contained less than half the markers of its complete marker set, yet produced statistically indistinguishable prediction accuracies for 6 traits of varying heritability in soybean. In summary, the minimum required marker density for GS training to appropriately capture causal polymorphisms is dependent on LD decay, wherein shorter LD decay distance requires greater minimum marker density and vice versa.

The significance of LD intersects the effects of marker density and population relatedness, as closely related genotypes share common (larger) LD blocks, whereas more distant genotypes exhibit uncommon (smaller) LD blocks (Hickey et al. 2014). To compensate for uncommon LD blocks in distantly related TP and VPs, prediction accuracy may be improved by increasing marker set density (Norman et al. 2018).

2.9.1.3 Trait complexity

Complex traits are often controlled by a combination of additive and non-additive (e.g. dominant/recessive, epistatic etc.) allele effects, which are characterized by many small-effect QTL and environmental effects (Korte and Farlow 2013). Furthermore, the ratio of additive to non-additive genetic variance may differ even for highly correlated traits (Farooq et al. 2023). Although GS is considered clearly advantageous to MAS for the improvement of complex traits, the heritability of specific traits remains a significant factor affecting GS prediction accuracy. The many minor additive and non-additive effects within complex traits makes genomic prediction difficult; i.e., low heritability traits often lead to reduced prediction accuracy (Farooq et al. 2023; Heffner et al. 2009; Jannink et al. 2010; Kaler et al. 2022).

The effect of trait on prediction accuracy also intersects with the effects of several factors including TP size, TP relatedness to VP, and the quality/quantity of phenotypic data available for each genotype (Habier et al. 2007; Daetwyler et al. 2008; Goddard 2009; Hickey et al. 2014; Heslot et al. 2015; Werner et al. 2020). The trait of interest also affects the minimum marker density required to reach a prediction accuracy plateau. Compared to high-heritability traits, which exhibited maximum prediction accuracy plateaus with marker sets containing ~1,000 markers, Aalborg et al. (2024) observed improved accuracy for low-heritability traits (yield, senescence) when using several increasingly large marker sets over 10,000 markers. Norman et al. (2018) observed a similar response in GS accuracy for low-heritability traits when using increasingly high-density marker sets.

2.9.1.4 Genomic selection models

The effectiveness of GS relies on accurate estimations of GEBVs using computer models which perform whole-genome regression (WGR). Many of these models have been developed; however, each contains different assumptions providing their own respective advantages and disadvantages (Table 2.1). The notion regarding GS is emphasized: no singular model can be considered “best.” (Lorenz et al. 2011).

The earliest GS model proposed by Meuwissen et al. (2001) was the linear least squares regression. As opposed to common GS models today, linear least squares regression is not classified as a WGR model, as it first performs regression on individual chromosome segments separately, subsequently

only utilizing the chromosomal segments containing significant marker effects. (Howard et al. 2022). Since their proposal, Meuwissen et al. (2001) and many other researchers have aimed to enhance the accuracy, computational efficiency, and applicability of GS models for their implementation in breeding programs.

These GS approaches can be considered in different classifications, such as parametric and non-parametric models, machine learning and deep learning models, single-trait and multi-trait models, multi-trait-multi-environment models, GWAS-guided GS, and more. (Budhlakoti et al. 2022; Emmert-Streib et al. 2020; Jubair and Domratzki 2023). Each of these GS approaches contain unique parameters to consider population structure, epistasis, missing data, environmental effects, and optimizing computational efficiency with large datasets. The accuracy of different models varies with the genetic architecture of traits due to their assumptions and treatment of marker effects (Larkin et al. 2019).

2.9.1.4.1 Parametric models

Best linear unbiased prediction (BLUP) is one of the most implemented methods for GS in plant science, using mixed model–based WGR approaches to estimate marker effects (Habier et al. 2013). Common variants of BLUP include ridge regression BLUP (rrBLUP) and genomic BLUP (GBLUP) (Endelman 2011; VanRaden 2008). Both rrBLUP and GBLUP are considered computationally unelaborate and perform relatively similarly, each assuming that marker effects are homogeneously distributed across the genome, contributing to genetic variance with at least a small but non-zero effect (Clark et al. 2011; Goddard et al. 2009; Tan et al. 2017; Whittaker et al. 2000; Würschum et al. 2014). The assumptions of rrBLUP and GBLUP exemplify their classification as parametric models. The rrBLUP model assumes that all markers have equal variances and small but non-zero effects, whereas GBLUP uses a genomic relationship matrix (G matrix) to predict phenotypes (Endelman 2011). Parametric model assumptions imply a certain relationship between the trait of interest (response variable) and the amount of genetic variance that can explain the trait, limiting their predictive ability to additive effects (Clark et al. 2011; Pérez-Rodríguez et al. 2012; Tan et al. 2017). Several parametric models are considered advantageous to basic linear models which often suffer from overfitting resulting from the “large p small n” problem (Endelman 2011; Neves et al. 2012). The assumptions of parametric models, however, cause difficulty in recognizing high-order effects such as epistasis (Howard et al. 2022).

Other commonly used parametric models include the least absolute shrinkage and selector operator (LASSO), reproducing kernels Hilbert spaces regression (RKHS), and an array of models classified as the "Bayesian alphabet" (Desta and Ortiz 2014; Habier et al. 2011).

In contrast to BLUP models, Bayesian methods are constructed to induce different variances among markers and to perform marker selection, shrinkage of the marker effects, or both (Habier et al. 2011; Howard et al. 2022). The prior density (i.e. the type of probability distribution assigned to parameters of priorly understood genotypic/phenotypic relationships) determines the magnitude and type of shrinkage to be induced, acting as a distinguishing factor amongst particular Bayesian models (Howard et al. 2022). Meuwissen et al. (2001) introduced two Bayesian methods (BayesA and BayesB) which enabled genomic regions containing larger-effect markers to be assigned greater importance. BayesA assumes that all markers contribute effects, but the effects are drawn from a scaled-t distribution controlling the degree of shrinkage towards zero effect (Howard et al. 2022; Meuwissen et al. 2001). BayesB builds upon BayesA by incorporating a mixed prior, where some markers have effects drawn from a scaled-t distribution with shrinkage, while others have zero effects with a prior probability drawn from a beta distribution (Howard et al. 2022; Meuwissen et al. 2001). Several Bayesian models containing variants of prior distributions have been developed, including BayesC, BayesCπ, Bayesian LASSO (BL), and Bayesian ridge regression (BRR) (Hoerl and Kennard 2000; de los Campos et al. 2009; Pérez et al. 2010).

2.9.1.4.2 Non-parametric models

Machine learning (ML) models are becoming increasingly popular for WGR. Methods include both semiparametric and non-parametric models (Reinoso-Peláez et al. 2022). These models differ from parametric models in that ML models do not require an explicit or implicit distribution of the data to be characterized (Reinoso-Peláez et al. 2022). The design of ML models intends to capture signals from high-order interactions such as epistasis (Pérez-Rodríguez et al. 2012). Proponents of ML suggest that many limitations in parametric WGR can, in principle, be resolved with ML (Reinoso-Peláez et al. 2022).

Gianola et al. (2006) first proposed a nonparametric treatment of SNPs with the reproducing kernel Hilbert spaces (RKHS) method, which combines a classical additive genetic model with a kernel function. Kernel methods utilize no assumptions of linearity, providing great flexibility in

simultaneously capturing nonadditive effects of all kinds (Howard et al. 2022). Other common non-parametric models include random forest (RF), support vector machine (SVM), and artificial neural networks (ANN) (De los Campos et al. 2010; Reinoso-Peláez et al. 2022)

Relative to parametric approaches, ML is said to be advantageous regarding three main limitations: the "large p small n" problem, high-order genetic interactions, and non-normal input data. Section 2.9.2 explores these concepts further.

Chapter 4 of this research investigates two decision tree-based methods: random forest (RF) (Breiman 2001) and extreme gradient boosting (XGB) (Chen and Guestrin 2016; Friedman 2001). Decision trees act as flowcharts based on binary decisions for single independent features (e.g. SNP markers). The name RF represents its utilization of an "ensemble" of decision trees that are created with a bootstrapping (AKA bagging) method, wherein random subsets of SNPs and individuals are analyzed (Montesinos- López et al. 2022a). By providing trees with different portions of training data, bootstrapping allows the predictions to be combined (ensembled) providing error compensation and therefore improved prediction generalization (Montesinos-López et al. 2022a). At each branch division (across x number of trees), RF randomly selects a subset of SNPs, and the SNP which minimizes loss of function is chosen to split the observations for the next branch (Reinoso-Peláez et al. 2022). By searching for the best independent variable within a random subset (instead of the most important independent variables overall), RF generates a wide heterogeneity that improves predictive power (Montesinos-López et al. 2022a).

Gradient boosting (GB) also utilizes ensembled decision trees to inform regression (Farooq et al. 2023). However, contrasting RF's ensemble of independent trees trained on bootstrapped subsets of training data, GB builds decision trees in a sequential (dependent) manner, boosting predictive power by training based on the results of previous trees (Farooq et al. 2023). Aiming to reduce the residual error at each stage of training, GB calculates the gradient of the predictive loss of function made by the current ensemble of trees, and the residuals are then adjusted for the next predictor (Reinoso-Peláez et al. 2022).

The research in Chapter 4 also investigates support vector machine (SVM) which is considered a kernel method, similar to RKHS (Wahba 2002). Since its proposition by Cortes and Vapnik (1995), SVM has performed especially well as a classification method but has also been implemented for

Table 2.1 Classification of genomic selection models[1]

Parametric regressions	Penalized approach	Ridge regression best linear unbiased prediction (rrBLUP)
		Least absolute shrinkage and selector operator (LASSO)
		Elastic net (EN)
		Support vector regression (SVR)
		Neural networks (NN)
		Reproducing kernels Hilbert spaces regression (RKHS)
		Genomic best linear unbiased prediction (GBLUP)
	Bayesian approach	Genomic best linear unbiased prediction (GBLUP)
		Bayesian ridge regressions (BRR)
		Bayesian LASSO (BL)
		BayesA
		BayesB
		BayesC
		Reproducing kernels hilbert space (RKHS)
Non-parametric regressions	Ensemble	Boosting (BOOST)
		Bagging (BAGG)
		Random Forest (RF)
	Kernel	Reproducing kernels hilbert space (RKHS)
		Support vector machine (SVM)
	Deep learning	Bayesian neural network (BNN)
		Artificial neural network (ANN)
		Convolutional neural network (CNN)

[1]Adapted from Desta and Ortiz 2014; Reinoso-Peláez et al. 2022

Table 2.2 Main features of genomic selection models[1]

Model	Features
rrBLUP	Assumes that all markers have equal variances with small but non-zero effect. Applies homogeneous shrinkage of predictors towards zero but allows for markers to have uneven effects. Computed from a realized-relation matrix based on markers. Some QTL are in LD to marker loci, whereas others are not.
LASSO	Combines both shrinkage and variable selection methods. rrBLUP does not use variable selection but outsmarts LASSO when there is multi-colinearity between the predictors.
BRR	Induces homogeneous shrinkage of all marker effects towards zero and yields a Gaussian distribution of marker effects. Similar to RR-BLUP, there is a problem of QTL linkages to the marker loci.
BL	Applies to both shrinkage and variable selection. Has an exponential prior on marker variances resulting in a double exponential (DE) distribution. The DE distribution has a higher mass density at zero and heavier prior tails compared with a Gaussian distribution.
BayesA	Utilizes an inverse chi-square ($\chi 2$) on marker variances yielding a scaled t-distribution for marker effects. Similar to BL and in contrast to BRR, it shrinks tiny marker effects towards zero and larger values survive. Has a higher peak of mass density zero compared with the DE distribution.
BayesB	Similar to BayesA, uses an inverse $\chi 2$ resulting in a scaled t-distribution. Unlike BayesA, utilizes both shrinkage and variable selection methods. When $\pi = 0$, then it is similar to BayesA.
BayesC	Applies both shrinkage and variable selection methods. Characterized by a Gaussian distribution. BayesB and BayesC consist of point of mass at zero in their slab priors.
RF	Uses the regression model rooted in bootstrapping sample observations. Takes the average of all tree nodes to find the best prediction model. Captures the interactions between markers.
BOOST	Builds decision trees sequentially, "boosting" predictive power based on previous trees.
RKHS	Based on genetic distance and a kernel function with a smoothing parameter to regulate the distribution of QTL effects. Effective for detecting nonadditive gene effects.
SVM	Uses non-linear kernel function to reclassify training data to high-dimensional features, facilitating separability by hyperplane. Combines linear regression with non-linear reclassification of training data Highly capable of generalization.

[1]Adapted from Desta and Ortiz 2014; Farooq et al. 2023; Merrick and Carter 2021; Noble 2006; Awad and Khanna 2015

regression (i.e. support vector regression (SVR)) (Reinoso-Peláez et al. 2022). SVR analyzes genotypic and phenotypic training data to determine a hyperplane – a line that best fits phenotypic variation by genotypic data, thereby creating a decision boundary based on many features (Noble 2006). Training data is reclassified to high-dimensional features using a non-linear kernel function, thereby allowing separability in that higher dimensional space, followed by linear regression within that feature space (Merrick and Carter 2021; Noble 2006). The main advantages of SVR are its capability for generalization, allowing high prediction accuracy, while computational complexity is not restricted to the dimensionality of the input data (Awad and Khanna 2015).

2.9.2 Advantages and limitations of genomic selection

The advantages of GS can be considered as improvements relative to its predecessor, MAS. Breeders have implemented major-effect QTL to MAS for a wide variety of traits; however, complex traits controlled by minor-effect QTL demonstrate the breakdown of MAS' effectiveness due to inconsistent QTL effect estimates (Bernardo 2008). Furthermore, QTL identification, mapping, and inter-population validation are costly preparatory operations inherent to MAS which precede any actual genetic gains (Collard and Mackill 2008; Kaler et al. 2022; Krishnappa et al. 2021). In contrast to MAS, GS follows a "black-box" approach, inputting genome-wide markers and phenotypic data in a single analysis capable of implementing both major- and minor-effect QTL in calculating predictions (Krishnappa et al. 2021). By providing GEBVs for genotypes from which only genotypic data is required, GS provides a breeder with informed selections while reducing the reliance on repeated phenotyping over multiple generations (Habier et al. 2007; Jannink et al. 2010). The use of GEBVs to inform the removal of certain germplasm (which would otherwise require generations of phenotypic data to inform) is a significant benefit to plant breeders, facilitating the allocation of resources to desirable germplasm more likely to contribute to the final cultivar. By optimizing the required volume of phenotypic data to inform selections, GS therefore improves genetic gain per unit time while reducing costs per breeding cycle and (potentially) the overall number of breeding cycles (Hickey et al. 2014). These advantages are especially important regarding complex agronomic traits that are difficult/costly to phenotype and for particular species that exhibit long generation times (Merrick and Carter 2021; Neyhart et al. 2017). For example, tree or perennial crop breeding (e.g. apple) can implement GS to inform

selections that bypass several years of vegetative growth required to phenotype at sexual maturity (Kumar et al. 2012).

Practical usage of GS has demonstrated its advantages over MAS and phenotypic selection methods in several crop species. In maize and wheat, GS prediction accuracy (correlation) of only 0.53 provided multifold annual genetic gains relative to MAS and pedigree selection methods (Heffner et al. 2010). In soft red winter wheat, yield increase was 10 % greater when combining GS with phenotypic selection relative to solely phenotypic selection (Lozada et al. 2019b). When predicting soybean seed weight, genomic selection accuracy was found to be 0.11 greater than MAS at their respective peak performances (Zhang et al. 2016). In some scenarios, GS does not significantly improve responses to genetic gain, however, the optimization of phenotyping procedures allows breeders to achieve similar gains with significantly reduced costs, such as observed by Beyene et al. (2019) whose hybrid maize breeding costs reduced by 32 % when incorporating GS. The reduction of phenotyping volume across breeding cycles not only provides economic benefits, but also environmental benefits when considering land, water, fossil fuel, and crop input resources across long-term multi-environment field testing (Krishnappa et al. 2021). Alongside GS, speed breeding (SB) techniques (implementing particular lighting and growth conditions to significantly reduce crop plant generation time relative to normal field/greenhouse conditions) have been integrated to produce "Speed GS" (Ćeran et al. 2024; Krishnappa et al. 2021). Using Speed GS, genetic gains per unit time are compounded by both GS' reduction in number of breeding cycles and SB's reduction of generation time (Krishnappa et al. 2021). Crops with standardized SB procedures, such as wheat and tall fescue, have been investigated using Speed GS, exhibiting significant improvements in breeding cycle efficiency and overall costs relative to phenotypic selection methods and MAS (Jighly et al. 2019; Watson et al. 2019).

Despite the success of parametric approaches exampled above, advocates for ML-based WGR suggest its advantages over parametric models regarding the "large p small n" problem, complex traits controlled by high-order genomic interactions, and non-normal phenotypic data. The "large p small n" problem (p: number of markers, n: sample size) is one which has arisen with improved genotyping technologies (González-Recio et al. 2014). Despite yielding extremely dense SNP data at low costs, the inclination to utilize all available genotypic data may bring on costs to GS effectiveness; overfitting causes decreases in prediction accuracy observed in many situations

wherein p:n exceeds 50–100 (González-Recio et al. 2014). Nonparametric models are purported to provide great flexibility overcoming the large p small n problem by utilizing different forms of regularization (Reinoso-Peláez et al. 2022). For complex traits, nonparametric models are also suggested to provide greater flexibility in recognizing both additive and nonadditive effects (e.g. dominance, epistasis) simultaneously (Desta and Ortiz 2014; Howard et al. 2022; Varona et al. 2018). Improvements are highlighted when considering the assumptions of parametric models. Some of these assumptions include the partitioning of genetic variance into additive, dominance, additive × additive, and additive × dominance, which only hold under random mating of male and female parents, no inbreeding, no assortative mating, and no (natural or artificial) selection – all violated in practical plant breeding operations (Howard et al. 2014). Lastly, although parametric models are known to provide acceptable goodness of fit when phenotypes exhibit normal distributions, phenotypic distributions often deviate from normality in practice (e.g. environmental effects), with which, the assumptions of parametric models may limit their effectiveness relative to other approaches (Luaces et al. 2010).

The effectiveness of GS is also limited by a breeder's foresight, considering that a target breeding population should inform the design of the training population which precedes it (Lorenz and Nice 2017). To avoid poor prediction accuracies and wasted resources, breeders must consider a balance between two goals: 1) minimizing costs for TP phenotyping/genotyping, and 2) maximizing prediction accuracy for the target breeding population (Lorenz and Nice 2017). For example, a training population may provide very high accuracies for a small target population containing closely related individuals, however the overall costs of TP phenotyping/genotyping would then merit only a small number of selection decisions (Hickey et al. 2014). As a breeder's germplasm develops over several cycles, novel LD patterns are also expected to develop as a result of recombination, selection, and drift, causing declines in GS prediction accuracy (Lorenz et al. 2011; Neyhart et al. 2017). To ensure the long-term effectiveness of GS, breeders must recognize the need to update TP which effectively captures novel LD generated over breeding cycles. Neyhart et al. (2017) evaluated several methods of updating TP, demonstrating that all methods of updating TP exhibit similar long-term patterns of maintaining prediction accuracy and genetic gain, meanwhile only a stagnant TP significantly damages long-term GS effectiveness. Other methods of maintaining long-term effectiveness with GS should consider maintaining genetic diversity throughout generations. To address the issue of some conventional GS methods which provide

short-term gain at the cost of the long-term potential (reduction in genetic diversity), several approaches have been developed, such as weighted genomic selection (WGS), optimal haploid value (OHV) selection, genotype building (GB), and optimal population value (OPV) (Goiffon et al. 2017; Moeinizade et al. 2019).

2.9.3 Use of genomic selection in plant breeding

The "no one size fits all" outlook regarding GS has stimulated wide research applications across crop species and traits to consider the many variables which affect GS accuracy. These variables have been investigated using both practical applications and simulation data from historical germplasm.

Wheat has been a major subject of GS research. Many of the contributors to yield, such as thousand kernel weight, weight per spike, heading date, days to maturity, and yield itself, have been extensively investigated using GS (Lozada et al. 2019b; Song et al. 2017; Verges and Van Sanford 2020; Belamkar et al. 2018; Sun et al. 2019; Rife et al. 2018; Sandhu et al. 2021a; Sandhu et al. 2021b; Crain et al. 2018; Dunckel et al. 2017; Ward et al. 2019; Dreisigacker et al. 2023; Kehel et al. 2020; Ibba et al. 2020; Guo et al. 2020; Herter et al. 2019; Montesinos-López et al. 2023). Other important agronomic traits predicted by GS include height and resistance to biotic and abiotic stresses such as fusarium head blight, septoria tritici blotch, snow mold, winter hardiness, and frost tolerance (Kehel et al. 2020; Borrenpohl et al. 2020; Jiang et al. 2017; Dong et al. 2018; Herter et al. 2019; Lozada et al. 2019a; Beil et al. 2019; Michel et al. 2019). Considering the long catalogue of important processing quality traits in wheat, GS has even been used to predict water uptake, protein content, gluten content, flour yield, and many more (Gill et al. 2021; Zhang-Biehn et al. 2021; Tsai et al. 2020; Ibba et al. 2020; Michel et al. 2017; Veenstra et al. 2020).

The effectiveness of GS has also been exhibited in soybean research. Yield and its related traits including seed weight, days to maturity, and pods per plant have shown varying degrees of prediction accuracy (Ravelombola et al 2021; Stewart-Brown et al. 2019; Duhnen et al. 2017; Yoosefzadeh-Najafabadi et al. 2022; Miller et al. 2023; Smallwood et al. 2019; Wen et al. 2015; Matei et al. 2018). Other seed quality traits such as protein content, fatty acid composition, and overall oil content have exemplified GS' effectiveness in application to important oilseed traits (Stewart-Brown et al. 2019; Duhnen et al. 2017; Miller et al. 2023; Smallwood et al. 2019; Wen et

al. 2015). Responses to biotic and abiotic stresses, such as sudden death syndrome, cyst nematode, and lodging, have also been investigated in soybean research using GS (Ravelombola et al. 2020; Bao et al. 2015; Wen et al. 2015). Relative to phenotypic selection, Matei et al. (2018) demonstrated a 123 % increase in selection efficiency for yield and 212 % increase in selection efficiency for thousand kernel weight while reducing time from 6 years to 3 years.

Würschum et al. (2014) began exploring the potential of GS in *B. napus* breeding, observing high accuracies by rrBLUP when predicting plant height, but low accuracies when predicting yield and glucosinolate content. Since then, Würschum and many other researchers have investigated genomic prediction for traits related to seed quality, yield, morphology, disease resistance, and responses to environmental stresses, widely utilizing the *Brassica* 60K SNP array and other next-generation sequencing methods (Fikere et al. 2018; Koscielny et al. 2020; Hu et al. 2021a Hu et al. 2021b; Rajkumar and Panjabi 2022; Roy et al. 2021; Wang et al. 2021b; Werner et al. 2017; Werner et al. 2018; Würschum et al. 2014; Zou et al. 2016). Prediction accuracy responses in *B. napus* are shown to be consistent with previous studies in other crops, exhibiting negative correlation to trait complexity while accuracies can be improved through the increase in TP size and TP/VP relatedness (Rajkumar and Panjabi 2022).

A study by Hu et al. (2021b) explored the effectiveness of GS in *B. napus* hybrid breeding/prebreeding with diverse germplasm containing exotic introgressions, attaining high prediction accuracies for heterosis in traits including yield, yield-related traits, and seed quality traits. Hu et al. (2021b) observed that trait complexity, TP size, TP/VP relatedness, exotic introgression, and environments were significant factors affecting GS accuracy, but the choice of GS model was not. Another study by Wang et al. (2021) also predicted heterosis in a diverse set of *B. napus* with machine learning models, emphasizing GS' ability to improve the efficiency of heavily resource-dependent combining ability trials which precede hybrid breeding. Major QTL for blackleg resistance analyzed by Fikere et al. (2018) explained less than 30 % of genetic variance, however prediction accuracies attained in this study demonstrated GS' ability to capture minor-effect QTL, emphasizing improvement upon MAS and the feasibility for genomic breeding in canola disease resistance. Furthermore, using low-density marker sets filtered by LD, Werner et al. (2018) attained prediction accuracies that were comparable to those achieved by multifold larger marker sets.

Studies in *B. napus* have utilized GS to predict yield and yield-related traits such as seeds per pod, silique number, thousand kernel weight, flowering time, maturation time, shattering, and more (Fikere et al. 2020; Werner et al. 2017; Luo et al. 2017; Koscielny et al. 2020; Hu et al. 2021a). Seed-quality traits including glucosinolates, protein content, oil content, and fatty acid profile components (erucic, linoleic, linolenic, oleic, stearic) have also been predicted with GS (Fikere et al. 2020; Werner et al. 2017; Werner et al. 2018; Koscielny et al. 2020; Hu et al. 2021a; Zou et al. 2016). The effectiveness of GS for disease resistance predictions, namely sclerotinia stem rot and blackleg, have also been investigated (Derbyshire et al. 2021; Fikere et al. 2018; Roy et al. 2021).

Considering GS' great potential to improve plant breeding, further advancements in WGR have included the development of multi-trait models, multi-environment models, and incorporation of high-throughput phenotyping (HTP) platforms (Bhatta et al. 2020; Jia and Jannink 2012; Jiang et al. 2015; Lado et al. 2018; Sandhu et al. 2021a; Jubair and Domaratzki 2023). These advanced WGR methods have demonstrated their potential to further improve genetic gains while reducing resource dependence, particularly for traits of lower heritability that correlate with traits of higher heritability (Guo et al. 2020). Multi-environment data (e.g. precipitation, temperature, soil fertility, etc.) can be integrated to further improve GS accuracy (Khaki and Wang 2019; Washburn et al. 2021). Lado et al. (2018) suggested that a single correlated trait can be leveraged, providing a resource-efficient method which achieves prediction accuracies that are only marginally different than when incorporating multiple traits. Sandhu et al. (2021a) found that multi-trait GS models outperformed uni-trait GS models in predicting grain yield and protein content. Comparing the Bayesian multi-trait multi-environment model to other models, Guo et al. (2020) observed that low heritability traits, such as grain yield, exhibited more significant advantages than higher heritability traits, such as thousand kernel weight, achieving a 46 % prediction accuracy increase for the former, compared to 11 % for the latter. The use of HTP platforms such as drone imaging have also demonstrated their potential in enhancing breeding efficiency and effectiveness when combined with GS. By accurately, efficiently, and cost-effectively capturing large-scale phenotypic variation, the incorporation of HTP improves genetic gain by accelerating breeding cycles, minimizing the loss of important alleles due to linkage drag, and capturing genotype-environment associations (Jubair and Domaratzki 2023).

2.10 Genome-wide association study-guided genomic selection

Since its proposal in 2001, genomic selection (GS) methodology has developed towards improving the effectiveness of plant and animal breeding. The integration of next-generation sequencing and high-throughput phenotyping techniques has made GS particularly attractive to breeders by capturing extensive genotypic and phenotypic variation, optimizing resource allocation, reducing costs, and shortening overall breeding cycles (Heffner et al. 2010; Jubair and Domaratzki 2023; Krishnappa et al. 2021; Varshney et al. 2017). Despite being broadly evaluated in plant breeding methodologies, GS can be considered as still in its early stages wherein the multitude of inherent factors that influence GS prediction accuracy has thus far prevented a universal approach (i.e. no "one size fits all") (Lorenz et al. 2011; Zhang et al. 2023). Although not a solution to the no "one size fits all" problem, one method known as "GWAS-guided GS" has arisen as an option to improve the accuracy and efficiency of GS.

The central concept of GWAS-guided GS is the selection of optimally predictive markers (based on GWAS results) as inputs for GS model training and phenotypic predictions. The GWAS-guided GS method differs from MAS which relies on only one or very few major-effect QTL requiring prior identification, mapping, and diverse germplasm validation prior to implementation in plant breeding programs (Collard and Mackill 2008; Kaler et al. 2022; Krishnappa et al. 2021). Werner et al. (2018) describes the capacity of GWAS-guided GS as allowing fewer representative markers selected in respect of linkage disequilibrium (LD) to capture the association between genomic regions and a phenotypic trait. Low-density, GWAS-based marker sets enable GS prediction accuracies in breeding populations comparable to those achieved with high-density genotyping (Werner et al. 2018). Importantly, GWAS-guided GS mitigates model overfitting by systematically excluding the bulk of non-causative markers from GS model training, thereby avoiding erroneous identification of these random variations as causative for the trait of interest (Hickey et al. 2014; Montesinos-López et al. 2022b; Zhang et al. 2023). Simultaneously, although the marker subsets are typically small, GWAS-guided GS mitigates underfitting by ensuring that the selected markers provide sufficient effect information for the model to perform accurate phenotypic predictions (Montesinos-López et al. 2022b; Zhang et al. 2023).

The critical question arises though: how does one choose the subset of markers which provides optimal predictability for a trait of interest? Several methods have been developed to answer this

question. Zhang et al. (2014) proposed the use of preexisting GWAS results from public databases to build trait-specific genomic relationship matrices for phenotypic predictions, later described as "GS + historic GWAS". Spindel et al. (2016) proposed the "GS + de novo GWAS" method which inputs significant markers identified through GWAS within the specific training population of interest (i.e. de novo) to train the GS models as fixed effects. More recently, Jeong et al. (2020) developed a program known as "GMStool" which uses machine learning to select optimal marker sets and optimal WGR models for phenotypic prediction (Jeong et al. 2020).

Using historic GWAS results to guide GS in a population of interest can be considered problematic given the often population-structure-dependent nature of GWAS results, such that significant loci identified within an independent population might not reflect the true genetic structure of traits in the population of interest (Spindel et al. 2016). Alternatively, by implementing predictive markers identified within the actual population of interest, de novo approaches should theoretically be more effective than the GS + historical GWAS methods or conventional GS (Spindel et al. 2016).

2.10.1 Use of genome-wide association study-guided genomic selection in plant breeding

Several studies have investigated the benefit of incorporating GWAS results to training GS models. These studies have compared a variety of GWAS models, marker subset methods, and GS models in application to several species and traits. Rice, the first crop whose genome was sequenced, has been a focus of GWAS-guided GS research, starting with the proposal of GS + historic GWAS by Zhang et al. (2014). Spindel et al. (2016) proposed the GS + de novo GWAS method in rice, observing that the combination of simple Multiple Linear Regression (MLR) GWAS results with rrBLUP outperformed several other WGR models. Since then, the GS + de novo GWAS method has been evaluated for rice breeding, investigating the predictive ability of GS models (including GBLUP, rrBLUP, RF, RKHS, BayesA, BayesB, BRR, Deep Neural Network (DNN), etc.) for important agronomic and quality traits such as yield, panicle number, flowering time, plant height, protein content, and more (Jeong et al. 2020; Zhang et al. 2023). Wheat has also been the subject of GWAS-guided GS research wherein traits such as yield and septoria blotch resistance have been predicted using rrBLUP and GBLUP (Odilbekov et al. 2019; Sehgal et al. 2020). Even Norway spruce, an important tree for lumber and holiday celebrations (and the first gymnosperm with a sequenced genome), has seen GWAS-guided GS predictions towards frost damage, wood density, and budburst stage, comparing the accuracy of pedigree-based BLUP (PBLUP) and GBLUP (Chen

et al. 2023; Nystedt et al. 2013). *Brassica napus* has also been the subject of GWAS-guided GS research. Werner et al. (2017) utilized the method by Spindel et al. (2016), combining GWAS results from GenABEL P+K to evaluate the ability of rrBLUP, BayesB, and BRR to predict general and specific combining ability in *B. napus*. Next, Werner et al (2018) used a similar method to predict oil content, glucosinolate content, and plant height. More recently, a study by Sun (2021) compared the prediction accuracies of conventional GS to GWAS-guided GS for several agronomic and seed quality traits including yield, plant height, oil content, glucosinolate content, and protein content.

2.11 Research objectives

In Canada, the demand for canola is projected to reach 26 million metric tonnes by 2025. Fulfillment of this goal is driven not only by increasing seed yield, but also by filling seed quality niches such as high protein meal and specialty oils. Conventional selection methods and MAS are limited in their ability to meet these demands expeditiously or sustainably. The overall objective of this research was to demonstrate the potential of two bioinformatics methods, GWAS and GS, to streamline the development of *B. napus* cultivars that address both consumer needs and agricultural challenges.

The primary objective of the first study (Chapter 3) was to identify seed quality QTL for potential application in HEAR breeding while also examining the effects of population structure, marker density, and model selection on QTL discovery. The second study (Chapter 4) assessed GS in HEAR, exploring a broad range of potential breeding scenarios for seed quality traits. A total of 135 unique analyses were performed, evaluating prediction accuracy across nine regression models, five population designs, and three marker densities. The third study (Chapter 5) evaluated GWAS-guided GS, a method proposed to improve the accuracy of GS predictions, by applying a similar range of analyses used in Chapter 4 to compare prediction accuracies with conventional GS. Plant breeders may use this information regarding responses in QTL discovery and GS prediction accuracy to inform decisions on factors (population design, genotyping density, choice of model, etc.) to meet their breeding goals and optimize resource allocation.

3. GENOME-WIDE ASSOCIATION STUDY FOR OIL AND FATTY ACID PROFILE IN RAPESEED (*Brassica napus* L.)

3.1 Abstract

Genome-wide association studies (GWAS) have been widely implemented for *Brassicas* and other crop species to investigate the genetic architecture of important agronomic and quality traits. Significant marker trait associations (MTAs) identified by GWAS are often used for marker assisted selection (MAS); however, they may also be helpful in molecular dissection of the trait in question. This study examined a population of 454 high erucic acid rapeseed (HEAR) genotypes (92 parents, 362 hybrids) by inputting various population designs, marker sets, and GWAS models to discover QTL for five seed quality traits: erucic acid content (ERU), linoleic acid content (LLE), linolenic acid content (LLN), oleic acid content (OLE), and overall oil content (OIL). FarmCPU identified 89 peak QTL across the five traits. Several peak QTL identified in this study co-localize with QTL described for the same traits in literature. Thirty previously discovered fatty acid biosynthesis candidate genes coincide with significant loci identified in this study. One region on chromosome C04 was highly significant to all three of ERU, OLE, and LLE, indicating a novel candidate gene that warrants further research for oil profile improvement. This study demonstrated the effects of population design, marker density, and choice of model specifically in application to GWAS for *Brassica napus* L. fatty acid profiles. Significant effects were observed in population design, choice of GWAS model, and (to a lesser extent) marker set density. Population design especially highlighted performance differences among GWAS models, identifying FarmCPU as the optimal model. This study emphasized the effectiveness of GWAS for investigating seed quality traits in *B. napus* improvement.

3.2 Introduction

Brassica napus L. has surpassed wheat (*Triticum aestivum* L.) as Canada's most widely sown crop species (Statistics Canada 2024). *B. napus* crops represent the second-largest oilseed commodity in the world, producing 84.34 million metric tons globally in 2022, second only to soybeans (USDA 2023). The term "canola" describes seeds of the genus *Brassica* (*B. napus*, *B. rapa*, or *B. juncea*) which are "double-low," more specifically, contain less than 2% erucic acid

in its fatty acid profile and less than 30 micromoles of any glucosinolate per gram of the oil-free solid component (Canola Council of Canada 2024c). In 2023, canola was sown across 22.1 million acres, contributing an average of $43.7 billion to the Canadian economy between 2020 and 2023 while employing around 206,000 Canadians during that period (Canola Council of Canada 2024a; GlobalData 2024; Statistics Canada 2024). Canada is the largest producer of canola, accounting for over 20 % of worldwide production and 70 % of international exports (AgMRC 2022). The value of canola is derived from separating its high-quality oil (primarily for food industry) from the protein-rich meal (primarily for livestock feed).

While canola dominates the Canadian oilseed market, high erucic acid rapeseed (HEAR) grows alongside it, contributing to a wide range of industries beyond food and feed. High erucic acid oil is growing in demand for its use in biofuels, biodegradable plastics, industrial lubricants, surfactants, cosmetics, and pharmaceuticals like behenyl alcohol, the antiviral agent in Abreva® (Aoki et al. 2015; Aukema et al. 2011; Leonard 1992; Wang et al. 2022). The United States is the largest consumer of HEAR oil, followed by the European Union and Japan (Bhardwaj and Hamama 2000; USDA 1996). Predominantly used in biofuel production, rapeseed oil constituted around 45 % of EU biodiesel consumption in 2016 (Dusser, 2019). Key goals for HEAR cultivar improvements include enhancing yield, oil content, meal quality, erucic acid content, and disease resistance (Zelmer and McVetty 2009).

Canola oil is mainly constituted by three fatty acids: oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3), along with small proportions of saturated and other trace monounsaturated and polyunsaturated fatty acids (Barthet 2022; Przyblyski and Eskin 2011). The HEAR fatty acid profile is very similar to that of canola in that it also contains oleic acid, linoleic acid, and linolenic acid, however erucic acid constitutes a significant proportion of the total oil profile, accompanied by a reduction in other fatty acids relative to canola (Przyblyski and Eskin 2011). The demand for *B. napus* oil in food, feed, and industry has prompted breeders to develop specialty cultivars providing improved shelf life, heat stability, and functional foods (high linoleic:linolenic ratio), for higher-value meal in food and feed (increased protein quality and quantity) or biofuels and other industrial uses (high erucic acid content) (Abbadi and Leckband 2011; Fitzpatrick and Scarth 1998; Nesi et al. 2008; Wittkop et al. 2009).

Genome-wide association studies (GWAS) describe a range of methods that are widely implemented in plant improvement to investigate the genetic architecture of important agronomic and quality traits (Cortes et al. 2021). Results from GWAS are often implemented for marker-assisted selection (MAS). Additionally, MTAs identified by GWAS can be used in linkage mapping, investigating candidate gene function, genomic selection, and more (Bian and Holland 2017; Deng et al. 2017; Fiedler et al. 2017; Liu et al. 2019; Tsai et al. 2020; Wang et al. 2016; Zhu et al. 2011). The effectiveness of GWAS, however, is affected by several factors, including population size and diversity, marker coverage, and the choice of GWAS model (Alqudah et al. 2020; Cortes et al. 2021; Khan et al. 2019; Lander and Schork 1994; Liu et al. 2016b; Purcell et al. 2003; Risch and Merikangas 1996; Segura et al. 2012; Uffelmann et al. 2021; Visscher et al. 2012; Yu et al. 2005). The power of GWAS analyses to detect small-effect genes in high resolution relies on sufficient phenotypic and genotypic diversity provided by large and diverse populations (Lander and Schork 1994; Purcell et al. 2003; Risch and Merikangas 1996; Uffelmann et al. 2021; Visscher et al. 2012). Sufficient marker coverage for GWAS is often determined by the LD decay rate, ensuring that allelic segregation is captured across haplotype blocks (Alqudah et al. 2020). Single-locus GWAS (SL-GWAS) models have become commonplace in GWAS research as they correct for population structure and kinship using relatively simple computation (Yu et al. 2005). Multi-locus GWAS (ML-GWAS) models, such as FarmCPU and MLMM, have been developed to improve on SL-GWAS models, providing high statistical power and maintaining control of false positive and false negative signals caused by population structure (Cortes et al. 2021; Khan et al. 2019; Liu et al. 2016; Segura et al. 2012).

Like many crop species, GWAS has become a common tool in genetic research and improvement of *B. napus* cultivars. A wide range of QTL and candidate genes have been identified for important traits affecting yield, physiology, disease resistance, abiotic stress resistance, and more. Seed quality traits have been investigated using GWAS, revealing QTL and candidate genes for oil content (Li et al. 2014; Pal et al. 2021; Tang et al. 2021; Wang et al. 2021a; Xiao et al. 2019; Yao et al. 2021; Zhao et al. 2022), erucic acid content (Korber et al. 2016; Li et al. 2014; Wang et al. 2018a; Xiang et al. 2023), oleic acid content (Xiang et al. 2023; Zhao et al. 2019), glucosinolate content (Bhinder et al. 2022; Tan et al. 2021; Tang et al. 2023a; Xiang et al. 2023), fiber content (Gacek et al. 2021), and seed coat color (Wang et al. 2017a).

The main objective of this study was to discover QTL for *B. napus* oil and fatty acid profile traits by conducting an extensive range of GWAS analyses. This study also aims to highlight the effects of population design, marker density, and choice of GWAS model by inputting various combinations of these factors. Broadly, this study aims to demonstrate the benefits and drawbacks granted to *B. napus* breeders as a consequence of the various materials and methods selected prior to conducting GWAS. We hypothesize that several markers will be identified in significant association with the traits of interest. We also hypothesize that population design, marker density, and the choice of model will impose significant effects on GWAS' ability to identify true marker-trait associations.

3.3 Materials and Methods

3.3.1 Plant materials and phenotypic data

This study evaluated a total of 454 HEAR genotypes consisting of 31 females (B-lines) and 61 males (R-lines) which were crossed to develop 362 F_1 hybrid genotypes (H). Hybrids were developed using the *Ogu* cytoplasmic male sterility (CMS) system (Ogura 1968) developed by the Institute for Agricultural Research (INRA) (now National Research Institute for Agriculture, Food and Environment, INREA). This population was previously analyzed by Sun (2021), from which these materials and methods are summarized. Parents were evaluated in five site-years across southern Manitoba, Canada (Supplementary S3.1). Each site-year contained three sub-experiments. In each sub-experiment one third of the parental genotypes and control genotypes were tested using a randomized complete block design (RCBD) with three replicates (blocks). Hybrids were evaluated in 43 site-years across Alberta, Saskatchewan, and Manitoba, Canada (Supplementary Table S3.2). Of the 362 hybrid genotypes, 279 were replicated five to nine times, 34 were replicated ten to fifteen times, and 49 were replicated more than 15 times (unequal replications due to selection in the HEAR breeding program, wherein favourable genotypes were selected for further field experiments). All site-years were sown using 6 m x 1.6 m plots with six rows at 0.20 m row spacing (except Glenlea 2016, which used 3 m double-nursery rows at 0.40 m row spacing).

Edge® Granular Herbicide (5 %) (Gowman Canada, Winnipeg) was incorporated into the soil at 30.9 kg ha^{-1} in the fall before each field experiment. Soil tests were conducted in the fall to determine appropriate fertilizer application rates to achieve a yield goal of 2.5 t ha^{-1} based on the

nutrient recommendations from the Canola Council of Canada (2020) (Table 3.1). Fertilizer applications varied depending upon the year, but averaged 110 kg N, 44 kg P, 13 kg S and 18 kg K per hectare.

All seeds were treated with HELIX XTra® (Syngenta Canada Inc., Calgary) at 15 mL kg-^{1} seed before seeding. Insecticides were applied five days after emergence followed by another application one week later. When temperature at the time of application was below 25 °C, Decis® (Bayer CropScience, Leverkusen, Germany) was applied at 0.2 L ha^{-1} for control of flea beetles. When temperature was above 25 °C, Matador® (Syngenta Canada Inc., Calgary) was applied at 0.08 L ha^{-1}. Herbicides were applied at the 2-4 leaf stages (BBCH 12-14). A mixture of Poast® Ultra (BASF, Ludwigshafen, Germany), Muster® (DuPont Canada, Mississauga, ON), and Lontrel™ 360 (Dow AgroSciences, Indianapolis, IN) were applied during BBCH 12-16 at 0.67 L ha^{-1}, 4.9 g ha^{-1} and 0.67 L ha^{-1}, respectively, for grassy weed and broadleaf weed control. Fungicide PROLINE® (Bayer CropScience, Leverkusen, Germany) was applied during flowering time (BBCH 61-65) at 0.37 L ha^{-1} for management of sclerotinia stem rot. When necessary, Pounce® 384EC insecticide was applied at 0.18 L ha^{-1} for control of flea beetles prior to swathing.

Table 3.1 Canola nutrient requirements of N, P, S, K for a target yield of 2.5 t ha^{-1}. Calculation based on recommendations from Canola Council of Canada (2020).

Soil nutrient	Recommended rate for one kg seed yield	Recommended rate for target yield (kg ha^{-1})
Nitrogen (N)	0.13 to 0.88	100.80 to 151.20
Phosphate (P)	0.06 to 0.07	63.00 to 75.60
Sulphur (S)	0.02 to 0.04	25.20 to 40.30
Potassium (K)	0.10 to 0.11	115.92 to 126.00

Table 3.2 A summarized biologische bundesanstalt, bundessortenamt, and chemical industry (BBCH) scale of *Brassica napus*[1].

Stage	BBCH code	Description	
		Beginning of stage	End of stage
0: Germination	00-09	Dry seed	Emergence
1: Leaf development	10-19	Cotyledons completely unfold	≥ 9 leaves
2: Formation of side shoots	20-29	No side shoots	≥ 9 side shoots
3: Stem elongation	30-39	No internodes	≥ 9 visibly extended internodes
4: Development of harvestable vegetative plant parts	40-49	N/A	N/A
5: Inflorescence emergence	50-59	Flower buds present, yet still enclosed by leaves	First petals visible, flower buds still closed
6: Flowering	60-69	First flower open	End of flowering
7: Development of fruit	70-79	10 % of pods reach final size	Nearly all pods reach final size
8: Ripening	80-89	Seed green, filling pod cavity	Nearly all pods ripe, seeds black and hard
9: Senescence	90-99	Plants dead and dry	Harvested product

[1] Adapted from Lancashire et al. (1991).

Plots were swathed at growth stage BBCH 87 when the majority of seeds had reached physiological maturity (30-35 % seed moisture). Plots were harvested by combining at growth stage BBCH 99 (8-10 % seed moisture). Seeds from each plot were dried and cleaned manually.

Overall oil content (OIL) was measured using a FOSS NIR System (Model 6500, Foss NIR Systems Inc., Maryland, USA) at the Canadian Grain Commission certified seed quality lab at the University of Manitoba. Calibration of the NIR instrument was performed based on protocol by DeClercq et al. (1998) and measurements were performed based on protocol by Elahi et al. (2016). Fatty acid contents of erucic acid (ERU), linoleic acid (LLE), linolenic acid (LLN), and oleic acid (OLE) were analyzed using gas chromatography (GC) (Varian, Walnut Creek USA) as documented by Hougen and Bodo (1973).

Best linear unbiased predictions (BLUPs) were calculated from compiled site-year phenotypic data of parental and hybrid genotypes using META-R (Alvarado et al. 2020). Phenotypic data across experiments were analyzed as a Randomized Complete Block Design (RCBD) and environment (site-year) was input as a random effect. All remaining settings were set to default.

3.3.2 Genotypic data

All genotypes were grown in the University of Manitoba Department of Plant Science greenhouse facility (day temperature 25 °C; night temperature 22 °C; relative humidity 40-50 %; light cycle 16 h light, 8 h dark). From the first two true leaves on one plant of each genotype, a total of ~0.05 g leaf tissue was sampled and stored at -80 °C. Genomic DNA extraction was performed following CTAB protocol (Porebski et al. 1997) with modifications that excluded the use of polyvinylpyrrolidone and 2-mercaptoethanol, and the replacement of octanol with phenol. A NanoDrop^{TM} 2000 Spectrophotometer (Thermo Scientific, MA, USA) was used to check DNA sample quality and quantify DNA. Isolated DNA samples were then adjusted to 50 ng/µL for genotyping. All 454 lines were genotyped at Agriculture and Agri-Food Canada (AAFC) Saskatoon (Dr. Isobel Parkin's lab) using the *Brassica* 60K Illumina Infinium SNP array (Illumina Inc., CA, USA) (Clarke et al. 2016). Raw intensity data files and a custom cluster file were obtained from AAFC.

Three SNP marker sets were developed. Herein, all marker sets are denoted MS1 (high-density), MS2 (intermediate-density), and MS3 (low-density). Genotypic data files were input to

GenomeStudio (Illumina Inc., CA, USA), from which a full report of all markers and genotypes was exported and rearranged to hapmap format. TASSEL 5 (Bradbury et al. 2007) converted the full report from hapmap to PLINK format for import to PLINK v1.90b7 (Purcell et al. 2007). All markers with call rate < 0.95 were filtered out.

All markers were blasted to the long-read assembly of the *Darmor bzh* reference genome by Rousseau-Gueutin et al. (2020). Quality control was performed by filtering the hits with more than 70 percent coverage, 0 gaps, and 1 or less mismatches. The high-density MS1 contains all coinciding makers that passed both the samples' call-rate filter and the assembly's quality control filter. A total of 22,941 markers coincided to produce MS1, providing high confidence in both true physical location and sample quality. The intermediate-density MS2 was produced by applying an additional filter to the markers in MS1: all markers with minor allele frequency (MAF) < 0.05 were removed, producing MS2 with 14,699 markers. The low-density MS3 was produced by applying the PLINK indep-pairwise filter to MS1 with parameters 50 5 0.2 (in a scanning window of 50 markers and scanning interval of 5 markers, remove markers with LD above the r^2 threshold of 0.2) producing a final marker set containing 1,098 markers.

The genome-wide density of all marker sets was visualized using the SNP Density plot function in SRplot (Tang et al. 2023b).

The 454 *B. napus* genotypes were filtered for sample quality. All genotypes whose marker call rate < 0.8 were removed, leaving 427 genotypes (91 parents, 336 hybrids).

3.3.3 Linkage disequilibrium

Linkage disequilibrium (LD) decay was calculated based on MS2 using the TASSEL linkage disequilibrium function. The output of the LD function was exported and visualized in R Studio using the method by Ali (2021), adapted from Remington et al. (2001). The threshold LD decay distance was indicated as the distance at which the r^2 value between markers dropped below 0.2. Threshold LD decay distance was calculated for each individual chromosome, the A subgenome, the C subgenome, and the whole genome.

3.3.4 Principal component analysis

Population structure was evaluated using PCA analysis. The LD-filtered MS3 was input to TASSEL's PCA analysis with parameters set to default and number of components set to 3, as suggested by Sun (2021). The eigenvalues array was then exported and visualized in RStudio using ggplot2 (Wickham 2016).

3.3.5 Genome-wide association study

Marker set files were subset to compare GWAS results based on population. Two populations were compared: combined (all 427 genotypes) and parents (only 91 parent genotypes). Thus, four marker set files were used: MS1 × combined, MS1 × parents, MS2 × combined, and MS2 × parents.

Three GWAS models were run through GAPIT V. 3.1 (Wang and Zhang 2020) including CMLM (Compressing Mixed Linear Model), MLMM (Multi-Locus Mixed Linear Model), and FarmCPU (Fixed and random model Circulating Probability Unification). Each GAPIT model was run with total principal components set to 3. Two mixed linear models were run through TASSEL (Bradbury et al. 2007) including Mixed Linear Model with Kinship (MLM+K) and Mixed Linear Model with Kinship and Principal Component Analysis (MLM+K+PCA). Kinship analyses were run using TASSEL defaults, while PCA analyses were run with all parameters set to default and number of components set to 3. MLM models were run using optimum level compression and P3D variance component estimation. Finally, GenABEL P+K (Aulchenko et al. 2007) was run through RStudio V. 3.6.3 (RStudio Team 2020). Code shared in communication with Dr. Chakrabarty (2023) was adapted to nullify marker data filters and run using 3 principal components. Four of these models (MLM+K, MLM+K+PCA, CMLM, and GenABEL P+K) are classified as single-locus GWAS (SL-GWAS) models. The remaining two models (FarmCPU and MLMM) are classified as multi-locus GWAS (ML-GWAS) models.

Bonferroni thresholds were calculated for both marker sets using the equation shown in [1].

$$Bonferroni - corrected\ p = -\log\left(\frac{\alpha}{n}\right) \qquad [1]$$

where α = p value, in this study, 0.05;

n = total number of markers, in this study, were either 22,941 (MS1) or 14,699 (MS2)

The Bonferroni threshold significance values ($-\log_{10}(p)$) for MS1 and MS2 were calculated as 5.661642 and 5.468318, respectively.

Overall, 120 unique GWAS analyses were conducted by combining the various factors described above (Figure 3.1). Results from all GWAS analyses were visualized in RStudio with Q-Q plots using "CMplot" (Yin et al. 2021). Indicators of systematic bias (from stratified population structure, model approach, genotyping errors, etc.) are visualized on Q-Q plots by visualizing the extent to which the observed distribution of p-values follows the expected (null) distribution (Adkins and Shabalin 2023). Early separation of the observed $-\log_{10}(p)$ values above the expected line indicates that many moderately significant p-values are more significant than expected with the null hypothesis – that is, many p-values produced by the GWAS analysis are smaller (more significant) than expected from chance alone (i.e., false positives) (Ehret 2010).

For each trait, significant SNPs that were identified within close proximity (within the corresponding chromosomal LD decay threshold) by separate GWAS analyses were grouped as "consensus groups", denoting significant markers which are likely associated to the same candidate gene. Within consensus groups, the marker with the lowest p-value (greatest significance) was identified as the peak QTL. Peak QTL also include significant SNPs that were not in consensus with other significant SNPs from separate analyses.

3.3.6 Coinciding QTL and candidate genes in literature

Confidence intervals (CIs) around each peak QTL were created based on chromosomal LD decay thresholds. For each peak QTL, the CI was identified as the SNP marker's position minus/plus the corresponding chromosome's LD decay distance. Coinciding QTL were considered as QTL identified in literature which fall within the CI of peak QTL for the same seed quality trait identified in this study. Coinciding candidate genes were denoted as oil and fatty acid biosynthesis candidate genes identified in literature whose significant association regions align with peak QTL CIs identified in this study.

Most candidate genes identified in literature were published using the AGI (Arabidopsis Genome Initiative) or *Darmor-bzh* v5 naming conventions. For consistency, we have referenced those genes with the *Darmor-bzh* v5 naming convention. Table S3.10 provides a conversion (based on Rousseau-Gueutin et al. 2020) to convert gene IDs to the *Darmor-bzh* v10 naming convention.

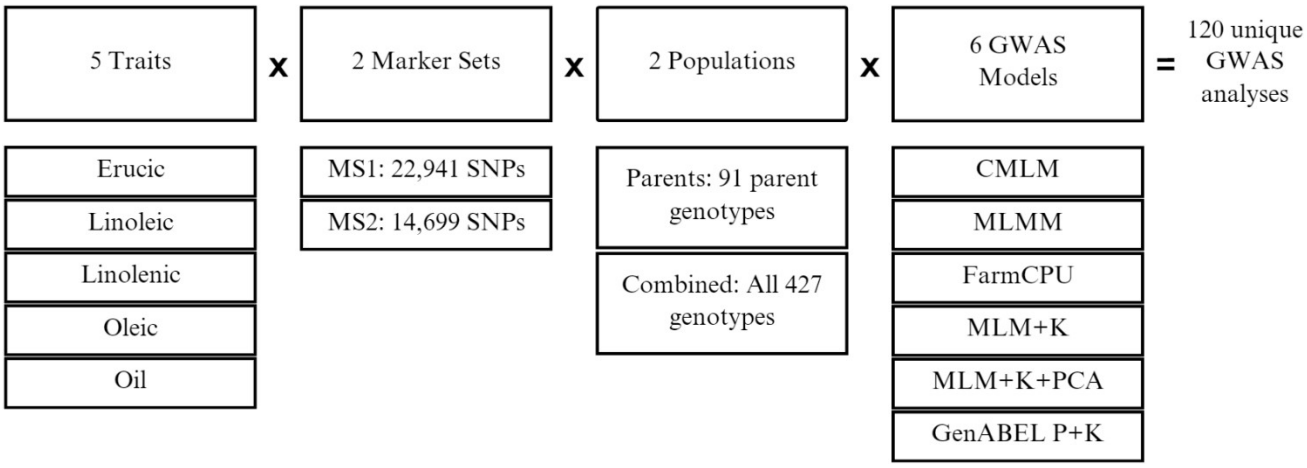


Figure 3.1 Schematic summary of all GWAS analyses in this study. Abbreviations: MS: Marker Set; SNP: Single Nucleotide Polymorphism; CMLM: Compressing Mixed Linear Model; MLMM: Multi-Locus Mixed Linear Model; FarmCPU: Fixed and random model Circulating Probability Unification; MLM + K: Mixed Linear Model with Kinship; MLM + K + PCA: Mixed Linear Model with Kinship and Principal Component Analysis.

3.4 Results

3.4.1 Phenotypic variation and correlation

According to Shapiro-Wilk tests of normality, raw phenotypic data across site-years exhibited normal distributions for traits in both the parent and hybrid genotypes. Mean overall oil content (OIL) across all site-years was 45.0 %, erucic acid (ERU) content mean was 51.1 %, linoleic acid (LLE) content mean was 10.8 %, linolenic acid (LLN) content mean was 8.4 %. and oleic acid (OLE) content mean was 14.0 %. Parent genotypes varied in OLE content from 9.7 % to 49.1 % while hybrids varied from 10.4 % to 46.2 % across all site-years. The range of OIL varied from 32.7 % to 55.0 % in parents and from 37.0 % to 57.0 % in hybrids. A comprehensive summary of the raw data across all site-years can be found in Table 3.3.

Summarization of site-year data by best linear unbiased predictions (BLUPs) takes into account site-year effects and unequal replication numbers. A singular BLUP value is provided for the five traits measured in each genotype (Table S3.3). Across all genotypes, BLUPs for OIL ranged from 36.85 % to 49.02 % with a mean of 45.05 %. Mean ERU content was 51.30 %, ranging in content from 41.98 % to 58.86 %. Again, a comprehensive summary of BLUPs can be found in Table 3.4. Herein, all discussion of phenotypic data refer to BLUP values rather than raw site-year data.

The seed quality traits in this study exhibited correlations amongst each other. All individual fatty acids demonstrated significant correlations with one another, however interestingly, OIL did not correlate significantly with any particular fatty acid (Figure 3.2). The strongest negative correlation exhibited was -0.672 between ERU and LLE. Significant negative correlations were detected between ERU and all three other fatty acids. Similarly, OLE was also negatively correlated with all three other fatty acid traits. The only positive correlation found was between LLN and LLE at 0.258, exactly opposite to OLE's correlation to LLE at -0.258.

3.4.2 Marker density

Mean SNP marker density was calculated for each chromosome, the A-subgenome, the C-subgenome, and the whole genome. By expressing marker density as kb/marker, greater values indicate greater genomic distances between SNP markers (low coverage), while smaller values indicate shorter genomic distances between SNP markers (high coverage) (Table 3.5 and Figure 3.3).

Table 3.3 Summary of raw site-year data from a *Brassica napus* L. population consisting of 31 B-lines, 60 R-lines, and 336 hybrids derived from parental genotypes. Data collected from 48 site-years across Canadian prairie provinces.

Genotypes	Statistic	ERU[1]	LLE[2]	LLN[3]	OLE[4]	OIL[5]
All genotypes	Minimum	14.3	5.42	4.72	9.72	32.7
	Maximum	60.6	15.87	11.27	49.18	57
	Mean	51.1	10.8	8.4	14.0	45.0
	Median	51.5	10.9	8.45	13.6	44.7
	Standard deviation	3.3	1.2	0.9	1.9	3.2
	Variance	11.2	1.5	0.7	3.7	10.5
	Coefficient of variation (%)	6.55	11.12	10.22	13.73	7.20
Parent genotypes	Minimum	14.3	5.4	4.7	9.7	32.7
	Maximum	60.6	15.7	11.1	49.2	55.0
	Mean	50.4	10.9	8.1	14.7	44.1
	Median	50.8	11.21	8.14	14.3	44.3
	Standard deviation	4.2	1.8	1.0	2.4	3.3
	Variance	18.0	3.2	1.0	5.9	10.9
	Coefficient of variation (%)	8.41	16.37	12.47	16.49	7.48
Hybrid genotypes	Minimum	16.0	7.8	6.4	10.4	37.0
	Maximum	59.6	15.9	11.3	46.3	57.0
	Mean	51.4	10.8	8.5	13.6	45.5
	Median	51.7	10.9	8.6	13.4	44.9
	Standard deviation	2.7	0.8	0.7	1.5	3.1
	Variance	7.5	0.6	0.5	2.1	9.7
	Coefficient of variation (%)	5.32	7.12	8.48	10.70	6.84

[1] Erucic acid content (% of total fatty acid)
[2] Linoleic acid content (% of total fatty acid)
[3] Linolenic acid content (% of total fatty acid)
[4] Oleic acid content (% of total fatty acid)
[5] Overall oil content (% of dry mass)

Table 3.4 Summary of best linear unbiased predictions (BLUPs) from a *Brassica napus* L. population consisting of 31 B-lines, 60 R-lines, and 336 hybrids derived from parental genotypes. BLUPs summarize raw site-year data collected from 48 site-years.

Genotypes	Statistic	ERU[1]	LLE[2]	LLN[3]	OLE[4]	OIL[5]
All genotypes	Minimum	41.98	6.06	5.92	10.34	36.85
	Maximum	58.86	13.42	10.19	19.24	49.02
	Mean	51.30	10.72	8.46	13.97	45.05
	Median	51.47	10.85	8.52	13.86	45.11
	Standard deviation	2.68	1.05	0.59	1.18	1.26
	Variance	7.18	1.11	0.35	1.40	1.60
	Coefficient of variation (%)	5.22	9.84	6.98	8.48	2.81
Parent genotypes	Minimum	41.98	6.06	5.92	10.34	39.81
	Maximum	58.86	13.42	9.86	19.24	49.02
	Mean	51.23	10.78	8.21	14.21	44.77
	Median	51.38	11.30	8.24	14.25	44.42
	Standard deviation	3.79	1.82	0.84	1.85	2.01
	Variance	14.39	3.31	0.71	3.44	4.02
	Coefficient of variation (%)	7.40	16.88	10.28	13.05	4.48
Hybrid genotypes	Minimum	42.26	8.63	7.08	11.54	36.85
	Maximum	57.10	12.40	10.19	17.64	47.99
	Mean	51.32	10.71	8.53	13.90	45.13
	Median	51.48	10.81	8.54	13.83	45.13
	Standard deviation	2.29	0.72	0.48	0.91	0.96
	Variance	5.23	0.52	0.23	0.83	0.91
	Coefficient of variation (%)	4.46	6.71	5.61	6.56	2.12

[1] Erucic acid content (% of total fatty acid)
[2] Linoleic acid content (% of total fatty acid)
[3] Linolenic acid content (% of total fatty acid)
[4] Oleic acid content (% of total fatty acid)
[5] Overall oil content (% of dry mass)

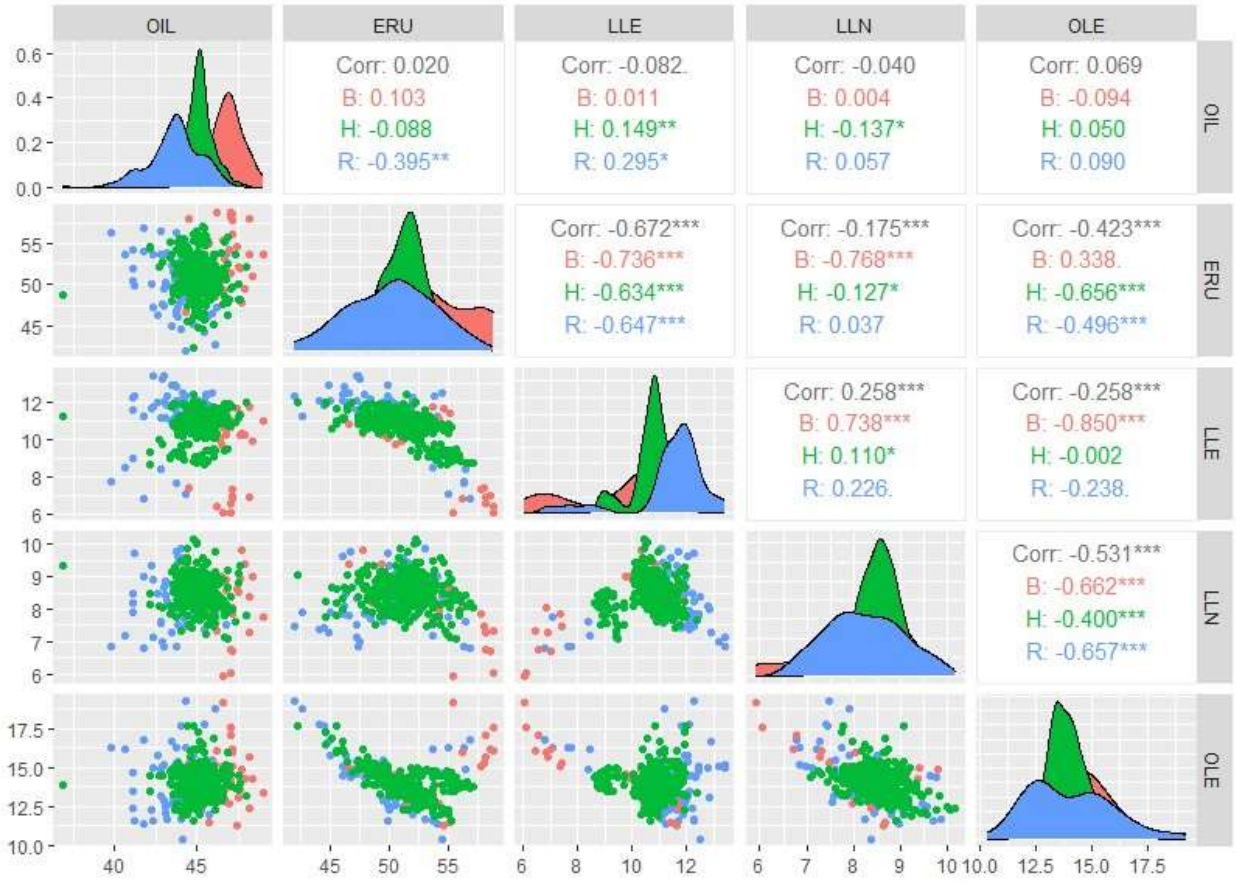
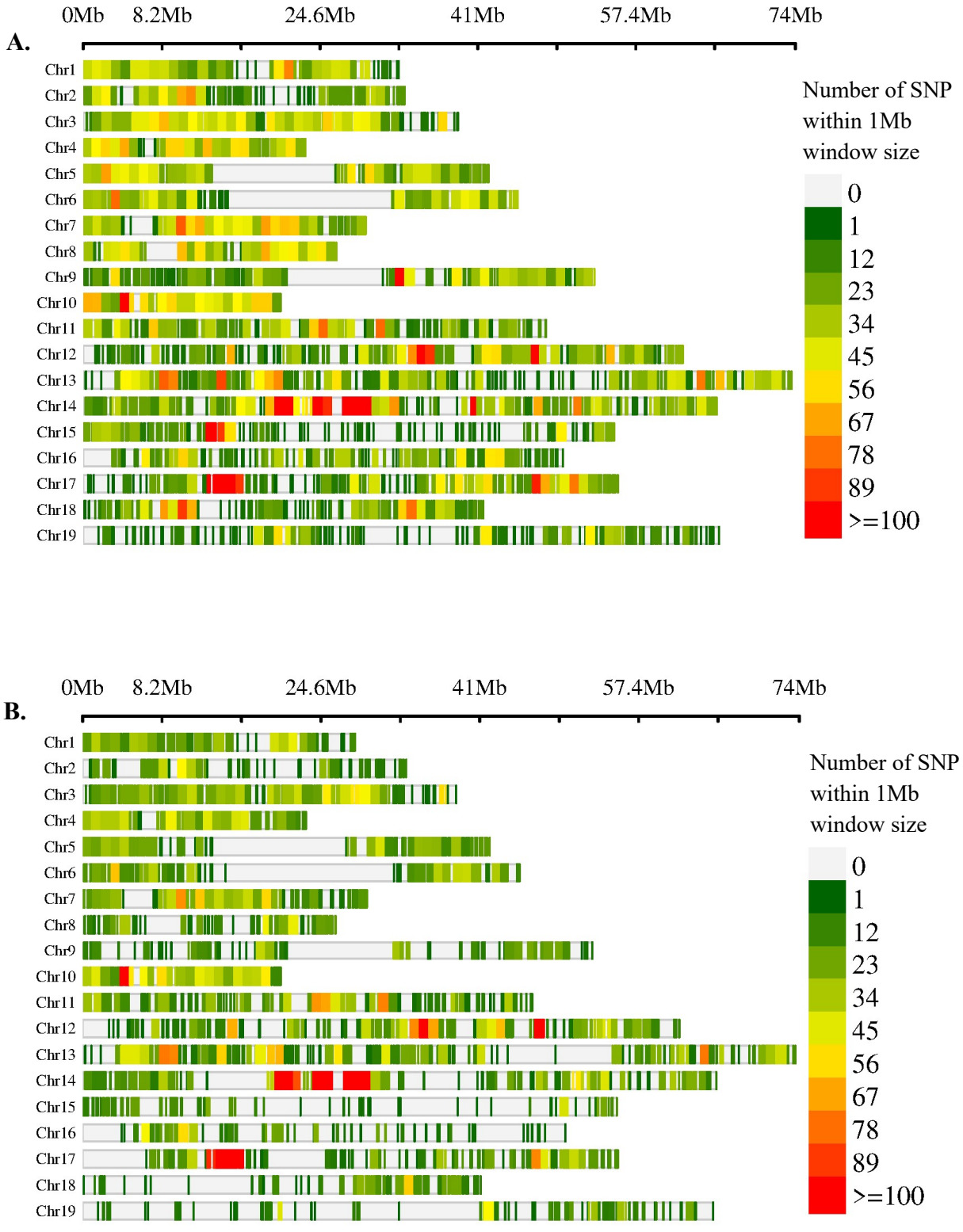


Figure 3.2 Correlation matrix of seed quality traits based on best linear unbiased predictions (BLUPs) for a *Brassica napus* L. population consisting of 31 B-lines, 60 R-lines, and 336 hybrids (H). Level of significance is noted by asterisks. The diagonal graphs the distribution of BLUPs for all traits. Abbreviations: OIL: overall oil content; ERU: erucic acid content; LLE: linoleic acid content; LLN: linolenic acid content; OLE: oleic acid content

Table 3.5 Marker density of three SNP marker sets (MS) on each *Brassica napus* L. chromosome, subgenome, and whole genome in a population consisting of 31 B-lines, 60 R-lines, and 336 hybrids developed from parental genotypes. Chromosome lengths based on the *Darmor-bzh* assembly by Rousseau-Gueutin et al. (2020).

		MS1	MS2	MS3	MS1	MS2	MS3
		Total # markers			Mean marker density (kb/marker)		
Chromosome	Length (kb)						
Whole genome	866,916	22,941	14,699	1,098	37.8	59.0	789.5
A-subgenome	346,466	10,113	6,157	570	34.3	56.3	607.8
C-subgenome	520,450	12,828	8,542	528	40.6	60.9	985.7
A01	32,959	946	579	56	34.8	56.9	588.6
A02	33,433	782	355	41	42.8	94.2	815.4
A03	39,686	1,350	990	86	29.4	40.1	461.5
A04	23,102	940	659	54	24.6	35.1	427.8
A05	42,112	1,044	586	68	40.3	71.9	619.3
A06	45,146	912	604	56	49.5	74.7	806.2
A07	29,391	1,101	710	70	26.7	41.4	419.9
A08	26,309	914	373	34	28.8	70.5	773.8
A09	53,550	1,099	452	54	48.7	118.5	991.7
A10	20,778	1,025	849	51	20.3	24.5	407.4
C01	48,239	1,274	982	51	37.9	49.1	945.9
C02	62,297	1,629	1,343	55	38.2	46.4	1,132.7
C03	73,670	1,848	1,466	83	39.9	50.3	887.6
C04	65,838	2,846	2,134	80	23.1	30.9	823.0
C05	56,383	815	313	43	69.2	180.1	1,311.2
C06	50,219	1,050	437	47	47.8	114.9	1,068.5
C07	55,657	1,604	1,082	70	34.7	51.4	795.1
C08	41,682	985	374	49	42.3	111.4	850.7
C09	66,465	777	411	50	85.5	161.7	1,329.3


A.
0Mb
8.2Mb
24.6Mb
41Mb
57.4Mb
74Mb
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Number of SNP within 1Mb window size
0
1
12
23
34
45
56
67
78
89
>=100
B.
0Mb
8.2Mb
24.6Mb
41Mb
57.4Mb
74Mb
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Number of SNP within 1Mb window size
0
1
12
23
34
45
56
67
78
89
>=100

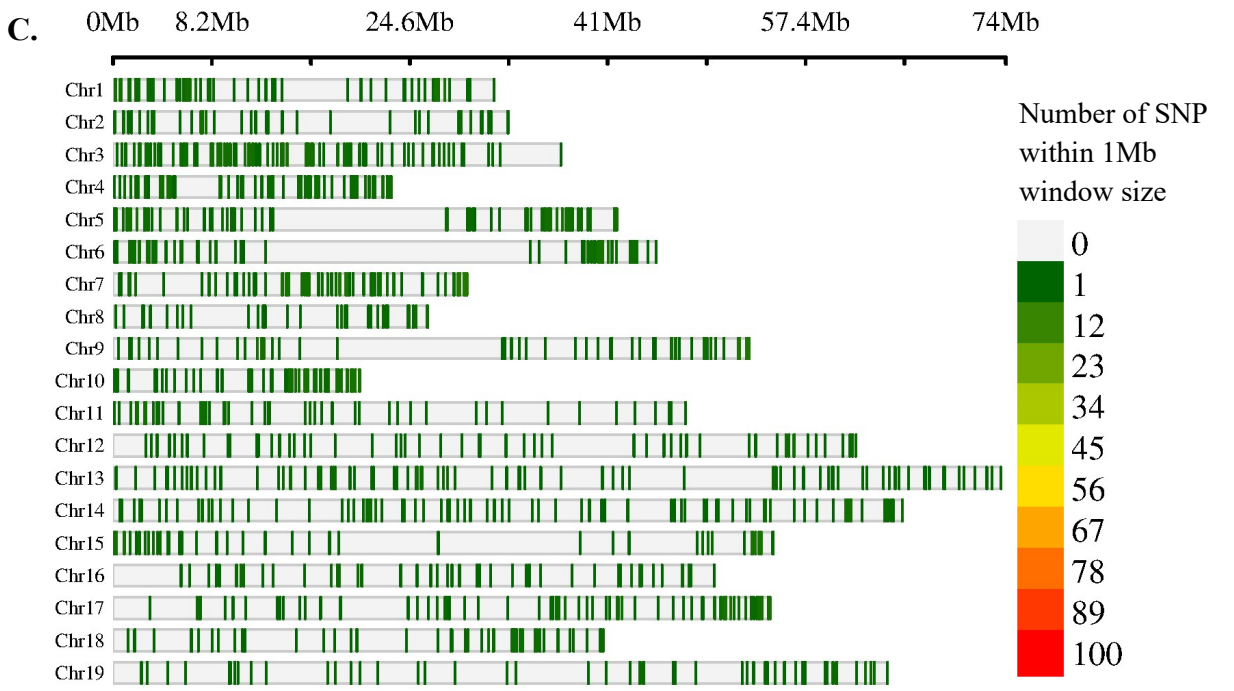


Figure 3.3 Marker density distribution in a *Brassica napus* L. population consisting of 31 B-lines, 60 R-lines, and 336 hybrids developed from parental genotypes. (A) Marker density based on MS1 (22,941 SNP markers); (B) marker density based on MS2 (14,699 SNP markers); (C) marker density based on MS3 (1,098 SNP markers).

Across the whole 866,916 kb genome, the high-density MS1 contains 37.8 kb/marker, whereas MS2 contains 59.0 kb/marker, and the low-density MS3 contains a significantly less dense 789.5 kb/marker. The shorter A-subgenome contained more dense SNP markers than the longer C-subgenome across all three marker sets. Regarding specific chromosomes, the A10 chromosome contained the most dense SNP markers across all marker sets, while the C09 chromosome contained the least dense SNP markers for MS1 and MS3.

The MAF filter which produced MS2 from MS1 reduced the number of SNP markers from 22,941 to 14,699 – a 35.9 % reduction. Chromosome C08 exhibited a reduction from 985 to 374 SNP markers, indicating that over 62 % of markers on that chromosome exhibited MAF < 0.05. Chromosome A10 exhibited only a 17.2 % reduction in markers.

The LD filter which produced MS3 from MS1 reduced the number of SNP markers from 22,941 to 1,098 – a 95.2 % reduction. All chromosomes exhibited SNP marker number decreases between 93.5 % to 97.2 %, indicating that most markers in the high-density marker set were in linkage disequilibrium with each other.

3.4.3 Linkage disequilibrium

A threshold for linkage disequilibrium (LD) decay was calculated for each chromosome, the A-subgenome, the C-subgenome, and the whole genome (Table 3.6). The A-subgenome LD decay threshold was calculated as 1,827 kb, and its constituent chromosomes ranged from 716 kb (A07) to 8,175 kb (A09). The C-subgenome LD decay threshold was calculated as 9,648 kb, much longer than the A-subgenome. The C-subgenome's constituent chromosomes ranged in LD decay threshold from 1,138 kb (C08) to 11,618 kb (C04).

3.4.4 Population structure

A PCA analysis revealed that PC1, PC2 and PC3 explained 7.08 %, 5.29 %, and 4.72 % of total genotypic variance, respectively. Together, the first three principal components explain 17.08 % of the total genotypic variance, while the first fifteen principal components explained 38.07 %. Figure 3.4 demonstrates that no clear clusters were formed based on the type of individuals (i.e. B-lines, R-lines or hybrids), therefore, no clear stratification was present in the population. Figure 3.4 also demonstrates that the hybrid genotypes were evenly distributed among the parental genotypes.

Table 3.6 Linkage disequilibrium (LD) decay threshold of all chromosomes, both subgenomes, and whole genome for a *Brassica napus* L. population consisting of 31 B-lines, 60 R-lines, and 336 hybrids. LD decay threshold calculated based on MS2 (14,699 SNP markers). LD decay threshold identified as the point where LD $r^2 = 0.2$

Chromosome	LD decay threshold (kb)
A-subgenome	1,827
C-subgenome	9,648
Whole genome	5,235
A01	2,247
A02	1,819
A03	1,686
A04	1,434
A05	1,224
A06	798
A07	716
A08	7,578
A09	8,175
A10	2,258
C01	6,375
C02	6,514
C03	5,247
C04	11,618
C05	4,819
C06	1,696
C07	11,387
C08	1,138
C09	3,004

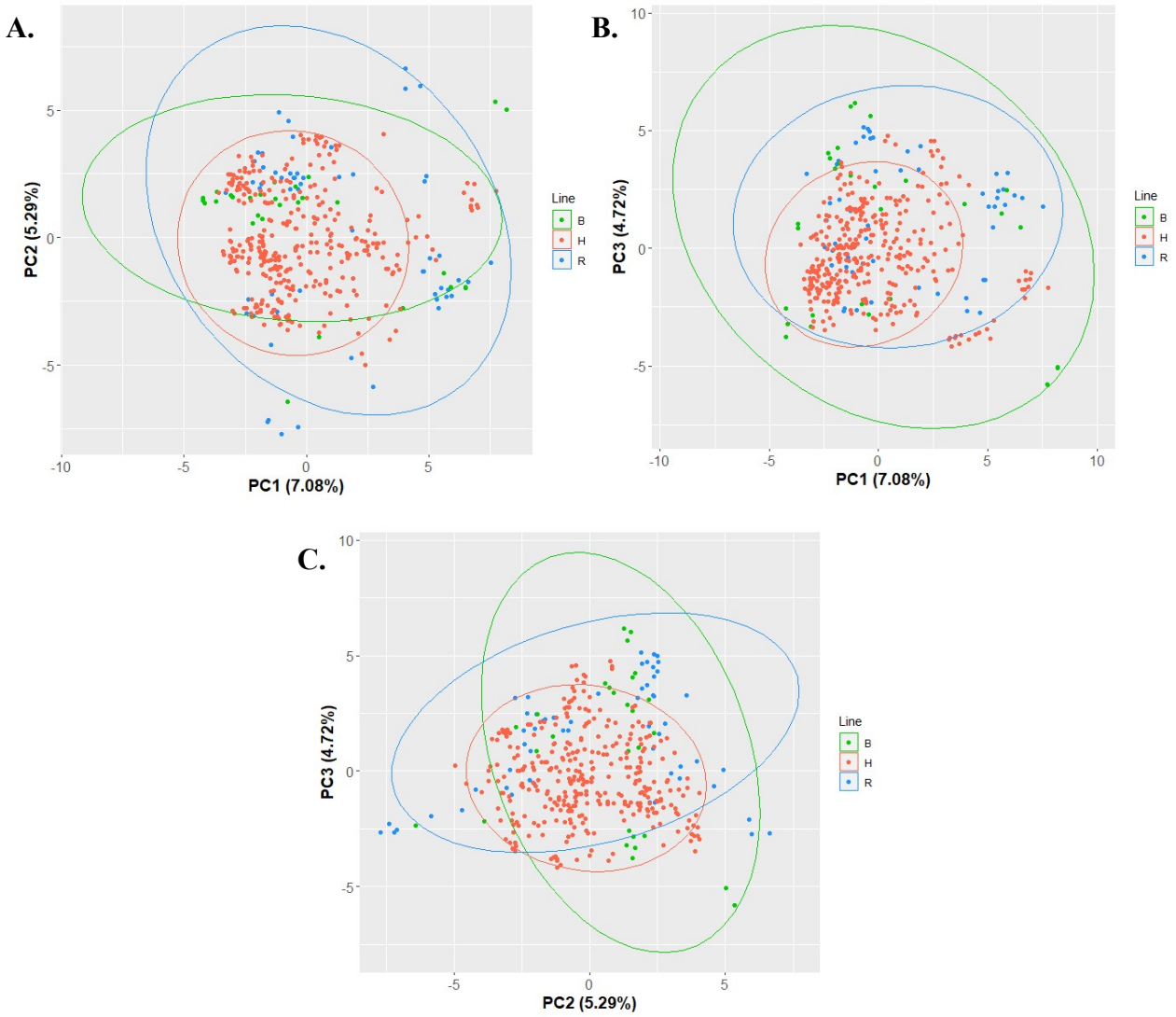


Figure 3.4 Principal component analysis (PCA) on a *Brassica napus* L. population consisting of 31 B-lines, 60 R-lines and 336 hybrids based on the LD-filtered marker set MS3 (1,098 SNP markers). (A) PC1 vs. PC2; (B) PC1 vs. PC3; (C) PC2 vs. PC3

3.4.5 Identification of QTL

Across all 120 GWAS analyses, a total of 1,400 significant SNPs were identified (using corresponding Bonferroni thresholds). FarmCPU analyses identified 89 peak QTL for all traits, producing Q-Q plots that exhibit the most consistent control of false negatives and false positives among all models. Consensus groups (groups of multiple significant SNPs identified within close physical distance by separate GWAS analyses, denoting significant markers which are likely associated to the same candidate gene) contained 34 of 89 peak QTL, while the remaining 55 peak QTL were not identified in consensus groups (Table 3.8 and Tables S3.4 – S3.6) (Bento 2024a). The method of separating significant markers by LD decay and identifying peak QTL from consensus groups promotes the representation of only distinct QTL corresponding to one significant genomic region. The identical method of identifying consensus groups and peak QTL was performed for significant SNPs identified by all 120 GWAS analyses, demonstrating additional consensus with those QTL identified by FarmCPU (Tables S3.7 – S3.9) (Bento 2024b).

Consensus among separate FarmCPU GWAS analyses provides added assurance to the true significance of a genomic region for a trait of interest (Table 3.8). Among the five seed quality traits, peak QTL for OIL occurred within the greatest number of consensus groups, identified across chromosomes A01, A02, A04, A07, A09, C01, C05, C06, and C09. Peak QTL for ERU occurred within five consensus groups identified on chromosomes A01, A02, A09, A10, and C06. Across all FarmCPU analyses, LLN and OLE exhibited the greatest number of total significant SNPs and peak QTL. Peak QTL for LLN and OLE occurred in 9 and 7 consensus groups, respectively, whereas LLE exhibited peak QTL in only 3 consensus groups. Large proportions of phenotypic variance in oil and fatty acid traits could be explained by peak QTL effects. Peak QTL identified by FarmCPU explained as much as 33.48 % of variance for OIL, 8.08 % for ERU, 40.10 % for OLE, 68.86 % for LLE, and 27.17 % for LLN.

Furthermore, peak QTL were detected within the same genetic region for multiple seed quality traits. For example, on chromosome A07, three peak QTL for OLE, LLN, and OIL occur less than 0.5 Mb apart, while on chromosome C04, three peak QTL for OLE, LLE, and ERU occur less than 0.4 Mb apart. There are twenty of these regions where peak QTL for two or more traits are identified. These suggest fatty acid biosynthesis genes that contribute to the phenotypic expression of multiple seed quality traits.

Table 3.7 Summary of quantitative trait loci (QTL) for five seed quality traits in *Brassica napus* L. identified by FarmCPU GWAS analyses. Two populations were input to FarmCPU: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to FarmCPU: MS1 (22,941 markers) and MS2 (14,699 markers).

Trait	Total significant SNPs	# peak QTL	# peak QTL in consensus groups
ERU[1]	23	16	5
LLE[2]	23	15	3
LLN[3]	31	22	9
OLE[4]	31	22	7
OIL[5]	25	14	10
Total	133	89	34

[1] Erucic acid content (% of total fatty acid)
[2] Linoleic acid content (% of total fatty acid)
[3] Linolenic acid content (% of total fatty acid)
[4] Oleic acid content (% of total fatty acid)
[5] Overall oil content (% of dry mass)

Table 3.8 Peak quantitative trait loci (QTL) for five seed quality traits in *Brassica napus* L. identified by FarmCPU GWAS analyses. Two populations were input to FarmCPU: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two marker sets were input to FarmCPU: MS1 (22,941 SNP markers) and MS2 (14,699 SNP markers). Peak QTL denote the most significant markers identified within genomic regions across all FarmCPU analyses.

QTL name	Trait	Chr**	Position (bp)	Effect	PVE %***	p-value
ERU.A02.1	Erucic	A02	10,511,413*	1.14	1.07	1.06E-08
ERU.A03.1	Erucic	A03	11,577,136*	0.69	2.60	3.48E-10
ERU.A03.2	Erucic	A03	24,409,405	0.54	2.68	2.17E-08
ERU.A04.1	Erucic	A04	8,514,267	0.75	5.26	4.47E-07
ERU.A07.1	Erucic	A07	19,335,355	-0.78	0.23	1.41E-07
ERU.A10.1	Erucic	A10	6,924,242	-0.62	1.07	6.21E-07
ERU.A10.2	Erucic	A10	16,385,236	-0.57	1.14	5.87E-08
ERU.C02.1	Erucic	C02	51,623,095	-0.78	3.95	1.09E-09
ERU.C04.1	Erucic	C04	7,564,979*	-2.23	8.08	9.56E-23
ERU.C05.1	Erucic	C05	4,643,372*	0.69	0.61	3.52E-11
ERU.C06.1	Erucic	C06	20,477,079	-0.70	1.88	1.92E-07
ERU.C07.1	Erucic	C07	55,261,373	-0.68	2.41	1.28E-09
ERU.C08.1	Erucic	C08	15,214,319	-1.27	3.44	5.09E-11
ERU.C08.2	Erucic	C08	19,681,528	0.74	2.54	5.98E-11
ERU.C08.3	Erucic	C08	33,060,340	0.55	4.41	1.15E-06
ERU.C09.1	Erucic	C09	53,083,082*	-0.68	3.58	2.47E-07
LLE.A03.1	Linoleic	A03	6,895,572	-0.43	0.90	1.91E-10
LLE.A05.1	Linoleic	A05	8,046,563	-0.48	6.24	2.10E-06
LLE.A05.2	Linoleic	A05	31,982,138*	-0.29	1.63	2.74E-18
LLE.A06.1	Linoleic	A06	10,130,705	-0.20	1.63	7.19E-07
LLE.A06.2	Linoleic	A06	34,244,475	0.19	2.10	2.31E-07
LLE.A06.3	Linoleic	A06	43,390,042*	-0.33	2.46	3.41E-16
LLE.A07.1	Linoleic	A07	14,181,792	0.18	0.78	6.46E-08
LLE.A07.2	Linoleic	A07	15,321,915	0.21	1.88	1.19E-07

LLE.C01.1	Linoleic	C01	36,436,106	-0.59	4.97	6.00E-08
LLE.C03.1	Linoleic	C03	5,356,113	-0.21	0.36	3.09E-08
LLE.C03.2	Linoleic	C03	13,966,612	0.12	0.93	1.56E-06
LLE.C03.3	Linoleic	C03	35,367,718	-0.34	14.63	2.14E-08
LLE.C04.1	Linoleic	C04	7,564,979*	1.94	68.86	1.22E-88
LLE.C06.1	Linoleic	C06	36,690,496	0.28	0.66	2.75E-07
LLE.C08.1	Linoleic	C08	30,359,844	-0.16	0.90	7.89E-08
LLN.A01.1	Linolenic	A01	9,484,245	0.36	22.09	4.24E-07
LLN.A02.1	Linolenic	A02	2,756,523	-0.17	9.45	2.08E-06
LLN.A02.2	Linolenic	A02	28,656,592*	0.31	7.53	4.16E-09
LLN.A03.1	Linolenic	A03	6,518,274	0.37	27.17	9.87E-14
LLN.A03.2	Linolenic	A03	16,008,448	-0.42	8.22	5.53E-08
LLN.A04.1	Linolenic	A04	19,850,117	0.18	3.93	4.34E-09
LLN.A05.1	Linolenic	A05	2,848,661*	0.14	0.89	3.52E-07
LLN.A05.2	Linolenic	A05	7,273,735	-0.25	2.56	1.92E-06
LLN.A05.3	Linolenic	A05	34,635,187*	-0.38	8.67	4.40E-13
LLN.A06.1	Linolenic	A06	1,225,742*	0.47	0.00	8.46E-12
LLN.A07.1	Linolenic	A07	854,890	-0.12	7.19	9.93E-08
LLN.A07.2	Linolenic	A07	12,247,869	0.20	3.23	5.02E-08
LLN.A10.1	Linolenic	A10	15,889,091	0.18	3.34	2.02E-13
LLN.C01.1	Linolenic	C01	10,249,055*	0.14	3.78	4.34E-08
LLN.C01.2	Linolenic	C01	24,509,631*	0.31	12.63	4.22E-24
LLN.C02.1	Linolenic	C02	21,767,429*	-0.22	13.69	1.20E-11
LLN.C02.2	Linolenic	C02	47,180,884*	0.20	9.78	1.15E-12
LLN.C03.1	Linolenic	C03	9,530,072	-0.24	9.47	4.44E-11
LLN.C03.2	Linolenic	C03	38,347,210	0.15	8.14	1.09E-09
LLN.C04.1	Linolenic	C04	55,668,135	0.35	21.34	7.80E-08
LLN.C07.1	Linolenic	C07	53,007,499*	0.17	1.73	1.80E-08
LLN.C09.1	Linolenic	C09	59,191,666	-0.42	6.68	1.25E-07
OLE.A01.1	Oleic	A01	1,004,440	-0.30	1.27	5.50E-10
OLE.A01.2	Oleic	A01	10,018,954*	0.37	1.63	6.67E-11

OLE.A02.1	Oleic	A02	29,476,802	-0.24	0.16	2.13E-06
OLE.A04.1	Oleic	A04	3,839,457	0.42	2.09	2.16E-08
OLE.A05.1	Oleic	A05	35,012,221	-0.26	0.37	2.75E-06
OLE.A06.1	Oleic	A06	32,128,133	-0.32	0.72	2.94E-10
OLE.A06.2	Oleic	A06	36,035,495*	1.77	40.10	1.76E-15
OLE.A06.3	Oleic	A06	40,428,199	-0.62	7.44	6.55E-07
OLE.A07.1	Oleic	A07	854,890*	0.51	4.65	5.74E-13
OLE.A07.2	Oleic	A07	3,499,153	-0.43	1.01	8.14E-09
OLE.A07.3	Oleic	A07	9,825,427	-0.65	6.73	3.16E-06
OLE.A09.1	Oleic	A09	48,422,971	-0.25	0.09	9.41E-08
OLE.C02.1	Oleic	C02	24,890,286*	0.52	3.21	1.86E-14
OLE.C03.1	Oleic	C03	3,503,806*	0.55	2.81	7.25E-13
OLE.C04.1	Oleic	C04	7,925,988	0.32	0.47	1.75E-09
OLE.C04.2	Oleic	C04	55,668,135	-0.64	4.04	3.91E-07
OLE.C05.1	Oleic	C05	47,351	0.58	2.20	5.55E-07
OLE.C06.1	Oleic	C06	20,278,233	0.30	1.97	2.13E-06
OLE.C07.1	Oleic	C07	54,467,793	-0.42	1.38	5.35E-12
OLE.C08.1	Oleic	C08	15,214,319*	0.64	10.68	1.38E-08
OLE.C08.2	Oleic	C08	19,698,597*	0.35	1.03	3.59E-08
OLE.C09.1	Oleic	C09	52,715,903	0.56	4.52	1.79E-09
OIL.A01.1	Oil	A01	6,284,652*	-0.41	2.43	1.76E-09
OIL.A01.2	Oil	A01	13,219,631*	0.71	17.41	8.41E-07
OIL.A02.1	Oil	A02	7,685,955*	-0.65	0.56	5.64E-08
OIL.A04.1	Oil	A04	7,761,689*	-0.88	2.10	3.13E-13
OIL.A06.1	Oil	A06	3,278,513	-0.43	1.13	2.65E-06
OIL.A07.1	Oil	A07	1,274,336*	0.41	0.82	2.30E-07
OIL.A07.2	Oil	A07	8,815,064	0.31	0.87	7.79E-07
OIL.A09.1	Oil	A09	1,353,329*	-2.21	33.48	6.89E-13
OIL.C01.1	Oil	C01	44,790,053*	-0.55	2.29	4.94E-08
OIL.C04.1	Oil	C04	48,978,218	-0.75	1.72	1.11E-07
OIL.C05.1	Oil	C05	54,732,137*	0.73	3.70	4.68E-09

OIL.C06.1	Oil	C06	6,241,025*	0.67	4.82	1.92E-07
OIL.C08.1	Oil	C08	7,028,186	-0.58	6.12	2.66E-06
OIL.C09.1	Oil	C09	54,428,974*	0.56	7.61	6.95E-11

* Peak QTL found within consensus group

** Chromosome

*** % Phenotypic variance explained

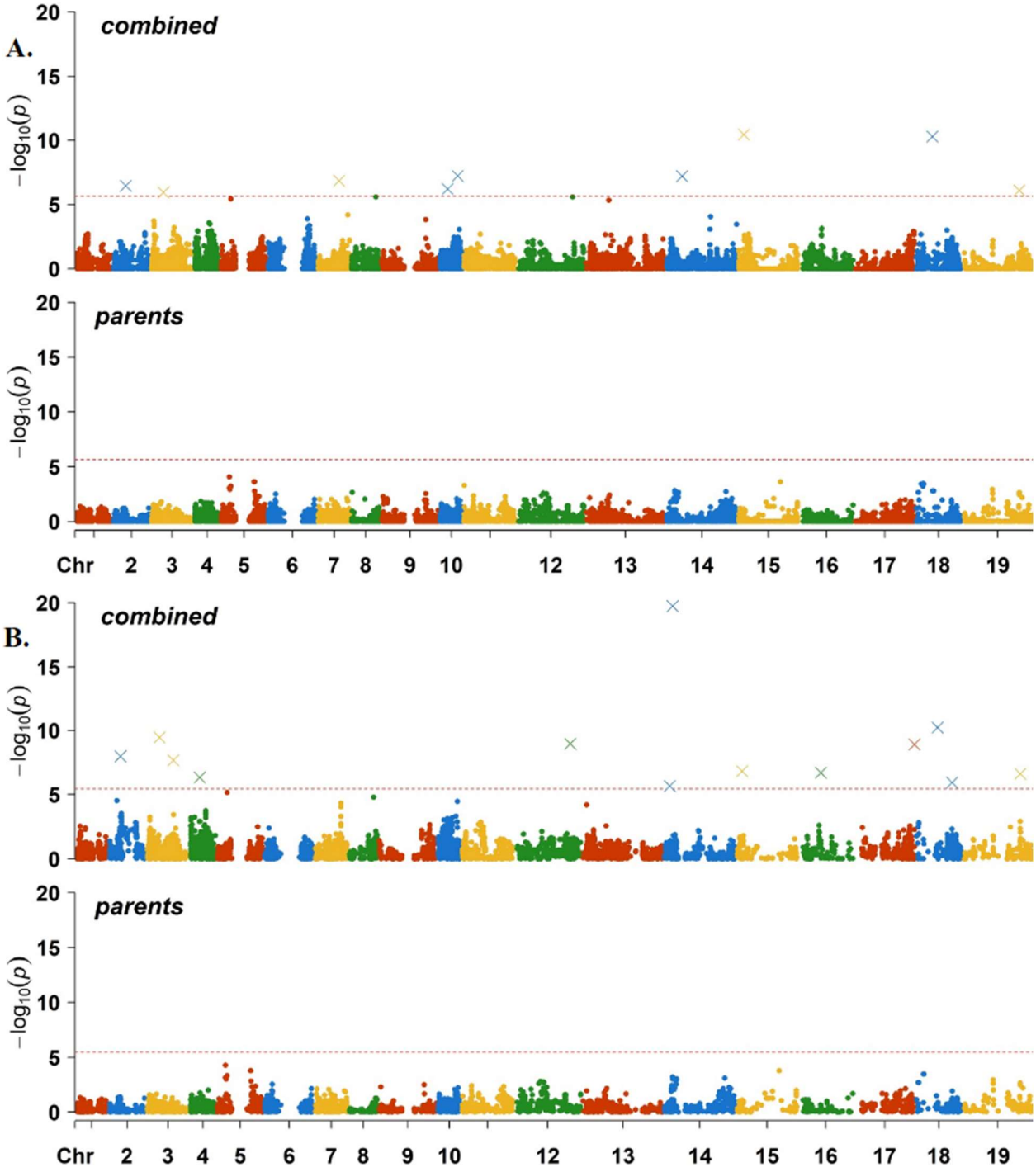


Figure 3.5 Manhattan plots produced by FarmCPU GWAS analyses for erucic acid content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

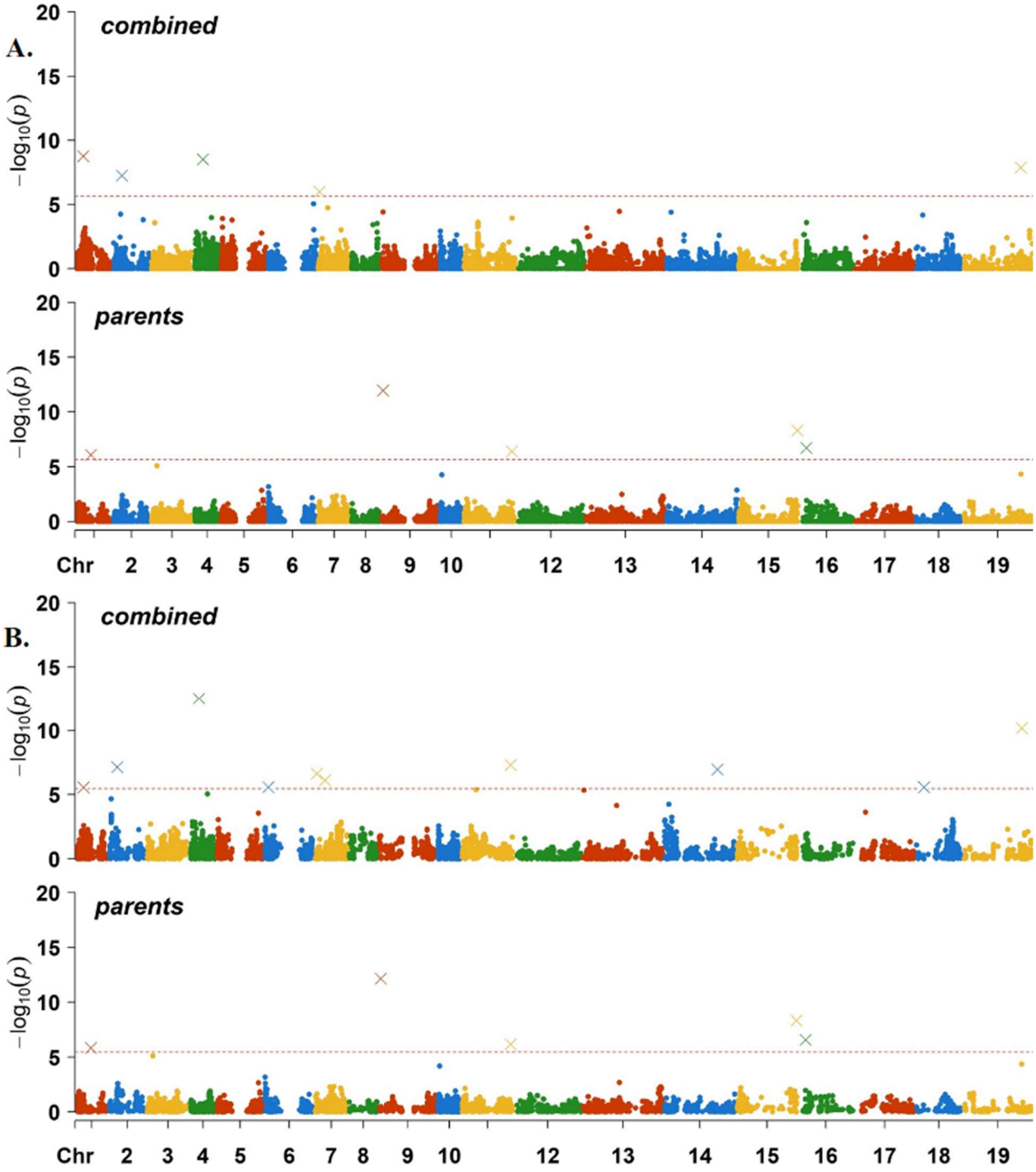


Figure 3.6 Manhattan plots produced by FarmCPU GWAS analyses for oil content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

An obvious disparity can be observed when comparing GWAS results from the two populations. Of 133 total significant SNPs identified by FarmCPU, 28 and 105 of them were identified in the parent and combined populations, respectively. This disparity contributed to only 5 of 34 consensus groups containing peak QTL identified in both parent and combined populations.

Of those 133 significant SNPs, 62 and 71 of them were identified using MS1 and MS2, respectively. Consensus between the two marker sets was prevalent: 33 of 34 consensus groups contained QTL identified by both MS1 and MS2. Of the 36 peak QTL identified using MS1, 30 of them were SNP markers whose MAF > 0.05 (that is, these 30 markers passed the MAF filter and were also present in MS2). Two of the peak QTL whose MAF < 0.05 were still in consensus with a peak QTL identified using MS2. The remaining four peak QTL identified by MS1 (ERU.A10.2, LLN.A03.1, LLE.A03.1, and LLE.C03.3) exhibit MAF < 0.05 and do not signal a region in consensus with another FarmCPU GWAS analysis.

3.4.6 Model comparison

By inputting two marker sets (MS1, MS2), two population designs (parents, combined), and six GWAS models, 24 unique analyses were conducted to identify associations between SNP markers and five seed quality traits. Different GWAS analyses were compared through Q-Q plots.

The Q-Q plots from all FarmCPU GWAS analyses of ERU and OIL are available in Figures 3.7 and 3.8, respectively (see Figures S3.6 – S3.10 for all Q-Q plots). These plots depict varying control of false positives and false negatives. The most notable trend is that, in nearly all analyses upon the combined population, the CMLM, MLM+K, and MLM+K+PCA models consistently deviate early from the expected. The FarmCPU, MLMM, and GenABEL P+K models exhibited consistent control of false positives across analyses of the combined population. Regarding GWAS analyses upon the parent population, most models exhibited no noticeable early deviations above the expected, however, some models (especially MLM+K and MLM+K+PCA) instead depicted trends below the expected, indicating false negatives.

Comparing Q-Q plots between marker sets, the pattern of early deviation of observed from expected is not affected by the use of MS1 as opposed to MS2. That is, analyses which tended to exhibit early deviations (e.g. CMLM x combined population) and analyses which did not (e.g. FarmCPU x combined population) exhibit those patterns regardless of marker set. The most

notable difference between marker sets is that far more significant SNPs were identified in MS1 than MS2. This can most likely be attributed to the high density of MS1, i.e. a causal locus is likely to have more markers in linkage than a lower-density marker set, resulting in more significant markers per causal locus.

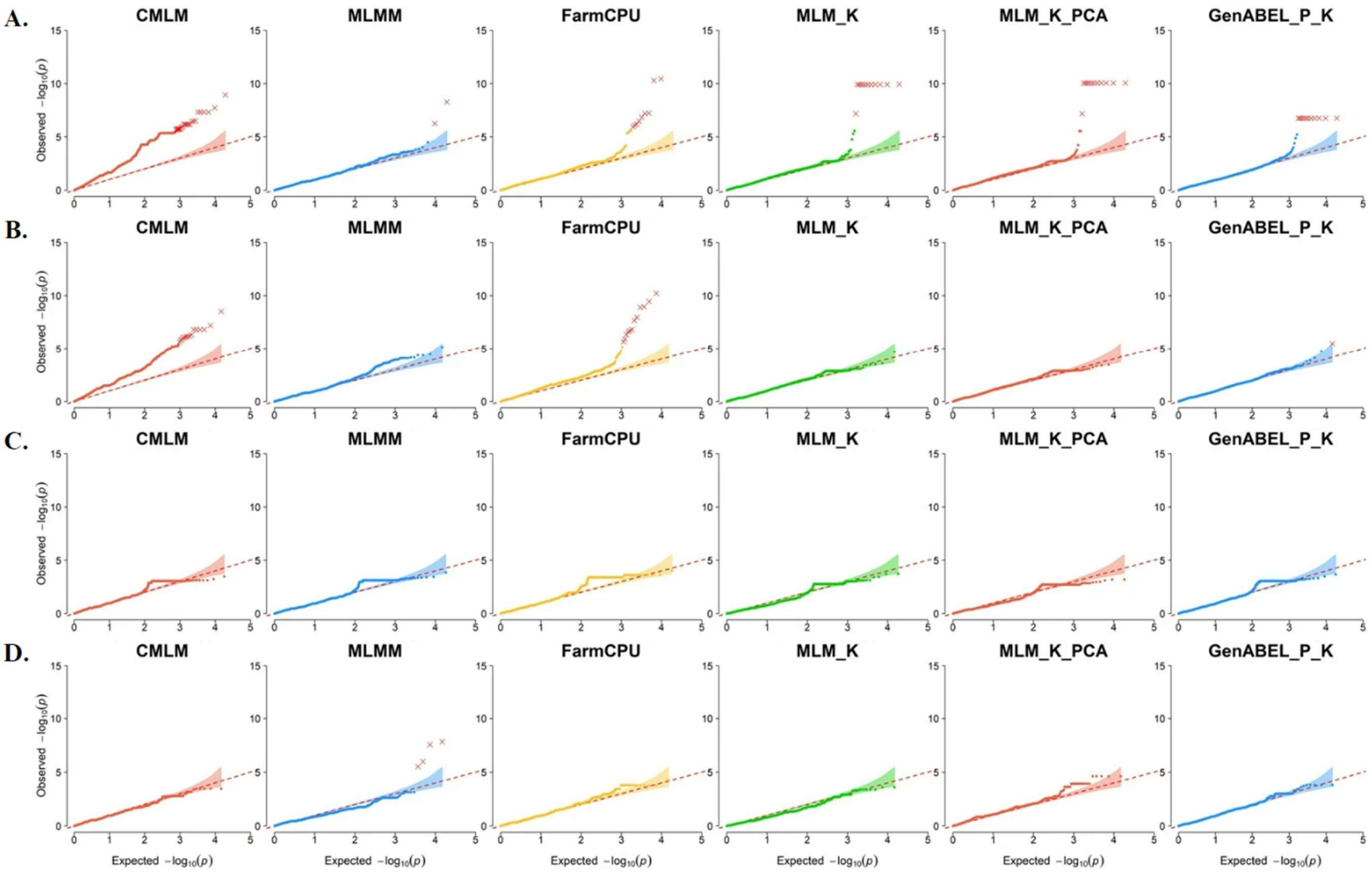
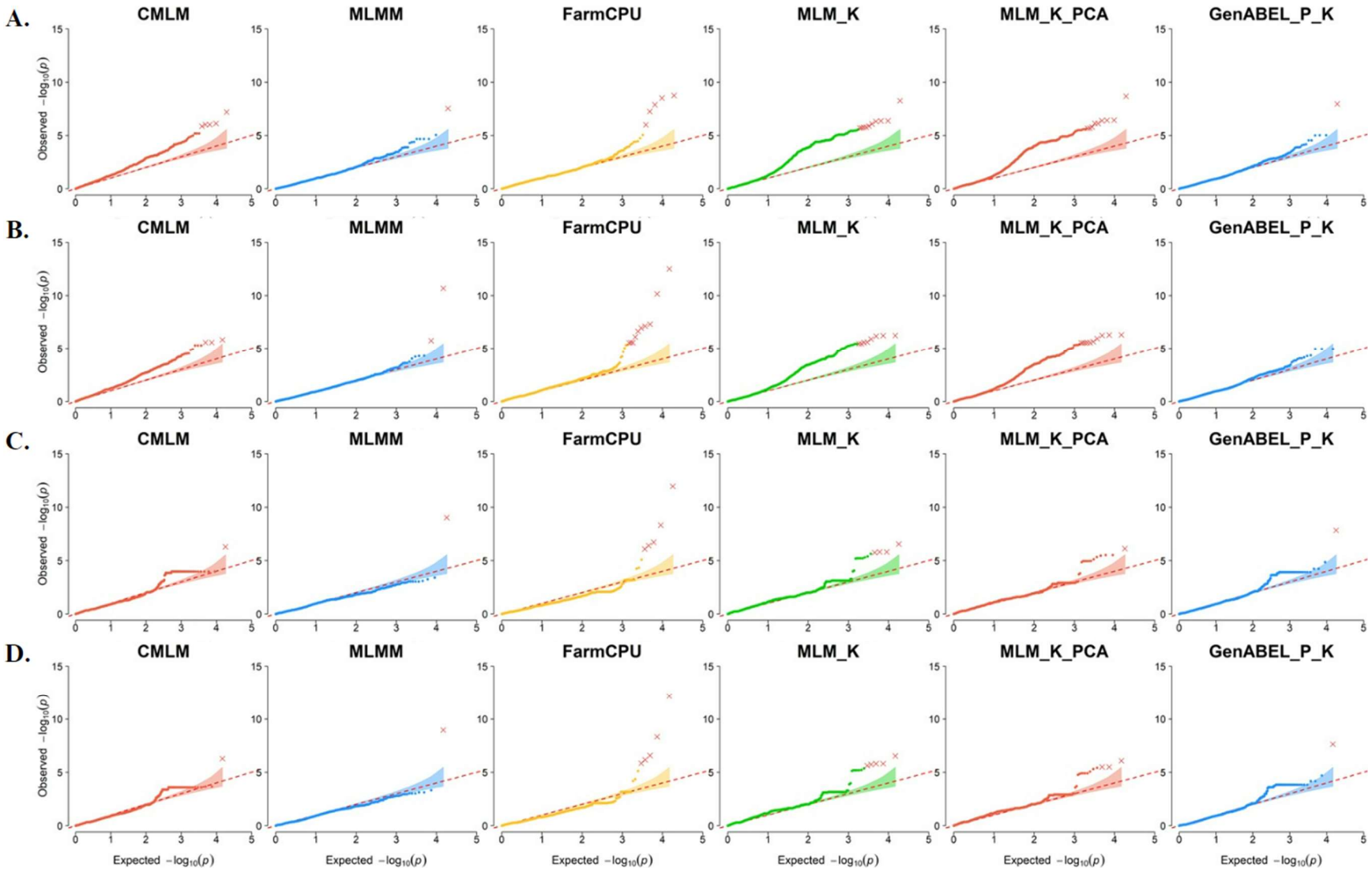


Figure 3.7 Q-Q plots of GWAS results for erucic acid. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.



2 Figure 3.8 Q-Q plots of GWAS results for oil content. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.

3.5 Discussion

As the most widely grown crop species in Canada and the second-largest oilseed commodity in the world, the improvement of *B. napus* oil quality remains a crucial objective for plant breeders (Statistics Canada 2024; USDA 2023). The improvement of *B. napus* has continuously adapted to rising populations and diversity of consumer demand, marked by the introduction of specialty cultivars for improved shelf life, heat stability, and functional foods (high LLE:LLN ratio), for higher-value meal in food and feed (increased protein storage), and for biofuels or other industrial uses (high erucic acid) (Abbadi and Leckband 2011; Fitzpatrick and Scarth 1998; Nesi et al. 2008; Wittkop et al. 2009). The *Brassica* 60K Illumina Infinium^{TM} SNP array used in this study is an example of advancements in biotechnology which facilitate plant breeders' response to shifting demands (Khan et al. 2021; Thomson 2014). The *Brassica* 60K has served as an extremely effective, high-throughput, and low-cost genotyping method assisting in the discovery of markers and candidate genes for important traits throughout an extensive catalogue of studies (Clarke et al. 2016; Kumar et al. 2017; Liu et al. 2022). In this study, all combinations of marker set, population design, and GWAS model produced 24 unique analyses for each of the five seed quality traits.

Within this *B. napus* breeding population, seed quality traits exhibited strong relationships among themselves (Figure 3.2). We detected significant negative correlations between OLE, ERU, and LLE, demonstrating their direct association in fatty acid biosynthesis which has been observed in other *B. napus* studies (Katavic et al. 2001; Sasongko et al. 2005; Tang et al. 2019). Strong negative correlations between OLE and ERU indicate that genes which contribute to high ERU function at the expense of desaturation to LLE and LLN (Sharafi et al. 2015). Interestingly, OIL was not significantly correlated to proportions of any particular fatty acid in this study. That is, although these four fatty acids constitute most of the fatty acid profile, overall OIL accumulation appears more complex than the biosynthetic pathways which control the proportions of its main fatty acids – most likely attributed to other minor-effect genes and genotype × environment interactions, while other studies note significant correlations between OIL and particular fatty acids (Delourme et al. 2006; Jiang et al. 2014; Sharafi et al. 2015; Tang et al. 2019; Zhao et al. 2007). Later in this discussion, we also identify a candidate gene known to alter the fatty acid profile as part of its response to salt stress. Such relationships among traits identified in this study represent important

considerations for *B. napus* breeders aiming to minimize linkage drag and introgression of unwanted traits (e.g. abiotic stress intolerance) from distant accessions.

Our genotypic data quality control produced a total of 22,941 SNP markers with more than 70 % coverage, 0 gaps, and 1 or less mismatches. These 22,941 SNPs, which constitute MS1, provide an overall mean of 37.8 kb between each marker (Table 3.6). Many GWAS studies filter markers whose minor allele frequency (MAF) is below 5% due to the supposedly low statistical power of such variants (Cortes et al. 2021; Turner et al. 2011b; Uffelmann et al. 2021). Filtering MS1 on the basis of MAF produced MS2, providing an overall mean of 59.0 kb between each marker. Required marker densities for GWAS are often determined by LD decay: the distance at which LD between loci declines below a certain threshold (typically when $r^2 < 0.2$) (Alqudah et al. 2020). The *Brassica* 60K array generates marker coverage above the required density indicated by *B. napus'* overall LD decay rate, however LD decay has been shown to vary dramatically across the genome (Liu et al. 2016a; Qian et al. 2014; Wang et al. 2014). In this study, LD decay was significantly lower in the A-subgenome than the C-subgenome (1,827 kb and 9,648 kb, respectively), consistent with observations of LD decay in other *B. napus* studies (Liu et al. 2016a; Qian et al. 2014; Wang et al. 2014). Contrasting LD decays between A- and C-subgenomes in this population reflect some of the main hypotheses driving these LD patterns such as the large selective bottlenecks during early adaptation of canola-quality *B. napus* that produced extensive conserved haplotype blocks, interspecific hybridizations between *B. rapa* and *B. napus* giving rise to genotypes exhibiting elite agronomic and quality traits with greater recombination and shorter haplotype blocks in the A-subgenome, and the C-subgenome's high content of transposable elements which are known to exhibit lower recombination rates (Chalhoub et al. 2014; Liu et al. 2016a; Qian et al. 2014; Wang et al. 2014). The lowest density marker set developed in this study, MS3, contains only 1,098 markers, however was filtered to minimal marker density within scanning windows while maintaining LD above the $r^2 = 0.2$ threshold. Therefore, each of the increasingly dense marker sets (MS2 and MS1) provide sufficient marker coverage for GWAS.

The ability of GWAS to detect significant associations between traits and genotypes relies largely on the allelic variation within the population of interest. Allelic variation is incorporated as population size grows, with the required power of a GWAS experiment determining the population size threshold above which true significant associations are likely to be detected (Purcell et al.

2003; Uffelmann et al. 2021; Visscher et al. 2012). In addition to size, GWAS populations are designed to incorporate diversity to avoid false positive or negative signals attributed to population structure (Uffelmann et al. 2021). In the present study, no distinct clusters were formed based on the type of individuals (i.e. B-lines, R-lines or hybrids), indicating no clear population stratification while hybrid genotypes were evenly distributed among parental genotypes (Figure 3.4). The effects of population design on QTL discovery are exemplified in this study by comparing a small "parent" population (91 inbred parent genotypes) to the large and diverse "combined" population (427 genotypes, combining the parents and their 336 hybrid progeny). By including hybrid genotypes, the larger and more diverse population offers greater power to detect significant associations, as reflected in the results: of 133 total significant SNPs detected across FarmCPU GWAS analyses, 105 (79 %) were identified when inputting the combined population.

Peak QTL which directly coincide (occur within confidence intervals based on chromosome-specific LD decay distance) with other previously identified QTL for the same traits were identified across seven chromosomes in the A-subgenome and three chromosomes in the C-subgenome (Table 3.9). Furthermore, 30 previously identified candidate genes involved in oil and fatty acid biosynthesis coincide closely with significant loci detected by FarmCPU in this study (Table 3.10). On chromosome C08, one marker (Bn-scaff_19242_1-p390109) induced the second-greatest effect on ERU and a significant effect on OLE, simultaneously. In this same region of chromosome C08, two independent studies identified QTL which were simultaneously associated to both ERU and OLE (Qu et al. 2017; Zhu et al. 2019). Zhu et al. (2019) indicated enoyl-CoA hydratase domain-containing protein 3, mitochondrial-like (BnaC08g09570D) (controlling CoA isomerase involved in unsaturated fatty acid degradation) as a candidate gene in this region (TAIR 2013). On chromosome A07, we identified a QTL for LLN which closely coincides with a QTL identified by Körber et al. (2019) for the same trait. Körber et al. (2019) predicted phospholipase A2-delta (BnaA07g01090D) as the candidate gene for this region. Around the 3 Mb region of chromosome A09, in close proximity to our largest-effect QTL for OIL, two studies also identified QTL (Qu et al. 2017; Wang et al. 2018a). These two studies subsequently both identified acyl-CoA oxidase 2 (BnaA09g07080D) as a candidate gene (Qu et al. 2017; Wang et al. 2018a). ACP5 (acyl carrier protein 5) also lies in close proximity to OIL.A09.1 – the ACP gene family is important to fatty acid chain elongation, but the overexpression of ACP5 in particular has been shown to improve salt stress tolerance at the expense of altered fatty acid profile (Huang et al.

2017). One of the most significant associations in this study, LLN.C01.2, is present near WSD1 (O-acyltransferase family protein), encoding a bifunctional enzyme that catalyzes the synthesis of both wax esters and triacylglycerols (TAGs). WSD1 was identified as a candidate gene in *B. napus* fatty acid biosynthesis by Qu et al. (2017) and exhibits orthologous forms in *Arabidopsis thaliana*, peanuts (*Arachis hypogaea*), and oleaginous yeast (Li et al. 2008; Rani et al. 2013; Saha et al. 2006; Turchetto-Zolet et al. 2016). Finally, the most significant QTL detected in this study, LLE.C04.1, was located on chromosome C04 in close proximity to highly significant QTL for two other traits: OLE and ERU (ERU.C04.1 was also the third-most significant QTL detected in this study). This region likely represents a single candidate gene which is very important in the biosynthesis of all three fatty acids listed. To our knowledge, a candidate gene for fatty acid biosynthesis has not yet been identified in this region of the *B. napus* C04 chromosome. This region and many others described above represent significant contributors to seed quality traits. These loci warrant further investigation towards potential applications in *B. napus* improvement. By coinciding significant regions with literature, the QTL described in Tables 3.9 and 3.10 provide assurance in the method's ability to identify loci which are associated to seed quality phenotypes.

As a note, considering this study analyzed a HEAR breeding population, the major genes controlling erucic acid production (two FAE1 copies located on A08 and C03) were fixed. The absence of variation within these loci resulted in the absence of significant associations for ERU on chromosomes A08 and C03.

The models used in this experiment can be distinguished by either single-locus GWAS (SL-GWAS) (CMLM, MLM+K, MLM+K+PCA, and GenABEL P+K) or multi-locus GWAS (ML-GWAS) (MLMM and FarmCPU). Comparing Q-Q plots amongst GWAS models, we observed patterns of early deviations above the expected (often attributed to false positives from systematic biases present within GWAS analyses) and below the expected (attributed to false negatives from models' over-correction for multiple testing) (Ehret 2010; Kaler et al. 2020). It was the population design which most clearly distinguished model performance (not marker set) since models that tended to exhibit early deviations (e.g. CMLM × combined population) and analyses which did not (e.g. FarmCPU × combined population) exhibited those patterns regardless of marker set. In both MS1 and MS2, the Q-Q plots for CMLM, MLM+K, and MLM+K+PCA demonstrate patterns of early deviations from the expected (the fourth SL-GWAS model, GenABEL P+K, did not display such

Table 3.9 Peak quantitative trait loci (QTL) identified in this study that coincide with previously identified QTL for the same traits in *Brassica napus* L.

Trait	Chr*	QTL name	QTL position	Reference QTL position	Reference
ERU[1]	A10	ERU.A10.2	16,385,236	15,381,881	Qu et al. (2017)
ERU	C08	ERU.C08.1	15,214,319	14,268,490	Qu et al. (2017)
ERU	C08	ERU.C08.1	15,214,319	14,472,595	Zhu et al. (2019)
LLE[2]	A06	LLE.A06.1	10,130,705	11,711,750	Zhu et al. (2019)
LLN[3]	A02	LLN.A02.1	2,756,523	3,541,030	Qu et al. (2017)
LLN	A05	LLN.A05.2	7,273,735	7,115,874	Qu et al. (2017)
LLN	A07	LLN.A07.1	854,890	278,027	Körber et al. (2016)
LLN	A10	LLN.A10.1	15,889,091	14,591,650	Zhu et al. (2019)
LLN	C02	LLN.C02.2	47,180,884	44,856,112	Körber et al. (2016)
OLE[4]	A01	OLE.A01.1	1,004,440	2,717,777	Körber et al. (2016)
OLE	C02	OLE.C02.1	24,890,286	27,031,784	Qu et al. (2017)
OLE	C08	OLE.C08.1	15,214,319	14,268,490	Qu et al. (2017)
OLE	C08	OLE.C08.1	15,214,319	14,472,595	Zhu et al. (2019)
OIL[5]	A01	OIL.A01.1	6,284,652	6,244,998	Xiao et al. (2019)
OIL	A02	OIL.A02.1	7,685,955	9,446,115	Sun et al. (2016)
OIL	A09	OIL.A09.1	1,353,329	2,850,145	Wang et al. (2018a)
OIL	C06	OIL.C06.1	6,241,025	7,620,296	Tang et al. (2021)
OIL	C08	OIL.C08.1	7,028,186	7,894,416	Zhao et al. (2022)

[1] Erucic acid content (% of total fatty acid)
[2] Linoleic acid content (% of total fatty acid)
[3] Linolenic acid content (% of total fatty acid)
[4] Oleic acid content (% of total fatty acid)
[5] Overall oil content (% of dry mass)
* Chromosome

Table 3.10 Candidate genes for fatty acid biosynthesis identified in previous *Brassica napus* L. literature which coincide with peak quantitative trait loci (QTL) regions identified in this study.

Chr	Reference's associated QTL position	Gene ID	Description	Reference
A01	6,244,998	BnaA01g12350D	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	Xiao et al. (2019)
A02	524,995	BnaA02g00830D	SHN Transcription Factors	Wang et al. (2018a)
A02	3,668,843	BnaA02g07780D	Uric acid degradation bifunctional protein TTL	Tang et al. (2021)
A02	6,411,341	BnaA02g12050D	Acyl carrier protein	Yao et al. (2020)
A02	7,310,000	BnaA02g13310D	Beta-ketoacyl reductase 1 (KCR1)	Qu et al. (2017)
A03	15,727,479	BnaA03g32450D	Dienoyl-CoA Reductase	Wang et al. (2018a)
A04	3,064,696	BnaA04g04400D	Inositolphosphorylceramide synthase	Wang et al. (2018a)
A04	3,064,696	BnaA04g04410D	Immunophilin-like FK506 binding protein	Wang et al. (2018a)
A04	3,064,696	BnaA04g04450D	Lipid transfer protein	Wang et al. (2018a)
A07	278,027	BnaA07g01090D	Phospholipase A2-delta	Körber et al. (2016)
A07	23,923,367	BnaA07g35680D	Cyclin-A2-4	Tang et al. (2021)
A09	1,752,479	BnaA09g03610D	Acyl carrier protein 5, chloroplastic	Zhu et al. (2019)
A09	2,539,185	BnaA09g04980D	Probable carboxylesterase 18	Zhu et al. (2019)

A09	2,539,185	BnaA09g05270D	Zeaxanthin epoxidase, chloroplastic	Zhu et al. (2019)
A09	2,640,000	BnaA09g05410D	TRANSPARENT TESTA16 (TT16)	Qu et al. (2017)
A09	2,850,145	BnaA09g06090D	MYB transcription factors	Wang et al. (2018a)
A09	2,949,845	BnaAnng14390D	Fatty acyl-CoA reductase	Körber et al. (2016)
A09	3,030,000	BnaA09g06170D	Holocarboxylase synthase 1 (HCS1)	Qu et al. (2017)
A09	3,329,514	BnaA09g07080D	Acyl-CoA oxidase 2 (ACX2)	Wang et al. (2018a)
A09	3,460,000	BnaA09g07080D	Acyl-CoA oxidase 2 (ACX2)	Qu et al. (2017)
A10	6,648,809	BnaA10g08110D	Putative U-box domain-containing protein 53	Tang et al. (2021)
A10	14,591,650	BnaA10g21150D	Zeaxanthin epoxidase, chloroplastic	Zhu et al. (2019)
A10	15,598,961	BnaA10g23790D	Fatty acid amide hydrolase	Wang et al. (2018a)
C01	24,624,245	BnaC01g25960D	O-acyltransferase (WSD1-like) family protein	Qu et al. (2017)
C01	24,624,245	BnaC01g26460D	Fatty acid biosynthesis 1 (FAB1)	Qu et al. (2017)
C01	24,624,245	BnaC01g26600D	HXXXD-type acyl-transferase family protein	Qu et al. (2017)
C05	5,886,965	BnaC05g10520D	Sphingoid base hydroxylase 2	Xiao et al. (2019)
C06	7,620,296	BnaC06g07070D	Protein translation factor SUI1 homolog 2	Tang et al. (2021)

C06	7,620,296	BnaC06g07070D	Protein translation factor SUI1 homolog 2	Tang et al. (2021)
C08	14,472,595	BnaC08g09570D	Enoyl-CoA hydratase domain-containing protein 3, mitochondrial-like	Zhu et al. (2019)

* Chromosome

distinct patterns) (Figures 3.7 and 3.8). In the parent population, most models either exhibited no early deviations from the expected followed by a small upwards tail indicating true positives (FarmCPU, MLMM, and GenABEL P+K) while other models exhibited trends below the expected without significant signals, demonstrating model overcorrection and false negatives (especially MLM+K and MLM+K+PCA). The combined population instead highlighted consistent early deviations above the expected, indicating false positives in nearly all SL-GWAS analyses. Meanwhile, FarmCPU, MLMM, and GenABEL P+K exhibited relatively consistent control of false positives in the combined population. Regarding false negatives, SL-GWAS MLM models (such as CMLM, MLM+K, and MLM+K+PCA) are known to perform overly stringent corrections in some cases, failing to detect the significance of truly significant loci (Khan et al. 2021). One possible explanation for the prevalence of false positives by SL-GWAS may be *B. napus*' slow LD decay rate: approaching marker-trait associations by calculating one marker at a time may falsely detect significant associations to the hundreds of genes located within extremely large haplotype blocks. (Kaler et al. 2020; Khan et al. 2021; Liu et al. 2022).

In contrast, the consistent control of false positives and false negatives demonstrated by FarmCPU in both populations aligns with the findings of previous studies, prompting us to utilize FarmCPU's results for QTL reporting (Kaler et al. 2020; Liu et al. 2016b; Pal et al. 2021). In a study investigating QTL for oil content in *B. napus*, Wang et al. (2013) also observed improved QTL identification and more accurate QTL effect estimations when using larger, more diverse populations like the combined population in this study. In an actual hybrid breeding program, even if an appropriately large parent population could be assembled, a large hybrid population remains more appropriate and powerful for GWAS in hybrid cultivar research given that GWAS on inbred parents does not reveal the hidden variables contributing to the heterosis of their progeny (Wang et al. 2017b). Several hybrid population designs exist which offer plant breeders great flexibility and effectiveness for hybrid cultivar development, adding to the growing toolbox of population designs to improve QTL discovery and lessen the spurious effects of population structure (Dell'Acqua et al. 2015; Liu et al. 2022; Wang et al. 2017b; Xiao et al. 2017). FarmCPU's superiority is emphasized especially when comparing pooled results across all 120 GWAS analyses. Despite making up only 133 of the 1400 significant SNPs overall, FarmCPU constituted nearly half (78/161) of the peak QTL detected across all analyses (Tables S3.7 - S3.9) (Bento 2024b). This demonstrates FarmCPU's ability to identify singular and highly significant SNP closest to the

causal locus (Kaler et al. 2020). FarmCPU's outperformance of other models can likely be attributed to its multi-locus (ML-GWAS) approach, simultaneously incorporating multiple markers as covariates in a modified MLMM which iteratively conducts fixed effect and random effect correction to remove the confounding between markers and kinship (Kaler et al. 2020; Liu et al. 2016b). The other ML-GWAS model in this study, MLMM, also showed excellent control of false positives and identified 23 significant QTL, however represented only 7 of the peak QTL when compared in the pool of results from all GWAS analyses.

Marker density exhibits less of an effect on QTL discovery than population design (Liu et al. 2016a; Long and Langley 1999; Zhu et al. 2008). Of 133 total significant SNPs identified by FarmCPU, 62 and 71 of them were identified using MS1 and MS2, respectively. Considering MS1's marker coverage was around double that of MS2 for each chromosome, it was expected that far more significant SNPs are identified in MS1 as a consequence of additional markers in linkage to any causal locus than a lower-density marker set. The discussion on marker set density's effect on GWAS intersects here with the choice of GWAS model. Four of the six GWAS models fulfilled the above expectation regarding marker coverage: when using MS1 and MS2, respectively, 812 and 432 significant SNPs were detected for the five traits across all GWAS analyses by SL-GWAS models (CMLM, MLM+K, MLM+K+PCA, and GenABEL P+K). The Q-Q plots for CMLM, MLM+K, and MLM+K+PCA especially demonstrate a pattern of false positive signals and several false negative signals (Figures 3.7 and 3.8). The prevalence of false positives in SL-GWAS cannot simply be attributed to marker density nor rare variants when considering the overwhelming proportions of markers which deviate above the expected in both MS2 (intermediate density, containing no markers with MAF < 0.05) and MS1 (high density, containing markers with MAF < 0.05). If rare variants were responsible for this behaviour, then such patterns would only be visible in MS1 analyses. Similarly, early deviations due to extreme marker density would only appear in MS1 analyses, but they are also prevalent in MS2 analyses.

Where ML-GWAS models are promising alternatives that provide the power to investigate complex trait genetic architecture and flexible selection criteria to reduce false positive and negative signals, SL-GWAS models also remain firmly established in GWAS research with proven effectiveness in detecting major associations and controlling population structure bias using less computationally demanding methods (Alseekh et al. 2021; Khan et al. 2019; Wang et al. 2016;

Wen et al. 2018). By integrating both approaches, *B. napus* improvement may leverage the strengths of SL-GWAS and ML-GWAS to enhance the reliability of GWAS results, thereby offering more comprehensive understandings of genetic architecture (Khan et al. 2021).

3.6 Conclusion

This study identified 89 peak QTL for five important seed quality traits in a HEAR breeding population. Many of these peak QTL align closely with other QTL and candidate genes identified in literature. The identified loci offer potential future investigation and utilization in *B. napus* improvement. Significant effects on GWAS results were observed in population design and choice of GWAS model. Population design especially highlighted the major performance differences among GWAS models. Marker set density and rare variants did not induce significant effects on GWAS results.

4. GENOMIC SELECTION FOR OIL AND FATTY ACID PROFILE IN RAPESEED (*Brassica napus* L.) USING PARAMETRIC AND MACHINE LEARNING METHODS

4.1 Abstract

Genomic selection (GS) describes a wide range of bioinformatics methods that produce phenotypic predictions for plant breeders, potentially reducing the overall number of breeding cycles by improving selection efficiency and lessening in-field phenotyping. Considering the highly subjective nature of GS' predictive ability, this study was prompted to demonstrate GS analyses in application to *B. napus* oil and fatty acid profile traits (erucic acid (ERU), linoleic acid (LLE), linolenic acid (LLN), oleic acid (OLE), and overall oil content (OIL)). The effects of training/validation population design, marker set density, and choice of GS model on prediction accuracy were observed across 675 unique GS analyses. Prediction accuracies (given as Pearson's *r* correlation between predictions and actual phenotypes) ranged from negative values to as high as 0.89, exhibiting distinguishable trends that included prediction accuracies above 0.70 for all five traits. ERU, LLE, and LLN generally exhibited the greatest prediction accuracies, whereas OLE and OIL were generally lower. Prediction accuracies generally decreased in response to trait complexity, increased in response to training population size and degree of training/validation population relatedness, and exhibited no significant differences among marker densities or parametric models. Machine learning (ML) models performed either equivalent to or poorer than the relatively simple parametric models. The identification of highly accurate analyses for all five seed quality traits supports the utility of GS in future *Brassica* breeding programs.

4.2 Introduction

Since its origin in livestock breeding, GS has been extensively applied to a variety of crops, traits, environments, population types, and more (Budhlakoti et al. 2022; Meuwissen et al. 2001). In contrast to traditional marker-assisted selection (MAS), which informs selections based on major-effect QTL, GS estimates major- and minor-effect QTL across genome-wide marker sets to calculate phenotypic predictions known as genomic estimated breeding values (GEBVs) (Krishnappa et al. 2021). The use of GEBVs to inform selections (which would otherwise be informed by phenotypic data collected over multiple years/locations) is a major benefit to plant

breeders, facilitating the allocation of resources to desirable germplasm. By optimizing the required amount of phenotypic data collection, GS therefore improves genetic gain per unit time while reducing costs per breeding cycle and (potentially) the overall number of breeding cycles (Hickey et al. 2014). These advantages are especially important regarding complex agronomic traits that are difficult/costly to phenotype and for species that exhibit long generation times (Merrick and Carter 2021; Neyhart et al. 2017).

Cross validation procedures are often used to simulate plant breeding methodology, evaluating GS' effectiveness before practical applications. The practical GS workflow includes two populations: a training population (TP) and a test population (represented by a validation population (VP) in cross validation procedures). The TP (e.g. parent genotypes, preliminary hybrids) is sown, phenotypic and genotypic data is collected, and this data is used to "train" a GS model (estimating the effects of all markers in a genome-wide marker set) (Bernardo 2008; Budhlakoti et al. 2022). The test population's genotypic data is input to the trained GS model, producing GEBVs that facilitate more efficient selections based on genotypic data rather than long-term multi-environment field testing (Budhlakoti et al. 2022).

The effectiveness of GS methods is highly subjective to the unique factors inherent across breeding programs, producing wide variations in accuracy even among traits within same species. Given its subjectivity, the general notion regarding GS methodology is that there is no "one size fits all" (Lorenz et al. 2011). An understanding of the factors which impact GS' effectiveness within large-scale plant breeding programs is crucial to achieve increased genetic gain per unit of time without increased costs relative to conventional selection or MAS (Crossa et al. 2017; Hickey et al. 2017; Werner et al. 2020). This study focuses on four major factors that are known to affect GS prediction accuracy: trait complexity, TP/VP design, marker density, and choice of model.

Although GS is considered clearly advantageous to MAS for the improvement of complex traits, the heritability of specific traits remains a significant factor affecting GS prediction accuracy. The many minor additive and non-additive effects within complex traits makes genomic prediction difficult. In general, highly complex traits controlled by minor-effect QTL cause reduced prediction accuracy in GS models (Farooq et al. 2023; Heffner et al. 2009; Jannink et al. 2010; Kaler et al. 2022).

The accuracy of GS model marker effect estimation is strongly affected by the allelic variation provided within the TP. An ideal TP should include wide phenotypic and genotypic variation to effectively capture haplotype combinations and their effects on traits of interest (Calus 2010; Werner et al. 2020). Furthermore, GEBV accuracy is subjective to the genetic relationship between the TP and the test population (Calus 2010; Lorenz and Nice 2017). Designs which incorporate greater genetic relatedness between TP and test population enhance GS accuracy by ensuring linkage phase between markers and QTL is similar for model training and model application (Werner et al. 2020). The inclusion of such relatedness between the TP and VP has been demonstrated as a significant factor to GS prediction accuracy (Bastiaansen et al. 2012; Clark et al. 2011; Makowsky et al. 2011; Werner et al. 2020). In addition to genetic relatedness, increasing TP size more extensively captures genotypic and phenotypic variation for GS model training, generally increasing prediction accuracy (Daetwyler et al. 2008; Desta and Ortiz 2014; Edwards et al. 2019; Hickey et al. 2014; Zhang et al. 2017).

Optimal marker density for GS is largely dependent on linkage disequilibrium (LD) decay (Desta and Ortiz 2014; Kaler et al. 2022). LD refers to the non-random association of alleles at different loci within a population, while LD decay refers to the rate at which non-random association "decays" over genetic distance (i.e. approaches random segregation, unlinked) (Vos et al. 2017). As an indirect selection method, GS accuracy relies on marker linkage to causal genes. It is not simply "marker density" but instead the maximum distance/distribution of markers to causal genes which determines optimal marker coverage. For example, in a potato breeding population whose LD block size is estimated 1 Mb, the presence of markers within a maximum 500 kb span on either side of causal genes can be considered a reliable marker density (Aalborg et al. 2024; Vos et al. 2017). A plateau effect on GS prediction accuracy was demonstrated by Kaler et al. (2022) whose LD-filtered marker set contained less than half the markers of its complete marker set, yet produced statistically indistinguishable prediction accuracies for 6 traits of varying heritability in soybean. In summary, the minimum required marker density for GS training to appropriately capture causal polymorphisms is dependent on LD decay, wherein shorter LD decay distance requires greater minimum marker density and vice versa.

No singular GS model is considered universally superior (Larkin et al. 2019). Early GS models, such as linear least squares regression, have evolved into more sophisticated whole-genome

regression (WGR) models, each with unique assumptions and parameters that affect their performance (Budhlakoti et al. 2022; Emmert-Streib et al. 2020; Jubair and Domaratzki 2023). The choice of parametric or non-parametric models, machine learning or deep learning approaches, single-trait or multi-trait approaches, or other variants are guided by the goals and available resources of a breeding program. By implementing current advancements in next-generation sequencing and high-throughput phenotyping, breeders may also capture vast genotypic and phenotypic variation, optimizing resource allocation and reducing the overall number of breeding cycles (Heffner et al. 2010; Jubair and Domaratzki 2023; Krishnappa et al. 2021; Varshney et al. 2017).

For *B. napus* improvement, researchers have utilized GS to predict yield and yield-related traits such as seeds per pod, silique number, thousand kernel weight, flowering time, maturation time, shattering, and more (Fikere et al. 2020; Hu et al. 2021a; Koscielny et al. 2020; Luo et al. 2017; Werner et al. 2017). Seed-quality traits including glucosinolates, protein content, overall oil content, and specific fatty acid profile components (erucic, linoleic, linolenic, oleic, stearic) have also been predicted with GS (Fikere et al. 2020; Hu et al. 2021a; Koscielny et al. 2020; Werner et al. 2017; Werner et al. 2018; Zou et al. 2016). The effectiveness of GS for disease resistance predictions, namely sclerotinia stem rot and blackleg, has also been investigated (Fikere et al. 2018; Derbyshire et al. 2021; Roy et al. 2021).

The objective of this study is to investigate the potential effectiveness of GS across a wide range of breeding scenarios for *B. napus* oil and fatty acid profile improvement. This array of unique GS analyses is designed to demonstrate the effects of trait complexity, TP/VP design, marker density, and choice of model on GS model accuracy. We hypothesize that trait complexity, TP/VP design, and the choice of model will exhibit significant effects on GS prediction accuracy. Since the three marker sets in this study are filtered for coverage above an LD decay threshold, we hypothesize that marker set density will not play a significant role in GS prediction accuracy.

4.3 Materials and Methods

4.3.1 Plant materials, phenotypic data, and genotypic data

Details of phenotypic and genotypic data collection were described in section 3.3.1. Briefly, a population of 454 high erucic acid rapeseed (HEAR) genotypes (91 parental genotypes and 345

hybrid genotypes) were evaluated in 48 site-years across Manitoba, Saskatchewan and Alberta. Field experiments for parental genotypes were performed using RCBD with three replicates per site-year for five site-years. Hybrid genotypes were evaluated in 43 site-years, however due to the nature of selection within the breeding program where only favoured genotypes were selected for further field experiments, hybrid genotypes did not have an equal number of replicates. Unbalanced phenotypic data replicates were corrected by calculating best linear unbiased prediction (BLUP) values for all parent and hybrid genotypes. These BLUP values were used as phenotypic data inputs for this study.

DNA samples of all genotypes were extracted following a modified CTAB protocol (Porebski et al. 1997). Genotyping was performed using the *Brassica* 60K Illumina Infinium SNP array (Illumina Inc., CA, USA) (Clarke et al. 2016). The same marker sets developed in Chapter 3 are utilized here. Briefly, markers were mapped to the *Darmor bzh* reference genome by Rousseau-Gueutin et al. (2020). High-density MS1 contains a total of 22,941 markers. Intermediate-density MS2 was produced by applying an MAF < 0.05 filter to MS1, leaving 14,699 markers. The low density MS3 was produced by a filter applied to MS1 which removes markers with LD above r^2 threshold = 0.2, producing a final marker set of 1,098 markers.

4.3.2 Genomic selection analyses

This experiment investigates the effects of TP/VP design, marker density, and choice of model on GS prediction accuracy. Overall, there were five TP/VP designs utilized in combination with three marker sets and nine GS models, producing 135 unique analyses upon five seed-quality traits (Figure 4.1). The nine GS models used in this study can be categorized as either parametric or non-parametric (ML) models.

Parametric models included Ridge Regression Best Linear Unbiased Prediction (rrBLUP), Genomic Best Linear Unbiased Prediction (GBLUP), BayesA, BayesB, BayesC, and Bayesian Ridge Regression (BRR). All six parametric models were run through RStudio V. 4.3.0 (RStudio Team 2023). The GBLUP and Bayesian models were run using default settings of the BGLR package V. 1.1.0 (Perez and de los Campos 2022). The rrBLUP model was run using the rrBLUP package V. 4.6.2 (Endelman 2011).

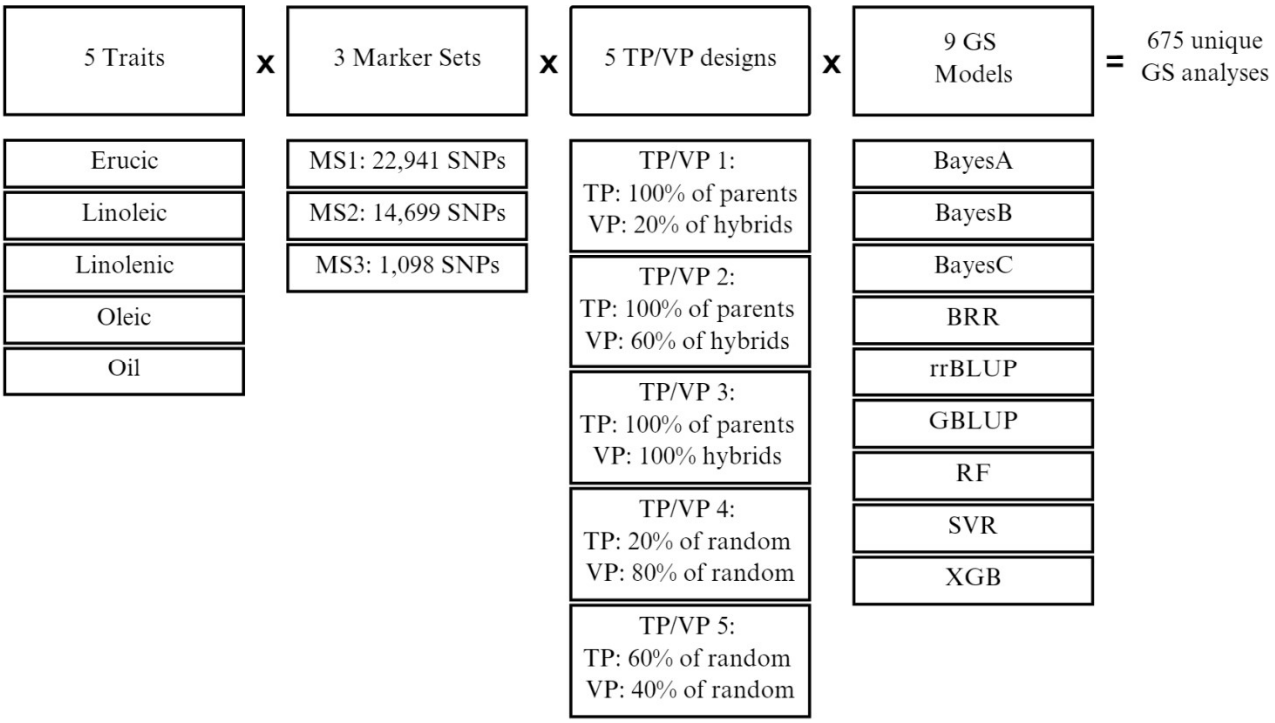


Figure 4.1 Schematic summary of all genomic selection (GS) analyses in this study. Abbreviations: MS: Marker Set; SNP: Single Nucleotide Polymorphism; TP: Training Population; VP: Validation Population; BRR: Bayesian Ridge Regression; rrBLUP: Ridge Regression Best Linear Unbiased Prediction; GBLUP: Genomic Best Linear Unbiased Prediction; RF: Random Forest; SVR: Support Vector Regression; XGB: XGBoost.

For parametric models, missing marker data were imputed using the "A.mat" function in the rrBLUP package which replaced missing data of a particular marker with its mean value across the population.

The effect of TP/VP design was investigated by comparing prediction accuracy response to five different splits of the 427 genotypes into TP and VP. These five TP/VP designs resemble different potential scenarios in a rapeseed hybrid breeding program (Figure 4.1).

Bayesian and GBLUP model cross validation was performed by randomly sampling phenotypic and genotypic data according to each TP/VP design. For each GS analysis (trait × marker set × TP/VP design × model), the model training was iterated 12,000 times with burn-in set to 5,000 (first 5,000 samples discarded) followed by GEBV prediction. For each rrBLUP analysis (trait × marker set × TP/VP design × rrBLUP), 500 iterations of cross validation were performed. Prediction accuracy for each analysis is presented as the Pearson's *r* correlation between GEBV predictions and actual phenotypes.

The ML models used in this study included Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support-Vector Regression (SVR). All hyperparameter tuning, training, and cross validation was performed in Python V. 3.11.5 (Python Software Foundation 2023). The RF and SVR models were imported from the Scikit_learn package V. 1.3.1 (Pedregosa et al. 2011), while XGB was imported from the XGBoost package V. 2.0.0 (Chen and Guestrin 2016).

For ML models, missing data for any particular marker was imputed with the most frequent value for that marker using the "fillna" function in pandas V. 2.2.1 (The pandas development team 2024). Following imputation, marker sets were then encoded with the OneHotEncoder function from Scikit_learn. Hyperparameter optimization was performed through the Optuna package V. 3.6.1 (Akiba et al. 2019). Hyperparameters for each model are specified in Table 4.1. For each TP/VP design, 10 replicates of hyperparameter optimization trials and prediction correlation calculations were performed. For each replicate, hyperparameter optimization included 100 trials performed with Optuna's "maximize" function, indicating the trial whose prediction accuracy was greatest for the trait and saving that trial's hyperparameters as the best parameters. GEBV predictions were performed using best parameters, and prediction correlations were calculated. Prediction accuracy was presented as the mean Pearson's *r* correlation across the 10 replicates for each combination of

Table 4.1 Hyperparameters for optimization using three machine learning algorithms.

Machine learning algorithm	Hyperparameter	Range of values
Extreme Gradient Boosting (XGBoost)	Number of Trees	100 – 400
	Maximum Depth of Tree	5 – 15
	Learning Rate	10^{-4} – 10^{-1}
	Minimum child weight	1.0 – 10
	Gamma	10^{-8} – 1.0
	Subsample	0.5 – 1.0
	Col sample	0.5 – 1.0
Random Forest (RF)	Number of Trees	100 – 400
	Maximum Depth of Tree	5 – 15
Support Vector Regression (SVR)	RBF[1] Kernel - gamma	2^{-17} – 2
	RBF[1] Kernel - C	2^{-5} – 2^{16}

[1] radial basis function

trait × marker set × TP/VP design × model. Due to the heavy computational demand of hyperparameter optimization and GEBV predictions through ML, the above procedure was run through UM's high-performance computing platform "Grex".

Prediction accuracies were compared visually using the ggplot2 package V. 3.5.0 (Wickham 2016) and statistically using Tukey's Honestly Significant Difference (HSD) test in the stats package V. 4.3.0 (RStudio Team 2023). Tukey's HSD calculates and compares the mean GS prediction accuracies of all seed quality traits, TP/VP designs, marker sets, or models, then identifies significant differences among all possible pairs within each aforementioned factor.

4.4 Results

This study compares the prediction accuracies (correlations between GEBVs and actual phenotypes) of 135 unique GS analyses for five seed quality traits. The slate of unique GS analyses was produced by each combination of 5 TP/VP designs, 3 marker sets, and 9 GS models. Prediction accuracy results from all GS analyses are visualized in Figure 4.2 – as results are further discussed, trends of prediction accuracy response to factors are emphasized. It is also important to keep in mind when comparing GS analyses that prediction accuracies are presented as Pearson's *r* correlations which range only from 0 to 1, therefore the significance of differences are regarded as the variation within that small range.

4.4.1 Prediction accuracy response among traits

The most distinguishable response in prediction accuracy was caused by trait. Figure 4.2 demonstrates how prediction accuracies for each trait appear tiered, a pattern which is visible through all other factors. Using a Tukey's HSD test to compare mean prediction accuracies of each trait across all analyses, OIL exhibited significantly poorer prediction accuracies than the other four traits. Alternatively, LLE exhibited significantly greater prediction accuracies than all traits besides LLN. High prediction accuracies were also exhibited for ERU – significantly greater than OLE and OIL, not significantly different than LLN, and only less than LLE by an average difference of 0.056 ($p = 2.81e\text{-}3$). That is, LLE outperformed all traits besides LLN, however ERU prediction accuracies were also not significantly different than LLN (Figure 4.2).

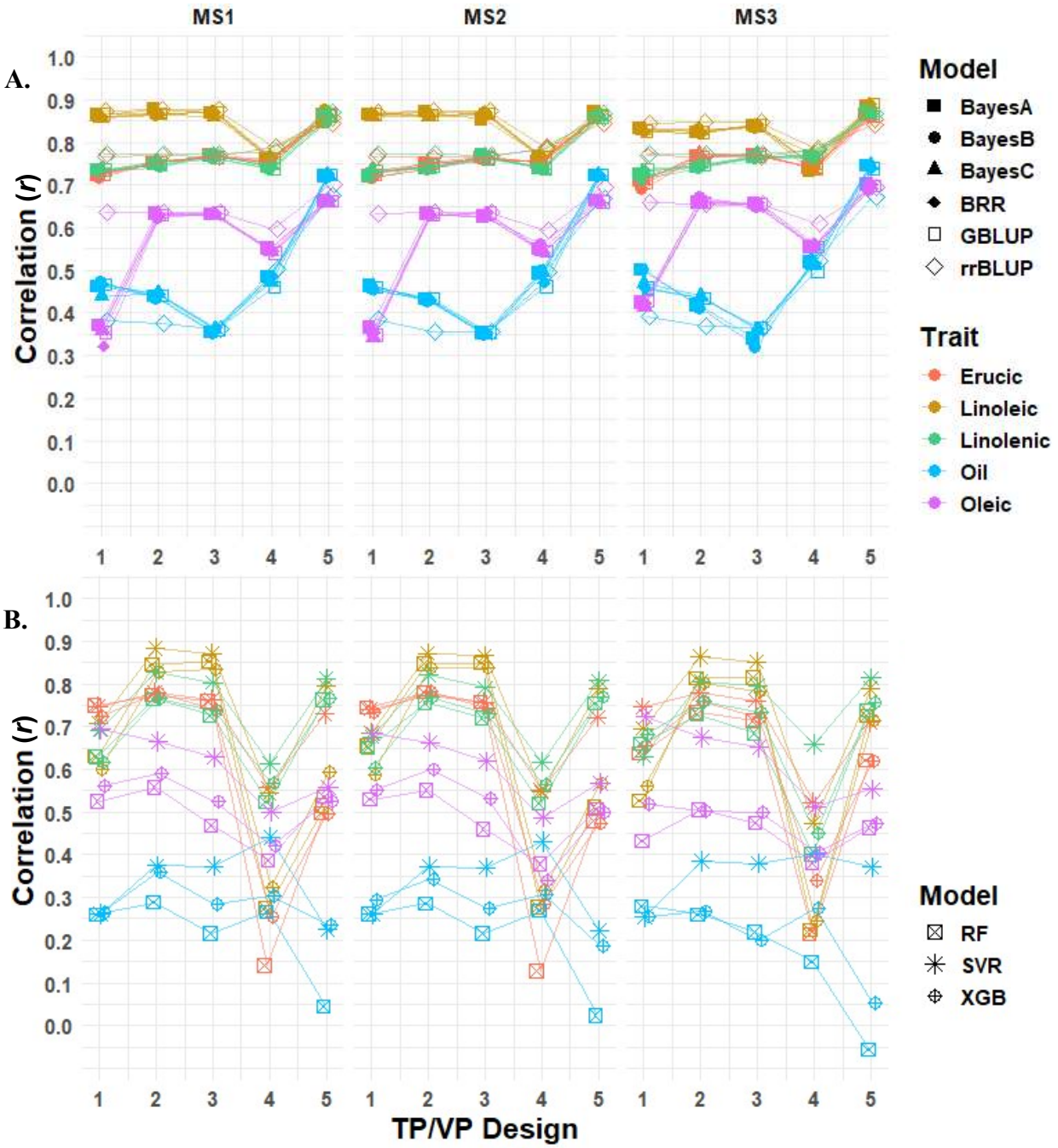


Figure 4.2 Genomic selection (GS) prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five seed quality traits based on three marker sets (MS1: 22941 SNPs; MS2: 14,699 SNPs; MS3: 1,098 SNPs) and five training/validation population designs (design 1: TP = 100 % of parent genotypes, VP = 20 % of hybrid genotypes; design 2: TP = 100 % of parent genotypes, VP = 60 % of hybrid genotypes; design 3: TP = 100 % of parent genotypes, VP = 100 % of hybrid genotypes; design 4: TP = 20 % of random genotypes, VP = 80 %

of random genotypes; design 5: TP = 60 % of random genotypes, VP = 40 % of random genotypes). Two categories of GS models were compared: (A) Six parametric GS models; (B) Three non-parametric (machine learning) GS models. Abbreviations: BRR: Bayesian Ridge Regression; GBLUP: Genomic Best Linear Unbiased Prediction; rrBLUP: Ridge Regression Best Linear Unbiased Prediction; RF: Random Forest; SVR: Support Vector Regression; XGB: XGBoost.

4.4.2 Prediction accuracy response among TP/VP designs

Some TP/VP designs exhibited similar prediction accuracies when their means were compared pairwise. For example, TP/VP designs 2 and 3 (which input all parent genotypes as TP and 202 or 336 of the hybrid genotypes as VP, respectively) showed no significant differences. However, TP/VP design 1, which had the same TP but a small VP of only 68, exhibited prediction accuracies significantly worse than design 2.

In contrast to the aforementioned TP/VP designs, designs 4 and 5 included random samples of both parents and hybrids in the TP and VP (thereby incorporating greater relatedness between TP and VP). TP/VP design 5 exhibited significantly greater mean prediction accuracies than TP/VP design 4. Both TP/VP designs 4 and 1, which respectively included a small TP (86) and a small VP (68) exhibited mean prediction accuracies that were not significantly different from each other, however were significantly poorer than all other TP/VP designs. Finally, TP/VP designs 2, 3, and 5 were not significantly different overall, however TP/VP design 5 exhibited the greatest overall accuracy across four of five traits (and second-greatest in the remaining trait, OLE).

Figure 4.2 further emphasizes the significance of trait towards response in GS prediction accuracy. Prediction accuracies are distinguishable by trait, however it is TP/VP design which most clearly affects the increases or decreases in prediction accuracy within each trait. Each trait produces a unique response to TP/VP regardless of marker set and choice of parametric model, meanwhile ML models also exhibit similar patterns among themselves.

4.4.3 Prediction accuracy response among marker sets

Marker set induces the least significant response in prediction accuracies. Comparing prediction accuracies among marker sets with Tukey's HSD test, MS1, MS2 and MS3 show no significant differences. Figure 4.2 demonstrates the similarity in response for all traits and models across all three marker set densities. Herein, figures are simplified by displaying prediction accuracies only from GS analyses which utilized MS2.

4.4.4 Prediction accuracy response among models

The response of prediction accuracy was different between parametric and ML models (Figure 4.3). Among the six parametric models, none performed significantly differently from one another. Figure 4.2 emphasizes the near-indistinguishable performance of parametric models across all marker sets and TP/VP designs. Given the parametric models' similarity in performance and rrBLUP's consideration as one of the most common GS models in practice, rrBLUP can serve as a representative of parametric models in the following comparisons (Habier et al. 2007; Howard et al. 2022).

The three ML models in this study showed great variation in prediction accuracies, both among themselves and compared to parametric models. Mean SVR prediction accuracies performed significantly better than RF by 0.121 ($p = 8.18e\text{-}4$) and XGB by 0.094 ($p = 0.026$), meanwhile RF and XGB models did not perform significantly differently from one another. Prediction accuracies by rrBLUP were greater than SVR, however a significant difference between the two was not detected ($p = 0.182$). Furthermore, SVR prediction accuracies were not found to be significantly different than any of the other five parametric models. Both RF and XGB performed significantly poorer than parametric models, 0.195 and 0.169 lower than rrBLUP ($p = <1e\text{-}7$ and $p = <1e\text{-}7$) means, respectively. The heat map in Figure 4.5 reemphasizes the trends observed in the scatterplot in Figure 4.4 – ML models tend to exhibit prediction accuracies which are either similar or less than those of parametric models while SVR tends to exhibit the greatest prediction accuracy among the three ML models. Figure 4.5 also emphasizes how prediction accuracies of all three ML models tend to be less than rrBLUP when using TP/VP designs 4 and 5.

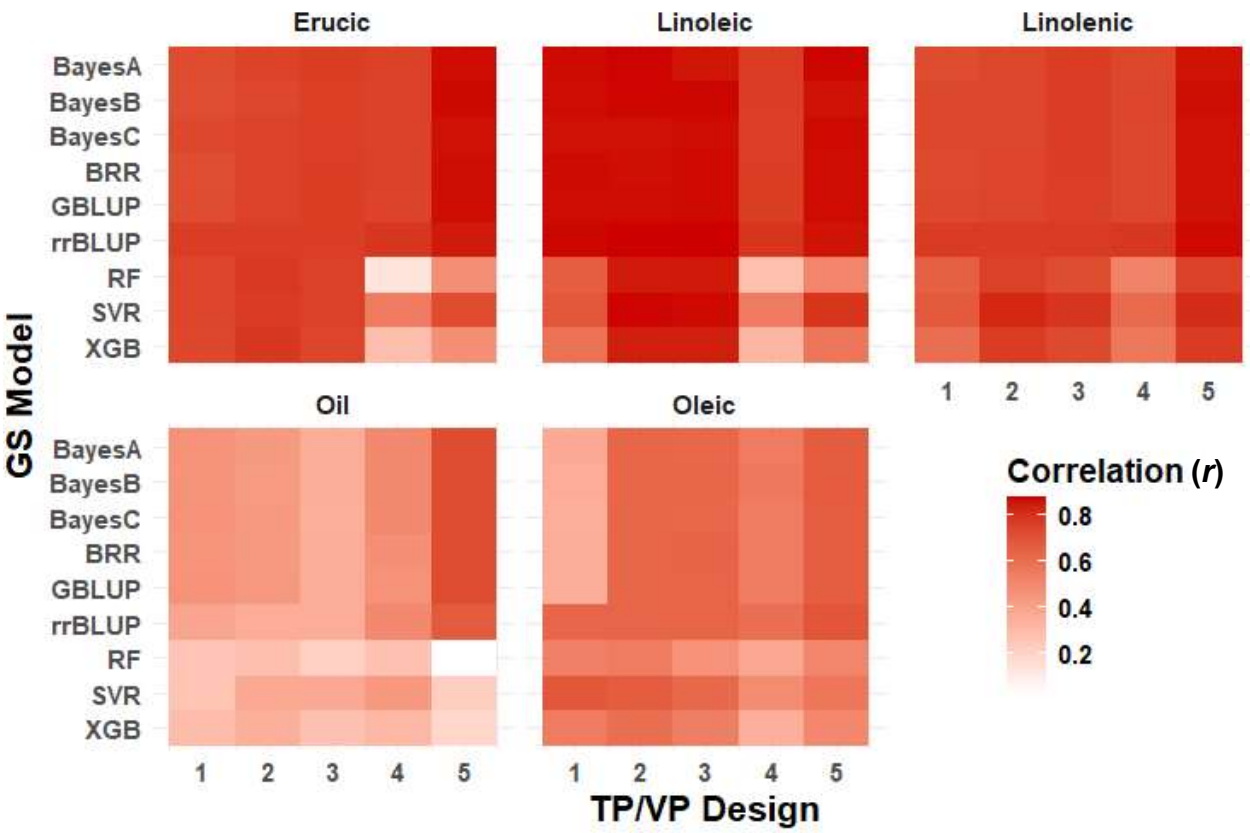


Figure 4.3 Heat maps of genomic selection (GS) model prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five seed quality traits based on 14,699 SNP markers and five training/validation population designs (design 1: TP = 100 % of parent genotypes, VP = 20 % of hybrid genotypes; design 2: TP = 100 % of parent genotypes, VP = 60 % of hybrid genotypes; design 3: TP = 100 % of parent genotypes, VP = 100 % of hybrid genotypes; design 4: TP = 20 % of random genotypes, VP = 80 % of random genotypes; design 5: TP = 60 % of random genotypes, VP = 40 % of random genotypes). Six parametric models (BayesA, BayesB, BayesC, BRR, GBLUP, rrBLUP) and three nonparametric models (RF, SVR, XGB) were compared. Abbreviations: BRR: Bayesian Ridge Regression; GBLUP: Genomic Best Linear Unbiased Prediction; rrBLUP: Ridge Regression Best Linear Unbiased Prediction; RF: Random Forest; SVR: Support Vector Regression; XGB: XGBoost.

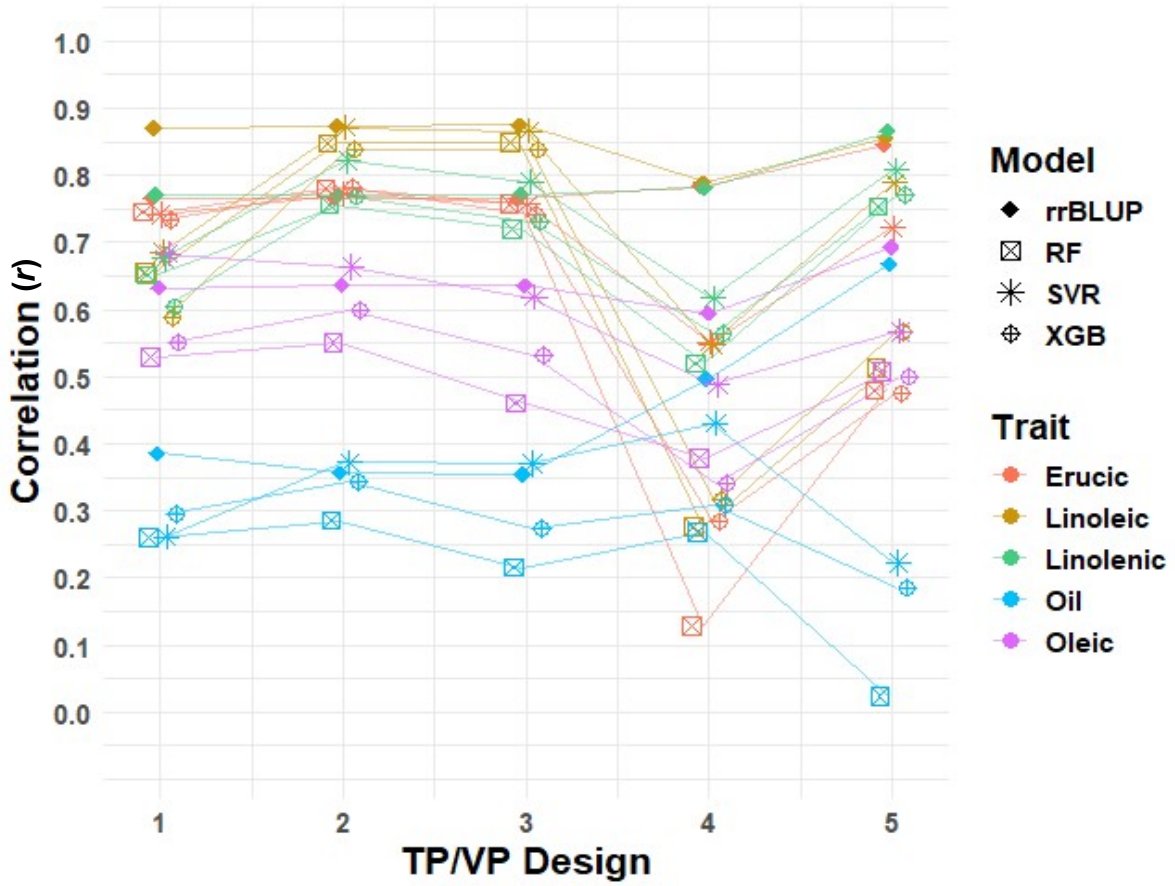


Figure 4.4 Genomic selection (GS) model prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five seed quality traits based on 14,699 SNP markers and five training/validation population designs (design 1: TP = 100 % of parent genotypes, VP = 20 % of hybrid genotypes; design 2: TP = 100 % of parent genotypes, VP = 60 % of hybrid genotypes; design 3: TP = 100 % of parent genotypes, VP = 100 % of hybrid genotypes; design 4: TP = 20 % of random genotypes, VP = 80 % of random genotypes; design 5: TP = 60 % of random genotypes, VP = 40 % of random genotypes). Parametric models are represented by rrBLUP for comparison to three nonparametric (machine learning) models. Abbreviations: rrBLUP: Ridge Regression Best Linear Unbiased Prediction; RF: Random Forest; SVR: Support Vector Regression; XGB: XGBoost.

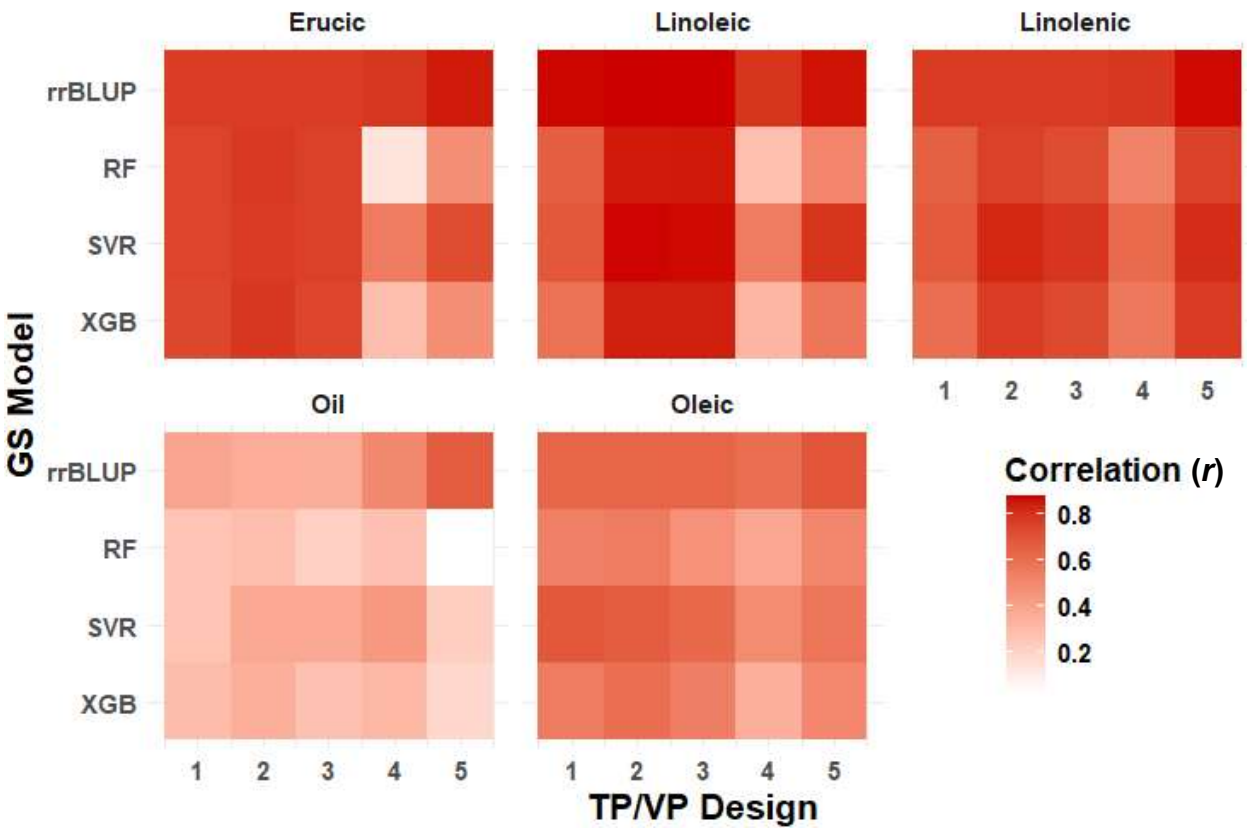


Figure 4.5 Heat maps of genomic selection (GS) model prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five seed quality traits based on 14,699 SNP markers and five training/validation population designs (design 1: TP = 100 % of parent genotypes, VP = 20 % of hybrid genotypes; design 2: TP = 100 % of parent genotypes, VP = 60 % of hybrid genotypes; design 3: TP = 100 % of parent genotypes, VP = 100 % of hybrid genotypes; design 4: TP = 20 % of random genotypes, VP = 80 % of random genotypes; design 5: TP = 60 % of random genotypes, VP = 40 % of random genotypes). Parametric models are represented by rrBLUP for comparison to three nonparametric models. Abbreviations: rrBLUP: Ridge Regression Best Linear Unbiased Prediction; RF: Random Forest; SVR: Support Vector Regression; XGB: XGBoost.

4.5 Discussion

For crop improvement, GS provides the potential to improve selection efficiency, thereby reducing the overall number of breeding cycles and aiding in sustainability (Heffner et al. 2010). Despite its introduction in 2001, GS is still considered in its early stages (Zhang et al. 2023). The notion regarding GS methodology is that "no one size fits all" – the inherent differences among plant breeding programs cause significant variations in predictive ability, prompting researchers to evaluate GS in their own context (Lorenz et al. 2011; Zhang et al. 2023).

This study conducted an extensive range of 675 GS analyses to demonstrate the effects of TP/VP design, marker set density, and choice of GS model on the prediction of five seed quality traits in *B. napus*. Overall, we demonstrated that GS can provide highly accurate predictions of seed quality traits, identifying analyses which consistently produce prediction accuracies above 0.70 for all five traits. Considering that a prediction accuracy threshold above 0.50 can provide multifold advantages in genetic gain and resource reduction compared to conventional selection or MAS, the results of this study indicate that GS may fit well in the *B. napus* improvement toolset (Heffner et al. 2010; Longin et al. 2015).

Prediction accuracies in this study were significantly impacted in response to trait complexity, TP/VP design, and some GS models, however marker set density did not produce any significant effects. Figure 4.2 emphasizes GS' reduced predictive ability for more complex traits, as also reported by Farooq et al. (2023), Heffner et al. (2009), and Hu et al. (2021b). Previous studies have described the high complexity of OIL, controlled by a multitude of minor-effect QTL rather than any reliable major-effect QTL (Delourme et al. 2006; Jiang et al. 2014; Zhao et al. 2007). Many biosynthetic pathways are considered to contribute to the overall accumulation of OIL, which may be further affected by environment (Delourme et al. 2006; Hu et al. 2021b; Koscielny et al. 2019). This is consistent with Zou et al. (2016) who also observed generally lower prediction accuracies for OIL than specific fatty acids including ERU. Furthermore, Figure 3.2 demonstrates that OIL is not significantly correlated to proportions of any of the four fatty acids. That is, although these four fatty acids constitute most of HEAR's fatty acid profile, overall OIL accumulation appears more complex than the biosynthetic pathways which control the proportions of its main fatty acids, most likely attributed to other minor-effect genes and genotype × environment interactions. Alternatively, OLE was significantly negatively correlated to each of the other three fatty acids in

this study, consistent with its role as a source to those fatty acids (Kaur et al. 2019). Prediction accuracy for OLE was generally greater than OIL but lower than the other fatty acids, perhaps attributed to its complex sink-source relationship with LLE, LLN, and ERU through inter-organelle transport (plastid to endoplasmic reticulum) split into two independent pathways (desaturation to produce LLE and LLN or elongation to produce eicosenoic acid and ERU) (Kaur et al. 2019).

Despite OLE and OIL typically exhibiting the lowest prediction accuracies among traits, several analyses improved their prediction accuracy well above the 0.50 threshold. Prediction accuracies responded most distinguishably by altering TP/VP designs, meanwhile marker set density induced negligible differences (focusing on results only from MS2 facilitates clearer comparisons). The observations here are consistent with previous GS studies in rapeseed and other crops where the size and relatedness between TP and VP induce strong effects on prediction accuracy (Beukert et al. 2017; Hu et al. 2021b; Jan et al. 2016; Reif et al. 2012; Technow et al. 2014; Werner et al. 2020; Zhao et al 2015). Werner et al. (2020) suggested that incorporating greater relatedness between the TP and VP helps to ensure similar linkage phase between markers and QTL in the training and testing of GS models, meanwhile increasing population size and diversity ensures that many different haplotypic combinations are captured in the TP. In this study, TP/VP design 5 (largest TP size (257 lines) and high TP/VP relatedness (TP containing both parents and hybrids) produced the greatest prediction accuracy across nearly all traits while utilizing a majority of the GS models. Despite the fact TP/VP design 4 also incorporated high TP/VP relatedness, its small TP size consistently produced poorer prediction accuracies than TP/VP design 5 across all traits and models (especially ML models). The poorer performance of TP/VP design 4 demonstrates that although TP/VP relatedness can improve prediction accuracy, TP size must also be sufficiently large to capture the diversity of haplotypic combinations possible within parents and hybrids. Hu et al. (2021b) noted that medium to high accuracy predictions can be made if both parental lines were included in the TP, but observed significantly decreased prediction accuracies when only one or neither parents were included. Although TP/VP designs 1-4 contained similarly small TP sizes, TP/VP design 4 likely excluded one or both of a hybrid's parents across many iterations of its random sampling of parent and hybrid lines, incorporating poor relatedness between the TP and VP, therefore poor predictive ability. Alternatively, TP/VP designs 1-3 included all 91 parents in their TP, producing generally good prediction accuracies based on model training which

considered both parents of all predicted hybrids. TP/VP design 5 either produced similar or improved prediction accuracies – that is, although TP containing only parents can produce good prediction accuracies, it is overall more reliable to include both hybrids and parents in model training. Our results emphasize the importance of TP/VP design described in previous studies: the predictive power of a GS model relies on the diversity of haplotypic combinations provided to it and how closely those combinations resemble the VP whose phenotypes they will be predicting.

This study compared the prediction accuracy response to various marker densities in MS1, MS2, and MS3 (22,941 SNPs, 14,699 SNPs, and 1,098 SNPs, respectively). In another GS study on *B. napus*, Werner et al. (2018) observed effective prediction accuracies when utilizing low density markers (a few hundred to a few thousand) which were comparable to prediction accuracies when utilizing manyfold larger marker sets. Zou et al. (2016) observed insignificant differences when predicting rapeseed traits with 1000, 5000, and 13,678 SNPs (while some traits even exhibited similar prediction accuracies with only 100 SNPs). Similarly, predictive abilities did not significantly increase or decrease for Kaler et al. (2022) or Poland et al. (2012) when comparing full marker sets to several LD-filtered marker sets in corn, soybean, rice, and wheat. The similarity among marker sets in Figure 4.2 illustrates how, as long as LD decay is accounted for, sufficient linkage to causal loci is available to estimate marker effects in GS model training. Kaler et al. (2022) described this as a haplotype block performing similar to a single marker, where adding more markers does not improve accuracy but instead can increase error (overfitting). The ability to produce reliable predictions with low-density marker sets may be advantageous to reduce both genotyping costs and computational burden, however, the effect of marker density on GS prediction accuracy can be nuanced. For example, when the TP and VP contain more distantly related accessions, improved prediction accuracy has been observed by increasing marker density (Norman et al. 2018). Alternatively, lower marker densities may provide sufficient information if the VP is composed of TP progeny since the similar, larger haplotype blocks reduce the need for extensive marker coverage (Meuwissen 2009; Hickey et al. 2014). Therefore, the decision of optimal marker density for GS in rapeseed intersects with both LD decay and breeding population design. With this in mind, breeders may elect to keep dense marker sets available to buffer the effects of LD disruption from recombination over long-term breeding cycles (DoVale et al. 2022).

All six parametric models in this study are designed with different assumptions and treatments of marker effects. In practice, however, the parametric models exhibited statistically equivalent prediction accuracies across TP/VP designs and marker sets when predicting respective seed quality traits. Our results are supported by previous studies which demonstrate that these parametric models can produce statistically equivalent accuracies for various traits in soybean, barley, and wheat (Heslot et al. 2012; Howard et al. 2022; Lorenz et al. 2012; Merrick and Carter 2021; Rolling et al. 2020; Zou et al. 2016). For *B. napus*, Zou et al. (2016) observed very little performance differences across a similar set of parametric models when predicting OIL, ERU, and other seed quality traits. Hu et al. (2021b) compared several models with various parametric marker effect treatments, observing that simple parametric models provide similarly effective prediction accuracies for traits including OIL, OLE, ERU, and yield. These results both reflect and contrast the relative performance of parametric models in other GS studies. Clearly, the inherent differences in parametric model design are known to cause performance differences. For example, Bayesian models are shown to outperform BLUP-based models when traits are controlled by fewer QTLs, meanwhile rrBLUP and GBLUP perform similarly or differently to one another and Bayesian models across various traits in *B. napus* and other species (Bhering et al. 2015; Daetwyler et al. 2010; Tan et al. 2017; Wang et al. 2015; Wang et al. 2018b; Zou et al. 2016). More than anything, the contrasting relative performances of models emphasize the "no one size fits all" notion of GS that prompted this study in application to *B. napus* oil quality (Larkin et al. 2019; Lorenz et al. 2011). Bayesian models produced the highest prediction accuracy in four out of five traits, however rrBLUP or GBLUP were always less than 0.03 below the highest accuracy. Considering its historical popularity and overall consistency among other parametric models in this study, rrBLUP served as an appropriate representative of parametric models in comparison to the non-parametric (machine learning) models (Habier et al. 2007; Howard et al. 2022). Similar to the parametric models, prediction accuracies by ML demonstrated distinguishable patterns across TP/VP designs for each respective trait, regardless of marker set. SVR's performance was significantly greater than RF and XGB, however was not distinguishable from rrBLUP or other parametric models. For some traits and TP/VP designs, SVR was marginally more accurate than rrBLUP, however rrBLUP outperformed SVR in all analyses which used TP/VP design 5 – the optimal design in this study. Merrick and Carter (2021) described how ML models may provide some slight improvements when using combined site-year data, however parametric models

consistently performed better than ML on a year-to-year or environment-to-environment basis. Our study emphasizes that, despite enormous interest in ML and deep learning, simpler parametric GS models can provide accurate and reliable predictions for marker-based improvement of complex seed quality traits (Hu et al. 2021b; Kaler et al. 2022; Merrick and Carter 2021; Zou et al. 2016). A potential explanation for the generally poor performance of ML in this study may be the relatively small TP sizes for sufficient learning. To optimize ML predictions, Sandhu et al (2021b) suggested increasing TP size (larger than their population of 666 lines). Despite the relatively poor performance of ML in this study, other investigations of GS demonstrate some advantages of ML for predicting agronomic and seed quality traits, especially those controlled by epistatic interactions (Farooq et al. 2023; Howard et al. 2014; Sandhu et al. 2021a; Sandhu et al. 2021b). In general, our results support the recommendation that both parametric and nonparametric GS models provide viable options for the improvement of complex traits (Merrick and Carter 2021).

4.6 Conclusion

This study demonstrates that GS can be a useful tool for *B. napus* improvement – all five seed quality traits exhibited prediction accuracies above a threshold that facilitates improved genetic gains relative to conventional selection or MAS. Trait complexity was a major factor affecting prediction accuracy, however the most complex trait, OIL, could still be predicted with accuracy as high as 0.745, whereas simpler traits such as LLE exhibited prediction accuracies as high as 0.888. The TP/VP design imposed significant effects on prediction accuracy, producing distinguishable responses among traits depending on TP size and relatedness to the VP. All six parametric models performed similarly to one another, whereas ML models either performed similarly to or poorer than parametric models. Marker density did not significantly affect prediction accuracy.

5. INTEGRATING GENOME-WIDE ASSOCIATION STUDY RESULTS TO IMPROVE GENOMIC SELECTION FOR OIL AND FATTY ACID PROFILE IN RAPESEED (*Brassica napus* L.)

5.1 Abstract

The advantage granted by GWAS-guided GS relies on selecting small subsets of informative markers based on GWAS results. This method aims to exclude non-causative markers from GS model training while ensuring that the selected markers provide sufficient effect information for accurate phenotypic predictions – simultaneously mitigating model overfitting/underfitting and lessening computational burden. This study evaluated GWAS-guided GS in *Brassica napus* L. breeding scenarios by inputting three marker sets to GWAS of parent genotypes, selecting trait-specific subsets of informative markers, and predicting five seed quality traits of hybrid genotypes using six parametric GS models. We identified GWAS-guided GS analyses which slightly outperformed conventional GS across all seed quality traits, however significant differences between the two methods were not identified. Parametric GS models produced equivalent results among themselves, similar to conventional GS applications for the same seed quality traits (erucic acid content (ERU), linoleic acid content (LLE), linolenic acid content (LLN), oleic acid content (OLE), and overall oil content (OIL)). Prediction accuracy response to individual traits was consistent between conventional GS and GWAS-guided GS, suggesting that the marker subsetting approach sufficiently captured the heritability of all five seed quality traits. Prediction accuracies achieved in this study rose above thresholds in other studies that facilitated greater genetic gains than conventional selection or MAS, demonstrating the potential of GWAS-guided GS to improve *B. napus* breeding while simultaneously reducing genotyping and phenotyping costs in practical applications.

5.2 Introduction

Despite being broadly evaluated in plant breeding applications, GS can still be considered in its early stages – a universally acceptable approach has been prevented by several inherently different factors among breeding programs (i.e. no “one size fits all”) (Lorenz et al. 2011; Zhang et al. 2023).

Although not a complete solution to this problem, GWAS-guided GS has been suggested as an option to improve the accuracy and cost-efficiency of GS methodology.

The central concept of GWAS-guided GS is the selection of optimally informative markers (based on GWAS results) as inputs for GS model training and phenotypic predictions. Werner et al. (2018) describes the capacity of GWAS-guided GS to allow fewer representative markers selected to capture sufficient association between genomic regions and a phenotypic trait. Low-density, GWAS-based marker sets enable GS prediction accuracies comparable to those achieved with high-density genotyping (Kaler et al. 2022; Werner et al. 2018). Importantly, GWAS-guided GS mitigates model overfitting by systematically excluding non-causative markers from GS model training, thereby avoiding erroneous identification of these random variations as causative for the trait of interest (Hickey et al. 2014; Montesinos-López et al. 2022b; Zhang et al. 2023). Simultaneously, although the marker subsets are typically very small, GWAS-guided GS mitigates underfitting by ensuring that the selected markers provide sufficient effect information for the model to perform accurate phenotypic predictions (Montesinos-López et al. 2022b; Zhang et al. 2023).

The critical question arises though: how does one select the subset of markers which provides optimal trait predictability? Several methods have been developed to answer this question. Zhang et al. (2014) proposed a method later described as “GS + historic GWAS” which used preexisting GWAS results from public databases to build trait-specific matrices for phenotypic predictions. Spindel et al. (2016) proposed the “GS + de novo GWAS” method to input significant markers identified from GWAS within the training population of interest (i.e. de novo) to train the GS models as fixed effects. More recently, Jeong et al. (2020) developed the “GMStool” program, using machine learning to select optimal marker sets and optimal WGR models for phenotypic prediction. Werner et al. (2018) and Kaler et al. (2022) compared GS accuracies when trained with markers identified by GWAS results filtered at increasingly stringent levels of significance.

Several studies have evaluated GWAS-guided GS by comparing a variety of GWAS models, GS models, and marker subsetting methods in application to several species and traits. Rice has been a focus of GWAS-guided GS research, starting with the proposal of GS + historic GWAS by Zhang et al. (2014). Spindel et al. (2016) proposed the GS + de novo GWAS method in rice, observing that the combination of simple Multiple Linear Regression (MLR) GWAS results with rrBLUP

outperformed several other WGR models. Since then, the GS + de novo GWAS method has been evaluated for rice breeding, investigating the predictive ability of GS for important agronomic and quality traits such as yield, panicle number, flowering time, plant height, protein content, and more (Jeong et al. 2020; Zhang et al. 2023). Wheat has also been the subject of GWAS-guided GS research wherein traits such as yield and septoria blotch resistance have been predicted with rrBLUP and GBLUP (Odilbekov et al. 2019; Sehgal et al. 2020). Studies on forest tree breeding have also implemented GWAS-guided GS methods, predicting wood quality, growth, and abiotic stress resistance traits in Norway spruce and *Eucalyptus* (Chen et al. 2023; Tan and Ingvarsson 2022). *B. napus* has also been the subject of GWAS-guided GS research. Werner et al. (2017) utilized the method by Spindel et al. (2016), combining GWAS results from GenABEL P+K to evaluate the ability of rrBLUP, BayesB, and BRR to predict general and specific combining ability in *B. napus*. Werner et al. (2018) also used a similar method to predict oil content, glucosinolate content, and plant height.

The objective of this study is to evaluate the potential of GWAS-guided GS to improve prediction accuracy of *B. napus* seed quality traits as compared to conventional GS. This study conducted a wide array of analyses which observed the effects of trait complexity, marker set density, and choice of GS model in order to identify methods which may improve prediction accuracy. The long-term objective of this study is to assess whether general implementation of GWAS-guided GS methods could improve genetic gains and lower costs of *B. napus* breeding programs. We hypothesize that trait complexity, marker set density, and the choice of GS model will significantly affect GWAS-guided GS prediction accuracy relative to conventional GS.

5.3 Materials and methods

5.3.1 Plant materials, phenotypic data, and genotypic data

Details of phenotypic and genotypic data collection in this study were described in Chapter 3.3.1. Briefly, this study investigates GWAS and GS by inputting phenotypic and genotypic data from a population of 454 high erucic acid rapeseed (HEAR) genotypes (91 parental genotypes and 345 hybrid genotypes) evaluated across a total of 48 site-years. Unbalanced phenotypic data replicates were corrected by calculating best linear unbiased prediction (BLUP) values for all parent and hybrid genotypes. These values were used as phenotypic data inputs for GWAS and GS analyses.

The DNA of all genotypes were extracted following a modified CTAB protocol (Porebski et al. 1997). Genotyping was performed using the *Brassica* 60K Illumina Infinium SNP array (Illumina Inc., CA, USA) (Clarke et al. 2016). The 454 genotypes were filtered for genotypic data quality by culling all with < 0.8 call rate, leaving 427 genotypes (91 parents, 336 hybrids). The same marker sets developed in Chapter 3 are utilized here: high-density MS1 (22,941 SNP markers), intermediate-density MS2 (14,699 SNP markers), and low-density MS3 (1,098 SNP markers).

5.3.2 GWAS-guided GS analyses

This experiment investigates the effects of trait, marker density, and model on the predictive ability of GWAS-guided GS. To simulate a practical plant breeding scenario, informative SNP subsets were selected based on the parent population's GWAS results. Each of the five seed quality traits were analyzed in the parent population by inputting MS1, MS2, and MS3 to FarmCPU, giving rise to 15 GWAS results. Each of the 15 GWAS results were filtered to include the 200 most significant markers for each trait (on the basis of p-value). Each GWAS-guided marker subset was subsequently input to GS models for training and GEBV predictions. Six GS models were employed in combination with the marker subsets using all 91 parents as the training population and all 336 hybrids as the validation population, matching TP/VP design 3 utilized in Chapter 4. The above factors produced 90 unique GWAS-guided GS analyses (18 for each seed quality trait) (Figure 5.1).

The six parametric GS models used in this experiment are the same as used in Chapter 4, including rrBLUP, GBLUP, BayesA, BayesB, BayesC, and BRR. All six parametric models were run through RStudio V. 4.3.0 (RStudio Team 2023). The GBLUP and Bayesian models were run using default settings of the BGLR package V. 1.1.0 (Perez and de los Campos 2014). The rrBLUP model was run using the rrBLUP package V. 4.6.2 (Endelman 2011). Missing marker data was imputed using the "A.mat" function in the rrBLUP package which replaced missing data of a particular marker with its mean value across the population.

For each GWAS-rrBLUP analysis (trait × marker set × rrBLUP), 500 iterations of cross validation were performed. For each GWAS-guided GS analysis (trait × GWAS-guided marker subset × GS model) besides GWAS-rrBLUP, the model training was iterated 12,000 times with burn-in set to 5,000 (first 5,000 samples discarded) followed by GEBV prediction. Prediction accuracy for each

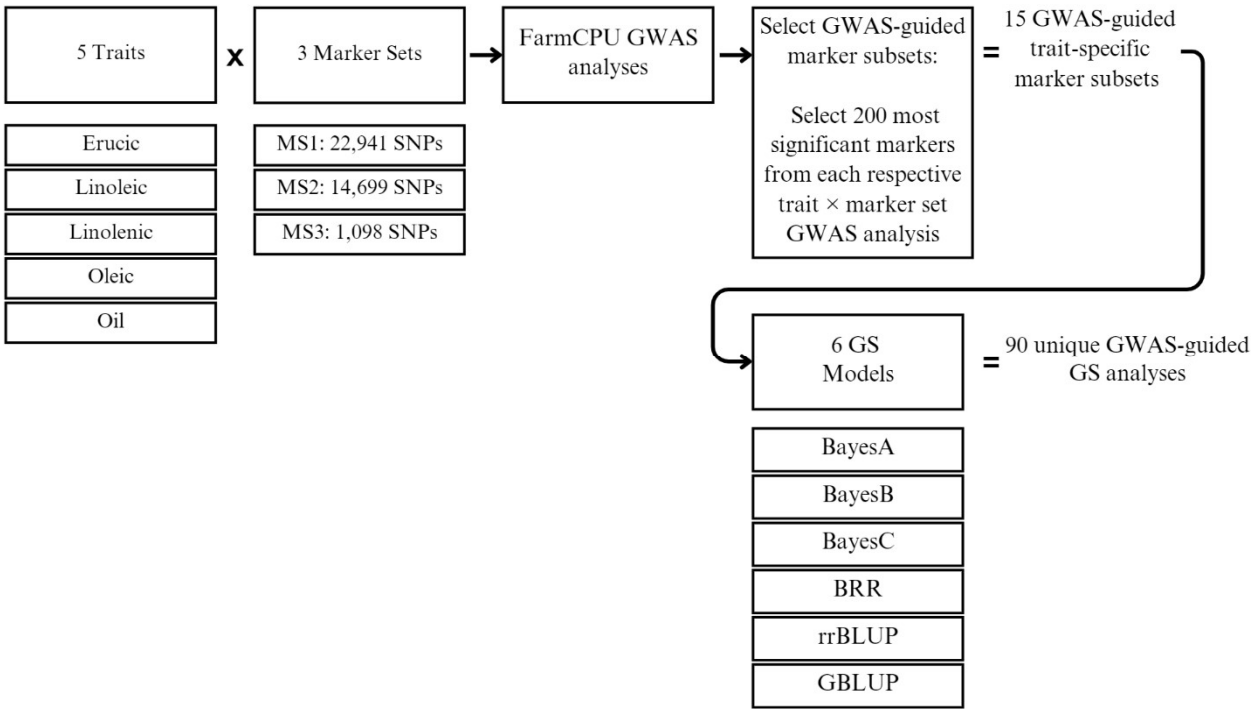


Figure 5.1 Schematic summary of all GWAS-guided genomic selection analyses in this study. Abbreviations: MS: Marker Set; SNP: Single Nucleotide Polymorphism; FarmCPU: Fixed and random model Circulating Probability Unification; BRR: Bayesian Ridge Regression; rrBLUP: Ridge Regression Best Linear Unbiased Prediction; GBLUP: Genomic Best Linear Unbiased Prediction.

analysis is presented as the Pearson's *r* correlation between GEBV predictions and actual phenotypes. The prediction accuracies from GWAS-guided GS analyses were compared to their corresponding conventional GS analyses from Chapter 4.

Prediction accuracies were compared visually using the ggplot2 package V. 3.5.0 (Wickham 2016) and statistically using Tukey's Honestly Significant Difference (HSD) test in the stats package V. 4.3.0 (RStudio Team 2023).

5.4 Results

This experiment performed 18 unique GWAS-guided GS analyses for each of the five seed quality traits. The unique analyses were produced by each combination of 3 marker sets, 1 TP/VP design, and 6 GS models (Figure 5.1). Trends in GS prediction accuracy were observed in both GWAS-guided GS and conventional GS respectively, followed by direct comparisons of performance between the two methods.

5.4.1 GWAS-guided GS prediction accuracy response among factors

As observed in Chapter 4 where GS prediction accuracy is clearly tiered by trait, the accuracies produced by GWAS-guided GS also vary significantly by trait. The GWAS-guided GS prediction accuracies exhibit the same tiering as seen in Figure 4.2 – that is, prediction accuracies for LLE were the greatest among the four other traits, ranging from 0.791 to as high as 0.901 across all models and marker sets. Below LLE, the overall prediction accuracies of LLN and ERU were statistically indistinguishable and respectively ranged from 0.712 to 0.811 and from 0.701 and 0.804. Prediction accuracies for OIL were significantly poorer than all four other traits, ranging from 0.323 to 0.463, while OLE was also significantly poorer than both LLN and ERU, ranging from 0.513 to 0.692 across all models and marker sets.

Regarding the choice of prediction model, Figure 5.2 demonstrates how prediction accuracies among the six parametric models are statistically equivalent. For any trait × marker set combination, the prediction accuracies by all six parametric models varied little, often grouping within the same ~0.05 range.

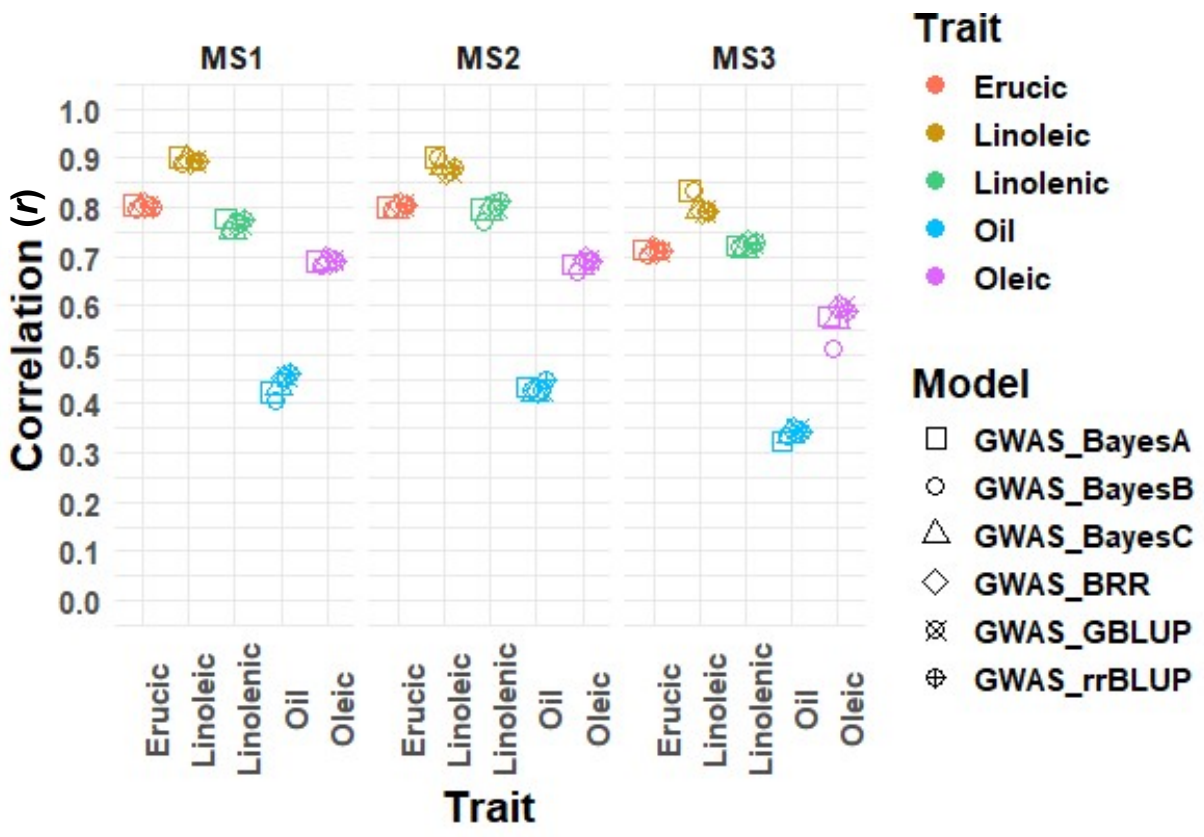


Figure 5.2 GWAS-guided genomic selection (GWAS-guided GS) prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five *B. napus* L. seed quality traits based on three marker sets (MS1: 22,941 SNPs; MS2: 14,699 SNPs; MS3: 1,098 SNPs) predicted by six parametric models. Training and validation populations consisted of 91 parental genotypes and 336 hybrid genotypes, respectively. GWAS-guided GS analyses of each trait utilized the 200 most significant markers from respective FarmCPU GWAS results as unique informative marker subsets. Abbreviations: BRR: Bayesian Ridge Regression; GBLUP: Genomic Best Linear Unbiased Prediction; rrBLUP: Ridge Regression Best Linear Unbiased Prediction.

In further consensus with the trends in Chapter 4, we also observe that GWAS-guided GS prediction accuracy does not vary significantly across the three marker sets (Figure 5.2). That is, when GWAS results from analyses using MS1, MS2, and MS3 (22,941 markers, 14,699 markers, and 1,098 markers, respectively) were subset to small groups of 200 informative markers, GWAS-guided GS performed statistically similarly across all marker sets. Of note, GWAS-guided GS utilizing MS3 did exhibit slightly lower prediction accuracies than MS1 and MS2 on average, however not significantly poorer than either – a Tukey's HSD test indicated that MS3 prediction accuracies were lower than those of MS1 on average by a difference of 0.087 ($p = 0.097$) and lower than MS2 by a difference of 0.088 ($p = 0.090$).

5.4.2 GWAS-guided GS versus conventional GS

For each trait, 18 GWAS-guided GS analyses were conducted by combining all choices of model and marker set (Figure 5.1) and directly compared to corresponding conventional GS analyses (i.e. those with the same combinations of above factors).

Figure 5.3 demonstrates how similarly GWAS-guided GS and conventional GS behave within themselves. As detailed in 5.4.1, prediction accuracies by both methods exhibit similar prediction accuracies on a trait-by-trait basis but vary little on the basis of marker set and model. Slight differences arise when comparing the two GS methods directly. Figure 5.3 demonstrates some distinguishable grouping among the two methods. GWAS-guided GS slightly outperformed conventional GS for all traits and models when inputting both MS1 and MS2, however performed slightly poorer than conventional GS when inputting MS3. That is, GWAS results from analyses of MS1 and MS2 that were sorted to identify the 200 most informative markers (of 22,941 and 14,699 markers from MS1 and MS2, respectively) for each trait provided slightly increased prediction accuracies when used to guide GS, however, prediction accuracies slightly decreased when guiding GS with the top 200 markers from GWAS analyses of MS3 (1,098 markers). Despite being somewhat visually distinguishable in both Figure 5.3 and Figure 5.4, the overall performance differences between GWAS-guided GS and conventional GS are small and not statistically significant.

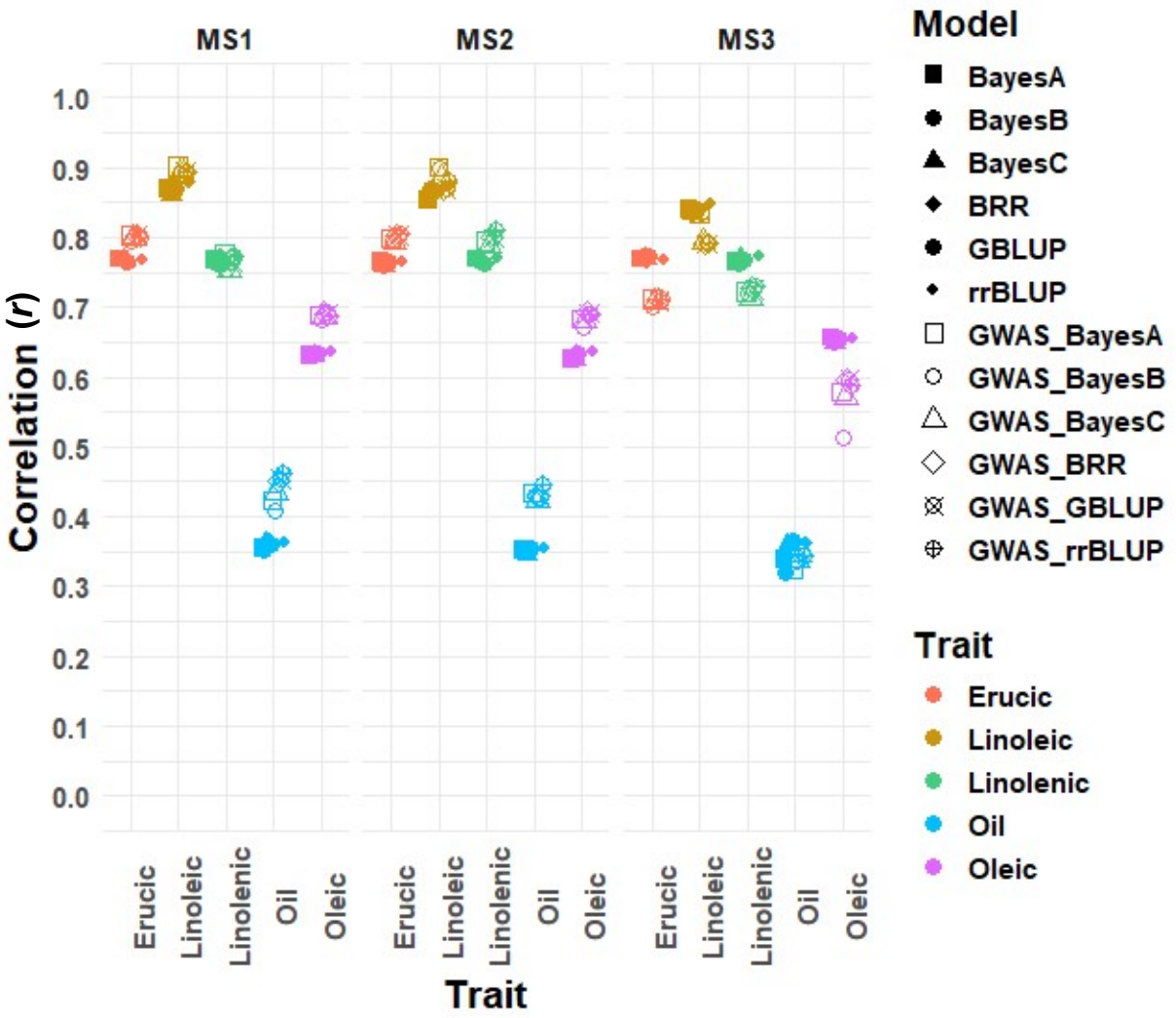


Figure 5.3 Conventional genomic selection (GS) and GWAS-guided GS prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five *Brassica napus* L. seed quality traits based on three marker sets (MS1: 22,941 SNPs; MS2: 14,699 SNPs; MS3: 1,098 SNPs) predicted by six parametric models. Training and validation populations consisted of 91 parental genotypes and 336 hybrid genotypes, respectively. Conventional GS input all markers within respective marker sets. GWAS-guided GS analyses of each trait utilized the 200 most significant markers from respective FarmCPU GWAS results as unique informative marker subsets. Abbreviations: BRR: Bayesian Ridge Regression; GBLUP: Genomic Best Linear Unbiased Prediction; rrBLUP: Ridge Regression Best Linear Unbiased Prediction.

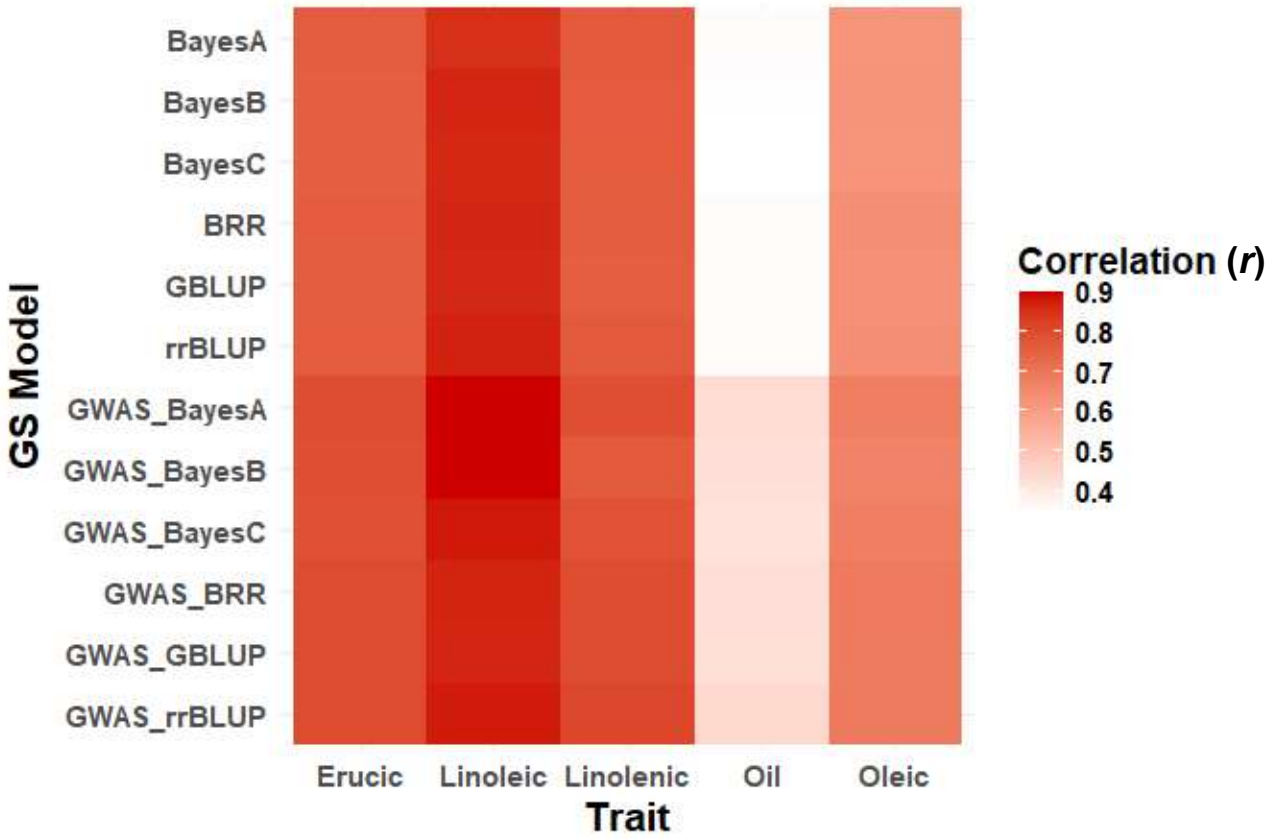


Figure 5.4 Heatmap of conventional genomic selection (GS) and GWAS-guided GS model prediction accuracies (Pearson's *r* correlation between predictions and actual phenotypes) for five *Brassica napus* L. seed quality traits based on MS2 (14,699 SNP markers) predicted by six parametric models. Training and validation populations consisted of 91 parental genotypes and 336 hybrid genotypes, respectively. Conventional GS input all markers within respective marker sets. GWAS-guided GS analyses of each trait utilized the 200 most significant markers from respective FarmCPU GWAS results as unique informative marker subsets. Abbreviations: BRR: Bayesian Ridge Regression; GBLUP: Genomic Best Linear Unbiased Prediction; rrBLUP: Ridge Regression Best Linear Unbiased Prediction.

5.5 Discussion

To investigate GWAS-guided GS methods for *B. napus* improvement, we predicted five seed quality traits in hybrid genotypes using GS models trained by small subsets of informative markers. These marker subsets were identified based on results from GWAS analyses of parent genotypes. GWAS-guided GS approaches aim to improve prediction accuracy by excluding non-causative markers from GS model training while ensuring that the included markers provide sufficient effect information for accurate phenotypic predictions, simultaneously mitigating model overfitting and underfitting while lessening computational burden (Hickey et al. 2014; Montesinos-López et al. 2022b; Zhang et al. 2023). Although slight differences were observed in several analyses, our study demonstrates that the predictive power of GWAS-guided GS does not outperform conventional GS across these five seed quality traits.

The six parametric models did not exhibit any significant differences among themselves when implemented to GWAS-guided GS, consistent with their behaviour in conventional GS (Figure 5.3). It was expected that drastic reductions in marker coverage would emphasize differences in predictive ability across traits with different genetic architectures as these parametric models each address marker variance and effect size uniquely. However, our results coincide with Zhang et al. (2023) who also observed no significant differences in prediction accuracy among the same six models in several GWAS-guided GS analyses. Considering the overall equivalent performance among parametric models, rrBLUP and GWAS-rrBLUP serve as representative models for conventional GS and GWAS-guided GS in the following comparisons.

Relative to conventional GS, the GWAS-guided GS analyses exhibited slightly (not significantly) increased prediction accuracies for all traits when informative markers were subset from both MS1 and MS2. Similarly small increases were exhibited using both MS1 and MS2. These small numeric differences may be attributed to the improved predictive power of GWAS-GS relative to conventional GS (Hickey et al. 2014; Montesinos-López et al. 2022b; Zhang et al. 2023). However, considering the high accuracies achieved by conventional GS for ERU, LLE, and LLN (> 0.75), the very small improvements exhibited with GWAS-guided GS may be attributed to the less complex genetic architecture of these traits perhaps approaching a prediction accuracy plateau.

When using MS3 for GWAS-GS analyses (i.e. top 200 most significant of 1,098 SNPs), nearly all prediction accuracies were slightly (not significantly) lower than conventional GS (Figure 5.3). This result was expected considering that GWAS-GS relies on capturing subsets of markers which provide more predictive power for phenotypic variation than the majority of markers which are unassociated to the trait (residing in unlinked haplotype blocks). In this example of a small marker set filtered by LD, only a very small proportion of markers (many of which represent entire haplotype blocks) would be considered in any way associated to a trait. Considering that the marker subsets included nearly 1/5 of all SNPs in MS3, many of the SNPs in those subsets were likely entirely unassociated to the traits. Thus, much of the noise excluded by subsets of MS1 and MS2 were not excluded with subsets of MS3 (Tan and Ingvarsson 2022). Again, the differences in prediction accuracy were slight – only a reduction of 0.055 for ERU and 0.020 for OIL – not a significant difference from conventional GS. However, as a credit to GWAS-GS, these prediction accuracies were based on supposedly poor training data: extremely small subsets of markers whose significance were determined by GWAS of only the inbred parent population, subsequently used to predict phenotypes of their hybrids. Considering that GWAS of inbred parents does not reveal hidden variables contributing to the heterosis of their progeny, the equivalently effective results produced by GWAS-GS may be considered for its use in *B. napus* improvement (Wang et al. 2017b).

Instead of relying solely on significant markers determined by a Bonferroni threshold, we chose to use the top 200 most significant markers from each GWAS analysis. This approach was taken for two reasons. First, focusing exclusively on significant SNPs can fail to explain a sufficient portion of genetic variation for a trait (i.e., the "missing heritability" problem) (Makowsky et al. 2011). This problem was especially prominent in preliminary tests for this study given that GWAS of parent genotypes yielded five or less significant markers for any one trait, whereas relaxing the significance threshold yielded wildly different numbers of markers for respective traits. Interestingly, marker subset approaches similar to this study have been shown to explain larger proportions of genetic variation than when using all SNPs, allowing improved prediction accuracies (Tan and Ingvarsson, 2022). Using this approach, sufficient heritability could be captured while also reducing noise and dimensionality. Second, this approach allowed prediction of all traits with the same number of markers – a balance that would not be possible using a significance threshold or percentage of markers from the different-sized marker sets.

As is the nature of GS, the effects of certain methods are nuanced. To further investigate GWAS-guided GS and move towards significant improvements, one might consider that Kaler et al. (2022) observed increased prediction accuracy for six traits across maize, rice, and soybean as the significance level of selected markers from GWAS became increasingly stringent to $p < 0.05$, but other studies in maize and *Eucalyptus* observe unimproved or reduced GS prediction accuracy when using significantly associated SNPs (Gowda et al. 2015; Tan and Ingvarsson 2022; Wallace et al. 2016). Kaler et al. (2022) suggests that significant markers should be selected up to a significance level where sufficient genomic coverage is maintained, whereas others suggest that the method of selecting markers depends largely on the trait of interest (Rice and Lipka 2019; Tan and Invgarsson 2022).

When selecting markers identified by GWAS for subsequent cross validation in the same population, the non-independence of experiments introduces significant bias, causing inflated prediction accuracies (McGowan et al. 2021). This study replicated a practical breeding scenario to demonstrate the true accuracy of GWAS-GS with minimal bias. Independence was maintained by separating parental genotypes from hybrids for the selection of informative markers, assuring uninflated results. The danger of biased approaches for plant breeders is similar to that of overfitting: predictions appear ideal with high accuracy in the current cross validation; however, such methods have not demonstrated their generalizability and may underperform relative to other methods in practical applications (McGowan et al. 2021; Zhang et al. 2023).

The results of this study support both GS and GWAS-guided GS as methods to improve the efficiency of practical breeding programs relative to conventional selection or MAS. Achieving effective prediction accuracies, not only with genome-wide marker sets (ranging from ~1,000 to 20,000+ SNPs) but also extremely small subsets of only 200 informative markers, lessens reliance on high-density genotyping. *B. napus*' slow LD decay rate contributes to its compatibility with lower marker coverage requirements in GS methodology (Desta and Ortiz 2014; Kaler et al. 2022). In this study, GWAS-guided GS provided consistent accuracies above 0.50, a threshold which represents potential advantages in genetic gain over conventional selection or MAS while simultaneously reducing phenotyping and genotyping costs (Heffner et al. 2010; Longin et al. 2015). The use of GEBVs (in both conventional and GWAS-guided GS) to inform selections against certain genotypes (which would otherwise require generations of phenotypic data) allows

resources to be allocated to more promising genotypes, therefore improving genetic gain per unit time, optimizing the volume of phenotyping/genotyping per breeding cycle, and potentially reducing the overall number of breeding cycles relative to conventional breeding or MAS (Habier et al. 2007; Heffner et al. 2010; Hickey et al. 2014; Jannink et al. 2010). However, when compared to conventional GS, the practical benefits of GWAS-guided GS are likely minimized due to three main reasons. Firstly, the identification of informative marker subsets based on GWAS still requires field phenotyping and genome-wide genotyping of the training population. Therefore, GWAS-guided GS provides no difference in cost-effectiveness regarding the training population. Secondly, informative marker subsets from one training population are not practical in long-term breeding programs. As LD blocks rearrange throughout breeding cycles and uncommon LD blocks are introduced with distantly related accessions, the predictive power of previously identified informative marker subsets decline, prompting the development of new training populations (DoVale et al. 2022; Norman et al. 2018). Thirdly, developments in genotyping technologies (e.g. *Brassica* 60K, PacBio, Oxford Nanopore) continually reduce the cost per marker, such that the choice between genome-wide or low-density genotyping is insignificant compared to costs associated with field phenotyping (Clarke et al. 2016; Magdy et al. 2020; Rhoads and Au 2015). Although results indicate GWAS-guided GS could approach improved prediction accuracies, the actual advantages (granted towards genetic gain, overall cost, environmental impact of field seasons, etc.) are unclear when considering other practical implications of this method.

5.6 Conclusion

This study demonstrated that GWAS-guided GS is not likely to replace conventional GS methods. By identifying small subsets of informative markers from GWAS of respective traits, several GWAS-guided GS analyses exhibited slight differences in prediction accuracy relative to conventional GS (models trained using genome-wide marker sets). Although numerically distinguishable, accuracy improvements were not statistically significant when selecting small subsets from MS1 or MS2 (high- or intermediate-density marker sets), meanwhile slight decreases in accuracy were also not significantly different from conventional GS when using MS3 (low-density, LD-filtered marker set). All six parametric models produced statistically equivalent results, as was seen in conventional GS. Furthermore, prediction accuracy response to individual traits followed an identical pattern in conventional GS and GWAS-guided GS, suggesting that the

marker subsetting approach used in this study sufficiently captured the heritability of all five seed quality traits. Barriers to general adoption of GWAS-guided GS (over conventional GS) include insignificant prediction accuracy improvements, impracticality for long-term breeding programs due to rearranging LD blocks, and the availability of cost-effective genotyping technologies which minimize the benefit of low-density genotyping.

6. GENERAL DISCUSSION

Rapeseed (*Brassica napus* L.) is now the second-largest oilseed commodity in the world, behind only soybean (USDA 2023). Canada developed canola in the 1970s and continues to drive the industry as principal producer and exporter of seed and oil (Barthet 2016; Canola Council of Canada 2024b). Conventional selection and early advancements in MAS have been major contributors to *B. napus* genetic improvements (Huang et al. 2012; Morrison 2016). However, global food demand is projected to increase as production shifts towards sustainability, demanding more from less. Canada's demand for canola is estimated to be 26 million metric tonnes by 2025 (Canola Council of Canada 2014). This demand is more complex than just seed yield, as the applications of *B. napus* have evolved to include niches in seed quality such as high protein meal, low polyunsaturated fatty acid content, and high erucic acid rapeseed (HEAR). Simple conventional selection and MAS will struggle to sustainably produce cultivars which promptly respond to consumers' demands and farmers' production challenges, encouraging the adoption of new selection methods. This research applied two bioinformatics methods, GWAS and GS, to HEAR breeding scenarios, demonstrating their potential to enhance genetic gain.

Seed quality traits in this *B. napus* breeding population exhibited significant correlations between OLE, ERU, and LLE, reflecting their involvement in biosynthetic pathways as described in previous studies (Katavic et al. 2001; Sasongko et al. 2005; Tang et al. 2019). Interestingly, OIL was not significantly correlated to proportions of any of the four specific fatty acids, demonstrating the highly complex nature of oil accumulation which is influenced by many minor-effect genes and genotype × environment interactions (Delourme et al. 2006; Jiang et al. 2014; Zhao et al. 2007). These relationships are critical for plant breeders seeking to improve quality traits while minimizing linkage drag and introgression of unwanted traits from distant accessions. A prime example of such a relationship is the ACP5 gene: identified near a peak QTL for OIL, this gene is known to alter the fatty acid profile in response to salt stress (Huang et al. 2017).

Several QTL identified in this study coincide with QTL for the same traits in previous studies (Table 3.9). Furthermore, candidate genes for fatty acid biosynthesis identified in previous studies also coincide with significant regions detected here (Table 3.10). These candidate genes were identified across 11 *B. napus* chromosomes, including WSD1: an enzyme synthesizing both wax esters and triacylglycerols with orthologs in *Arabidopsis thaliana*, peanuts (*Arachis hypogaea*),

and oleaginous yeast (Li et al. 2008; Rani et al. 2013; Saha et al. 2006; Turchetto-Zolet et al. 2016). Accordance with QTL in literature provides added assurance to the significance of novel associations. The most significant QTL detected in this study, LLE.C04.1, was located in close proximity to two highly significant QTL for OLE and ERU (ERU.C04.1 was the third-most significant QTL detected overall). A candidate gene for fatty acid biosynthesis has not yet been identified in this region, warranting further investigation to dissect the genetic control of such significant effects on these three important seed quality traits.

A total of 24 unique GWAS analyses for each of the five seed quality traits compared various population designs, marker sets, and models. Across the five traits, 89 peak QTL were identified by FarmCPU. The remaining GWAS models also detected significant SNPs; however, were not as consistently reliable in controlling false positives and negatives. Population design most clearly emphasized performance differences among models. Among all QTL identified by FarmCPU, 79 % were identified in the larger combined population (Tables S3.4 - S3.6) (Bento 2024a). In another study of *B. napus* oil content, Wang et al. (2013) observed improved QTL discovery and more accurate effect estimations while using larger, more diverse populations like the combined population in this study; the power to detect alleles with heterotic effects is greater in large hybrid populations than populations of their more homozygous parents (Wang et al. 2017b). FarmCPU was particularly effective in accurately identifying QTL while controlling false positives and negatives. FarmCPU is classified as a multi-locus GWAS (ML-GWAS) model which is designed to estimate all marker-effects simultaneously, providing higher statistical power and reliability against false positive/negative signals than single-locus GWAS (SL-GWAS) models which consider each marker as an individual test (Vikas et al. 2022; Wang et al. 2016). This study compared two ML-GWAS models (FarmCPU, MLMM) and four SL-GWAS models (CMLM, MLM+K, MLM+K+PCA, GenABEL P+K). In the parent population, SL-GWAS models (besides GenABEL P+K) exhibited false negatives (Figures S3.6 - S3.10). Some SL-GWAS analyses are known to perform overly stringent corrections, failing to detect truly significant loci (false negatives) (Khan et al. 2021). Alternatively, consistent false positives were observed in SL-GWAS analyses of the combined population, whereas ML-GWAS models and GenABEL P+K generally exhibited consistent control. One possible explanation for the prevalence of false positives by SL-GWAS may be *B. napus*' slow LD decay rate: approaching marker-trait associations by calculating one marker at a time may falsely detect significant associations to the hundreds of genes located

within extremely large haplotype blocks, whereas this issue is mitigated by ML-GWAS (Kaler et al. 2020; Khan et al. 2021; Liu et al. 2022). Our results especially emphasize FarmCPU's superiority when comparing pooled results across all 120 GWAS analyses. Despite making up only 133 of the 1400 significant SNPs overall, FarmCPU constituted nearly half (78/161) of the peak QTL, demonstrating its ability to identify singular and highly significant SNPs closest to the causal locus (Kaler et al. 2020).

The prevalence of false positives in SL-GWAS could not simply be attributed to marker density nor rare variants present in MS1. Overwhelming proportions of markers deviated above the expected in both MS2 (intermediate density, not containing markers with MAF < 0.05) and MS1 (high density, containing markers with MAF < 0.05). If rare variants were responsible for this behaviour, then such patterns would only be appear in the MS1 analyses. Similarly, false positives due to extreme marker density would only appear in MS1 analyses, but they are also prevalent in MS2 analyses. These observations (and simultaneously strong performance of ML-GWAS and GenABEL P+K) indicate that poorly controlled false positives and negatives were not attributed to marker density or population composition, but instead SL-GWAS models themselves.

ML-GWAS approaches provide flexible selection criteria to reduce false positive and negative signals while investigating complex traits. However, SL-GWAS approaches show proven effectiveness in detecting major associations with less computationally demanding correction (Alseekh et al. 2021; Khan et al. 2019; Wang et al. 2016; Wen et al. 2017). By leveraging both approaches, GWAS for *B. napus* improvement could benefit from identifying consensus among results, thereby more comprehensive understandings of genetic architecture (Khan et al. 2021). Additionally, several population designs exist for plant breeders to improve QTL discovery and lessen the spurious effects of population structure (Liu et al. 2022; Dell'Acqua et al. 2015; Wang et al. 2017b; Xiao et al. 2017).

The findings from Chapter 3 demonstrate how marker-based models rely strongly on the allelic variation provided to them. This variety may be achieved by increasing population size/diversity; however, no clear benefit arises by increasing marker density above the LD decay threshold. These factors are crucial for a breeder's consideration as they relate to phenotyping costs, genotyping costs, and computational resources for both GWAS (Chapter 3) and GS (Chapter 4).

In contrast to traditional MAS, which informs selections based only on major-effect QTL, GS estimates major- and minor-effect QTL to calculate GEBVs (Krishnappa et al. 2021). The use of GEBVs to inform selections (which would otherwise require multi-year, multi-location phenotypic data) is a major benefit to plant breeders, facilitating the allocation of resources to desirable germplasm and potentially lessening the number of breeding cycles (Heffner et al. 2010; Jubair and Domaratzki 2023; Krishnappa et al. 2021; Varshney et al. 2017). The application of GS in Chapter 4 serves to evaluate its potential effectiveness for *B. napus* improvement. The broad spectrum of analyses were designed to simulate practical breeding scenarios – a breeder may elect to use any combination of training population or validation population (TP/VP) designs, marker densities, or models surveyed in this experiment. Overall, we demonstrated that GS can provide highly accurate predictions of seed quality traits, identifying analyses which accurately predicted all five traits above 0.70 (i.e. correlation of predictions to actual phenotypes). Overall oil content was considered the most complex trait in this study, difficult to predict as a consequence of many minor-effect QTL potentially affected further by environment (Delourme et al. 2006; Hu et al. 2021b; Jiang et al. 2014; Koscielny et al. 2019; Zhao et al. 2007). By combining different TP/VP designs and models, OIL prediction accuracy ranged from negative values to as high as 0.745 (TP/VP design 5 input to BayesC). It was TP/VP design which most clearly affected the increases or decreases in OIL prediction accuracy. Each trait responded significantly to TP/VP design regardless of marker set and choice of parametric model, meanwhile ML models exhibited similar responses to TP/VP design among themselves.

This study compared the effects of 5 unique TP/VP designs on GS prediction accuracy, each inputting different proportions of parent genotypes, hybrid genotypes, or random samples of all genotypes (design 1: TP = 100 % of parent genotypes, VP = 20 % of hybrid genotypes; design 2: TP = 100 % of parent genotypes, VP = 60 % of hybrid genotypes; design 3: TP = 100 % of parent genotypes, VP = 100 % of hybrid genotypes; design 4: TP = 20 % of random genotypes, VP = 80 % of random genotypes; and design 5: TP = 60 % of random genotypes, VP = 40 % of random genotypes). Despite the similarity between designs 4 and 5 both incorporating parents and hybrids in model training (incorporating relatedness between TP/VP), the larger TP in design 5 produced the greatest prediction accuracy among nearly all traits and models, whereas the smaller TP in TP/VP design 4 often produced poorer accuracies than other designs which only trained models using parents. Incorporating TP/VP relatedness may improve prediction accuracy; however, TP

size must also remain sufficiently large to capture the diversity of haplotypic combinations possible within parents and hybrids (Daetwyler et al. 2008; Desta and Ortiz 2014; Werner et al. 2020). Although TP/VP designs 1-4 contained similarly small TP sizes, TP/VP design 4 likely excluded one or both of a hybrid's parents from the TP across many iterations of its random sampling, demonstrating poor predictive ability. Hu et al. (2021b) also observed declines in prediction accuracy when model training incorporated neither parent of *B. napus* hybrids versus when incorporating one or both parents. Although TP/VP designs 1-3 did not incorporate any hybrids in training, they incorporated both parents of all predicted hybrids, producing generally good prediction accuracies.

Of the factors compared in this study, marker set density was least significant to GS prediction accuracy. As long as marker set density is developed in accordance to LD blocks, sufficient linkage to causal loci is available for GS models to estimate marker effects on traits (Figure 4.3). Kaler et al. (2022) described this as a haplotype block performing similar to a single marker, where adding more markers does not improve accuracy but instead can increase error (overfitting). The ability to produce reliable predictions with low-density marker sets may be promising for plant breeders looking to reduce genotyping costs and computational burden.

All six parametric models performed statistically equivalently, maintaining more reliable prediction accuracies than the more computationally demanding ML models. The ML models used in this study either performed similarly (SVR) or poorer (RF and XGB) to parametric models (Figure 4.6). A potential explanation for the generally poor performance of ML in this study may be TP sizes that are too small for sufficient machine learning (Montesinos-Lopez et al. 2021; Sandhu et al. 2021b). Despite exponentially growing interest in ML and deep learning, our results demonstrate how simpler parametric GS models can provide relatively accurate and reliable predictions for marker-based improvement of complex seed quality traits (Hu et al. 2021b; Kaler et al. 2022; Zou et al. 2016, Merrick and Carter 2021).

Breeders may also consider GWAS-guided GS, a technique proposed to improve GS accuracy by training models on optimal selections of informative markers. These small marker subsets aim to mitigate model overfitting and underfitting while reducing computational burden (Hickey et al. 2014; Montesinos-López et al. 2022b; Zhang et al. 2023). Our application of GWAS-guided GS to *B. napus* improvement did not exhibit significantly improved prediction accuracies relative to

conventional GS. When 200 informative markers were subset from GWAS results of MS1 and MS2, the GWAS-guided GS prediction accuracies for each trait only slightly increased numerically relative to conventional GS. With MS3, small numeric decreases in prediction accuracy were observed when using the top 200 markers of the low-density marker set. This difference among marker sets could be expected considering that GWAS- guided GS relies on capturing informative markers with more predictive power than the majority of markers which are unassociated to the trait – the 200-marker subsets incorporated nearly 1/5 of SNPs in MS3, many of which were likely unassociated to the traits. In Chapters 3 and 4, we discussed how GWAS of inbred parents has difficulty in revealing hidden the variables which contribute to the heterosis of their progeny. As a credit to the GWAS-guided GS method, effective prediction accuracies of hybrid traits were achieved based on supposedly poor training data: extremely small subsets of markers whose significance were determined by GWAS of only the inbred parental population (Hu et al. 2021b; Wang et al. 2017b). This result demonstrates the potential benefits of GWAS-guided GS relative to conventional breeding or MAS. While GWAS-guided GS offers theoretical advantages, its practical benefits relative to conventional GS are limited by the continued need for costly phenotyping and genotyping in new training populations, the declining predictive power of marker subsets due to LD block rearrangements throughout breeding cycles, and advancements in genotyping technologies that diminish cost differences between genome-wide and low-density approaches (Clarke et al. 2016; DoVale et al. 2022; Magdy et al. 2020; Norman et al. 2018; Rhoads and Au 2015). Broad integration of GWAS-guided GS in *B. napus* breeding is unlikely considering it does not significantly outperform conventional GS in cost effectiveness or long-term practicality.

This research demonstrates the combined strengths of GWAS and GS. Informing genetic gains while reducing reliance on field resources and genotypic data draws a direct connection to the overarching objective in agriculture to produce more with less. Several QTL for each seed quality trait were identified through GWAS, including novel QTL and those in accordance with previous literature. Effective prediction accuracies were produced using both GS and GWAS-guided GS methods, providing potential genetic gain advantages over conventional selection or MAS. These results encourage the adoption and further development of selection methods to drive *B. napus* improvement amidst future challenges.

7. FUTURE RESEARCH RECOMMENDATIONS

While the results of this research are promising for the implementation of GWAS and GS in *B. napus* improvement, they also demonstrate the subjectivity of marker-based modelling. Many methods are available to improve GWAS and GS effectiveness; however, such methods remain largely unexplored in application to *B. napus*. Further research may progress towards standard operating procedures which are effective in most plant breeding scenarios.

Regarding GWAS analyses, greater genetic information could be extracted by conducting GWAS based on next-generation sequencing data (e.g. including rare variants). Results from this study also demonstrate how consensus among SL-GWAS and ML-GWAS models can improve the reliability of GWAS results. Future research in *B. napus* improvement should integrate both approaches, leveraging the strengths of SL-GWAS and ML-GWAS to enhance the reliability of GWAS results, thereby offering more comprehensive understandings of genetic architecture. Additionally, several QTL identified in Chapter 3 could be further investigated using BLAST, transcriptomics, or gene knockout to understand their role in oil and fatty acid profile traits.

Given that each of the four fatty acids were significantly correlated with each other, leveraging data from these traits could further improve their prediction accuracy. Follow-up research on *B. napus* GS could explore multi-trait GS models to enhance prediction accuracy for complex traits that are also correlated to seed quality traits. To buffer the effects of selection and drift on LD between QTL throughout breeding cycles, additional research could investigate strategies of updating TP data to maintain or improve GS accuracy throughout cycles of *B. napus* breeding.

Future applications of GWAS-guided GS should focus on simulating other practical *B. napus* breeding scenarios. For example, including parents and preliminary hybrids in training, replicated with both groups, only hybrids, or random proportions of either group for independent GWAS analyses and marker subsetting. Future GWAS-guided GS research may also coincide with conventional GS, incorporating methods such as multi-trait models, environmental data, high throughput phenotyping, or advanced TP updating strategies to maintain GS prediction accuracy while reducing costs for large-scale breeding programs.

8. REFERENCE MATTER

8.1 Literature cited

Aalborg, T., Sverrisdóttir, E., Kristensen, H. T., & Nielsen, K. L. (2024). The effect of marker types and density on genomic prediction and GWAS of key performance traits in tetraploid potato. *Frontiers in Plant Science*, *15*, 1340189. https://doi.org/10.3389/fpls.2024.1340189

Abbadi, A., & Leckband, G. (2011). Rapeseed breeding for oil content, quality, and sustainability. *European Journal of Lipid Science and Technology*, *113*(10), 1198–1206. https://doi.org/10.1002/ejlt.201100063

Adkins, D., & Shabalin, A. (2023). GWAS Post Processing. *Introduction to Statistical Genetics*.

AgMRC. (2022). Canola profile. Iowa State University. https://www.agmrc.org/commodities-products/grains-oilseeds/canola-profile

Ahmad, N., Su, B., Ibrahim, S., Kuang, L., Tian, Z., Wang, X., Wang, H., & Dun, X. (2022). Deciphering the genetic basis of root and biomass traits in rapeseed (*Brassica napus* L.) through the integration of GWAS and RNA-seq under nitrogen stress. *International Journal of Molecular Sciences*, 23(14), 7958. https://doi.org/10.3390/ijms23147958

Ahmadi, N. (2022). Genetic bases of complex traits: From quantitative trait loci to prediction. In N. Ahmadi & J. Bartholomé (Eds.), *Genomic Prediction of Complex Traits* (Vol. 2467, pp. 1–44). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_1

Ahmar, S., Gill, R. A., Jung, K.-H., Faheem, A., Qasim, M. U., Mubeen, M., & Zhou, W. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: Recent advances and future outlook. *International Journal of Molecular Sciences*, *21*(7), 2590. https://doi.org/10.3390/ijms21072590

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. https://doi.org/10.48550/ARXIV.1907.10902

Alcock, T. D., Havlickova, L., He, Z., Bancroft, I., White, P. J., Broadley, M. R., & Graham, N. S. (2017). Identification of candidate genes for calcium and magnesium accumulation in *Brassica napus* L. by association genetics. *Frontiers in Plant Science*, 8, 1968. https://doi.org/10.3389/fpls.2017.01968

Ali, M. (2021). LD decay plot from Tassel LD output. GitHub. https://github.com/bikashpoudel123/mohsinali1990/blob/main/LD%20decay%20Plot%20from%20TASSEL%20LDoutput.R

Alqudah, A. M., Sallam, A., Stephen Baenziger, P., & Börner, A. (2020). GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from

Barley – A review. *Journal of Advanced Research*, *22*, 119–135. https://doi.org/10.1016/j.jare.2019.10.013

Alseekh, S., Kostova, D., Bulut, M., & Fernie, A. R. (2021). Genome-wide association studies: Assessing trait characteristics in model and crop plants. *Cellular and Molecular Life Sciences*, *78*(15), 5743–5754. https://doi.org/10.1007/s00018-021-03868-w

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., Pérez-Rodríguez, P., & Lopez-Cruz, M. A. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *The Crop Journal*, *8*(5), 745–756. https://doi.org/10.1016/j.cj.2020.03.010

Aoki, F., Bennet, J., Dolin, R., & Blaser, M. (2015). In *Mandell, Douglas, and Bennett's principles and practice of infectious diseases* (8th ed., Vol. 1, pp. 546–562). essay, Elsevier/Saunders.

Aukema, H., Campbell, L., Daun, J. K., Eskin, M. N. A., & Hickling, D. (2011). Chapter 9 - Oil Nutrition and Utilization. In *Canola chemistry, production, processing, and Utilization* (pp. 245–280). AOCS Press.

Aulchenko, Y., Ripke, S., Isaacs, A., van Duijn, C. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 15;23(10),1294-6. doi: 10.1093/bioinformatics/btm108

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., & Nordborg, M. (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, *1*(5). https://doi.org/10.1371/journal.pgen.0010060

Awad, M., & Khanna, R. (2015). Support vector regression. In M. Awad & R. Khanna, *Efficient Learning Machines* (pp. 67–80). Apress. https://doi.org/10.1007/978-1-4302-5990-9_4

Bährle-Rapp, M. (2007). Behenic acid. *Springer Lexikon Kosmetik Und Körperpflege*, 61–61. https://doi.org/10.1007/978-3-540-71095-0_1046

Bao, Y., Kurle, J. E., Anderson, G., & Young, N. D. (2015). Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Molecular Breeding*, *35*(6), 128. https://doi.org/10.1007/s11032-015-0324-3

Barthet, V. J. (2016). Canola: Overview. *Encyclopedia of Food Grains*, 237–241. https://doi.org/10.1016/b978-0-12-394437-5.00029-2

Barthet, V. (2022). Quality of western Canadian canola. Canadian Grain Commission. https://publications.gc.ca/collections/collection_2023/ccg-cgc/A92-14-2022-eng.pdf

Bastiaansen, J. W., Coster, A., Calus, M. P., Van Arendonk, J. A., & Bovenhuis, H. (2012). Long-term response to genomic selection: Effects of estimation method and reference

population structure for different genetic architectures. *Genetics Selection Evolution*, *44*(1), 3. https://doi.org/10.1186/1297-9686-44-3

Bayer, P. E., Hurgobin, B., Golicz, A. A., Chan, C. K., Yuan, Y., Lee, H., Renton, M., Meng, J., Li, R., Long, Y., Zou, J., Bancroft, I., Chalhoub, B., King, G. J., Batley, J., & Edwards, D. (2017). Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnology Journal*, 15(12), 1602–1610. https://doi.org/10.1111/pbi.12742

Beil, C. T., Anderson, V. A., Morgounov, A., & Haley, S. D. (2019). Genomic selection for winter survival ability among a diverse collection of facultative and winter wheat genotypes. *Molecular Breeding*, *39*(2), 29. https://doi.org/10.1007/s11032-018-0925-8

Belamkar, V., Guttieri, M. J., Hussain, W., Jarquín, D., El-basyoni, I., Poland, J., Lorenz, A. J., & Baenziger, P. S. (2018). Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3 Genes|Genomes|Genetics*, *8*(8), 2735–2747. https://doi.org/10.1534/g3.118.200415

Bell, J. M. (1982). From rapeseed to canola: A brief history of research for superior meal and edible oil. *Poultry Science*, 61(4), 613–622. https://doi.org/10.3382/ps.0610613

Bento, J. (2024a). Supplementary Tables S3.4 - S3.6 FarmCPU QTL. *Borealis*. https://doi.org/10.5683/SP3/XBNFLT

Bento, J. (2024b). Supplementary Tables S3.7 - S3.9 All QTL. *Borealis*. https://doi.org/10.5683/SP3/YNFPS1

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*, *48*(5), 1649–1664. https://doi.org/10.2135/cropsci2008.03.0131

Bernardo, A., Wang, S., St. Amand, P., & Bai, G. (2015). Using next generation sequencing for multiplexed trait-linked markers in wheat. *PLOS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0143890

Beukert, U., Li, Z., Liu, G., Zhao, Y., Ramachandra, N., Mirdita, V., Pita, F., Pillen, K., & Reif, J. C. (2017). Genome-based identification of heterotic patterns in rice. *Rice*, *10*(1), 22. https://doi.org/10.1186/s12284-017-0163-4

Beversdorf, W. D., & Kott, L. S. (1987). Development of triazine resistance in crops by classical plant breeding. *Weed Science*, *35*(S1), 9–11. https://doi.org/10.1017/s0043174500060975

Beyene, Y., Gowda, M., Olsen, M., Robbins, K. R., Pérez-Rodríguez, P., Alvarado, G., Dreher, K., Gao, S. Y., Mugo, S., Prasanna, B. M., & Crossa, J. (2019). Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. *Frontiers in Plant Science*, *10*, 1502. https://doi.org/10.3389/fpls.2019.01502

Bhardwaj, H. L., & Hamama, A. A. (2000). Oil, erucic acid, and glucosinolate contents in winter hardy rapeseed germplasms. *Industrial Crops and Products*, *12*(1), 33–38. https://doi.org/10.1016/S0926-6690(99)00043-6

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://doi.org/10.3389/fgene.2016.00221

Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., Pardo, M. F., Lanaro, V., Sayas, M., & Castro, A. J. (2020). Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (*Hordeum vulgare* L.). *G3 Genes|Genomes|Genetics*, *10*(3), 1113–1124. https://doi.org/10.1534/g3.119.400968

Bhering, L. L., Junqueira, V. S., Peixoto, L. A., Cruz, C. D., & Laviola, B. G. (2015). Comparison of methods used to identify superior individuals in genomic selection in plant breeding. *Genetics and Molecular Research*, *14*(3), 10888–10896. https://doi.org/10.4238/2015.September.9.26

Bhinder, G., Sharma, S., Kaur, H., Akhatar, J., Mittal, M., & Sandhu, S. (2022). Genomic regions associated with seed meal quality traits in *Brassica napus* germplasm. *Frontiers in Plant* Science, 13, 882766. https://doi.org/10.3389/fpls.2022.882766

Bian, Y., & Holland, J. B. (2017). Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity*, *118*(6), 585–593. https://doi.org/10.1038/hdy.2017.4

Borrenpohl, D., Huang, M., Olson, E., & Sneller, C. (2020). The value of early-stage phenotyping for wheat breeding in the age of genomic selection. *Theoretical and Applied Genetics*, *133*(8), 2499–2520. https://doi.org/10.1007/s00122-020-03613-0

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

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Browse, J., Spychalla, J., Okuley, J., & Lightner, J. (1998). Altering the fatty acid composition of vegetable oils. In *Plant Lipid Biosynthesis: Fundamentals and Agricultural Applications* (pp. 131–153). essay, Cambridge University Press.

Brzyski, D., Peterson, C. B., Sobczyk, P., Candès, E. J., Bogdan, M., & Sabatti, C. (2017). Controlling the rate of GWAS false discoveries. *Genetics*, *205*(1), 61–75. https://doi.org/10.1534/genetics.116.193987

Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., Kumar, U., Kumar, R. R., Juliana, P., Mishra, D. C., & Kumar, S. (2022). Genomic selection: A tool for accelerating the efficiency of molecular breeding for development of climate-

resilient crops. *Frontiers in Genetics*, *13*, 832153. https://doi.org/10.3389/fgene.2022.832153

Calus, M. P. L. (2010). Genomic breeding value prediction: Methods and procedures. *Animal*, *4*(2), 157–164. https://doi.org/10.1017/S1751731109991352

Canola Council of Canada (2014) Keep it coming 2025. https://www.canolacouncil.org/about-us/strategic-plan/

Canola Council of Canada (2020) How much fertilizer does canola need?

Canola Council of Canada. (2024a) Economic impact of canola. https://www.canolacouncil.org/about-canola/economic-impact/

Canola Council of Canada. (2024b). Canadian canola export statistics. https://www.canolacouncil.org/markets-stats/exports/

Canola Council of Canada. (2024c). Canola cultivar traits. https://www.canolacouncil.org/canola-encyclopedia/history-of-canola-seed-development/canola-seed-traits/

Canola Council of Canada. (2024d). History of canola seed development: Canola encyclopedia. https://www.canolacouncil.org/canola-encyclopedia/history-of-canola-seed-development/

Canadian Food Inspection Agency. (1989). Registered Variety: HYOLA 40. https://inspection.canada.ca/active/netapp/regvar/regvare.aspx?id=1118

Canadian Oilseed Processors Association. (2023). Crush & Oil & Meal Production. https://copacanada.com/crush-oil-meal-production/

Cantila, A. Y., Thomas, W. J. W., Saad, N. S. M., Severn-Ellis, A. A., Anderson, R., Bayer, P. E., Edwards, D., Van De Wouw, A. P., & Batley, J. (2023). Identification of candidate genes for LepR1 resistance against *Leptosphaeria maculans* in *Brassica napus*. *Frontiers in Plant Science*, 14, 1051994. https://doi.org/10.3389/fpls.2023.1051994

Cavanagh, C., Morell, M., Mackay, I., & Powell, W. (2008). From mutations to magic: Resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*, *11*(2), 215–221. https://doi.org/10.1016/j.pbi.2008.01.002

Ćeran, M., Miladinović, D., Đorđević, V., Trkulja, D., Radanović, A., Glogovac, S., & Kondić-Špika, A. (2024). Genomics-assisted speed breeding for crop improvement: Present and future. *Frontiers in Sustainable Food Systems*, *8*, 1383302. https://doi.org/10.3389/fsufs.2024.1383302

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., Falentin, C., Koh, C. S., Le Clainche, I., 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

Chao, W. S., Horvath, D. P., Stamm, M. J., & Anderson, J. V. (2021). Genome-wide association mapping of freezing tolerance loci in canola (*Brassica napus* L.). *Agronomy*, 11(2), 233. https://doi.org/10.3390/agronomy11020233

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. https://doi.org/10.48550/ARXIV.1603.02754

Chen, L., Wan, H., Qian, J., Guo, J., Sun, C., Wen, J., Yi, B., Ma, C., Tu, J., Song, L., Fu, T., & Shen, J. (2018). Genome-wide association study of cadmium accumulation at the seedling stage in rapeseed (*Brassica napus* L.). *Frontiers in Plant Science*, 9, 375. https://doi.org/10.3389/fpls.2018.00375

Chen, Z.-Q., Klingberg, A., Hallingbäck, H. R., & Wu, H. X. (2023). Preselection of QTL markers enhances accuracy of genomic selection in Norway spruce. *BMC Genomics*, *24*(1), 147. https://doi.org/10.1186/s12864-023-09250-3

Cheung, W. Y., Gugel, R. K., & Landry, B. S. (1998). Identification of RFLP markers linked to the white rust resistance gene (Acr) in mustard (*Brassica juncea L.*) czern. and Coss.). *Genome*, *41*(4), 626–628. https://doi.org/10.1139/g98-043

Clark, S. A., Hickey, J. M., & Van Der Werf, J. H. (2011). Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution*, *43*(1), 18. https://doi.org/10.1186/1297-9686-43-18

Clarke, W. E., Higgins, E. E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., Batley, J., Edwards, D., Meng, J., Li, R., Lawley, C. T., Pauquet, J., Laga, B., Cheung, W., Iniguez-Luy, F., Dyrszka, E., Rae, S., Stich, B., Snowdon, R. J., Parkin, I. A. (2016). A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theoretical and Applied Genetics*, *129*(10), 1887–1899. https://doi.org/10.1007/s00122-016-2746-7

Collard, B. C., Jahufer, M. Z., Brouwer, J. B., & Pang, E. C. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for Crop Improvement: The basic concepts. *Euphytica*, *142*(1–2), 169–196. https://doi.org/10.1007/s10681-005-1681-5

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

Cobb, J. N., Biswas, P. S., & Platten, J. D. (2019). Back to the future: Revisiting MAS as a tool for modern plant breeding. *Theoretical and Applied Genetics*, 132(3), 647–667. https://doi.org/10.1007/s00122-018-3266-4

Combs, E., & Bernardo, R. (2013). Accuracy of genome-wide selection for different traits with constant population size, heritability, and number of markers. *The Plant Genome*, 6(1), plantgenome2012.11.0030. https://doi.org/10.3835/plantgenome2012.11.0030

Canola Info (n.d.). Comparison of dietary fats https://www.canolainfo.org/quadrant/media/files/downloads/pdfs/DietaryFatChartprint2016.pdf

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

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

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., & Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *The Plant Genome*, *11*(1), 170043. https://doi.org/10.3835/plantgenome2017.05.0043

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

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, *3*(10), e3395. https://doi.org/10.1371/journal.pone.0003395

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, *185*(3), 1021–1031. https://doi.org/10.1534/genetics.110.116855

Dakouri, A., Lamara, M., Karim, Md. M., Wang, J., Chen, Q., Gossen, B. D., Strelkov, S. E., Hwang, S.-F., Peng, G., & Yu, F. (2021). Identification of resistance loci against new pathotypes of Plasmodiophora *Brassicae* in *Brassica napus* based on genome-wide association mapping. *Scientific Reports*, 11(1), 6599. https://doi.org/10.1038/s41598-021-85836-9

Daun, J. K., Eskin, N. A. M., & Hickling, D. (Eds.). (2011). *Canola: Chemistry, production, processing, and utilization*. AOCS Press.

De Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K., & Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, *182*(1), 375–385. https://doi.org/10.1534/genetics.109.101501

de Roos, A. P. W., Hayes, B. J., & Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, *183*(4), 1545–1553. https://doi.org/10.1534/genetics.109.104935

DeClercq DR, Daun JK, Tipples K (1998) Quality of western Canadian canola. Canadian Grain Commission

Delourme, R., Falentin, C., Huteau, V., Clouet, V., Horvais, R., Gandon, B., Specel, S., Hanneton, L., Dheu, J. E., Deschamps, M., Margale, E., Vincourt, P., & Renard, M. (2006). Genetic control of oil content in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics*, *113*(7), 1331–1345. https://doi.org/10.1007/s00122-006-0386-z

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in Zea mays. *Genome Biology*, *16*(1), 167. https://doi.org/10.1186/s13059-015-0716-z

Deng, M., Li, D., Luo, J., Xiao, Y., Liu, H., Pan, Q., Zhang, X., Jin, M., Zhao, M., & Yan, J. (2017). The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnology Journal*, *15*(10), 1250–1263. https://doi.org/10.1111/pbi.12712

Derbyshire, M. C., Khentry, Y., Severn-Ellis, A., Mwape, V., Saad, N. S. M., Newman, T. E., Taiwo, A., Regmi, R., Buchwaldt, L., Denton-Giles, M., Batley, J., & Kamphuis, L. G. (2021). Modeling first order additive × additive epistasis improves accuracy of genomic prediction for sclerotinia stem rot resistance in canola. *The Plant Genome*, *14*(2), e20088. https://doi.org/10.1002/tpg2.20088

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: Genome-wide prediction in plant improvement. *Trends in Plant Science*, 19(9), 592–601. https://doi.org/10.1016/j.tplants.2014.05.006

Dong, H., Wang, R., Yuan, Y., Anderson, J., Pumphrey, M., Zhang, Z., & Chen, J. (2018). Evaluation of the potential for genomic selection to improve spring wheat resistance to fusarium head blight in the pacific northwest. *Frontiers in Plant Science*, *9*, 911. https://doi.org/10.3389/fpls.2018.00911

Dong, Z., Tang, M., Cui, X., Zhao, C., Tong, C., Liu, Y., Xiang, Y., Li, Z., Huang, J., Cheng, X., & Liu, S. (2022). Integrating GWAS, linkage mapping and gene expression analyses reveal the genetic control of first branch height in *Brassica napus* L. *Frontiers in Plant Science*, 13, 1080999. https://doi.org/10.3389/fpls.2022.1080999

DoVale, J. C., Carvalho, H. F., Sabadin, F., & Fritsche-Neto, R. (2022). Genotyping marker density and prediction models effects in long-term breeding schemes of cross-pollinated crops. *Theoretical and Applied Genetics*, *135*(12), 4523–4539. https://doi.org/10.1007/s00122-022-04236-3

Downey, R. K., & Rimmer, S. R. (1993). Agronomic improvement in oilseed Brassicas. In *Advances in Agronomy* (Vol. 50, pp. 1–66). Elsevier. https://doi.org/10.1016/S0065-2113(08)60831-7

Dreisigacker, S., Pérez-Rodríguez, P., Crespo-Herrera, L., Bentley, A. R., & Crossa, J. (2023). Results from rapid-cycle recurrent genomic selection in spring bread wheat. *G3: Genes, Genomes, Genetics*, *13*(4), jkad025. https://doi.org/10.1093/g3journal/jkad025

Drucker, H., Burges, C., Kaufma, L., & Vapnik, V. (1996). Support Vector Regression Machines. In *Advances in Neural Information Processing Systems*.

Du, H., Raman, H., Kawasaki, A., Perera, G., Diffey, S., Snowdon, R., Raman, R., & Ryan, P. R. (2022). A genome-wide association study (GWAS) identifies multiple loci linked with the natural variation for Al. *Functional Plant Biology*, 49(10), 845–860. https://doi.org/10.1071/FP22073

Duan, X., Wang, X., Jin, K., Wang, W., Liu, H., Liu, L., Zhang, Y., Hammond, J. P., White, P. J., Ding, G., Xu, F., & Shi, L. (2021). Genetic dissection of root angle of *Brassica napus* in response to low phosphorus. *Frontiers in Plant Science*, 12, 697872. https://doi.org/10.3389/fpls.2021.697872

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., & Mangin, B. (2017). Genomic selection for yield and seed protein content in soybean: A study of breeding program data and assessment of prediction accuracy. *Crop Science*, *57*(3), 1325–1337. https://doi.org/10.2135/cropsci2016.06.0496

Duncan, R. W., McVetty, P. B. E., Nugent-Rigby, J. A., Fernando, W. G. D., & Li, G. (2020). Evolve Roundup Ready high erucic acid, low glucosinolate hybrid summer rape. *Canadian Journal of Plant Science*, *101*(1), 140–142. https://doi.org/10.1139/cjps-2019-0331

Dunckel, S., Crossa, J., Wu, S., Bonnett, D., & Poland, J. (2017). Genomic selection for increased yield in synthetic-derived wheat. *Crop Science*, *57*(2), 713–725. https://doi.org/10.2135/cropsci2016.04.0209

Dusser, P. (2019). The European Energy Policy for 2020–2030 Red II: What future for vegetable oil as a source of bioenergy? *OCL*, *26*, 51. https://doi.org/10.1051/ocl/2019040

Eckardt, N. A. (2010). Evolution of domesticated bread wheat. *The Plant Cell*, 22(4), 993–993. https://doi.org/10.1105/tpc.110.220410

Edwards, S. M., Buntjer, J. B., Jackson, R., Bentley, A. R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Gaynor, R. C., Gorjanc, G., Howell, P., Ober, E., Mackay, I. J., & Hickey, J. M. (2019). The effects of training population design on genomic prediction accuracy in wheat. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-019-03327-y

EFSA Panel on Contaminants in the Food Chain (CONTAM), Knutsen, H. K., 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., Schwerdtle, T., Vollmer, G., … Vleminckx, C. (2016). Erucic acid in feed and food. *EFSA Journal*, 14(11). https://doi.org/10.2903/j.efsa.2016.4593

Ehret, G. B. (2010). Genome-wide association studies: Contribution of genomics to understanding blood pressure and essential hypertension. *Current Hypertension Reports*, *12*(1), 17–25. https://doi.org/10.1007/s11906-009-0086-6

Elahi, N., Duncan, R. W., & Stasolla, C. (2016). Modification of oil and glucosinolate content in canola seeds with altered expression of *Brassica napus* LEAFY COTYLEDON1. *Plant Physiology and Biochemistry*, *100*, 52–63. https://doi.org/10.1016/j.plaphy.2015.12.022

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, *3*, 4. https://doi.org/10.3389/frai.2020.00004

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with r package rrblup. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Farooq, M., Van Dijk, A. D. J., Nijveen, H., Mansoor, S., & De Ridder, D. (2023). Genomic prediction in plants: Opportunities for ensemble machine learning based approaches. *F1000Research*, *11*, 802. https://doi.org/10.12688/f1000research.122437.2

Fehr, W. (1987). *Principles of cultivar development* (Vol. 1). Macmillan.

Fiedler, J. D., Salsman, E., Liu, Y., Michalak de Jiménez, M., Hegstad, J. B., Chen, B., Manthey, F. A., Chao, S., Xu, S., Elias, E. M., & Li, X. (2017). Genome-wide association and prediction of grain and semolina quality traits in durum wheat breeding populations. *The Plant Genome*, *10*(3). https://doi.org/10.3835/plantgenome2017.05.0038

Fikere, M., Barbulescu, D. M., Malmberg, M. M., Shi, F., Koh, J. C. O., Slater, A. T., MacLeod, I. M., Bowman, P. J., Salisbury, P. A., Spangenberg, G. C., Cogan, N. O. I., & Daetwyler, H. D. (2018). Genomic prediction using prior quantitative trait loci information reveals a large reservoir of underutilised blackleg resistance in diverse canola (*Brassica napus* L.) lines. *The Plant Genome*, 11(2), 170100. https://doi.org/10.3835/plantgenome2017.11.0100

Fikere, M., Barbulescu, D. M., Malmberg, M. M., Maharjan, P., Salisbury, P. A., Kant, S., Panozzo, J., Norton, S., Spangenberg, G. C., Cogan, N. O. I., & Daetwyler, H. D. (2020). Genomic prediction and genetic correlation of agronomic, blackleg disease, and seed quality traits in canola (*Brassica napus* L.). *Plants*, *9*(6), 719. https://doi.org/10.3390/plants9060719

Fitzpatrick, K., & Scarth, R. (1998). Improving the health and nutritional value of seed oils. *PBI Bulletin*, 15–19.

Fletcher, R. S., Mullen, J. L., Heiliger, A., & McKay, J. K. (2014). QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in *Brassica napus*. *Journal of Experimental Botany*, *66*(1), 245–256. https://doi.org/10.1093/jxb/eru423

Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., & Valè, G. (2005). Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture*, *82*(3), 317–342. https://doi.org/10.1007/s11240-005-2387-z

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Friedt, W., & Lühs, W. (1998). Recent developments and perspectives of industrial rapeseed breeding. *Lipid - Fett*, *100*(6), 219–226. https://doi.org/10.1002/(sici)1521-4133(199806)100:6<219::aid-lipi219>3.3.co;2-p

Friedt, W., & Snowdon, R. (2010). Oilseed rape. In J. Vollmann & I. Rajcan (Eds.), *Oil Crops* (pp. 91–126). Springer. https://doi.org/10.1007/978-0-387-77594-4_4

Fu, Y.-B., Yang, M.-H., Zeng, F., & Biligetu, B. (2017). Searching for an accurate marker-based prediction of an individual quantitative trait in Molecular Plant Breeding. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01182

Fuller, D. Q., Denham, T., Arroyo-Kalin, M., Lucas, L., Stevens, C. J., Qin, L., Allaby, R. G., & Purugganan, M. D. (2014). Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences*, *111*(17), 6147–6152. https://doi.org/10.1073/pnas.1308937110

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. https://doi.org/10.3389/fpls.2016.02062

Gacek, K., Bayer, P., Anderson, R., Severn-Ellis, A., Wolko, J., Łopatyńska, A., Matuszczak, M., Bocianowski, J., Edwards, D., Batley, J. (2021). QTL Genetic Mapping Study for Traits Affecting Meal Quality in Winter Oilseed Rape (*Brassica Napus* L.). *Genes (Basel), 12(8)* 1235. doi: 10.3390/genes12081235

Gao, S., Martinez, C., Skinner, D. J., Krivanek, A. F., Crouch, J. H., & Xu, Y. (2008). Development of a seed DNA-based genotyping system for marker-assisted selection in maize. *Molecular Breeding*, *22*(3), 477–494. https://doi.org/10.1007/s11032-008-9192-4

Ghafoorunissa, Vani, A., Laxmi, R., & Sesikeran, B. (2002). Effects of dietary α-linolenic acid from blended oils on biochemical indices of coronary heart disease in Indians. *Lipids*, *37*(11), 1077–1086. https://doi.org/10.1007/s11745-002-1003-2

Gianola, D., Fernando, R. L., & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, *173*(3), 1761–1776. https://doi.org/10.1534/genetics.105.049510

Gill, H. S., Halder, J., Zhang, J., Brar, N. K., Rai, T. S., Hall, C., Bernardo, A., Amand, P. S., Bai, G., Olson, E., Ali, S., Turnipseed, B., & Sehgal, S. K. (2021). Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Frontiers in Plant Science*, *12*, 709545. https://doi.org/10.3389/fpls.2021.709545

GlobalData (2024). *The Economic Impact of Canola on the Canadian Economy: 2024 update*. https://www.canolacouncil.org/download/131/economic-impact/37664/gd_the-economic-impact-of-canola-on-the-canadian-economy_august-2024

Gnoni, G. V., Natali, F., Geelen, M. J. H., & Siculella, L. (2010). Oleic acid as an inhibitor of fatty acid and cholesterol synthesis. *Olives and Olive Oil in Health and Disease Prevention*, 1365–1373. https://doi.org/10.1016/b978-0-12-374420-3.00152-2

Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, *136*(2), 245–257. https://doi.org/10.1007/s10709-008-9308-0

Goiffon, M., Kusmec, A., Wang, L., Hu, G., & Schnable, P. S. (2017). Improving response in genomic selection with a population-based selection strategy: Optimal population value selection. *Genetics*, *206*(3), 1675–1682. https://doi.org/10.1534/genetics.116.197103

Gómez-Campo, C. (1999). *Biology of Brassica coenospecies*. Elsevier.

González-Recio, O., Rosa, G. J. M., & Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livestock Science*, *166*, 217–231. https://doi.org/10.1016/j.livsci.2014.05.036

Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., Stritt, C., Roulin, A. C., Schackwitz, W., Tyler, L., Martin, J., Lipzen, A., Dochy, N., Phillips, J., Barry, K., Geuten, K., Budak, H., Juenger, T. E., Amasino, R., … Vogel, J. P. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-02292-8

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M. S., Bright, J. M., Beyene, Y., & Prasanna, B. M. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theoretical and Applied Genetics*, 128(10), 1957–1968. https://doi.org/10.1007/s00122-015-2559-0

Granado-Casas, M., & Mauricio, D. (2019). Oleic acid in the diet and what it does: Implications for diabetes and its complications. *Bioactive Food as Dietary Interventions for Diabetes*, 211–229. https://doi.org/10.1016/b978-0-12-813822-9.00014-x

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

Guo, Z., Tucker, D. M., Basten, C. J., Gandhi, H., Ersoz, E., Guo, B., Xu, Z., Wang, D., & Gay, G. (2014). The impact of population structure on genomic prediction in stratified populations. *Theoretical and Applied Genetics*, *127*(3), 749–762. https://doi.org/10.1007/s00122-013-2255-x

Guo, J., Khan, J., Pradhan, S., Shahi, D., Khan, N., Avci, M., Mcbreen, J., Harrison, S., Brown-Guedira, G., Murphy, J. P., Johnson, J., Mergoum, M., Esten Mason, R., Ibrahim, A. M. H., Sutton, R., Griffey, C., & Babar, M. A. (2020). Multi-trait genomic prediction of yield-related traits in us soft wheat under variable water regimes. *Genes*, *11*(11), 1270. https://doi.org/10.3390/genes11111270

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2014). Association mapping in crop plants. *Advances in Genetics*, 109–147. https://doi.org/10.1016/b978-0-12-800271-1.00002-0

Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177(4), 2389–2397. https://doi.org/10.1534/genetics.107.081190

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12(1), 186. https://doi.org/10.1186/1471-2105-12-186

Habier, D., Fernando, R. L., & Garrick, D. J. (2013). Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics*, *194*(3), 597–607. https://doi.org/10.1534/genetics.113.152207

Hall, R., Peters, R. D., & Assabgui, R. A. (1993). Occurrence and impact of blackleg on oilseed rape in Ontario. *Canadian Journal of Plant Pathology*, *15*(4), 305–313. https://doi.org/10.1080/07060669309501928

Hameed, N. (2023). W51 canola/rapeseed update: Global rapeseed oil consumption hits record high in 2023, and Canada anticipates robust canola harvest in 2023/24. *Tridge*. https://www.tridge.com/weekly-product-updates/w51-canolarapeseed-update-global-rapeseed-oil-consumption-hits-record-high-in#:~:text=In%202023%2C%20global%20rapeseed%20oil,attractive%20prices%20and%20substantial%20production

Hansen, M., Halldén, C., Nilsson, N.-O., & Säll, T. (1997). Marker-assisted selection of restored male-fertile *Brassica napus* plants using a set of dominant RAPD markers. *Molecular Breeding*, 3(6), 449–456. https://doi.org/10.1023/A:1009647417936

Harker, K., Blackshaw, R., Kirkland, K., Derksen, D., & Wall, D. (2000). Herbicide-tolerant canola: Weed control and yield comparisons in western Canada. *Canadian Journal of Plant Science*, 80(3), 647–654.

Hatzig, S. V., Frisch, M., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.-H., Leckband, G., Abbadi, A., & Snowdon, R. J. (2015). Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Frontiers in Plant Science*, 6. https://doi.org/10.3389/fpls.2015.00221

Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., & Goddard, M. E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, *41*(1), 51. https://doi.org/10.1186/1297-9686-41-51

Hayward, A., Mason, A. S., Dalton-Morgan, J., Zander, M., Edwards, D., & Batley, J. (2012). SNP discovery and applications in *Brassica napus*. *Journal of Plant Biotechnology*, *39*(1), 49–61. https://doi.org/10.5010/jpb.2012.39.1.049

Health Canada, Novel Food Information - Imidazolinone Tolerant (Clearfield') Canola Quality *Brassica juncea* S006 (2016).

Heffner, E. L., Sorrells, M. E., & Jannink, J. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1–12. https://doi.org/10.2135/cropsci2008.08.0512

Heffner, E. L., Lorenz, A. J., Jannink, J., & Sorrells, M. E. (2010). Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science*, *50*(5), 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Helal, M., Gill, R. A., Tang, M., Yang, L., Hu, M., Yang, L., Xie, M., Zhao, C., Cheng, X., Zhang, Y., Zhang, X., & Liu, S. (2021). SNP- and haplotype-based GWAS of flowering-related traits in *Brassica napus*. *Plants*, *10*(11), 2475. https://doi.org/10.3390/plants10112475

Herter, C. P., Ebmeyer, E., Kollers, S., Korzun, V., & Miedaner, T. (2019). An experimental approach for estimating the genomic selection advantage for Fusarium head blight and *Septoria tritici* blotch in winter wheat. *Theoretical and Applied Genetics*, *132*(8), 2425–2437. https://doi.org/10.1007/s00122-019-03364-7

Heslot, N., Jannink, J., & Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science*, *55*(1), 1–12. https://doi.org/10.2135/cropsci2014.03.0249

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M., Grondona, M., Zambelli, A., Windhausen, V. S., Mathews, K., & Gorjanc, G. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, *54*(4), 1476–1488. https://doi.org/10.2135/cropsci2013.03.0195

Hickey, J. M., Chiurugwi, T., Mackay, I., & Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics*, *49*(9), 1297–1303. https://doi.org/10.1038/ng.3920

Higgins, E. E., Clarke, W. E., Howell, E. C., Armstrong, S. J., & Parkin, I. A. (2018). Detecting de novo homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. *G3 Genes|Genomes|Genetics*, *8*(8), 2673–2683. https://doi.org/10.1534/g3.118.200118

Hirani, A. H., Gao, F., Liu, J., Fu, G., Wu, C., Yuan, Y., Li, W., Hou, J., Duncan, R., & Li, G. (2016). Transferring clubroot resistance from Chinese cabbage (*Brassica rapa*) to canola (*B. napus*). *Canadian Journal of Plant Pathology*, *38*(1), 82–90. https://doi.org/10.1080/07060661.2016.1141799

Hoerl, A. E., & Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *42*(1), 80–86. https://doi.org/10.1080/00401706.2000.10485983

Hougen, F. W., & Bodo, V. (1973). Extraction and methanolysis of oil from whole or crushed rapeseed for fatty acid analysis. *Journal of the American Oil Chemists' Society*, 50(7), 230–234. https://doi.org/10.1007/BF02641791

Howard, R., Carriquiry, A. L., & Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3 Genes|Genomes|Genetics*, *4*(6), 1027–1046. https://doi.org/10.1534/g3.114.010298

Howard, R., Jarquin, D., & Crossa, J. (2022). Overview of genomic prediction methods and the associated assumptions on the variance of marker effect, and on the architecture of the target trait. In N. Ahmadi & J. Bartholomé (Eds.), *Genomic Prediction of Complex Traits* (Vol. 2467, pp. 139–156). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_5

Hu, D., Jing, J., Snowdon, R. J., Mason, A. S., Shen, J., Meng, J., & Zou, J. (2021a). Exploring the gene pool of *Brassica napus* by genomics-based approaches. *Plant Biotechnology Journal*, 19(9), 1693–1712. https://doi.org/10.1111/pbi.13636

Hu, D., Zhao, Y., Shen, J., He, X., Zhang, Y., Jiang, Y., Snowdon, R., Meng, J., Reif, J. C., & Zou, J. (2021b). Genome-wide prediction for hybrids between parents with distinguished difference on exotic introgressions in *Brassica napus*. *The Crop Journal*, *9*(5), 1169–1178. https://doi.org/10.1016/j.cj.2020.11.002

Huang, Z., Xiao, L., Dun, X., Xia, S., Yi, B., Wen, J., Shen, J., Ma, C., Tu, J., Meng, J., & Fu, T. (2012). Improvement of the recessive genic male sterile lines with a subgenomic background in *Brassica napus* by molecular marker-assisted selection. *Molecular Breeding*, 29(1), 181–187. https://doi.org/10.1007/s11032-010-9536-8

Huang, J., Xue, C., Wang, H., Wang, L., Schmidt, W., Shen, R., & Lan, P. (2017). Genes of acyl carrier protein family show different expression profiles and overexpression of acyl

carrier protein 5 modulates fatty acid composition and enhances salt stress tolerance in arabidopsis. *Frontiers in Plant Science*, *8*, 987. https://doi.org/10.3389/fpls.2017.00987

Huang, Y.-J., Paillard, S., Kumar, V., King, G. J., Fitt, B. D., & Delourme, R. (2019). Oilseed rape (*Brassica napus*) resistance to growth of *Leptosphaeria maculans* in leaves of young plants contributes to quantitative resistance in stems of adult plants. *PLOS ONE*, *14*(9). https://doi.org/10.1371/journal.pone.0222540

Huang, L., Min, Y., Schiessl, S., Xiong, X., Jan, H. U., He, X., Qian, W., Guan, C., Snowdon, R. J., Hua, W., Guan, M., & Qian, L. (2021). Integrative analysis of GWAS and transcriptome to reveal novel loci regulation flowering time in semi-winter rapeseed. *Plant Science*, 310, 110980. https://doi.org/10.1016/j.plantsci.2021.110980

Hwang, S.-F., Strelkov, S., Peng, G., Ahmed, H., Zhou, Q., & Turnbull, G. (2016). Blackleg (*Leptosphaeria maculans*) severity and yield loss in canola in Alberta, Canada. *Plants*, *5*(3), 31. https://doi.org/10.3390/plants5030031

Hwang, S. F., Strelkov, S. E., Ahmed, H. U., Manolii, V. P., Zhou, Q., Fu, H., Turnbull, G., Fredua-Agyeman, R., & Feindel, D. (2017). Virulence and inoculum density-dependent interactions between clubroot resistant canola (*Brassica napus*) and *Plasmodiophora Brassicae*. *Plant Pathology*, *66*(8), 1318–1328. https://doi.org/10.1111/ppa.12688

Ibba, M. I., Crossa, J., Montesinos-López, O. A., Montesinos-López, A., Juliana, P., Guzman, C., Delorean, E., Dreisigacker, S., & Poland, J. (2020). Genome-based prediction of multiple wheat quality traits in multiple years. *The Plant Genome*, *13*(3), e20034. https://doi.org/10.1002/tpg2.20034

Ibrahim, S., Li, K., Ahmad, N., Kuang, L., Sadau, S. B., Tian, Z., Huang, L., Wang, X., Dun, X., & Wang, H. (2021). Genetic dissection of mature root characteristics by genome-wide association studies in rapeseed (*Brassica napus* L.). *Plants*, 10(12), 2569. https://doi.org/10.3390/plants10122569

Iggman, D., Gustafsson, I.-B., Berglund, L., Vessby, B., Marckmann, P., & Risérus, U. (2011). Replacing dairy fat with rapeseed oil causes rapid improvement of hyperlipidaemia: A randomized controlled study. *Journal of Internal Medicine*, *270*(4), 356–364. https://doi.org/10.1111/j.1365-2796.2011.02383.x

Illumina. (n.d.). Illumina microarray solutions. https://www.illumina.com/content/dam/illumina-marketing/documents/applications/genotyping/Microarray_Solutions.pdf

Jan, H. U., Abbadi, A., Lücke, S., Nichols, R. A., & Snowdon, R. J. (2016). Genomic prediction of Testcross performance in Canola (*Brassica napus*). *PLOS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0147769

Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. *Briefings in Functional Genomics*, *9*(2), 166–177. https://doi.org/10.1093/bfgp/elq001

Javed, N., Geng, J., Tahir, M., McVetty, P. B., Li, G., & Duncan, R. W. (2015). Identification of QTL influencing seed oil content, fatty acid profile and days to flowering in *Brassica napus* L. *Euphytica*, *207*(1), 191–211. https://doi.org/10.1007/s10681-015-1565-2

Jaworski, J. G., Clough, R. C., & Barnum, S. R. (1989). A cerulenin insensitive short chain 3-ketoacyl-acyl carrier protein synthase in *Spinacia oleracea* leaves. *Plant Physiology*, *90*(1), 41–44. https://doi.org/10.1104/pp.90.1.41

Jeong, S.-C., Moon, J.-K., Park, S.-K., Kim, M.-S., Lee, K., Lee, S. R., Jeong, N., Choi, M. S., Kim, N., Kang, S.-T., & Park, E. (2019). Genetic diversity patterns and domestication origin of soybean. *Theoretical and Applied Genetics*, 132(4), 1179–1193. https://doi.org/10.1007/s00122-018-3271-7

Jeong, S., Kim, J.-Y., & Kim, N. (2020). GMStool: GWAS-based marker selection tool for genomic prediction from genomic data. *Scientific Reports*, *10*(1), 19653. https://doi.org/10.1038/s41598-020-76759-y

Jia, Y., & Jannink, J.-L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, *192*(4), 1513–1522. https://doi.org/10.1534/genetics.112.144246

Jiang, C., Shi, J., Li, R., Long, Y., Wang, H., Li, D., Zhao, J., & Meng, J. (2014). Quantitative trait loci that control the oil content variation of rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics*, *127*(4), 957–968. https://doi.org/10.1007/s00122-014-2271-5

Jiang, J., Zhang, Q., Ma, L., Li, J., Wang, Z., & Liu, J.-F. (2015). Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity*, *115*(1), 29–36. https://doi.org/10.1038/hdy.2015.9

Jiang, Y., Schulthess, A. W., Rodemann, B., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Ganal, M. W., Röder, M. S., & Reif, J. C. (2017). Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. *Theoretical and Applied Genetics*, *130*(3), 471–482. https://doi.org/10.1007/s00122-016-2827-7

Jighly, A., Lin, Z., Pembleton, L. W., Cogan, N. O. I., Spangenberg, G. C., Hayes, B. J., & Daetwyler, H. D. (2019). Boosting genetic gain in allogamous crops via speed breeding and genomic selection. *Frontiers in Plant Science*, *10*, 1364. https://doi.org/10.3389/fpls.2019.01364

Jourdren, C., Barret, P., Brunel, D., Delourme, R., & Renard, M. (1996a). Specific molecular marker of the genes controlling linolenic acid content in rapeseed. *Theoretical and Applied Genetics*, 93(4), 512–518. https://doi.org/10.1007/BF00417942

Jourdren, C., Barret, P., Horvais, R., Foisset, N., Delourme, R., & Renard, M. (1996b). Identification of RAPD markers linked to the loci controlling erucic acid level in rapeseed. *Molecular Breeding*, *2*(1). https://doi.org/10.1007/bf00171352

Jubair, S., & Domaratzki, M. (2023). Crop genomic selection with deep learning and environmental data: A survey. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.1040295

Kachroo, A., Shanklin, J., Whittle, E., Lapchyk, L., Hildebrand, D., & Kachroo, P. (2006). The Arabidopsis stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant Molecular Biology*, *63*(2), 257–271. https://doi.org/10.1007/s11103-006-9086-y

Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Frontiers in Plant Science*, *10*, 1794. https://doi.org/10.3389/fpls.2019.01794

Kaler, A. S., Purcell, L. C., Beissinger, T., & Gillman, J. D. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biology*, *22*(1), 87. https://doi.org/10.1186/s12870-022-03479-y

Katavic, V., Friesen, W., Barton, D. L., Gossen, K. K., Giblin, E. M., Luciw, T., An, J., Zou, J., MacKenzie, S. L., Keller, W. A., Males, D., & Taylor, D. C. (2001). Improving erucic acid content in rapeseed through biotechnology: What can the *Arabidopsis FAE1* and the yeast *SLC1-1* genes contribute? *Crop Science*, *41*(3), 739–747. https://doi.org/10.2135/cropsci2001.413739x

Kaur, H., Wang, L., Stawniak, N., Sloan, R., van Erp, H., Eastmond, P., & Bancroft, I. (2019). The impact of reducing fatty acid desaturation on the composition and thermal stability of rapeseed oil. *Plant Biotechnology Journal*, *18*(4), 983–991. https://doi.org/10.1111/pbi.13263

Kawasaki, M., Ohara, T., Ishida, M., Takahata, Y., & Hatakeyama, K. (2021). Development of novel clubroot resistant rapeseed lines (*Brassica napus* L.) effective against Japanese field isolates by marker assisted selection. *Breeding Science*, 71(5), 528–537. https://doi.org/10.1270/jsbbs.21014

Kehel, Z., Sanchez-Garcia, M., El Baouchi, A., Aberkane, H., Tsivelikas, A., Charles, C., & Amri, A. (2020). Predictive characterization for seed morphometric traits for genebank accessions using genomic selection. *Frontiers in Ecology and Evolution*, *8*, 32. https://doi.org/10.3389/fevo.2020.00032

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

Khan, S. U., Saeed, S., Khan, M. H. U., Fan, C., Ahmar, S., Arriagada, O., Shahzad, R., Branca, F., & Mora-Poblete, F. (2021). Advances and challenges for QTL analysis and GWAS in the plant-breeding of high-yielding: A focus on rapeseed. *Biomolecules*, *11*(10), 1516. https://doi.org/10.3390/biom11101516

Khangura, R. K., & Barbetti, M. J. (2001). Prevalence of blackleg (*Leptosphaeria maculan*s) on canola (*Brassica napus*) in Western Australia. *Australian Journal of Experimental Agriculture*, *41*(1), 71. https://doi.org/10.1071/ea00068

Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, *10*, 621. https://doi.org/10.3389/fpls.2019.00621

Kimber, D. S., & MacGregor, D. I. (Eds.). (1995). *Brassica* oilseeds: Production and utilization. CAB International.

Korber, N., Bus, A., Li, J., Parkin, I. A. P., Wittkop, B., Snowdon, R. J., & Stich, B. (2016). Agronomic and seed quality traits dissected by genome-wide association mapping in *Brassica napus*. *Frontiers in Plant Science*, 7. https://doi.org/10.3389/fpls.2016.00386

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Koscielny, C. B., Gardner, S. W., Technow, F., & Duncan, R. W. (2020). Linkage mapping and whole-genome predictions in canola (*Brassica napus*) subjected to differing temperature treatments. *Crop and Pasture Science*, *71*(3), 229. https://doi.org/10.1071/CP19387

Koski, A., Pekkarinen, S., Hopia, A., Wahala, K., & Heinonen, M. (2003). Processing of rapeseed oil: Effects on sinapic acid derivative content and oxidative stability. *European Food Research and Technology*, *217*(2), 110–114. https://doi.org/10.1007/s00217-003-0721-4

Krishnappa, G., Savadi, S., Tyagi, B. S., Singh, S. K., Mamrutha, H. M., Kumar, S., Mishra, C. N., Khan, H., Gangadhara, K., Uday, G., Singh, G., & Singh, G. P. (2021). Integrated genomic selection for rapid improvement of crops. *Genomics*, *113*(3), 1070–1086. https://doi.org/10.1016/j.ygeno.2021.02.007

Kroll, H. (1994). Ein archaologischer rapsfund des 16. Jahrhunderts, entdeckt in heide in holstein, norddeutschland. *Journal of Agronomy and Crop Science*, 173(1), 17–21. https://doi.org/10.1111/j.1439-037X.1994.tb00569.x

Kumar, S., Chagné, D., Bink, M. C. A. M., Volz, R. K., Whitworth, C., & Carlisle, C. (2012). Genomic selection for fruit quality traits in apple (Malus×domestica Borkh.). *PLoS ONE*, *7*(5), e36674. https://doi.org/10.1371/journal.pone.0036674

Kumar, J., Gupta, D. S., Gupta, S., Dubey, S., Gupta, P., & Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports*, *36*(8), 1187–1213. https://doi.org/10.1007/s00299-017-2127-y

Kutcher, H. R., Yu, F., & Brun, H. (2010). Improving blackleg disease management of *Brassica napus* from knowledge of genetic interactions with *Leptosphaeria maculans*. *Canadian Journal of Plant Pathology*, 32(1), 29–34. https://doi.org/10.1080/07060661003620961

Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., & Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theoretical and Applied Genetics*, *131*(12), 2719–2731. https://doi.org/10.1007/s00122-018-3186-3

Lancashire, P. D., Bleiholder, H., Boom, T. V. D., Langelüddeke, P., Stauss, R., Weber, E., & Witzenberger, A. (1991). A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology*, *119*(3), 561–601. https://doi.org/10.1111/j.1744-7348.1991.tb04895.x

Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science*, *265*(5181), 2037–2048. https://doi.org/10.1126/science.8091226

Landry, B. S., Hubert, N., Etoh, T., Harada, J. J., & Lincoln, S. E. (1991). A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. *Genome*, *34*(4), 543–552. https://doi.org/10.1139/g91-084

Langridge, P., Lagudah, E. S., Holton, T. A., Appels, R., Sharp, P. J., & Chalmers, K. J. (2001). Trends in genetic and genome analyses in wheat: A review. *Australian Journal of Agricultural Research*, 52(12), 1043. https://doi.org/10.1071/AR01082

Larkin, Lozada, & Mason. (2019). Genomic selection—Considerations for successful implementation in wheat breeding programs. *Agronomy*, *9*(9), 479. https://doi.org/10.3390/agronomy9090479

Lasky-Su, J. (2017). Statistical techniques for genetic analysis. *Clinical and Translational Science*, 347–362. https://doi.org/10.1016/b978-0-12-802101-9.00019-3

Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: Study designs and statistical tests. *The American Journal of Human Genetics*, 95(1), 5–23. https://doi.org/10.1016/j.ajhg.2014.06.009

Leonard, C. (1992). High-erucic vegetable oils. *Industrial Crops and Products*, *1*(2-4), 119–123. https://doi.org/10.1016/0926-6690(92)90009-k

Li, G., & Quiros, C. F. (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in *Brassica*. *Theoretical and Applied Genetics*, *103*(2–3), 455–461. https://doi.org/10.1007/s001220100570

Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., & Kunst, L. (2008). Identification of the wax ester synthase/acyl-coenzyme a:diacylglycerol acyltransferase

wsd1 required for stem wax ester biosynthesis in arabidopsis. *Plant Physiology*, *148*(1), 97–107. https://doi.org/10.1104/pp.108.123471

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*, 21(4), 355–367. https://doi.org/10.1093/dnares/dsu002

Li, L., Long, Y., Zhang, L., Dalton-Morgan, J., Batley, J., Yu, L., Meng, J., & Li, M. (2015). Genome wide analysis of flowering time trait in multiple environments via high-throughput genotyping technique in *Brassica napus* L. *PLOS ONE*, 10(3), e0119425. https://doi.org/10.1371/journal.pone.0119425

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://doi.org/10.3389/fpls.2016.01483

Li, H., Cheng, X., Zhang, L., Hu, J., Zhang, F., Chen, B., Xu, K., Gao, G., Li, H., Li, L., Huang, Q., Li, Z., Yan, G., & Wu, X. (2018). An integration of genome-wide association study and gene co-expression network analysis identifies candidate genes of stem lodging-related traits in *Brassica napus*. *Frontiers in Plant Science*, 9, 796. https://doi.org/10.3389/fpls.2018.00796

Li, K., & Du, D. (2023). Breeding of determinate inflorescence new varieties of *Brassica napus* by molecular marker-assisted selection (MAS). *Chinese Journal of Oil Crop Sciences*, 45(1), 30–37. https://doi.org/10.19802/j.issn.1007-9084.2021332

List, G.R. (2016). Oilseed composition and modification for Health and Nutrition. *Functional Dietary Lipids*, 23–46. https://doi.org/10.1016/b978-1-78242-247-1.00002-8

List, G.R. (2022). Minor high-oleic oils. *High Oleic Oils*, 125–142. https://doi.org/10.1016/b978-0-12-822912-5.00010-1

Liu, Z., Fu, T., Tu, J., Ma, C., Chen, B., & Zhou, Y. (2005). Molecular marker-assisted selection for development of yellow seeded Pol CMS restorer lines in *Brassica napus* L. *Biotechnology.* 350-353. https://www.gcirc.org/fileadmin/documents/Proceedings/IRCWuhan2007vol2/Biotechnology/350-353.pdf

Liu, L, Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., Snowdon, R. J., & Li, J. (2013). A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0083052

Liu, LZ, & Li, J. (2014). QTL Mapping of Oleic Acid, Linolenic Acid and Erucic Acid Content in *Brassica napus* by Using the High Density SNP Genetic Map. *Scientia Agricultura Sinica*, *47*(1), 24–32. https://doi.org/10.3864/j.issn.0578-1752.2014.01.003

Liu, S., Fan, C., Li, J., Cai, G., Yang, Q., Wu, J., Yi, X., Zhang, C., & Zhou, Y. (2016a). A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. *Theoretical and Applied Genetics*, *129*(6), 1203–1215. https://doi.org/10.1007/s00122-016-2697-z

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016b). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genetics*, *12*(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Liu, M., Tan, X., Yang, Y., Liu, P., Zhang, X., Zhang, Y., Wang, L., Hu, Y., Ma, L., Li, Z., Zhang, Y., Zou, C., Lin, H., Gao, S., Lee, M., Lübberstedt, T., Pan, G., & Shen, Y. (2019). Analysis of the genetic architecture of maize kernel size traits by combined linkage and association mapping. *Plant Biotechnology Journal*, *18*(1), 207–221. https://doi.org/10.1111/pbi.13188

Liu, H., Wang, J., Zhang, B., Yang, X., Hammond, J. P., Ding, G., Wang, S., Cai, H., Wang, C., Xu, F., & Shi, L. (2021). Genome-wide association study dissects the genetic control of plant height and branch number in response to low-phosphorus stress in *Brassica napus*. Annals of Botany, 128(7), 919–930. https://doi.org/10.1093/aob/mcab115

Liu, H., Wang, W., Yang, M., Yuan, P., Hammond, J. P., King, G. J., Ding, G., White, P. J., Wang, S., Cai, H., Wang, C., Lu, C., Xu, F., & Shi, L. (2022a). Genome-wide association studies of important agronomic traits in *Brassica napus*: What we have learned and where we are headed. In J. A. Roberts (Ed.), *Annual Plant Reviews online* (1st ed., pp. 151–180). Wiley. https://doi.org/10.1002/9781119312994.apr0801

Liu, H., Zou, M., Zhang, B., Yang, X., Yuan, P., Ding, G., Xu, F., & Shi, L. (2022b). Genome-wide association study identifies candidate genes and favorable haplotypes for seed yield in *Brassica napus*. *Molecular Breeding, 42*(10), 61. https://doi.org/10.1007/s11032-022-01332-6

Long, A. D., & Langley, C. H. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research*, *9*(8), 720–731. https://doi.org/10.1101/gr.9.8.720

Longin, C. F. H., Mi, X., & Würschum, T. (2015). Genomic selection in wheat: Optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theoretical and Applied Genetics*, *128*(7), 1297–1306. https://doi.org/10.1007/s00122-015-2505-1

LMC International. (2020, December). *The economic impact of canola on the Canadian Economy: 2020 update*. https://www.canolacouncil.org/download/131/economic-impact/17818/economic-impact-report-canada_december-2020

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., Smith, K. P., Sorrells, M. E., & Jannink, J.-L. (2011). Genomic selection in plant breeding. In *Advances in Agronomy* (Vol. 110, pp. 77–123). Elsevier. https://doi.org/10.1016/B978-0-12-385531-2.00002-5

Lorenz, A. J., Smith, K. P., & Jannink, J. -L. (2012). Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. *Crop Science*, *52*(4), 1609–1621. https://doi.org/10.2135/cropsci2011.09.0503

Lorenz, A., & Nice, L. (2017). Chapter 2 Training Population Design and Resource Allocation for Genomic Selection in Plant Breeding. In *Genomic selection for crop improvement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-63170-7

Lozada, D., Godoy, J. V., Murray, T. D., Ward, B. P., & Carter, A. H. (2019a). Genetic dissection of snow mold tolerance in us pacific northwest winter wheat through genome-wide association study and genomic selection. *Frontiers in Plant Science*, *10*, 1337. https://doi.org/10.3389/fpls.2019.01337

Lozada, D., Mason, R. E., Sarinelli, J. M., & Brown-Guedira, G. (2019b). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genetics*, *20*(1), 82. https://doi.org/10.1186/s12863-019-0785-1

Lu, K., Xiao, Z., Jian, H., Peng, L., Qu, C., Fu, M., He, B., Tie, L., Liang, Y., Xu, X., & Li, J. (2016). A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in *Brassica napus*. *Scientific Reports*, 6(1), 36452. https://doi.org/10.1038/srep36452

Luaces, O., Quevedo, J. R., Pérez-Enciso, M., Díez, J., Del Coz, J. J., & Bahamonde, A. (2010). Explaining the genetic basis of complex quantitative traits through prediction models. *Journal of Computational Biology*, *17*(12), 1711–1723. https://doi.org/10.1089/cmb.2009.0161

Luo, X., Ding, Y., Zhang, L., Yue, Y., Snyder, J. H., Ma, C., & Zhu, J. (2017). Genomic prediction of genotypic effects with epistasis and environment interactions for yield-related traits of rapeseed (*Brassica napus* L.). *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00015

Luo, T., Zhang, Y., Zhang, C., Nelson, M. N., Yuan, J., Guo, L., & Xu, Z. (2021). Genome-wide association mapping unravels the genetic control of seed vigor under low-temperature conditions in rapeseed (*Brassica napus* L.). *Plants*, 10(3), 426. https://doi.org/10.3390/plants10030426

Magdy, T., Kuo, H., & Burridge, P. W. (2020). Precise and cost-effective nanopore sequencing for post-GWAS fine-mapping and causal variant identification. *iScience*, *23*(4), 100971. https://doi.org/10.1016/j.isci.2020.100971

Makowsky, R., Pajewski, N. M., Klimentidis, Y. C., Vazquez, A. I., Duarte, C. W., Allison, D. B., & Campos, G. De los. (2011). Beyond missing heritability: Prediction of complex traits. *PLOS Genetics*, *7*(4), e1002051. https://doi.org/10.1371/journal.pgen.1002051

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. https://doi.org/10.1038/nature08494

Mason, A. S., Higgins, E. E., Snowdon, R. J., Batley, J., Stein, A., Werner, C., & Parkin, I. A. (2017). A user guide to the *Brassica* 60K Illumina Infinium SNP genotyping array. *Theoretical and Applied Genetics*, *130*(4), 621–633. https://doi.org/10.1007/s00122-016-2849-1

Matei, G., Woyann, L. G., Milioli, A. S., De Bem Oliveira, I., Zdziarski, A. D., Zanella, R., Coelho, A. S. G., Finatto, T., & Benin, G. (2018). Genomic selection in soybean: Accuracy and time gain in relation to phenotypic selection. *Molecular Breeding*, *38*(9), 117. https://doi.org/10.1007/s11032-018-0872-4

McGee, D., & Petrie, G. (1978). Variability of *Leptosphaeria maculans* in Relation to Blackleg of Oilseed Rape. *Phytopathology*, 68, 625–630.

McGowan, M., Wang, J., Dong, H., Liu, X., Jia, Y., Wang, X., Iwata, H., Li, Y., Lipka, A. E., & Zhang, Z. (2021). Ideas in genomic selection with the potential to transform plant molecular breeding: A review. In I. Goldman (Ed.), *Plant Breeding Reviews* (1st ed., pp. 273–319). Wiley. https://doi.org/10.1002/9781119828235.ch7

McNaughton, I.H. (1976). Swedes and rapes *Brassica napus* (Cruciferae). In *Evolution of Crop Plants.* (pp. 53-56). Longman.

McVetty, P., Rimmer, S., & Scarth, R. (2001). Cartier BX summer oilseed rape. *Canadian Journal of Plant Science*, 81(1), 105–106.

McVetty, P., Zelmer, C., Tahir, M., Fernando, D., & Li, G. (2009). High-erucic acid, low-glucosinolate rapeseed (HEAR) cultivar development in Canada. *Biocatalysis and Agricultural Biotechnology*, 43–61. https://doi.org/10.1201/9781420077070.ch4

McVetty, P., & Duncan, R. W. (2015). Canola, rapeseed, and mustard: For biofuels and bioproducts. In V. M. V. Cruz & D. A. Dierig (Eds.), *Industrial Crops: Breeding for BioEnergy and Bioproducts* (pp. 133–156). Springer. https://doi.org/10.1007/978-1-4939-1447-0_7

Melchinger, A. E., Utz, H. F., & Schön, C. C. (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, 149(1), 383–403. https://doi.org/10.1093/genetics/149.1.383

Mérida-García, R., Liu, G., He, S., Gonzalez-Dugo, V., Dorado, G., Gálvez, S., Solís, I., Zarco-Tejada, P. J., Reif, J. C., & Hernandez, P. (2019). Genetic dissection of agronomic and quality traits based on association mapping and genomic selection approaches in durum

wheat grown in Southern Spain. *PLOS ONE*, *14*(2), e0211718. https://doi.org/10.1371/journal.pone.0211718

Merrick, L. F., & Carter, A. H. (2021). Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *The Plant Genome*, *14*(3), e20158. https://doi.org/10.1002/tpg2.20158

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

Meuwissen, T. H. (2009). Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. *Genetics Selection Evolution*, *41*(1), 35. https://doi.org/10.1186/1297-9686-41-35

Michel, S., Gallee, M., Löschenberger, F., Buerstmayr, H., & Kummer, C. (2017). Improving the baking quality of bread wheat using rapid tests and genomics: The prediction of dough rheological parameters by gluten peak indices and genomic selection models. *Journal of Cereal Science*, *77*, 24–34. https://doi.org/10.1016/j.jcs.2017.07.012

Michel, S., Löschenberger, F., Hellinger, J., Strasser, V., Ametz, C., Pachler, B., Sparry, E., & Bürstmayr, H. (2019). Improving and maintaining winter hardiness and frost tolerance in bread wheat by genomic selection. *Frontiers in Plant Science*, *10*, 1195. https://doi.org/10.3389/fpls.2019.01195

Miller, M. J., Song, Q., & Li, Z. (2023). Genomic selection of soybean (*Glycine max*) for genetic improvement of yield and seed composition in a breeding context. *The Plant Genome*, *16*(4), e20384. https://doi.org/10.1002/tpg2.20384

Moeinizade, S., Hu, G., Wang, L., & Schnable, P. S. (2019). Optimizing selection and mating in genomic selection with a look-ahead approach: An operations research framework. *G3 Genes|Genomes|Genetics*, *9*(7), 2123–2133. https://doi.org/10.1534/g3.118.200842

Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W. R., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., & Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Genomics*, *22*(1), 19. https://doi.org/10.1186/s12864-020-07319-x

Montesinos-López, O. A., Montesinos-López, A., & Crossa, J. (2022a). Random forest for genomic prediction. In O. A. Montesinos-López, A. Montesinos-López, & J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 633–681). Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_15

Montesinos-López, O. A., Montesinos-López, A., & Crossa, J. (2022b). Overfitting, model tuning, and evaluation of prediction performance. In O. A. Montesinos-López, A. Montesinos-López, & J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 109–139). Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_4

Montesinos-López, O. A., Bentley, A. R., Saint Pierre, C., Crespo-Herrera, L., Salinas Ruiz, J., Valladares-Celis, P. E., Montesinos-López, A., & Crossa, J. (2023). Integrating parental

phenotypic data enhances prediction accuracy of hybrids in wheat traits. *Genes*, *14*(2), 395. https://doi.org/10.3390/genes14020395

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st Century crop improvement. *Plant Physiology*, *147*(3), 969–977. https://doi.org/10.1104/pp.108.118232

Minnesota Canola Council. (1998). *Brassica napus* variety trials. Minnesota Canola Council. http://www.mncanola.org/pdf/1998%20CPC%20Final%20Report.pdf

Morrison, M. J., Harker, K. N., Blackshaw, R. E., Holzapfel, C. J., & O'Donovan, J. T. (2016). Canola yield improvement on the Canadian Prairies from 2000 to 2013. *Crop and Pasture Science*, *67*(4), 245. https://doi.org/10.1071/cp15348

Naczk, M., Amarowicz, R., Sullivan, A., & Shahidi, F. (1998). Current research developments on Polyphenolics of Rapeseed/Canola: A Review. *Food Chemistry*, *62*(4), 489–502. https://doi.org/10.1016/s0308-8146(97)00198-2

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. (2017). 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

Nath, U. K., Wilmer, J. A., Wallington, E. J., Becker, H. C., & Möllers, C. (2008). Increasing erucic acid content through combination of endogenous low polyunsaturated fatty acids alleles with LD-LPAAT + BN-fae1 transgenes in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics*, *118*(4), 765–773. https://doi.org/10.1007/s00122-008-0936-7

National Library of Medicine. (2024). Compound Summary: Erucic Acid. https://pubchem.ncbi.nlm.nih.gov/compound/Erucic-Acid

Nesi, N., Delourme, R., Brégeon, M., Falentin, C., & Renard, M. (2008). Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. *Comptes Rendus. Biologies*, *331*(10), 763–771. https://doi.org/10.1016/j.crvi.2008.07.018

Neves, H. H., Carvalheiro, R., & Queiroz, S. A. (2012). A comparison of statistical methods for genomic selection in a mice population. *BMC Genetics*, *13*(1), 100. https://doi.org/10.1186/1471-2156-13-100

Neyhart, J. L., Tiede, T., Lorenz, A. J., & Smith, K. P. (2017). Evaluating methods of updating training data in long-term genomewide selection. *G3 Genes|Genomes|Genetics*, *7*(5), 1499–1510. https://doi.org/10.1534/g3.117.040550

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565

Norman, A., Taylor, J., Edwards, J., & Kuchel, H. (2018). Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. *G3 Genes|Genomes|Genetics*, *8*(9), 2889–2899. https://doi.org/10.1534/g3.118.200311

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., … Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, *497*(7451), 579–584. https://doi.org/10.1038/nature12211

O'Donovan, J. T., Harker, K. N., Clayton, G. W., & Blackshaw, R. E. (2006). Comparison of a glyphosate-resistant canola (*Brassica napus* L.) system with traditional herbicide regimes. *Weed Technology*, 20(2), 494–501. https://doi.org/10.1614/WT-05-092R.1

Odilbekov, F., Armonienė, R., Koc, A., Svensson, J., & Chawade, A. (2019). Gwas-assisted genomic prediction to predict resistance to septoria tritici blotch in nordic winter wheat at seedling stage. *Frontiers in Genetics*, *10*, 1224. https://doi.org/10.3389/fgene.2019.01224

OECD. (1997). Consensus Document on the Biology of *Brassica napus* L. (Oilseed Rape). Series on Harmonization of Regulatory Oversight of Biotechnology. *Organisation for Economic Co-operation and Development*. https://one.oecd.org/document/OCDE/GD(97)63/en/pdf

Ogura H (1968) Studies on the new male sterility in Japanese radish, with special references on the utilization of this sterility towards the practical raising of hybrid seeds. *Mem Fac Agric Kagoshima Univ* 6:40-75

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., & Browse, J. (1994). *Arabidopsis* FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *The Plant Cell*, *6*(1), 147. https://doi.org/10.2307/3869682

Onogi, A., Ideta, O., Inoshita, Y., Ebana, K., Yoshioka, T., Yamasaki, M., & Iwata, H. (2015). Exploring the areas of applicability of whole-genome prediction methods for Asian rice (Oryza sativa L.). *Theoretical and Applied Genetics*, *128*(1), 41–53. https://doi.org/10.1007/s00122-014-2411-y

Orr, H. A. (2009). Fitness and its role in Evolutionary Genetics. *Nature Reviews Genetics*, *10*(8), 531–539. https://doi.org/10.1038/nrg2603

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y., & Tanaka, T. (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics*, *32*(4), 650–654. https://doi.org/10.1038/ng1047

Pageau, D., Lajeunesse, J., & Lafond, J. (2006). Effect of clubroot [Plasmodiophora *Brassicae*] on the productivity and quality of canola. *Canadian Journal of Plant Pathology*, *1*(28), 137–143.

Pal, L., Sandhu, S. K., Bhatia, D., & Sethi, S. (2021a). Genome-wide association study for candidate genes controlling seed yield and its components in rapeseed (*Brassica napus* subsp. *napus*). *Physiology and Molecular Biology of Plants*, *27*(9), 1933–1951. https://doi.org/10.1007/s12298-021-01060-9

Pal, L., Sandhu, S. K., & Bhatia, D. (2021b). Genome-wide association study and identification of candidate genes for seed oil content in *Brassica napus*. *Euphytica*, 217(4), 66. https://doi.org/10.1007/s10681-021-02783-2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). *Scikit-learn: Machine learning in python*. https://doi.org/10.48550/ARXIV.1201.0490

Pérez, P., De Los Campos, G., Crossa, J., & Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in R. *The Plant Genome*, *3*(2), plantgenome2010.04.0005. https://doi.org/10.3835/plantgenome2010.04.0005

Pérez, P., & De Los Campos, G. (2014). Genome-wide regression and prediction with the bglr statistical package. *Genetics*, *198*(2), 483–495. https://doi.org/10.1534/genetics.114.164442

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., & Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes|Genomes|Genetics*, *2*(12), 1595–1605. https://doi.org/10.1534/g3.112.003665

Petruzzello, M. (2023). List of plants in the family Brassicaceae. *Encyclopædia Britannica*. https://www.britannica.com/topic/list-of-plants-in-the-family-Brassicaceae-2004620

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., & Jannink, J. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome*, *5*(3), plantgenome2012.06.0006. https://doi.org/10.3835/plantgenome2012.06.0006

Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, *15*(1), 8–15. https://doi.org/10.1007/BF02772108

Potter, C. (1999). 10th International Rapeseed Congress. In P. Salisbury, T. Potter, & G. McDonald (Eds.), *Canola in Australia: The first thirty years*. Canberra.

Prohens, J. (2011). Plant breeding: A success story to be continued thanks to the advances in genomics. *Frontiers in Plant Science*, 2. https://doi.org/10.3389/fpls.2011.00051

Przybylski, R., & Eskin, N. A. M. (2011). Oil composition and properties. In *Canola* (pp. 189–227). Elsevier. https://doi.org/10.1016/B978-0-9818936-5-5.50011-5

Purcell, S., Cherny, S. S., & Sham, P. C. (2003). Genetic Power Calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, *19*(1), 149–150. https://doi.org/10.1093/bioinformatics/19.1.149

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. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Python Software Foundation (2023) Python 3.11.5 https://www.python.org/downloads/release/python-3115/

Qi, W., Tinnenbroek-Capel, I. E. M., Salentijn, E. M. J., Zhang, Z., Huang, B., Cheng, J., Shao, H., Visser, R. G. F., Krens, F. A., & Van Loo, E. N. (2018). Genetically engineering CRAMBE ABYSSINICA —A potentially high-value oil crop for salt land improvement. *Land Degradation & Development*, 29(4), 1096–1106. https://doi.org/10.1002/ldr.2847

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

Qin, M., Song, J., Guo, N., Zhang, M., Zhu, Y., Huang, Z., & Xu, A. (2022). Genome-wide association analyses reveal candidate genes controlling harvest index and related agronomic traits in *Brassica napus* L. *Agronomy*, 12(4), 814. https://doi.org/10.3390/agronomy12040814

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

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). https://doi.org/10.1186/s12864-017-3607-8

Raboanatahiry, N., Chao, H., Dalin, H., Pu, S., Yan, W., Yu, L., Wang, B., & Li, M. (2018). QTL alignment for seed yield and yield related traits in *Brassica napus*. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.01127

Raboanatahiry, N., Li, H., Yu, L., & Li, M. (2021). Rapeseed (*Brassica napus*): Processing, utilization, and genetic improvement. *Agronomy*, *11*(9), 1776. https://doi.org/10.3390/agronomy11091776

Rahman, M., Sun, Z., McVetty, P. B. E., & Li, G. (2008). High throughput genome-specific and gene-specific molecular markers for erucic acid genes in *Brassica napus* (L.) for marker-assisted selection in plant breeding. *Theoretical and Applied Genetics*, 117(6), 895–904. https://doi.org/10.1007/s00122-008-0829-9

Rahman, Habibur. (2013). Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Canadian Journal of Plant Science*, *93*(3), 363–373. https://doi.org/10.4141/cjps2012-074

Rahman, H., Peng, G., Yu, F., Falk, K. C., Kulkarni, M., & Selvaraj, G. (2014). Genetics and breeding for Clubroot Resistance in Canadian Spring Canola (*Brassica napus* L.). *Canadian Journal of Plant Pathology*, *36*(sup1), 122–134. https://doi.org/10.1080/07060661.2013.862571

Rajkumar, B., & Panjabi, P. (2022). Genomic Selection in Oilseed Brassica: Potential, Prospects and Challenges. In *Genomic selection in plants: A guide for breeders* (First edition). CRC Press, Taylor & Francis Group.

Rakow, G., Raney, J., & Relf-Eckstein, J. (1999). Agronomic performance and seed quality of a new source of yellow seeded *Brassica napus*. Agronomic Performance and Seed Quality of a New Source of Yellow Seeded *Brassica napus*. 10th International Rapeseed Congress, Canberra, Australia.

Raman, H., Raman, R., Coombes, N., Song, J., Diffey, S., Kilian, A., Lindbeck, K., Barbulescu, D. M., Batley, J., Edwards, D., Salisbury, P. A., & Marcroft, S. (2016). Genome-wide association study identifies new loci for resistance to *leptosphaeria maculans* in canola. *Frontiers in Plant Science*, 7. https://doi.org/10.3389/fpls.2016.01513

Raman, H., Raman, R., Qiu, Y., Yadav, A. S., Sureshkumar, S., Borg, L., Rohan, M., Wheeler, D., Owen, O., Menz, I., & Balasubramanian, S. (2019). GWAS hints at pleiotropic roles for FLOWERING LOCUS T in flowering time and yield-related traits in canola. *BMC Genomics*, 20(1), 636. https://doi.org/10.1186/s12864-019-5964-y

Rani, S. H., Saha, S., & Rajasekharan, R. (2013). A soluble diacylglycerol acyltransferase is involved in triacylglycerol biosynthesis in the oleaginous yeast *Rhodotorula glutinis*. *Microbiology*, *159*(Pt_1), 155–166. https://doi.org/10.1099/mic.0.063156-0

Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., Orf, J. H., Young, N. D., & Chen, S. (2020). Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. *PLOS ONE*, *15*(7), e0235089. https://doi.org/10.1371/journal.pone.0235089

Ravelombola, W., Qin, J., Shi, A., Song, Q., Yuan, J., Wang, F., Chen, P., Yan, L., Feng, Y., Zhao, T., Meng, Y., Guan, K., Yang, C., & Zhang, M. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. *PLOS ONE*, *16*(8), e0255761. https://doi.org/10.1371/journal.pone.0255761

Reif, J. C., Zhao, Y., Würschum, T., Gowda, M., & Hahn, V. (2012). Genomic prediction of sunflower hybrid performance. *Plant Breeding*, *132*(1), 107–114. https://doi.org/10.1111/pbr.12007

Reinoso-Peláez, E. L., Gianola, D., & González-Recio, O. (2022). Genome-enabled prediction methods based on machine learning. In N. Ahmadi & J. Bartholomé (Eds.), *Genomic Prediction of Complex Traits* (Vol. 2467, pp. 189–218). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_7

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*, 98(20), 11479–11484. https://doi.org/10.1073/pnas.201394398

Ren, Z., Pae, M., Dao, M. C., Smith, D., Meydani, S. N., & Wu, D. (2010). Dietary supplementation with tocotrienols enhances immune function in C57BL/6 mice. *The Journal of Nutrition*, *140*(7), 1335–1341. https://doi.org/10.3945/jn.110.121434

Ren, W., Si, J., Chen, L., Fang, Z., Zhuang, M., Lv, H., Wang, Y., Ji, J., Yu, H., & Zhang, Y. (2022). Mechanism and utilization of Ogura cytoplasmic male sterility in cruciferae crops. *International Journal of Molecular Sciences*, 23(16), 9099. https://doi.org/10.3390/ijms23169099

Rhoads, A., & Au, K. F. (2015). Pacbio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, 13(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Ribaut, J.-M., & Hoisington, D. (1998). Marker-assisted selection: New Tools and Strategies. *Trends in Plant Science*, *3*(6), 236–239. https://doi.org/10.1016/s1360-1385(98)01240-0

Rice, B., & Lipka, A. E. (2019). Evaluation of rr‐BLUP genomic selection models that incorporate peak genome‐wide association study signals in maize and sorghum. *The Plant Genome*, 12(1), 180052. https://doi.org/10.3835/plantgenome2018.07.0052

Rife, T. W., Graybosch, R. A., & Poland, J. A. (2018). Genomic analysis and prediction within a us public collaborative winter wheat regional testing nursery. *The Plant Genome*, *11*(3), 180004. https://doi.org/10.3835/plantgenome2018.01.0004

Rio, S., Charcosset, A., Mary-Huard, T., Moreau, L., & Rincent, R. (2022). Building a calibration set for genomic predictions (Gp), characteristics to be considered, and optimization approaches. In N. Ahmadi & J. Bartholomé (Eds.), *Genomic Prediction of Complex Traits: Methods and Protocols* (pp. 77–112). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_3

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516–1517. https://doi.org/10.1126/science.273.5281.1516

Risérus, U., Willett, W. C., & Hu, F. B. (2009). Dietary fats and prevention of type 2 diabetes. *Progress in Lipid Research*, *48*(1), 44–51. https://doi.org/10.1016/j.plipres.2008.10.002

Robertsen, C., Hjortshøj, R., & Janss, L. (2019). Genomic selection in cereal breeding. *Agronomy*, 9(2), 95. https://doi.org/10.3390/agronomy9020095

Rolling, W. R., Dorrance, A. E., & McHale, L. K. (2020). Testing methods and statistical models of genomic prediction for quantitative disease resistance to Phytophthora sojae in soybean [Glycine max (L.) Merr] germplasm collections. *Theoretical and Applied Genetics*, *133*(12), 3441–3454. https://doi.org/10.1007/s00122-020-03679-w

Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., Falentin, C., Boideau, F., Boutte, J., Delourme, R., Deniot, G., Engelen, S., Carvalho, J. F., Lemainque, A., Maillet, L., Morice, J., Wincker, P., Denoeud, F., Chèvre, A.-M., Aury, J.-M. (2020). Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*, 9(12). https://doi.org/10.1093/gigascience/giaa137

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), 21773. https://doi.org/10.1038/s41598-021-01272-9

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA

RStudio Team (2023) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., & Allard, R. W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences, 81*(24), 8014–8018. https://doi.org/10.1073/pnas.81.24.8014

Saha, S., Enugutti, B., Rajakumari, S., & Rajasekharan, R. (2006). Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiology*, *141*(4), 1533–1543. https://doi.org/10.1104/pp.106.082198

Sandhu, K., Patil, S. S., Pumphrey, M., & Carter, A. (2021a). Multitrait machine- and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome*, *14*(3), e20119. https://doi.org/10.1002/tpg2.20119

Sandhu, K. S., Aoun, M., Morris, C., & Carter, A. H. (2021b). Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. *Biology, 10*(7), 689. https://doi.org/10.3390/biology10070689

Sanyal, A., Pinochet, X., Merrien, A., Laustriat, M., Decocq, G., & Fine, F. (2015). Erucic acid rapeseed: 1. Prospects of improvements. *OCL, 22*(3), D303. https://doi.org/10.1051/ocl/2015011

Sasongko, N. D., & Möllers, C. (2005). Toward increasing erucic acid content in oilseed rape (*Brassica napus* L.) through the combination with genes for high oleic acid. *Journal of the American Oil Chemists' Society*, *82*(6), 445–449. https://doi.org/10.1007/s11746-005-1091-4

Scarth, R., McVetty, P. B. E., Rimmer, S. R., & Stefansson, B. R. (1988). Stellar low linolenic-high linoleic acid summer rape. *Canadian Journal of Plant Science*, 68(2), 509–511. https://doi.org/10.4141/cjps88-061

Scheben, A., Verpaalen, B., Lawley, C. T., Chan, C. K., Bayer, P. E., Batley, J., & Edwards, D. (2018). CropSNPdb: A database of SNP array data for *Brassica* crops and hexaploid bread wheat. *The Plant Journal*, *98*(1), 142–152. https://doi.org/10.1111/tpj.14194

Schiessl, S., Samans, B., Huttel, B., Reinhard, R., & Snowdon, R. J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica* napus. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00404

Schiessl, S., Iniguez-Luy, F., Qian, W., & Snowdon, R. J. (2015). Diverse regulatory factors associate with flowering time and yield responses in winter-type *Brassica* napus. *BMC Genomics*, 16(1), 737. https://doi.org/10.1186/s12864-015-1950-1

Schmidt, M., Kollers, S., Maasberg-Prelle, A., Großer, J., Schinkel, B., Tomerius, A., Graner, A., & Korzun, V. (2016). Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. *Theoretical and Applied Genetics*, *129*(2), 203–213. https://doi.org/10.1007/s00122-015-2639-1

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, *44*(7), 825–830. https://doi.org/10.1038/ng.2314

Sehgal, D., Rosyara, U., Mondal, S., Singh, R., Poland, J., & Dreisigacker, S. (2020). Incorporating genome-wide association mapping results into genomic prediction models for grain yield and yield stability in CIMMYT spring bread wheat. *Frontiers in Plant Science*, *11*, 197. https://doi.org/10.3389/fpls.2020.00197

Shahbandeh, M. (2024, February 13). *Worldwide production major vegetable oils 2012-2024*. Statista. https://www.statista.com/statistics/263933/production-of-vegetable-oils-worldwide-since-2000/

Sharafi, Y., Majidi, M. M., Goli, S. A. H., & Rashidi, F. (2015). Oil content and fatty acids composition in *Brassica* species. *International Journal of Food Properties*, 18(10), 2145–2154. https://doi.org/10.1080/10942912.2014.968284

Slocum, M. K., Figdore, S. S., Kennard, W. C., Suzuki, J. Y., & Osborn, T. C. (1990). Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. *Theoretical and Applied Genetics*, *80*(1), 57–64. https://doi.org/10.1007/bf00224016

Smallwood, C. J., Saxton, A. M., Gillman, J. D., Bhandari, H. S., Wadl, P. A., Fallen, B. D., Hyten, D. L., Song, Q., & Pantalone, V. R. (2019). Context-specific genomic selection strategies outperform phenotypic selection for soybean quantitative traits in the progeny row stage. *Crop Science*, *59*(1), 54–67. https://doi.org/10.2135/cropsci2018.03.0197

Smith, W.L., & Borgeat, P. (1985). The eicosanoids: prostaglandins, thromboxanes, leukotrienes, and hydroxy-eicosaenoic acids. In *Biochemistry of Lipids and Membranes* (Vance, D.E. and Vance, J.E., eds.), pp. 325–360, Benjamin/Cummings Publishing Co. Inc., Menlo Park.

Somers, D., Rakow, G., Raney, P., Prabhu, V., Seguin-Swartz, G., Rimmer, R., Gugel, R., Lydiate, D., & Sharp, A. (1999). 10th International Rapeseed Congress. In *Developing marker-assisted breeding for quality and disease resistance traits in Brassica oilseeds*. Canberra.

Song, K. M., Suzuki, J. Y., Slocum, M. K., Williams, P. M., & Osborn, T. C. (1991). A linkage map of *Brassica rapa* (syn. campestris) based on restriction fragment length polymorphism loci. *Theoretical and Applied Genetics*, *82*(3), 296–304. https://doi.org/10.1007/bf02190615

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2013). Development and evaluation of SOYSNP50K, a high-density genotyping array for soybean. *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0054985

Song, J., Carver, B. F., Powers, C., Yan, L., Klápště, J., El-Kassaby, Y. A., & Chen, C. (2017). Practical application of genomic selection in a doubled-haploid winter wheat breeding program. *Molecular Breeding*, *37*(10), 117. https://doi.org/10.1007/s11032-017-0715-8

Song, J., Li, B., Cui, Y., Zhuo, C., Gu, Y., Hu, K., Wen, J., Yi, B., Shen, J., Ma, C., Fu, T., & Tu, J. (2021). QTL mapping and diurnal transcriptome analysis identify candidate genes regulating *Brassica napus* flowering time. *International Journal of Molecular Sciences*, *22*(14), 7559. https://doi.org/10.3390/ijms22147559

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

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., & McCouch, S. R. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice

breeding lines. *PLOS Genetics*, *11*(2), e1004982. https://doi.org/10.1371/journal.pgen.1004982

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J.-L., & McCouch, S. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity*, *116*(4), 395–408. https://doi.org/10.1038/hdy.2015.113

Starosta, E., Szwarc, J., Niemann, J., Szewczyk, K., & Weigt, D. (2023). *Brassica napus* haploid and double haploid production and its latest applications. *Current Issues in Molecular Biology*, *45*(5), 4431–4450. https://doi.org/10.3390/cimb45050282

Statistics Canada. (2024). Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and Imperial units. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210035901

Stefansson, B., Hougen, F. W., & Downey, R. K. (1961). Note on the isolation of rape plants with seed oil free from Erucic acid. *Canadian Journal of Plant Science*, *41*(1), 218–219. https://doi.org/10.4141/cjps61-028

Stefansson, B. R., & Kondra, Z. P. (1975). Tower summer rape. *Canadian Journal of Plant Science*, 55(1), 343–344. https://doi.org/10.4141/cjps75-053

Stefansson, B., & Downey, R. (1995). Rapeseed. In *Harvest of Gold : The History of Field Crop Breeding in Canada* (pp. 140–152). essay, University of Calgary.

Stein, A., Coriton, O., Rousseau-Gueutin, M., Samans, B., Schiessl, S. V., Obermeier, C., Parkin, I. A. P., Chèvre, A., & Snowdon, R. J. (2017). Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant Biotechnology Journal*, *15*(11), 1478–1489. https://doi.org/10.1111/pbi.12732

Stewart-Brown, B. B., Song, Q., Vaughn, J. N., & Li, Z. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3 Genes|Genomes|Genetics*, *9*(7), 2253–2265. https://doi.org/10.1534/g3.118.200917

Stricker, H., Duchini, F., Facchini, M., & Mombelli, G. (2008). Canola oil decreases cholesterol and improves endothelial function in patients with peripheral arterial occlusive disease – a pilot study. *Artery Research*, *2*(2), 67. https://doi.org/10.1016/j.artres.2008.02.001

Sun, F., Liu, J., Hua, W., Sun, X., Wang, X., & Wang, H. (2016). Identification of stable QTLs for seed oil content by combined linkage and association mapping in *Brassica napus*. *Plant Science*, *252*, 388–399. https://doi.org/10.1016/j.plantsci.2016.09.001

Sun, J., Poland, J. A., Mondal, S., Crossa, J., Juliana, P., Singh, R. P., Rutkoski, J. E., Jannink, J.-L., Crespo-Herrera, L., Velu, G., Huerta-Espino, J., & Sorrells, M. E. (2019). High-throughput phenotyping platforms enhance genomic selection for wheat grain yield

across populations and cycles in early stage. *Theoretical and Applied Genetics*, *132*(6), 1705–1720. https://doi.org/10.1007/s00122-019-03309-0

Sun, J. (2021) *Genome-wide association and genomic selection in Brassica napus L.* [Doctoral dissertation, University of Manitoba]. MSpace University of Manitoba Institutional Repository. http://hdl.handle.net/1993/36265

Swanson, E. B., Herrgesell, M. J., Arnoldo, M., Sippell, D. W., & Wong, R. S. C. (1989). Microspore mutagenesis and selection: Canola plants with field tolerance to the imidazolinones. *Theoretical and Applied Genetics*, 78(4), 525–530. https://doi.org/10.1007/BF00290837

Syvänen, A.-C. (2001). Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2(12), 930–942. https://doi.org/10.1038/35103535

TAIR - The Arabidopsis Information Resource. (2013). Locus: AT4G14440. https://www.arabidopsis.org/locus?key=128729

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., & Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. *BMC Plant Biology*, 17(1), 110. https://doi.org/10.1186/s12870-017-1059-6

Tan, Z., Xie, Z., Dai, L., Zhang, Y., Zhao, H., Tang, S., Wan, L., Yao, X., Guo, L., & Hong, D. (2021). Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in *Brassica napus*. *Plant Biotechnology Journal*, *20*(1), 211–225. https://doi.org/10.1111/pbi.13707

Tan, B., & Ingvarsson, P. K. (2022). Integrating genome‐wide association mapping of additive and dominance genetic effects to improve genomic prediction accuracy in Eucalyptus. *The Plant Genome*, 15(2), e20208. https://doi.org/10.1002/tpg2.20208

Tanhuanpää, P. K., Vilkki, J. P., & Vilkki, H. J. (1996). Mapping of a QTL for oleic acid concentration in spring turnip rape (*Brassica rapa* ssp. *oleifera*). *Theoretical and Applied Genetics*, *92*(8), 952–956. https://doi.org/10.1007/bf00224034

Tang, M., Zhang, Y., Liu, Y., Tong, C., Cheng, X., Zhu, W., Li, Z., Huang, J., & Liu, S. (2019). Mapping loci controlling fatty acid profiles, oil and protein content by genome-wide association study in *Brassica napus*. *The Crop Journal*, *7*(2), 217–226. https://doi.org/10.1016/j.cj.2018.10.007

Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., Yang, Q.-Y., Zhou, Y., Wang, X., Ma, W., Xie, W., & Guo, L. (2021). Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Molecular Plant*, 14(3), 470–487. https://doi.org/10.1016/j.molp.2020.12.003

Tang, Y., Zhang, G., Jiang, X., Shen, S., Guan, M., Tang, Y., Sun, F., Hu, R., Chen, S., Zhao, H., Li, J., Lu, K., Yin, N., & Qu, C. (2023a). Genome-wide association study of glucosinolate metabolites (MGWAS) in *Brassica napus* L. *Plants*, 12(3), 639. https://doi.org/10.3390/plants12030639

Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., & Wang, Y. (2023b). SRplot: A free online platform for data visualization and graphing. *PLOS ONE*, *18*(11), e0294236. https://doi.org/10.1371/journal.pone.0294236

Taylor, D. C., Smith, M. A., Fobert, P., Mietkiewska, E., & Weselake, R. J. (2011). Metabolic Engineering of higher plants to produce bio-industrial oils. *Comprehensive Biotechnology*, 67–85. https://doi.org/10.1016/b978-0-08-088504-9.00256-7

Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., & Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics*, *197*(4), 1343–1355. https://doi.org/10.1534/genetics.114.165860

Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breeding and Biotechnology*, *2*(3), 195–212. https://doi.org/10.9787/PBB.2014.2.3.195

The pandas development team (2024) What's new in 2.2.1 https://pandas.pydata.org/docs/whatsnew/v2.2.1.html

Tian, Y., Yang, L., Lu, H., Zhang, B., Li, Y., Liu, C., Ge, T., Liu, Y., Han, J., Li, Y., & Qiu, L. (2022). QTL analysis for plant height and fine mapping of two environmentally stable QTLs with major effects in soybean. *Journal of Integrative Agriculture*, *21*(4), 933–946. https://doi.org/10.1016/s2095-3119(21)63693-6

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences, 108(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108Tsai, H.-Y., Janss, L. L., Andersen, J. R., Orabi, J., Jensen, J. D., Jahoor, A., & Jensen, J. (2020). Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60203-2

Turchetto-Zolet, A. C., Christoff, A. P., Kulcheski, F. R., Loss-Morais, G., Margis, R., & Margis-Pinheiro, M. (2016). Diversity and evolution of plant diacylglycerol acyltransferase (Dgats) unveiled by phylogenetic, gene structure and expression analyses. *Genetics and Molecular Biology*, *39*(4), 524–538. https://doi.org/10.1590/1678-4685-gmb-2016-0024

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., de Andrade, M., Doheny, K. F., Haines, J. L., Hayes, G., Jarvik, G., Jiang, L., Kullo, I. J., Li, R., Ling, H., Manolio, T. A., Matsumoto, M., McCarty, C. A., McDavid, A. N., … Ritchie, M. D. (2011a). Quality Control Procedures for genome-wide association studies. *Current Protocols in Human Genetics*, *68*(1). https://doi.org/10.1002/0471142905.hg0119s68

Turner, S. D., Berg, R. L., Linneman, J. G., Peissig, P. L., Crawford, D. C., Denny, J. C., Roden, D. M., McCarty, C. A., Ritchie, M. D., & Wilke, R. A. (2011b). Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent EMR-linked biobanks. *PLoS ONE*, 6(5), e19586. https://doi.org/10.1371/journal.pone.0019586

U, N. (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany*, *7*, 389-452.

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

USDA. (1996) Crambe, Industrial Rapeseed, and Tung Provide Valuable Oils. https://www.ers.usda.gov/webdocs/outlooks/37342/32910_ius6c_002.pdf?v=2645.4

USDA. (2023). Oilseeds: World markets and trade. https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/k643cc49f/41688v39p/oilseeds.pdf

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Varona, L., Legarra, A., Toro, M. A., & Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Frontiers in Genetics*, *9*, 78. https://doi.org/10.3389/fgene.2018.00078

Varshney, R. K., Roorkiwal, M., & Sorrells, M. E. (2017). Genomic selection for crop improvement. *Springer International Publishing.* https://doi.org/10.1007/978-3-319-63170-7

Veenstra, L. D., Poland, J., Jannink, J., & Sorrells, M. E. (2020). Recurrent genomic selection for wheat grain fructans. *Crop Science*, *60*(3), 1499–1512. https://doi.org/10.1002/csc2.20130

Verges, V. L., & Van Sanford, D. A. (2020). Genomic selection at preliminary yield trial stage: Training population design to predict untested lines. *Agronomy*, *10*(1), 60. https://doi.org/10.3390/agronomy10010060

Vikas, V. K., Pradhan, A. K., Budhlakoti, N., Mishra, D. C., Chandra, T., Bhardwaj, S. C., Kumar, S., Sivasamy, M., Jayaprakash, P., Nisha, R., Shajitha, P., Peter, J., Geetha, M., Mir, R. R., Singh, K., & Kumar, S. (2022). Multi-locus genome-wide association studies (ML-GWAS) reveal novel genomic regions associated with seedling and adult plant stage leaf rust resistance in bread wheat (*Triticum aestivum* L.). *Heredity*, 128(6), 434–449. https://doi.org/10.1038/s41437-022-00525-1

Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., Van Eck, H. J., & Van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics*, *130*(1), 123–135. https://doi.org/10.1007/s00122-016-2798-8

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Wahba, G. (2002). Soft and hard classification by reproducing kernel Hilbert space methods. *Proceedings of the National Academy of Sciences*, *99*(26), 16524–16530. https://doi.org/10.1073/pnas.242574899

Wainschtein, P., Jain, D., Zheng, Z., Cupples, L. A., Shadyab, A. H., McKnight, B., Shoemaker, B. M., Mitchell, B. D., Psaty, B. M., Kooperberg, C., Liu, C.-T., Albert, C. M., Roden, D., Chasman, D. I., Darbar, D., Lloyd-Jones, D. M., Arnett, D. K., Regan, E. A., Boerwinkle, E., … Visscher, P. M. (2019). Recovery of trait heritability from whole genome sequence data. *Europe PMC*. https://doi.org/10.1101/588020

Wallace, J. G., Zhang, X., Beyene, Y., Semagn, K., Olsen, M., Prasanna, B. M., & Buckler, E. S. (2016). Genome-wide association for plant height and flowering time across 15 tropical maize populations under managed drought stress and well-watered conditions in sub-Saharan Africa. *Crop Science*, 56(5), 2365–2378. https://doi.org/10.2135/cropsci2015.10.0632

Wang, X., Wang, H., Long, Y., Li, D., Yin, Y., Tian, J., Chen, L., Liu, L., Zhao, W., Zhao, Y., Yu, L., & Li, M. (2013). Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in b. Napus. *PLoS ONE*, 8(12), e80569. https://doi.org/10.1371/journal.pone.0080569

Wang, N., Li, F., Chen, B., Xu, K., Yan, G., Qiao, J., Li, J., Gao, G., Bancroft, I., Meng, J., King, G. J., & Wu, X. (2014). Genome-wide investigation of genetic changes during modern breeding of *Brassica napus*. *Theoretical and Applied Genetics*, *127*(8), 1817–1829. https://doi.org/10.1007/s00122-014-2343-6

Wang, X., Yang, Z., & Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. *Science Bulletin*, *60*(10), 925–935. https://doi.org/10.1007/s11434-015- 0791-2

Wang, S., Feng, J., Ren, W., Huang, B., Zhou, L., Wen, Y., Zhang, J., Dunwell, J. M., Xu, S., & Zhang, Y. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep19444

Wang, J., Xian, X., Xu, X., Qu, C., Lu, K., Li, J., & Liu, L. (2017a). Genome-wide association mapping of seed coat color in *Brassica napus*. *Journal of Agricultural and Food* Chemistry, 65(26), 5229–5237. https://doi.org/10.1021/acs.jafc.7b01226

Wang, H., Xu, C., Liu, X., Guo, Z., Xu, X., Wang, S., Xie, C., Li, W.-X., Zou, C., & Xu, Y. (2017b). Development of a multiple-hybrid population for genome-wide association

studies: Theoretical consideration and genetic mapping of flowering traits in maize. *Scientific Reports*, *7*(1), 40239. https://doi.org/10.1038/srep40239

Wang, B., Wu, Z., Li, Z., Zhang, Q., Hu, J., Xiao, Y., Cai, D., Wu, J., King, G. J., Li, H., & Liu, K. (2018a). Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. *Plant Biotechnology Journal*, 16(7), 1336–1348. https://doi.org/10.1111/pbi.12873

Wang, J., Zhou, Z., Zhang, Z., Li, H., Liu, D., Zhang, Q., Bradbury, P. J., Buckler, E. S., & Zhang, Z. (2018b). Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. *Heredity*, *121*(6), 648–662. https://doi.org/10.1038/s41437-018-0075-0

Wang J., Zhang Z. (2020) GAPIT Version 3: boosting power and accuracy for genomic association and prediction. bioRxiv:2020.2011.2029.403170

Wang, H., Wang, Q., Pak, H., Yan, T., Chen, M., Chen, X., Wu, D., & Jiang, L. (2021a). Genome-wide association study reveals a patatin-like lipase relating to the reduction of seed oil content in *Brassica napus*. *BMC Plant Biology*, 21(1), 6. https://doi.org/10.1186/s12870-020-02774-w

Wang, Q., Yan, T., Long, Z., Huang, L. Y., Zhu, Y., Xu, Y., Chen, X., Pak, H., Li, J., Wu, D., Xu, Y., Hua, S., & Jiang, L. (2021b). Prediction of heterosis in the recent rapeseed (*Brassica napus*) polyploid by pairing parental nucleotide sequences. *PLOS Genetics*, 17(11), e1009879. https://doi.org/10.1371/journal.pgen.1009879

Wang, P., Xiong, X., Zhang, X., Wu, G., & Liu, F. (2022). A review of Erucic acid production in *Brassicaceae* Oilseeds: Progress and prospects for the genetic engineering of high and low-erucic acid rapeseeds (*Brassica napus*). *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.899076

Wang, Y., Strelkov, S. E., & Hwang, S.-F. (2023). Blackleg yield losses and interactions with verticillium stripe in canola (*Brassica napus*) in Canada. *Plants*, *12*(3), 434. https://doi.org/10.3390/plants12030434

Ward, B. P., Brown-Guedira, G., Tyagi, P., Kolb, F. L., Van Sanford, D. A., Sneller, C. H., & Griffey, C. A. (2019). Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. *Crop Science*, *59*(2), 491–507. https://doi.org/10.2135/cropsci2018.03.0189

Washburn, J. D., Cimen, E., Ramstein, G., Reeves, T., O'Briant, P., McLean, G., Cooper, M., Hammer, G., & Buckler, E. S. (2021). Predicting phenotypes from genetic, environment, management, and historical data using CNNs. *Theoretical and Applied Genetics*, *134*(12), 3997–4011. https://doi.org/10.1007/s00122-021-03943-7

Wassan, G. M., Khanzada, H., Zhou, Q., Mason, A. S., Keerio, A. A., Khanzada, S., Solangi, A. M., Faheem, M., Fu, D., & He, H. (2021). Identification of genetic variation for salt

tolerance in *Brassica napus* using genome-wide association mapping. *Molecular Genetics and Genomics*, 296(2), 391–408. https://doi.org/10.1007/s00438-020-01749-8

Watson, A., Hickey, L. T., Christopher, J., Rutkoski, J., Poland, J., & Hayes, B. J. (2019). Multivariate genomic selection and potential of rapid indirect selection with speed breeding in spring wheat. *Crop Science*, *59*(5), 1945–1959. https://doi.org/10.2135/cropsci2018.12.0757

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

Wen, Z., Boyse, J. F., Song, Q., Cregan, P. B., & Wang, D. (2015). Genomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics*, *16*(1), 671. https://doi.org/10.1186/s12864-015-1872-y

Wen, Y.-J., Zhang, H., Ni, Y.-L., Huang, B., Zhang, J., Feng, J.-Y., Wang, S.-B., Dunwell, J. M., Zhang, Y.-M., & Wu, R. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics*, *19*(4), 700–712. https://doi.org/10.1093/bib/bbw145

Werner, C. R., Qian, L., Voss-Fels, K. P., Abbadi, A., Leckband, G., Frisch, M., & Snowdon, R. J. (2017). Genome-wide regression models considering general and specific combining ability predict hybrid performance in oilseed rape with similar accuracy regardless of trait architecture. *Theoretical and Applied Genetics*, *131*(2), 299–317. https://doi.org/10.1007/s00122-017-3002-5

Werner, C. R., Voss-Fels, K. P., Miller, C. N., Qian, W., Hua, W., Guan, C., Snowdon, R. J., & Qian, L. (2018). Effective genomic selection in a narrow-genepool crop with low-density markers: Asian rapeseed as an example. *The Plant Genome*, *11*(2). https://doi.org/10.3835/plantgenome2017.09.0084

Werner, C. R., Gaynor, R. C., Gorjanc, G., Hickey, J. M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R. J., & Stahl, A. (2020). How population structure impacts genomic selection accuracy in cross-validation: Implications for practical breeding. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.592977

Whelan, J., & Fritsche, K. (2013). Linoleic acid. *Advances in Nutrition*, *4*(3), 311–312. https://doi.org/10.3945/an.113.003772

Whittaker, J. C., Thompson, R., & Denham, M. C. (2000). Marker-assisted selection using ridge regression. *Genetical Research*, *75*(2), 249–252. https://doi.org/10.1017/S0016672399004462

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J.-L., Sorrells, M. E., Raman, B., Cairns, J. E., Tarekegne, A., Semagn, K., Beyene, Y., Grudloyma, P., Technow, F., Riedelsheimer, C., & Melchinger, A. E. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 Genes|Genomes|Genetics*, *2*(11), 1427–1436. https://doi.org/10.1534/g3.112.003699

Wittkop, B., Snowdon, R. J., & Friedt, W. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica*, *170*(1–2), 131. https://doi.org/10.1007/s10681-009-9940-5

Wolinetz, M. H., & Hein, M. (2017). Tracking biofuel consumption, feedstocks and avoided greenhouse gas emissions. https://www.naviusresearch.com/wp-content/uploads/2022/11/Biofuels-in-Canada-2022-Final-2022-10-27.pdf

Wu, G., Lang, C., & Chen, J. (2007). Breeding and application of high erucic acid oleaginous rape for industrial use. *J. Nucl. Agric.*, *21*, 374–377. https://doi.org/10.3969/j.issn.1000-8551.2007.04.013

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://doi.org/10.3389/fpls.2016.01418

Wu, J., Lin, L., Xu, M., Chen, P., Liu, D., Sun, Q., Ran, L., & Wang, Y. (2018). Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4966-5

Würschum, T., Abel, S., & Zhao, Y. (2014). Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. *Plant Breeding*, 133(1), 45–51. https://doi.org/10.1111/pbr.12137

Xiang, Y., Helal, M. M. U., Liang, L., Zhang, M., Han, H., Dai, X., Zhao, J., Chen, D., Wang, X., Li, M., & Zhao, C. (2023). Genome-wide association study exposed the pleiotropic genes for yield-related and oil quality traits in *Brassica napus* L. *Oil Crop Science*, *8*(3), 156–164. https://doi.org/10.1016/j.ocsci.2023.08.001

Xiao, Y., Liu, H., Wu, L., Warburton, M., & Yan, J. (2017). Genome-wide association studies in maize: Praise and stargaze. *Molecular Plant*, *10*(3), 359–374. https://doi.org/10.1016/j.molp.2016.12.008

Xiao, Z., Zhang, C., Tang, F., Yang, B., Zhang, L., Liu, J., Huo, Q., Wang, S., Li, S., Wei, L., Du, H., Qu, C., Lu, K., Li, J., & Li, N. (2019). Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop *Brassica napus*. *Biotechnology for Biofuels*, 12(1), 216. https://doi.org/10.1186/s13068-019-1557-x

Xiao, Y., Xi, Z., Wang, F., & Wang, J. (2023). Genomic asymmetric epigenetic modification of transposable elements is involved in gene expression regulation of allopolyploid *Brassica napus*. *The Plant Journal*, *117*(1), 226–241. https://doi.org/10.1111/tpj.16491

Xin, S., Dong, H., Cui, Y., Liu, Y., Tian, G., Deng, N., Wan, H., Liu, Z., Li, X., & Qian, W. (2023). Identification of a candidate QTG for seed number per silique by integrating QTL mapping and RNA-seq in *Brassica napus* L. *The Crop Journal*, *11*(1), 189–197. https://doi.org/10.1016/j.cj.2022.07.012

Xiong, Z., Gaeta, R. T., & Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences*, *108*(19), 7908–7913. https://doi.org/10.1073/pnas.1014138108

Xu, Yunbi, & 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

Xu, Y. (2010). Molecular plant breeding. *Experimental Agriculture*, *47*(1), 734. https://doi.org/10.1017/s0014479710000669

Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., & Prasanna, B. M. (2012). Whole-genome strategies for marker-assisted plant breeding. *Molecular Breeding*, *29*(4), 833–854. https://doi.org/10.1007/s11032-012-9699-6

Xu, L., Hu, K., Zhang, Z., Guan, C., Chen, S., Hua, W., Li, J., Wen, J., Yi, B., Shen, J., Ma, C., Tu, J., & Fu, T. (2016). Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Research*. https://doi.org/10.1093/dnares/dsv035

Xu, P., Li, H., Xu, K., Cui, X., Liu, Z., & Wang, X. (2022). *Genetic Variation in Bngrp1 Contributes to Low Phosphorus Tolerance in Brassica napus*. https://doi.org/10.1101/2022.07.14.500146

Xue, Y., Chen, B., Wang, R., Win, A. N., Li, J., & Chai, Y. (2018). Genome-wide survey and characterization of fatty acid desaturase gene family in *Brassica napus* and its parental species. *Applied Biochemistry and Biotechnology*, 184(2), 582–598. https://doi.org/10.1007/s12010-017-2563-8

Yadav, R., Humphreys, G., Laroche, A., Randhawa, H., Soto-Cerda, B., Domaratzki, M., McCartney, C., & Ragupathy, R. (2023). Recent advances in genomic selection (GS) in wheat: Translating GS advances into actual genetic gains in wheat breeding programs [Literature Review].

Yan, S., He, J., Tang, M., Ming, B., Li, H., Fan, S., Xiong, Y., Chao, H., Zhang, L., Wang, A., & Li, M. (2023). Dissecting the meiotic recombination patterns in a *Brassica napus* double

haploid population using 60k SNP Array. *International Journal of Molecular Sciences*, *24*(5), 4469. https://doi.org/10.3390/ijms24054469

Yang, Y., Lai, K., Tai, P., & Li, W. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *Journal of Molecular Evolution*, 48(5), 597–604. https://doi.org/10.1007/PL00006502

Yang, Y., Wang, W., Hu, Q., Raman, H., & Liu, J. (2023). Genome-wide association and RNA-seq analyses identify loci for pod orientation in rapeseed (*Brassica napus*). *Frontiers in Plant Science*, 13, 1097534. https://doi.org/10.3389/fpls.2022.1097534

Yao, M., Guan, M., Zhang, Z., Zhang, Q., Cui, Y., Chen, H., Liu, W., Jan, H. U., Voss-Fels, K. P., Werner, C. R., He, X., Liu, Z., Guan, C., Snowdon, R. J., Hua, W., & Qian, L. (2020). GWAS and co-expression network combination uncovers multigenes with close linkage effects on the oleic acid content accumulation in *Brassica napus*. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-020-6711-0

Yao, M., Guan, M., Yang, Q., Huang, L., Xiong, X., Jan, H. U., Voss-Fels, K. P., Werner, C. R., He, X., Qian, W., Snowdon, R. J., Guan, C., Hua, W., & Qian, L. (2021). Regional association analysis coupled with transcriptome analyses reveal candidate genes affecting seed oil accumulation in *Brassica napus*. *Theoretical and Applied Genetics*, 134(5), 1545–1555. https://doi.org/10.1007/s00122-021-03788-0

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). Rmvp: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*, 19(4), 619–628. https://doi.org/10.1016/j.gpb.2020.10.007

Yoosefzadeh-Najafabadi, M., Rajcan, I., & Eskandari, M. (2022). Optimizing genomic selection in soybean: An important improvement in agricultural genomics. *Heliyon*, *8*(11), e11873. https://doi.org/10.1016/j.heliyon.2022.e11873

Young, N. D. (1994). Constructing a plant genetic linkage map with DNA markers. *Advances in Cellular and Molecular Biology of Plants*, 39–57. https://doi.org/10.1007/978-94-011-1104-1_3

Young, A. I. (2019). Solving the missing heritability problem. *PLOS Genetics*, *15*(6). https://doi.org/10.1371/journal.pgen.1008222

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. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Yu, F., Zhang, X., Peng, G., Falk, K. C., Strelkov, S. E., & Gossen, B. D. (2017). Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of

Plasmodiophora Brassicae in *Brassica rapa*. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-04903-2

Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., Li, J., & Simianer, H. (2014). Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS ONE*, *9*(3), e93017. https://doi.org/10.1371/journal.pone.0093017

Zhang, J., Mason, A. S., Wu, J., Liu, S., Zhang, X., Luo, T., Redden, R., Batley, J., Hu, L., & Yan, G. (2015). Identification of putative candidate genes for water stress tolerance in canola (*Brassica napus*). *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.01058

Zhang, J., Song, Q., Cregan, P. B., & Jiang, G.-L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics*, *129*(1), 117–130. https://doi.org/10.1007/s00122-015-2614-x

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., Cui, Z., Ruan, Y., Burgueño, J., San Vicente, F., Olsen, M., Prasanna, B. M., Crossa, J., Yu, H., & Zhang, X. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in Plant Science*, *8*, 1916. https://doi.org/10.3389/fpls.2017.01916

Zhang, H., Yin, L., Wang, M., Yuan, X., & Liu, X. (2019). Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Frontiers in Genetics*, 10, 189. https://doi.org/10.3389/fgene.2019.00189

Zhang, G., Zhou, J., Peng, Y., Tan, Z., Li, L., Yu, L., Jin, C., Fang, S., Lu, S., Guo, L., & Yao, X. (2022). Genome-wide association studies of salt tolerance at seed germination and seedling stages in *Brassica napus*. *Frontiers in Plant Science*, 12, 772708. https://doi.org/10.3389/fpls.2021.772708

Zhang, Y., Zhang, M., Ye, J., Xu, Q., Feng, Y., Xu, S., Hu, D., Wei, X., Hu, P., & Yang, Y. (2023). Integrating genome-wide association study into genomic selection for the prediction of agronomic traits in rice (*Oryza sativa* L.). *Molecular Breeding*, *43*(11), 81. https://doi.org/10.1007/s11032-023-01423-y

Zhang-Biehn, S., Fritz, A. K., Zhang, G., Evers, B., Regan, R., & Poland, J. (2021). Accelerating wheat breeding for end-use quality through association mapping and multivariate genomic prediction. *The Plant Genome*, *14*(3), e20164. https://doi.org/10.1002/tpg2.20164

Zhao, J., Dimov, Z., Becker, H. C., Ecke, W., & Möllers, C. (2007). Mapping QTL controlling fatty acid composition in a doubled haploid rapeseed population segregating for oil content. *Molecular Breeding*, *21*(1), 115–125. https://doi.org/10.1007/s11032-007-9113-y

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H., & Reif, J. C. (2014). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity*, 112(6), 638–645. https://doi.org/10.1038/hdy.2014.1

Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Würschum, T., Mock, H.-P., Matros, A., Ebmeyer, E., Schachschneider, R., Kazman, E., Schacht, J., Gowda, M., Longin, C. F. H., & Reif, J. C. (2015). Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proceedings of the National Academy of Sciences*, *112*(51), 15624–15629. https://doi.org/10.1073/pnas.1514547112

Zhao, Q., Wu, J., Cai, G., Yang, Q., Shahid, M., Fan, C., Zhang, C., & Zhou, Y. (2019). A novel quantitative trait locus on chromosome A9 controlling oleic acid content in *Brassica napus*. *Plant Biotechnology Journal*, 17(12), 2313–2324. https://doi.org/10.1111/pbi.13142

Zhao, W., Lai, X., Liu, D., Zhang, Z., Ma, P., Wang, Q., Zhang, Z., & Pan, Y. (2020). Applications of support vector machine in genomic prediction in pig and maize populations. *Frontiers in Genetics*, *11*, 598318. https://doi.org/10.3389/fgene.2020.598318

Zhao, C., Xie, M., Liang, L., Yang, L., Han, H., Qin, X., Zhao, J., Hou, Y., Dai, W., Du, C., Xiang, Y., Liu, S., & Huang, X. (2022). Genome-wide association analysis combined with quantitative trait loci mapping and dynamic transcriptome unveil the genetic control of seed oil content in *Brassica napus* L. *Frontiers in Plant Science*, 13, 929197. https://doi.org/10.3389/fpls.2022.929197

Zhou, H., Xiao, X., Asjad, A., Han, D., Zheng, W., Xiao, G., Huang, Y., & Zhou, Q. (2022). Integration of GWAS and transcriptome analyses to identify SNPs and candidate genes for aluminum tolerance in rapeseed (*Brassica napus* L.). *BMC Plant Biology*, *22*(1). https://doi.org/10.1186/s12870-022-03508-w

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences*, 420–427. https://doi.org/10.7150/ijbs.3.420

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, *1*(1), plantgenome2008.02.0089. https://doi.org/10.3835/plantgenome2008.02.0089

Zhu, C., Li, X., & Yu, J. (2011). Integrating rare-variant testing, function prediction, and gene network in composite resequencing-based genome-wide association studies (CR-GWAS). *G3 Genes|Genomes|Genetics*, *1*(3), 233–243. https://doi.org/10.1534/g3.111.000364

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

Zou, J., Zhao, Y., Liu, P., Shi, L., Wang, X., Wang, M., Meng, J., & Reif, J. C. (2016). Seed quality traits can be predicted with high accuracy in *Brassica napus* using genomic data. *PLOS ONE*, *11*(11). https://doi.org/10.1371/journal.pone.0166624

8.2 Appendices

8.2.1 List of abbreviations

AAFC	Agriculture and Agri-Food Canada
BBCH-scale	Biologische bundesanstalt bundessortenamt und chemische industrie scale
BayesA	Bayesian A regression method
BayesA	Bayesian B regression method
BayesC	Bayesian C regression method
BGLR	Bayesian generalized linear regression
BLUP	Best linear unbiased prediction
bp	Base pairs
BRR	Bayesian ridge regression
CI	Confidence interval
CMLM	Compression mixed linear model
CMS	Cytoplasmic male sterility
CoA	Coenzyme A
CTAB	Cetyl methylammonium bromide
CV	Cross validation
CV %	Coefficient of variation percentage
DH	Doubled haploid
DNA	Deoxyribonucleic acid
ERU	Erucic acid content
EU	European Union

FarmCPU	Fixed and random model circulating probability unification
GAPIT	Genome association and prediction integrated tool
GC	Gas chromatography
GBLUP	Genomic best linear unbiased prediction
GBS	Genotyping-by-sequencing
GEBV	Genomic estimated breeding value
GS	Genomic selection
GWAS	Genome-wide association study
HEAR	High erucic acid rapeseed
INRA	Institut National de la Recherche Agronomique
Kb	Kilobase pairs
kg ha^{-1}	Kilograms per hectare
L ha^{-1}	Litres per hectare
LD	Linkage disequilibrium
LLE	Linoleic acid content
LLN	Linolenic acid content
MAF	Minor allele frequency
MAS	Marker-assisted selection
Mb	Megabase pairs
ML	Machine learning
ML-GWAS	Multi-locus genome-wide association study
MLM	Mixed linear model
MLM+K	Mixed linear models considering kinship

MLM+K+PCA	Mixed linear models considering kinship and subpopulation structure via principle component analyses
MLMM	Multi-locus mixed model
MS	Marker set
MTA	Marker-trait association
ng/μL	Nanograms per microlitre
NIR	Near infrared
OIL	Overall oil content
OLE	Oleic acid content
PCA	Principal component analysis
QTL	Quantitative trait loci
Q-Q	Quantile-quantile
RCBD	Randomized complete block design
RDF	Radial basis function
RF	Random forest
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
rrBLUP	Ridge regression best linear unbiased prediction
SB	Speed breeding
SL-GWAS	Single-locus genome-wide association study model
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
SVR	Support vector regression

t ha^{-1}	Tonnes per hectare
TASSEL	Trait analysis by association, evolution and linkage
TAG	Triacylglycerol
TP	Training population
VP	Validation population/test population
WGR	Whole-genome regression
XGB	Extreme gradient boosting

8.2.2 Supplemental tables and figures

Table S3.1 Summary of locations and soil types of field experiments for 92 *Brassica napus* L. parent genotypes tested across Manitoba in randomized complete block designs from 2016-2018.

Site-year #	Year	Site	Soil type
1	2016	Glenlea	Rego black chernozem (typic hapludert) of scantenbury series
2	2017	Carman	Orthic blacks developed on sandy, coarse loamy, fine loamy and clayey sediments
3	2017	Portage la Prairie	Chernozemic with occurrence of regosolic and gleysolic
4	2018	Glenlea	Rego black chernozem (typic hapludert) of scantenbury series
5	2018	Portage la Prairie	Chernozemic with occurrence of regosolic and gleysolic

Adapted from Sun, J. (2021) Genome-wide association and genomic selection in *Brassica napus* L. University of Manitoba

Table S3.2 Summary of locations and number of 362 *Brassica napus* L. hybrid genotypes tested across Alberta, Saskatchewan, and Manitoba in 2014-2018.

Site-year #	Year	Site	Number of genotypes evaluated
1	2014	Bison (Winnipeg), MB	34
2	2014	Carman, MB	37
3	2015	Arboretum (Winnipeg), MB	31
4	2015	Bison (Winnipeg), MB	182
5	2015	Carman, MB	182
6	2015	Portage la Prairie, MB	182
7	2016	Arboretum (Winnipeg), MB	35
8	2016	Bison (Winnipeg), MB	198
9	2016	Carman, MB	236
10	2016	Holland, MB	3
11	2016	Killam, AB	3
12	2016	North Battleford, SK	3
13	2016	Pense, SK	3
14	2016	Portage la Prairie, MB	233
15	2016	Rosebank, MB	3
16	2016	Rosetown, SK	3
17	2016	Saint Albert, AB	3
18	2016	Thornhill, MB	3
19	2016	Wawanesa, MB	3
20	2016	Yellow Grass, SK	3
21	2017	Bison (Winnipeg), MB	35
22	2017	Carman, MB	45
23	2017	Holland, MB	10
24	2017	Killam, AB	10
25	2017	Lake Lenore, SK	10
26	2017	Marquis, SK	10
27	2017	Portage la Prairie, MB	35
28	2017	Rosebank, MB	10
29	2017	Rosetown, SK	10
30	2017	Saint Albert, AB	10
31	2017	Thornhill, MB	10
32	2017	Vanscoy, SK	10
33	2017	Watrous, SK	10
34	2018	Carman, MB	15
35	2018	Carstairs, AB	9
36	2018	Killam, AB	1
37	2018	Lake Lenore, SK	9

38	2018	Marquis, SK	9
39	2018	Portage la Prairie, MB	15
40	2018	Rosebank, MB	9
41	2018	Saint Albert, AB	10
42	2018	Vanscoy, SK	10
43	2018	Watrous, SK	10

Adapted from Sun, J. (2021) Genome-wide association and genomic selection in *Brassica napus* L. University of Manitoba

Table S3.3 Best linear unbiased prediction (BLUP) values of five seed quality traits in a population of 427 *Brassica napus* L. genotypes. BLUPs summarize raw site-year data collected from 48 site-years.

Name	Line	ERU[1]	LLE[2]	LLN[3]	OLE[4]	OIL[5]
P1	R-line	46.04	12.43	7.86	15.08	45.61
P2	R-line	46.24	12.79	8.11	15.01	43.73
P3	R-line	44.56	11.85	7.44	17.65	44.05
P4	R-line	44.08	12.09	7.40	17.74	45.43
P5	R-line	42.62	11.21	7.91	18.75	46.13
P6	R-line	46.55	12.48	7.81	14.76	45.95
P7	R-line	44.90	12.92	7.27	16.70	43.44
P8	R-line	49.85	12.87	8.01	13.08	43.63
P9	R-line	41.98	12.28	7.47	19.24	44.32
P10	R-line	47.96	12.37	7.85	14.33	45.54
P11	R-line	45.06	12.03	7.89	15.82	44.78
P12	R-line	48.43	12.14	7.69	14.48	44.19
P13	R-line	47.23	11.54	8.33	14.59	46.18
P14	R-line	47.73	11.83	8.71	15.18	43.98
P15	R-line	53.75	11.70	7.53	12.78	46.78
P16	R-line	50.10	12.11	7.95	13.52	45.11
P17	R-line	49.88	11.84	8.83	12.58	43.14
P18	R-line	54.07	11.25	8.99	11.39	41.77
P19	R-line	52.01	12.23	8.47	12.37	41.07
P20	R-line	53.68	11.18	8.45	12.08	43.10
P21	R-line	49.94	12.30	9.72	11.60	41.18
P22	R-line	51.67	11.37	8.30	12.79	43.79
P23	R-line	52.37	10.89	8.76	12.71	43.44
P24	R-line	50.54	11.56	8.21	13.43	41.07
P25	R-line	50.58	11.72	8.58	12.88	45.40
P26	R-line	50.57	11.30	8.83	12.79	44.06
P27	R-line	50.55	11.91	8.90	12.71	44.91
P28	R-line	51.49	11.33	9.05	12.30	45.22
P29	R-line	51.42	11.35	9.15	12.29	45.00
P30	R-line	53.70	11.70	8.93	11.57	42.37
P31	R-line	51.67	11.08	8.63	13.00	43.51
P32	R-line	50.44	10.65	8.21	14.14	44.42
P33	R-line	49.44	10.98	7.60	15.45	45.93
P34	R-line	48.90	11.24	8.24	15.29	43.73
P35	R-line	47.23	12.07	7.51	15.34	43.86
P36	R-line	47.42	13.42	6.84	15.15	42.97
P37	R-line	47.27	13.41	7.12	14.95	42.37

P38	R-line	47.56	13.19	7.00	15.05	42.90
P39	R-line	46.34	11.96	8.50	15.75	42.65
P40	R-line	50.04	11.68	9.33	12.29	42.20
P41	R-line	49.47	11.99	8.69	12.90	42.99
P42	R-line	50.19	11.91	8.89	12.48	43.13
P44	R-line	46.77	11.73	9.77	14.38	43.75
P45	R-line	46.04	12.06	9.65	14.46	43.87
P46	R-line	51.17	11.40	9.59	12.46	44.25
P47	R-line	50.57	11.73	9.09	12.61	43.67
P48	R-line	51.61	11.13	9.46	12.66	43.83
P49	R-line	52.98	11.10	9.86	11.67	44.04
P50	R-line	52.00	11.86	8.82	11.95	45.76
P51	R-line	54.28	8.99	7.91	14.64	41.09
P52	R-line	56.50	8.40	7.62	14.83	42.78
P53	R-line	56.86	6.81	6.79	16.76	41.76
P54	R-line	52.04	12.04	7.75	13.13	43.79
P55	R-line	56.32	7.76	6.85	16.29	39.81
P56	R-line	53.73	8.46	7.20	16.20	40.58
P57	R-line	53.80	12.30	8.23	11.70	43.16
P58	R-line	54.43	12.48	8.38	10.34	44.08
P59	R-line	56.31	7.65	6.86	16.32	43.76
P60	R-line	55.02	7.08	7.74	16.24	44.34
P61	R-line	53.73	11.16	7.59	13.21	42.76
P62	B-line	46.65	10.83	9.39	15.00	44.34
P63	B-line	47.81	10.32	9.82	14.90	47.71
P64	B-line	55.10	11.40	8.68	11.42	45.77
P65	B-line	50.49	9.75	9.02	14.55	46.45
P66	B-line	49.28	10.83	8.00	15.08	46.55
P67	B-line	52.11	10.25	8.91	13.72	46.82
P68	B-line	51.38	10.63	8.38	14.17	47.04
P69	B-line	51.30	10.81	7.94	14.21	45.85
P70	B-line	53.64	11.71	8.37	12.39	48.20
P71	B-line	49.94	10.05	8.85	14.48	46.36
P72	B-line	48.63	10.92	8.20	15.20	46.09
P73	B-line	53.06	10.57	8.81	13.02	46.68
P74	B-line	51.53	10.41	8.57	14.02	46.75
P75	B-line	51.49	11.65	8.34	12.80	46.19
P76	B-line	51.99	10.23	8.95	13.61	47.59
P77	B-line	58.09	7.43	7.46	15.28	44.49
P78	B-line	58.60	6.99	6.70	16.17	47.13
P79	B-line	58.12	6.92	7.28	15.68	48.22

P80	B-line	51.01	9.94	8.99	14.25	48.41
P81	B-line	49.43	10.27	9.39	14.28	47.96
P82	B-line	54.27	11.04	7.97	12.66	47.02
P83	B-line	52.11	11.27	8.52	13.03	47.50
P84	B-line	56.23	6.86	8.06	15.99	47.20
P85	B-line	58.86	6.12	6.04	17.63	47.04
P86	B-line	58.86	6.43	7.31	16.10	46.25
P87	B-line	57.80	7.34	7.87	15.03	47.20
P88	B-line	58.16	6.55	6.77	17.04	47.02
P89	B-line	55.47	6.06	5.92	19.16	46.56
P90	B-line	52.64	10.67	8.50	13.39	45.41
P91	B-line	53.64	11.00	7.74	13.38	49.02
P92	B-line	54.65	11.67	8.64	11.23	47.42
H1	Hybrid	51.59	10.99	9.78	12.30	46.10
H2	Hybrid	50.84	10.39	9.59	13.44	45.39
H5	Hybrid	48.65	11.27	9.34	13.91	36.85
H6	Hybrid	47.82	10.89	9.22	14.93	46.84
H7	Hybrid	46.22	11.19	8.67	15.99	46.91
H8	Hybrid	45.13	10.54	8.67	16.90	46.03
H9	Hybrid	48.13	11.29	9.33	13.93	47.68
H10	Hybrid	47.54	11.28	8.91	14.91	46.60
H11	Hybrid	50.04	11.19	8.81	13.57	47.50
H12	Hybrid	46.45	11.57	9.06	15.34	45.41
H13	Hybrid	48.13	11.12	9.11	14.48	47.52
H14	Hybrid	45.61	11.02	8.63	16.01	45.45
H16	Hybrid	50.72	11.15	8.77	13.34	46.35
H17	Hybrid	49.15	10.77	8.80	14.40	46.10
H18	Hybrid	50.11	11.14	8.92	13.20	46.06
H19	Hybrid	48.13	10.79	9.06	14.43	45.82
H20	Hybrid	48.63	10.65	8.28	15.12	46.02
H21	Hybrid	51.92	11.15	8.27	13.16	46.62
H22	Hybrid	50.80	10.90	8.40	13.96	46.19
H23	Hybrid	48.52	11.87	7.90	14.70	46.73
H24	Hybrid	47.71	11.62	8.13	15.13	46.74
H25	Hybrid	44.63	11.94	8.25	16.50	45.32
H26	Hybrid	50.58	11.37	8.08	13.86	46.75
H27	Hybrid	49.67	11.72	8.00	13.76	47.15
H28	Hybrid	49.45	10.78	8.37	14.55	47.10
H29	Hybrid	50.34	11.00	8.11	14.35	46.32
H30	Hybrid	50.91	10.79	8.27	14.23	46.42
H31	Hybrid	49.08	11.22	8.12	14.71	46.35

H32	Hybrid	48.10	11.90	7.96	15.08	46.98
H33	Hybrid	51.60	11.25	8.30	13.02	46.56
H34	Hybrid	50.74	10.91	8.47	13.79	45.33
H35	Hybrid	49.67	10.97	8.52	14.15	45.61
H36	Hybrid	50.54	11.56	8.54	13.37	44.46
H37	Hybrid	50.04	11.47	8.66	13.17	45.29
H42	Hybrid	53.10	10.84	8.33	13.09	44.41
H43	Hybrid	53.29	10.98	8.64	12.87	44.94
H44	Hybrid	52.29	10.85	8.65	13.09	45.33
H45	Hybrid	54.50	10.81	8.74	12.32	44.35
H46	Hybrid	53.05	10.74	8.33	13.24	45.33
H47	Hybrid	53.81	10.46	8.79	12.78	45.05
H48	Hybrid	53.26	10.72	8.66	12.96	45.26
H49	Hybrid	51.88	11.13	8.08	13.79	44.39
H50	Hybrid	51.99	11.34	8.37	13.45	44.59
H53	Hybrid	53.61	10.93	8.45	12.88	45.45
H54	Hybrid	51.70	11.31	8.10	13.57	45.29
H56	Hybrid	53.51	11.21	8.54	12.59	44.30
H57	Hybrid	55.32	10.70	8.92	11.54	44.83
H58	Hybrid	55.00	10.32	9.09	11.95	45.19
H59	Hybrid	52.42	10.97	8.21	13.45	45.44
H60	Hybrid	50.86	10.74	9.04	13.47	44.17
H62	Hybrid	51.19	11.92	9.30	12.27	43.69
H66	Hybrid	52.63	10.05	8.90	13.39	45.44
H67	Hybrid	51.51	10.38	8.76	13.83	45.41
H68	Hybrid	51.33	10.08	9.14	13.66	45.45
H69	Hybrid	48.07	11.43	8.74	15.42	45.14
H71	Hybrid	52.43	10.48	9.17	13.01	46.16
H74	Hybrid	53.87	10.64	8.60	12.67	45.04
H76	Hybrid	52.51	10.85	8.97	12.69	45.28
H78	Hybrid	52.04	10.04	8.33	14.57	45.98
H79	Hybrid	52.12	10.02	8.11	14.50	42.74
H80	Hybrid	50.43	10.75	8.49	14.13	46.17
H81	Hybrid	53.68	11.09	8.30	12.84	45.27
H82	Hybrid	50.59	10.89	7.87	14.37	45.13
H83	Hybrid	50.69	11.03	7.97	14.26	45.34
H84	Hybrid	48.98	11.01	8.26	14.56	45.15
H85	Hybrid	51.13	11.62	8.07	13.59	45.36
H87	Hybrid	50.61	11.77	7.81	14.05	44.39
H88	Hybrid	49.85	11.90	7.99	14.04	44.81
H90	Hybrid	50.37	11.56	7.76	14.23	45.02

H91	Hybrid	50.61	11.21	7.99	14.21	44.87
H92	Hybrid	48.90	11.78	7.51	14.78	44.50
H93	Hybrid	52.16	11.88	7.75	13.23	44.99
H94	Hybrid	51.33	11.59	7.44	14.11	44.68
H95	Hybrid	50.99	11.51	7.67	14.26	45.19
H96	Hybrid	49.35	11.87	7.91	14.35	44.17
H97	Hybrid	51.45	12.40	7.68	13.39	44.81
H98	Hybrid	49.37	10.79	8.56	14.75	44.89
H99	Hybrid	49.52	10.54	8.78	14.82	45.45
H100	Hybrid	47.87	10.85	8.70	15.10	45.44
H102	Hybrid	51.36	10.89	8.98	13.27	43.97
H103	Hybrid	52.10	11.05	9.06	12.50	44.66
H104	Hybrid	49.96	10.97	9.03	13.38	43.75
H106	Hybrid	51.48	11.02	8.74	13.06	45.36
H107	Hybrid	51.89	11.03	8.79	12.95	45.84
H109	Hybrid	49.20	11.43	8.52	14.29	45.77
H110	Hybrid	49.19	10.54	9.13	14.72	45.22
H111	Hybrid	50.14	10.74	9.03	14.05	45.23
H112	Hybrid	47.55	10.85	9.00	15.22	44.74
H113	Hybrid	49.36	10.85	9.04	14.85	45.93
H114	Hybrid	49.14	10.54	8.94	14.74	45.08
H115	Hybrid	49.68	10.87	9.17	14.09	44.88
H116	Hybrid	47.44	11.00	9.12	14.87	44.45
H117	Hybrid	47.74	10.37	8.74	16.15	45.08
H118	Hybrid	45.53	10.85	7.79	17.64	44.37
H119	Hybrid	48.71	10.72	8.26	15.31	45.19
H120	Hybrid	46.02	11.05	7.99	16.67	45.14
H121	Hybrid	44.68	11.96	7.75	17.30	45.05
H122	Hybrid	52.22	10.35	8.61	13.47	44.66
H123	Hybrid	52.33	10.26	8.99	13.22	45.11
H124	Hybrid	51.47	11.15	8.17	13.70	45.25
H125	Hybrid	52.07	10.30	8.59	13.77	44.94
H126	Hybrid	51.90	10.30	8.65	13.54	44.94
H127	Hybrid	52.31	10.10	8.78	13.71	45.02
H128	Hybrid	51.98	10.76	8.32	13.62	44.76
H129	Hybrid	52.01	10.74	8.40	13.54	44.57
H130	Hybrid	52.31	10.53	8.29	13.66	44.86
H131	Hybrid	50.41	10.66	8.37	14.28	44.92
H132	Hybrid	51.70	10.58	8.31	13.92	45.29
H133	Hybrid	51.01	10.95	8.57	13.62	45.29
H134	Hybrid	52.61	10.62	8.33	13.52	45.35

H135	Hybrid	52.62	10.95	8.57	13.01	45.10
H136	Hybrid	52.41	10.30	8.73	13.36	45.69
H137	Hybrid	53.09	10.33	8.50	13.34	45.29
H138	Hybrid	52.44	10.54	9.00	13.08	44.41
H139	Hybrid	52.40	10.58	9.19	12.64	43.77
H140	Hybrid	51.61	11.49	8.49	12.99	44.72
H141	Hybrid	50.72	10.53	8.61	14.17	44.09
H142	Hybrid	49.72	10.53	8.89	14.36	44.52
H143	Hybrid	52.28	11.32	8.12	13.21	46.02
H144	Hybrid	52.03	10.22	8.80	13.59	43.80
H145	Hybrid	51.34	10.19	9.15	13.51	44.14
H146	Hybrid	52.21	11.18	8.41	13.31	45.09
H147	Hybrid	48.91	10.76	8.54	14.90	43.56
H148	Hybrid	49.04	10.32	8.63	15.09	43.93
H149	Hybrid	49.57	11.43	8.36	14.10	44.58
H150	Hybrid	47.88	10.67	8.62	15.24	45.60
H151	Hybrid	49.90	10.72	8.65	14.05	45.71
H152	Hybrid	48.79	10.61	8.87	14.58	45.59
H153	Hybrid	49.91	11.49	8.51	13.90	46.00
H154	Hybrid	48.46	10.76	8.54	15.27	44.82
H155	Hybrid	50.61	10.84	8.91	13.88	45.52
H156	Hybrid	47.40	10.82	8.87	15.42	45.16
H157	Hybrid	49.57	11.66	8.36	14.42	46.38
H158	Hybrid	51.74	10.93	8.16	13.67	46.38
H159	Hybrid	53.87	10.97	8.08	12.84	46.48
H160	Hybrid	51.36	10.77	8.27	13.83	45.78
H161	Hybrid	52.26	11.98	7.83	13.14	47.99
H162	Hybrid	49.84	10.72	8.23	14.54	45.53
H163	Hybrid	49.35	10.43	8.55	14.91	45.72
H164	Hybrid	49.67	10.47	8.79	14.29	44.97
H165	Hybrid	49.03	10.64	8.92	14.41	44.90
H166	Hybrid	53.77	10.71	8.85	12.51	45.21
H167	Hybrid	55.18	8.99	8.09	13.99	43.38
H168	Hybrid	55.95	8.91	7.79	13.99	45.53
H169	Hybrid	55.48	9.04	8.31	13.61	44.86
H170	Hybrid	53.77	10.99	9.23	11.85	44.15
H171	Hybrid	54.56	9.29	8.49	13.49	42.13
H172	Hybrid	55.95	9.13	7.95	13.62	45.31
H173	Hybrid	55.17	9.21	8.42	13.46	43.87
H174	Hybrid	52.95	10.91	8.80	12.60	45.02
H175	Hybrid	53.32	9.08	7.88	14.81	43.34

H176	Hybrid	55.34	8.84	7.54	14.67	45.61
H177	Hybrid	53.89	9.73	8.20	13.85	44.33
H178	Hybrid	52.82	10.69	9.16	12.51	44.30
H179	Hybrid	54.39	8.68	8.15	14.43	42.17
H180	Hybrid	55.14	8.86	8.06	14.26	45.02
H181	Hybrid	54.53	9.13	8.53	13.89	43.83
H182	Hybrid	50.19	11.27	8.55	13.85	43.98
H185	Hybrid	52.73	9.73	8.04	14.47	42.87
H186	Hybrid	50.48	11.36	8.71	13.44	44.84
H187	Hybrid	52.43	9.12	8.03	14.93	44.14
H188	Hybrid	53.70	9.06	7.58	14.77	45.99
H189	Hybrid	51.92	9.44	8.09	14.94	45.71
H190	Hybrid	49.71	11.51	8.67	14.01	45.29
H191	Hybrid	52.47	9.37	8.04	14.89	43.66
H193	Hybrid	51.57	10.48	8.11	14.41	45.77
H194	Hybrid	50.25	10.85	8.87	13.45	45.08
H195	Hybrid	52.99	11.56	8.14	12.79	45.76
H196	Hybrid	54.76	9.40	7.32	14.60	44.07
H197	Hybrid	55.74	9.49	7.08	14.36	46.88
H198	Hybrid	54.34	9.71	7.54	14.45	44.54
H199	Hybrid	51.66	11.22	8.31	13.32	46.60
H200	Hybrid	54.21	9.35	7.77	14.47	43.78
H201	Hybrid	55.13	9.40	7.13	14.50	46.69
H202	Hybrid	53.65	9.40	7.57	14.73	46.09
H203	Hybrid	50.47	11.20	8.83	13.26	44.76
H204	Hybrid	53.66	9.16	8.04	14.40	43.49
H205	Hybrid	54.01	8.94	7.77	14.77	45.09
H206	Hybrid	52.47	9.70	8.17	14.39	44.04
H207	Hybrid	47.96	11.75	9.16	13.83	43.45
H208	Hybrid	49.26	11.90	8.79	13.26	43.57
H209	Hybrid	50.23	11.19	8.87	13.77	44.06
H210	Hybrid	51.54	11.11	9.03	13.15	44.49
H211	Hybrid	51.24	11.26	8.27	14.05	44.14
H212	Hybrid	51.47	10.60	8.29	14.09	44.23
H213	Hybrid	52.78	10.91	8.50	13.32	44.92
H214	Hybrid	50.98	10.25	8.55	14.25	45.27
H215	Hybrid	49.39	11.12	8.56	14.49	45.00
H216	Hybrid	48.79	10.74	8.52	15.09	45.19
H217	Hybrid	50.59	10.65	8.52	13.84	45.88
H218	Hybrid	51.19	10.44	8.68	14.13	45.56
H219	Hybrid	53.07	10.80	8.84	12.99	45.10

H220	Hybrid	52.46	10.36	9.10	13.02	45.43
H221	Hybrid	49.95	11.32	8.84	13.87	44.70
H222	Hybrid	50.90	10.67	8.82	13.68	45.27
H223	Hybrid	50.72	10.97	8.89	13.32	45.20
H224	Hybrid	52.33	10.39	8.82	13.52	46.03
H225	Hybrid	53.70	10.57	9.07	12.71	46.13
H226	Hybrid	52.83	10.74	8.71	13.18	44.70
H227	Hybrid	52.40	11.29	8.55	13.25	45.39
H229	Hybrid	48.60	11.37	8.47	14.45	44.93
H230	Hybrid	55.02	10.83	8.80	12.31	45.11
H231	Hybrid	51.01	10.56	8.00	14.75	45.34
H232	Hybrid	48.96	11.45	8.14	15.03	45.27
H233	Hybrid	49.93	10.82	7.81	15.12	45.91
H234	Hybrid	49.29	11.10	7.78	14.98	45.37
H235	Hybrid	51.66	10.94	7.93	14.19	46.70
H236	Hybrid	51.48	10.82	8.78	13.37	45.42
H237	Hybrid	48.97	11.70	8.87	13.89	44.64
H238	Hybrid	47.41	11.21	8.40	15.47	45.52
H239	Hybrid	49.97	11.38	8.54	13.98	45.32
H240	Hybrid	50.86	10.99	8.81	13.16	46.22
H241	Hybrid	51.52	11.32	8.78	12.98	44.30
H242	Hybrid	50.63	11.46	7.93	14.13	45.21
H244	Hybrid	54.88	9.39	7.25	14.74	45.52
H245	Hybrid	53.24	9.46	7.48	14.87	46.01
H246	Hybrid	53.26	10.82	8.20	13.40	46.04
H247	Hybrid	51.03	11.61	7.78	14.00	46.13
H248	Hybrid	51.12	11.35	7.82	14.33	46.66
H249	Hybrid	51.49	11.70	7.89	13.38	46.06
H250	Hybrid	51.71	11.19	7.64	14.27	45.80
H251	Hybrid	53.73	10.90	7.62	13.66	47.36
H253	Hybrid	52.56	10.77	7.89	14.01	46.87
H255	Hybrid	51.48	10.70	8.27	14.07	45.13
H256	Hybrid	53.57	11.11	8.76	12.59	44.36
H257	Hybrid	52.42	11.18	8.65	12.68	43.88
H258	Hybrid	52.82	10.75	8.62	13.23	44.83
H259	Hybrid	53.93	10.83	8.78	12.57	44.36
H260	Hybrid	50.95	11.14	8.28	14.12	45.03
H261	Hybrid	52.03	11.42	8.14	13.22	44.93
H262	Hybrid	51.85	10.28	8.53	13.91	45.79
H263	Hybrid	53.09	10.60	8.32	13.42	45.91
H264	Hybrid	52.30	10.90	8.52	13.46	44.57

H265	Hybrid	52.41	10.91	8.52	13.22	44.65
H266	Hybrid	52.57	10.26	8.90	13.39	45.09
H267	Hybrid	53.75	10.32	8.62	13.21	44.82
H268	Hybrid	48.70	10.95	8.27	15.32	43.83
H269	Hybrid	49.47	10.91	8.08	14.93	43.55
H270	Hybrid	50.82	10.37	8.38	14.51	44.76
H271	Hybrid	51.08	10.52	8.39	14.33	44.09
H272	Hybrid	52.80	11.03	8.59	12.82	45.28
H273	Hybrid	51.76	10.74	8.90	13.24	45.44
H274	Hybrid	50.26	11.02	8.51	13.89	44.60
H275	Hybrid	48.68	11.34	8.65	14.50	44.73
H276	Hybrid	46.44	11.73	8.59	15.43	45.20
H277	Hybrid	48.12	11.63	8.38	14.88	45.09
H278	Hybrid	50.78	10.83	8.79	13.66	45.01
H279	Hybrid	51.81	11.28	8.64	12.94	44.83
H280	Hybrid	51.47	11.00	8.46	13.37	45.12
H281	Hybrid	49.50	10.95	8.78	14.04	44.19
H282	Hybrid	49.21	11.06	8.50	14.71	44.14
H283	Hybrid	50.82	10.92	8.60	13.56	45.51
H284	Hybrid	49.71	11.29	8.46	14.03	44.49
H285	Hybrid	49.59	10.74	8.90	14.08	44.83
H286	Hybrid	50.66	10.96	8.58	13.65	45.78
H287	Hybrid	52.00	10.79	8.37	13.63	45.40
H288	Hybrid	52.16	10.64	8.68	13.32	45.29
H289	Hybrid	49.65	10.69	8.54	14.47	44.81
H290	Hybrid	49.68	11.37	8.47	14.28	44.68
H291	Hybrid	51.71	11.02	8.44	13.44	45.92
H292	Hybrid	50.95	11.26	8.43	13.44	45.61
H293	Hybrid	51.19	10.85	8.41	13.88	45.04
H294	Hybrid	51.95	10.83	8.51	13.59	44.83
H296	Hybrid	52.01	10.68	8.87	13.25	45.67
H297	Hybrid	50.53	10.51	8.64	14.26	45.51
H299	Hybrid	51.19	10.68	8.25	14.07	45.59
H300	Hybrid	50.92	10.96	8.14	13.99	45.51
H301	Hybrid	51.21	10.93	8.55	13.81	45.45
H302	Hybrid	52.16	10.65	8.67	13.55	45.85
H303	Hybrid	51.99	11.10	8.89	12.89	45.11
H304	Hybrid	51.47	10.86	8.97	13.39	45.11
H305	Hybrid	50.37	11.12	8.76	13.74	44.89
H306	Hybrid	49.79	11.57	8.55	14.31	44.43
H307	Hybrid	42.26	12.01	9.08	17.64	44.79

H308	Hybrid	46.88	11.95	8.70	14.72	44.89
H309	Hybrid	51.65	10.95	8.76	13.31	46.31
H310	Hybrid	51.74	11.19	8.73	13.15	45.06
H311	Hybrid	52.62	10.51	8.24	13.52	45.66
H312	Hybrid	51.60	10.29	8.26	14.23	44.63
H313	Hybrid	48.15	10.59	8.29	15.65	44.63
H314	Hybrid	49.09	10.77	8.35	15.06	44.47
H315	Hybrid	46.06	11.18	8.35	16.14	44.78
H316	Hybrid	49.89	10.68	8.14	14.87	44.98
H317	Hybrid	51.72	10.13	8.50	14.10	45.79
H318	Hybrid	51.87	10.40	8.18	14.19	45.16
H319	Hybrid	51.88	10.64	8.42	14.04	44.84
H320	Hybrid	51.27	10.49	8.60	14.20	45.09
H321	Hybrid	49.44	10.66	8.54	14.87	45.09
H322	Hybrid	49.58	10.75	8.33	15.28	44.81
H323	Hybrid	45.59	11.64	8.17	16.13	44.56
H325	Hybrid	50.94	10.59	8.63	14.12	45.58
H326	Hybrid	51.87	10.83	8.61	13.79	45.16
H327	Hybrid	52.76	10.60	8.48	13.44	44.34
H329	Hybrid	50.60	10.47	9.35	13.63	45.46
H331	Hybrid	55.76	8.95	8.29	13.98	45.53
H332	Hybrid	54.90	8.63	8.19	14.42	45.47
H333	Hybrid	50.21	10.27	8.87	14.29	45.15
H334	Hybrid	51.43	10.91	8.48	13.75	45.48
H335	Hybrid	52.13	10.10	8.91	13.68	46.07
H337	Hybrid	52.76	10.90	8.82	12.83	45.21
H339	Hybrid	51.17	10.51	9.43	13.19	45.65
H340	Hybrid	56.10	8.84	8.23	13.96	44.78
H341	Hybrid	55.07	9.00	8.16	13.99	45.22
H342	Hybrid	50.44	10.81	8.57	14.10	45.32
H343	Hybrid	51.57	10.98	8.66	13.32	45.35
H344	Hybrid	51.75	10.35	9.02	13.48	46.22
H345	Hybrid	52.89	10.44	8.85	13.10	45.23
H346	Hybrid	52.32	10.99	8.61	13.47	44.34
H347	Hybrid	50.92	10.73	9.21	13.64	44.50
H348	Hybrid	52.36	10.57	10.06	12.25	45.00
H350	Hybrid	56.62	8.76	8.10	14.09	44.96
H351	Hybrid	55.44	9.28	8.37	13.51	44.09
H352	Hybrid	51.40	10.32	8.81	13.89	45.58
H353	Hybrid	52.74	11.20	8.71	12.91	44.38
H356	Hybrid	54.57	10.46	9.43	12.04	44.41

H358	Hybrid	52.21	10.42	10.19	12.35	44.76
H360	Hybrid	56.89	8.79	8.21	13.67	45.43
H361	Hybrid	55.60	8.89	8.53	13.70	44.25
H362	Hybrid	51.92	10.15	9.30	13.45	45.33
H363	Hybrid	52.59	11.01	8.93	12.96	44.96
H364	Hybrid	52.60	10.10	9.47	13.18	45.52
H365	Hybrid	53.05	10.29	9.05	13.32	45.06
H366	Hybrid	54.38	10.65	9.57	11.92	44.45
H368	Hybrid	52.18	10.52	10.01	12.47	44.94
H369	Hybrid	57.10	8.77	8.07	13.77	45.33
H370	Hybrid	55.55	9.10	8.56	13.49	43.91
H371	Hybrid	51.21	10.40	9.22	13.69	44.81
H372	Hybrid	51.98	10.91	9.09	13.05	44.77
H373	Hybrid	52.21	10.26	9.45	13.30	44.99
H374	Hybrid	53.09	10.37	8.99	13.27	45.11
H375	Hybrid	53.75	10.75	9.49	12.16	44.05
H377	Hybrid	51.01	10.47	9.85	13.19	44.45
H378	Hybrid	56.46	8.72	8.45	13.77	45.03
H379	Hybrid	55.92	8.95	8.42	13.64	45.10
H381	Hybrid	53.07	11.07	8.95	12.54	45.36
H383	Hybrid	52.63	10.48	8.81	13.38	45.55

[1] Erucic acid content (% of total fatty acid)

[2] Linoleic acid content (% of total fatty acid)

[3] Linolenic acid content (% of total fatty acid)

[4] Oleic acid content (% of total fatty acid)

[5] Overall oil content (% of dry mass)

Table S3.10 Gene ID conversions for fatty acid biosynthesis candidate genes identified in previous *Brassica napus* L. literature which coincide with peak quantitative trait loci (QTL) regions identified in this study. Conversion based on the method by Rousseau-Gueutin et al. (2020).

Darmor v5 gene ID	*Darmor* v10 gene ID
BnaA01g12350D	A01p13900.1_BnaDAR
BnaA02g00830D	A02p00830.1_BnaDAR
BnaA02g07780D	A02p07860.1_BnaDAR
BnaA02g12050D	A02p13160.1_BnaDAR
BnaA02g13310D	A02p14430.1_BnaDAR
BnaA03g32450D	A03p37810.1_BnaDAR
BnaA04g04400D	A04p05770.1_BnaDAR
BnaA04g04410D	A04p05770.1_BnaDAR
BnaA04g04450D	A04p05800.1_BnaDAR
BnaA07g01090D	A07p01040.1_BnaDAR
BnaA07g35680D	A07p45290.1_BnaDAR
BnaA09g03610D	A09p05160.1_BnaDAR
BnaA09g04980D	A09p06590.1_BnaDAR
BnaA09g05270D	A09p06910.1_BnaDAR
BnaA09g05410D	A09p07040.1_BnaDAR
BnaA09g06090D	A09p07690.1_BnaDAR
BnaAnng14390D	A09p49530.1_BnaDAR
BnaA09g06170D	A09p07790.1_BnaDAR
BnaA09g07080D	A09p08920.1_BnaDAR
BnaA09g07080D	A09p08920.1_BnaDAR
BnaA10g08110D	A10p09050.1_BnaDAR
BnaA10g21150D	A10p26400.1_BnaDAR
BnaA10g23790D	A10p29110.1_BnaDAR
BnaC01g25960D	C01p31110.1_BnaDAR
BnaC01g26460D	C01p30220.1_BnaDAR
BnaC01g26600D	C01p29930.1_BnaDAR
BnaC05g10520D	C05p12040.1_BnaDAR
BnaC06g07070D	C06p10040.1_BnaDAR
BnaC06g07070D	C06p10040.1_BnaDAR

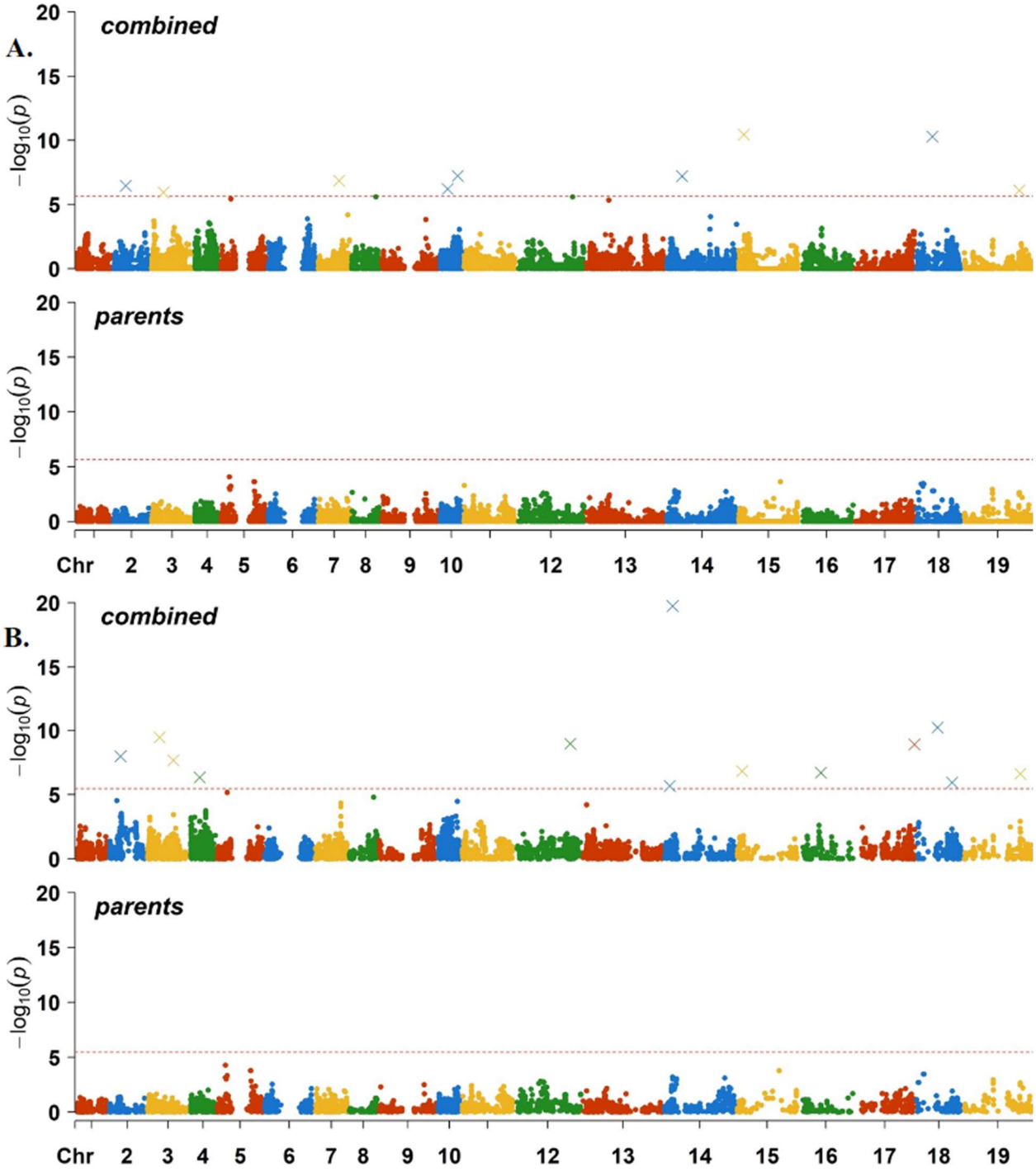


Figure S3.1 Manhattan plots produced by FarmCPU GWAS analyses for erucic acid content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

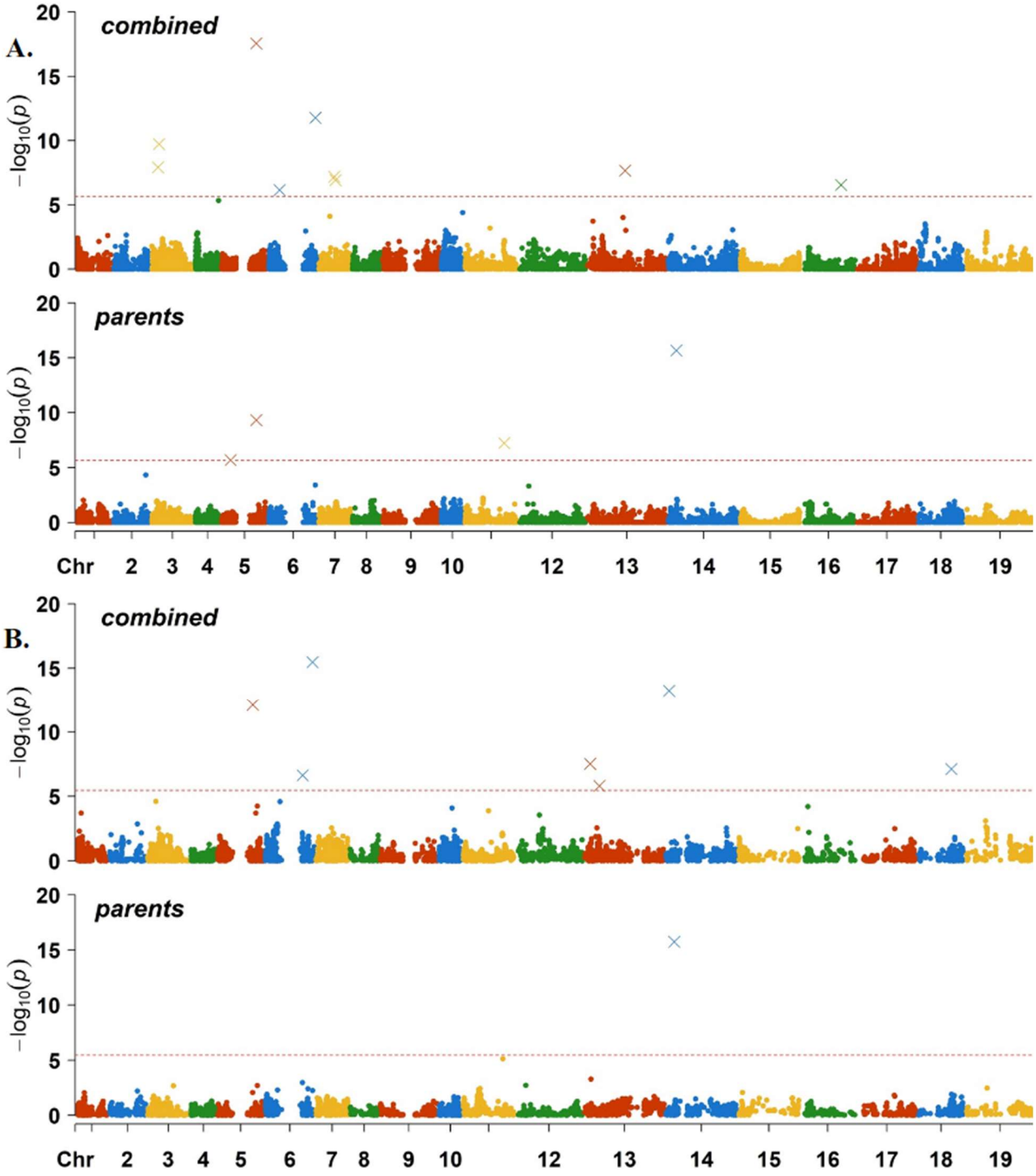


Figure S3.2 Manhattan plots produced by FarmCPU GWAS analyses for linoleic acid content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

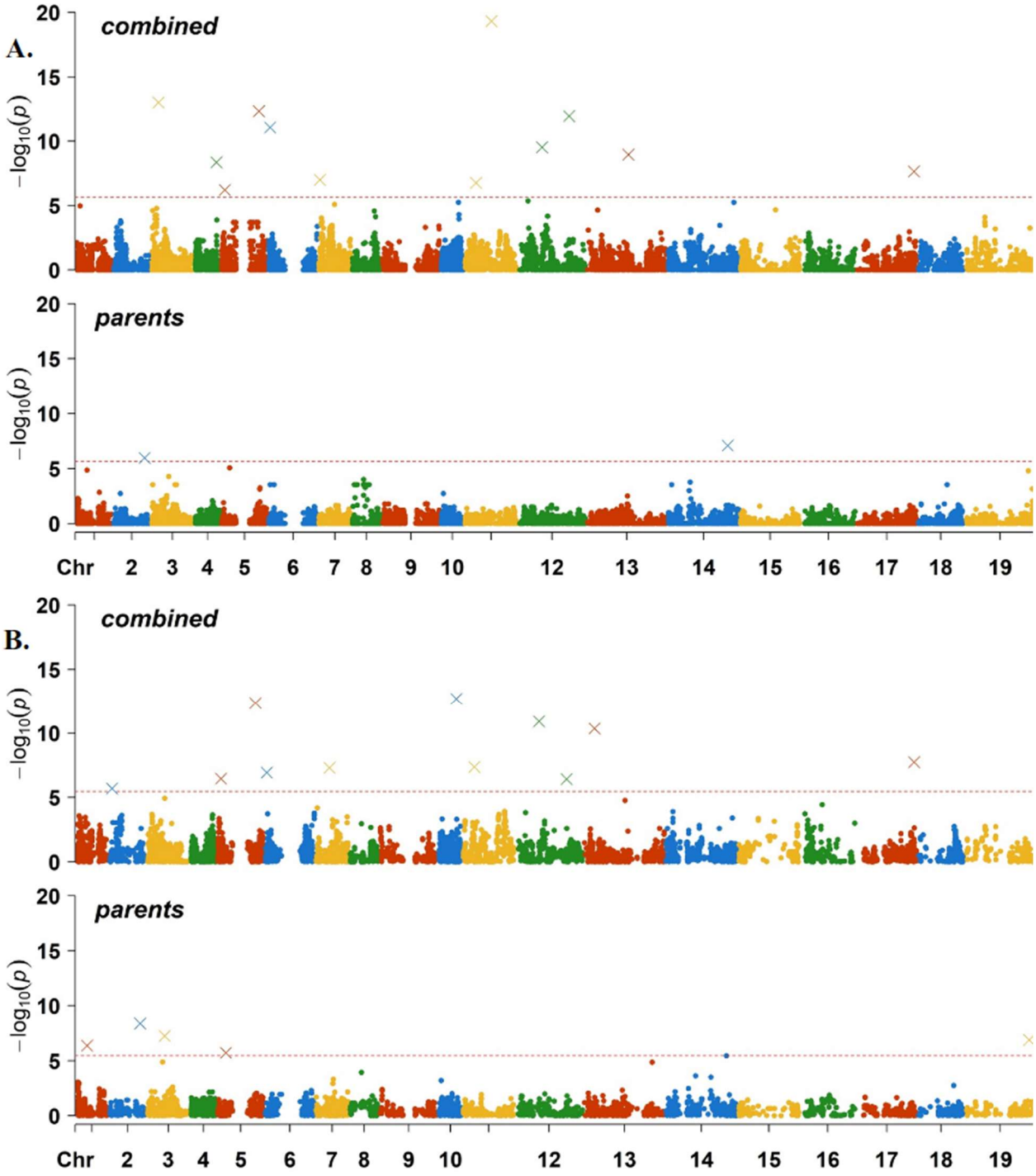


Figure S3.3 Manhattan plots produced by FarmCPU GWAS analyses for linolenic acid content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

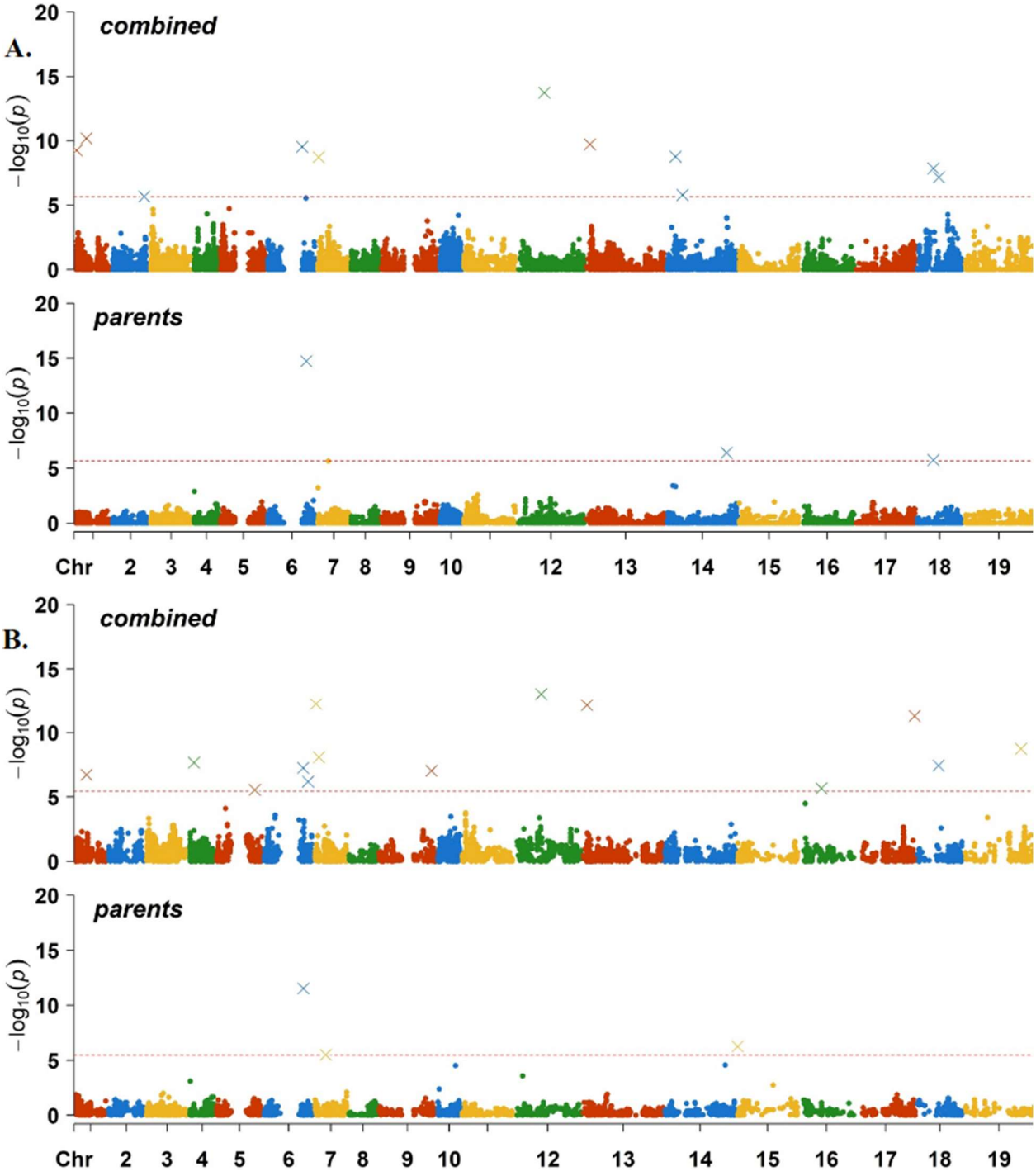


Figure S3.4 Manhattan plots produced by FarmCPU GWAS analyses for oleic acid content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

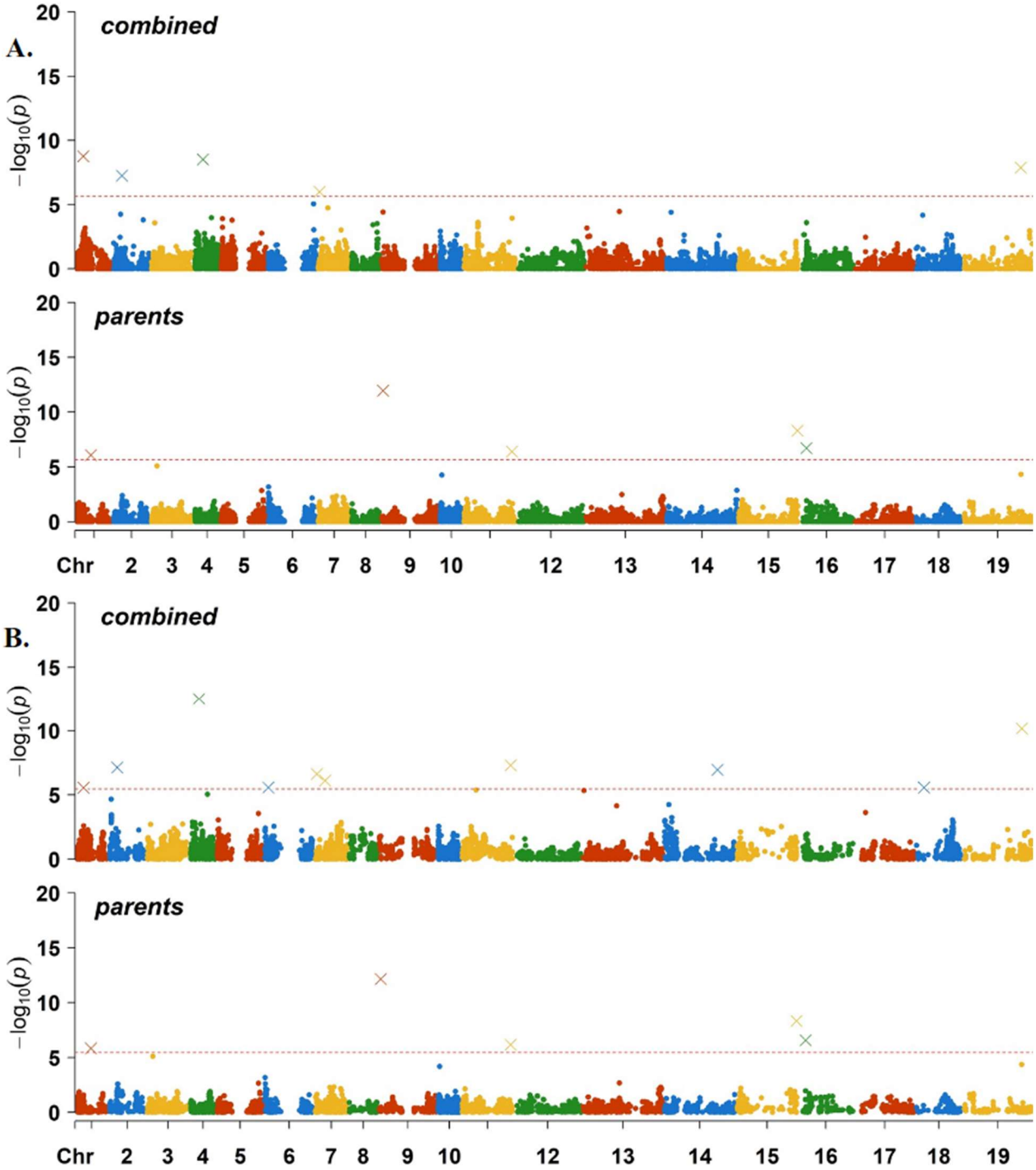


Figure S3.5 Manhattan plots produced by FarmCPU GWAS analyses for oil content in *Brassica napus* L. Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). Two SNP marker sets were input to GWAS: (A) MS1 (22,941 markers); (B) MS2 (14,699 markers). Each "×" above the dotted line denotes a significant SNP.

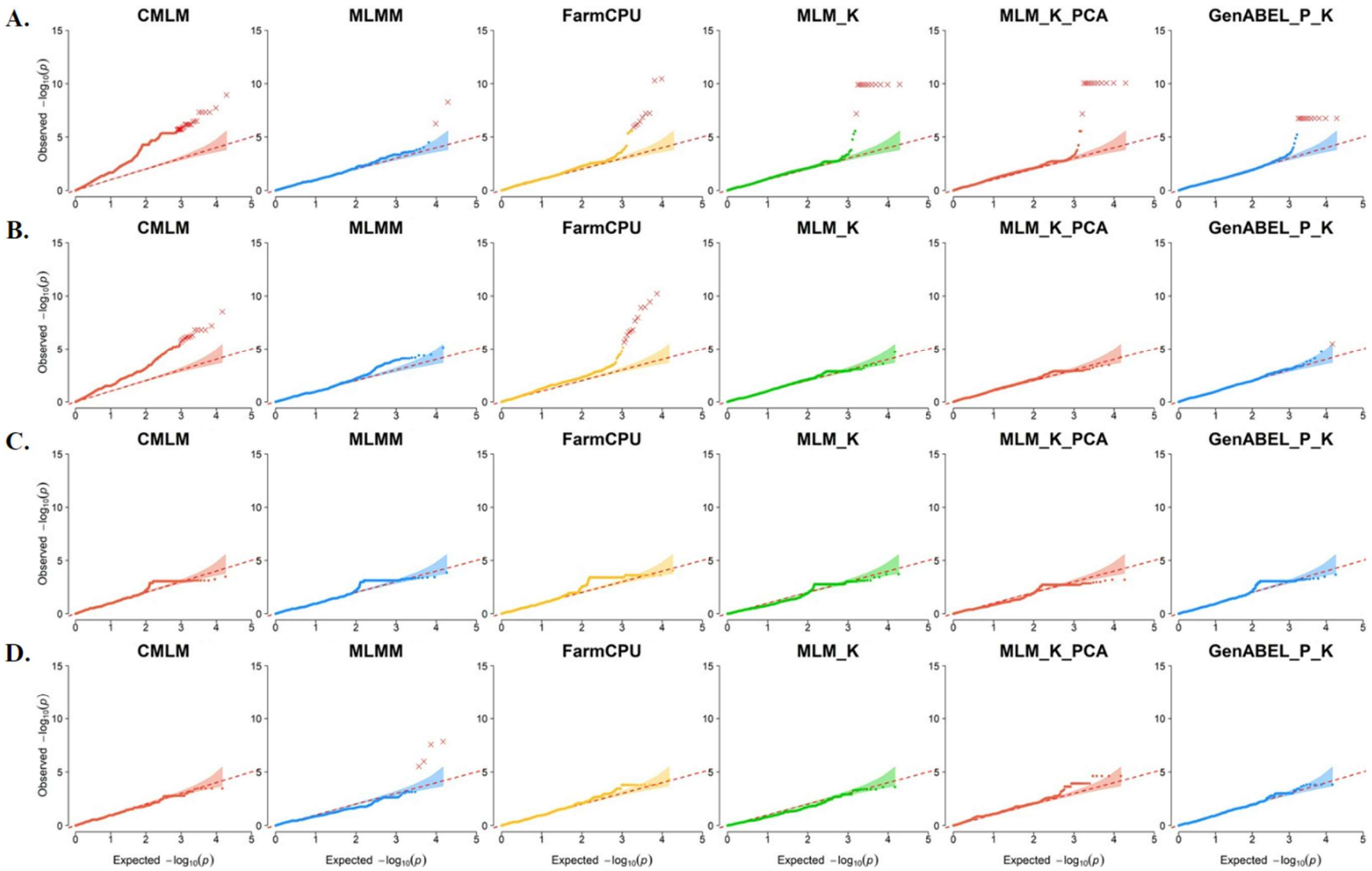


Figure S3.6 Q-Q plots of GWAS results for erucic acid content. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.

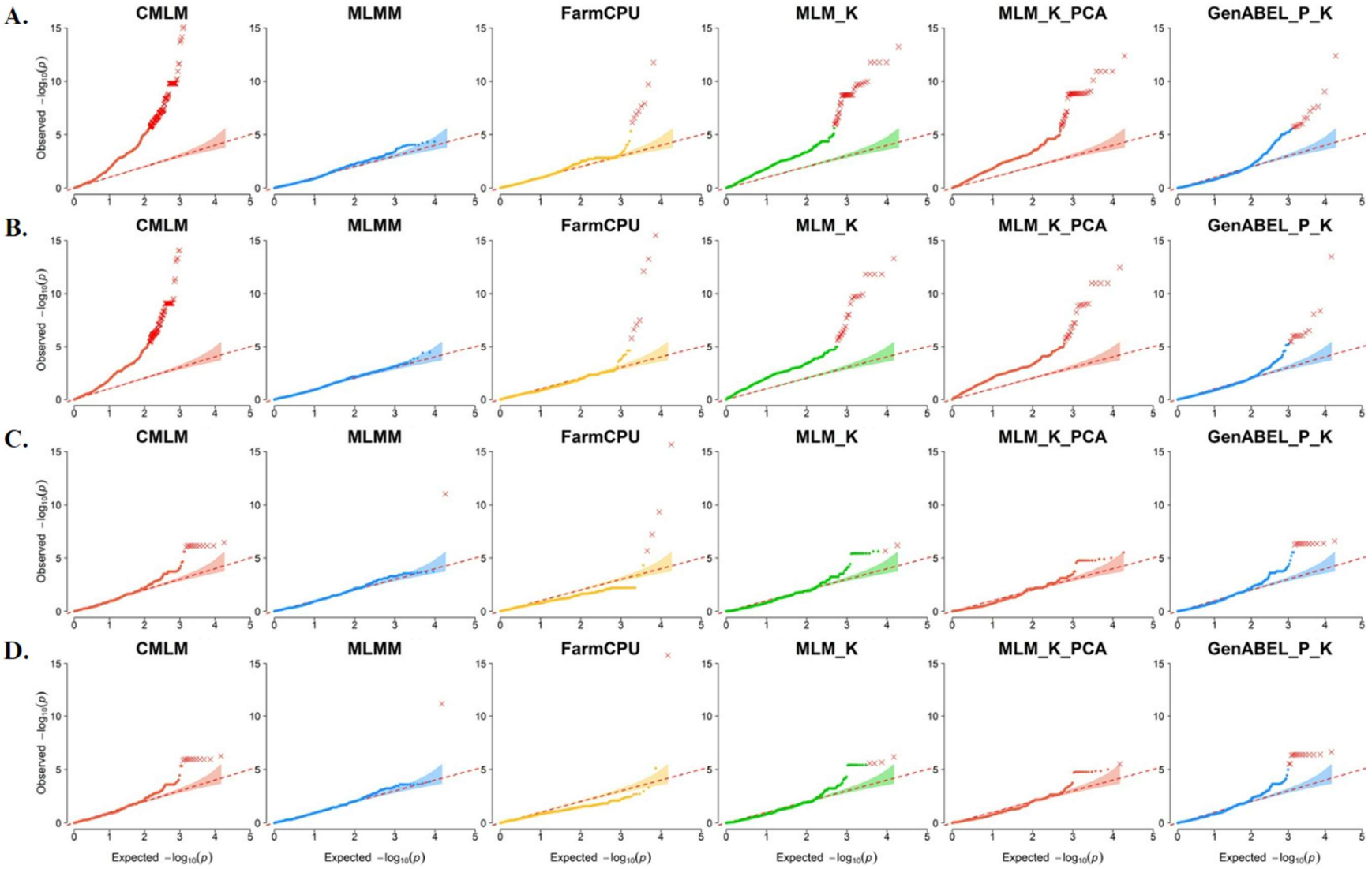


Figure S3.7 Q-Q plots of GWAS results for linoleic acid content. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.

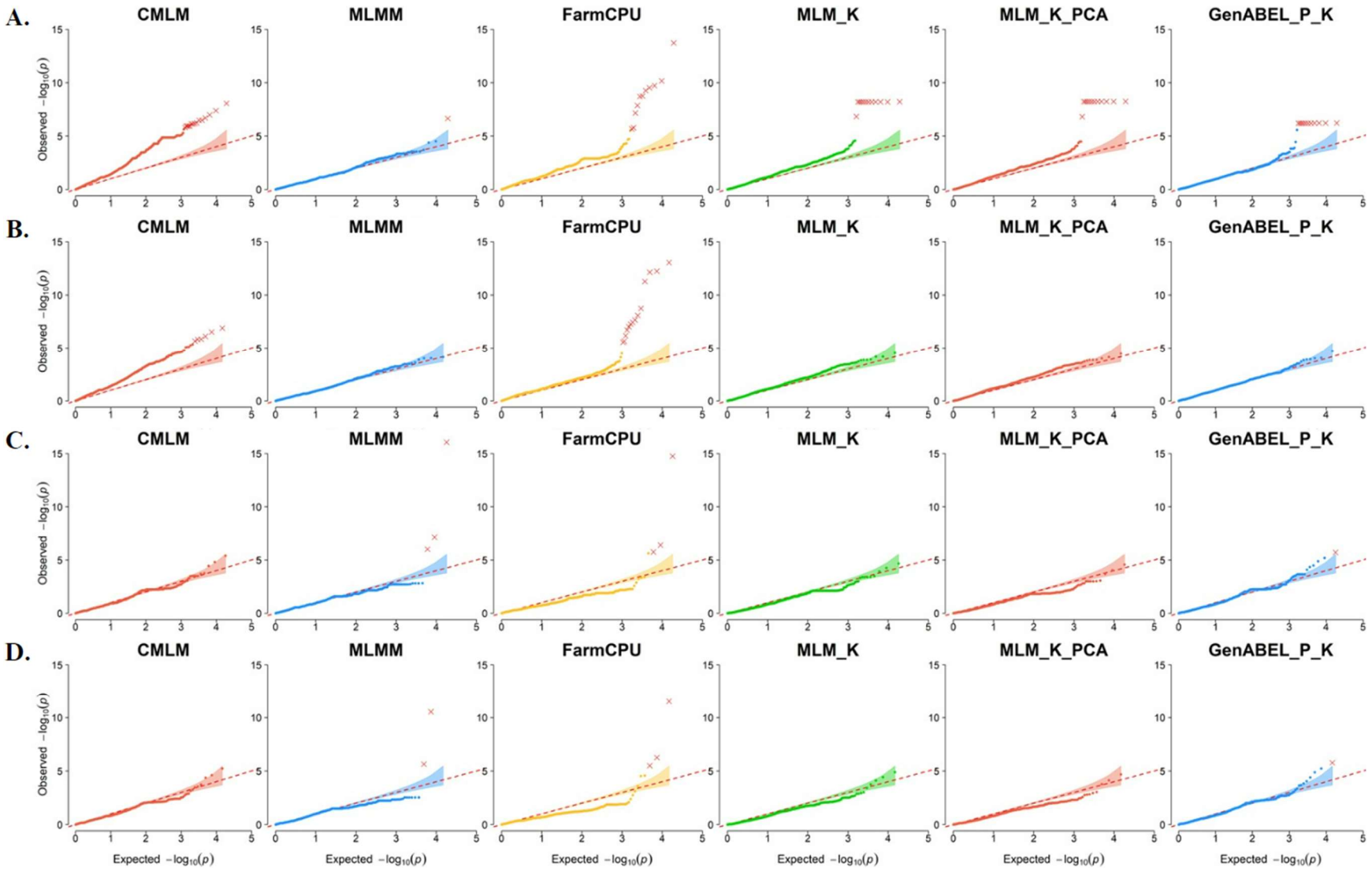


Figure S3.8 Q-Q plots of GWAS results for linolenic acid content. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.

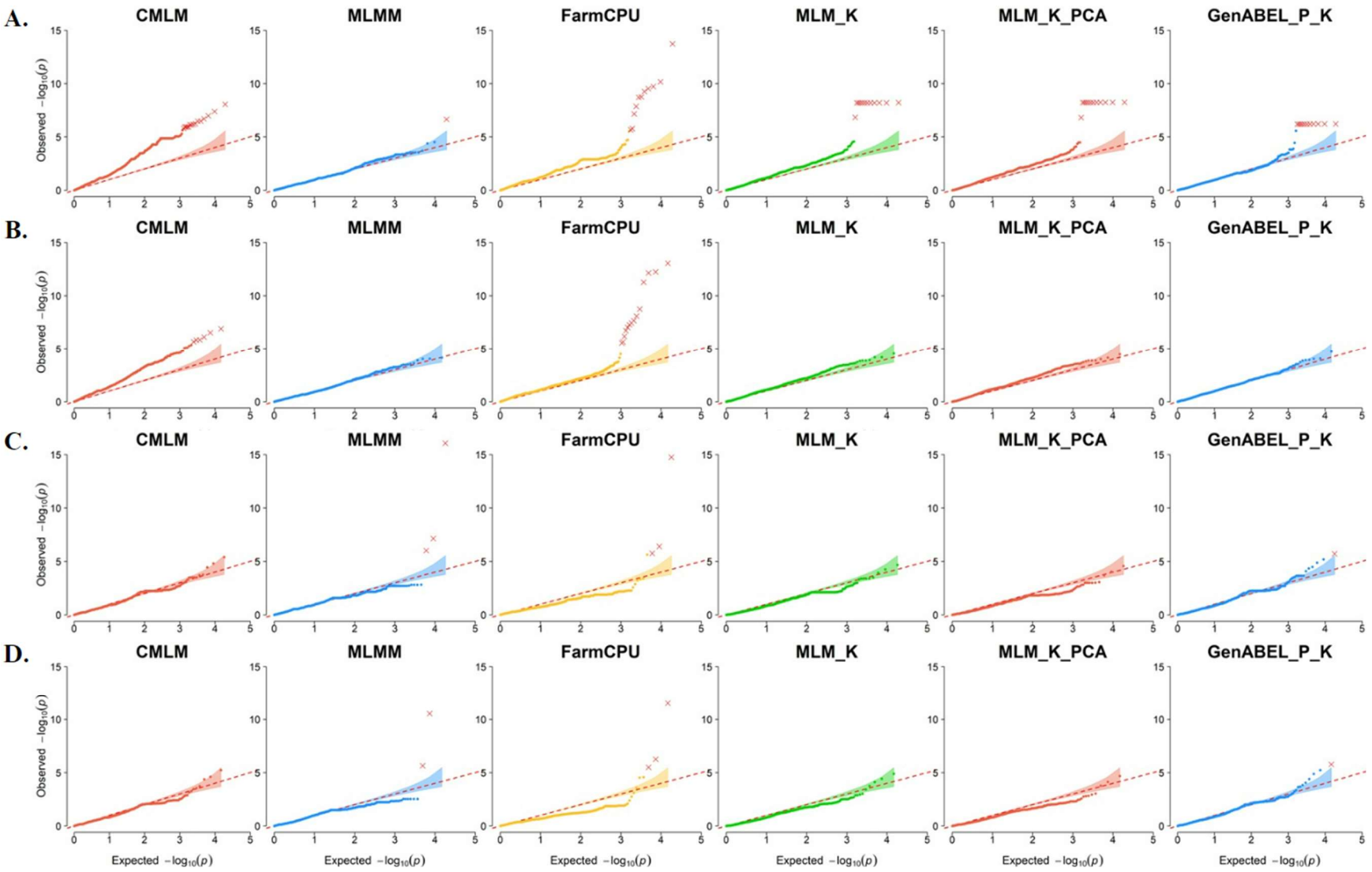


Figure S3.9 Q-Q plots of GWAS results for oleic acid content. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.

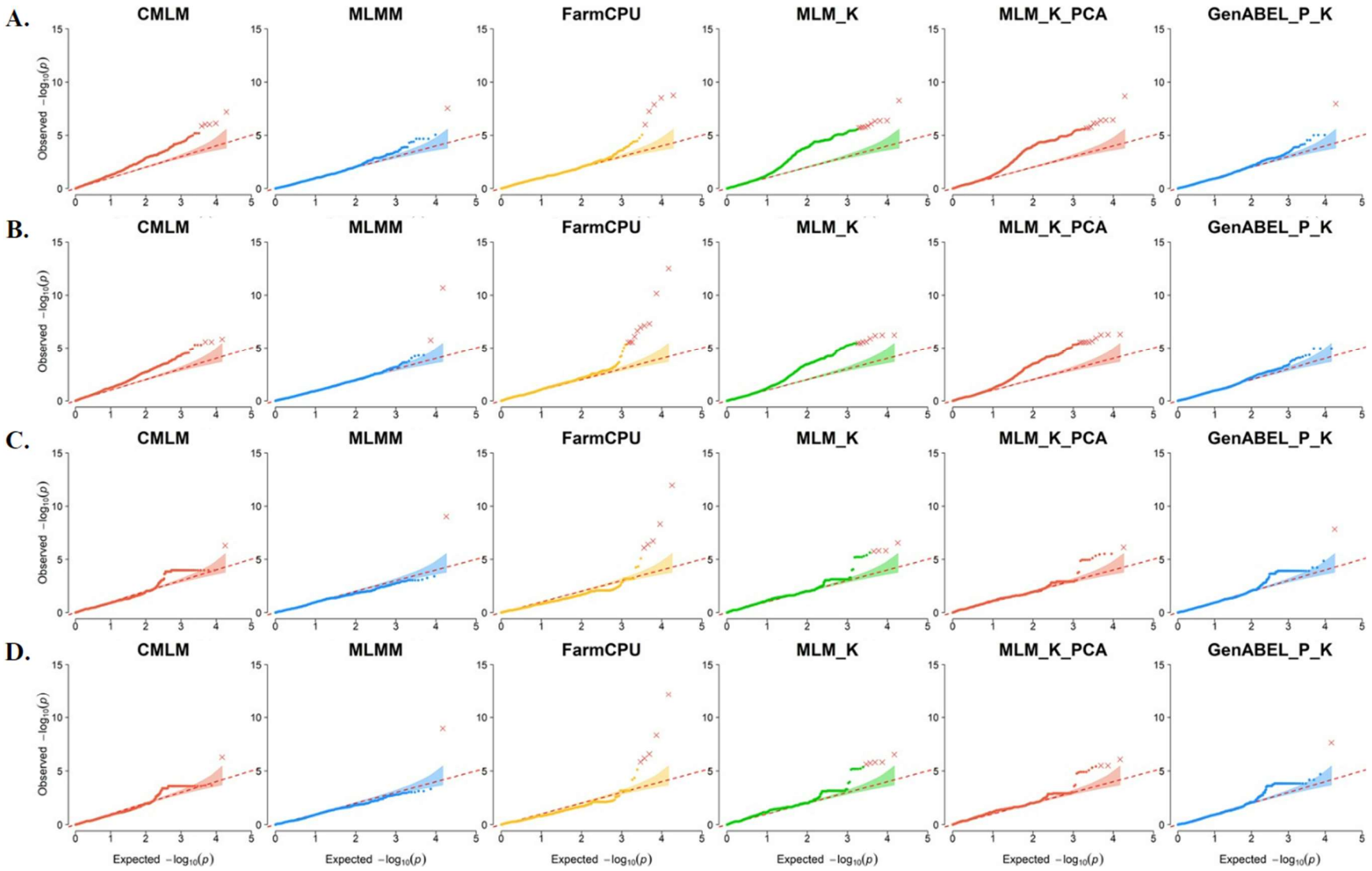


Figure S3.10 Q-Q plots of GWAS results for oil content. Red "×" denotes significant SNP based on Bonferroni correction. Two SNP marker sets were input to GWAS: MS1 (22,941 markers) and MS2 (14,699 markers). Two populations were input to GWAS: parents (31 B-lines, 60 R-lines) and combined (parents, 336 hybrids). (A) MS1+combined; (B) MS2+combined; (C) MS1+parents; (D) MS2+parents.