Breeding improvement of cruciferin content in the meal protein of spring *Brassica napus* L.

By

Kenny Kin Yang So

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Plant Science

University of Manitoba

Winnipeg, Manitoba

Copyright ©2021 by Kenny Kin Yang So

Acknowledgements

This dissertation would not have been possible without the support of many people.

To my advisor Dr. Robert Duncan, I am much obliged for the opportunity to pursue this Ph.D. I am indebted to my committee members Dr. Claudio Stasolla, Dr. Dilantha Fernando, and Dr. Alejandro Costamagna, for their time and expertise. I am also thankful to Dr. Mark Belmonte for his guidance and insightful discussions.

To Ralph Kowatsch, and Dr. Mohamed Elhiti, thank you for your company in the lab, friendship, and invaluable life lessons. I am thankful for the support of the Brassica breeding group, especially Debbie Witko and Kelsey Dickson, for their help with seed quality analysis. To the support staff in Plant Science, especially Martha Blouw and Jackie Edelson, thank you for the many laughs we shared and guiding me through the administrative side of grad school.

To my fellow grad students and colleagues, thank you for being there to share my successes and frustrations throughout the last six years. To my friends and family at home, thank you for your continued support and unwavering encouragement to complete this dissertation.

I gratefully acknowledge the Alexander Graham Bell Canada Graduate Scholarship provided by the Natural Sciences and Engineering Research Council of Canada and financial support from the University of Manitoba, Bunge, DL seeds, the Canola Council of Canada, and the Canola Cluster Funding.

Table of Contents

Acknowledgements .... **II**

List of figures .... **VIII**

List of Tables .... **XIII**

List of Abbreviations .... **XV**

Abstract .... **XVIII**

Foreword .... **XX**

Chapter 1. General introduction **20**

Chapter 2. Literature review **6**

2.1 Addressing food security for the growing global population 6

2.2 Transition from rapeseed to canola 11

2.3 Extracting oil from raw seed 15

2.4 Effects of processing on meal quality 16

2.5 Nutritional value of canola meal protein 17

2.6 Anti-nutrients in canola meal 21

2.7 Seed development in Canola 22

2.8 Seed storage proteins of the Brassicaceae family 23

2.8.1 Cruciferin of *B. napus* 26

2.8.2 Napin of *B. napus* 29

2.9 Genetic control of seed storage proteins in canola 33

Non-genetic control of seed storage proteins in canola 35

2.10 Canola protein as a novel food product 35

2.10.1 Allergenicity of canola seed storage proteins 36

2.10.2 Regulation of novel foods in the United States 38

2.10.3 Regulation of novel food in the European Union 39

2.10.4 Regulation of novel food in Canada 39
2.11 Canola meal as a protein source in animal husbandry 42
2.11.1 Cattle 43
2.11.2 Swine 43
2.11.3 Poultry 44
2.11.4 Aquaculture 45
Seed storage proteins in food processing 45
2.12 Separation and quantification of seed storage proteins 46
2.13 Immunodection and quantification of seed storage proteins 48
2.14 Manipulation of seed storage proteins in select crop species 48
2.14.1 Seed storage protein manipulation through conventional breeding 49
2.14.2 Soybean 50
2.14.3 Wheat 51
2.14.4 Canola 52
2.15 Plant breeding in the omics era 53
2.15.1 Genomics 53
2.15.2 Transcriptomics 61
2.15.3 Proteomics 62
2.16 Marking of a new era of crop improvement with genome editing 63
2.17 Regulatory consideration for genome editing 64
Chapter 3. Development and optimization of a Western blot protocol to phenotype cruciferin content in spring *Brassica napus* L. 67
3.1 Abstract 67
3.2 Introduction 67
3.3 Materials and methods 72
3.3.1 Plant material and defatted meal preparation 72
3.3.2 Total soluble protein extraction and preparation 73
3.3.3 Polyclonal anti-cruciferin antibody production 73
3.3.4 Antibody validation 74

3.3.5 Western blot optimization 75
3.4 Results 77
3.4.1 Optimizing extraction of total soluble protein 77
3.4.2 Antibody validation 77
3.4.3 Western blot optimization 80
3.5 Discussion 83
3.6 Conclusion 85
Chapter 4. Assessing genetic diversity of cruciferin content in spring *Brassica napus* L.. 87
4.1 Abstract 87
4.2 Introduction 87
4.3 Materials and Methods 90
4.3.1 Plant material 90
4.3.2 Plant growth parameters and experimental design 90
4.3.3 Non-destructive measurement of total oil, total protein, and glucosinolate content.... 96
4.3.4 Defatting and total soluble protein extraction 97
4.3.5 Phenotyping cruciferin content by Western Blotting and densitometry 98
4.3.6 Statistical analyses 99
4.4 Results 100
4.4.1 Non-destructive measurement of total oil, total protein, and glucosinolates 100
4.4.2 Phenotyping cruciferin content by Western Blotting and densitometry 100
4.4.3 Phenotypic correlation analyses 100
4.5 Discussion 107
4.5.1 Seed quality of the population 107
4.5.2 Effect of greenhouse environment on seed quality 108
4.5.3 Genetic variation in cruciferin content exists within spring *B. napus* germplasm 108
4.5.4 Cruciferin content was positively correlated with both protein content and glucosinolate content 109
4.6 Conclusion 110

Chapter 5. Identification of genomic regions associated with cruciferin content in spring-type *Brassica napus* L. 113

5.1 Abstract 113

5.2 Introduction 114

5.3 Material and Methods 117

5.3.1 Plant material and growth parameters 117
5.3.2 Genotyping and data quality control 117
5.3.3 Estimating linkage disequilibrium 118
Population structure 119
5.3.4 Phenotyping for protein content, oil content, and cruciferin content 120
5.3.5 Generating best linear unbiased predictors 120
Association mapping and model comparison 121
5.3.6 Candidate gene search 122

5.4 Results 123

5.4.1 Marker quality control and density 123
5.4.2 Linkage disequilibrium 123
5.4.3 Population structure 129
5.4.4 Phenotypic evaluation 129
5.4.5 Model selection for association mapping 133
5.4.6 Association mapping results for seed protein content 135
5.4.7 Association Mapping Results for Oil Content 141
5.4.8 Association Mapping Results for Relative Cruciferin Content 145

5.5 Discussion 149

5.6 Conclusion 154

Chapter 6. Transcriptomic differences during early seed development of *Brassica napus* L. cultivars divergent in cruciferin content 156

Chapter 6: 156

6.1 Abstract 156

6.2 Introduction 157

6.3 Materials and Methods 160

6.3.1 Plant material and growth conditions 160
6.3.2 Sample preparation and RNA sequencing 160
6.3.3 Data analysis 161
6.3.4 Validation with qRT-PCR 163

6.4 Results 164

6.4.1 RNA sequencing 164
6.4.2 Cruciferin transcripts 164
6.4.3 Differential gene expression 164
6.4.4 Gene ontology enrichment 169
6.4.5 qRT-PCR validation 172

6.5 Discussion 172

6.6 Conclusion 177

Chapter 7. General Discussion 178

Chapter 8. Future research 181

Chapter 9. Reference material 183

9.1 Literature cited 183

9.2 Appendix 230

List of figures

Figure 2.1. The world population is projected to reach nine billion by the mid-21st century. Data points up to 2020 are annual estimates. Black data points from 2021 to 2075 are estimates based on median fertility. Red data trends indicate high and low fertility variant estimates (Raw data source: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, Online Edition. Rev. 1, accessed November 2019). 7

Figure 2.2. Global rapeseed (*Brassica napus* L.) harvested acres have increased annually since the introduction of the first low-erucic acid, low-glucosinolate canola variety 'Tower' in 1974 (Data source: FAOSTAT, Food and Agriculture Organization of the United Nations, accessed February 2019). 10

Figure 2.3. Rapeseed (including canola) was the fourth largest oil crop in 2019 based on global production quantity. (Data source: FAOSTAT, Food and Agriculture Organization of the United Nations, accessed January 2021) 13

Figure 2.4. Canola production in Canada has increased steadily since the introduction of the first low-erucic acid, low-glucosinolate canola variety 'Tower' in 1974 (Source: Statistics Canada. Table 32-10-0359-01, accessed November 2019). 14

Figure 2.5. Coincidence of zygotic embryogenesis stages in spring *Brassica napus* L. with spatial-temporal patterns in the accumulation of seed storage protein transcripts and gene products during the first six weeks seed of development. Abbreviations are as follows: G, globular stage; H, heart stage; T, torpedo stage; C, cotyledon stage. A single asterisk (*) indicates initial detection and three asterisks (***) indicate peak levels... 24

Figure 3.1. Sequence alignment of all five *Brassica napus* L. cruciferin subunits. Curated amino acid sequences were obtained from the UniProt Knowledgebase. The peptide sequence (highlighted in the red box) is within the α-chain of the cruciferin subunit and was used as an antigen for antibody production.. 69

Figure 3.2. Proteins recovered from the sequential extraction of defatted *Brassica napus* L. 'Westar' meal resolved under reducing conditions. Soluble protein was repeatedly extracted from the meal a total of six times. Equal volumes were loaded in each lane. Numbered lanes refer to the ordinal extraction order. The ladder was resolved in the M lane.. 78

Figure 3.3. Coomassie-staining and Western blotting of soluble seed protein from *Brassica carinata* A. Braun '194251', *B. napus* L. 'Qing2', *B. juncea* (L.) Czern. 'Cutlass', and *B. rapa* L. 'K880.' One microgram of protein was resolved under reducing conditions in each lane on duplicate gels. Resolved proteins were visualized by staining with Coomassie Brilliant Blue (A). Western blotting was performed using a custom anti-cruciferin antibody and directly colorimetric detection (B).. 79

Figure 3.4. Western blot detection of cruciferin in soluble protein from defatted *Brassica napus* L. 'Westar' meal. Soluble protein was extracted; equal volumes of each serial dilution were resolved

in each lane. Colorimetric detection of cruciferin was performed using an anti-cruciferin antibody (1:4000 dilution) directly (A) or indirectly with a secondary antibody (1:8000 dilution) (B). Lanes are labelled by dilution factor. Lane M indicated ladder. 81

Figure 3.5. Optimization of protein loading for cruciferin quantification by Western blotting. Varying quantities of soluble protein from the defatted meal of *Brassica napus* L. 'Westar' were resolved under reducing conditions (A). Resolved proteins were blotted and cruciferin content was determined by colorimetric detection using a custom anti-cruciferin antibody (B). Signal intensity of each band was plotted against protein load and a natural logarithm regression was fitted (C).82

Figure 4.1. Frequency distribution of protein content (A), oil content (B), glucosinolate content (C), and cruciferin content (D) for spring *Brassica napus* L. cultivars phenotyped in 2019 under greenhouse conditions in Winnipeg, Canada. The y axis represents percent of the population. 101

Figure 4.2. Cruciferin content in spring *Brassica napus* L. grown in a randomized complete block design, with three replications and repeated in two experiments, under greenhouse conditions at the Crop Technology Centre, University of Manitoba, Winnipeg in 2019. Cruciferin content was assessed by Western blotting relative to a field-grown reference sample from the cultivar 'Westar.' Boxes represent the value of all biological replicates across the two experiments. The horizontal bar above identifies genotypes classified into low, medium, and high-cruciferin classes by the Tukey-Kramer method.. 104

Figure 4.3. Correlations between seed protein content, oil content, glucosinolate content, and cruciferin content were assessed in 44 spring *Brassica napus* L. genotypes grown in a randomized complete block design with three blocks, replicated twice, under greenhouse conditions at the Crop Technology Centre, University of Manitoba, Winnipeg in 2019. Histograms along the diagonal axis represent the distribution of the respective trait. Data points in the scatterplots are coloured by genotype. Values in the matrix indicate the Pearson correlation coefficient of the two traits and its associated p value directly below.. 106

Figure 5.1. Distribution and density of single nucleotide polymorphism (SNP) markers across the chromosomes of *Brassica napus* L. in a population of 43 spring genotypes. Colour scale represents number of SNP markers per megabase... 124

Figure 4.5. Linkage disequilibrium (LD) decay in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions at the University of Manitoba in 2019, calculated at the whole genome and subgenome levels, respectively... 127

Figure 5.3. Intrachromosomal linkage disequilibrium (LD) decay in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions at the University of Manitoba in 2019. The LD decay threshold of $R^2 = 0.2$ is indicated by the dotted red line. The panel shows individual decay plots for each chromosome (Chr) on the A subgenome (Chr A1 to A10) and the C subgenome (Chr C1 to C9).. 128

Figure 4.7. Subpopulation structure of 43 spring *Brassica napus* L. genotypes as determined by Bayesian clustering (A). Each genotype is represented by a vertical bar and subpopulations are designated k1 to k6. Information about the year of introduction (B), geographic origin (C), and oil quality (D) of each genotype has been superimposed on the horizontal axis. Abbreviations are

as follows: AU, Australia; CA, Canada; CN, China; DE, Germany; DK, Denmark; GB, England; SE, Sweden; 00, canola; HEAR, high erucic acid rapeseed. 130

Figure 4.8. Subpopulation structure of 43 spring *Brassica napus* L. genotypes as determined by principal component analysis using three principal components (A). Each datapoint in the biplot represents one genotype. Information about the year of introduction (B), geographic origin (C), and oil quality (D), respectively, of each genotype has been superimposed over each biplot. Abbreviations are as follows: AU, Australia; CA, Canada; CN, China; DE, Germany; DK, Denmark; GB, England; SE, Sweden; 00, canola; HEAR, high erucic acid rapeseed. 131

Figure 4.9. Quantile-quantile plots show the distribution of the observed p values plotted against the expected values for different GWAS models in a population of 43 *Brassica napus* L. genotypes grown in the greenhouse at the University of Manitoba in 2019. The mean squared difference of the deviation is presented in parentheses. Plots for seed protein content (A), oil content (B), and cruciferin content (C) are presented. Abbreviations are as follows: GLM, general linear model; MLM, mixed linear model; PCA, principal component analysis; Q, population structure matrix from Bayesian clustering; K, kinship matrix. 134

Figure 4.10. Marker-trait associations for seed protein content in 43 *Brassica napus* L. genotypes grown in a greenhouse at the University of Manitoba in 2019. Manhattan plots are presented for the whole genome (A) and chromosome C5 (B). Genome-wide Manhattan plots from various mixed linear models considering kinship (MLM.K) and subpopulation structure via either principal component analysis (MLM.K.PCA) or Bayesian clustering (MLM.K.Q) are superimposed (C) to show overlaps in the plots. Each dot on the plots represents one single nucleotide polymorphism marker. The solid and dashed red lines represent significance thresholds based on the Bonferonni corrected p = 0.05/30290 and the modified Bonferonni p = 1/30290. The dotted black line represents a suggestive threshold of p = 0.001... 136

Figure 4.11. Haplotype blocks containing single nucleotide polymorphism markers associated with seed protein content in a panel of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions. The suggestively associated marker Bn-scaff_18181_1-p1745427 (A), and markers Bn-scaff_18181_1-p962029, Bn-scaff_18181_1-p960226, Bn-scaff_18181_1-p954167 (B) were identified by a mixed linear model with Bayesian clustering for subpopulation structure. The marker Bn-A04-p4601980 (C) was identified by a mixed linear model with principal component analysis for subpopulation structure. 138

Figure 4.12. Marker-trait associations for oil content in 43 *Brassica napus* L. genotypes grown in a greenhouse at the University of Manitoba in 2019. Manhattan plots are presented for the whole genome (A) and chromosome C5 (B). Genome-wide Manhattan plots from various mixed linear models considering kinship (MLM.K) and subpopulation structure via either principal component analysis (MLM.K.PCA) or Bayesian clustering (MLM.K.Q) are superimposed (C) to show overlaps in the plots. Each dot on the plots represents one single nucleotide polymorphism marker. The solid and dashed red lines represent significance thresholds based on the Bonferonni corrected p = 0.05/30290 and the modified Bonferonni p = 1/30290. The dotted black line represents a suggestive threshold of p = 0.001. 142

Figure 4.13. Haplotype blocks containing single nucleotide polymorphism markers associated with oil content in a panel of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions. Markers identified by a mixed linear model were either not in linkage disequilibrium with neighbouring markers (chrC05p196753) (A) or clustered in a separate block (chrC05p4942983, chrC05p4943192, chrC05p4945701) (B). Additional markers identified across all mixed linear models were located within a separate haploblock (c). 144

Figure 4.14. Marker-trait associations for cruciferin content in 43 *Brassica napus* L. genotypes grown in a greenhouse at the University of Manitoba in 2019. Manhattan plots are presented for the whole genome (A) and chromosome A6 (B). Genome-wide Manhattan plots from various mixed linear models considering kinship (MLM.K) and subpopulation structure via either principal component analysis (MLM.K.PCA) or Bayesian clustering (MLM.K.Q) are superimposed (C) to show overlaps in the plots. Each dot on the plots represents one single nucleotide polymorphism marker. The solid and dashed red lines represent significance thresholds based on the Bonferonni corrected p = 0.05/30290 and the modified Bonferonni p = 1/30290. The dotted black line represents a suggestive threshold of p = 0.001. Markers with strongly inflated significance values (-log(p)>7) were removed before plotting.. 146

Figure 4.15. Haplotype blocks containing single nucleotide polymorphism markers associated with cruciferin content in a panel of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions. The suggestively associated markers Bn-A06-p5780982, Bn-A06-p5693735, and Bn-A06-p5830298 were identified by a mixed linear model with consideration for population structure through principal component analysis (A). Additional markers (Bn-A07-p9791002 and Bn-A07-p9831545) were identified by a mixed linear model with consideration for population structure through Bayesian clustering.. 148

Figure 6.1. Multidimensional scaling of sequenced RNA-seq libraries prepared from developing seeds of the high-cruciferin *Brassica napus* L. cultivar ‘Red River 1852’ (cultivar 1000) and the low-cruciferin cultivar ‘Wizzard’ (cultivar 2000), respectively, collected at 15, 20, and 25 days after pollination (DAP). Each data point is numerically coded as cultivar, replicate, and DAP time point. Colours denote unique genotype-time point combinations. 166

Figure 6.2. Normalized transcript abundance (counts per million) of all seven cruciferin biosynthesis genes across the developing seeds of high-cruciferin *Brassica napus* L. cultivar ‘Red River 1852’ and the low-cruciferin cultivar ‘Wizzard,’ respectively, at 15, 20, and 25 days after pollination, as measured by RNA-seq.. 167

Figure 6.3. Count summary of differentially expressed genes in the developing seeds of high-cruciferin *Brassica napus* L. cultivar ‘Red River 1852’ and the low-cruciferin cultivar ‘Wizzard,’ respectively, at 15, 20, and 25 days after pollination as measured by RNA-seq. Differential expressed was assessed relative to ‘Red River 1852.’ 168

Figure 6.4. Enrichment analysis of GO terms assigned to genes that were up-regulated in developing seeds at 15, 20, and 25 days after pollination (DAP) between two cultivars of *Brassica napus* L. with different cruciferin content. Terms in red relate to the high-cruciferin genotype ‘Red River 1852’ while terms in blue relate to the low-cruciferin genotype ‘Wizzard.’ 170

Figure 6.5. Enrichment analysis of GO terms assigned to genes that were down-regulated in developing seeds at 15, 20, and 25 days after pollination (DAP) between two cultivars of *Brassica napus* L. with different cruciferin content. Terms in red relate to the high-cruciferin genotype 'Red River 1852' while terms in blue relate to the low-cruciferin genotype 'Wizzard.' 171

Figure 6.6. Quantitative reverse transcription PCR (qRT-PCR) of select, differentially expressed genes identified by RNA sequencing of developing seeds between the high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard' at 25 days after pollination. RNA sequencing results are presented as bars while qRT-PCR results are presented as dots. Error bars represent standard error of three biological replicates for qRT-PCR................ 173

List of Tables

Table 2.1. Comparison of seed quality between Canadian soybeans and western Canadian canola presented as the average value of 2013-2017 crops, inclusive. Table compiled with soybean data (Barthet, 2018b) and canola data (Barthet, 2018a) published by the Canadian Grain Commission. Non applicable measurements indicated with n.a. 18

Table 2.2. Comparison of moisture content, crude protein, and amino acid profiles between soybean meal and western Canadian canola meal. Values presented for canola meal are the mean and standard deviations of samples collected from five different crushing plants across western Canada. The origin of the soybean meal was not disclosed (Le Thanh et al., 2019)........ 19

Table 2.3. Five manually curated entries for *Brassica napus* L. cruciferin are listed in the Uniprot database (Source: https://www.uniprot.org, accessed November 2019). Length is presented as amino acid count. 27

Table 2.4. Seven manually annotated entries for *Brassica napus* L. napin are listed in the Uniprot database (Source: https://www.uniprot.org, accessed November 2019). Length is presented as amino acid length. 31

Table 2.5. Six canola (*Brassica napus* L.) products are listed in the United States Food and Drug Administration's (FDA) Generally Recognized as Safe (GRAS) notice inventory since the inventory was established in 1998. In all applications, the FDA had no questions regarding the GRAS status of the product. Current as of December 2020. 40

Table 2.6. Genome-wide association studies have been successfully implemented in *Brassica napus* L. to identify single nucleotide polymorphism (SNP) markers associated with various disease resistance, morphological/phenological, physiological, and seed quality traits. Studies listed in this table were performed using populations of various sizes and genotypic data acquired using the Brassica 60K Illumina Infinium™ SNP genotyping array (Clarke et al., 2016)........ 57

Table 3.1. Size summary of five curated accessions of *Brassica napus* L. cruciferin subunits in UniProt KB. Lengths are given as number of amino acids (AA). The expected molecular mass of each peptide is presented in parentheses and estimated using a factor of 1 amino acid = 0.11 kDa. 70

Table 4.1. Registration details of the spring *Brassica napus* L. genotypes used to assess genetic diversity in cruciferin content. Registration numbers are from the Canadian Food Inspection Agency database. N.A. indicates information was not available. 91

Table 4.2. Summary statistics for protein content, oil content, glucosinolate content, and cruciferin content for a population of spring *Brassica napus* L. cultivars phenotyped in 2019 under greenhouse conditions in Winnipeg, Canada. Values presented are least-squared means. Cruciferin content was assessed relative to a reference protein sample from the cultivar 'Westar.' 102

Table 4.3. Analysis of variance tables for protein content (A), oil content (B), glucosinolate content (C), and cruciferin content (D) of spring *Brassica napus* L. phenotyped in 2019 under greenhouse conditions at the University of Manitoba, Winnipeg, Canada. Cruciferin content was assessed relative to a reference protein sample from the cultivar 'Westar.' 103

Table 5.1. Marker density and linkage disequilibrium (LD) decay for each *Brassica napus* L. chromosome. Linkage disequilibrium decay values for the individual sub-genomes and whole genome were calculated using all markers in the sub-genome or whole genome, respectively. 125

Table 5.2. Summary of best linear unbiased predictions for seed protein content, oil content, and cruciferin content in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019. Protein and oil content, respectively, are described as percentage of total, dry seed. Cruciferin content is presented as percentage relative to cruciferin content in a reference sample of *B. napus* 'Westar.' C.V. represent the coefficient of variation. 132

Table 5.3. Candidate genes for single nucleotide polymorphism markers associated with seed protein content, oil content, and cruciferin content, respectively, in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019. 140

Table 6.1. Alignment rate of RNA-seq library reads prepared from developing seed of the high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard' harvested at 15, 20, and 25 days after pollination (DAP). 165

List of Abbreviations

HEAR	High erucic acid rapeseed
PDCAAS	Protein digestibility corrected amino acid score
SSP	Seed storage protein
QTL	Quantitative trait loci
IgE	Immunoglobulin E
FDA	Food and Drug Administration
GRAS	Generally recognized as safe
EU	European Union
EFSA	European Food Safety Authority
CFIA	Canadian Food Inspection Agency
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
RNAi	RNA interference
SNP	Single nucleotide polymorphism
GWAS	Genome-wide association study
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas9	CRISPR-associated protein 9

USDA	United States Department of Agriculture
CJEU	Court of Justice of the European Union
PNT	Plant with novel trait
TSP	Total soluble protein
TBS	Tris-buffered saline
TBST	TBS with 0.05% Tween-20
HRP	Horseradish peroxidase
PVDF	Polyvinylidene fluoride
CTAB	Cetyltrimethylammonium bromide
LD	Linkage disequilibrium
BLUP	Best linear unbiased predictor
GLM	General linear model
MLM	Mixed linear model
QQ	Quantile-quantile
MSD	Mean squared difference
MTA	Marker-trait association
GO	Gene ontology

PCA	Principal component analysis
Chr	Chromosome
DAP	Days after pollination
DEG	Differentially-expressed gene
BBCH	Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie
WAP	Weeks after pollination

Abstract

The oilseed *Brassica napus* L. is typically grown for edible oil; once the oil has been removed, the resulting seed meal is high in protein and has potential to be used as a dietary protein for human consumption. Specifically, the seed storage protein cruciferin is a candidate for development into a protein ingredient for food processing given its functional properties. Therefore, increasing cruciferin content in *B. napus* can add value to the crop by enabling the meal to be marketed as a secondary product. Traditional breeding efforts to increase cruciferin content in *B. napus* are hindered by the lack of information on the genetic variation of the trait and its underlying molecular mechanisms. The research presented in this dissertation aims to address these knowledge gaps. In the first project of the dissertation, a phenotyping protocol for cruciferin quantification was developed. The Western blotting method employed a custom anti-cruciferin antibody and was optimized to quantify cruciferin in soluble seed protein in *B napus*. The second project in the dissertation aimed to explore the genetic diversity of cruciferin content across a collection of spring *B. napus* cultivars that spanned 38 years of variety development. Genetic variation was observed in the population and varied from 44% to 93% of a reference cultivar. With genetic variation present, the third project of the dissertation used an association mapping approach to identify single nucleotide polymorphism molecular markers that were associated with cruciferin content to facilitate marker-assisted selection for the trait. Molecular markers on chromosomes A06 and A07 were found to be associated with cruciferin content and were in linkage disequilibrium with genes that may serve a regulatory role on cruciferin accumulation. Finally, to elucidate potential molecular mechanisms that govern differential cruciferin accumulation, a transcriptomics approach was taken in the fourth project of the dissertation. Using developing seeds from a high- and low-cruciferin cultivar, RNA sequencing revealed differential expression

of regulatory genes, rather than cruciferin biosynthesis genes. Taken together, the knowledge gained from this dissertation will facilitate the breeding improvement of cruciferin content in spring *B. napus* and add value to the crop.

Foreword

This thesis has been written in manuscript style according to the format outlined by the Faculty of Graduate Studies of the University of Manitoba. This thesis consists of a general introduction followed by a literature review, four chapter, a general discussion, and recommendations for future research. Reference materials, consisting of a list of literature cited and data appendices, are presented thereafter. The literature review has been published in Plants titled "Breeding Canola (*Brassica napus* L.) for Protein in Feed and Food" (So and Duncan. (2021) 10(10): 2220; doi:10.3390/plants10102220). Chapter one and chapter two will be combined into a single manuscript to be titled "Genetic diversity of cruciferin content in spring *Brassica napus* L." and will be submitted to the journal Genetic Resources and Crop Evolution or another suitable journal. Chapter three, to be prepared as a manuscript titled "Identification of genetic loci associated with cruciferin content in spring *Brassica napu* L." will be submitted to the journal Euphytica or another suitable journal. Chapter four, to be prepared as a manuscript titled "Transcriptomic analysis of early seed development in spring *Brassica napus* L. genotypes with divergent cruciferin content" will be submitted to the Canadian Journal of Plant Science or another suitable journal.

Chapter 1. General introduction

The ability to supply sufficient food and nutrients to support a growing global population is often cited as the largest challenge that modern agriculture faces (Bahadur KC et al. 2018; Tamburino et al. 2020). Of particular concern is the ability to produce sufficient, high-quality protein (de Gavelle et al. 2017). Proteins from plant sources have been suggested as nutritionally-comparable alternatives to dietary protein from animal sources with the added benefit of being less

environmentally destructive to produce (Röös et al. 2017a; Shepon et al. 2018; Gerten et al. 2020). Specifically, protein from canola/rapeseed (*Brassica napus* L.) has been identified as a potential candidate for development into a novel dietary protein source given its nutritional profile and availability (Bos et al. 2007; Fleddermann et al. 2013; Mupondwa et al. 2018).

Since the introduction of the first low glucosinolate and low erucic acid canola cultivar half a century ago (Stefansson and Kondra 1975), canola has become the fourth largest oilseed crop globally with over 70 million metric tonnes produced in 2019 (Food and Agriculture Organization of the United Nations 2020). Domestic production of canola totaled in excess of 18 million metric tonnes in 2019 (Food and Agriculture Organization of the United Nations 2020), contributing approximately $29.9 billion to the Canadian economy (LMC International 2020).

After the oil has been extracted from canola seed, the oil-free meal byproduct containing approximately 40% protein is used as livestock feed (Shepon et al. 2016; Mottet et al. 2017; Kim et al. 2019b). In 2020 alone, approximately five million metric tonnes of canola meal was produced in Canada (Statistics Canada 2021) and represents a substantial quantity of potential dietary protein for human consumption.

One way to introduce the consumption of canola protein is through its incorporation as a protein ingredient into processed food products. The seed storage protein cruciferin is the largest component in the *B. napus* seed protein pool, accounting for approximately 60% (Crouch and Sussex 1981) and possesses functional properties that render it suitable as a functional ingredient in human food products (Malabat et al. 2001; Cheung et al. 2014a; Perera et al. 2016; Mupondwa et al. 2018). However, to realize its maximum processing potential, cruciferin must be used as an isolate. The need to isolate and purify cruciferin (Bérot et al. 2005) from the total seed protein adds

technical challenges to manufacturing and renders the ingredient less economical; however, these limitations can be mitigated by increasing the amount of cruciferin in the seed and proportionally decreasing other constituents of the protein pool.

Crop cultivars with seed protein compositions tailored for various end uses have been produced through traditional breeding (Zarkadas et al. 2007; Kim et al. 2013b) and high-cruciferin varieties of *B. napus* can similarly be developed. Breeding for improved cruciferin content in *B. napus* first requires an accurate phenotyping platform with sufficient throughput to screen segregating progeny. Current electrophoresis-based cruciferin phenotyping protocols (Schatzki et al. 2014) are constrained by the inability to discern between non-target proteins of similar molecular mass and the differential staining of protein isoforms due to variation in amino acid composition. Thus, the development of a phenotyping platform for cruciferin quantification with improved accuracy is required.

Further, continued improvements in breeding are only possible through the presence of genetic diversity (Fu 2015; Henderson and Salt 2017). An understanding of the genetic diversity that exists for cruciferin content in *B. napus*, therefore, is essential prior to initiating a breeding program with this goal in mind. While the genetic diversity of cruciferin content in winter *B. napus* has previously been explored (Malabat et al. 2003), similar research has yet to be untaken in the spring ecotype, thus limiting the ability to improve cruciferin content in the crop.

Finally, an understanding of the molecular mechanisms that control variation in cruciferin content is necessary to facilitate breeding efficiency. Molecular markers associated with cruciferin content can be used for marker-assisted selection to identify high-cruciferin progeny early in the growth cycle, abating the need to phenotype mature seed (Cobb et al. 2019). Although quantitative

trait loci controlling cruciferin content were previously reported in winter rapeseed (Schatzki et al. 2014), their application is limited to the mapping population and are unlikely to extend to spring *B. napus* (Cobb et al. 2019). To ensure universal applicability, molecular markers can be identified through association studies using diverse populations of unrelated genotypes. Alternatively, transcriptomic approaches can be undertaken to identify differentially expressed genes and molecular markers can be developed from polymorphisms within coding sequences (Huang and Han 2014; Scheben et al. 2017). Such transcriptomic studies can offer further insight into the transcriptional regulation that underlies differential cruciferin accumulation.

Recognizing the paucity of information on the genetic diversity of cruciferin content in spring *B. napus* and the molecular mechanisms that control it, the following hypotheses were proposed:

1. Hypothesis: Cruciferin content in the seed protein of *B. napus* can be reliably quantified using immunological assays.

   Objective: Develop a phenotyping assay to quantify cruciferin content in the soluble protein of *B. napus* seeds.
2. Hypothesis: Phenotypic variation in cruciferin content exists in the spring-type *B. napus* germplasm and is correlated with breeding history.

   Objective: Assess the genetic diversity of cruciferin content within spring *B. napus* germplasm using a collection of genotypes that span the breeding history of the crop.
3. Hypothesis: Phenotypic variation due to genetic variation enables discovery of single nucleotide polymorphism molecular markers associated with cruciferin content in *B. napus.*

   Objective: Identify single nucleotide polymorphism molecular markers that are associated with cruciferin content in *B. napus* using a genome-wide association study approach.

4. Hypothesis: Differential gene expression patterns during seed development contributes to variation in cruciferin accumulation in *B. napus.*

   Objective: Evaluate transcriptomic differences between high- and low-cruciferin genotypes of *B. napus* during seed development to identify potential transcriptional mechanisms that underlie differential cruciferin accumulation.

Chapter 2. Literature review

So, KKY and Duncan, RW. (2021) Breeding Canola (Brassica napus L.) for Protein in Feed and Food. 10(10): 2220; doi:10.3390/plants10102220.

2.1 Addressing food security for the growing global population

The global population is projected by the United Nations to increase beyond nine billion by the mid-21st century (Figure 2.1) (United Nations Department of Economic and Social Affairs Population Division 2019) and scientists have been tasked with ensuring food security to sustain this growth (Godfray et al. 2010; Fedoroff et al. 2010; Tamburino et al. 2020). Agriculture will be able to feed the global population provided improvements are made to the sustainability of current agricultural practices along with a concomitant shift in dietary preferences (Röös et al. 2017b; Springmann et al. 2018; Tamburino et al. 2020; Gerten et al. 2020). Complicating the task of food production are increasingly severe and unpredictable climatic patterns (Myers et al. 2017; Röös et al. 2017b; Nelson et al. 2018) as well as the continued reduction of arable land (Röös et al. 2017b, 2017a; Ramankutty et al. 2018; Fitton et al. 2019). The concept of food security was initially defined by the United Nations at the World Food Conference of 1974 as the availability of food at reasonable prices at all times (United Nations 1975). The definition of food security has since been expanded to encompass the nutritional and social aspects of food (Food and Agriculture Organization of the United Nations 2003; Food Security Information Network 2018). Dietary protein has been a focus of nutrition programs as it is often a limiting macronutrient in malnourished and food insecure populations (Hwalla et al. 2016; Semba 2016; Crichton et al. 2019).

Dietary protein is primarily acquired through either the consumption of meat (animal protein) or legumes. Historically, meat consumption was associated with economic wealth and

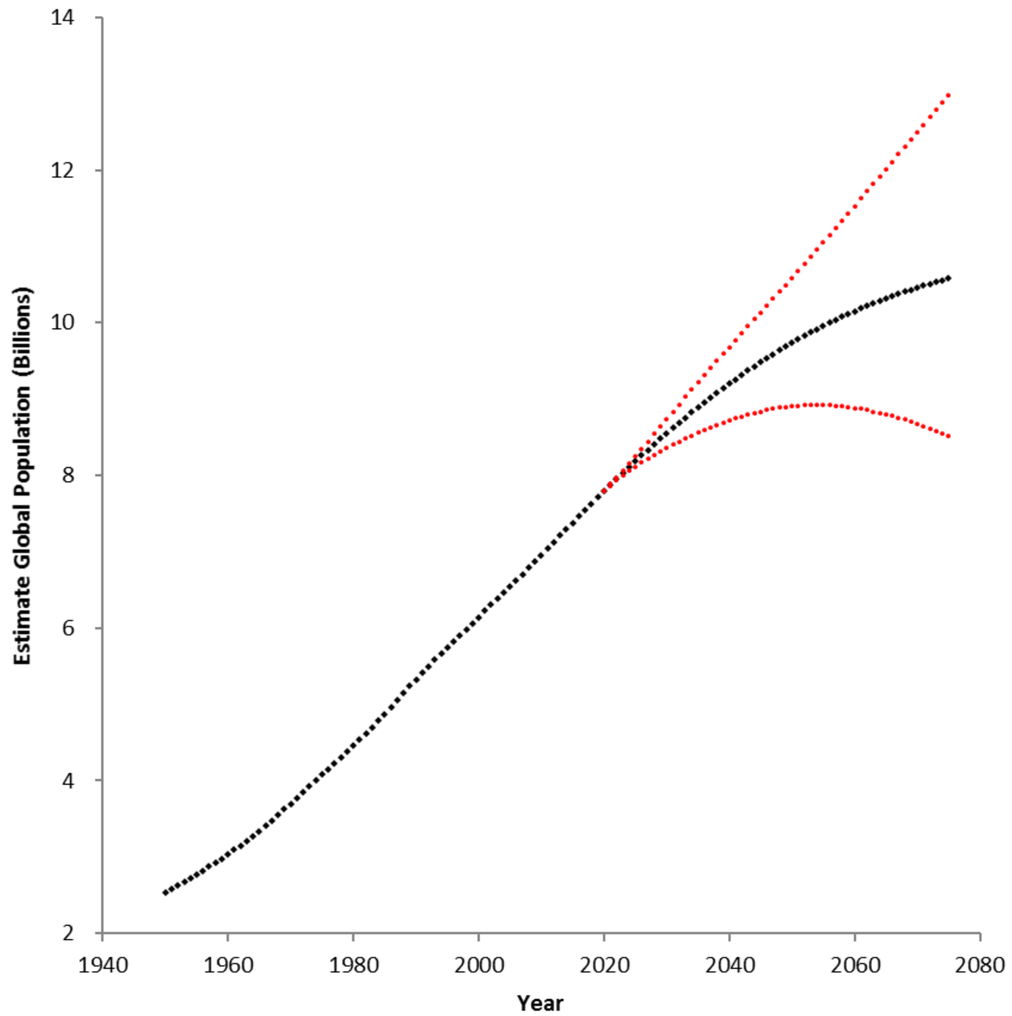

Figure 2.1. The world population is projected to reach nine billion by the mid-21st century. Data points up to 2020 are annual estimates. Black data points from 2021 to 2075 are estimates based on median fertility. Red data trends indicate high and low fertility variant estimates (Raw data source: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, Online Edition. Rev. 1, accessed November 2019).

recent shifts in developing economies have dramatically increased meat consumption (Gerbens-Leenes et al. 2010; Godfray et al. 2010; He et al. 2016). Intensive animal husbandry for meat production has been repeatedly cited as being environmentally destructive, unsustainable, and an inefficient method of converting protein feedstock into dietary protein (Pimentel and Pimentel 2003; Wu et al. 2014b; Sabaté and Soret 2014; Sabaté et al. 2015). The production of plant-based protein sources, such as kidney bean (*Phaseolus vulgaris* L.), requires substantially less input and generates less waste than an equivalent unit of red meat such as beef (Sabaté et al. 2015). Given the large amount of feed required to produce animal protein (Röös et al. 2017b), the redistribution of land from feed grain to crop production enables larger quantities of dietary protein to be produced per unit area (Shepon et al. 2018; Fitton et al. 2019). Evidently, even partial replacement of animal proteins with plant-based alternatives will have considerable impact on long term sustainability (Graham et al. 2019).

Although plant-based proteins are prominent in many cultures, the European and North American diets show a proclivity for animal proteins in their diet (Gerbens-Leenes et al. 2010; Röös et al. 2017b, 2017a; de Gavelle et al. 2017; Bentham et al. 2020). However, a recent assessment of food supplies across 171 countries found a decrease in animal protein supply in six countries across North America, Europe, and Australia (Bentham et al. 2020), suggesting a gradual replacement of animal protein with alternative proteins such as those from plants. Recent work has demonstrated that the composition of the human gut microbiome is in rapid flux with changes in diet (David et al. 2014), suggesting that humans are amenable to adapting to new diets with relative ease. The challenge of adopting plant-based proteins, therefore, may largely be cultural. In an effort to promote vegetable proteins as a nutritious, sustainable, and a secure alternative to animal protein, 2016 was designated as the International Year of Pulses (Food and

Agriculture Organization 2016). Consumer food choices are influenced by dietary guidelines published by national food authorities, and different recommendations show varying degrees of sustainability based on the resources required for production (Blackstone et al. 2018). Given their far-reaching influence, incorporating considerations for sustainability in government food recommendations may substantially affect food security for many people (Tuomisto 2018). In Canada, a recent revision to the food guide changing a "meat and alternatives" food group to "protein foods" with an emphasis on plant-based proteins reinforced the suitability of plant proteins as a replacement for animal proteins (Health Canada 2019). In expanding the available sources of plant-based proteins, rapeseed/canola (*Brassica napus* L.) protein has been suggested as a possible plant-based protein source. With increasing global acreages annually (Figure 2.2), canola that is currently grown and processed for edible oil production generates a large quantity of protein-rich seed meal as a byproduct of the oil extraction process (Wanasundara et al. 2016b). The ease of access to and availability of canola meal makes it a logically suitable and sustainable candidate for development into a plant-based protein for human consumption (Tan et al. 2011). Such a proposal has been supported by both industry and public institutions worldwide, as reflected by numerous abstracts at the recent International Rapeseed Congresses (International Rapeseed Congress 2019; Pilorgé et al. 2019).

In addition to serving as a dietary protein supplement, the functional properties of canola protein render them useful in a diverse range of food processing applications (Aider and Barbana 2011; Wanasundara et al. 2016b; Mupondwa et al. 2018). In such a case, individual canola proteins with the desired functionality would need to be isolated with additional processing steps from the crude seed protein to maximize their functional properties (Aider and Barbana 2011; Fetzer et al. 2018).

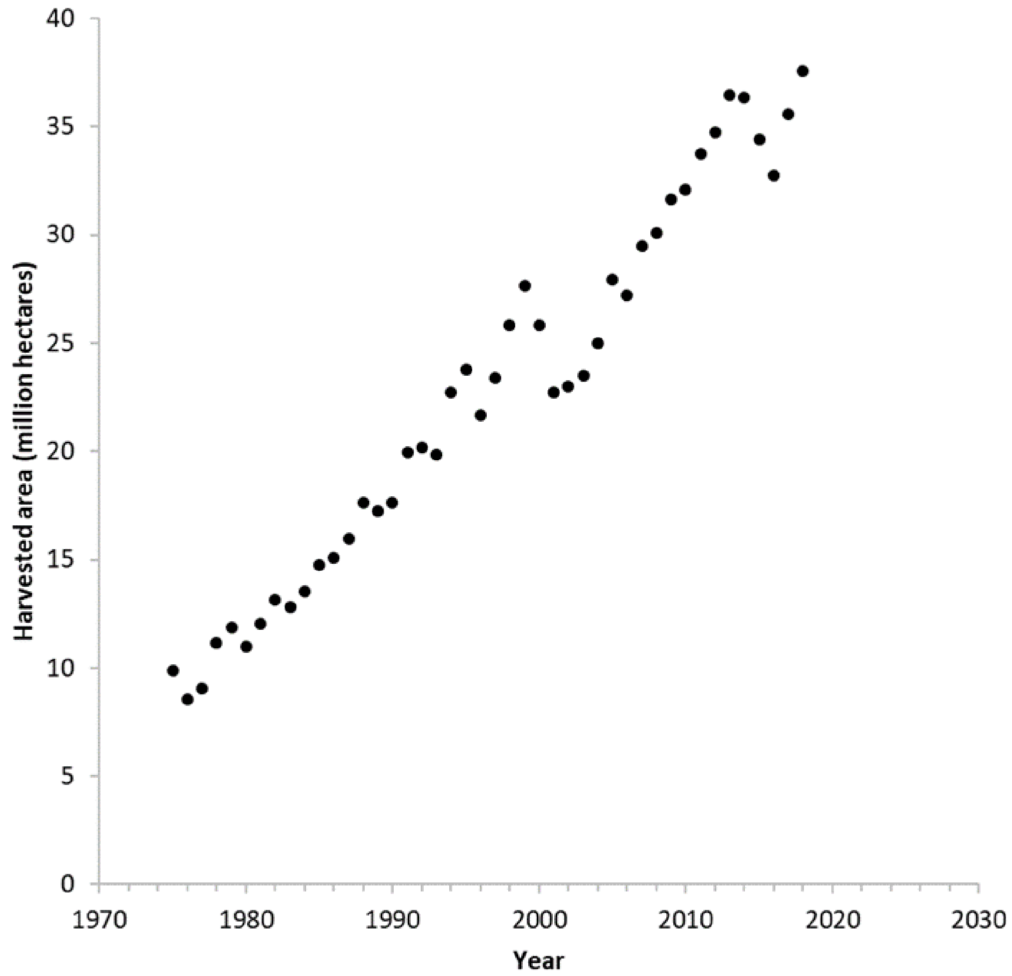

Figure 2.2. Global rapeseed (*Brassica napus* L.) harvested acres have increased annually since the introduction of the first low-erucic acid, low-glucosinolate canola variety 'Tower' in 1974 (Data source: FAOSTAT, Food and Agriculture Organization of the United Nations, accessed February 2019).

2.2 Transition from rapeseed to canola

Rapeseed is an economically important allotetraploid oilseed crop derived from *B. oleraceae* L. and *B. rapa* L. (Chalhoub et al. 2014). Rapeseed was introduced into Canada in 1943 initially as a source of vegetable oil for industrial applications (Craig 1971; Nieschlag and Wolff 1971). Despite the high oil content of rapeseed and its adaptability to be grown in Western Canada, the high percentage of erucic acid in rapeseed oil (approximately 40%) and high levels of glucosinolates (80 µmol/g seed) rendered the oil unfit for the human diet (Stefansson et al. 1961; Daun 1986). Erucic acid consumption had been implicated in the cause of multiple pathologies in various animal models and thus its elimination was a prerequisite for the adoption of rapeseed oil for human consumption (de Wildt and Speijers 1984; EFSA Panel on Contaminants in the Food Chain 2016). Glucosinolates are generally deemed innocuous when intact; however, their degradation products have been implicated as the causal agent of various health problems and are pungent, rendering both the oil and meal unpalatable (Jensen et al. 1995). Specifically, studies in rats repeatedly showed glucosinolates to adversely affect thyroid function through antagonism against iodine (Lo and Bell 1972; Vermorel et al. 1987). Thus, the elimination of glucosinolates was necessary before rapeseed oil could be incorporated in the human diet (Tripathi and Mishra 2007). Canadian breeding efforts led to the development of the first low erucic-acid and low glucosinolate rapeseed cultivar, named Tower (Stefansson and Kondra 1975). This and subsequent low-erucic acid, low glucosinolate cultivars were termed “canola” in North America, and double-low or 00 rapeseed in Europe, to distinguish them from wild rapeseed (Rakow 2011). Current commercial canola cultivars have near undetectable levels of glucosinolates (less than 20 µmol/g seed) and typically have less than 1% erucic acid, well below the maximum allowable levels of 2% erucic acid and 30 µmol glucosinolate/g seed as established by international law (Environment

Directorate and Organisation for Economic Co-operation and Development 2001; Rakow 2011). The near-elimination of erucic acid and glucosinolates deemed canola to be generally recognized as safe by the United States Food and Drug Administration (Direct Food Substances Affirmed as Generally Recognized as Safe, 21 C.F.R. §184.1555, [50 FR 3755, Jan. 28, 1985]) and is the fourth largest vegetable oil crop produced globally behind oil palm, soybean, and seed cotton (Figure 2.3) (Food and Agriculture Organization of the United Nations 2020). Domestically, canola production in Canada has grown steadily since its introduction in 1974 to approximately 20 million metric tonnes (Figure 2.4) and was estimated to contribute just under $30 billion to the Canadian economy (LMC International 2020).

Though unfit for human consumption, the potential value of erucic acid as an industrial lubricant and raw material for chemical manufacturing was realized by the USDA in the late 1960s (Goering et al. 1965; Nieschlag and Wolff 1971). At the same time, high erucic acid rapeseed (HEAR) cultivars with low glucosinolate levels had been developed in Sweden and HEAR breeding genotypes were developed in Canada (Calhoun et al. 1975). The first Canadian HEAR variety 'Reston' was registered by the University of Manitoba in 1982 (Registration number 2190).

Although canola remains primarily a crop grown for oil, its potential value as a renewable and sustainable protein source has been gaining attention. Whereas extensive breeding efforts continue to be focused on improving the quality and quantity of oil in canola (Tang et al. 2019), minimal attention has been directed to improving protein quality and content due to their inverse relationship with oil content (Chung et al. 2003; Hu et al. 2013). A canola cultivar with improved meal quality was assigned a patent to Agrigenetics Inc. and Dow AgroSciences LLC in 2016 claiming 45% crude seed protein and less than 18% fiber (US20120213909A1). To better

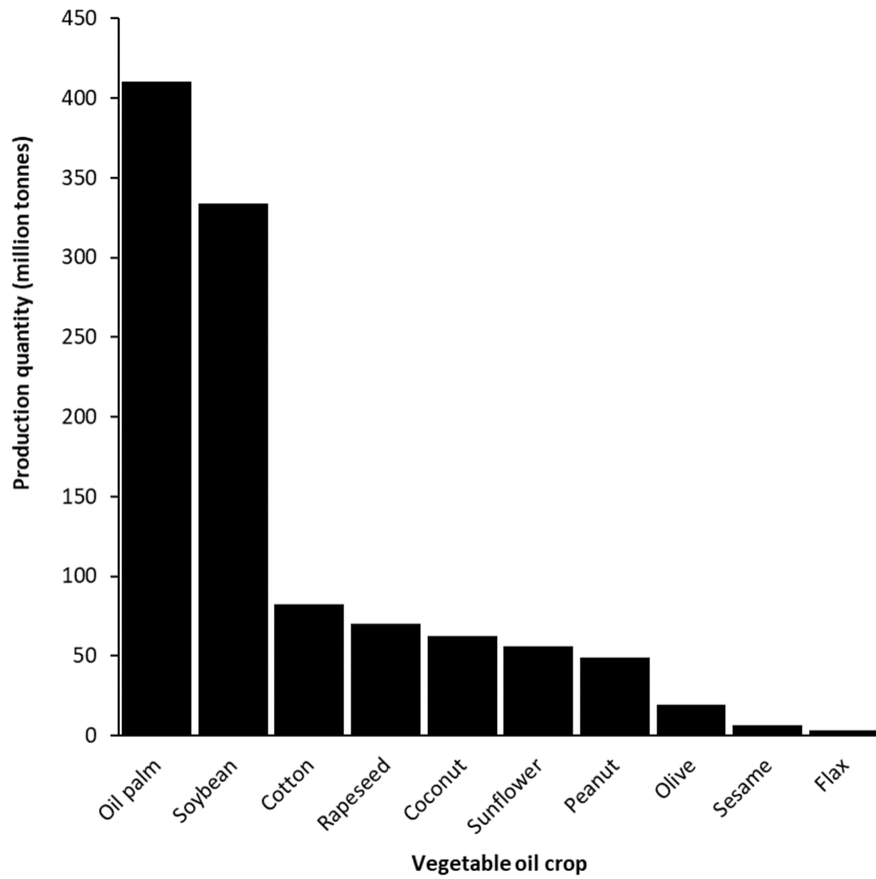

Figure 2.3. Rapeseed (including canola) was the fourth largest vegetable oil crop in 2019 based on global production quantity. (Data source: FAOSTAT, Food and Agriculture Organization of the United Nations, accessed January 2021)

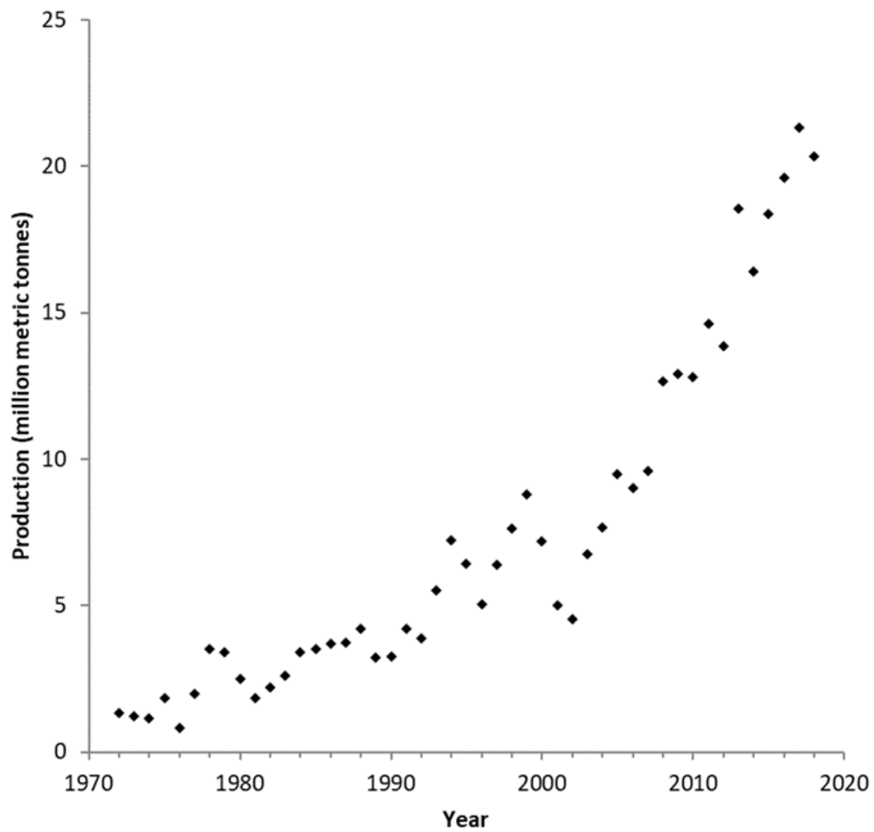

Figure 2.4. Canola production in Canada has increased steadily since the introduction of the first low-erucic acid, low-glucosinolate canola variety 'Tower' in 1974 (Source: Statistics Canada. Table 32-10-0359-01, accessed November 2019).

understand the challenges of developing canola protein as a dietary protein source, we need to first consider the technical, biochemical, and genetic factors that affect proteins in canola seed.

2.3 Extracting oil from raw seed

Canola/rapeseed meal is the residual portion of the seed that remains after the oil has been extracted (Tan et al. 2011). The oil extraction process is largely divided into a pre-processing step and an extraction step (Ohlson 1992). Pre-processing ensures maximum oil recovery from the seed: the canola seeds are pre-heated, after which they are flattened by rollers into flakes, and subsequently cooked (Ohlson 1992; Unger 2011; Mupondwa et al. 2018). Heating the seeds increases the pliability of the seed to ensure they can be thoroughly flaked without shattering while the flaking and cooking steps function primarily to rupture the seeds to allow oil to be released and to increase the surface area of the seed (Unger 2011).

The extraction step begins with the physical removal of oil from the cooked flakes by screw press (Ohlson 1992; Unger 2011). Remaining oil in the seed cake is then repeatedly extracted using a mixture of solvents collectively known as isohexane (Unger 2011). The meal and oil diverge in their processing after this solvent extraction; the oil is diverted for refining while the seed cake is sent to a desolventizer toaster where a combination of steam and heat remove residual solvent from the cake (Mag 1983; Ohlson 1992; Unger 2011). The resulting oil-free meal is then typically pressed into pellets for animal feed. Alternatively, protein can be isolated from the meal. The key determinate used to differentiate canola from rapeseed is its oil quality, however, in the oil-free meal, such a differentiation cannot be made. Henceforth, all protein products that are purified from meal originating from canola and rapeseed will be referred to as canola protein. The potential uses of canola protein as a food processing additive as well as non-food, non-feed applications of canola proteins have been recently reviewed (Wanasundara et al. 2016b).

2.4 Effects of processing on meal quality

The quality of canola meal destined for animal feed is judged largely on its amino acid profile and digestibility (Bailey and Stein 2019; Le Thanh et al. 2019). Conversely, the quality of canola protein destined for food processing is judged on its technical functionality (Foegeding and Davis 2011; Tandang-Silvas et al. 2011; Jones 2016), which is conferred by the molecular structure and content of the individual proteins contained within; thus, the quality of canola protein relates to the integrity of its components.

Extraction conditions can be altered to selectively extract different protein products with different purities and technical functionalities from canola meal (McCurdy 1990; Tan et al. 2011; Wanasundara 2011) and multiple patents have been assigned to different institutes to protect the intellectual property rights associated with these optimized conditions (Mupondwa et al. 2018). Harsh extraction processes can compromise protein quality (McCurdy 1990; Ohlson 1992; Unger 2011; Mupondwa et al. 2018). Seed proteins can be denatured by exposure to solvents (Fukushima 1969). High temperatures alter the digestibility of canola meal and can affect the structure of individual canola proteins (Jensen et al. 1995; Newkirk and Classen 2002; Thacker and Newkirk 2011; Mosenthin et al. 2016; Salazar-Villanea et al. 2017). Structural changes in canola proteins can abate their functional properties (Khattab and Arntfield 2009; He et al. 2014). As a feed supplement, protein content is routinely determined by its nitrogen content through combustion analysis (ISO 16634-1:2008; AOAC Method 990.03) thus, denatured proteins should not affect the protein determination of the meal. However, prolonged desolventizing can negatively impact the bioavailability of essential amino acids such as lysine, due to their modification from Maillard reactions (Lund and Ray 2017; Salazar-Villanea et al. 2017). In addition, the digestibility of canola meal decreases with prolonged toasting, rendering it less nutritious as a feed supplement (Almeida

et al. 2014; Mosenthin et al. 2016; Salazar-Villanea et al. 2016). If the meal protein is to be used as a functional ingredient in food processing, it is crucial that the structures of the proteins are not compromised.

2.5 Nutritional value of canola meal protein

Despite the potential reduction in quality during processing, canola meal and protein both have a nutritional profile that is comparable with soy, a common plant protein source. When considering the meal as a whole, the protein content of oil-free canola meal is roughly 38% (Barthet 2018a), compared to 44% for its soy counterpart (Barthet 2018b) (Table 2.1). Canola meal tends to have higher fiber, resulting in lower digestibility in animals which makes the meal less competitive compared to soy meal models (Adewole et al. 2017; Kasprzak et al. 2018). Currently, canola meal in its entirety is only used for animal feed; however, isolated protein products have the potential to be used as a dietary protein source for humans (Wanasundara et al. 2016b).

Protein quality for human nutrition is most commonly measured with a Protein Digestibility Corrected Amino Acid Score (PDCAAS) (Schaafsma 2000; Hughes et al. 2011; Millward 2012) or, more recently, with a Digestible Indispensable Amino Acid Score (Bailey and Stein 2019). The PDCAAS for a given protein is generally a ratio of its amino acid composition relative to that of a reference protein, then normalized to its digestibility; effectively, a maximum score of 1.0 indicated that one unit of the protein in question is able to supply all the essential amino acids after digestion (Hughes et al. 2011; Millward 2012). When considering only the protein fraction, canola protein is comparable with its soy counterpart in its amino acid profile (Table 2.2) (Le Thanh et al. 2019) and can generally satisfy human dietary requirements

Table 2.1. Comparison of seed composition between Canadian soybeans and western Canadian canola presented as the average value of 2013-2017 crops, inclusive. Table compiled with soybean data (Barthet 2018b) and canola data (Barthet 2018a) published by the Canadian Grain Commission. Non applicable measurements indicated with n.a.

	Soybean	Canola
Protein content (%) [1]	34.50	20.20
Oil content (%) [1]	18.40	44.50
Oil-free protein of the meal (%) [2]	43.80	37.80
Oleic acid (% in oil)	21.90	62.90
Linoleic acid (% in oil)	53.60	18.70
α-Linolenic acid (% in oil)	9.00	9.40
Total saturated fatty acids (% in oil)	15.20	6.70
Erucic acid (% in oil)	n.a.	0.01
Total seed glucosinolates (μmol/g, 8.5% moisture)	n.a.	10.00
Oil-free total glucosinolates of the meal (μmol/g, dry basis)	n.a.	21.00

[1]Based on 13% moisture in soybean and 8.5% moisture in canola

[2]Based on 13% moisture in soybean and 12% moisture in canola

Table 2.2. Comparison of moisture content, crude protein, and amino acid profiles between soybean meal and western Canadian canola meal. Values presented for canola meal are the mean and standard deviations of samples collected from five different crushing plants across western Canada. The origin of the soybean meal was not disclosed (Le Thanh et al. 2019).

	Soybean	Canola
Quality (%)		
Moisture	9.65	8.89 ± 0.43
Crude protein	46.10	37.82 ± 2.09
Amino acids (%)		
Alanine	1.97	1.64 ± 0.08
Arginine	3.27	2.15 ± 0.11
Aspartic acid	5.04	2.51 ± 0.12
Cysteine	0.61	0.81 ± 0.04
Glutamic acid	8.34	6.2 ± 0.41
Glycine	1.98	1.87 ± 0.10
Histidine	1.11	0.91 ± 0.05
Isoleucine	1.81	1.26 ± 0.09

Table 2.2 continued

Amino acids (%)	Soybean	Canola
Leucine	3.45	2.54 ± 0.14
Lysine	2.90	2.09 ± 0.08
Methionine	0.58	0.69 ± 0.03
Phenylalanine	2.25	1.44 ± 0.07
Proline	2.39	2.21 ± 0.11
Serine	2.31	1.47 ± 0.09
Threonine	1.73	1.52 ± 0.06
Tryptophan	0.69	0.52 ± 0.03
Tyrosine	1.74	1.06 ± 0.04
Valine	1.88	1.63 ± 0.09

for essential amino acids (Khattab and Arntfield 2009; Aider and Barbana 2011; Tan et al. 2011). The postprandial response to canola protein was empirically demonstrated to be equivalent to that of milk protein (Bos et al. 2007) and soy protein (Fleddermann et al. 2013).

2.6 Anti-nutrients in canola meal

The major drawback of canola protein nutrition is the presence of anti-nutritive compounds that negatively impact health, protein digestion, and amino acid availability (Gilani et al. 2005). The presence of glucosinolates, phytic acid, sinapine, and tannins reduce the nutritional value of canola meal (Bell 1993; Gilani et al. 2005).

Glucosinolates are sugar-amino acid conjugates and the content of the aliphatic class of these molecules was successfully reduced during the development of canola; however, residual quantities of phenolic glucosinolates in the meal can be broken down into off-tasting compounds that affect thyroid function (Tripathi and Mishra 2007) and show similar bioactivity to dioxins (Hannoufa et al. 2014). Phytic acid (phytates) is the main phosphorus-storage compound in seeds and functions as an antinutrient by binding mineral nutrients and inhibiting digestive enzymes (Hannoufa et al. 2014). Sinapine is an unpalatable phenolic compound that functions as an antinutrient by promoting feed-avoidance in sensitive animals (Harloff et al. 2012; Hannoufa et al. 2014). Tannins are polyphenolic compounds which, by interacting with proteins and the gastrointestinal tract, reduce the overall digestibility of proteins and bioavailability of amino acids (Gilani et al. 2005).

Progresses in traditional breeding and biotechnology have led to the reduction (Harloff et al. 2012; Hannoufa et al. 2014; Zhai et al. 2019) and sometimes near elimination of antinutritive compounds such as in the case of phytic acid (Sashidhar et al. 2019). For isolated protein products,

anti-nutritive compounds are not typically considered a limitation of use as during the process of protein isolation anti-nutritive factors are excluded (Schmidt et al. 2004; Das Purkayastha et al. 2013; Wanasundara et al. 2016b).

2.7 Seed development in Canola

Prior to examining seed storage proteins, a general understanding of seed development is necessary. A canola seed consists of an embryo and endosperm encased in a seed coat (Steeves 1983). Seed development in angiosperms is well characterized primarily based on empirical work using the model plant *Arabidopsis thaliana* (L.) Heynh. Seed development has been extensively reviewed in detail by numerous authors and will only be succinctly summarized below. Seed development in canola is broadly divided into three phases: the initial phase where rapid cell division occurs but seed growth is slow; the second phase where cells expand quickly and accumulate both storage protein and lipids; and the third phase where the seed matures and desiccates (Norton and Harris 1975; Crouch and Sussex 1981).

Embryo development begins with successful double fertilization: the fusion of sperm from the pollen with the egg forms the zygote while the endosperm results from the fusion of a second sperm with the polar nuclei (Lord and Russell 2002). The zygote then undergoes a series of coordinated divisions to establish polarity and generate tissue layers (Capron et al. 2009; Lau et al. 2012). Embryo developmental stages are described based on their visual morphology; the stages of embryogenesis in canola have been recorded in detailed drawings from light microscopy (Tykarska 1976, 1979) and scanning electron microscopy (Ilić-Grubor et al. 1998). The initial globular embryo becomes heart-shaped upon initiation of the cotyledon primordia; subsequently the embryo elongates into a torpedo shape; and as the cotyledons develop, they eventually bend over the embryo forming the bent cotyledon stage (Tykarska 1979; Capron et al. 2009; Lau et al.

2012). Although the progression from the globular stage to the bent cotyledon stage is ubiquitous across all studied canola genotypes, it is worth noting that the time of occurrence and duration of each stage shows genotypic variation (Tykarska 1980; Crouch and Sussex 1981; Fernandez et al. 1991; Yeung et al. 1996; Ilić-Grubor et al. 1998) (Figure 2.5).

Based on Norton and Harris (1975), the first phase of seed development is characterized by expeditious growth of the hull (silique) and slow seed growth; this phase of seed development persisted for up to four weeks (28 days) after pollination when examining field-grown plants of the winter canola cultivar 'Panter'. When grown under controlled environments, developing seeds of the spring canola cultivar 'Tower' complete the first phase of seed development at approximately 23 days after pollination (Crouch and Sussex 1981). The second phase of seed development as described by Norton and Harris (1975) occurs five to six weeks post pollination. This phase is characterized by the initiation of embryo development and the aggregation of storage compounds within the seed. However, light microscopy (Tykarska 1980) and scanning microscopy (Ilić-Grubor et al. 1998) experiments have shown embryo formation to initiate two to three days post pollination, with globular embryos visible one week after pollination, and torpedo-shaped embryos visible 11-12 days after pollination; Norton and Harris (1975) likely were referring to the expansion of the whole seed rather than the embryo. It is during this second phase of seed development that storage proteins begin to accumulate and continue to do so into the third phase of seed development at approximately 6-12 weeks after pollination (Norton and Harris 1975; Crouch and Sussex 1981).

2.8 Seed storage proteins of the Brassicaceae family

Seed storage proteins (SSP) are a unique class of proteins that are specifically expressed in the developing seed and accumulated within protein storage vacuoles in the mature seed

Weeks after pollination	1	2	3	4	5	6
Development						
Crouch and Sussex (1981) Planta 153: 64-74			H T	C		
Custers et al. (1999) Protoplasma 208: 257-264		G H T	C			
Fernandez et al. (1991) Development 111: 299-313			H	T	C	
Ilic-Grubor et al. (1998) Ann. Bot. 82: 157-165		H T C				
Yeung et al. (1996) Int. J. Plant Sci. 157(1): 27-39		H T		C		
CRU mRNA						
DeLisle and Crouch (1989) Plant Physiol. 91: 617-623				*		***
Finkelstein et al. (1985) Plant Physiol. 78: 630-636			*			***
Sjödahl et al. (1993) Plant Mol. Biol. 23: 1165-1176				*	***	
CRU protein						
Crouch and Sussex (1981) Planta 153: 64-74				*	***	
NAPIN mRNA						
DeLisle and Crouch (1989) Plant Physiol. 91: 617-623				*	***	
Finkelstein et al. (1985) Plant Physiol. 78: 630-636			*	***		
NAPIN protein						
Crouch and Sussex (1981) Planta 153: 64-74				*	***	

Figure 2.5. Coincidence of zygotic embryogenesis stages in spring *Brassica napus* L. with spatial-temporal patterns in the accumulation of seed storage protein transcripts and gene products during the first six weeks seed of development. Abbreviations are as follows: G, globular stage; H, heart stage; T, torpedo stage; C, cotyledon stage. A single asterisk (*) indicates initial detection and three asterisks (***) indicate peak levels.

(Shewry and Halford 2002; Wanasundara 2011), though contemporary evidence of synthesis during germination has been reported (Galland et al. 2014). These proteins function as a nutrient reservoir by converting nitrogen into storage-stable proteins which are mobilized during seed germination to support early seedling growth (Herman and Larkins 1999; Müntz 2007; Wanasundara 2011). Recently, SSP transcripts have been transiently detected in vegetative tissues in canola when plants were grown under different nitrogen fertility conditions (Bieker et al. 2019) which corroborates the role of these proteins in nitrogen storage.

In the Brassicaceae, the seed protein pool is dominated by the globulin-type SSP cruciferin and the albumin-type SSP napin, along with minor quantities of proteins that play a diverse array of metabolic and physiological roles, respectively (Wanasundara 2011). The globulin and albumin designation refers to the classical work of Osborne (Osborne and Osbornb 1924) who differentiated plant proteins based on their differential solubility in various solvents while the numeric values denote the sedimentation coefficient value of the protein. Late embryogenesis abundant (LEA) proteins, for example, are present in the seeds of most plants and function to mediate dehydration tolerance during seed maturation (Hundertmark and Hincha 2008). Specific to the seed protein pool of oilseed species may be substantial quantities of oleosins, proteins that enhance the stability of oil bodies in which these plants accumulate oil (Jolivet et al. 2009). Extensive proteomic studies on the seeds of *B. napus* at various developmental stages have systematized the profusion of proteins present in the seed protein pool (Hajduch et al. 2006; Jolivet et al. 2009; Yin et al. 2015; Kubala et al. 2015). Furthermore, these works report remarkable plasticity in the species composition of the seed protein pool through various environmental stimuli (Borisjuk et al. 2013; Yin et al. 2015) and physiological processes (Hajduch et al. 2006; Kubala et al. 2015). The spatial-temporal distribution patterns of SSP synthesis and accumulation in *B. napus*

during seed development have been described in detail and exhibits genotypic variation (Crouch and Sussex 1981; Hoglund et al. 1992; Sjödahl et al. 1993; Borisjuk et al. 2013).

2.8.1 Cruciferin of *B. napus*

Cruciferin is a salt-soluble globulin-type protein that accounts for approximately 60% of the total protein pool in canola seed (Crouch and Sussex 1981). Related 11S/12S globulin-type SSP from other crop species include cucurbitin from pumpkin (*Cucurbita maxima* Duchesne) and glycinin from soybean (*Glycine max* (L.) Merr.), with the latter being the most similar to cruciferin (Tandang-Silvas et al. 2010; Wanasundara 2011). Mature cruciferin is a large hexameric protein, whose subunits are each comprised of a disulfide-bridged α and β subunit. Recent work has suggested that cruciferin subunits may be reassembled into octameric structure for storage inside protein storage vacuoles (Nietzel et al. 2013). The general biosynthesis and deposition of globulin-type SSP has been extensively reviewed (Müntz 1998) and is summarized below.

Globulins are encoded by multi-gene families. Specifically, in canola, cruciferin is encoded by 9 - 12 genes (Wanasundara 2011). A search on the Uniprot knowledgebase reveals five curated accessions of *B. napus* cruciferin (Table 2.3) (https://www.uniprot.org, accessed November 2019). Globulin-encoding genes are translated into pre-propolypeptides on the rough endoplasmic reticulum (rER) and at minimum consist of three elements (listed from N to C terminus): a signal peptide, the α subunit, and the β subunit (Müntz 1998; Nietzel et al. 2013). The pre-propolypeptide is simultaneously translated and transported into the lumen of the rER and the signal peptide is detached resulting in a propolypeptide (Müntz 1998). As the propolypeptide is shuttled through the lumen of the rER, post-translational modifications such as glycosylation occur followed by the formation of disulfide bridges (Müntz 1998). The polypeptides are then oligomerized into

Table 2.3. Five manually curated entries for *Brassica napus* L. cruciferin are listed in the Uniprot database (Source: https://www.uniprot.org, accessed November 2019). Length is presented as amino acid (AA) count.

Entry	**Entry name**	**Protein names**	**Length (AA)**
P33522	CRU4_BRANA	Cruciferin CRU4 (11S globulin) (12S storage protein) [Cleaved into: Cruciferin CRU4 alpha chain; Cruciferin CRU4 beta chain]	465
P11090	CRUA_BRANA	Cruciferin (11S globulin) (12S storage protein) [Cleaved into: Cruciferin subunit alpha; Cruciferin subunit beta]	488
P33524	CRU2_BRANA	Cruciferin BnC2 (11S globulin) (12S storage protein) [Cleaved into: Cruciferin BnC2 subunit alpha; Cruciferin BnC2 subunit beta]	496
P33525	CRU3_BRANA	Cruciferin CRU1 (11S globulin) (12S storage protein) [Cleaved into: Cruciferin CRU1 alpha chain; Cruciferin CRU1 beta chain]	509

Table 2.3 continued

Entry	Entry name	Protein names	Length
P33523	CRU1_BRANA	Cruciferin BnC1 (11S globulin) (12S storage protein) [Cleaved into: Cruciferin BnC1 subunit alpha; Cruciferin BnC1 subunit beta]	490

entropically-favourable trimeric structures that transit to the Golgi apparatus (Müntz 1998; Tandang-Silvas et al. 2010). After sorting into vesicles at the trans-Golgi, the trimeric structures are transported to storage protein vacuoles where the prepolypeptides are enzymatically processed by endopeptidases to yield trimers of mature globulin subunits, each consisting of an α and β chain, respectively (Jung et al. 1998; Müntz 1998). Only upon production of the mature globulin subunits can mature hexameric globulins form (Jung et al. 1998). Across different canola genotypes, variation in the temporal accumulation of both total cruciferin transcripts and proteins is observed (Figure 2.5), thought the spatial-temporal distribution of individual cruciferin isoforms have yet to be explored.

In food processing, cruciferin functions as a good gelling agent (Aider and Barbana 2011) and has the ability to form stronger gels than napin (Kim et al. 2016), enabling it to be used in a wide array of food products (Wanasundara et al. 2016b). Cruciferin is also able to improve foaming stability in oil-containing mixtures (Pudel et al. 2015). Additional information on the functional properties of cruciferin are presented in Section 0.

2.8.2 Napin of *B. napus*

The second most abundant SSP in canola seed is napin which accounts for approximately 20% of the seed protein pool (Crouch and Sussex 1981; Byczyńska and Barciszewski 1999). Napin is classified as an albumin-type protein reflecting its solubility in water (Osborne and Osbornb 1924). Low molecular weight proteins were initially isolated from canola by Lönnerdal and Janson (1972). Characterization of these strongly basic proteins found them to be composed of two polypeptides, approximately 90 and 30 amino acids in length respectively, linked by disulfide bridges with a total molecular mass of 12 – 14 kDa (Lönnerdal and Janson 1972). Subsequent

experiments demonstrated that both napin chains were generated from the cleavage of a common precursor polypeptide (Ericson et al. 1986).

Seed storage albumins are encoded by multigene families (Boutilier et al. 1999; Mylne et al. 2014). In canola, a minimum of 10 to 16 genes were estimated by Southern blotting to encode napin (Josefsson et al. 1987; Scofield and Crouch 1987). Seven manually annotated entries for *B. napus* napin are currently listed in the UniProt Knowledgebase (Table 2.4) (https://www.uniprot.org, accessed November 2019).

The biosynthesis of napin mirrors that of cruciferin (Müntz 1998) but is initiated before the latter (Figure 2.5) and was reviewed in detail by Mylne *et al.* (2014).Translation of the napin-encoding genes at the ribosomes on the rough endoplasmic reticulum (rER) generates a single pre-proalbumin polypeptide consisting of a signal peptide and the two mature napin chains with linker peptides between each element (Ericson et al. 1986; Müntz 2007; Mylne et al. 2014). The signal peptide is removed during translation and the resulting proalbumin is directed into the rER lumen (Mylne et al. 2014). Inside the lumen, a pattern of eight cysteine residues conserved across similar plant storage albumins allow four disulfide bonds to be formed: two bonds are formed within the large subunit, and two intermolecular bonds are formed between the large and small subunits (José-Estanyol et al. 2004; Wanasundara 2011; Mylne et al. 2014). The polypeptide is then folded with the aid of chaperone proteins and can be subject to post-translational modifications (Mylne et al. 2014). Subsequently, the proalbumin is transported to the protein storage vacuole where the linker peptides are removed by proteases to form the mature protein (Byczyńska and Barciszewski 1999; Mylne et al. 2014). The tertiary and quaternary structures of mature napin contribute to its structural stability and resistance to digestion (Byczyńska and Barciszewski 1999; Wanasundara 2011; Mylne et al. 2014).

Table 2.4. Seven manually annotated entries for *Brassica napus* L. napin are listed in the Uniprot database (Source: https://www.uniprot.org, accessed November 2019). Length is presented as amino acid length.

Entry	Entry name	Protein names	Length
P24565	2SSI_BRANA	Napin-1A (Napin BnIa) [Cleaved into: Napin-1A small chain; Napin-1A large chain]	110
P09893	2SSE_BRANA	Napin embryo-specific (1.7S seed storage protein) [Cleaved into: Napin embryo-specific small chain; Napin embryo-specific large chain]	186
P17333	2SS4_BRANA	Napin (1.7S seed storage protein) [Cleaved into: Napin small chain; Napin large chain]	180
P27740	2SSB_BRANA	Napin-B (1.7S seed storage protein) [Cleaved into: Napin-B small chain; Napin-B large chain]	178
P01090	2SS2_BRANA	Napin-2 (1.7S seed storage protein) [Cleaved into: Napin-2 small chain; Napin-2 large chain]	178

Table 2.4 continued

Entry	Entry name	Protein names	Length
P01091	2SS1_BRANA	Napin-1 (1.7S seed storage protein) [Cleaved into: Napin-1 small chain; Napin-1 large chain] (Fragment)	133
P80208	2SS3_BRANA	Napin-3 (1.7S seed storage protein) (Napin BnIII) (Napin nIII) [Cleaved into: Napin-3 small chain; Napin-3 large chain]	125

Napin is water soluble thus enabling it to be incorporated into many food products (Pudel et al. 2015). In addition, napin also has exceptional foaming capacity and good emulsifying properties (Pudel et al. 2015) though these properties vary depending on the extraction process (Malabat et al. 2001). These properties enable napin to be used to partially replace more-costly milk proteins (Schwartz et al. 2015) and egg white (Wanasundara et al. 2016b) in food processing. Additional information on the functional properties of napin is presented in Section 0.

2.9 Genetic control of seed storage proteins in canola

Given the potential value of cruciferin and napin in food processing (Wanasundara et al. 2016b; Mupondwa et al. 2018), the development of cultivars with improved accumulation of either protein can add value to the crop. In order to facilitate breeding for improved seed storage protein in canola, knowledge of the existing genetic variation of the trait is required, in addition to an understanding of the genetic and non-genetic determinates that affect their synthesis and accumulation. To date, a single study examining genetic variation in seed storage protein was conducted in winter rapeseed (Malabat et al. 2003) while similar studies have yet to be undertaken in spring rapeseed germplasm. Seed storage protein content within the seed is dependent on not only the expression of SSP-encoding genes but also on the post-translational processes that generate the quaternary structure from precursor polypeptides, and their subsequent transfer of assembled SSP into storage vacuoles (Kawakatsu and Takaiwa 2010; Mylne et al. 2014; Le Signor et al. 2017). The quantitative nature of SSP content suggests complex genetic regulation of the trait and indeed many genes are involved in the synthesis and processing of SSP.

Quantitative trait loci (QTL) are regions within the genome whose occurrences are associated with specific quantitative phenotypes; multiple QTL each contribute some variation to the phenotype, and multiple genes may reside within each locus (Members of the Complex Trait

Consortium 2003). The successful identification of QTL associated with important seed quality traits in canola such as oil content (Zhao et al. 2006; Würschum et al. 2012; Teh and Möllers 2016; Chao et al. 2017; Behnke et al. 2018), fatty acid composition (Bao et al. 2018; Chen et al. 2018a), and glucosinolate levels (Toroser et al. 1995; Uzunova et al. 1995; Howell et al. 2003; Würschum et al. 2012; Schatzki et al. 2014) have been reported by numerous groups. In contrast, comparatively few reports have focused on the identification of QTL associated with total seed protein in canola (Zhao et al. 2006; Schatzki et al. 2014; Chao et al. 2017; Behnke et al. 2018). Even fewer are reports of QTL associated with protein quality traits such as SSP composition (Schatzki et al. 2014) and amino acid content (Wen et al. 2016).

Conserved motifs exist in the promoter regions of seed storage protein genes across diverse plant species (Fauteux and Strömvik 2009) suggesting the regulation of seed storage protein biosynthesis is governed by transcriptional mechanisms that evolved early in the evolution of the plant kingdom. The disruption of some of these conserved motifs in the canola cruciferin-encoding *cru1* gene (Sjödahl et al. 1995) and the napin-encoding *napA* gene (Stålberg et al. 1996) led to reduced accumulation of the respective protein product. Multiple transcription factors are known to regulate the expression of seed storage protein genes in *Arabidopsis*: *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) (Parcy et al. 1994; Ezcurra et al. 2000; Kroj et al. 2003; Lara et al. 2003), *FUSCA 3* (*FUS3*) (Vicient et al. 2000; Kroj et al. 2003), *LEAFY COTYLEDON 1* (*LEC1)* (Vicient et al. 2000; Kagaya et al. 2005; Pelletier et al. 2017), *LEC2* (Kroj et al. 2003), and multiple *MYC* transcription factors (Gao et al. 2016a). A detailed summary on the transcriptional regulation of seed storage proteins was reviewed by Verdier and Thompson (2008).

Non-genetic control of seed storage proteins in canola

Plant growth regulators are capable of modifying transcriptional activity and consequently, seed storage protein accumulation may in part be influenced by plant hormones (Verdier and Thompson 2008). Early works on seed storage proteins in canola implicated abscisic acid in the direct regulation of cruciferin and napin (Crouch and Sussex 1981; Crouch 1982; Finkelstein et al. 1985). In addition to its role in inducing the accumulation of seed storage proteins, abscisic acid is also associated with abiotic stress tolerance in plants (Sah et al. 2016). Thus, seed storage protein accumulation is influenced by the environment in which the plant is grown, as reported in canola (Barthet 2016) and winter wheat (Triboï et al. 2003). Indeed, from a breeding perspective, seed storage protein content is a quantitative trait which suggests the phenotype is influenced by the environment. Research in soybeans has demonstrated that differences in soil fertility were capable of increasing total seed protein content (Staswick et al. 1991) as well as altering storage protein composition (Boutilier et al. 1999; Paek et al. 2000). In canola, crop production on sulfur-limited land not only results in a reduction in glucosinolates and sulfur-containing amino acids (Grant et al. 2003; Wang et al. 2008; Borpatragohain et al. 2016) but was can also alter seed storage protein profiles in the crop (Poisson et al. 2019). Recent work has also correlated nitrogen supply with the expression of seed storage protein genes in canola (Bieker et al. 2019). Though the extent to which environmental factors influence seed storage protein accumulation has yet to be empirically determined, the aforementioned studies suggest that agronomic practices can be exploited to alter seed protein content and quality.

2.10 Canola protein as a novel food product

Currently, canola protein is primarily used for animal feed but has the potential to be directly consumed in food products (Chardigny and Walrand 2016; Mupondwa et al. 2018). For canola

protein to be permitted to be used as in human food, regardless of whether it is for dietary purposes or functional purposes, its safety must first be determined. Canola protein does not have a history of human consumption and thus makes it a novel food in the context of most food safety regulations. The procedures and standards for assessing the safety of novel foods, and the authorization required for their marketing, vary by country. Toxicology studies on the safety of a novel food are based on animal feeding experiments. While safety concerns were raised in early animal studies using rapeseed meal (non-canola quality), recent studies using canola-quality meal products concluded the ingredient to be safe. Collectively, the literature suggests the safety of rapeseed meal was compromised by anti-nutritive compounds and the strict use of canola-quality *B. napus* alleviates safety concerns. Specific safety studies on a cruciferin-rich (Mejia et al. 2009a) and a napin-rich (Mejia et al. 2009b) canola protein isolate, respectively, found no adverse effects in rats fed with a diet supplemented with up to 20% of either product. Due to the high costs associated with acquiring regulatory approval, authorization of novel foods is typically only sought for countries with potential markets. Canola protein is currently a fledgling in the plant protein sector whose commercialization has been spearheaded by three companies; the history of canola protein commercialization and its challenges have been recently analyzed by Mupondwa et al. (2018).

2.10.1 Allergenicity of canola seed storage proteins

The ability of napin to resist digestion suggests its potential as a food allergen (Moreno and Clemente 2008; Masilamani et al. 2012). Allergens are foreign proteins that are capable of evoking an adverse response from the immune system after they enter the body (Sicherer and Sampson 2018; Waserman et al. 2019). Generally, an immediate allergic response happens when an allergen has been encountered twice. The first time an allergen is encountered, the body produces a type of

antibody that can bind the allergen called immunoglobulin E (IgE); this first encounter and production of IgE is collectively called sensitization (Waserman et al. 2019). Following sensitization, if the allergen is encountered a subsequent time, an immune response occurs in which symptoms typically associated with an allergic response, such as urticaria (hives) and rhinitis (runny nose), are experienced (Sicherer and Sampson 2018; Waserman et al. 2019). Repeated exposure to an allergen can result in increasing severity of the symptoms. The molecular mechanisms behind how IgE function during allergies have been reviewed by Gould and Sutton (2008).

Napin from canola has been identified as a potential allergen (Aider and Barbana 2011) based on its ability to bind IgE from sensitized patients (Monsalve et al. 1997). Further, napin, in addition to cruciferin, have both been identified as allergenic proteins in cold-pressed canola oil that reacted with IgE from sensitized children (Puumalainen et al. 2015). These results had been considered by the regulatory bodies when determining the safety of canola protein products, but were not deemed to be a substantial risk (EFSA Panel on Dietetic Products Nutrition and Allergies 2013).

An allergenic reaction can be caused by a protein, to which a patient has not been previously sensitized, if it is sufficiently similar in structure to a known allergen (Bonds et al. 2008; Fiocchi et al. 2016). This cross-reactivity suggests that patients who are allergic to other plant globulins and plant albumins, respectively, may react to cruciferin and napin and vice versa (Bonds et al. 2008; Moreno and Clemente 2008; L'Hocine et al. 2019). Mustard was added to a list of priority allergens in Canada (2011) following a systematic review completed by Health Canada (2009) who found sufficient scientific evidence of its allergenicity to be relevant to the Canadian public. The 2S albumin of mustard seed is the major allergen of mustard (Marambe et

al. 2014; L'Hocine et al. 2019) and is similar in its amino acid sequence to napin (Moreno and Clemente 2008; Wanasundara 2011) suggesting that people who are sensitized to mustard may also be allergic to napin.

Although the allergenicity of proteins in canola oil have been demonstrated and concerns of their potential to cross-react in patients sensitized to related protein allergens have been raised, it is noteworthy to consider that there have been no reports on allergies to canola oil, likely due to the refining process (Puumalainen et al. 2015). Furthermore, the long history of mustard oil consumption in certain parts of the world and the nutritional value of canola oil outweigh concerns of its allergenicity (Gylling 2006); however, questions regarding the safety of concentrated canola protein products need to be evaluated given their novel nature.

2.10.2 Regulation of novel foods in the United States

Canola meal and canola proteins have not had an extensive history of consumption and are considered novel foods by the United States Food and Drug Administration (FDA). Novel foods in the United States that are deemed generally recognized as safe (GRAS) by the FDA fall outside the purview of the Federal Food, Drug, and Cosmetic Act (Paul et al. 2017). For a novel food to be granted GRAS status, its safety must be demonstrated empirically and be recognized as safe by the scientific community. Under the current GRAS Notification program, the FDA does not conduct scientific testing on the novel food to determine its safety, but rather, the Agency reviews the safety data it receives (Paul et al. 2017). The general process required for an ingredient to acquire GRAS status involves three steps: first, the sponsor of the product performs or submits relevant scientific studies on the safety of the novel food in the form and dose it is intended to be used; second, the data is submitted as part of a GRAS notice to the FDA for review; and third, the FDA will respond with a letter stating whether or not it has questions regarding the GRAS

conclusion of the novel food based on the submitted notice (Paul et al. 2017). A letter indicating the agency has no questions equates to acknowledgement and acceptance of the GRAS status of the ingredient. To date, six GRAS notices regarding *B. napus* products, of which three pertain to meal protein, and the FDA's response to each one is found in the FDA's online GRAS Notice Inventory (Table 2.5) (https://www.fda.gov/food/generally-recognized-safe-gras/gras-notice-inventory).

2.10.3 Regulation of novel food in the European Union

As in the United States, novel foods must be deemed safe before they can be used in the European Union. Canola protein and canola protein isolates are considered novel foods in the EU, which by definition are foods that are not "used for consumption to a significant degree" prior to May 15, 1997 (Regulation (EC) No 258/97) (van der Spiegel et al. 2013). The safety determination process in the EU is generally similar to that of the GRAS Notification Program: sponsors of novel food submit data on the safety of the ingredient under its intended use to the European Commission, who then forwards the application to the European Food Safety Authority (EFSA). The EFSA assesses the data and publish its finding as a Scientific Opinion; a committee votes whether to accept the finding; and the Commission issues authorization based on the vote. To date, the EFSA has published only one Scientific Opinion on protein isolate derived from a mix of *B. napus* and *B. rapa*, finding the product to be safe under its intended use (EFSA Panel on Dietetic Products Nutrition and Allergies 2013).

2.10.4 Regulation of novel food in Canada

In Canada, Health Canada is the regulatory body overseeing the safety of food additives and its decisions are enforced by the Canadian Food Inspection Agency (CFIA). Purified canola

Table 2.5. Six canola (*Brassica napus* L.) products are listed in the United States Food and Drug Administration's (FDA) Generally Recognized as Safe (GRAS) notice inventory since the inventory was established in 1998. In all applications, the FDA had no questions regarding the GRAS status of the product. Current as of December 2020.

GRAS Number	Notifier	Substance	Intended Use	Date of Closure	FDA Response
683	DSM Innovation Company	Canola protein isolate	Dietary protein Food processing	2017	no questions
682	Cargill Inc.	Lecithin from canola	Food processing Dietary fat Dietary choline	2017	no questions
533	American Lecithin Company	Lecithin from canola	Food processing Dietary fat Dietary choline	2015	no questions
425	Danone Trading B.V.	Canola oil (low erucic acid rapeseed oil)	Dietary fat (infant formula)	2012	no questions

Table 2.5 continued

GRAS Number	Notifier	Substance	Intended Use	Date of Closure	FDA Response
386	BioExx Specialty Proteins, Ltd.	Canola protein isolate and hydrolyzed canola protein isolate	Food processing Dietary protein	2011	no questions
327	Archer Daniels Midland Company	Cruciferin-rich canola/rapeseed protein isolate and napin-rich canola/rapeseed protein isolate	Dietary protein	2010	no questions

protein and canola protein isolates are considered novel foods in Canada as they do not have a history of use for human consumption (Day 2013). The process to request approval for novel foods approximates that of the United States and European Union: first, the applicant submits pertinent scientific studies and data on the safety of the ingredient to the Food Directorate at Health Canada; second, the application is forwarded for scientific review after the Food Directorate has verified the information is complete; and third, the results of the review are sent back to the Food Directorate who then decides on the authorization of the ingredient in question. Similar to the GRAS Notice Program, the decision of the Food Directorate is communicated as a Letter of No Objection for approved novel foods. An inventory of novel food decisions that received no objection (https://www.canada.ca/en/health-canada/services/food-nutrition/genetically-modified-foods-other-novel-foods/approved-products.html) is available online. Currently, no canola protein products have been approved as a novel food in Canada.

2.11 Canola meal as a protein source in animal husbandry

Protein is an integral part of animal diets; however, it is typically a costly component of feed (Kim et al. 2019b) and efforts to explore low-cost protein supplements have been untaken. The low cost and abundance of rapeseed and canola meal renders it an economical protein source (Lomascolo et al. 2012; Wanasundara et al. 2016b; Mupondwa et al. 2018). Mammals that are commercially raised for food production largely can be classified based on their ability to acquire nutrients from plant material: ruminants, such as cattle, have the capacity to ferment plant material prior to digestion allowing for successful nutrient uptake, while non-ruminants, such as swine, lack this ability (Wu et al. 2014b). This difference in digestive anatomy has implications in the use of canola meal as a protein supplement in the respective feeds of these animals: specifically, differing sensitivities to glucosinolates between ruminants and non-ruminants generally dictate the relative

quantity of canola meal that can be incorporated into feed (Thomke 1981; Tripathi and Mishra 2007). In addition to cattle and swine, poultry and fish are also raised for food on commercial scales and the feasibility of using canola meal in these production systems has also been studied.

2.11.1 Cattle

The United States is the largest importer of Canadian canola meal (Table 980-0012, Statistics Canada, 2019). The use of canola meal as a protein supplement has been widely adopted by the cattle (*Bos taurus* Linnaeus, 1758) industry since the development of canola (Thomke 1981). Specifically, canola meal is one of the standard protein supplements for dairy cattle production (Swanepoel et al. 2014). Given the ruminant nature of cattle, the incorporation of plant-based protein supplements such as soybean meal is a well-established practice. A meta-analysis conducted on dairy cattle found that the use of canola meal was superior to soybean meal as a protein supplement as measured by feed intake, milk yield, and milk protein yield (Huhtanen et al. 2012). In cattle produced for meat, the substitution of barley as the protein source with canola meal during the growing period did not increase feed efficacy compared to the use of barley (He et al. 2013). Although the conversation of feed to animal protein was not improved, the use of canola meal in feedlot cattle production may still be a more economical and sustainable alternative than using barley. Taken together, these studies support the continued use of canola meal as a protein supplement in cattle production.

2.11.2 Swine

Soybean meal is widely used as an economical protein source in swine (*Sus scrofa domesticus* Erxleben, 1777) production and remains a reference for novel plant-based protein supplements. Canola meal can partially replace soybean meal in the swine diet without ill-effect on animal health and production efficiency (Thacker and Newkirk 2011); however complete

replacement is hampered by its high fibre content and the presence of various secondary metabolites, especially glucosinolates based on feeding experiments (Mejicanos et al. 2016). Swine are more sensitive to glucosinolates compared to cattle (Tripathi and Mishra 2007; Mejicanos et al. 2016) and residual glucosinolates in canola meal-containing feed will drive the animals to choose soybean meal-supplemented feed when presented with the option (Landero et al. 2012). Additional processing steps during meal production such as toasting were able to mitigate feed rejection presumably due to the decomposition of residual glucosinolates; however, production efficacy did not supersede that of soybean meal (Thacker and Newkirk 2011). Technical advances in canola meal processing will enable it to be a competitive and cost-efficient alternative to soybean meal (Mejicanos et al. 2016)

2.11.3 Poultry

Poultry (*Gallus gallus domesticus* Linnaeus, 1758) ranks second as the most consumed animal protein globally (Chai et al. 2017). Protein in the chicken diet is often supplied from a mix of animal and plant sources (Hossain et al. 2012; Kim et al. 2019b), the latter of which is primarily from soybean due to its nutritional properties (Ravindran et al. 2014). Although chickens grow faster with some animal protein in their diet, work has been conducted to observe the effects of using solely vegetable-sourced protein for chicken (Proudfoot et al. 1983; Newkirk and Classen 2002; Vieira and Lima 2005; Hossain et al. 2012; Kasprzak et al. 2017). Canola meal has been explored as a plant-based protein source for chicken production and was found to be generally comparable to soybean meal (Proudfoot et al. 1983; Bryan et al. 2019). An examination of the lower digestive tract of chickens fed canola meal failed to identify intact proteins typical of canola seed suggesting that the meal protein had been completely digested by the animal (Kasprzak et al. 2017). Of cattle, swine, and chicken, the growth of the latter appears to be least affected by the use

of canola meal as the sole source of feed protein, as evidenced by the possibility of using canola meal as the sole protein source in feed (Adewole et al. 2017).

2.11.4 Aquaculture

Fish currently account for approximately 20% of the animal protein consumed globally and its production is split approximately in half between aquaculture (farmed fish) and capture fisheries (wild caught fish) (Food and Agriculture Organization of the United Nations and FAO 2018). The standard protein source for aquaculture has historically been fish meal; however, high costs, limited supply, and sustainability concerns led to a decline in the use of fish meal in favour of plant protein sources and other novel feedstuff (Food and Agriculture Organization of the United Nations and FAO 2018; Kim et al. 2019b; Turchini et al. 2019). The complete replacement of fish meal with other protein sources has been reported to be successful across various aquaculture species, particularly those of lower trophic levels (Kim et al. 2019b). As with land animals, fish are sensitive to glucosinolates and high levels in their feed from the inclusion of canola meal can negatively impact production efficiency (Tripathi and Mishra 2007). Nonetheless, canola meal and different canola protein isolates have proven successful as a protein supplement in various aquaculture species (Turchini et al. 2019) including rainbow trout (*Oncorhynchus mykiss* Walbaum. 1792) (Nagel et al. 2012; Slawski et al. 2012; Adem et al. 2014), and high-value crustaceans such as mitten crab (*Eriocheir sinensis* H. Milne-Edwards, 1853) (Luo et al. 2011) and shrimp species (*Litopenaeus stylirostris* Stimpson, 1871; *L. vannamei* Boone, 1931) (Cruz-Suarez et al. 2001; Suárez et al. 2009).

Seed storage proteins in food processing

In addition to fulfilling its role as a macronutrient, protein can also be supplemented in food products to improve its nutritional profile. Plant-based protein isolates, namely derived from

soybean and pea and comprised mostly of SSP, are ubiquitously available on the health food market in either its pure form or supplemented in food products (Kumar et al. 2017). The continued availability of these products on supermarket shelves speaks to the palatability of plant-based proteins and their acceptance by the general consumer.

Proteins also serve a non-nutritional role in food by directly controlling its physical properties (Day 2013; Mupondwa et al. 2018). To ensure food products, particularly meat analogues, meet the textural expectations of consumers (Hoek et al. 2011), commercial processors use different proteins (Kumar et al. 2017) in addition to different mechanical process (Dekkers et al. 2018). As different SSP have different physiochemical properties based on their amino acid composition and consequently their structure, not all SSP can fulfill the same functional role (Wu and Muir 2008; Tan et al. 2011). More importantly, plant-based protein isolates contain a mixture of SSP whose constituents may be mutually antagonistic in their functionalities and raw isolates may therefore be unfit for use as a structural additive in commercial food processing despite its nutritive contribution (Wu and Muir 2008). One solution towards the adoption of plant protein for food processing applications is to supplement single SSP, or single SSP-enriched isolates rather than raw isolates. In this way, control over the texture of the final food product can be better maintained. Furthermore, SSP can be strategically used to improve not only the absolute protein content of the food product, but also the amino acid profile of the product (Kohno-Murase et al. 1995; Falco et al. 1995), allowing for marketing to more diverse consumer bases.

2.12 Separation and quantification of seed storage proteins

Globulins and albumins can be separated based on their difference in solubility in dilute saline solution (Osborne and Osbornb 1924). Indeed, napin and cruciferin have been successfully separated on the basis of solubility in salt (Tan et al. 2011; Cheung et al. 2014a, 2014b; Yang et

al. 2014; Gerzhova et al. 2016). Separation of cruciferin and napin can also be facilitated by the differences in solubility in different pH (Bérot et al. 2005; Wanasundara et al. 2012; Cheung et al. 2014b, 2014a; Gerzhova et al. 2016). This appears to be the basis, at least partially, of patents granted for the commercial separation of these major canola SSP. A comprehensive review on separation technologies specific to canola protein and a selected list of relevant patents was recently compiled by Mupondwa et al. (2018).

On a bench scale, the separation of cruciferin and napin can be accomplished by means of chromatography (Crouch and Sussex 1981). Although chromatographic separation and purification results in a highly purified single-SSP product, the cost of such methods, the technical expertise required for their operation, and the low recovery rate (Bérot et al. 2005) limit such methods from being used on a commercial scale until broader market demand for these proteins develop. The ease of protein separation can be improved if the native SSP pool has reduced species complexity or consists of a single protein species.

Breeding efforts towards altering SSP composition requires a method for quantifying individual canola proteins. Most breeding programs lack the expertise and infrastructure required to perform chromatography, and although accurate, such analytical methods lack the throughput required for phenotyping large populations. In current breeding programs, SSP quantification is performed by SDS-PAGE followed by Coomassie staining and densitometry (Schatzki et al. 2014). However, this method fails to distinguish between different proteins with similar molecular masses and preferential binding of the dye to aromatic amino acids can skew the accuracy of the assay (Congdon et al. 1993). Immunological methods to quantify individual SSP may offer the accuracy and throughput needed for breeding programs.

2.13 Immunodection and quantification of seed storage proteins

Electrophoretic separation of proteins followed by immunodetection of target antigens was first described by Towbin *et al.* (1979) to analyze ribosomal proteins. To date, this variant of Western blotting remains a prevalent technique in the literature being employed in approximately 10% of all protein research articles (Moritz 2020). Despite criticisms of its reproducibility due to variation in antibody quality (Moritz 2020), Western blotting can be a valuable technique for the relative quantification of proteins in mixtures provided proper extraction protocols, standard curves, and normalization protocols are implemented (Janes 2015). Specifically, the use of Western blotting to quantify individual SSP within total seed protein mirrors conventional size exclusion chromatography (SEC) methods (Malabat et al. 2003) in that both techniques rely on the separation of proteins by molecular mass prior to estimating the abundance of target proteins of a given theoretical mass. Arguably, Western blotting requires less infrastructure to perform and the use of an antibody allows for greater discrimination of proteins compared to mass alone. Conversely, Western blotting has been implemented to qualitatively elucidate basic knowledge of SSP structure (Rödin and Rask 1990; Jung et al. 1998; Kawagoe et al. 2005) and post-translational modifications (Wan et al. 2007). Similarly, immunological methods have also been employed to study the spatial distribution of SSP (Hoglund et al. 1992) and quantify temporal patterns in their accumulation (Crouch and Sussex 1981; Murphy and Cummins 1989; Murphy et al. 1989) through seed development.

2.14 Manipulation of seed storage proteins in select crop species

The challenges of chemical separation of SSP can be circumvented altogether with plants whose seed protein pool contains a single protein species. Seed storage proteins are regulated on a genetic level, and therefore genetic technologies can be employed to alter SSP composition in

plants. Efforts undertaken to genetically alter the SSP profile of various crop species, for functional (Gil-Humanes et al. 2012; Sikdar et al. 2016) or nutritive purposes (Kim et al. 2013a; Wu and Messing 2014; Sikdar et al. 2016) have been successful.

2.14.1 Seed storage protein manipulation through conventional breeding

Conventional breeding relies on cycles of crossing and selection to generate distinct genotypes with improvements in desired traits and genetic variability is required in plant breeding to facilitate genetic gain in the trait of interest (Henderson and Salt 2017). To date, only one survey of SSP variability in rapeseed has been conducted in France (Malabat et al. 2003) which found historic rapeseed varieties to be relatively richer in napin content compared to modern canola-quality varieties. Currently, the genetic variation that exists for SSP in Canadian germplasm remains unknown.

Diversity in SSP composition has been reported in various crop species such as wheat (Carmona et al. 2010) and soybean (Liu et al. 2006), implying similar diversity may exist in canola. Wild soybean accessions with unique SSP profiles have been successfully incorporated into breeding programs to generate segregating populations with variable seed protein profiles (Kim et al. 2013b) indicating the manipulation of storage protein composition can be achieved by conventional breeding.

Intentional selection for SSP composition in canola has never been reported; however, selection for oil content is believed to have also inadvertently selected for increased cruciferin content (Malabat et al. 2003). A transcriptomic study of the breeding response in Chinese winter rapeseed over two decades of selection for improved fatty acid profiles and oil content revealed minimal changes to the transcription of cruciferin and napin, though actual levels of SSP were not

determined (Hu et al. 2009). Interestingly, protein and oil content both increased concomitantly with overall yield (Hu et al. 2009) suggesting the inverse relationship between the two traits can be broken. Furthermore, cruciferin and napin levels were found to be highly heritable suggesting the possibility for the successful genetic improvement of the levels of either protein through conventional breeding (Schatzki et al. 2014).

2.14.2 Soybean

Soybean is globally cultivated as an oilseed, and is an important protein source in Asia (Rizzo and Baroni 2018). Approximately 70% of the soybean SSP pool is composed collectively of the multimeric globulins glycinin (11S) and β-conglycinin (7S) (Nishinari et al. 2014, 2018). While soy protein is versatile in both its functional and edible properties (Singh et al. 2008), a significant portion is directed towards tofu production. The marketability and consumer acceptance of tofu is mostly based on its texture; thus, research has focused on the relationship between soybean protein and its functional properties in tofu production. Seed storage protein composition is known to play a critical role in controlling the texture of tofu (Saio et al. 1969), and more recent work has demonstrated that the subunit composition of individual SSP also plays a critical role (Poysa et al. 2006). An effort to change the SSP profile of soybeans through a combination of mutagenesis and conventional breeding have led to the development of genotypes with improved tofu-making qualities (Takahashi et al. 2003; Poysa et al. 2006; Zarkadas et al. 2007; Kim et al. 2013b).

Concomitant with improving protein functionality, efforts to improve the nutritional value of soybean protein have also been undertaken. The improvement of lysine content in the total seed protein of soybean was made possible by altering its feedback regulatory mechanism during biosynthesis (Falco et al. 1995). Effort aimed at improving the methionine content of soybean

protein was made through the expression of a chimeric gene encoding a methionine-rich δ-zein SSP from corn under the control of an endogenous soybean β-conglycinin promoter (Kim and Krishnan 2004). Increased accumulation of foreign proteins was enabled by the successful suppression of β-conglycinin production by RNAi (Schmidt and Herman 2008; Kim et al. 2014). These works demonstrate the possibility for the continued improvement of soy protein nutrition by enabling the improved accumulation of higher-nutrient or therapeutic proteins (Schmidt and Herman 2008) at the expense of lower-nutrient endogenous SSP.

2.14.3 Wheat

In wheat, the seed protein pool, colloquially referred to as gluten, is comprised of the SSP gliadins and glutelins (Shewry and Halford 2002; Wieser 2007). Wheat SSP primarily serves a structural rather than nutritive role in food and as such and breeding efforts have centered around improving its functionality. Despite the prevalence of wheat in the diet, the protein constituents of gluten elicits immunogenic reactions in susceptible patients and recent work has focused on the elimination of its reactivity (Gil-Humanes et al. 2010; Altenbach et al. 2015). Efforts to improve the functional properties of wheat have been made by various groups through the overexpression of glutenins (Altpeter et al. 1996; Blechl and Anderson 1996; Barro et al. 1997). The ectopic expression of genes encoding high-molecular-weight glutenin subunits resulted in an increased accumulation of the protein (Altpeter et al. 1996; Blechl and Anderson 1996) as well as increased dough elasticity (Barro et al. 1997) in a dose-dependent manner. The downregulation of gliadins using RNAi has also been shown to improve the functionality of flour (Pistón et al. 2011; Altenbach et al. 2014), demonstrating the direct relationship between SSP composition and protein functionality. Furthermore, flour from gliadin-deficient wheat genotypes was demonstrated to have reduced immunogenicity (Gil-Humanes et al. 2010; Altenbach et al. 2015) suggesting that

the manipulation of SSP can be a strategy to generate food products to cater to specialized dietary needs. Collectively, these studies in soybean and wheat provide empirical evidence on the possibility of manipulating seed storage proteins in canola.

2.14.4 Canola

To date, canola remains primarily an oilseed crop and its meal protein largely remains a by product (Troise et al. 2018). Meal protein has only been considered for use as a nutritional supplement, and consequently efforts to alter the SSP profile of canola has only been examined from a nutritional standpoint. To improve the methionine content of canola meal protein, a chimeric gene containing the coding sequence of a methionine-rich 2S SSP from Brazil nut (*Bertholletia excels* Humb. & Bonpl.) under the control of either a phaseolin promoter (Altenbach et al. 1992) or soybean lectin promoter (Guerche et al. 1990) (both promoters whose native functions drives SSP expression in the seed), was transformed into canola. Seeds from the resultant transgenic plants showed elevated levels of methionine (Guerche et al. 1990; Altenbach et al. 1992), suggesting that the amino acid profile of canola meal protein was amenable to change and could be altered through the expression of foreign proteins in the seed. Efforts to improve the lysine content of canola meal protein were made by disrupting the feedback-regulation of lysine during biosynthesis, which effectively doubled the total seed lysine content (Falco et al. 1995). This indicates that improvements to the levels of individual amino acids can be achieved. More importantly, these works suggest that the plant is able to accumulate higher than normal levels of individual amino acids without severe physiological consequences. Subsequent efforts to improve the quality of canola meal protein were made using advancements in antisense technology with the goal of selectively attenuating individual SSP (Kohno-Murase et al. 1994, 1995). By suppressing the accumulation of napin transcripts, an increase of cruciferin protein was observed

without major effects on fatty acid composition (Kohno-Murase et al. 1994); similarly, reducing the level of cruciferin transcripts resulted in an increase of napin protein content as well as improved levels of methionine, lysine, and cysteine (Kohno-Murase et al. 1995). In both cases, modifications to the SSP did not result in changes to the total macromolecular profile of the seed. Interestingly, co-suppression of both cruciferin and napin together did not lead to a compensatory increase in oil (Rolletschek et al. 2020) suggesting the inverse relationship between oil and protein content is not a simple competition for metabolic intermediates.

2.15 Plant breeding in the omics era

Advancements in omics technologies have enabled breeders to study the underlying mechanisms that govern desirable phenotypes on a genomic, transcriptomic, and proteomic level. These technologies improve the efficiency of breeding programs and improve the speed in which new cultivars can be generated.

2.15.1 Genomics

A major goal of genomics in crop agriculture is to correlate genotypic information with phenotypic data (Edwards et al. 2013). In this way, selections can be made early in the growth cycle on the basis of molecular markers that are impervious to environmental variation thus improving both efficiency and accuracy. Early genotyping efforts relied on restriction digestion and hybridization to discern allelic variation. The advent of PCR subsequently enabled the development of amplification-based genotyping platforms with improved throughput and efficiency (Edwards et al. 2013). Next, genotyping by single nucleotide polymorphism (SNP) markers was developed as the markers were numerous and widely distributed across plant genomes (Zhu et al. 2008). Multiplex SNP genotyping using crop-specific bead chip arrays such as the *Brassica napus* 60K Illumina Infinium™ SNP array (Clarke et al. 2016; Mason et al. 2017) further

enhanced the efficacy of acquiring genotypic information in crop plants (You et al. 2018) by enabling thousands of SNPs to be surveyed simultaneously. As next-generation sequencing becomes increasingly more accessible, genotyping with SNP markers may potentially be replaced by whole genome sequence data. While early whole-genome sequencing protocols such as restriction-site associated DNA sequencing (Baird et al. 2008) and genotyping-by-sequencing (Elshire et al. 2011) used restriction enzymes to reduce the complexity of the genome and cost, rapid improvements in sequencing technologies now enable whole genomes to be economically sequenced at unprecedented speeds (Scheben et al. 2017).

2.15.1.1 Linkage mapping

Linkage mapping identifies significant correlations between molecular markers and traits of interest in a structured population (Myles et al. 2009; Xu et al. 2017); such populations are typically generated by crossing parents with divergent phenotypes and subsequently fixing recombination events in the segregating progeny through doubled-haploid production or repeated cycles of selfing. The size and relatedness between individuals in the population directly affect the resolution of the mapping study (Cockram and Mackay 2018). The mapping population along with the parental genotypes are phenotyped in replicated experiments and genotyped by SNP-chip arrays (Rasheed et al. 2017). A linkage map displaying the physical order of the SNP markers and the genetic distance between them can then be built with various software packages by inferring recombination ratios (Lorieux 2012; Meng et al. 2015). Correlations between markers and phenotypic data are performed using various statistical methods (Meng et al. 2015). Linkage mapping has been successfully employed in canola to identify molecular markers that co-segregate with total seed protein (Zhao et al. 2006; Würschum et al. 2012; Teh and Möllers 2016; Chao et

al. 2017; Behnke et al. 2018), seed storage protein content (Schatzki et al. 2014), and non-essential amino acid content (Wen et al. 2016).

2.15.1.2 Association mapping

Similar to linkage mapping, association mapping also aims to identify significant correlations between genotypic information and phenotypes (Zhu et al. 2008; Myles et al. 2009). However, unlike linkage mapping which relies on recombination events occurring in a structured bi-parental population, association mapping (also referred to as genome-wide association studies; GWAS) takes into consideration all historic recombination events in a population of genetically diverse individuals (J. and Cloutier 2012). The practical implications of this difference is twofold: first, by using a population of genetically diverse individuals, no time investment to generate biparental mapping populations is needed, which effectively shortens the time required for association mapping; second, by using genetically diverse genotypes and considering all historic recombination events, the resolution of association mapping is higher than that of linkage mapping, and causal SNPs can be identified rather than loci (Yu and Buckler 2006; Myles et al. 2009; J. and Cloutier 2012).

Genome-wide association studies were initially proposed for use in human genetics in the 1990s (Risch and Merikangas 1996). At the same time the first association studies were reported on grain crops (J. and Cloutier 2012). Advancement in statistical methods to account for genetic challenges that are unique to crops were first incorporated into a GWAS for flowering time in maize a decade later (Thornsberry et al. 2001). To date, GWAS has been successfully and regularly implemented to study many major crop species (Xu et al. 2017; Liu and Yan 2019; Tian et al. 2020).

The population size required for GWAS varies depending on the goal of the study and the inherent genetic structure of the population: 300 individuals are sufficient for candidate gene validation and 1000 – 5000 individuals are recommended for marker discovery (Ersoz et al. 2009). Although larger populations are able to improve the power to detect QTL in canola by GWAS (Li et al. 2016b), populations of 200-500 individuals have been typical in recent studies (Table 2.6). Phenotypic data is collected from replicated experiments and genotypic data, often in the form of SNP markers, is typically acquired through high-throughput SNP-chip arrays that are crop specific (Rasheed et al. 2017). Recent studies have used whole genome sequencing and transcriptome sequencing in place of SNP-chip arrays to generate genotypic data (Huang and Han 2014). The significance of the association between each SNP and the phenotype is then tested using different statistical models while considering the relatedness between each individual of the population and possible cryptic population structures (Zhu et al. 2008; J. and Cloutier 2012; Xu et al. 2017). Finally, the results of GWAS are typically visualized in a Manhattan plot where the significance of the marker association is plotted against the physical position of the marker on the genome, thus offering a view of all the tested markers across the genome (Tang et al. 2016).

In canola, GWAS has been successfully implemented to identify molecular markers and loci associated with a variety of disease-resistance, stress-tolerance, and agronomic traits (Table 2.6); however, studies focusing on seed storage protein-related traits are lacking. In other major crops such as rice (Chen et al. 2018c) and legumes (Le Signor et al. 2017), GWAS identified markers that underlie variation in storage protein accumulation. Furthermore, markers associated with amino acid content were identified using GWAS in maize (Deng et al. 2017), wheat (Peng et al. 2018), and soybean (Lee et al. 2019), as well as *Arabidopsis* (Angelovici et al. 2013). The lack

Table 2.6. Genome-wide association studies have been successfully implemented in *Brassica napus* L. to identify single nucleotide polymorphism (SNP) markers associated with various disease resistance, morphological/phenological, physiological, and seed quality traits. Studies listed in this table were performed using populations of various sizes and genotypic data acquired using the Brassica 60K Illumina Infinium™ SNP genotyping array (Clarke et al., 2016).

Category	Trait	Population size	Reference
Disease Resistance	Blackleg	200	(Gabur et al. 2018)
	Verticillum	200	(Gabur et al. 2020)
	Clubroot	472	(Li et al. 2016c)
	Sclerotinia	347	(Wei et al. 2016)
	Sclerotinia	448	(Wu et al. 2016)
Morphology/Phenology	Branching	520	(Lu et al. 2017)
	Branching	327	(He et al. 2017)
	Branching	472	(Li et al. 2016a)
	Branching	472	(Li et al. 2017a)
	Branching	143	(Liu et al. 2016b)
	Branching	530	(Sun et al. 2016a)
	Branching	333	(Zheng et al. 2017)
	Flowering	1425	(Jan et al. 2019)
	Flowering	140	(Schiessl et al. 2015)
	Flowering	448	(Wang et al. 2016b)
	Flowering	523	(Xu et al. 2015)
	Germination	218	(Hatzig et al. 2015)

Table 2.6 continued

Category	Trait	Population	Reference
Morphology/Phenology	Hypocotyl elongation	210	(Luo et al. 2017)
	Ovule number	203	(Khan et al. 2019)
	Plant height	203	(Qian et al. 2016b)
	Plant height	296	(Werner et al. 2018)
	Root Structure	520	(He et al. 2019)
	Seed coat colour	521	(Wang et al. 2017)
	Seed size	520	(Khan et al. 2019)
	Seed size	472	(Lu et al. 2017)
	Seed size	143	(Li et al. 2014)
	Shattering	331	(Liu et al. 2016a)
	Silique number	520	(Li et al. 2020)
	Silique number	520	(Lu et al. 2017)
Physiology	Cadmium toxicity	472	(Zhang et al. 2018)
	Cadmium toxicity	419	(Chen et al. 2018b)
	Chlorophyll content	203	(Qian et al. 2016a)
	Chlorophyll content	203	(Qian et al. 2016b)
	Drought stress	66	(Zhang et al. 2015)
	Flooding	520	(Wang et al. 2020)
	Lead toxicity	572	(Zhang et al. 2020)
	Salt stress tolerance	368	(Wan et al. 2017a)
	Salt stress tolerance	214	(Wan et al. 2018)

Table 2.6 continued

Category	Trait	Population	Reference
Yield/Quality	Erucic acid content	215	(Hatzig et al. 2018)
	Erucic acid content	203	(Qian et al. 2014)
	Erucic acid content	472	(Li et al. 2014)
	Fatty acid composition	435	(Qu et al. 2017)
	Fatty acid composition	520	(Tang et al. 2019)
	Fatty acid content	370	(Guan et al. 2019)
	Fibre content	520	(Wang et al. 2015)
	Glucosinolate	521	(Liu et al. 2020)
	Glucosinolate	203	(Qian et al. 2014)
	Glucosinolate	520	(Qu et al. 2015)
	Glucosinolate	1425	(Jan et al. 2019)
	Glucosinolate	203	(Werner et al. 2018)
	Glucosinolate	203	(Qian et al. 2016a)
	Glucosinolates	215	(Hatzig et al. 2018)
	Harvest index	520	(Lu et al. 2016)
	Harvest index	155	(Luo et al. 2015)
	Oil content	472	(Li et al. 2014)
	Oil content	521	(Liu et al. 2016c)
	Oil content	105	(Sun et al. 2016b)

Table 2.6 continued

Category	Trait	Population	Reference
Yield/Quality	Oil content	370	(Tang et al. 2019)
	Oil content	521	(Liu et al. 2016c)
	Oil content	203	(Werner et al. 2018)
	Oil content	203	(Qian et al. 2016b)
	Protein content	370	(Tang et al. 2019)

of similar association studies in canola focused on seed storage protein traits represents a void in the literature that warrants investigation.

2.15.2 Transcriptomics

Transcriptomics aim to observe all transcripts that are present in a tissue at a given time point in order to better understand how gene expression correlates with biological processes (Wang et al. 2009). In plant biology, transcriptomic studies are particularly helpful for understanding phenomena such as how plants initially react to stresses and how equilibrium is achieved in complex biochemical pools. From a plant breeding perspective, transcriptomics can be used to monitor changes in gene expression and help to elucidate potential transcriptional mechanisms that confer desirable phenotypes. The combination of genomics and transcriptomics results in a robust platform to detect genomic regions linked to desirable agronomic traits (Li et al. 2019).

2.15.2.1 Quantitative Real-time PCR

Early transcriptomic work in plant biology relied on quantitative real-time PCR (qPCR) to quantify gene expression in plant tissues (Udvardi et al. 2008; Abdallah and Bauer 2016). Total RNA is first extracted from tissue samples that have been flash frozen to preserve the integrity of the transcripts (Udvardi et al. 2008; Abdallah and Bauer 2016). The RNA is reverse transcribed into cDNA which is subsequently used as template for PCR with gene specific primers (Bustin et al. 2009; Abdallah and Bauer 2016). Transcript levels are normalized against reference genes and gene expression is then statistically assessed (Bustin et al. 2009; Abdallah and Bauer 2016). To ensure the reproducibility and consistency of qPCR experiments, a Minimum Information for Publication of Quantitative Real-Time PCR Experiments guideline was adopted for clinical settings (Bustin et al. 2009) and similar guidelines have been suggested for plant biology (Udvardi et al. 2008).

2.15.2.2 RNA-sequencing

Despite its sensitivity, qPCR is a low throughput technique and the evaluation of multiple transcript targets quickly becomes cumbersome. The subsequent development of hybridization-based microarrays enabled the simultaneous assessment of multiple transcript targets, greatly improving throughput; however, such assays were inherently biased as only known transcripts could be assessed. RNA-sequencing (RNA-seq) is able to overcome the need for prior knowledge while simultaneously offering even higher throughput than that of microarrays (Wang et al. 2009). Sequencing the entire transcriptome enables the detection of rare transcripts as well as transcripts of genes not previously known to affect the biological processes under study (Wang et al. 2009). The technological advancements of RNA-seq have recently been reviewed by Stark *et al.* (2019). Briefly, RNA is converted to cDNA which is subsequently processed into libraries that are sequenced using next generation sequencing platforms (Stark et al. 2019). Sequencing data is typically received as a collection of reads (fragments) which are first mapped to a reference genome or transcriptome; downstream data analyses can then be conducted to interrogate differentially expressed genes, novel transcripts, and other expression phenomena (Van den Berge et al. 2019). Methods for expression analysis using RNA-seq have been comprehensively reviewed by Van den Berge et al. (2019).

2.15.3 Proteomics

Proteomics allows breeders to study the composition and relative abundance of proteins that are translated at any given point in time (Cox and Mann 2011). By directly studying gene products, the breeder can then identify target genes for selection. Of the three omics technologies available, proteomics requires the most technical expertise and thus its use in breeding is currently

limited; however, various groups are working to develop this technique as a tool for crop improvement (Eldakak et al. 2013; Rödiger and Baginsky 2018).

2.16 Marking of a new era of crop improvement with genome editing

Early efforts in the genetic manipulation of SSP profiles relied primarily on the use of antisense technology (Kohno-Murase et al. 1994, 1995; Goossens et al. 1999; Huang et al. 2004) and RNA interference (Schmidt et al. 2011; Kim et al. 2013a, 2014; Altenbach et al. 2014; Hegedus et al. 2014; Wu and Messing 2014). These techniques relied on the degradation of gene transcripts to reduce the accumulation of select SSP. Advancements in genome editing technologies (succinctly reviewed by Langner et al. (2018)), namely the Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)system (Jinek et al. 2012) and it derivatives, now enable the modification of SSP at the genetic level by inducing deleterious mutations in the associated genes. Briefly, the CRISPR/Cas9 system relies on a short single guide RNA (sgRNA) complex to guide the Cas9 endonuclease to the genomic region to which the former is complementary; after a double-stranded break is induced at the target site by Cas9, the host's innate DNA repair mechanism repairs the break in an error-prone fashion, resulting in deleterious mutations that effectively silence the gene (Jinek et al. 2012).

Initially applied to diploid animal and plant systems, CRISPR/Cas9 technology has also proven to be highly effective in targeting multiple homeologous alleles in polyploid *B. napus* (Yang et al. 2017). Successful editing of the *ALCATRAZ* gene to improve shattering resistance (Braatz et al. 2017), the*BnWRKY11* and *BnWRKY70* transcription factors to improve *Sclerotinia sclerotiorum* resistance (Sun et al. 2018), the *Fatty Acid Desaturase 2* gene to modify fatty acid profile (Okuzaki et al. 2018) and other gene targets (Hu et al. 2018; Li et al. 2018) in *B. napus* have recently been reported. Given the ease of use of the CRISPR/Cas9 system and its numerous successful

applications in *B. napus*, genome editing technology has potential to be used as a tool to modify the SSP profile of canola.

2.17 Regulatory consideration for genome editing

Concomitant with technological advances are ethical and regulatory hurdles that must be overcome before novel techniques can be widely deployed. Of primary concern is the presence of transgenes in the resultant plant. Keeping in mind that the mutation that arises from genome editing segregates independently of the editing vector, the subsequent removal of the vector through classical breeding results in a plant with the desired phenotype yet lacks any trace of foreign genetic elements (Gao et al. 2016b; Chen et al. 2019b). Methods to ensure the absence of foreign genetic elements in genome-edited plants arose with improvements to the delivery of genome editing systems. The use of pre-assembled Cas9/sgRNA ribonucleoprotein complexes allows genome editing to occur without the use of foreign genetic elements and has the potential added advantage of limiting spurious editing due to the short persistence of the endonuclease (Kim et al. 2015; Murovec et al. 2018; Metje-Sprink et al. 2019). Similarly, genome editing without stable introgression of the editing vectors may be possible by delivering genetic material encoding the editing system to the plant cell (Demirer et al. 2019), or more specifically to the chloroplast, for transient expression; upon the loss of endonuclease enzyme, edited cells can be regenerated into full plants.

A genetically modified plant is defined by the United States Department of Agriculture (USDA) as a plant that contains genetic material from plant pests (Waltz 2018) or is produced by means of a plant pest such as *Agrobacterium tumefaciens* (Ledford 2013). In the early 2010s the USDA indicated that plant varieties developed using early genome editing techniques would not be regulated as a plant pest (Ledford 2013). Recognizing that advances in genetic engineering

technology were capable of generating novel plants that circumvented its criteria for triggering a regulatory review, the USDA was mandated by the Office of Science and Technology Policy to modernize its regulatory system to encompass plants that were modified using new genome editing technologies (Waltz 2015). To date, at least five CRISPR/Cas9-edited plants in various stages of commercialization, including the oilseed *Camelina sativa* (L.) Crantz modified to increase seed oil content, have been deemed by the USDA to not require regulatory review (Waltz 2018).

Contrary to the view of the USDA towards genome edited plants, the European Union has maintained a current view of novel plants, subjecting them to regulatory approval required for conventional genetically modified organisms (Callaway 2018). Under European law, organisms modified by mutagenesis are not considered genetically modified; although genome-edited plants that are devoid of foreign genetic material are genotypically indistinguishable from a mutagenized plant. The Court of Justice of the European Union (CJEU) contended that genome editing technologies were unconventional and lacked a long safety record, thus, subjecting the resultant plants for regulation (Callaway 2018).

Distinct from the USDA and the CJEU whereby plants are regulated based on the process from which they are generated, the Canadian Food Inspection Agency (CFIA) regulates any plant which contains intentionally introduced traits that are either not found in natural populations in Canada or potentiates the plant to become an environmental hazard, regardless of method used; these plants are collectively termed 'Plants with Novel Traits' (PNT) (CFIA Directive 2009-09, s2.1). Using genome editing to alter levels of individual SSP should circumvent PNT regulation as existing traits are being improved and no additional traits have been introduced. However, PNT designation is also extended to plants with improved quantitative traits that are beyond the range currently observed in domestically cultivated populations (CFIA directive 2009-09, s 2.3.1). To date, data

has only been collected on total protein content of Canadian canola harvests (Barthet 2018a) and no comprehensive surveys have been conducted to establish existing ranges of individual SSP; thus, the improvement of individual SSP that do not alter the overall protein content of canola may not trigger PNT regulation as the upper limit of the natural variation of individual SSP is unknown.

Despite the increasing commercial interest in canola protein, breeding efforts towards improving canola meal protein quality is lacking. It is possible that the manipulation of cruciferin content can improve the value of canola protein. Cruciferin is a logical target for improvement given its potential commercial applications and that it accounts for more than half of the seed storage protein pool (Crouch and Sussex 1981). We hypothesize that an understanding of the genetic variation that exists in cruciferin content in Canadian spring *B. napus* germplasm will facilitate preliminary breeding efforts; furthermore, investigations into the mechanisms that underlie the differential accumulation of cruciferin can uncover molecular markers that can be applied in marker-assisted selection for the improvement of cruciferin through conventional breeding.

Chapter 3. Development and optimization of a Western blot protocol to phenotype cruciferin content in spring *Brassica napus* L.

3.1 Abstract

Defatted *Brassica napus* L. meal is a candidate source of renewable plant protein. Specifically, the seed storage protein cruciferin, has potential as a food ingredient given its desirable functional properties. The development and adoption of cruciferin as a food ingredient is hindered by the need to separate it from the crude meal protein. High-cruciferin *B. napus* can circumvent this limitation by rendering the separation process more efficient, thus making cruciferin economically competitive. To facilitate the improvement of cruciferin content through conventional breeding, a phenotyping protocol with high precision and reasonable throughput is required. To this end, we have raised a custom anti-cruciferin antibody and developed a Western blotting protocol for cruciferin quantification. The antibody was specific to the α chain of cruciferin in *B. napus* and cross reacted with cruciferin homologs in *B. carinata*, *B. juncea,* and *B. rapa*. The Western blotting protocol was optimized by determining optimal dilutions of primary and secondary antibodies, respectively, and determining optimal sample loading. The viability of using Western blotting as a phenotyping tool in plant breeding is discussed.

3.2 Introduction

Changes in consumer demand for animal protein (Melina et al. 2016; Poore and Nemecek 2018) in conjunction with the challenge of producing sufficient protein for the global population (Shepon et al. 2018; Gerten et al. 2020) have led to the need to identify novel sources of plant proteins for human consumption. Protein from the oilseed crop *Brassica napus* L. is a candidate

for development as a novel plant protein given its nutritional quality (Aachary and Thiyam 2012; Wanasundara et al. 2016b) and its abundance (Day 2013).

Protein from *B. napus* (canola protein or rapeseed protein) refers to the protein fraction of the seed; this crude protein mixture is contained in the seed meal after *B. napus* seed is crushed and the oil extracted (Mupondwa et al. 2018). Within crude *B. napus* protein, the seed storage protein cruciferin accounts for more than half the soluble seed protein (Crouch and Sussex 1981).

Cruciferin is a hexameric, globulin-like protein that is composed of various permutations of five different disulfide-bonded subunits (Rödin and Rask 1990; Wanasundara 2011). Each subunit consists of a single α and β polypeptide linked by a disulfide bond (Dalgalarrondo et al. 1986; Nietzel et al. 2013). Although regions of homology are interspersed within the peptide sequence (Figure 3.1), variation in both amino acid composition and size is present across the five subunits (Table 3.1).

Cruciferin has potential to be developed as a functional food ingredient given its ability to form stable gels and emulsions (Wu and Muir 2008; Yang et al. 2014). The technical functionality of cruciferin in food processing has been comprehensively reviewed by Wanasundara *et al.* (2016b). However, the adoption of cruciferin for food processing is marred by the need for its purification from the crude seed protein as the presence of another seed storage protein, napin, may render the protein an allergen (Fiocchi et al. 2016) while the presence of secondary metabolites such as glucosinolates and phytic acid may reduce the nutritional value of the protein (McCurdy 1990; Sashidhar et al. 2020). Commercial scale protein purification methods that exploit the physical differences between cruciferin and other components of the crude canola protein extract have been developed to selectively enrich and purify the cruciferin fraction

```
P33525_CRU3    1 MVKVPHLLVATFGVLLVLNGCLARQSLGVPPQLGNACNLDNLDVLQPTETIKSEAGRVEYWDHNNPQIRCAGVSVSRVII
P33522_CRU4    1 -MGPTSLLSFFFTFLTLFHGFTA-------QQWPNECQLDQLNALEPSQIIKSEGGRIEVWDHHAPQLRCSGFAFERFVI
P33524_CRU2    1 MARLSSLLYFSITVLIFLHGSTA-------QQFPNECQLDQLNALEPSHVLKAEAGRIEVWDHHAPQLRCSGVSFVRYII
P11090_CRUA    1 MARLSSLLSFSLALLTFLHGSTA-------QQFPNECQLDQLNALEPSHVLKAEAGRIEVWDHHAPQLRCSGVSFVRYII
P33523_CRU1    1 MARLSSLLSFSLALLIFLHGSTA-------QQFPNECQLDQLNALEPSHVLKAEAGRIEVWDHHAPQLRCSGVSFVRYII

P33525_CRU3   81 EQGGLYLPTFFSSPKISYVVQGMGISGRVVPGCAETFMDSQPMQGQQQGQPWQGQQGQQGQQGQQ---GQQGQQGQQGQ-
P33522_CRU4   73 EPQGLYLPTFLNAGKLTFVVHGHALMGKVTPGCAETFNDSPVFGQ-GQGQEQG---------------QG---------
P33524_CRU2   74 ESQGLYLPSFLNTANVSFVAKGQGLMGRVVPGCAETFQDSSVFQP-GSGSPFGEGQGQ-GQQGQGQGQQGQGKGQQGQG
P11090_CRUA   74 ESKGLYLPSFFSTARLSFVAKGEGLMGRVV-LCAETFQDSSVFQP-SGGSPFGEGQGQ-GQQGQGQGH-QGQGQGQQG--
P33523_CRU1   74 ESKGLYLPSFFSTAKLSFVAKGEGLMGRVVPGCAETFQDSSVFQP-SGGSPSGEGQGQ-GQQGQGQGH-QGQGQGQQG--

P33525_CRU3  157 QGQQGQQGQQGQQQQGFRDMHQKVEHVRGDIIAITAGSSHWIYNTGDQPLVIICLLDIANYQNQLDRNPRTFRLAGNNP
P33522_CRU4  127 --------QGQGQQQGFRDMHQKVEHLRGDTIATPPGVAQWFYNNGNEPLILVAAADIANNLNQLDRNLRPFLLAGNNP
P33524_CRU2  152 KGQQGQSQGQQGQQGFRDMHQKVEHIRGDTIATHPGVAQWFYNNGNQPLVIVAVMDLASHQNQLDRNPSQFYLAGKNP
P11090_CRUA  148 ---QQGQQGQQSQQGFRDMHQKVEHIRGDTIATHPGVAQWFYNDGNQPLVIVSVLDLASHQNQLDRNPRPFYLAGNNP
P33523_CRU1  149 ---QQGQQGQQSQQGFRDMHQKVEHIRGDTIATHPGVAQWFYNDGNQPLVIVSVLDLASHQNQLDRNPRPFYLAGNNP

P33525_CRU3  237 QGGSQ---QQQQQQNMLSGFDPQVLAQALKIDVRLAQELQNQQDSRGNIVRVKGPFQVVRPPLRQPYESEQWRHPRGPP
P33522_CRU4  199 QGQQWLQGRQQQKQNNIFNGFAPQILAQAFKISVETAQKLQNQQVNRGNIVKVQGQFGVIRPPLRQGQ--------GGQQ
P33524_CRU2  232 QGQSWLHGRGQQPQNNILNGFSPEVLAQAFKIDVRTAQQLQNQQDNRGNIVRVQGPFGVIRPPLKSQR-----------P
P11090_CRUA  225 QGQVWIEGREQQPQKNILNGFTPEVLAKAFKIDVRTAQQLQNQQDNRGNIIRVQGPFSVIRPPLRSQR-----------P
P33523_CRU1  226 QGQVWIEGREQQPQKNILNGFTPEVLAKAFKIDVRTAQQLQNQQDNRGNIIRVQGPFSVIRPPLRSQR-----------P

P33525_CRU3  314 QSPQDNGLEETICSMRTHENIDDPARADVYKPNLGRVTSVNSYTLPILQYIRLSATRGILQGNAMVLPKYNMNANEILYC
P33522_CRU4  271 PQEEGNGLEETLCTMRCTENLDDPSSADVYKPSLGYISTLNSYNLPILRFLRLSALRGSIHNNAMVLPQWNVNANAALYV
P33524_CRU2  301 QETEANGLEETICSARCTDNLDDPSNADVYKPQLGYISILNSYDLPILRVLRLSALRGSIRQNAMVLPQWKSKSNAVLYV
P11090_CRUA  294 QE-EVNGLEETICSARCTDNLDDPSNADVYKPQLGYISTLNSYDLPILRFLRLSALRGSIRQNAMVLPQWNANANAVLYV
P33523_CRU1  295 QETEVNGLEETICSARCTDNLDDPSNADVYKPQLGYISTLNSYDLPILRFLRLSALRGSIRQNAMVLPQWNANANAVLYV

P33525_CRU3  394 TQGQARIQVVNDNGQNVLDQQVQKGQLVVIPQGFAYVVQSHQNNFEWISFKTNANAMVSTLAGRTSALRALPLEVITNAF
P33522_CRU4  351 TKGKAHIQNVNDNGQRVFDQEISKGQLLVVPQGFAVVKRATSQQFQWIEFKSNDNAQINTLAGRTSVMRGLPLEVISNGY
P33524_CRU2  381 TDGEAQIQVVNDNGDRVFDGQVSQGQLLSIPQGFSVVKRATSDQFRWIEFKTNANAQINTLAGRTSVMRGLPLEVIANGY
P11090_CRUA  373 TDGEAHVQVVNDNGDRVFDGQVSQGQLLSIPQGFSVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVISNGY
P33523_CRU1  375 TDGEAHVQVVNDNGDRVFDGQVSQGQLLSIPQGFSVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVISNGY

P33525_CRU3  474 QISLEEARRIKFNTLETTLTRARGGQPQLIEEIVEA
P33522_CRU4  431 QISPQEARSVKFSTLETTLTQSSGPMGYGMPRVEA-
P33524_CRU2  461 QISLEEARRVKFNTIETTLTHSSGPASYGRPRKADA
P11090_CRUA  453 QISLEEARRVKFNTIETTLTHSSGPASYGGPRKADA
P33523_CRU1  455 QISLEEARRVKFNTIETTLTHSSGPASYGGPRKADA
```

Figure 3.1. Sequence alignment of all five *Brassica napus* L. cruciferin subunits. Curated amino acid sequences were obtained from the UniProt Knowledgebase. The peptide sequence (highlighted in the red box) is within the α-chain of the cruciferin subunit and was used as an antigen for antibody production.

Table 3.1. Size summary of five curated accessions of *Brassica napus* L. cruciferin subunits in UniProt KB. Subunit lengths are given as number of amino acids (AA). The expected molecular mass of each peptide is presented in parentheses under the subunit length and estimated using a factor of 1 amino acid = 0.11 kDa.

Subunit	UniProt Accession	Subunit length (AA)	α-chain length	β-chain length
CRU1	P33525	509 (56.0 kDa)	296 (32.6 kDa)	190 (20.9 kDa)
BnC1	P33523	490 (53.9 kDa)	277 (30.5 kDa)	190 (20.9 kDa)
BnC2	P33524	496 (54.6 kDa)	283 (31.1 kDa)	190 (20.9 kDa)
CRU2/3 (CRUA)	P11090	488 (53.7 kDa)	275 (30.3 kDa)	190 (20.9 kDa)
CRU4	P33522	465 (51.2 kDa)	254 (27.9 kDa)	189 (20.9 kDa)

(Bérot et al. 2005; Mupondwa et al. 2018). These methods typically require specialized chromatography equipment and the use of large quantities of water; they are therefore barriers to the wide-scale adoption of cruciferin for food processing as they pose potential technological, environmental, and cost limitations to food manufacturers.

The development of *B. napus* cultivars with high levels of cruciferin may render the purification process more economical and may even circumvent the need for purification altogether should cruciferin displace napin and other seed proteins. Genetic variation in meal protein functionality exists in *B. napus* (Aluko and McIntosh 2001). Using traditional breeding, this genetic variation is essential to improving cruciferin content, and consequently meal functionality.

Plant breeding efforts require accurate phenotyping platforms with high throughput to breed for a particular trait (Yang et al. 2020). A high throughput seed storage protein quantification method is not currently available in *B. napus*. Seed storage protein composition in crops have traditionally been phenotyped by electrophoretically resolving protein samples and subsequently staining the gel with a protein-binding dye to visualize the protein bands (Liu et al. 2006; Boehm et al. 2018). This method may meet the required throughput, but its accuracy is suboptimal due to the differential staining of protein bands (Tal et al. 1985; Congdon et al. 1993) and the inability to discriminate target bands from other proteins of similar molecular mass. Conversely chromatography methods (Bérot et al. 2005) offer high precision for protein quantification but are labor intensive and fail to provide the throughput required for a breeding program.

Immunological methods are potential phenotyping strategies for storage protein content by offering greater throughput than chromatography and greater accuracy than gel staining. While

studies using immunological detection to localize storage proteins to the subcellular level have been reported (Rödin and Rask 1990; Robin et al. 1991; Nietzel et al. 2013; Bieker et al. 2019), it's use as a phenotyping tool in breeding programs is currently limited.

Currently, a reliable immunoassay to phenotype cruciferin content has yet to be developed. To this end, we have raised a polyclonal antibody against the α-chain of all five cruciferin subunits in *B. napus*. After characterizing the antibody, a standardized Western blotting protocol was developed to allow the phenotyping of cruciferin content in *B. napus* and other Brassica species.

3.3 Materials and methods

3.3.1 Plant material and defatted meal preparation

Seed from *B. napus* 'Westar' was produced in an isolated field in Winnipeg, Manitoba in 2015 and used for both antibody validation and Western blot optimization. Seeds of *B. carinata* A. Braun '194251', *B. napus* L. 'Qing2', *B. juncea* (L.) Czern. 'Cutlass', and *B. rapa* L. 'K880' were produced in the field in Portage-la-Prairie, Manitoba in 2017 and used to assess cross-reactivity of the antibody with cruciferin from other Brassica species.

One gram of seed was crushed in a porcelain mortar and pestle (CoorsTek, Golden, USA) and transferred to a Kimax glass test-tube with a screw cap (DWK Life Science, Rockwood, USA). For defatting, approximately 7 mL of petroleum ether (ACS reagent grade; Sigma-Aldrich, Oakville, Canada) was added to the tube, the cap screwed on, and the tube was placed at an angle on a rotary shaker set to a moderate speed that permitted the crushed seed to continually mix gently with the solvent. The seed was defatted overnight (14 to 16 hrs) at room temperature in the fume hood. The solvent was then replaced with an equal volume of fresh solvent the next morning, and defatting continued for another eight hours. At the end of defatting, the solvent was decanted, the

oil-free meal was transferred to an open 15 mL Falcon tube, rinsed once with petroleum ether, and allowed to dry in the fume hood for 48 hrs. Once the meal was thoroughly dried, the tubes were capped and stored at room temperature.

3.3.2 Total soluble protein extraction and preparation

Total soluble protein (TSP) was extracted by combining 50 mg defatted meal with 1 mL of extraction buffer (50 mM Tris-HCl, pH 7.5, and 50 mM Na_2EDTA) and one copper coated steel 4.5 mm BB pellet (Crosman, Bloomfield, USA) in a 2 mL microcentrifuge tube. The meal was homogenized using a TissueLyser II (Qiagen, Toronto, Canada) for 5 minutes at 20 Hz and then centrifuged (16, 000 xg, 10 min, room temperature). The supernatant was collected as the TSP and held on ice. To assess complete extraction of soluble protein, the meal was repeatedly extracted in the same manner five additional times and each TSP was kept separate. Gel electrophoresis was used to qualitatively assess the presence of protein in the extract. Samples were prepared for electrophoresis by diluting 1:9 with tris-buffered saline (TBS; 10 mM Tris, 150 mM NaCl, pH 7.5) and combined with an equal volume of 2x Laemmli sample buffer containing 100 mM dithiothreitol before incubating at 65°C for ten minutes. Twenty µL of prepared sample were resolved on a 12 % TGX FastCast polyacrylamide gel (Bio-Rad Laboratories, Hercules, U.S.A.) and visualized using a Gel Doc XR+ system (Bio-Rad Laboratories, Hercules, U.S.A.) after applying the manufacturer-recommended five-minute UV activation. Exposure time was set to automatic for detection of low intensity bands. Soluble protein was considered completely extracted when no protein bands were visible in the lane.

3.3.3 Polyclonal anti-cruciferin antibody production

The production of a custom rabbit polyclonal antibody against cruciferin was contracted to GenScript (Piscataway, U.S.A.). The antigen was synthesized at Genscript as a 20 amino acid

peptide (FRDMHQKVEHIR) from a conserved region in the α-chain of the five *B. napus* cruciferin accessions in the UniProt database (https://www.uniprot.org/) (Figure 3.1); the peptide sequence was nested within an antigen peptide previously used by (Nietzel et al., 2013) for the generation of antibodies against the N-terminus of cruciferin subunits. Affinity purified polyclonal was received from Genscript.

3.3.4 Antibody validation

Antibody specificity was assessed by Western blotting following the conditions supplied in the quality control data from Genscript. Soluble protein from four *Brassica* species: *B. carinata* ‘194251’, *B. napus* ‘Qing2’, *B. juncea* ‘Cutlass’, and *B. rapa* L. ‘K880’ was extracted once as described previously. Protein concentration was determined by Bradford assay (Bio-Rad Laboratories, Hercules, U.S.A.) relative to a bovine serum albumin standard following the manufacturer’s protocol for microplate assays. Protein samples were prepared and 1 µg of protein was resolved as described above on duplicate gels.

To observe the protein profile of the samples, one gel was stained in Coomassie Brilliant Blue for 1 hr and de-stained (40% methanol, 10% acetic acid) for an additional hour at room temperature before an image was captured with a flatbed scanner. The second gel was used for Western blotting. Resolved proteins were wet transferred (25 mM Tris, 192 mM glycine, 20% (v/v) methanol; 100 V, 90 minutes, in ice bath) onto a polyvinylidene difluoride (PVDF) membrane (Bio-Rad Laboratories, Hercules, U.S.A.). All subsequent steps were performed at room temperature on an orbital shaker in a covered dish; all wash steps were performed as three, 5-min washes in tris buffered saline with 0.05% Tween-20 (TBST).

After transfer, the blot was washed and blocked for 1 hr in a block solution of 5% (w/v) blotting grade blocker (fat-free skim milk; Bio-Rad Laboratories, Hercules, U.S.A.) in TBST. The blot was washed and probed with horseradish peroxidase (HRP)-conjugated anti-cruciferin antibody at a 1:1000 dilution in block solution for 1 hr at room temperature. The blot was washed again and developed with fridge-cold (4°C) 1-Step™ Ultra TMB-Blotting Solution (Thermo Scientific, Waltham, U.S.A.) and incubated for 2 min at room temperature. The development was quenched and washed twice in deionized water before an image was captured on a gel-doc using Image Lab software version 6.1 (Bio-Rad Laboratories, Hercules, U.S.A.) with the exposure set to automatic detection for low intensity bands. Specificity of the antibody was inferred by visually assessing the captured blot for the presence of a bands that did not correspond to the expected molecular mass of the cruciferin α-chains.

3.3.5 Western blot optimization

Western blot optimization was performed using TSP from *B. napus* 'Westar': protein from defatted meal was repeatedly extracted four times as described previously, and all four fractions were pooled as TSP. To determine optimal antibody dilutions, TSP was serially diluted 2-fold (1/4, 1/8, 1/16, and 1/32-fold) with TBS to provide a range of antigen content and 10 µL were resolved on a 12 % TGX FastCast polyacrylamide gel. The resolved proteins were semi-dry blotted onto a low-absorbance PVDF membrane using a Trans-Blot Turbo system (Bio-Rad Laboratories, Hercules, U.S.A.) with the pre-set protocol for a single TGX gel (2.5 A, 25 V, 3 minutes, room temperature). After transfer, blots were washed in TBST as described above. Multiple blots were prepared in the same manner using the same TSP. To determine the optimal dilution of the primary HRP-conjugated anti-cruciferin antibody, dilutions at 1:1000 (recommended by manufacturer), 1:2000, 1:4000, and 1:8000 in block solution (5% milk in TBST) were evaluated, followed by

direct detection with fridge-cold (4°C) 1-Step™ Ultra TMB-Blotting Solution. The blots were developed as described in Section 3.3.4 and captured with exposure set to automatic detection for high intensity bands to prevent signal saturation.

The use of a secondary antibody for signal amplification was subsequently assessed. To determine the optimal dilution for the secondary antibody, the blot was first probed with either a 1:2000 or 1:4000 dilution of primary anti-cruciferin antibody. After washing, HRP-conjugated monoclonal mouse-anti-rabbit antibody (clone M205; GenScript, Piscataway, U.S.A.) at dilutions of 1:5000 (recommended by manufacturer), 1:8000, and 1:16000 in block solution were evaluated. Blots were then developed and captured as described above.

To determine the linearity of the Western blot protocol, 100 ng to 5 µg TSP were resolved and semi-dry blotted onto PVDF membrane as described previously with one modification: prior to blotting, the gel was UV-activated for 1 min on a gel-doc and an image was captured for downstream total protein normalization using a gel doc with exposure set to automatic detection for low intensity bands. The blot was washed and blocked as described previously before being probed with primary anti-cruciferin antibody at a dilution of 1:4000 and secondary HRP-conjugated mouse-anti-rabbit antibody at a dilution of 1:8000. The blot was developed and captured as described above. To determine the signal intensity of each lane, Image Lab software version 6.1 was used with the default parameters set to: sensitivity, 148.48; scale, 7; noise filter, 4; shoulder 1; and disc size, 3.8 mm. Band intensity was plotted against TSP load using Microsoft Excel 2016 (Microsoft, Redmond, U.S.A.) and fitted against a natural logarithmic function.

3.4 Results

3.4.1 Optimizing extraction of total soluble protein

Soluble protein was repeatedly extracted six times from defatted *B napus* 'Westar' meal and each extract was assessed by gel electrophoresis. Soluble protein resolved to form three prominent groups of bands under reducing conditions at approximately 30 kDa, 20 kDa, and 10 kDa in the first three extracts; however, from fourth extract onward, no protein was visible (Figure 3.2) indicating soluble protein was completely extracted after three extractions. Minor quantities of protein above the 37 kDa marker were observed in the first extract but diminished in subsequent extractions.

3.4.2 Antibody validation

Antibody specificity was assessed by Western blotting using TSP from defatted meal of *B. carinata* '194251', *B. napus* 'Qing2', *B. juncea* 'Cutlass', and *B. rapa* 'K880'. Total soluble protein from all Brassica species resolved to form similar profiles under reducing conditions as observed previously with prominent bands appearing at approximately 30 kDa, 20 kDa, and 10 kDa (Figure 3.3A); a faint band indicating high molecular mass proteins of approximately 75 kDa were observed across all samples. Western blotting with the anti-cruciferin antibody generated signal as a single band between the 25 kDa and 37 kDa markers (approximately 30 kDa) (Figure 3.3B) across all samples corresponding to the expected molecular mass of the cruciferin α-chain. The antibody was therefore specific to cruciferin α-chains in *B. napus* and was able to cross-react with homologs in related Brassica species. When the experiment was repeated using a longer development time, a second band in the same molecular weight bracket was detected indicating the antibody showed affinity towards multiple isoforms of the protein.

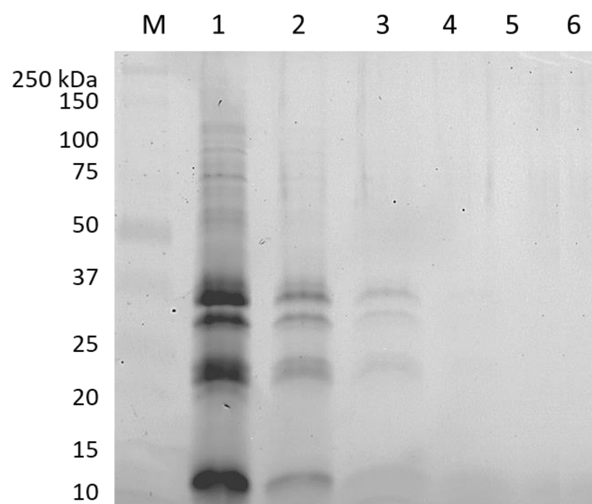

Figure 3.2. Proteins recovered from the sequential extraction of defatted *Brassica napus* L. 'Westar' meal resolved under reducing conditions. Soluble protein was repeatedly extracted from the meal a total of six times. Equal volumes were loaded in each lane. Numbered lanes refer to the ordinal extraction order. The ladder was resolved in the M lane.

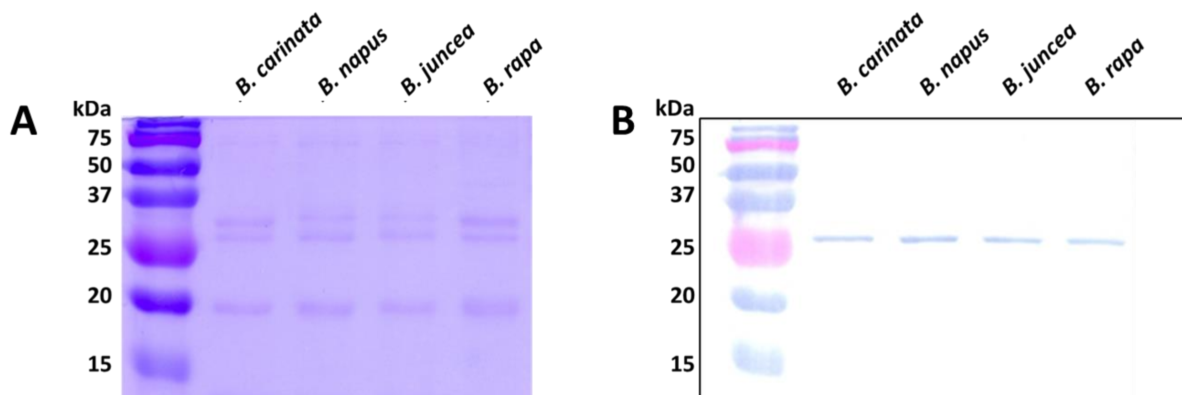

Figure 3.3. Coomassie-staining and Western blotting of soluble seed protein from *Brassica carinata* A. Braun '194251', *B. napus* L. 'Qing2', *B. juncea* (L.) Czern. 'Cutlass', and *B. rapa* L. 'K880.' One microgram of protein was resolved under reducing conditions in each lane on duplicate gels. Resolved proteins were visualized by staining with Coomassie Brilliant Blue (A). Western blotting was performed using a custom anti-cruciferin antibody and directly colorimetric detection (B).

3.4.3 Western blot optimization

The optimal dilution of primary antibody was determined to identify the largest dilution that would generate a clear signal for Western blotting. Direct detection with the primary antibody diluted at 1:1000, 1:2000, and 1:4000 yielded identical results with signal visible across all four TSP loads between the 25 kDa and 37 kDa markers (Figure 3.4A), while the 1:8000 dilution failed to generate signal at the two lowest loads; therefore a 1:4000 dilution of primary antibody was considered optimal.

The use of secondary antibody amplified signal intensity of the bands regardless of primary antibody dilution at 1:2000 or 1:4000 fold. A second protein band also between the 25 kDa and 37 kDa markers was observed across all samples on all blots indicating the presence of cruciferin α-chains from different cruciferin subunits (Figure 3.4B). Signal intensity appeared identical when secondary antibody was used at dilutions of 1:5000 and 1:8000, respectively, and two bands were clearly visible on the blot. Conversely, when secondary antibody was used at a 1:16000 dilution, only a single band was visible on the blot, effectively underestimating the quantity of cruciferin present; therefore, a dilution of 1:8000 secondary antibody was deemed optimal.

To determine the linear range of the Western blotting protocol for cruciferin quantification, varying quantities of TSP from 100 ng to 5 μg were tested. For each protein load, two bands between the 25 kDa and 37 kDa molecular markers were observed corresponding to different isoforms of the cruciferin α chain (Figure 3.5A). When more than 1 μg of TSP was blotted, faint bands between the 50 kDa and 70 kDa molecular markers were detected corresponding to procruciferin polypeptides (Figure 3.5B). Antibody signal plotted as a function of TSP load fitted a natural logarithmic function ($r^2 = 0.992$) with a linear phase between 100 ng and 400 ng TSP (Figure 3.5C).

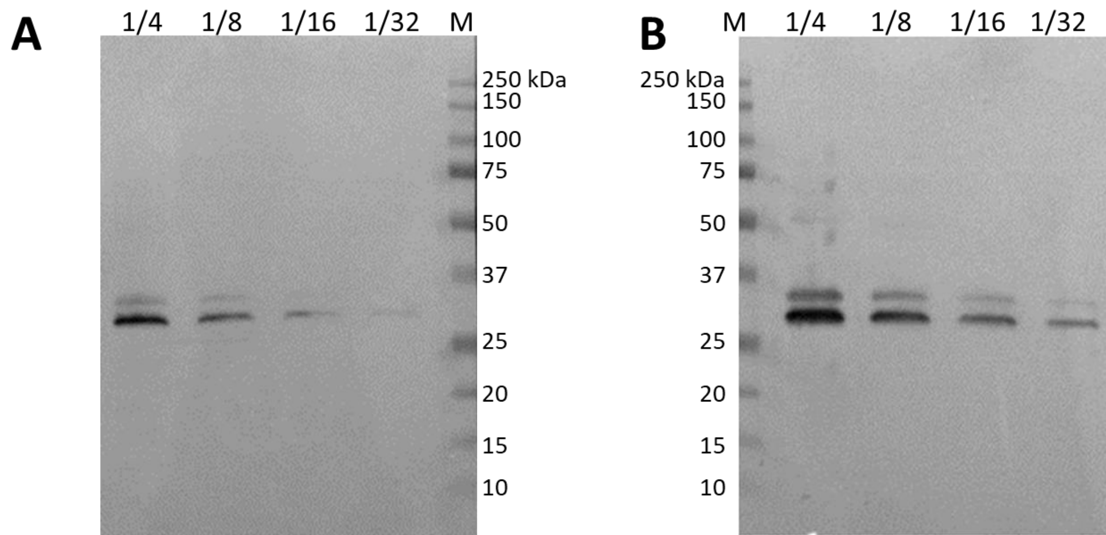

Figure 3.4. Western blot detection of cruciferin in soluble protein from defatted *Brassica napus* L. 'Westar' meal. Soluble protein was extracted, and a serial 2-fold dilution series was prepared using extraction buffer; equal volumes of each dilution were resolved in each lane. Colorimetric detection of cruciferin was performed using an anti-cruciferin antibody (1:4000 dilution) directly (A) or indirectly with a secondary antibody (1:8000 dilution) (B). Lanes are labelled by serial dilution factor. Lane M indicated ladder.

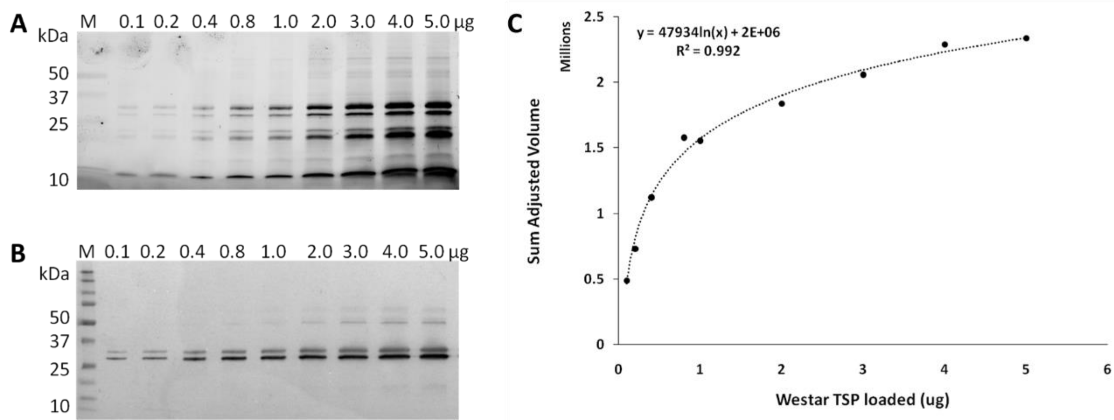

Figure 3.5. Optimization of protein loading for cruciferin quantification by Western blotting. Varying quantities of soluble protein from the defatted meal of *Brassica napus* L. 'Westar' were resolved under reducing conditions (A). Resolved proteins were blotted and cruciferin content was determined by colorimetric detection using a custom anti-cruciferin antibody (B). Signal intensity of each band was plotted against protein load and a natural logarithm regression was fitted (C).

3.5 Discussion

The development of high-cruciferin *B. napus* cultivars has the potential to accelerate the adoption of cruciferin as a food ingredient and establish a new market class of *B. napus* with specialty meal protein; however, these efforts are currently hindered by the lack of a suitable phenotyping protocol to quantify cruciferin. While storage proteins have previously been phenotyped by Coomassie staining of resolved proteins (Schatzki et al. 2014) and chromatography (Bérot et al. 2005), the precision and throughput of these methods, respectively, may not meet the needs of breeding programs. Immunological methods, however, offer a compromise between precision and throughput.

The α and β chains of the cruciferin subunits are synthesized as a single ca. 50 kDa procruciferin polypeptide (Ereken-Tumer et al. 1982), which undergoes processing to form the individual ca. 30 kDa α and ca. 20 kDa β chains linked together by a disulfide bond (Dalgalarrondo et al. 1986; Rödin and Rask 1990). Therefore, the α and β chains exist in an equal stoichiometric ratio and the quantity of either chain is proportional to total cruciferin content. In using the α chain as the antigen, the antibody should cross-react with native cruciferin as the α chain faces the exterior of the mature protein (Robin et al. 1991; Job et al. 2005). Preliminary enzyme-linked immunosorbent assay experiments indicated the antibody was functional against native cruciferin; therefore, immunological methods using the antibody are not limited by the conformation of cruciferin under study. Cross-reactivity of the antibody to cruciferin from *B. carinata, B. juncea*, and *B. rapa* was expected given their sequence homology and suggests immunological methods developed using this reagent may be directly translatable to other Brassica species.

In breeding programs, phenotyping assays must be easy to use and affordable for large populations and this was taken into consideration in developing this Western protocol.

Colorimetric detection was chosen as it is an economical alternative to chemiluminescent and fluorescent detection by not requiring film or detection equipment. Of the three methods, colorimetric blotting also requires arguably less technical skill to perform (Bass et al. 2017). Although colorimetric detection is not as sensitive as chemiluminescent detection (Constantine et al. 1994), cruciferin accounts for more than half the soluble seed meal protein (Crouch and Sussex 1981) and its abundance abates the need for a low detection limit.

Breeding programs often require phenotyping hundreds or thousands of samples; with a throughput of eight samples on a typical membrane, multiple researchers running multiple blots in parallel may be necessary to meet efficiency demands. Western blotting may therefore not be economically feasible for small breeding programs. Alternative quantitative immunoassays such as enzyme immunoassays, capillary electrophoresis, and immune-PCR (Chang et al., 2016) can be considered in the future. These plate-based assays enable 24 genotypes along with a standard curve to be assayed in triplicate on each 96-well plate, effectively tripling the 8-sample output of a single Western blot. Although Western blotting for the quantification of cruciferin may be difficult to implement in large breeding populations, the protocol nonetheless represents a useful tool to evaluate comparatively smaller populations such as germplasm collections and to confirm enzyme immunoassays results.

Following the successful development of the polyclonal anti-cruciferin antibody, anti-napin antibodies can be raised in the future to quantify napin, the second most abundant seed storage protein in *B. napus* (Crouch and Sussex, 1981). Despite its potential allergenicity (Moreno and Clemente, 2008), napin is functionally capable of replacing egg albumin (Wanasundara, McIntosh, et al., 2016) and milk casein (Schwartz et al., 2015) in food processing; thus, its improvement is a potential breeding focus in the future. Finally, monoclonal antibodies against

cruciferin and napin can be raised to establish a validated and standardized reagent for seed storage protein research in *B. napus*.

3.6 Conclusion

A Western blotting protocol using a custom anti-cruciferin antibody was developed as a phenotyping tool to quantify cruciferin content in plant breeding programs. The antibody has confirmed specificity to the α-chains of cruciferin in *B. napus* but was able to cross-react with cruciferin from related Brassica species. Antibody dilutions were optimized to maximize efficiency and sample load was optimized to ensure linearity of the assay. This protocol represents the first step in phenotyping cruciferin content on a scale suitable for breeding programs.

The previous chapter described the characterization of a new anti-cruciferin antibody and the development of a robust immunological phenotyping method to quantify cruciferin content within the soluble protein of *B. napus* seeds. Together these resources will be employed in the next chapter to evaluate cruciferin content across a diverse population of *B. napus* to explore the genetic variation of the trait within the species.

Chapter 4. Assessing genetic diversity of cruciferin content in spring *Brassica napus* L.

4.1 Abstract

The seed storage protein cruciferin is the most abundant protein in canola (*Brassica napus* L.) seed and has functional properties that make it an ideal candidate for food processing. Increasing cruciferin content in canola can improve the value of the crop by allowing the meal to be repurposed for food processing after the oil is extracted. To establish a breeding program focused on improving cruciferin content, knowledge of the genetic diversity of cruciferin content in spring *B. napus* is required. To this end, cruciferin content was phenotyped in a population of historic cultivars spanning the breeding history of *B. napus* under greenhouse conditions by Western blotting. Genotype had a significant effect on cruciferin content indicating genetic diversity for the trait was present in the population. Correlation analysis found cruciferin content to be positively correlated with protein and glucosinolate contents and negatively correlated with oil content. The implications of genetic variation in cruciferin content and the suitability of phenotyping under greenhouse conditions are discussed.

4.2 Introduction

Plant proteins have been recommended as a sustainable alternative to animal proteins to meet our dietary protein needs (Aachary and Thiyam 2012; Wanasundara et al. 2016a; Gerten et al. 2020). In the human diet, seeds serve as the primary source of dietary plant protein (Day 2013). Mature seeds store nitrogen as storage proteins, which are mobilized during germination to support the growth of the seedling (Shewry et al. 1995; Job et al. 2005).

The seed storage protein cruciferin in *Brassica napus* L. is a large hexameric protein that constitutes approximately 60% of the total seed protein content (Crouch and Sussex 1981). Five different subunits of cruciferin have been documented in the UniProt database (https://www.uniprot.org/) and mature cruciferin consists of varying combinations of the five subunits (Withana-Gamage et al. 2013). Each subunit of cruciferin consists of an α and β chain which are linked together via a disulfide bond (Dalgalarrondo et al. 1986; Rödin and Rask 1990); three subunits assemble into trimers, and subsequently two trimers form the mature hexameric protein through stabilization by non-covalent interactions (Shewry et al. 1995; Wanasundara 2011). The reduction of cruciferin results in a pool of peptides approximately 30 kDa representing various forms of the α chains and a pool approximating 20 kDa corresponding to the β chains (Rödin and Rask 1990).

Despite its nutritional (Bos et al. 2007; Fleddermann et al. 2013) and functional (Salleh et al. 2002; Wu and Muir 2008; Yang et al. 2014) similarities with soy protein, canola protein in general, is not prevalent in the Canadian marketplace with only one cruciferin-rich isolate product readily available (https://www.burcon.ca/products/canola-proteins/). The presence of anti-nutritive compounds and the storage protein napin in crude canola protein isolates reduce both the nutritional and functional properties of cruciferin (Gilani et al. 2005). Furthermore, napin has been suggested to be an allergen (Moreno and Clemente 2008; Fiocchi et al. 2016) and its presence may limit the use of canola protein products. Thus, the removal of these compounds is required before cruciferin can be adopted as an ingredient in food processing.

Various methods exploiting the physical differences between cruciferin and other components of the crude canola protein extract have been developed to selectively enrich and purify the cruciferin fraction (Bérot et al. 2005; Mupondwa et al. 2018). These methods typically

require specialized chromatography equipment and the use of large quantities of water; they are therefore barriers to the wide-scale adoption of cruciferin for food processing as they pose potential technological, environmental, and cost limitations to food manufacturers.

Consequently, canola cultivars with enriched levels of cruciferin can reduce the need for protein purification and facilitate its use as a functional protein ingredient. In soybean, cultivars with specialized functional protein profiles have been developed through traditional breeding for tofu production (Poysa et al. 2006; Zarkadas et al. 2007; Boehm et al. 2018). Similarly, in canola the development of genotypes with altered fatty acid profiles (Scarth et al. 1988) demonstrates the amenability of the nutritional content and functional properties of the canola seed to improvement through biotechnology and conventional breeding (Scarth and Tang 2006; Hannoufa et al. 2014). To date, no canola cultivars with special protein functionality have been developed and thus, the breeding of high-cruciferin cultivars is required prior to the establishment of a novel market class in the canola industry.

Genetic diversity is a prerequisite for plant breeding as it enables genetic recombinants with superior phenotypes (Fu 2015; Henderson and Salt 2017; Louwaars 2018). Classical quantitative genetic theory estimates phenotypic variation to be the sum of the genetic variation, environmental variation, and their interaction (Bernardo 2020); when environmental variation is minimized, the observed phenotypic variation can approximate genetic variation. To initiate a breeding program for cruciferin content improvement, the genetic variation for cruciferin content must first be assessed. To date, such studies have yet to be conducted in spring *B. napus*, thus, constituting a knowledge gap in the field. Therefore, the objective of this research is to assess the genetic variation of cruciferin content in spring *B. napus* within the University of Manitoba Brassica Germplasm Collection. The identification of high and low cruciferin genotypes will

enable breeding efforts towards the improvement of cruciferin content to commence in spring *B. napus* and will allow mapping populations to be generated for further genetic studies on cruciferin content.

4.3 Materials and Methods

4.3.1 Plant material

A population of 44 genotypes of spring *B. napus* from the University of Manitoba Brassica Germplasm Collection (Table 4.1) were evaluated in this research. The population was grown and self-pollinated for three cycles prior to the start of the experiment. Seed from the last cycle of self-pollination was used to grow the plant material for this study.

4.3.2 Plant growth parameters and experimental design

Growth stages are reported using the BBCH code (Lancashire et al. 1991). Plants were seeded and grown in Sunshine #4 potting mix (Sun Gro Horticulture, Vassar, Canada) for the duration of the experiment. Seeds were planted into standard 812 inserts in a growth room in the Department of Plant Science, University of Manitoba set to 22 °C day/18 °C night with a 16- hr photoperiod. One week after germination, approximately growth stage 10 – 11 on the BBCH scale (Lancashire et al. 1991), seedlings were fertilized with 20-20-20 fertilizer (Master Plant-Prod Inc., Brampton, Canada) at a rate of approximately 15 mL per 3.89 L of water. Two-week old seedlings, approximately growth stage 11 – 12, were transferred to the Crop Technology Centre greenhouse, University of Manitoba, and transplanted into six-inch standard pots. Greenhouse condition were set to 22°C day/18°C night with a 16-hr photoperiod. Plants were arranged in a randomized complete block design consisting of three blocks. To ensure sufficient seed yield for downstream analyses, three plants of each genotype were grown in each block and seeds from the three plants were bulked at harvest. The experiment was replicated a second time at a separate location

Table 4.1. Details of the spring canola and high erucic acid (HEAR) *Brassica napus* L. genotypes used to assess genetic diversity in cruciferin content. Registration numbers are from the Canadian Food Inspection Agency. N.A. indicated data not available.

Genotype	Year of registration	Registration number	Oil quality	Country of origin	Cultivar description/Reference
04R2026	2004	N.A.	HEAR	Canada	N.A.
46A65	1996	4288	Canola	Canada	N.A.
AC Excel	1990	3601	Canola	Canada	Rakow (1993) Can. J. Plant Sci. 73: 183-184
Allons	1994	3996	Canola	Canada	Scarth et al. (1997) Can. J. Plant Sci. 77: 125–126
Apollo	1993	3694	Canola	Canada	Scarth et al. (1995) Can. J. Plant Sci. 75:203-204
Castor	1997	4581	HEAR	Canada	McVetty et al. (1997) Can. J. Plant Sci. 78: 305–306
Cheetah	1990	N.A.	Canola	Canada	N.A.
Defender	1994	3911	Canola	Sweden	N.A.

Table 4.1 continued

Genotype	Year of registration	Registration number	Oil quality	Country of origin	Cultivar description/Reference
Ebony	1994	3936	Canola	Canada	N.A.
Hanna	1985	2524	Canola	Sweden	Akbar (1989) Hereditas. 111: 247-253
Hero	1989	3129	HEAR	Canada	Scarth et al. (1991) Can. J. Plant Sci. 71: 865-866
Hi-Q	1999	4898	Canola	Canada	Stringham et al. (2000) Can. J. Plant Sci. 80:835–836
Industry	1992	N.A.	HEAR	Denmark	N.A.
Karat	1978	N.A.	Canola	Sweden	Rakow (1993) Can. J. Plant Sci. 73: 181-182
Legacy	1993	3736	Canola	Sweden	N.A.
LG3310	1995	4087	Canola	Canada	N.A.
LG3333	1998	4724	Canola	Canada	N.A.

Table 4.1 continued

Genotype	Year of registration	Registration number	Oil quality	Country of origin	Cultivar description/Reference
LG3430	1997	4524	Canola	Canada	N.A.
Lynx	1990	N.A.	Canola	Germany	Raman et al. 2012 DNA Res. 19: 51-65
Mercury	1992	3532	HEAR	Canada	Scarth et al. (1995) Can. J. Plant Sci. 75: 205-206
MillenniUM03	1999	5012	HEAR	Canada	McVetty et al. (200) Can. J. Plant Sci. 80: 611–612
Mystic	1998	N.A.	Canola	Australia	Raman et al. 2012 DNA Res. 19: 51-65
Polo	1994	3883	Canola	England	N.A.
Profit	1989	3107	Canola	Canada	Rakow and Downey (1993) Can. J. Plant Sci. 73: 187-188
Quantum	1995	4062	Canola	Canada	Stringham et al. (1995) Can. J. Plant Sci. 75: 903-904
Red River 1826	2006	6030	HEAR	Canada	McVetty et al. (1996) Can. J. Plant Sci. 86: 1179–1180

Table 4.1 continued

Genotype	Year of registration	Registration number	Oil quality	Country of origin	Cultivar description/Reference
Red River 1852	2006	6029	HEAR	Canada	McVetty et al. (2006) Can. J. Plant Sci. 86: 1181–1182
Red River 1861	2012	7155	HEAR	Canada	McVetty et al. (2012) Can. J. Plant Sci. 92: 14071409
Red River 1997	2010	6733	HEAR	Canada	McVetty et al. (2010) Can. J. Plant Sci.90: 711-713
Regent	1977	1746	Canola	Canada	Canola Council of Canada; https://bit.ly/3kiqlro
Reston	1982	2190	HEAR	Canada	Canola Council of Canada; https://bit.ly/3kiqlro
Savery	2000	N.A.	HEAR	Germany	N.A.
Sentry	1996	4366	Canola	Canada	Rimmer et al. (1998) Can. J. Plant Sci. 78: 615–616
Skyhawk	2000	5101	Canola	Denmark	N.A.
SP Banner RR	2002	5460	Canola	Canada	N.A.

Table 4.1 continued

Genotype	Year of registration	Registration number	Oil quality	Country of origin	Cultivar description/Reference
SP Bucky RR	2002	5468	Canola	Canada	N.A.
SP Desirable	2004	5827	Canola	Canada	N.A.
Surpass400	2000	N.A.	Canola	Australia	Raman et al. 2012 DNA Res. 19: 51-65
SW Gladiator RR	2001	5308	Canola	England	N.A.
Topas	1987	I-65	Canola	Sweden	Hasan et al. (2006) Genet. Resour. Crop Ev. 53: 793-802
Tower	1974	1518	Canola	Canada	Stefansson and Kondra (1975) Can. J. Plant Sci. 55:343-344
Westar	1982	2238	Canola	Canada	Klassen et al. (1986) Can. J. Plant Sci. 67: 491-493
Wizzard	2008	5644	Canola	Sweden	N.A.
Z-821	2005	N.A.	Canola	China	He et al. (1987) Chinese J. Oil Crop Sci. 2: 11-15

within the greenhouse, staggered by a two-week period. Plants were irrigated to soil saturation when the top 5 cm of the substrate was dry and fertilized twice before bolting (growth stage 30) at the same rate as previously described. Powdery mildew was treated with a single application of powdered sulfur (N. M. Bartlett Inc., Beamsville, Canada) dusted onto the leaves of all plants at a rate of approximately 10 mL per adult flowering plant. Thrips were treated with a single application of Pylon (BASF, Mississauga, Canada) to all plants following the manufacturer's label rate. Inflorescences were covered with a perforated plastic bag to ensure self-pollination at growth stage 59 – 60; all open flowers present at the time of bagging were removed prior to bagging. Bagged plants were agitated every morning to maximize seed set. The pollination bag was removed once flowering was complete. Plants were irrigated to soil saturation when the top 5 cm of the medium was dry until the peduncle began to show signs of chlorosis (approximately growth stage 83 – 87) at which point irrigation was terminated and siliques were allowed to dry completely on the plant. Seeds were harvested and cleaned of plant debris before being stored at room temperature until analysis. Seed moisture was not measured.

4.3.3 Non-destructive measurement of total oil, total protein, and glucosinolate content

Total oil, total protein, and glucosinolate concentrations were measured by near infrared (NIR) spectrometry at the University of Manitoba Oilseed Quality Laboratory (certified by the Canadian Grain Commission) by Debbie Witko and Kelsey Dickson following established standard methods (Daun et al. 1994; Barthet et al. 2020). Briefly, seeds were packed into medium-sized sample cups and passed through a NIRSystems 6500 (FOSS, Eden Prairie, USA). Oil, protein, and glucosinolate contents, respectively, were determined based on reflectance spectra and reported as a percentage of the seed. Oil and protein content were determined at 0% moisture basis, respectively, while glucosinolate content was determined at 8.5% moisture.

4.3.4 Defatting and total soluble protein extraction

Total soluble protein (TSP) was extracted from defatted meal. To produce canola meal, one gram of whole seed was crushed in a porcelain mortar and pestle (CoorsTek, Golden, USA) and transferred to a Kimax glass test-tube with a screw cap (DWK Life Science, Rockwood, USA).

For defatting, approximately 7 mL of petroleum ether (ACS reagent grade; Sigma-Aldrich, Oakville, Canada) was added to the tube, the cap screwed on, and the tube was placed at an angle on a rotary shaker set to a moderate speed that permitted the crushed seed to continually mix gently with the solvent. The seed was defatted overnight (14 to 16 hrs) at room temperature in the fume hood. The solvent was then replaced with an equal volume of fresh solvent the next morning, and defatting continued for another eight hours. At the end of defatting, the solvent was decanted, the oil-free meal was transferred to an open 15 mL Falcon tube, rinsed once with petroleum ether, and allowed to dry in the fume hood for 48 hrs. Once the meal was thoroughly dried, the tubes were capped and stored at room temperature.

To extract total soluble protein (TSP), 50 mg defatted meal was combined with 1 mL of Tris-buffered saline (TBS; 10 mM Tris, 150 mM NaCl, pH 7.5) and one copper coated steel 4.5 mm BB pellet (Crosman, Bloomfield, USA) in a 2 mL microcentrifuge tube. The meal was homogenized using a TissueLyser II (Qiagen, Toronto, Canada) for 5 minutes at 20 Hz and then centrifuged (16, 000 xg, 10 min, room temperature). The supernatant was transferred to a 15 mL centrifuge tube and held on ice. The meal was repeatedly extracted three more times and the supernatant from all four extractions were pooled as TSP. The reference TSP sample was extracted from the cultivar 'Westar' produced in an isolated field in Winnipeg, Manitoba in 2015.

4.3.5 Phenotyping cruciferin content by Western Blotting and densitometry

Protein concentration was determined using the Bio-Rad protein assay (Bio-Rad, Hercules, USA) against a bovine serum albumin standard curve following the manufacturer's microtiter plate protocol. After quantification, TSP samples were diluted to 100 ng/µL with TBS and subsequently combined with an equal volume of 2x Laemmli buffer containing 100 mM DTT. The diluted samples were heated at 65 °C for 10 min. Samples were cooled to room temperature before storage at 4 °C until use. Residual TSP samples were stored at -80 °C for long term storage.

For each sample, 300 ng of TSP was resolved on a 12 % TGX polyacrylamide gel (Bio-Rad, Hercules, USA). A reference TSP sample from the genotype 'Westar' was included in each gel. The gel was UV-activated for 1 min with a Gel Doc XR+ System (Bio-Rad, Hercules, USA) and an image was captured for downstream total protein normalization with pre-set exposure for faint bands. The resolved proteins were semi-dry blotted onto PVDF membrane with a Trans-Blot Turbo Transfer System (Bio-Rad, Hercules, USA) using the pre-set Turbo protocol for 1 mini gel (25 V, 1.3 A, 7 min).

The blot was immediately blocked in a 5 % solution of fat-free milk in TBST for 1 hr at room temperature. After 3 5-min washes in TBST, the blot was probed with a custom polyclonal anti-cruciferin antibody (previously described in Chapter 3) diluted 1:4000 in blocking solution for 1 hr at room temperature. The membrane was washed again as described previously and secondary horseradish peroxidase-conjugated monoclonal mouse-anti-rabbit antibody (clone M205; GenScript, Piscataway, USA) was applied at a dilution of 1:8000 for 1 hr at room temperature. The blot was washed as described previously and 4 °C 1-Step Ultra TMB-Blotting Solution (ThermoFisher, Waltham, USA) was applied. Blots were developed for 2 min at room temperature with gentle shaking before the reaction was quenched with deionized water. The blot

was rinsed twice in deionized water before being captured on a Gel Doc with the pre-set colorimetric blot protocol and automatic exposure for low intensity bands.

Densitometry to quantify cruciferin content was performed using the Image Lab Software Suite (Bio-Rad, Hercules, USA). All bands showing signal on the Western blot were considered for densitometry. Signal intensity of each band was assessed using default parameters: sensitivity, 148.48; scale, 7; noise filter, 4; shoulder 1; and disc size, 3.8 mm. Total protein normalization was performed against the standard 'Westar' sample as the reference lane. Cruciferin content of each sample was expressed as a percent relative to the field-grown 'Westar' reference sample.

4.3.6 Statistical analyses

All statistical analyses were performed with SAS University Edition (SAS Institute Inc., Cary, USA) accessed via SAS Studio release 3.8 (SAS Institute Inc., Cary, USA). Analysis of variance for each trait at each location was initially performed using the *proc mixed* procedure following the model **phenotype = genotype + block + genotype*block + residual** where the genotype effect was fixed and the block effect was random. Data from both locations were combined based on visual assessment of residual plots generated using the *proc univariate* command and a pooled analysis was performed following the model **phenotype = genotype + location + block(location) + genotype*location + residual** where the genotype effect was fixed and all other effects considered random. Least squared means were calculated with the *lsmeans* command and applying the Tukey adjustment. Genotypes were separated into classes using the Tukey-Kramer method implemented with *proc glm.* A boxplot showing cruciferin content across the genotypes was generated using *proc sgplot vbox*. Correlation plots between traits were produced with the *proc sgscatter* command and correlation coefficients were calculated with the *proc corr* command.

4.4 Results

4.4.1 Non-destructive measurement of total oil, total protein, and glucosinolates

Total oil, total protein and glucosinolate content are presented in Figure 4.1. Summary statistics of the parameters are presented in Table 4.2. Oil content ranged from 50.7% to 56.5% with a mean of 54.7% and standard deviation (SD) of 1.9. Total protein ranged from 11.8% to 19.2% with a mean of 13.8% and SD of 1.4, and glucosinolate content ranged from 4.1 µmol/g to 51.8 µmol/g with a mean of 12.4 µmol/g and a SD of 7.4. Analysis of variance indicated a significant genotype effect for the three assayed traits (Table 4.3).

4.4.2 Phenotyping cruciferin content by Western Blotting and densitometry

Cruciferin content of the genotypes is presented in (Figure 4.2). Mean cruciferin content of the genotypes varied from 43.64% to 92.72% of the reference sample; cruciferin content differed by approximately two-fold between the highest and lowest-cruciferin cultivars in the population. The 'Westar' genotype within the experimental population has a lower cruciferin content than the field-grown reference sample indicating different growth conditions likely impact cruciferin content. Analysis of variance indicated genotype had a significant effect on cruciferin content (Table 4.3D). Genotypes were separated into low, medium, and high-cruciferin classes based on the Tukey-Kramer method with 39 of the 44 cultivars assigned to the medium-cruciferin class; the low-cruciferin class consisted of a single cultivar 'Wizzard' while the high-cruciferin class consisted of four cultivars: 'Red River 1852', '04R2026', 'Red River 1826', and 'LG3333' (Figure 4.2).

4.4.3 Phenotypic correlation analyses

The correlation coefficients of all pair-wise combinations of protein content, oil content, glucosinolate content, and cruciferin content are presented in (Figure 4.3). Protein content

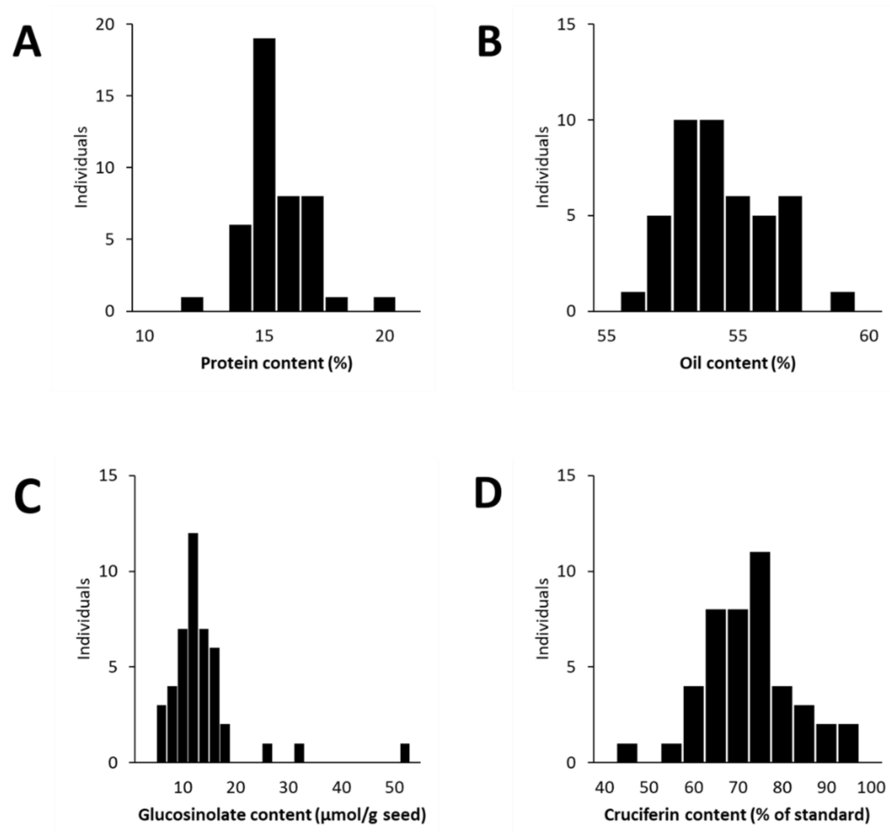

Figure 4.1. Frequency distribution of protein content (A), oil content (B), glucosinolate content (C), and cruciferin content (D) for spring *Brassica napus* L. cultivars phenotyped in 2019 under greenhouse conditions in Winnipeg, Canada. The y axis represents number of genotypes.

Table 4.2. Summary statistics for protein content, oil content, glucosinolate content, and cruciferin content for a population of spring *Brassica napus* L. cultivars phenotyped in 2019 under greenhouse conditions in Winnipeg, Canada. Values presented are least-squared means. Cruciferin content was assessed relative to a reference protein sample from the cultivar 'Westar.'

Trait	Minimum	Maximum	Mean	Standard Deviation	C.V.(%)
Protein content (%)	11.8	19.2	13.8	1.4	9.9
Oil content (%)	50.7	56.5	54.7	1.9	3.5
Glucosinolate content (μmol/g)	4.1	51.8	12.4	7.4	60.8
Cruciferin content	0.4	0.9	0.7	0.2	30.7

Table 4.3. Pooled analysis of variance tables for protein content (A), oil content (B), glucosinolate content (C), and cruciferin content (D) of spring *Brassica napus* L. phenotyped in 2019 under greenhouse conditions in Winnipeg, Canada. Cruciferin content was assessed relative to a reference protein sample from the cultivar 'Westar.'

A Protein content

Source	**DF**	**Mean Square**	**Error DF**	**F value**	**P value**
Genotype	43	10.235	43	5.56	<.0001
Location	1	376.249	4.3994	32.78	0.003
Genotype*Location	4	10.928	171	8.47	<.0001
Block(Location)	43	1.840	171	1.43	0.059
Residual	171	1.290			

B Oil content

Source	**DF**	**Mean Square**	**Error DF**	**F value**	**P value**
Genotype	43	10.235	43	5.56	<.0001
Location	1	376.249	4.3994	32.78	0.003
Genotype*Location	4	10.928	171	8.47	<.0001
Block(Location)	43	1.840	171	1.43	0.059
Residual	171	1.290			

C Glucosinolate content

Source	**DF**	**Mean Square**	**Error DF**	**F value**	**P value**
Genotype	43	10.235	43	5.56	<.0001
Location	1	376.249	4.3994	32.78	0.003
Genotype*Location	4	10.928	171	8.47	<.0001
Block(Location)	43	1.840	171	1.43	0.059
Residual	171	1.290			

D Cruciferin content

Source	**DF**	**Mean Square**	**Error DF**	**F value**	**P value**
Genotype	43	10.235	43	5.56	<.0001
Location	1	376.249	4.3994	32.78	0.003
Genotype*Location	4	10.928	171	8.47	<.0001
Block(Location)	43	1.840	171	1.43	0.059
Residual	171	1.290			

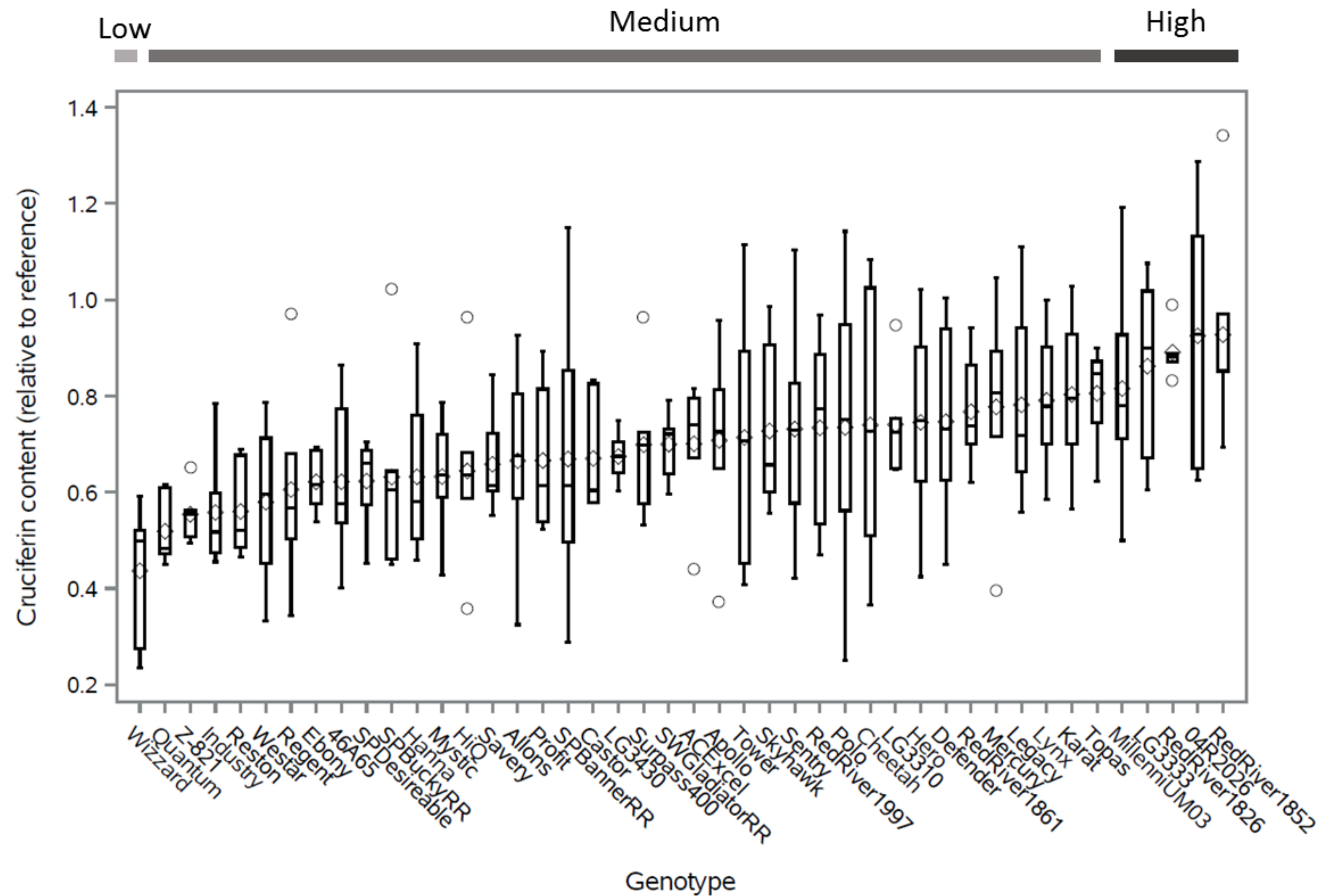

Figure 4.2. Cruciferin content in spring *Brassica napus* L. grown in a randomized complete block design, with three replications and repeated in two experiments, under greenhouse conditions in Winnipeg, Canada in 2019. Cruciferin content was assessed by Western

blotting relative to a field-grown reference sample from the cultivar 'Westar.' Boxes represent the value of all biological replicates across the two experiments. The horizontal bar above identifies genotypes classified into low, medium, and high-cruciferin classes by the Tukey-Kramer method.

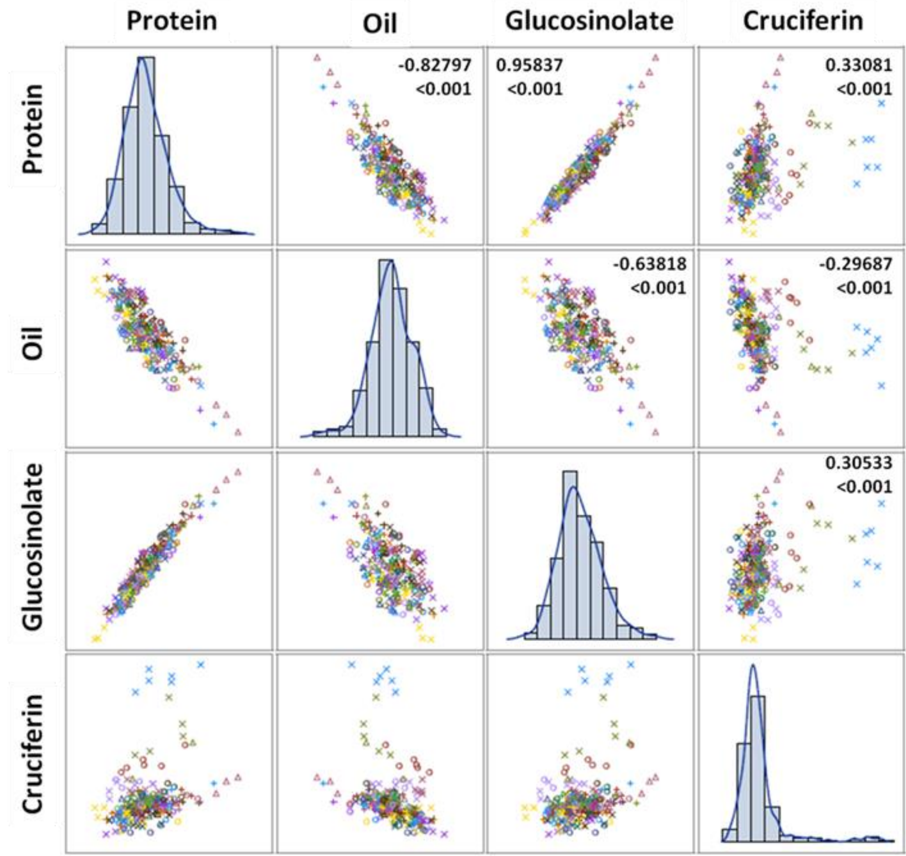

Figure 4.3. Correlations between seed protein content, oil content, glucosinolate content, and cruciferin content were assessed in 44 spring *Brassica napus* L. genotypes grown in a randomized complete block design with three blocks, replicated twice, under greenhouse conditions at the Crop Technology Centre, University of Manitoba, Winnipeg in 2019. Histograms along the diagonal axis represent the distribution of the respective trait. Data points in the scatterplots are coloured by genotype. Values in the matrix indicate the Pearson correlation coefficient of the two traits and its associated p value directly below.

exhibited a significant negative correlation ($r = -0.83$, $p < 0.001$, $n = 44$) with oil content and a significant positive correlation with glucosinolate content ($r = 0.96$, $p < 0.001$, $n = 44$). Cruciferin content was positively correlated with protein content ($r = 0.33$, $p < 0.01$, $n = 44$) and glucosinolate content ($r = 0.35$, $p < 0.01$, $n = 44$) but was negatively correlated with oil content ($r = -0.30$, $p < 0.001$, $n = 44$). The correlation between glucosinolate content and protein content appeared stronger than that of the latter with oil content.

4.5 Discussion

4.5.1 Seed quality of the population

Oil content and protein content of the 2019 western Canadian canola crop was 44.6 % and 20.4%, respectively, with 9 µmol/g total glucosinolates (Barthet 2019). In this experiment, mean oil content across the population was higher than field-grown samples across western Canada while total protein content was lower. Mean total glucosinolate content remained consistent with that which was reported for the 2019 western Canadian field harvest (Barthet 2019). As breeding efforts continually prioritize the improvement of oil content and quality in spring *B. napus* (Ton et al., 2020), the inclusion of historic cultivars in the population was expected to result in an underrepresentation of oil content relative to last season's crop. Further, historic cultivars were expected to have higher protein content than their contemporary counterparts as oil content and protein content are typically inversely related (Chao et al., 2017; Jasinski et al., 2018; Jolivet et al., 2013; Schatzki et al., 2014; Schwender and Hay, 2012). It would therefore be reasonable to conclude that the inverse trend observed in our quality data with that of recent harvest data may be due to differences in growing environment (Poorter et al., 2016).

4.5.2 Effect of greenhouse environment on seed quality

Variation in plant growth conditions can have profound effects on the accuracy and reproducibility of phenotyping (Poorter et al., 2012). Phenotyping the population under controlled greenhouse conditions attempted to minimize the environmental variation between experimental units and replicates of the experiment; understandably, such conditions are not representative of field conditions and deviations in seed quality from field-grown samples were expected (Hohmann et al., 2016; Poorter et al., 2016). Notably in the experiment, irrigation was applied until the seeds approached physiological maturity. In winter *B. napus*, a reduction in seed oil content with a concomitant increase in seed protein content was induced by drought under both field (Elferjani and Soolanayakanahally, 2018; Zhu et al., 2016) and greenhouse conditions (Hatzig et al., 2018). Therefore, in supplying irrigation to the point of physiological maturity of the seed, seed quality may have been biased towards oil accumulation and simultaneously against protein accumulation. It is worth noting, however, that mild drought stress under field conditions did not ostensibly impact seed quality in winter *B. napus* (Drebenstedt et al., 2020), perhaps due to unrestricted root growth which potentiates water to be acquired from deeper within the soil (Poorter et al., 2016). Despite differences between greenhouse and field conditions and its effect on protein and oil content respectively, selection for seed quality traits under greenhouse conditions in the early stages of *B. napus* cultivar development may still be a fruitful endeavour to expedite breeding programs (Bahrani and McVetty, 2008). To accurately reflect field conditions for phenotyping, the use of large planting containers instead of six-inch pots under greenhouse conditions should be considered for phenotyping studies (Hohmann et al., 2016).

4.5.3 Genetic variation in cruciferin content exists within spring *B. napus* germplasm

To our knowledge, this is the first assessment of the genetic diversity of cruciferin content in spring-type *B. napus*. Similar to winter rapeseed (Raab et al. 1992; Malabat et al. 2003), genetic

diversity in cruciferin content was observed in the population and alludes to the amenability of the trait in spring ecotypes of *B. napus* to be improved through conventional breeding. Within the population a single low-cruciferin canola-quality cultivar ‘Wizzard’ was identified while four high-cruciferin genotypes were identified of which ‘Red River 1852’ and ‘Red River 1826’ were high erucic acid rapeseed (HEAR) cultivars and both ‘LG3333’ and ‘04R2026’ were canola-quality cultivars. The presence of additional HEAR cultivars in the medium-cruciferin group suggests the divergence of cruciferin content was likely not a direct consequence of selecting for erucic acid. Further, the divergence of cruciferin content is also not likely to be an unintended result of selection against glucosinolate content as early work by Finlayson et al. (1973) demonstrated that storage protein profiles between ‘Bronowski,’ the source of the low glucosinolate trait in canola (Klassen et al. 1987), and ‘Target’, a wild-type *B. napus* with improved agronomic traits appeared similar based on gel electrophoresis.

4.5.4 Cruciferin content was positively correlated with both protein content and glucosinolate content

Breeding efforts focused on improving cruciferin content must not result in unintended changes in other seed quality traits critical to the market; therefore, an assessment of the correlation between cruciferin content and other seed parameters was warranted. In our population, cruciferin content was negatively correlated with oil content. This observation contradicts the work of Schatzki et al. (2014) who reported a positive correlation between cruciferin content and oil content in a bi-parental mapping population of winter rapeseed. Similarly, a positive correlation between cruciferin content and oil content was also reported in a panel of diverse winter rapeseed cultivars, suggesting that an increase in cruciferin content was evident across the germplasm and an unintended result of the improvement of oil content in the crop (Malabat et al. 2003). The discrepancy between the results reported herein with the work of Schatzki et al. (2014) and

Malabat et al. (2003) may have arisen as a result of conducting the experiment under greenhouse conditions where optimal conditions for plant growth favouring oil accumulation may have biased the correlation analyses.

A significant negative correlation was expected and observed between oil content and protein content in the population, which conforms to previous literature in *B. napus* (Schwender and Hay 2012; Jolivet et al. 2013; Schatzki et al. 2014; Chao et al. 2017; Jasinski et al. 2018) and other oilseed crops (Li et al. 2017b; Patil et al. 2018; Kambhampati et al. 2020). This negative correlation is often attributed to competition for carbon between oil and protein biosynthesis (Kambhampati et al. 2020) though emerging evidence suggests morphological differences in seed anatomy may be at play (Rolletschek et al. 2020). Glucosinolate and protein content in the seeds can be elevated in response to environmental conditions such as abiotic stress and nutrition (Elferjani and Soolanayakanahally 2018); furthermore, the domestication and continued breeding efforts for oil content have effectively selected for the concomitant decrease of both protein and glucosinolate content. Therefore, a positive correlation between protein and glucosinolate content was expected and observed in the data.

To confirm the correlations observed in this study and to ensure the results are translatable to western Canadian canola production, additional phenotyping experiments with larger populations under field conditions are warranted. Furthermore, determining the degree and consistency with which seed quality data differs between greenhouse and field environments will be useful to accelerate recurrent selection programs for cruciferin content through the use of controlled environments.

4.6 Conclusion

In this project, we report the first assessment of the genetic diversity of cruciferin content in spring *B. napus* and the identification of both high and low-cruciferin cultivars, respectively. Although cruciferin content was consistently lower than that of the field-grown reference sample, the high cruciferin cultivars ‘Red River 1852’, ‘04R2026’, ‘Red River 1826’, and ‘LG3333’ were found to have almost double the cruciferin content of the low-cruciferin cultivar ‘Wizzard.’ Further, cruciferin content was correlated with both protein and glucosinolate content, which contradicted previous results reported in the literature. These discrepancies may be a result of conducting the experiment under greenhouse conditions wherein prolonged irrigation through to seed maturity favoured oil accumulation at the expense of protein. With variation in cruciferin content observed in the population, the next step is to identify the mechanisms that control cruciferin accumulation. In addition, bi-parental mapping populations can be generated using high and low-cruciferin cultivars to facilitate genetic mapping studies to identify quantitative trait loci that correspond to cruciferin content. Future work should also focus on the development of phenotyping assays that facilitate greater throughput than Western blotting so that larger populations and additional field locations can be used to improve phenotyping accuracy.

In the previous chapter, genetic variation in cruciferin content was found to exist in a population of *B. napus* that spanned the breeding history of the crop. This finding suggested that improvements to cruciferin content can be achieved through conventional breeding. To facilitate breeding efforts and to better understand the basic molecular mechanisms that govern cruciferin accumulation, the subsequent chapter aims to identify genomic regions and molecular markers associated with cruciferin content.

Chapter 5. Identification of genomic regions associated with cruciferin content in spring-type *Brassica napus* L.

5.1 Abstract

The seed storage protein cruciferin in the oilseed *Brassica napus* L. has technical functionalities that enable it to be used in food processing in addition to being a source of quality dietary protein. Breeding for improved cruciferin content can add value to the crop by enabling the seed meal to be used for food after the oil has been extracted. Genetic variation in cruciferin content is present in spring *B. napus;* however, the genetic mechanisms that underlie this variation have yet to be explored in this economically important crop. To this end, an association study was conducted using a panel of 44 spring *B. napus* spanning 38 years of breeding history of the crop to identify molecular markers that are associated with this trait and other seed quality parameters. Seed protein and oil content were assessed by near-infrared spectroscopy while cruciferin content was quantified immunologically with custom antibodies; the population was genotyped using the 60K Illumina Infinium single nucleotide polymorphism genotyping array. Six association mapping models were evaluated and a mixed linear model with consideration for population structure through principal component analysis was deemed the most optimal to assess cruciferin content. Molecular markers associated with cruciferin content were detected within haplotype blocks on chromosome A06 and A07, respectively, and did not coincide with known cruciferin biosynthesis genes nor loci previously implicated in cruciferin content. Candidate genes were identified by gene ontology analysis and likely function to indirectly regulate cruciferin content. To our knowledge, this is the first genome-wide exploration of cruciferin content in spring *B. napus*.

5.2 Introduction

Plant-based proteins are an environmentally-friendly, complete, and potential alternative to animal proteins for the human diet (Sabaté et al. 2015), the majority of which are represented by seed storage proteins (SSP). In the oilseed *Brassica napus* L., the SSP pool consists of primarily a combination of the globulin-type cruciferin and the albumin-type napin, with the former constituting the largest fraction, at approximately 60% (Crouch and Sussex 1981). Cruciferin has desirable functional properties for food processing (Tan et al. 2011; Cheung et al. 2014a; Jones 2016) and can be incorporated into food products as a structural ingredient and as a protein supplement. Technological improvements have been made to produce canola protein products and advance their commercialization (Mupondwa et al. 2018). The isolation of cruciferin-rich protein fractions successfully removes napin, a potential allergen (Moreno and Clemente 2008; Aider and Barbana 2011), and other antinutritive compounds (Mupondwa et al. 2018); however, this comes at a cost for processors. The development of *B. napus* cultivars with high-cruciferin content may simplify separation technologies during extraction as an increase in cruciferin likely leads to a reduction of napin (Kohno-Murase et al. 1994). Consequently, high-cruciferin cultivars can render cruciferin-based protein products more competitively priced.

Altering the composition of SSP in plants typically does not result in changes to total seed protein content or oil content (Schmidt and Herman 2008; Withana-Gamage et al. 2013; Wu and Messing 2014). A similar nutrient rebalancing mechanism has been observed in *B. napus* whereby complete elimination of SSP resulted in the compensatory increase of other seed proteins without changing the overall seed protein content (Rolletschek et al. 2020). This phenomenon is particularly important for SSP improvement in *B. napus* as the crop is primarily grown as an oilseed: seed protein content is negatively correlated with oil content and improvements to the

former at the expense of the latter renders such cultivars unacceptable for production. Therefore, improving protein quality through modulating SSP composition can add value *to B. napus* protein without negatively impacting the oil content.

In *B. napus*, direct efforts to modulate SSP composition through the use of anti-sense technology to transcriptionally silence SSP expression have been successful (Kohno-Murase et al. 1994, 1995). To date, traditional breeding for SSP has not yet been reported in *B. napus*; however, breeding improvements to oil quality appear to have resulted in a simultaneous increase in cruciferin content (Malabat et al. 2003). Targeted efforts to improve cruciferin content in *B. napus* through conventional breeding requires an understanding of the genetic elements that underlie the trait in order for molecular markers to be identified for use in marker-assisted selection or genomic selection. Marker-assisted selection facilitates selection early in the breeding cycle, abating the need to grow plants to maturity, thus increasing the efficiency of the breeding process (Cobb et al. 2019).

Linkage mapping has successfully identified molecular markers that are associated with cruciferin content in winter *B. napus* (Schatzki et al. 2014). However, the use of this mapping approach is limited by the need for biparental populations derived from parents with divergent phenotypes (Mauricio 2001; Huang and Han 2014). For this reason, linkage mapping for cruciferin content in spring-type *B. napus* has not been possible as parents divergent in cruciferin content have not previously been identified until recently (Chapter 4).

Despite this, the identification of markers significantly associated with cruciferin content in spring-type canola is possible through association mapping (genome-wide association studies; GWAS). Association mapping was initially developed for the study of complex diseases in humans

(Tam et al. 2019) and has been successfully applied across crop species since its first implementation in corn (Thornsberry et al. 2001). Unlike linkage mapping, which relies on a limited number of recombination events produced in a biparental cross, GWAS evaluates genome-wide historic recombination events using panels of genetically diverse individuals (Huang and Han 2014). In addition to eliminating the need to generate mapping populations, GWAS in combination with high-throughput genotyping platforms such as the Brassica 60K Illumina Infinium genotyping array are able to map markers with higher resolution than linkage mapping due to high marker density across the genome (Clarke et al. 2016).

A major limitation in the accuracy of GWAS is the presence of population structure which can lead to the detection of markers associated with individual subpopulations rather than the trait of interest (Price et al. 2010; J. and Cloutier 2012). Population structure is particularly prominent in plant breeding as subpopulations are intentionally generated through controlled crossing and genetic drift occurs through segregation of different breeding programs (Huang and Han 2014; Moyers et al. 2018). To mitigate the effect of population structure, the use of mixed linear models with a kinship matrix describing the relatedness between individuals has been recommended (Zhang et al. 2010). Furthermore, population structure information can be included as a covariate to remove its confounding effects from the analysis (Zhao et al. 2007) and such models have been successfully employed to identify molecular markers statistically associated with various seed quality traits in *B. napus* (Guan et al. 2019; Tang et al. 2019).

The identification of molecular markers associated with cruciferin content will facilitate the development of high-cruciferin cultivars of *B. napus*. In addition, the identification of novel molecular markers associated with seed protein content and oil content may assist in the further breeding improvement of these economically important traits. To this end, the objective of this

research is to identify molecular markers associated with seed protein content, oil content, and cruciferin content, in spring-type *B. napus*. In addition, the value of including population structure data in the association mapping model is assessed.

5.3 Material and Methods

5.3.1 Plant material and growth parameters

The genotypes utilized have previously been described in Table 4.1; data on the year of registration, geographical origin, and oil quality of each genotype were collected from the literature. Briefly, the population consisted of forty-four registered, open-pollinated cultivars of *B. napus* from the University of Manitoba Brassica collection. Plant growth parameters and experimental design are described in detail in Chapter Four (section 4.3). In summary, plants were grown in a randomized complete block design with three replications under standard greenhouse conditions at the University of Manitoba Crop Technology Centre from January to April 2019. The experiment was repeated a second time staggered by a two-week period. Plants were cared for using standard greenhouse practices as outlined in Chapter Four (section 4.3) and irrigated as required until the base of the raceme began to show signs of yellowing (BBCH growth stage 83 – 87) (Lancashire et al. 1991); irrigation was terminated and siliques were allowed to dry completely on the plant before seeds were harvested. Cleaned seed were stored in paper envelopes at room temperature until analysis.

5.3.2 Genotyping and data quality control

Genomic DNA was extracted from the first true leaf of seedlings using a standard CTAB extraction protocol (Porebski et al. 1997) modified to exclude the use of both polyvinylpyrrolidone and 2-mercaptoethanol, and the use of phenol in place of octanol. DNA was sent to Agriculture

and Agri-Food Canada (AAFC), Saskatoon, Canada for genotyping using the Illumina Brassica 60K Infinium array (Illumina Inc., U.S.A.) (Clarke et al. 2016). Raw intensity data files and a custom cluster file were received from AAFC and both loaded into GenomeStudio software (Illumina Inc., USA). Quality control of the initial 52,157 markers on the SNP panel was performed by applying filters for minor allele frequency less than 5% and a call rate of less than 80% (Liu et al. 2016c). The 30,290 remaining markers were then mapped to the Darmor reference genome (Chalhoub et al. 2014) supplied by AAFC. Genotypes with a genotyping frequency less than 80% were removed from the population. Missing genotypes were not imputed. Only markers and genotypes that meet filtering criteria were used for analyses.

5.3.3 Estimating linkage disequilibrium

To estimate linkage disequilibrium (LD), correlation coefficients (r^2) were calculated for every possible pair of markers using PLINK version 1.90b6.16 (Chang et al. 2015) by applying the following flags: --r2 --ld-window-r2 0 --ld-window 999999. The --ld-window-kb flag was set to 71,850, the length of the longest chromosome in the Darmor reference genome (Bayer et al. 2017). For ease of visualization, marker pairs were binned by 10 kb distances across the genome and the average r^2 was calculated for each bin. Genome-wide LD decay plots were generated by plotting r^2 against distance by bins in R Studio version 1.2.1335. The R code to generate the plots (https://www.biostars.org/p/300381/) was modified to exclude the marker thinning step; filters for sub-genomes and chromosome number were applied to generate LD decay plots for sub genomes and individual chromosomes, respectively. The midpoint of the first bin at which r^2 dropped below 0.2 was assessed as the distance where LD decayed.

Population structure

Population structure was assessed by Bayesian clustering and principal component analysis (PCA) using a subset of independent markers selected based on LD. LD-pruning was performed in PLINK version 1.90b6.16 (Chang et al. 2015) using the --indep-pairwise flag with a window size of 50 markers, a shift interval of 5 markers, and a r^2 threshold of 0.2. LD-pruning was confirmed by visually inspecting LD heat maps generated using TASSEL v5.2.63 (Bradbury et al. 2007).

Bayesian clustering was performed using STRUCTURE version 2.3.4 (Pritchard et al. 2000). An admixture model with a 100,000 burn-in period and 100,000 Markov Chain Monte Carlo iterations was used to cluster the genotypes from one (k = 1) to ten (k = 10) theoretical subpopulations. Ten iterations were performed for each *k* value (Gajardo et al. 2015). The optimal number of subpopulations was determined by the method of Evanno *et al.* (2005) implemented in Structure Harvester web v0.6.94 (Earl and VonHoldt 2012). Briefly, delta k was plotted against k, the theoretical number of groups, and the optimal number of subpopulations was determined by observing the *k* value at which the graph peaked. Thereafter, the population was assessed again by running STRUCTURE with the *k* value set to 6 to generate the q matrix. The result of the clustering was displayed with the "order by q" option in TASSEL and visualized in Microsoft Excel 2016 (Microsoft Corporation, Redmond, U.S.A.). Clustering results with patterns of the year of registration (by decade), geographic origin, and oil quality superimposed were also visualized using Microsoft Excel. Subpopulation membership was assessed using a value of 70% and genotypes having less than 70% of any given population were deemed an admixture of more than one subpopulation (Gyawali et al. 2013; Liu et al. 2016c).

Principal component analysis (PCA) was conducted using TASSEL. A scree plot was used to determine the optimal number of principal components to retain. The first three principal components were selected for use in association mapping. Results of the PCA and the superimposition of year of registration, geographical origin, and oil quality data were visualized in R Studio.

5.3.4 Phenotyping for protein content, oil content, and cruciferin content

Phenotyping was conducted on cleaned, dry seed as described in Chapter Four (section 4.3): seed protein and oil content were estimated on a percent basis using near infrared spectroscopy at 0% moisture (Daun et al. 1994; Barthet et al. 2020). Cruciferin content was estimated by Western blotting using custom polyclonal antibody raised against a conserved region within the α chain of all five *B. napus* cruciferin accessions in the Uniprot KB (The UniProt Consortium et al. 2021) as described in the Materials and Methods of Chapter Two (Section 2.3). Cruciferin content was quantified relative to a reference seed sample of field grown *B. napus* 'Westar.'

5.3.5 Generating best linear unbiased predictors

Best linear unbiased predictors (BLUP) of each trait were calculated using a linear mixed model **phenotype = genotype + location + block(location) + genotype*location + residual** with genotype, location, block within location, and the genotype by location interaction as random effects. The model was implemented using the lme4 package (Bates et al. 2015) in R 3.6.3 (R Core Team 2020) with RStudio version 1.2.1335 (RStudio Team 2020) and code from GitHub (https://github.com/mighster/BLUPs_Heritability/blob/master/BLUP_Tutorial.R).

Association mapping and model comparison

Association mapping was conducted using TASSEL using standard workflows outlined by the program. Various general linear (GLM) and mixed linear models (MLM) (Zhang et al. 2010), each with different considerations for population structure were implemented to identify markers significantly associated with three seed quality traits: seed protein content, oil content, and cruciferin content. A total of six statistical models were tested: (1) a naïve general linear model (GLM), $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X} + \boldsymbol{\epsilon}$; (2) a GLM with population structure accounted for by PCA (GLM+PCA), $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X} + \boldsymbol{P}_{\boldsymbol{PCA}} + \boldsymbol{\epsilon}$; (3) a GLM with population structure accounted for by a q matrix (GLM+Q), $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X} + \boldsymbol{P}_{\boldsymbol{Q}} + \boldsymbol{\epsilon}$; (4) a MLM with consideration for kinship (MLM+K), $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X} + \boldsymbol{K} + \boldsymbol{\epsilon}$; (5) a MLM with consideration for kinship and population structure using PCA (MLM+K+PCA), $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X} + \boldsymbol{K} + \boldsymbol{P}_{\boldsymbol{PCA}} + \boldsymbol{\epsilon}$; (6) and a MLM with consideration for kinship and population structure using a q matrix (MLM+K+Q), $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X} + \boldsymbol{K} + \boldsymbol{P}_{\boldsymbol{Q}} + \boldsymbol{\epsilon}$. In the models above, y represents the phenotype; μ represents the population mean; X represents the marker effect (fixed); P represents population structure effects (fixed); K represents kinship effects (random), and ϵ represents the residual effects (random). The kinship matrix required for the MLM was calculated in TASSEL using the default Centered IBS algorithm and maximum of six alleles. All mixed linear models were implemented with optimal compression.

Quantile-quantile (QQ) plots of the observed p values plotted against the expected p values were generated in R 3.6.3 (R Core Team 2020) with RStudio version 1.2.1335 (RStudio Team 2020) for each model and visually assessed. The model where the plot most closely approximated a line with a slope of 1 was deemed the most accurate (Price et al. 2010). To differentiate between visually similar QQ plots and numerically quantify the relative accuracy of the models, the average deviation of the observed p values from their expected counterparts was assessed. Markers were

arranged in ascending order and the mean squared difference (MSD) for each model was calculated as:

$$MSD = \left\{\sum_{i=1}^{n}\left(p_{obs} - p_{exp}\right)^2\right\}/n$$

where n represents the total number of markers in the model. The model with the smallest MSD was deemed to be most accurate and, after visual confirmation of the QQ plot (Fredua-Agyeman et al. 2020), was selected for downstream use to identify marker-trait associations (MTA).

The significance of MTA were assessed using both a Bonferroni-corrected threshold of $a = 0.05/n$ where n equals the number of markers, and $a = 1/n$ to account for the inflated type I error caused by the conservative nature of the former statistic (Zhang et al. 2019). An additional threshold of $a = 0.001$ was implemented to assess suggestive MTA (Li et al. 2017a; Gabur et al. 2020). Manhattan plots were generated using the CMplot package (Yin 2020) implemented in R 3.6.3 (R Core Team 2020) with RStudio version 1.2.1335 (RStudio Team 2020).

5.3.6 Candidate gene search

Candidate genes were identified using methods modified from Hatzig et al. (2018). For chromosomes containing significant or suggestive MTA under the selected model, LD decay was calculated across the whole chromosome and visualized using Haploview v4.2 (Barrett et al. 2005). Markers were rescored without consideration for Hardy-Weinberg equilibrium (Bernardo 2020) and haplotype blocks were defined using the "Solid spine of LD" option. Flanking markers in strong LD (r^2>0.2) with the associated marker formed haplotype blocks from which genes inside were extracted from the *B. napus* genome browser (Chalhoub et al. 2014) and functionally annotated in Blast2Go (Conesa and Götz 2008) using the default annotation pipeline to assign gene ontology (GO) terms. For associated markers that were not in LD with flanking markers, the region

between the flanking markers was used to extract genes of interest. GO terms were visualized in Blast2GO after applying the recommended 10% sequence filter. Candidate genes were identified by GO terms related to the trait under study.

5.4 Results

5.4.1 Marker quality control and density

After quality filtering, a total of 30,290 markers mapping to the reference genome and 43 genotypes with satisfactory genotyping frequency were retained for analysis. Marker density was visually consistent across the length chromosomes in the A sub-genome while a reduction in marker density was observed in the centromeric regions of chromosomes in the C sub-genome (Figure 5.1). Two regions of high marker density were observed on A10 and C4, respectively. Single nucleotide polymorphism marker densities for each chromosome were calculated by dividing the number of markers on the chromosome by its length (Clarke et al. 2016) and are presented in Table 5.1. In general, marker density was higher in the A sub-genome (19 kb/marker) than in the C sub-genome (29 kb/marker) (Table 5.1). Across the whole genome, marker density averaged 24 kb/marker, with a range from 11 kb/marker (chromosome C09) to 78 kb/marker (chromosome A10).

5.4.2 Linkage disequilibrium

Genome-wide LD decayed at 4.1 Mb, with LD decaying quicker in the A sub-genome (350 kb) than in the C sub-genome (6.25 Mb) (Figure 5.2). Intrachromosomal LD within the A sub-genome varied from 250 kb to 1050 kb with a mean of 470 kb and median of 250 kb. Intrachromosomal LD decayed within 1 Mb for all A chromosomes with the exception of A09 which decayed at 1.05 Mb (Figure 5.3). Intrachromosomal LD decayed slower in the C sub-

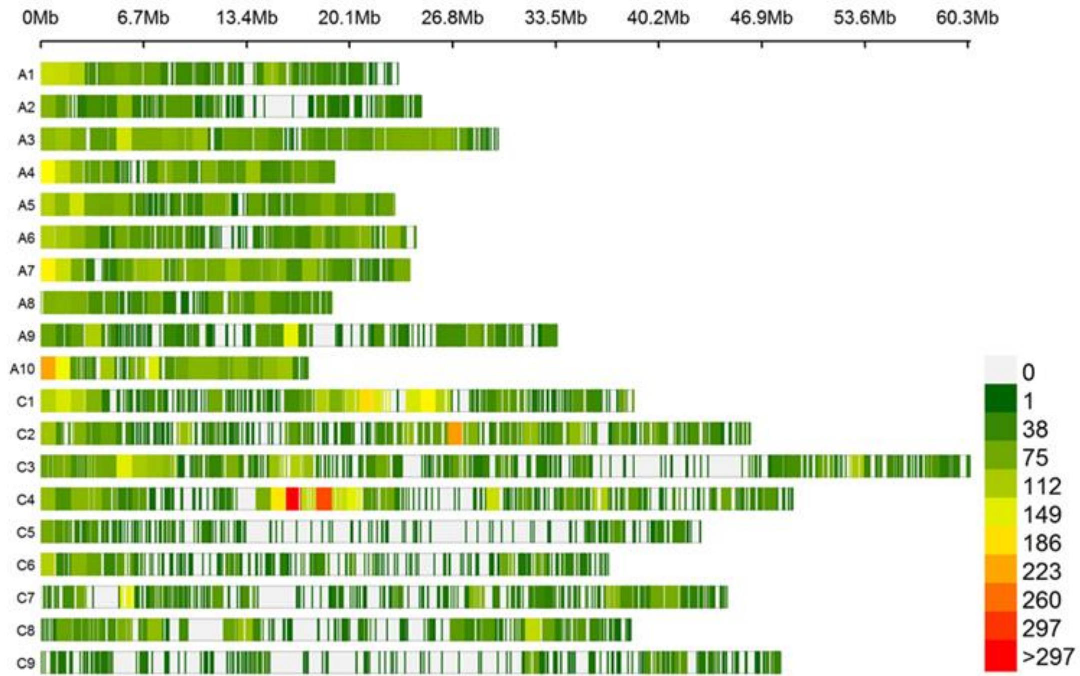

Figure 5.1. Distribution and density of single nucleotide polymorphism (SNP) markers across the chromosomes of *Brassica napus* L. in a population of 43 spring genotypes. Colour scale represents number of SNP markers per megabase.

Table 5.1. Marker density and linkage disequilibrium (LD) decay for each *Brassica napus* L. chromosome. Linkage disequilibrium decay values for the individual sub-genomes and whole genome were calculated using all markers in the sub-genome or whole genome, respectively.

Chromosome	Length (kb)	Number of Markers	Marker Density (kb/marker)	LD decay (kb)
Whole genome	725,833	30,290	24	4,150
A sub-genome	271,017	14,538	19	350
C sub-genome	454,809	15,752	29	6,250
A01	27,105	1,357	20	250
A02	29,627	851	35	650
A03	35,753	2,093	17	250
A04	21,080	1,420	15	250
A05	25,706	1,564	16	250
A06	26,146	1,447	18	250
A07	25,458	1,802	14	250
A08	21,685	1,043	21	950
A09	40,546	1,332	30	1,050

Table 5.1 continued

Chromosome	Length (kb)	Number of Markers	Marker Density (kb/marker)	LD decay (kb)
A10	17,911	1,629	11	550
A sub-genome mean				470
C01	45,604	2,443	19	8,050
C02	47,311	2,322	20	7,050
C03	67,777	2,589	26	450
C04	55,069	2,984	18	5,450
C05	48,717	677	72	250
C06	40,797	1,077	38	950
C07	48,823	1,486	33	850
C08	44,716	1,459	31	1,650
C09	55,995	715	78	450
C sub-genome mean				2,794

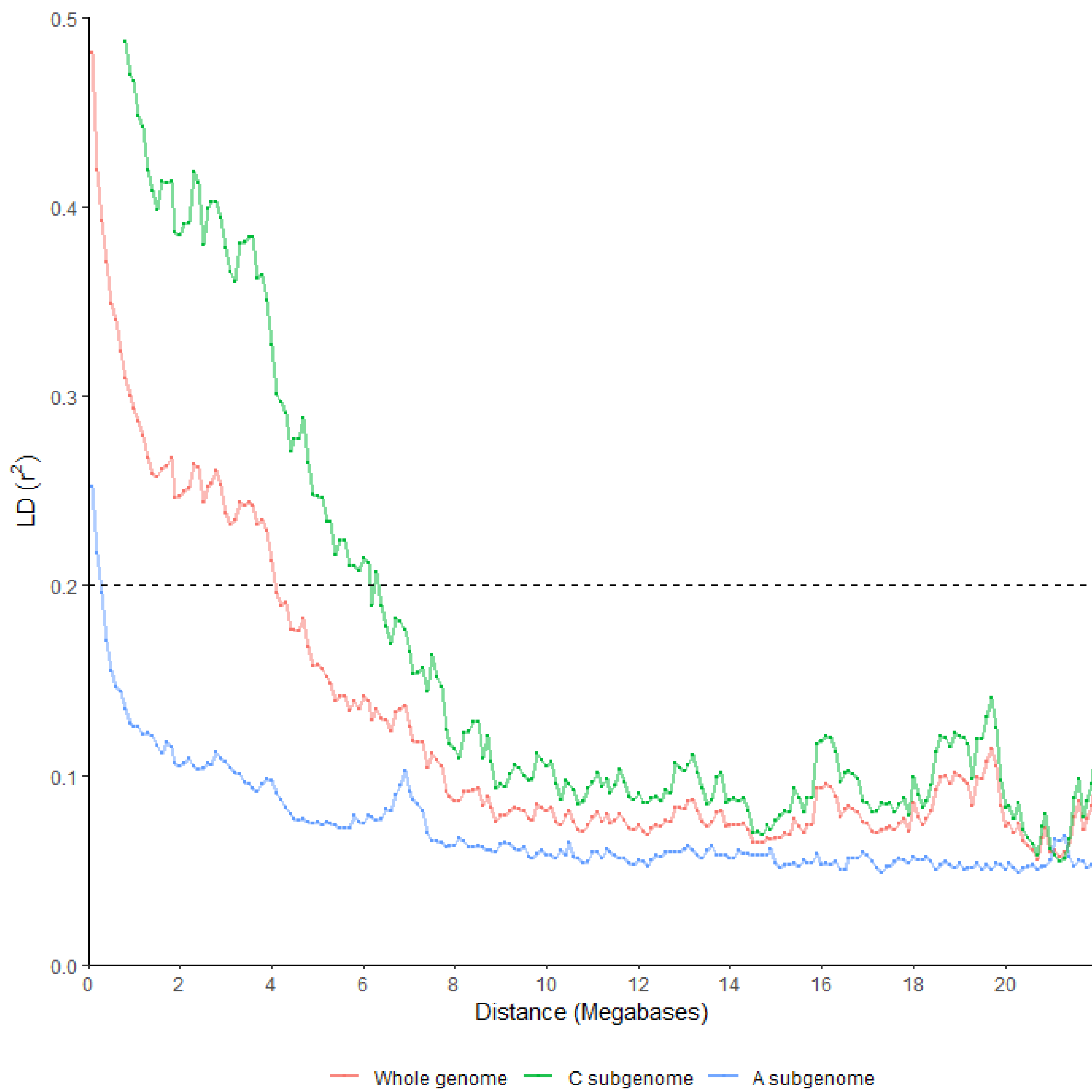

Figure 5.2. Linkage disequilibrium (LD) decay in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019, calculated at the whole genome and subgenome levels, respectively.

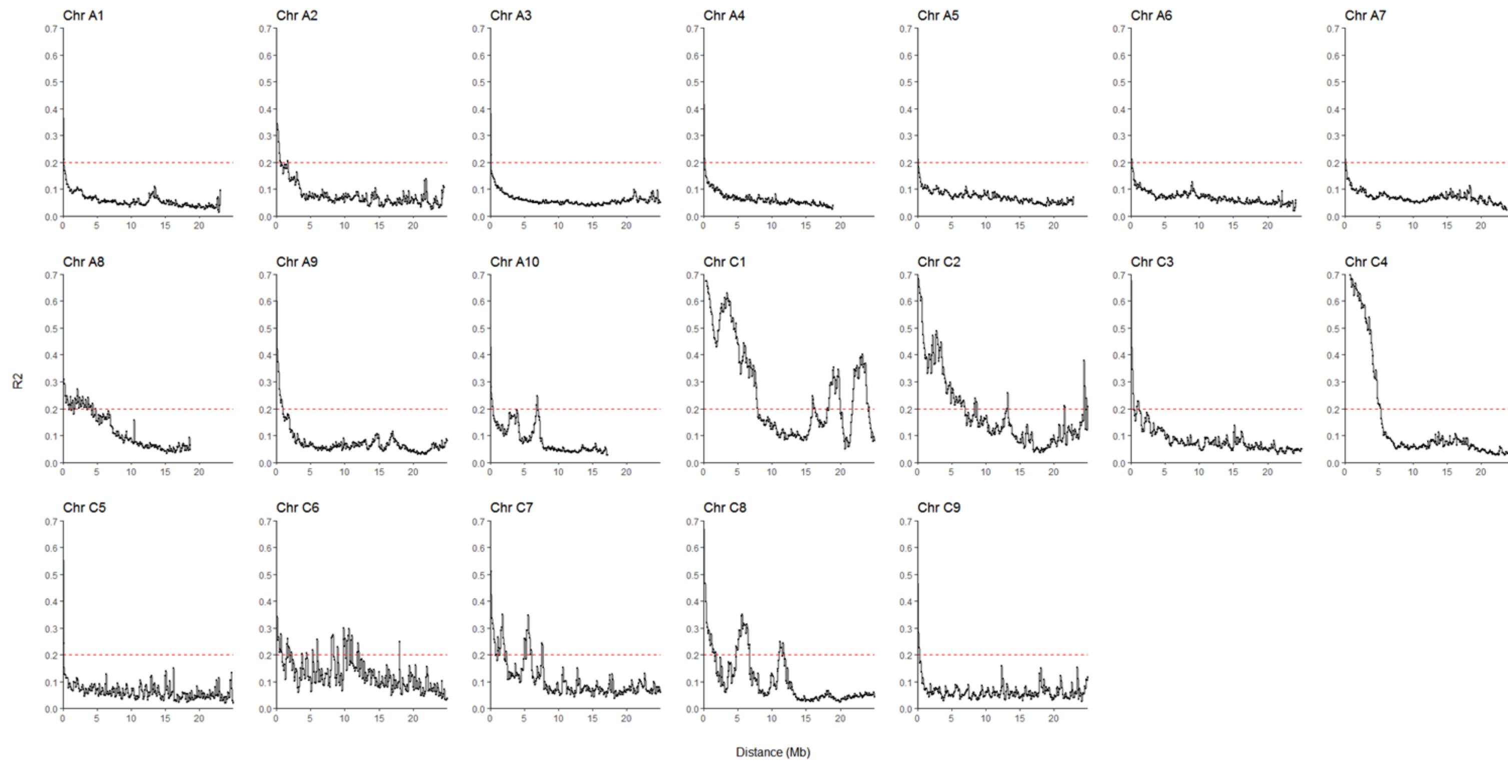

Figure 5.3. Intrachromosomal linkage disequilibrium (LD) decay in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019. The LD decay threshold of $R^2 = 0.2$ is indicated by the dotted red line. The panel shows individual decay plots for each chromosome (Chr) on the A subgenome (Chr A1 to A10) and the C subgenome (Chr C1 to C9).

genome than the A sub-genome, ranging from 250 kb to 8050 kb, with a mean of 2,794 kb and median of 950 kb (Figure 5.3).

5.4.3 Population structure

Population structure was first assessed by Bayesian clustering using a LD-pruned marker set. Six subpopulations were determined to be optimal to account for the population structure of the association panel based on behavior of Delta K relative to K (Figure 5.4). When subpopulation membership was assessed using a 70% threshold, 24 genotypes were assigned to subpopulations while the remaining 19 genotypes were deemed admixtures that did not belong to a subpopulation. The size of each subpopulation ranged from two to five genotypes.

Patterns in the year of introduction, geographical origin, and oil quality were not reflected in the subpopulation assignments (Figure 5.4). Subpopulation structure was then assessed using principal component analysis (PCA) (Figure 5.5) with three principal components. Genotypes did not form identifiable clusters that overlapped with patterns in the year of introduction, geographical origin, or oil quality of the genotypes (Figure 5.5). Principal component analysis did not cluster the genotypes into six subpopulations as determined by Bayesian clustering.

5.4.4 Phenotypic evaluation

The summary statistics of the BLUP values calculated for protein content, oil content, and cruciferin content are presented in Table 5.2. Seed protein content ranged from 10% to 24% of the seed with a mean of 15% and total oil content ranged from 45% to 60% of the seed with a mean of 54%. Cruciferin content of the genotypes were generally lower than that of the reference sample with BLUP values ranging from 24% to 134% relative to the reference sample, with a mean of

Figure 5.4. Subpopulation structure of 43 spring *Brassica napus* L. genotypes as determined by Bayesian clustering (A). Each genotype is represented by a vertical bar and subpopulations are designated k1 to k6. Information about the year of introduction (B), geographic origin (C), and oil quality (D) of each genotype has been superimposed on the horizontal axis. Abbreviations are as follows: AU, Australia; CA, Canada; CN, China; DE, Germany; DK, Denmark; GB, England; SE, Sweden; 00, canola; HEAR, high erucic acid rapeseed.

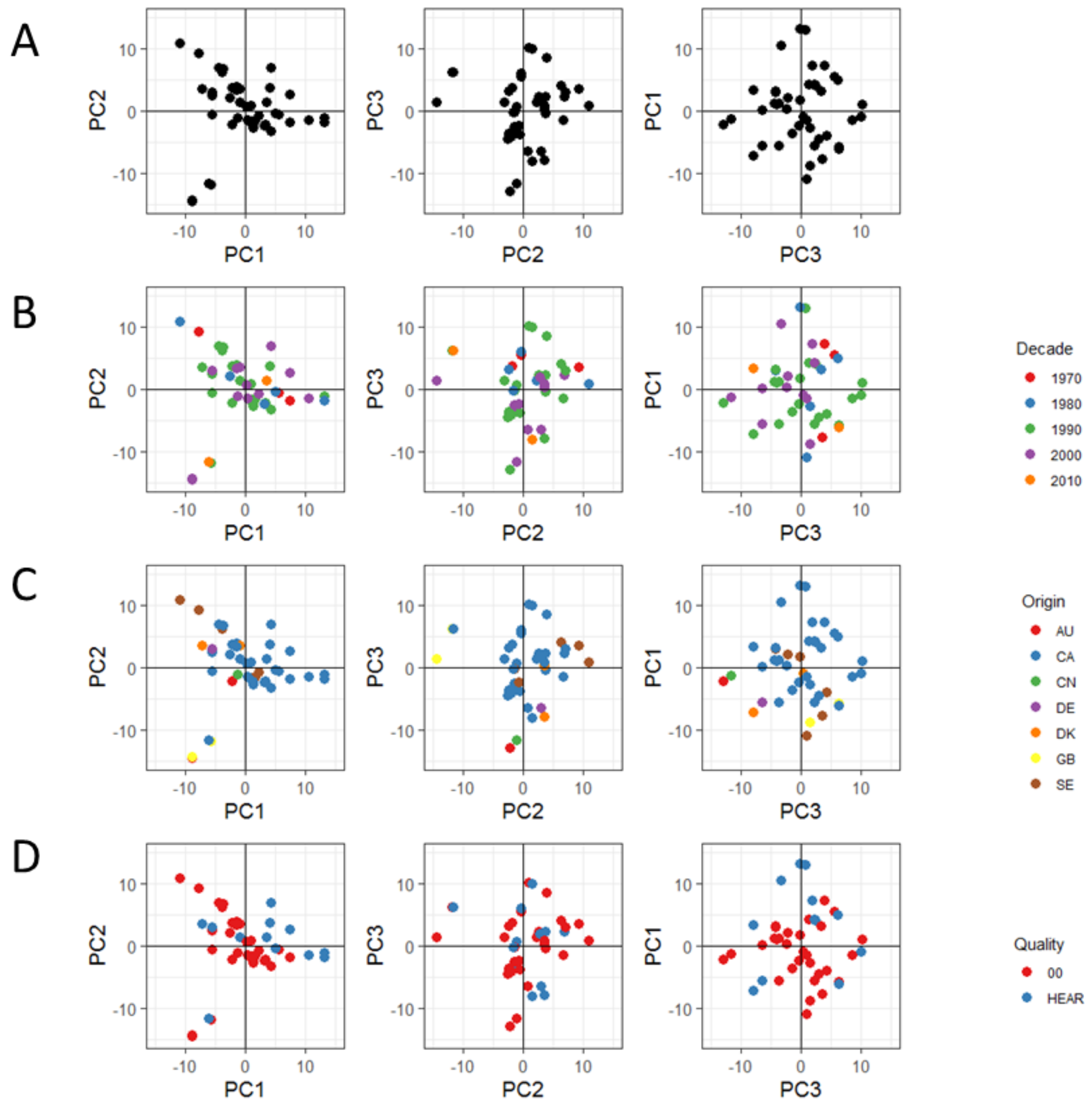

Figure 5.5. Subpopulation structure of 43 spring *Brassica napus* L. genotypes as determined by principal component analysis using three principal components (A). Each datapoint in the biplot represents one genotype. Information about the year of introduction (B), geographic origin (C), and oil quality (D), respectively, of each genotype has been superimposed over each biplot. Abbreviations are as follows: AU, Australia; CA, Canada; CN, China; DE, Germany; DK, Denmark; GB, England; SE, Sweden; 00, canola; HEAR, high erucic acid rapeseed.

Table 5.2. Summary of best linear unbiased predictions for seed protein content, oil content, and cruciferin content in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019. Protein and oil content, respectively, are described as percentage of total, dry seed. Cruciferin content is presented as percentage relative to cruciferin content in a reference sample of *B. napus* 'Westar.' C.V. represent the coefficient of variation.

Trait	Minimum	Maximum	Mean ± SD	C.V.(%)
Protein content (%)	10	24	15± 2	7
Oil content (%)	45	60	54± 2	3
Cruciferin content (%)	24	134	70 ± 19	11

70% . Of the three traits phenotyped, relative cruciferin content measurements showed the greatest variability (Table 5.2).

5.4.5 Model selection for association mapping

A total of six general linear models (GLM) and mixed linear models (MLM) with different consideration for population structure were tested to identify markers associated with seed protein content, oil content, and relative cruciferin content. To determine the most accurate model for candidate gene identification, the mean squared difference (MSD) was used to quantify the mean deviation of the observed *p* values from the null distribution across the six models for each trait. For the three traits of interest, seed protein content, oil content, and cruciferin content, the MSD of the mixed linear models (MLM) were generally lower than those of the general linear models (GLM) with the exception of the MLM + K model for cruciferin content (Figure 5.6).

The inclusion of population structure information in the form of PCA and Q with a MLM was most accurate for seed protein content (MSD = 0.001) and cruciferin content (MSD = 0.005) respectively. In contrast, a MLM without consideration for population structure was the most accurate with a MSD of 0.003 for oil content. Quantile-quantile (QQ) plots visually confirmed the relative accuracy of the different GWAS models for each trait (Figure 5.6). For each trait, the GLM tended to have markers with inflated significance (type one error), the MLM appeared to have over corrected the inflated significance as evidenced by deviation while of the data points below the identity line (Figure 5.6). In the MLM + K model for cruciferin content, a large number of markers with abnormally high association -log(*p*) values ranging from 16.95 to 59.43 were observed. Inflated *p* values were not observed for the same markers under other mixed models

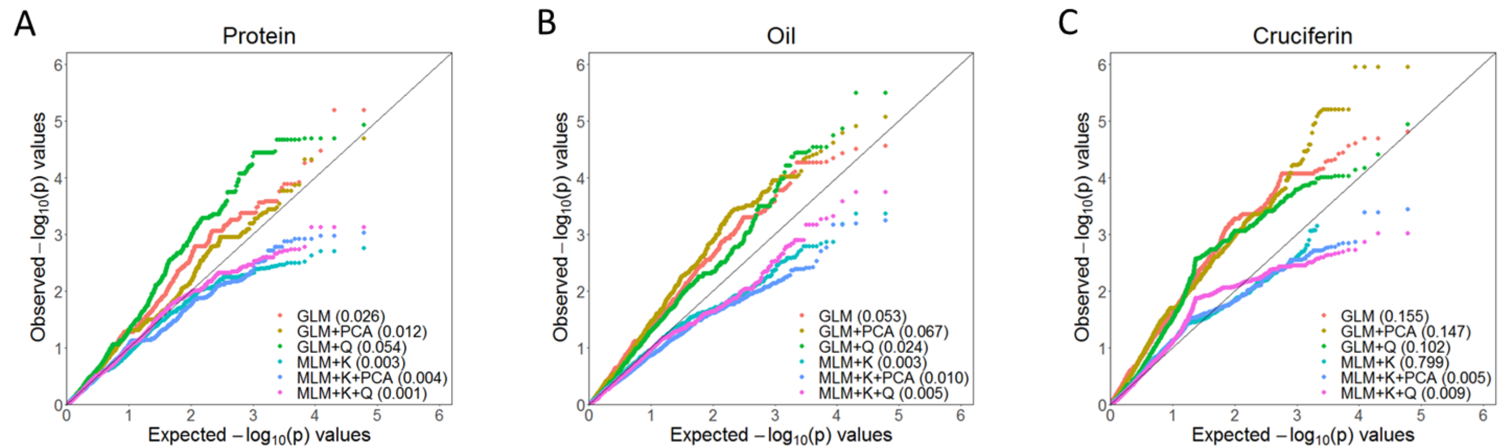

Figure 5.6. Quantile-quantile plots show the distribution of the observed p values plotted against the expected values for different GWAS models in a population of 43 *Brassica napus* L. genotypes grown in the greenhouse at the University of Manitoba in 2019. The mean squared difference of the deviation is presented in parentheses. Plots for seed protein content (A), oil content (B), and cruciferin content (C) are presented. Abbreviations are as follows: GLM, general linear model; MLM, mixed linear model; PCA, principal component analysis; Q, population structure matrix from Bayesian clustering; K, kinship matrix.

with consideration of population structure. Across both the GLM and MLM respectively, the inclusion of population structure data did not appear to significantly improve model accuracy. Taken together, MLM were able to better account for false positives in the association study compared to GLM; however, the inclusion of population structure data in the MLM did not greatly improve accuracy beyond the basal MLM + K model. Marker trait associations for each trait were then examined from the association model with the lowest MSD and Manhattan plots were generated. A list of markers showing suggestive association with each trait of interest cross-referenced to their physical location on the reference genome (Darmor ID) is presented (Appendix Table S1).

5.4.6 Association mapping results for seed protein content

The MLM + K + Q model was the most accurate model with an MSD of 0.001 (Figure 5.6A) and was used for assessing MTA for seed protein content. No significant MTA were detected using either the Bonferroni corrected threshold of $\alpha = 1.65 \times 10^{-6}$ (0.05/n, n = 30, 290) or the relaxed threshold of $\alpha = 3.30 \times 10^{-5}$ (1/n) (Figure 5.7). A third threshold of $\alpha = 0.001$ was implemented to capture suggestive MTA that may be lost due to type two error (false negatives). Four suggestive MTA for seed protein were detected on chromosome C05 (Figure 5.7A).

Of the four suggestive MTA, three markers (Bn-scaff_18181_1-p962029, Bn-scaff_18181_1-p960226, Bn-scaff_18181_1-p954167) were in close physical proximity and appeared as a single point on the Manhattan plot, while the fourth marker (Bn-scaff_18181_1-p1745427) appeared as a distinct location (Figure 5.7B). The three markers in close physical proximity were assigned to haplotype block 54 while the fourth marker was assigned to haplotype block 9 (Figure 5.8A). Haplotype block 54 spanned 8 kb and contained four genes (Figure 5.8B) while haplotype block 9 spanned 22 kb and contained three genes (Figure 5.8A). Genes identified

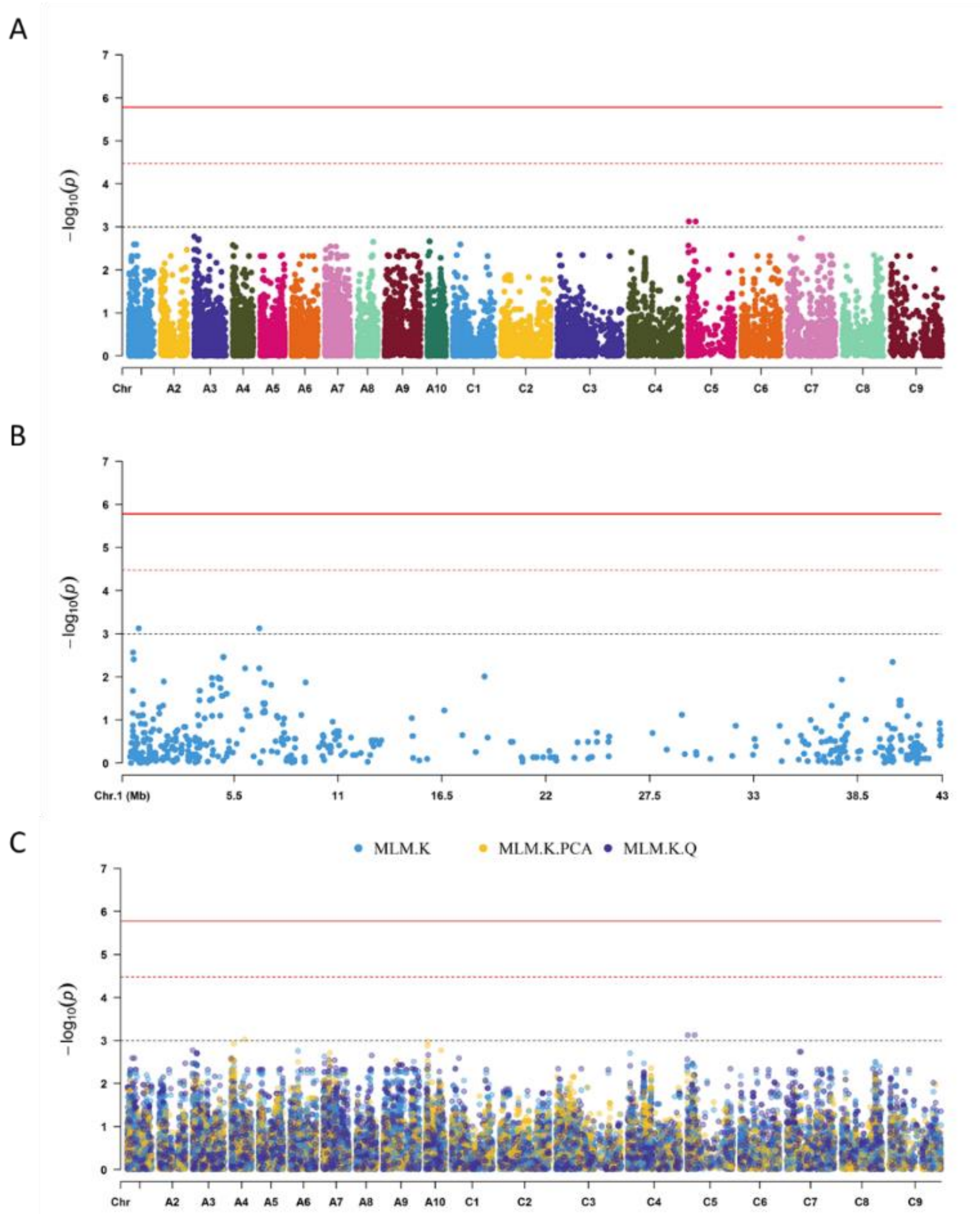

Figure 5.7. Marker-trait associations for seed protein content in 43 *Brassica napus* L. genotypes grown in a greenhouse at the University of Manitoba in 2019. Manhattan plots are presented for the whole genome (A) and chromosome C5 (B). Genome-wide Manhattan plots from various

mixed linear models considering kinship (MLM.K) and subpopulation structure via either principal component analysis (MLM.K.PCA) or Bayesian clustering (MLM.K.Q) are superimposed (C) to show overlaps in the plots. Each dot on the plots represents one single nucleotide polymorphism marker. The solid and dashed red lines represent significance thresholds based on the Bonferonni corrected p = 0.05/30290 and the modified Bonferonni p = 1/30290. The dotted black line represents a suggestive threshold of p = 0.001.

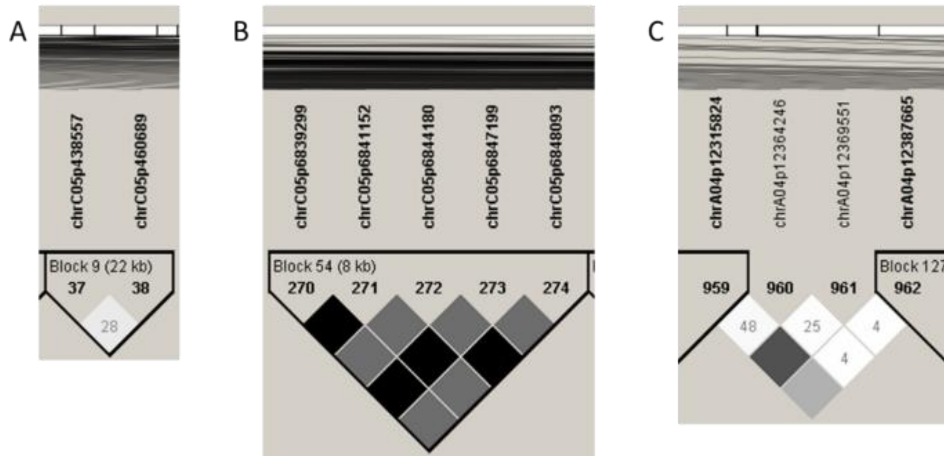

Figure 5.8. Haplotype blocks containing single nucleotide polymorphism markers associated with seed protein content in a panel of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions. The suggestively associated marker Bn-scaff_18181_1-p1745427 (A), and markers Bn-scaff_18181_1-p962029, Bn-scaff_18181_1-p960226, Bn-scaff_18181_1-p954167 (B) were identified by a mixed linear model with Bayesian clustering for subpopulation structure. The marker Bn-A04-p4601980 (C) was identified by a mixed linear model with principal component analysis for subpopulation structure.

within each haplotype blocks or flanking regions and their description are presented in Appendix Table S1.

Due to the low number of suggestive MTA from the MLM + K + Q model, the two other MLM models were considered to identify additional significant or suggestive MTA for seed protein despite their higher MSD values. While no significant MTA were detected using the two established significance thresholds, an additional suggestive MTA was detected using the MLM + K + PCA model on chromosome A04 (Bn-A04-p4601980, $p = 9.39 \times 10^{-4}$) (Figure 5.7C). The significant marker was not in LD with flanking markers and candidate genes were mined from the 54 kb region between the flanking markers: 11 genes were identified in this region (Figure 5.8C). A total of 18 genes were analyzed (Appendix Table S1).

Genes were functionally annotated and GO terms were assigned based on both sequence homology and protein domains. The assigned GO terms for biological process were enriched for transport (GO: 0006810) and defense response (GO: 0006952) while terms for molecular function were enriched for protein binding (GO:0005515) and DNA binding (GO: 0003677). Cellular components showed enrichment for cytoplasm (GO: 0005737). Nutrient reservoir activity, nitrogen metabolism, seed development, abiotic stress and other GO terms directly related to seed protein were not enriched in the list of genes. Four candidate genes were identified based on their similarity to transcription factors known to regulate seed and embryo development: B3 domain-containing transcription factor NGA3 (BnaC05g00870D), zinc finger protein CONSTANS-LIKE 3 (BnaA04g14640D), and agamous-like MADS-box protein AGL61 (BnaA04g14700D, BnaA04g14710D) (Table 5.3).

Table 5.3. Candidate genes for single nucleotide polymorphism markers associated with seed protein content, oil content, and cruciferin content, respectively, in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019.

Trait	Model	Linkage block	Markers	Genes in Block	Description
Protein	MLM+K+Q	9	Bn-chrC05_random-p460689	BnaC05g00870D	B3 domain-containing transcription factor NGA3
	ALL MLM	flanking	Bn-chrA04-p12364246	BnaA04g14640D	zinc finger protein CONSTANS-LIKE 3
				BnaA04g14700D	agamous-like MADS-box protein AGL61
				BnaA04g14710D	agamous-like MADS-box protein AGL61
Oil	MLM+K	flanking	Bn-chrC05_random-p196753	BnaC05g00380D	acyl-CoA thioesterase II
				BnaC05g00450D	glycerol-3-phosphate 2-O-acyltransferase 4-like
Cruciferin	MLM+K+PCA	94	Bn-chrA06-p5218016	BnaA06g09770D	sulfhydryl oxidase 1
			Bn-chrA06-p5250207	BnaA06g09810D	E3 ubiquitin-protein ligase RHA2A-like
			Bn-chrA06-p5258210		
	MLM+K+Q	129	Bn-chrA07-p11076585	BnaA07g12050D	AP2-like ethylene-responsive transcription factor TOE3
			Bn-chrA07-p11107078		

5.4.7 Association Mapping Results for Oil Content

The MLM + K model was used to identify significant and suggestive MTA for oil content as it had the lowest MSD (Figure 5.6B). While no significant MTA were detected, three suggestive MTA were detected on chromosomes C5 (Figure 5.9A). Two of the suggestive markers Bn-scaff_15712_10-p49746 ($p = 4.32 \times 10^{-4}$) and Bn-scaff_15712_10-p52253 ($p = 4.32 \times 10^{-4}$) occurred in haplotype block 51 (Figure 5.10B) while the third marker Bn-scaff_15712_10-p199582 ($p = 6.61 \times 10^{-4}$) was not found to be in LD with any flanking markers (Figure 5.10A). Haplotype block 51, which spanned 160 kb, contained 15 genes. The region between the markers that flanked Bn-scaff_15712_10-p199582 was 111 kb in length and contained 30 genes (Appendix Table S1).

In addition to the selected MLM + K model, two additional MLM models with consideration for population structure were examined to identify additional significant and suggestive markers. No significant markers were detected using the two additional models, but additional suggestive MTA were detected on chromosomes A07 (Bn-A07-p3921656, Bn-A07-p3929927, Bn-A07-p3930035, Bn-A07-p3930092, and Bn-A07-p3931338) and C05 (Bn-scaff_15712_10-p173981 and Bn-scaff_15712_10-p49537) (Figure 5.9B). Suggestive MTA were detected on chromosome C05 under all three MLM models (Figure 5.9C) and all localized to haplotype block 51 (Figure 5.10B); candidate gene analysis therefore focused on this chromosome.

GO terms for biological activity were enriched for cellular metabolic process (GO: 0044237) and transmembrane transport (GO: 0055085). GO terms for molecular function were enriched for protein binding (GO: 0005515) and hydrolase activity (GO: 0016787) suggesting that genes of interest identified by the suggestive MTA may be involved in the post-translational

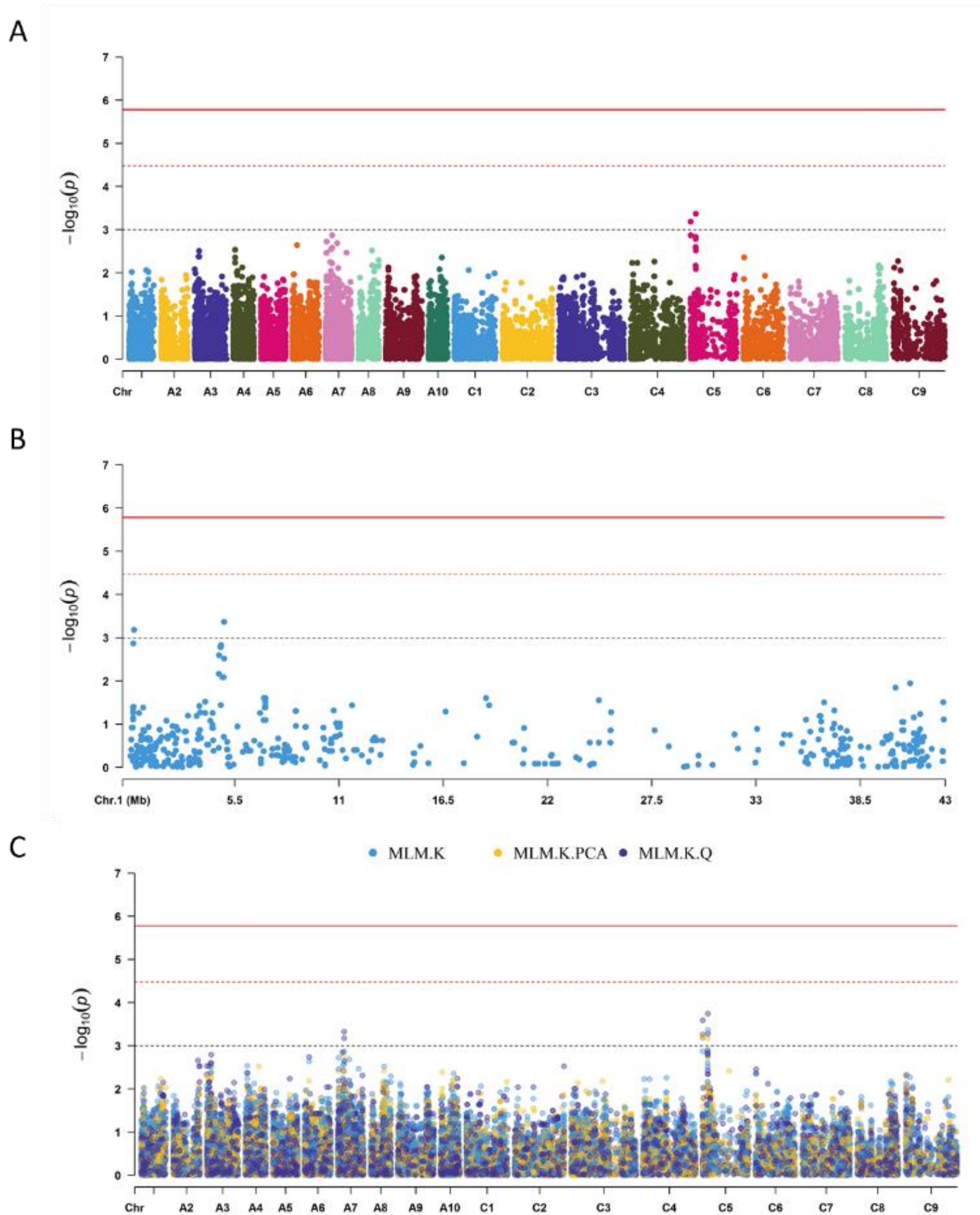

Figure 5.9. Marker-trait associations for oil content in 43 *Brassica napus* L. genotypes grown in a greenhouse at the University of Manitoba in 2019. Manhattan plots are presented for the whole genome (A) and chromosome C5 (B). Genome-wide Manhattan plots from various mixed linear

models considering kinship (MLM.K) and subpopulation structure via either principal component analysis (MLM.K.PCA) or Bayesian clustering (MLM.K.Q) are superimposed (C) to show overlaps in the plots. Each dot on the plots represents one single nucleotide polymorphism marker. The solid and dashed red lines represent significance thresholds based on the Bonferonni corrected $p = 0.05/30290$ and the modified Bonferonni $p = 1/30290$. The dotted black line represents a suggestive threshold of $p = 0.001$.

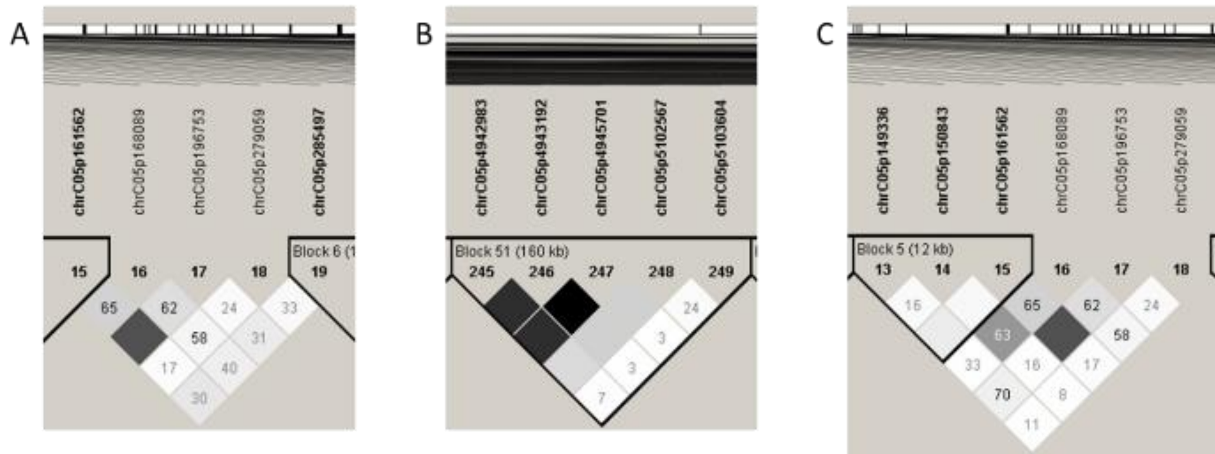

Figure 5.10. Haplotype blocks containing single nucleotide polymorphism markers associated with oil content in a panel of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions. Markers identified by a mixed linear model were either not in linkage disequilibrium with neighbouring markers (chrC05p196753) (A) or clustered in a separate block (chrC05p4942983, chrC05p4943192, chrC05p4945701) (B). Additional markers identified across all mixed linear models were located within a separate haploblock (c).

regulation of oil content. GO terms related to fatty acid metabolism, oil body, and other oil-related terms were not enriched in the list of genes. Two genes related to fatty acid metabolism were identified: acyl-CoA thioesterase II (BnaC05g00380D) and glycerol-3-phosphate 2-O-acyltransferase 4-like (BnaC05g00450D) (Table 5.3).

5.4.8 Association Mapping Results for Relative Cruciferin Content

The MSD of the MLM + K + PCA model was the lowest across the six models used to test MTA with cruciferin content (Figure 5.6C) and was thus selected as the most accurate model for the trait. Similar to protein and oil content, no significant MTA associations were detected for cruciferin content; however, three suggestive MTA (Bn-A06-p5780982, Bn-A06-p5693735, and Bn-A06-p5830298) were identified on chromosome A06 (Figure 5.11A) in close proximity to each other (Figure 5.11B). All three markers were assigned to haplotype block 94 (Figure 5.12A). Ten genes were detected within the 40 kb haplotype block (Appendix Table S1).

When all MLM models were considered, additional suggestive MTA were detected across multiple chromosomes. Under the MLM + K + Q model, two additional suggestive MTA in close physical proximity on chromosome A7 (Bn-A07-p9791002, $p = 9.73 \times 10^{-4}$; Bn-A07-p9831545, $p = 9.73 \times 10^{-4}$) were detected (Figure 5.11C). Using the MLM + K model, five suggestive MTA were detected on chromosome A06, three of which had previously been detected using the MLM + K + PCA model (Figure 5.11C). The two remaining suggestive MTA are located approximately 300 kb from the three detected previously. Fourteen markers with abnormally high association p values ranging from $p = 1.13 \times 10^{-17}$ ($-\log(p) = 16.95$) to $p = 3.74 \times 10^{-60}$ ($-\log(p) = 59.43$) were observed in the model. Three markers with abnormally high association significance were also located on chromosome A06 and flanked either side of the block of markers with suggestive

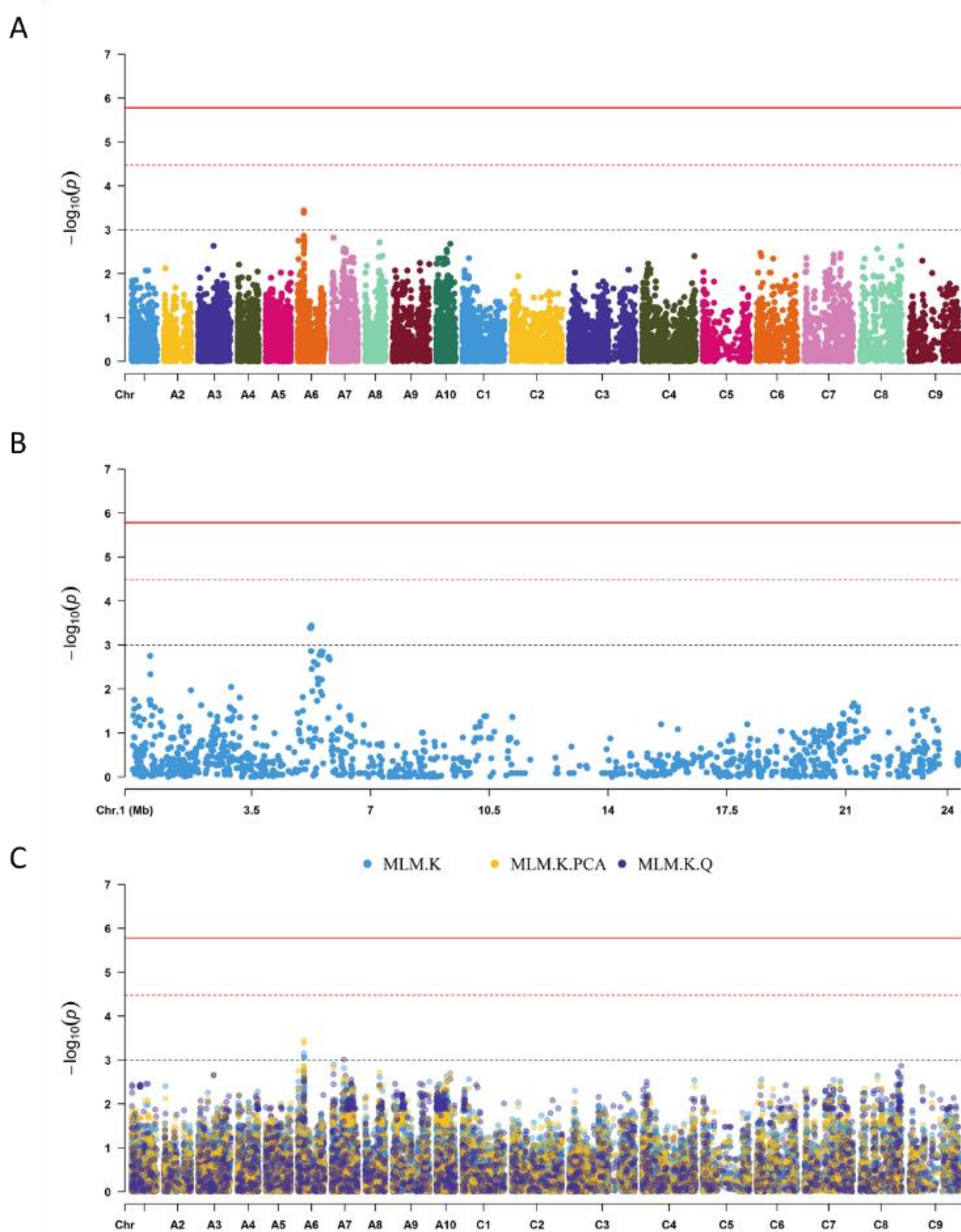

Figure 5.11. Marker-trait associations for cruciferin content in 43 *Brassica napus* L. genotypes grown in a greenhouse at the University of Manitoba in 2019. Manhattan plots are presented for the whole genome (A) and chromosome A6 (B). Genome-wide Manhattan plots from various

mixed linear models considering kinship (MLM.K) and subpopulation structure via either principal component analysis (MLM.K.PCA) or Bayesian clustering (MLM.K.Q) are superimposed (C) to show overlaps in the plots. Each dot on the plots represents one single nucleotide polymorphism marker. The solid and dashed red lines represent significance thresholds based on the Bonferonni corrected $p = 0.05/30290$ and the modified Bonferonni $p = 1/30290$. The dotted black line represents a suggestive threshold of $p = 0.001$. Markers with strongly inflated significance values ($-\log(p)>7$) were removed before plotting.

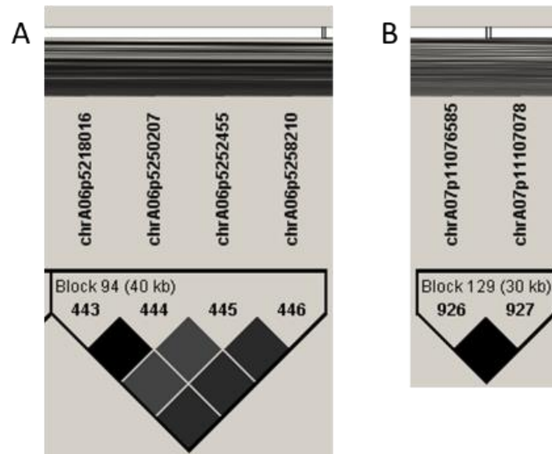

Figure 5.12. Haplotype blocks containing single nucleotide polymorphism markers associated with cruciferin content in a panel of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions. The suggestively associated markers Bn-A06-p5780982, Bn-A06-p5693735, and Bn-A06-p5830298 were identified by a mixed linear model with consideration for population structure through principal component analysis (A). Additional markers (Bn-A07-p9791002 and Bn-A07-p9831545) were identified by a mixed linear model with consideration for population structure through Bayesian clustering.

association. Due to the apparent, anomalously inflated association significance, these markers were not considered for candidate gene identification and the y- axis of the Manhattan plot was limited to –log(p) = 7. GO terms for biological processes were enriched for regulation of transcription, DNA-templated (GOL 0006355) and oxidation-reduction process (GO: 0055114) while GO terms for molecular function were enriched for DNA binding (GO: 0003677) and oxidoreductase activity (GO: 0016491). Three candidate genes were identified by assigned GO terms: the first, (BnaA06g09770D), was identified based on its protein disulfide isomerase activity; the second, E3 ubiquitin-protein ligase RHA2A-like (BnaA06g09810D), was identified based on its role in the positive regulation of abscisic acid (ABA)-activated signalling; and the third, AP2-like ethylene-responsive transcription factor TOE3 (BnaA07g12050D), was identified based on its similarity to the AP2 transcription factor (Table 5.3).

5.5 Discussion

An association study was conducted on a population of spring-type *B. napus* from the University of Manitoba Brassica collection to identify genomic regions associated with seed protein content, oil content, and cruciferin content in mature seeds. To date, this study is the first to explore genetic markers associated with levels of the seed storage protein cruciferin in spring-type germplasm.

For cruciferin content, markers with suggestive association did not coincide with QTL on linkage groups N2 (Chr A02) and N19 (Chr C09) for cruciferin content previously identified in winter *B. napus* (Schatzki et al. 2014) and genes identified based on haplotype blocks did not contain known cruciferin biosynthesis genes. Therefore, the loci identified from this study may be considered novel and candidate genes can be inferred to function in the indirect control of cruciferin content. Candidate genes with activities in different aspects of protein synthesis and

peptide maturation were identified: disulfide bond formation, ABA response, and a transcription factor similar to *APETALA2* (*AP2*). Cruciferin is a hexameric protein whose subunits are composed of an alpha and a beta chain linked together via disulfide bonds (Rödin and Rask 1990; Shewry et al. 1995; Jung et al. 1997). Disulfide bond formation in seed storage proteins is essential for proper sorting at the endoplasmic reticulum (Kawagoe et al. 2005) and defects may hinder their accumulation; therefore, the candidate gene *BnaA06g09770D* may function through the control of endomembrane sorting which precedes maturation and subsequent accumulation of hexameric cruciferin. Further, cruciferin expression and accumulation in *B. napus* require the presence of ABA (Crouch and Sussex 1981; Finkelstein et al. 1985) and *BnaA06g09810D* may function to promote the continued synthesis and accumulation of the protein. Specifically, the *AP2/EREBP* class of transcription factors has been implicated in the control of seed mass and the production of both oil and protein (Cernac and Benning 2004; Jofuku et al. 2005; Ohto et al. 2009) in addition to the regulation of ABA (Shu et al. 2018). Thus, *BnaA07g12050D* was also considered a candidate gene for the control of cruciferin content. Taken together, candidate genes were identified from suggestive MTA with ontology terms that reasonably implicate their role in the indirect regulation of protein, oil, and cruciferin contents, respectively.

Storage proteins constitute the majority of the seed protein in seeds and their expression is under strict transcriptional regulation (Fujiwara et al. 2002; Verdier and Thompson 2008). The candidate genes identified from suggestive markers associated with protein content showed homology to transcription factors known to regulate seed and embryo development. The *NGATHA* (*NGA*) transcription factors have been reported to play a role in leaf and flower development (Swaminathan et al. 2008), and *NGA1* is involved in the regulation of abscisic acid (ABA) production; however, the role of *NGA3* in storage protein regulation has not been reported. Other

transcription factors containing a B3 domain such as *LEAFY COTYLEDON*2, *FUSCA3*, and *ABSCISIC ACID INSENSITIVE3* are known to regulate the expression of seed storage protein genes (Kroj et al. 2003; Kagaya et al. 2005; Swaminathan et al. 2008); perturbations in these transcription factors result in altered seed protein content (Fujiwara et al. 2002; Wang et al. 2007; Angeles-Núñez and Tiessen 2011). The second candidate transcription factor *CONSTANS-LIKE 3* regulates light response (Datta et al. 2006) and flowering (Tripathi et al. 2017) but has also not been implicated in the direct control of storage protein in the seed. Finally, *AGAMOUS-LIKE61* (*AGL61*) is a MADS-box protein that regulates the development of the central cell in the female gametophyte (Steffen et al. 2008). Broadly, MADS-box transcription factors regulate many aspects of seed development (Smaczniak et al. 2012) and partake in ABA homeostasis (Castelán-Muñoz et al. 2019); therefore, they may indirectly control seed protein content. Specifically, in corn the MADS-box protein ZmMADS47 has been shown to bind directly to seed storage protein genes and jointly activates transcription of the gene with other transcription factors (Qiao et al. 2016).

Regarding oil content, two candidate genes related to fatty acid metabolism were identified from the association study: acyl-CoA thioesterase II and glycerol-3-phosphate 2-O-acyltransferase 4-like. Triglycerides consist of three fatty acids attached to a glycerol backbone and are the most prevalent form of lipid in seeds (Graham 2008; Manan et al. 2017). Acyl-CoA are fatty acids that are conjugated to coenzyme A via an ester bond (Graham 2008). Although acyl-CoA thioesterases catalyze the breakdown of fatty acyl-CoA into free fatty acids and coenzyme A (Cassin-Ross and Hu 2014), they are not believed to be involved in lipid oxidation in plants (Tilton et al. 2004) and their exact physiological function remains unclear (Graham 2008). While contemporary works demonstrated defects in acyl-CoA thioesterase activity and the introduction of foreign orthologues

into canola resulted in altered fatty acid profiles, these studies suggest the proteins function in part to regulate the species composition of the fatty acid pool (Scarth and Tang 2006; Cassin-Ross and Hu 2014) rather than exert direct control over oil production. Glycerol-3-phosphate (G3P) is comprised of a three-carbon glycerol molecule with a phosphate group attached; this molecule is required as a substrate for the biosynthesis of membrane lipids and triglycerides (Manan et al. 2017). Acyl-CoA are attached to G3P through the action of various glycerol-3-phosphate acyltransferases (GPAT) each of which show preference for different substrates (Chen et al. 2011a; Yang et al. 2012). These enzymes are therefore directly involved in the biosynthesis of triglycerides and can by extension directly regulate seed oil content. Specifically, GPAT4 catalyzes the attachment of acyl-CoA to the *sn-2* position of G3P and may also play a role in the synthesis of lipid polymers (Yang et al. 2010; Chen et al. 2011b).

Population structure in cultivated crops arising from domestication and selection (Huang and Han 2014) can confound association studies as markers associated with subpopulation structure can be misinterpreted as being associated with the trait of interest (Lander and Schork 1996; Liu and Yan 2019). Multiple methods have been proposed to mitigate the effect of population structure such as the use of mixed linear models and the inclusion of a q matrix representing the population structure in the model (Price et al. 2010); our results confirmed that these methods were more accurate than the naïve general linear model approach and at times resulted in overcorrection (Kaler et al. 2020). Although the use of Bayesian clustering and PCA have been suggested to be equally effective in accounting for population structure (Zhao et al. 2007) and often corroborate each other (Wang et al. 2012; Qian et al. 2014), the results of both analyses conducted on our population were incongruous, with the former identifying six subpopulations while no clear clusters were identified by the latter.

Subpopulation structure in *B. napus* coincides with patterns in geographical origin and growth habit (Gyawali et al. 2013; Gazave et al. 2016; Rahman et al. 2021), but has also been observed to coincide with historic development periods (Fu and Gugel 2010) in other crops (Sim et al. 2011; van Heerwaarden et al. 2012). Contrary to this, our study found concordance of the subpopulation structure with neither geographic origin, development period, nor seed quality. The lack of separation between the high and low erucic acid cultivars is likely attributed to the simple genetic differences that control the trait (Harvey and Downey 1964; Rahman et al. 2008), while selection for all other agronomic traits remained similar. This apparent lack of subpopulation structure should be interpreted with caution given the small number of genotypes in each subpopulation.

Recognizing that association studies in *B. napus* have employed from 66 (Zhang et al. 2015) to 950 genotypes (Jan et al. 2019) in the association panel, the population size employed in this study is undoubtedly suboptimal which likely resulted in a paucity of MTA. The population size was dictated by the difficulty in phenotyping cruciferin content for replicated experiments. The small population size may be exacerbated by the short domestication history of the crop having undergone intense selection for oil quality (Qian et al. 2014), possibly leading to the paucity of marker trait associations detect in the experiment. Nonetheless, this study represents a preliminary exploration into the genetic control of cruciferin content in spring *B. napus* and the results presented herein can be extended with additional studies employing larger population sizes. Finally, the use of multi-parent mapping populations should be considered to allow for the detection of rare, causal markers that directly control cruciferin content and removal of the confounding effects of subpopulation structure (Scott et al. 2020).

5.6 Conclusion

The presence of genetic variation in cruciferin content within spring *B. napus* enables the improvement of cruciferin content through traditional breeding. However, the genetic basis underlying this variation has previously not been explored due to the unavailability of suitable mapping populations in the crop. As such, an association mapping approach was undertaken to identify genomic regions that were associated with cruciferin content in a collection of spring *B. napus* that spanned the cultivar development history of the crop. From the study, candidate genes in linkage disequilibrium with markers associated with oil, protein, and cruciferin content were identified based on gene ontology terms related to their respective biosynthetic processes. Specifically, BnaA06g09770D, BnaA06g09810D, and BnaA07g12050D may play a role in the indirect control of cruciferin content through regulating disulfide bond formation, ABA signalling, and transcriptional control. This study is the first to evaluate and explore the genetic loci governing variation in cruciferin content in spring *B. napus* and may offer novel targets for the development of molecular markers to be employed in marker-assisted or genomic selection the trait.

In the previous chapter, genomic regions that were associated with cruciferin content were identified, and candidate genes within the identified regions likely functioned to indirectly mediate cruciferin accumulation. One limitation to the methods employed previously was that potential genomic regions of study were narrowed by the availability of molecular markers on the genotyping platform. To overcome this limitation, the subsequent work uses an unbiased transcriptomics approach to explore all differentially expressed genes in the seed to identify transcriptional patterns that correlate with improved cruciferin accumulation.

Chapter 6. Transcriptomic differences during early seed development of *Brassica napus* L. cultivars divergent in cruciferin content

6.1 Abstract

Protein from *Brassica napus* L. seed has potential as a novel and sustainable source of dietary protein. Specifically, the seed storage protein cruciferin has the potential to function as a structural ingredient in commercial food processing. Variation in cruciferin content exists across *B. napus* germplasm, yet the molecular mechanisms that underlie this variation have not previously been explored. To this end, a transcriptomics approach was undertaken to explore differential transcriptional patterns within the whole seed of two spring *B. napus* genotypes divergent in cruciferin content from 15 to 25 days after pollination (DAP) representing the onset of cruciferin accumulation. Sequencing reads had a mean overall alignment rate of 95%. Multidimensional scaling delineated sequencing libraries by genotype at each time point. Transcripts of cruciferin biosynthetic genes were not detected until 25 days after pollination, indicating cruciferin accumulation was initiated at this time. Further, cruciferin transcripts were only detected in the low-cruciferin genotype indicating precocious expression of cruciferin biosynthetic genes did not result in increased cruciferin content. Gene ontology analysis indicated differentially expressed genes (DEG) for the high-cruciferin cultivar 'Red River 1852' at 25 DAP were enriched for polyamine metabolism and glucan synthase; GO terms for DEG in the low cruciferin genotype 'Wizzard' at the same time point were enriched for photosynthesis-related processes and nutrient reservoir activity. These results represent an initial exploration into the molecular mechanisms that modulate differential cruciferin accumulation in *B. napus* and DEG are candidates for molecular marker development to facilitate the breeding improvement of cruciferin content.

6.2 Introduction

Consumer awareness for health and sustainability has furthered the development of plant proteins as an environmentally-friendly alternative to their animal counterparts (Perignon et al. 2017; Röös et al. 2017b; Shepon et al. 2018). To date, soy protein is the most prevalent form of plant protein available on the market (Kumar et al. 2017; Nishinari et al. 2018), but interest has burgeoned in the development of canola/rapeseed (*B. napus)* protein as both a dietary protein source and food ingredient (Mupondwa et al. 2018). Globally, *B. napus* is the fourth largest oilseed crop with over 18 metric tonnes harvested in Canada alone in 2020 (Food and Agriculture Organization of the United Nations 2020). Considering protein accounts for approximately 20% of the total seed, equivalent to 38% of the oil-free meal (Barthet 2019), canola protein is abundant and thus, a logical candidate for development into a sustainable protein source.

Seed storage compounds in the form of protein and oil serve as a reservoir of nutrients to support germination (Shewry et al. 1995). The biosynthesis and accumulation of these compounds are temporally and spatially confined to the developing seed (Shewry et al. 1995). Although extensive research has focused on understanding the molecular mechanisms that regulate oil quality (Scarth and Tang 2006; Hannoufa et al. 2014) and quantity (Li et al. 2014; Sun et al. 2016b) in *B. napus*, there is a relative paucity of equivalent knowledge on seed storage protein (SSP).

The SSP pool in *B. napus* primarily consists of two proteins: cruciferin and napin (Crouch and Sussex 1981). The globulin-type SSP cruciferin accounts for approximately 60% of the seed protein in *B. napus* (Crouch and Sussex 1981) and has functional properties that make it amenable for a variety of food processing applications (Malabat et al. 2001; Wanasundara et al. 2016b). The subunits that make up mature, hexameric cruciferin are encoded by a family of five genes (Sjödahl et al. 1993; Wanasundara 2011) and their transcriptional activation have previously been described

(Kroj et al. 2003; Verdier and Thompson 2008; Fauteux and Strömvik 2009). Despite the coordinated expression of cruciferin biosynthesis genes, variation in both total cruciferin content and cruciferin subunit composition is present in *B. napus* (Sjödahl et al. 1991; Wanasundara 2011; Withana-Gamage et al. 2013) and in the SSP of other crops (Poysa et al. 2006). This variation alludes to potentially multiple levels of regulation: at the genomic level, sequence polymorphisms and copy number variation can affect the accumulation of transcripts (Gabur et al. 2019), while post-transcriptional modifications can attenuate translation of the gene product (Walling et al. 1986; Verdier and Thompson 2008; Kawakatsu and Takaiwa 2010). Additionally, the presence of post-translational motifs within the peptide sequence of cruciferin further suggests the regulation of cruciferin processing, accumulation, and catabolism at the protein level (Jung et al. 1998; Gruis et al. 2002; Jolivet et al. 2009; Mylne et al. 2014). Finally, SSP expression is known to be modulated by environmental conditions (Brunei-Muguet et al. 2015; Tan et al. 2020) and a role for hormonal regulation has been implicated in early studies on SSP in *B. napus* (Finkelstein et al. 1985; Ezcurra et al. 2000).

Although the expression of cruciferin biosynthesis genes is largely coordinated within individual genotypes (Sjödahl et al. 1993), variation in seed development and maturation periods between genotypes may result in prolonged cruciferin production which in turn translates to increased cruciferin content in the mature seed. Variation in when cruciferin transcripts and protein first accumulate to detectable levels has been observed: specifically, cruciferin transcripts were first detectable between 21 to 25 days after pollination (DAP) by Northern blotting (Crouch and Sussex 1981; Finkelstein et al. 1985; Murphy and Cummins 1989; Murphy et al. 1989; Sjödahl et al. 1993). Alternatively, upregulation of SSP-related regulatory pathways may serve to indirectly modulate differential cruciferin accumulation. Thus, an understanding of the molecular

mechanisms that control SSP can be used to develop molecular markers to facilitate the breeding improvement of SSP content.

The albumin-type SSP napin is the second most abundant protein in the seed protein of *B. napus*, accounting for approximately 20% of the total pool (Crouch and Sussex 1981); despite its ability to replace casein (Schwartz et al. 2015), napin has been shown to be a potential allergen (Monsalve et al. 1997; Moreno and Clemente 2008; Marambe et al. 2015; Fiocchi et al. 2016). Therefore, the adoption of cruciferin as dietary protein or as a functional ingredient requires its separation from napin and the residual seed material. Industrial purification of cruciferin requires technical infrastructure (Bérot et al. 2005), adding cost to the product which hinders its adoption. To this end, the development of high-cruciferin cultivars can increase the cost effectiveness of protein separation to make cruciferin an economically competitive protein ingredient.

Crop cultivars with specialized seed storage protein composition have been developed previously through both biotechnology (Kohno-Murase et al. 1994, 1995; Kim et al. 2014; Barro et al. 2016; Rolletschek et al. 2020) and traditional breeding (Poysa et al. 2006), though few have been brought to commercialization. The successful generation of the above cultivars portends the potential for high-cruciferin *B. napus* cultivars to be developed, and the use of traditional breeding to improve cruciferin content enables access to diverse global markets.

To initiate the breeding improvement of cruciferin content, a survey of a subset of the University of Manitoba germplasm collection found *B. napus* 'Red River 1852' to be a high cruciferin cultivar (93% of the check cultivar 'Westar') while 'Wizzard' was identified as a low cruciferin cultivar (44% of the check cultivar 'Westar') (Chapter 4). A genome-wide association study subsequently identified single-nucleotide polymorphism markers that were significantly

associated with cruciferin content; the markers were not in linkage disequilibrium with known cruciferin biosynthesis genes and likely functioned in the indirect control of cruciferin accumulation (Chapter 5). To further characterize these divergent genotypes, the molecular mechanisms that underlie this trait require additional study. To explore potential transcriptional mechanisms that control cruciferin accumulation, a comparative transcriptomics approach was undertaken to observe differential gene expression at 15, 20 and 25 DAP, representing the early seed maturation phase at which cruciferin accumulation is initiated, between the high-cruciferin genotype ‘Red River 1852’ and the low-cruciferin genotypes ‘Wizzard.’

6.3 Materials and Methods

6.3.1 Plant material and growth conditions

Six plants each of the high cruciferin *Brassica napus* cultivar ‘Red River 1852’ and low cruciferin content cultivar ‘Wizzard’ were grown in a growth chamber at the Department of Plant Science, University of Manitoba, Winnipeg, Canada. Plants were cared for as previously described in Chapter 2 (Section 2.3) and were periodically rotated randomly within the growth chamber to minimize within-chamber variation in light and temperature (Measures et al. 1973; Porter et al. 2015). Growth chamber parameters were set to a 16-hour photoperiod and daily temperatures of 20/18°C. At bolting (BBCH growth stage 55) (Lancashire et al. 1991) three plants of each cultivar that appeared most uniform were selected for future sampling.

6.3.2 Sample preparation and RNA sequencing

On the day of anthesis, flowers were self-pollinated and tagged with the date of pollination using a label on the pedicel. This was repeated daily over a two-week period. At 15, 20, and 25 DAP, the developing seeds within 10 – 12 siliques of each plant were dissected out using a

disposable scalpel and immediately frozen in liquid nitrogen. Tissue samples were stored at -80°C until processing. In total, 18 samples were collected from three biological replicates of each of the two genotypes for each of the three time points.

Total RNA from 10 – 12 seeds was extracted using a RNeasy Plant Mini Kit with on-column DNase treatment (Qiagen, Hilden, Germany). RNA concentration was quantified using a Nanodrop 2000 (ThermoFisher Scientific. Waltham, U.S.A.) and quality was assessed by gel electrophoresis on a 2% agarose gel (Rong et al. 2021). RNA samples were sent to Genome Quebec (Montreal, Canada, http://gqinnovationcenter.com/) for quality control, library preparation, and sequencing: stranded mRNA libraries were prepared using NEBNext Dual adaptors (New England Biolabs, Ipswich, U.S.A.) and sequenced on the Illumina NovaSeq 6000 platform (https://www.illumina.com/systems/sequencing-platforms/novaseq.html) to generate 100-base pair, paired-end reads. An aliquot of each RNA sample was reserved for validation by reverse transcription - quantitative Polymerase Chain Reaction (RT-qPCR) (Section 6.3.4).

6.3.3 Data analysis

Raw read files were downloaded from Genome Quebec through the Nanuq web portal (https://cesgq.com/en-nanuq). The quality of the raw reads was assessed using FastQC version 0.11.9 (Andrews 2010). Adapters were trimmed using Trimmomatic version 0.39 (Bolger et al. 2014) with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Trimmed reads were re-analyzed using FastQC to confirm trimming was successful. Transcript alignment and abundance estimation were performed following Pertea et al. (2016). Briefly, trimmed reads were aligned to the *B. napus* genome version AST_PRJEB5043_v1.dna.toplevel from the ENSEMBL database (Yates et al. 2020) using HISAT2 (Kim et al. 2019a). The resultant SAM files were converted to

sorted BAM format using StringTie2 (Kovaka et al. 2019). Transcript abundances were estimated directly against the *B. napus* reference annotation version AST_PRJEB5043_v1.49 from ENSEMBL (Yates et al. 2019) using StringTie2 (Kovaka et al. 2019) with the -e and -B options. The count data were imported into R Studio version 1.2.1335 (RStudio Team 2020) using the tximport package (Soneson et al. 2016).

Differential gene expression was analyzed between the high and low-cruciferin genotypes separately for each time point following the protocol of Law et al. (2018). In short, genes that showed low expression were first filtered out using default parameters and normalized using the trimmed mean of M-values method (Robinson and Oshlack 2010), both implemented in edgeR (Robinson et al. 2009; McCarthy et al. 2012), before a voom transformation (Law et al. 2014) was applied. Fold change between the two genotypes was assessed by fitting a linear model for each gene using the limma package (Ritchie et al. 2015) and applying an empirical Bayes modification (Smyth 2004). Genes with an absolute fold change of at least 2.0 relative to the high-cruciferin genotype 'Red River 1852' and were statistically significant at a false-discovery rate adjusted α-value of 0.05 were deemed to be differentially expressed and carried forward for enrichment analysis. Heatmaps were generated using the pheatmap package (Kolde 2019) in R studio to qualitatively assess the variation in gene expression between biological replicates. Venn diagrams were generated online (http://bioinformatics.psb.ugent.be/webtools/Venn/) to identify differentially expressed genes that were unique to each time point.

Gene ontology (GO) terms (Ashburner et al. 2000; Carbon et al. 2021) were assigned to significant DEG with ShinyGO v0.61 (Ge et al. 2020) using the *B. napus* genome annotation from Ensembl Plants release 43. Enrichment analysis of GO terms was performed using the online platform g:Profiler (Raudvere et al. 2019); the query was conducted in order of increasing *p*-value

against all known genes including electronic GO annotations and the significance of the enrichment was assessed at a Bonferonni-corrected threshold of 0.05. Results of the GO enrichment analyses were plotted in R studio with the script modified from (https://sarahpenir.github.io/r/WEGO/).

6.3.4 Validation with qRT-PCR

Results of the RNA-seq were validated by qRT-PCR. Genes showing differential expression between the two genotypes at the 25 DAP timepoint were randomly selected (Wang et al. 2016a; Geng et al. 2018) using the *sample* function in R Studio version 1.2.1335 (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/) with the seed set to 1. The differentially expressed cruciferin gene (*BnaA10g02240D*) was included for confirmation. Primers were designed using Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye et al. 2012). Primer sequences are presented in Appendix Table S2**Error! Reference source not found.**. Reverse transcription was performed using SuperScriptIV VILO Master Mix (ThermoFisher Scientific. Waltham, U.S.A.) with 500 ng of input RNA following the manufacturer's protocol. RT-qPCR was performed on a CFX Real-Time System (Bio-Rad, Hercules, U.S.A.) with SsoFast EvaGreen Supermix (Bio-Rad, Hercules, U.S.A.); reactions were scaled to a 10 uL volume and run with technical triplicates using the following conditions: 95°C for 3 minutes; 40 cycles of 95°C for 10 seconds and 60°C for 10 seconds; and 65°C for 30 seconds. All runs were followed by a melt curve performed from 65°C to 95°C in 0.5°C increments. Relative transcript abundance of each gene was calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) with *Ubiquitin conjugating enzyme 21* as the reference gene (Chen et al. 2010) and 'Red River 1852' as the reference sample.

6.4 Results

6.4.1 RNA sequencing

To explore transcriptional differences early in seed development that may underlie differential cruciferin accumulation in *B. napus*, RNA sequencing was conducted on whole seeds at 15-, 20-, and 25-DAP from the high-cruciferin genotype 'Red River 1852' and the low-cruciferin genotype 'Wizzard.' All extracted RNA had RNA integrity numbers above 8. Total number of reads for the samples ranged from 59,026,850 to 82,434,549 with a median of 68,435,558. Overall alignment rate to the reference genome averaged 95% across all extracted RNA libraries. Multidimensional scaling indicated biological replicates clustered together and the two genotypes were separated along the diagonal axis (Figure 6.1).

6.4.2 Cruciferin transcripts

Transcripts of previously described *B. napus* cruciferin biosynthesis genes (Wanasundara 2011; Nietzel et al. 2013) from the Uniprot KB database (The UniProt Consortium et al. 2021) were not detected at 15 and 20 DAP indicating the biosynthesis of cruciferin was not initiated at these time points. At 25 DAP, while transcripts of two cruciferin biosynthesis genes *BnaA10g02240D* and *BnaC05g02160D* were detected in the genotype Wizzard.; only *BnaA10g02240D* was found to be differentially expressed.

6.4.3 Differential gene expression

At 15 DAP, a total of 3161 DEG were observed whereas 1565 and 3898 DEG were observed at 20 DAP and 25 DAP, respectively (Figure 6.3). Generally, differential expression was approximately symmetric at 15 DAP with 1492 genes upregulated and 1669 genes downregulated. At both 20 DAP and 25 DAP, differential expressed skewed towards downregulation: overall, the

Table 6.1. Alignment rate of RNA-seq library reads prepared from developing seed of the high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard' harvested at 15, 20, and 25 days after pollination (DAP).

Cultivar	Time point (DAP)	Biological Replicate	Overall Alignment Rate
Red River 1852	15	1	94.46%
		2	94.57%
		3	94.74%
	20	1	94.59%
		2	94.56%
		3	94.92%
	25	1	94.78%
		2	94.60%
		3	94.79%
Wizzard	15	1	94.37%
		2	94.60%
		3	94.77%
	20	1	94.46%
		2	94.58%
		3	94.14%
	25	1	94.45%
		2	94.56%
		3	94.57%

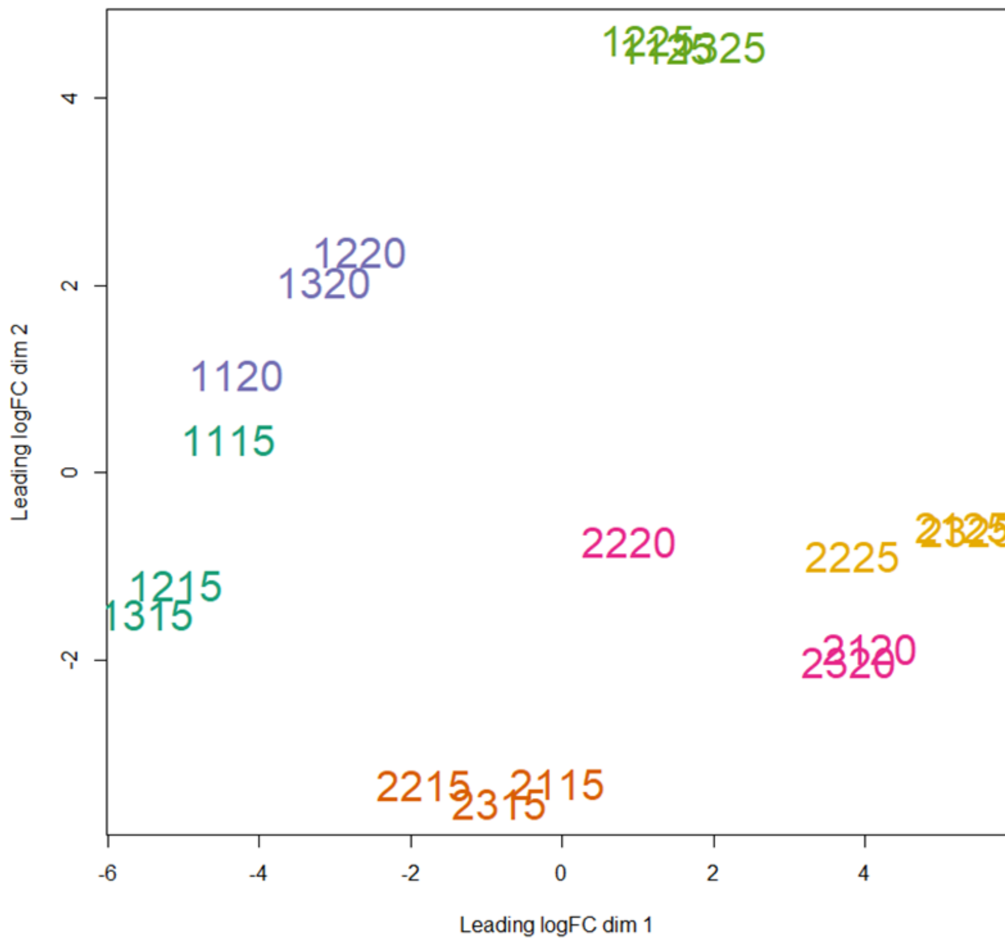

Figure 6.1. Multidimensional scaling of sequenced RNA-seq libraries prepared from developing seeds of the high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' (cultivar 1000) and the low-cruciferin cultivar 'Wizzard' (cultivar 2000), respectively, collected at 15, 20, and 25 days after pollination (DAP). Three biological replicates were analyzed at each time point for each cultivar. Each data point is numerically coded as cultivar (1000, 2000), biological replicate (1-3), and DAP time point (15, 20, 25). Colours denote unique genotype-time point combinations.

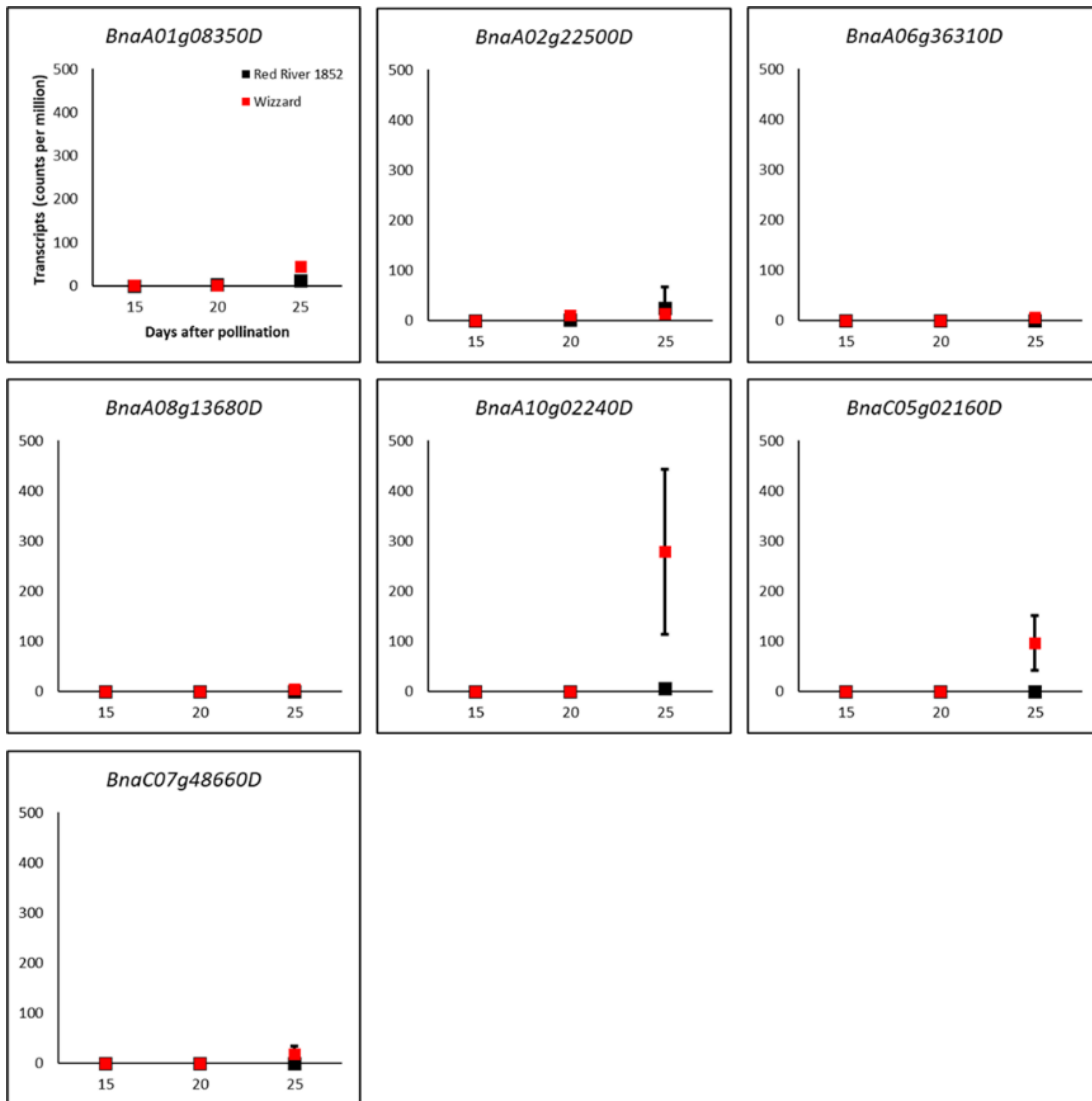

Figure 6.2. Normalized transcript abundance (counts per million) of all seven cruciferin biosynthesis genes across the developing seeds of high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard,' respectively, at 15, 20, and 25 days after pollination, as measured by RNA-seq.

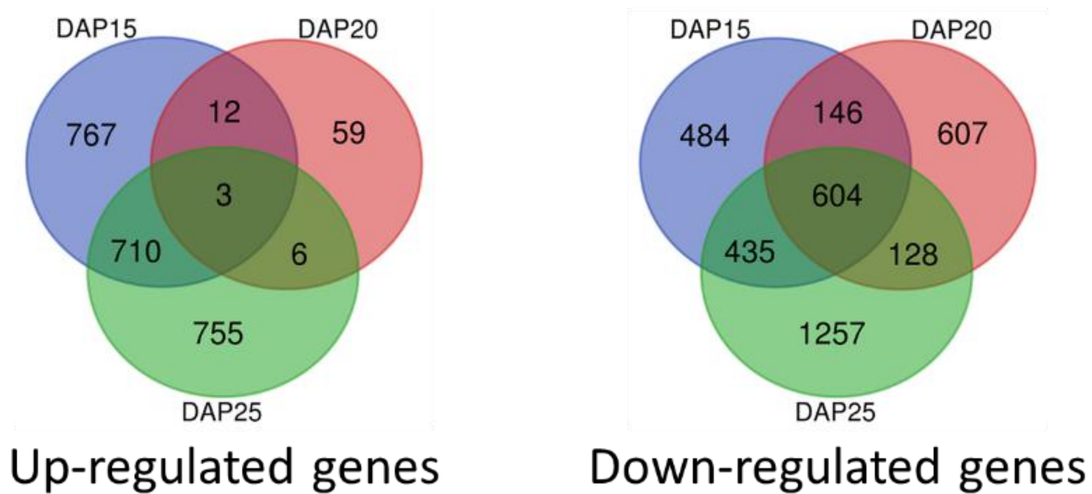

Figure 6.3. Count summary of differentially expressed genes in the developing seeds of high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard,' respectively, at 15, 20, and 25 days after pollination as measured by RNA-seq. Differential expressed was assessed relative to 'Red River 1852.'

number of DEG at 20 DAP was lowest of the three tested time points with 80 genes upregulated and 1285 genes downregulated. At 25 DAP, 1474 genes were upregulated while 2424 were downregulated. When the differentially expressed genes were compared across the three time points, 604 downregulated genes were common across the time points whereas only three genes were commonly upregulated. At 25 DAP when cruciferin transcripts were first detected, 755 genes were uniquely upregulated while 1257 were down-regulated.

6.4.4 Gene ontology enrichment

Gene ontology (GO) enrichment analysis was performed to identify overarching patterns in the differentially expressed genes for both the high- and low-cruciferin genotypes. At 15 DAP, upregulated genes in 'Red River 1852' were enriched for polyamine metabolism (Figure 6.4) while protein catabolism was enriched in the down-regulated genes (Figure 6.5); in 'Wizzard,' upregulated genes were enriched for protein-chromophore linkage (Figure 6.4) while down-regulated genes were primarily enriched for chlorophyll binding and nucleic acid polymerase activity. AT 20 DAP, no significant enrichment was detected for both up- and down-regulated genes in 'Red River 1852' (Figure 6.4). In 'Wizzard,' up-regulated genes at 20 DAP were again enriched for protein-chromophore linkage, in addition to photosynthetic electron transport and nucleic acid biosynthesis (Figure 6.4) while down-regulated genes were primarily enriched for both chlorophyll binding and electron transport (Figure 6.5). At 25 DAP, up-regulated genes in 'Red River 1852' were enriched for polyamine metabolism, while in 'Wizzard' photosynthesis-related processes were enriched (Figure 6.4). In 'Red River 1852,' down-regulated genes at 25 DAP were enriched for glucan synthase activity; in 'Wizzard' significant enrichment for nutrient reservoir activity, was observed at 25 DAP.

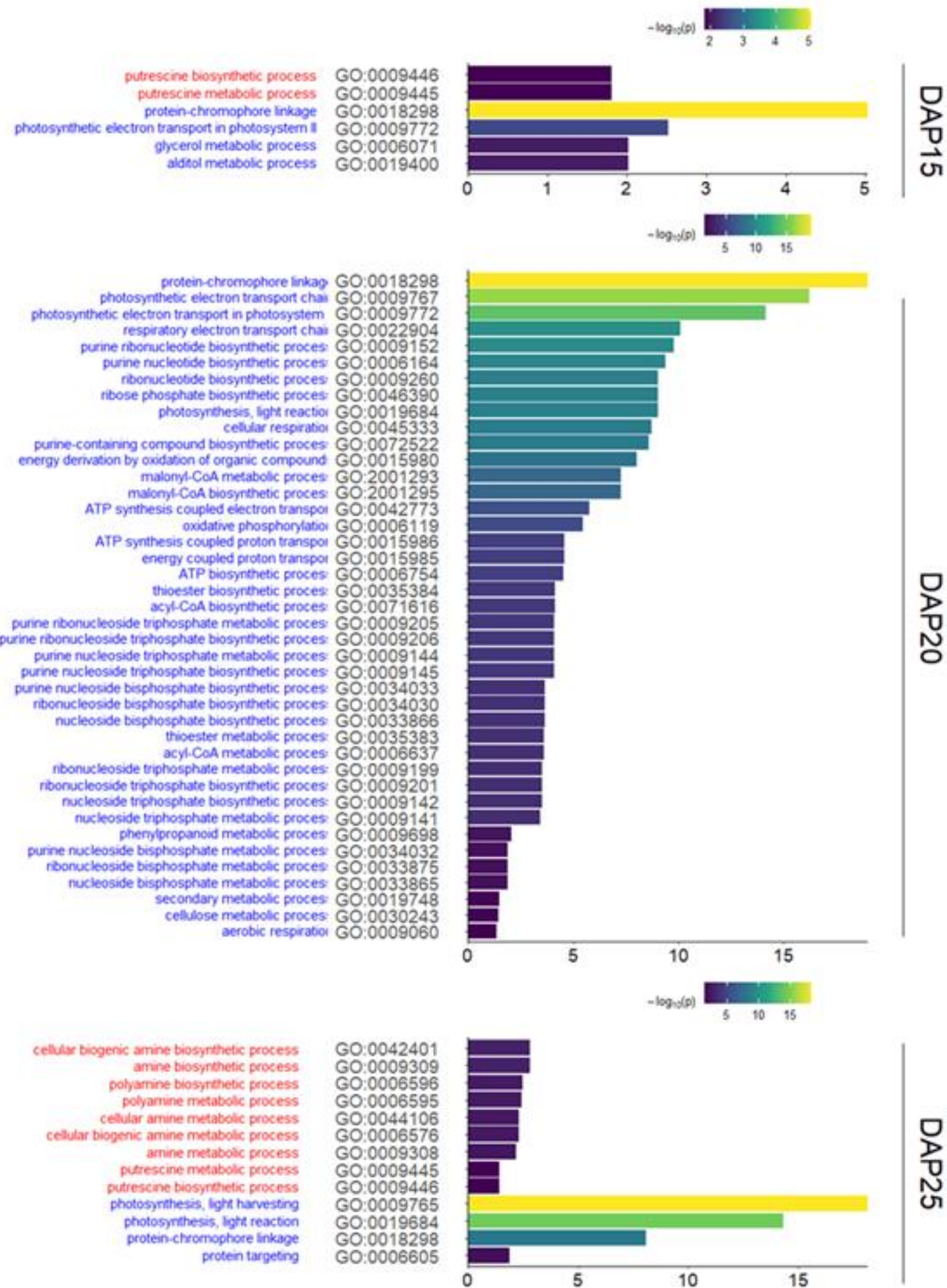

Figure 6.4. Enrichment analysis of gene ontology (GO) terms assigned to genes that were up-regulated in developing seeds at 15, 20, and 25 days after pollination (DAP) between two cultivars of *Brassica napus* L. with different cruciferin content. Terms in red relate to the high-cruciferin genotype 'Red River 1852' while terms in blue relate to the low-cruciferin genotype 'Wizzard.'

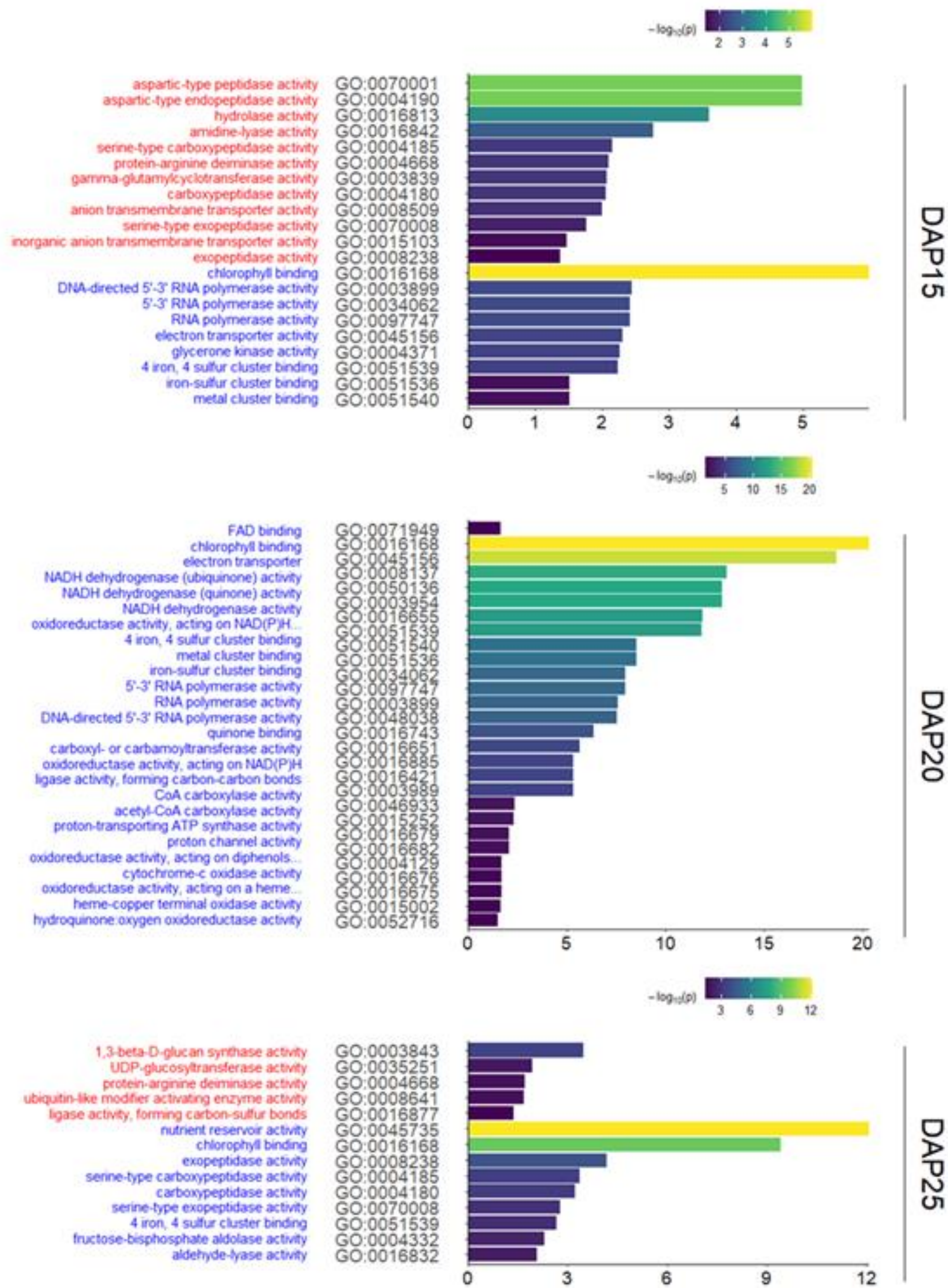

Figure 6.5. Enrichment analysis of gene ontology (GO) terms assigned to genes that were down-regulated in developing seeds at 15, 20, and 25 days after pollination (DAP) between two cultivars of *Brassica napus* L. with different cruciferin content. Terms in red relate to the high-cruciferin genotype 'Red River 1852' while terms in blue relate to the low-cruciferin genotype 'Wizzard.'

6.4.5 qRT-PCR validation

Transcriptomic results were validated by qRT-PCR at 25 DAP using ten randomly selected differentially expressed genes. Broadly, qRT-PCR results were generally concordant with those of the RNA seq (Figure 6.6); RNA seq data for the single differentially expressed cruciferin gene*BnaA10g02240D* matched the qRT-PCR data, both showing increased expression in the low-cruciferin genotype 'Wizzard.' However, some discrepancies were observed, particularly when large differences in expression were detected by one method and not the other, such as in the case of BnaA07g20190D and BnaA03g52820D.

6.5 Discussion

Interest in canola meal protein as a renewable, plant protein source and functional food ingredient has bourgeoned, but basic knowledge of the molecular mechanisms that underlie variation in canola seed protein remains limited. Building on the identification of high- and low-cruciferin genotypes of *B. napus*, this research aimed to identify transcriptional differences during early seed development between the high-cruciferin cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard.' 4

Previous time course transcriptomic studies in developing *B. napus* seeds have focused on elucidating mechanisms that underlie oil accumulation (Niu et al. 2009; Wang et al. 2016a; Wan et al. 2017b; Liao et al. 2019; Shahid et al. 2019) with little focus on storage protein accumulation. The proteomic study of Hajduch et al. (2006) conducted on developing *B. napus* seed revealed storage protein to be prevalent from 28 DAP onward. Given the temporal delay between transcription and translation (Liu et al. 2016d), transcripts of seed storage proteins genes can be

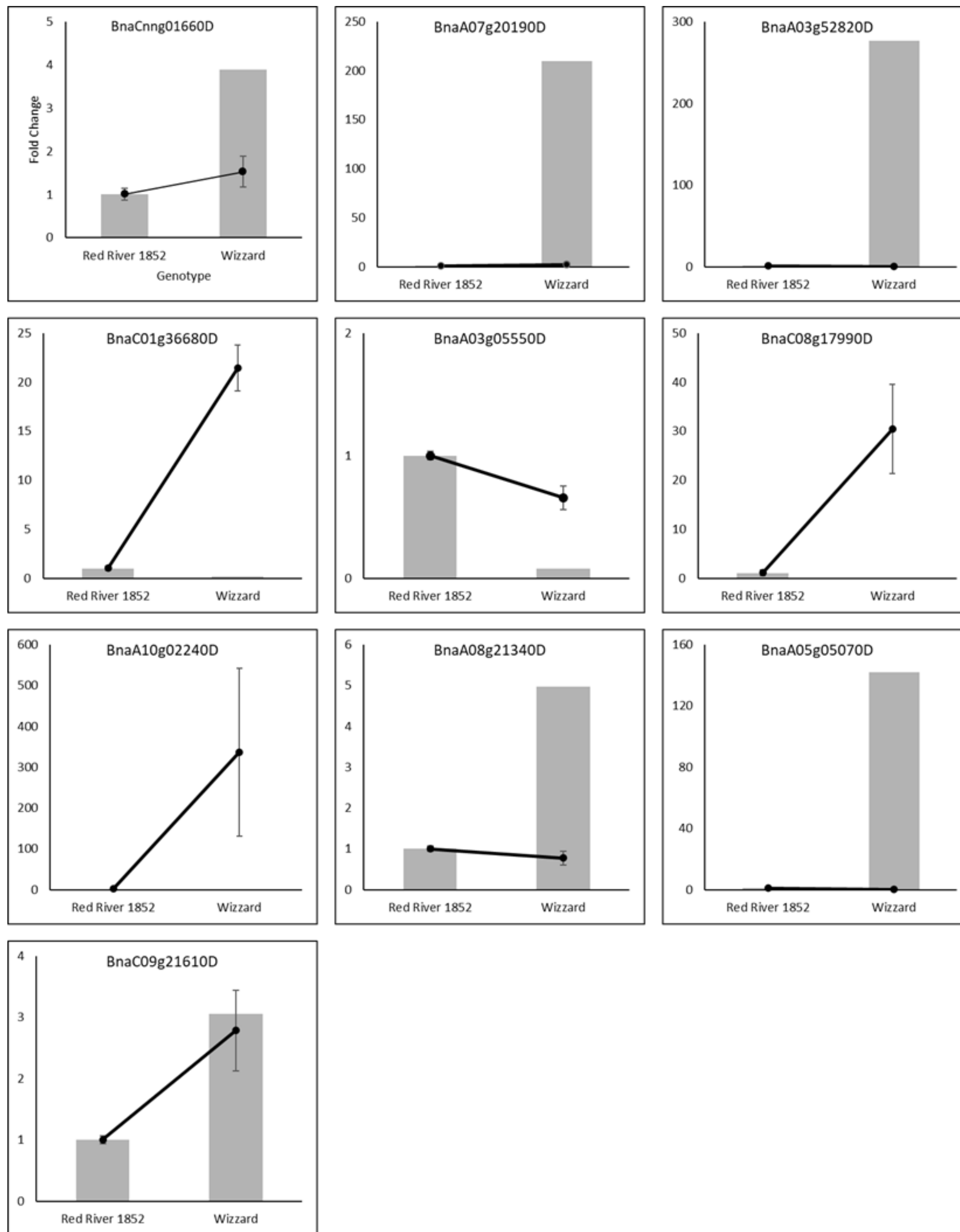

Figure 6.6. Quantitative reverse transcription PCR (qRT-PCR) of select, differentially expressed genes identified by RNA sequencing of developing seeds between the high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard' at 25 days after pollination. RNA sequencing results are presented as bars while qRT-PCR results are presented as dots. Error bars represent standard error (α=0.05) of three biological replicates for qRT-PCR.

expected to begin accumulating before this time point. Indeed, the accumulation of napin and cruciferin transcripts was found to peak before the accumulation of their translated product (Delisle and Crouch 1989; Gruis et al. 2002) in *B. napus*.

The accumulation of cruciferin is transcriptionally regulated and shows distinct spatial-temporal patterns across seed development (Fernandez et al. 1991; Hoglund et al. 1992; Sjödahl et al. 1993). In the present study, transcripts of cruciferin biosynthesis genes were undetected at 15 and 20 DAP, while those of two genes were present at 25 DAP, suggesting the biosynthesis of cruciferin was initiated at 25 DAP. Although these results mirrored the findings of (Finkelstein et al. 1985) who found cruciferin transcripts were detected in *B. napus* embryos starting at 21 DAP, they are incongruous with the work of Tan et al. (2020) who found protein content to already rapidly increase in *B. napus* embryos starting from 25 DAP. This temporal discrepancy may be attributed to genetic differences between the genotypes and the use of spring *B. napus* in the former work and the winter ecotype in the latter. Variation in seed maturity date and rate of development exists; therefore, equivalent time points measured relative to pollination may result in different developmental stages.

Interestingly, cruciferin transcripts at 25 DAP were only detected in the low cruciferin cultivar 'Wizzard' and not in the high-cruciferin cultivar 'Red River 1852.' Although both genotypes were planted and flowered at the same time, a visually noticeable difference in seed size was observed indicating the rate of seed development was not coherent between the two cultivars. A larger seed size in the low-cruciferin genotype 'Wizzard' suggested seed development in this cultivar was faster than that of 'Red River 1852.' Rapid progression through a condensed seed developmental timeline may alter when seed reserves are produced; therefore, differences in the

rate of seed and embryo development may explain the detection of cruciferin transcripts in the low-cruciferin genotype but not its high-cruciferin counterpart at 25 DAP. This in turn indicates early initiation of cruciferin biosynthesis is not the underlying mechanism by which these genotypes differentially accumulate cruciferin. Future studies to explore transcriptomic changes through the entirety of seed develop will aid in discerning whether this observation was anomalous or an effect of cruciferin subunit diversity (Dalgalarrondo et al. 1986; Wanasundara 2011; Withana-Gamage et al. 2013) between the genotypes.

To observe overarching patterns in the DEG at each time point, functional enrichment analysis was used to identify overrepresented GO terms. Generally, overrepresented GO terms related to functions known to be associated with embryo development. For example, polyamines have been implicated in the regulation of embryo development (Chen et al. 2019a) and genes related to its metabolism were found to be overrepresented in the upregulated genes of the high-cruciferin genotype ‘Red River 1852’ at 15 DAP. Similarly, GO terms related to photosynthesis and light response were found to be overrepresented at multiple time points. Immature embryos are capable of photosynthesis and defects in this function result in irregular post-germinative development (Sela et al. 2020); thus, overrepresentation of photosynthesis-related terms was reasonable. Further, light has been demonstrated to induce the biosynthesis of seed storage proteins (Tan et al. 2020) and accordingly, protein-chromophore linkage was overrepresented at multiple time points. The overrepresentation of nutrient reservoir activity in the upregulated genes of the low-cruciferin genotype at 25 DAP indicated that many of the differentially-expressed transcripts were related to storage protein and oil. Detection of cruciferin transcripts at 25 DAP suggests cruciferin protein accumulation should begin to accumulate shortly after this time point which is consistent with the timeline of Murphy and Cummins (1989).

In spring *B. napus,* the developing seed undergoes major metabolic changes from 2 to 6 weeks after pollination (WAP) where seed storage proteins dominate the proteome by 4 WAP (Hajduch et al. 2006); therefore, the accumulation of their transcripts was expected to commence prior to this point. However, it is now evident that the time points used in this study were too early to observe differential expression of cruciferin biosynthetic genes between the high and low-cruciferin genotypes.

In retrospect, time points for transcriptomic analysis should have been selected following preliminary experiments to quantify cruciferin transcript and protein levels across seed development in the genotypes of this study. Such preliminary work would have accounted for inherent differences in cruciferin accumulation due to genotypic variation in the rate of seed development, and more importantly, would have identified suitable time points to ensure the transcriptome was analyzed once cruciferin accumulation had been initiated. Future studies employing a broad range of time points from 25 DAP onwards to seed maturity will enable any temporal differences in cruciferin biosynthesis or subunit bias to be resolved between the two cultivars. Finally, some discrepancies between the transcriptomics data and the qRT-PCR validation data were observed. Although inconsistencies between both methods have previously been observed (Everaert et al. 2017), correlation between the two are typically reported to be high (Griffith et al. 2010; Wu et al. 2014a) and the use of additional genes for validation may aid in improving the correlation in this study.

Nonetheless, the work presented herein represents the first attempt to identify transcriptional patterns and mechanisms underlying differential cruciferin accumulation in *B. napus*. The regulation of cruciferin accumulation at the genomic and post-translational levels has yet to be fully explored and cannot be ruled out; newly-developed genomic resources such as the

pan-genome (Song et al. 2021) will assist in identifying copy number variation and other genomic features that contribute to differential cruciferin accumulation. Finally, an examination of the proteomic landscape over the course of seed development may provide novel insights into transcriptional dynamics that may be correlated with differential cruciferin accumulation.

6.6 Conclusion

The development of high-cruciferin cultivars of *B. napus* is necessary to increase the competitiveness of canola protein as a source of dietary protein. An understanding of the molecular mechanisms that underlie differential cruciferin accumulation in the crop facilitates the breeding improvement of cruciferin content by identifying targets for selection. Cruciferin transcripts begin to accumulate at 25 DAP indicating this may be the earliest time point in which selection can be made. Although differences in the accumulation of cruciferin subunit transcripts were detected between the high and low-cruciferin genotypes, precocious accumulation of these transcripts at 25 DAP did not result in a high cruciferin phenotype. This study offers a view of the transcriptional differences at the early stages of seed development between high and low-cruciferin genotypes, and future work focused on later time points in seed development will offer insight into the transcriptional control of cruciferin accumulation.

Chapter 7. General Discussion

The diversion of oil-free canola (*Brassica napus* L.) seed meal after crushing from livestock feed to a protein ingredient for food processing adds value to an already lucrative oilseed crop (LMC International 2020). The development of high-cruciferin canola cultivars suitable for food processing is a step towards ensuring the ability of modern agriculture to address posterior food and nutrition security (Nelson et al. 2018). More importantly, altering the composition of the seed protein in canola improves protein quality without compromising the primary commodity (edible oil) of the crop. Further, cultivars with improved cruciferin content can be utilized for the production of novel plant protein products to meet increasing consumer demand for sustainable protein alternatives (Kumar et al. 2017). Evidently, high-cruciferin canola cultivars are beneficial to multiple stakeholders across the agriculture value chain. The research presented in this thesis was untaken to establish protocols and basic knowledge of cruciferin content variation in spring canola to facilitate its breeding improvement.

Breeding for improved cruciferin content first requires a reliable phenotyping protocol to quantify cruciferin content. To this end, an immunoblotting protocol using a custom anti-cruciferin antibody was established in Chapter 3. The antibody was raised and confirmed to be specific to the alpha chain of the five cruciferin subunits in *B. napus* (Wanasundara 2011; Nietzel et al. 2013). Although phenotyping for cruciferin content had previously been achieved through chromatography (Malabat et al. 2003) and densitometry of electrophoretically resolved peptids (Schatzki et al. 2014), the immunoblotting protocol established in Chapter 3 is advantageous in requiring less technical infrastructure than the former and offering better specificity in identifying cruciferin peptides than the latter.

Subsequently, the immunoblotting protocol was employed to assess the genetic diversity in cruciferin content that exists in spring *B. napus*. The presence of genetic diversity is a prerequisite for breeding as it enables progeny with genetic recombinations encoding improved phentoypes to be generated (Fu 2015; Henderson and Salt 2017). Results from Chapter 4 indicated genetic variation in cruciferin content was present in spring *B. napus* genotypes that spanned the breeding history of the crop. When grown under greenhouse conditions to minimize possible environmental effects on the trait, cruciferin content varied from 44% to 93% of a reference sample of field-grown seed from the genotype 'Westar.' Although genetic variation in winter *B. napus* had previously been examined (Malabat et al. 2003), similar assessments have yet to be conducted in spring *B. napus* and this study represents a unique contribution to the canola literature.

Upon establishing the presence of genetic variation in cruciferin content within spring *B. napus*, the identification of molecular markers associated with the trait is necessary to facilitate efficient selection of segregating progeny. To this end, an association mapping approach was used in Chapter 5 to identify single nucleotide polymorphism molecular markers that were associated with cruciferin content in the crop. Multiple molecular markers on chromosomes A06 and A07 were found to be associated ($p < 0.01$) with cruciferin content and candidate genes found within haplotype blocks containing the markers suggested their role in the indirect control of cruciferin accumulation. The associated markers did not overlap with quantitative loci previously reported to be correlated with cruciferin content in winter *B. napus* (Schatzki et al. 2014) and therefore alludes to potentially novel mechanisms that govern cruciferin accumulation. In using an association mapping approach, the application of the identified markers are not limited to the association panel (Cobb et al. 2019).

Finally, to complement the association mapping study and to better understand the molecular regulation of cruciferin accumulation, an RNA sequencing study was performed in Chapter 6 to identify transcriptomic differences in the developing seeds between high and low cruciferin *B. napus* genotypes at the onset of cruciferin accumulation. Results of the study indicated cruciferin accumulation was initiated at 25 days after pollination which was generally in concordance with previous research (Finkelstein et al. 1985). The transcriptomic results further indicated that early initiation of the expression of cruciferin biosynthesis genes was not responsible for the differential cruciferin phenotype and differentially expressed genes between the high- and low-cruciferin genotypes were primarily enriched for regulatory functions.

Collectively, the studies presented in this thesis explore both the phenotypic and genotypic aspects of cruciferin accumulation in spring *B. napus*. The studies contribute an improved phenotyping protocol for cruciferin quantification, the first assessment of genetic variation of cruciferin content in spring *B. napus,* and insight into potential molecular mechanisms that control cruciferin accumulation. Together, the results of this thesis will facilitate the breeding development of high-cruciferin canola cultivars to mitigate food security and add value to the crop.

Chapter 8. Future research

The work presented in this dissertation provides basic knowledge on the genetic variation in cruciferin content that exists in spring *Brassica napus* L. and molecular mechanisms that govern this variation. Although this knowledge enables the breeding improvement of cruciferin content to be initiated, additional research is warranted to better understand the factors that control cruciferin accumulation and ensure long term breeding success.

To facilitate phenotyping cruciferin content, a Western blot protocol was established and optimized in Chapter 3. Improvements to the throughput of this phenotyping method will increase breeding efficiency by allowing larger populations to be screened. The specificity of immunoassays should be maintained and the use of plate-based assays such as dot blots and enzyme-linked immunosorbent assays should be considered to increase throughput. Further, validated plate-based assays can be commercialized to establish a phenotyping benchmark for the industry.

Using the phenotyping method previously optimized, genetic diversity in cruciferin content was assessed under greenhouse conditions in Chapter 4. To better assess genetic diversity in cruciferin content, replicated multi-year and multi-location research experiments are warranted in addition to greenhouse studies. Field research experiments reflect commercial production and will enable the elucidation of the environmental and the genotype-by-environmental effects on cruciferin accumulation.

Thereafter, molecular markers associated with cruciferin content were identified by association mapping in Chapter 5. To improve the accuracy of the identification of molecular markers associated with cruciferin content, additional genotypes, sourced globally, should be

included in the association panel; increased genetic diversity and population size will afford greater mapping resolution. In addition, the use of multi-parental populations and whole genome sequencing can further improve the accuracy of the association mapping. Forward genetic and cloning experiments will also be required to conclusively validate marker-trait associations.

An RNA sequencing approach was used in Chapter 6 to identify transcriptional differences at the initiation of cruciferin accumulation that may underlie diversity in the trait. Additional time points in the time course (up to seed dehiscence) should be considered to comprehensively evaluate transcriptional patterns across the entirety of seed development that underlie differential cruciferin accumulation. The use of additional genotypes is also warranted to confirm the ubiquity of the differential transcriptional patterns. Together, the above recommendations will contribute basic knowledge to facilitate the breeding improvement of cruciferin content in *B. napus* and contribute to the challenge of ensuring adequate global food and nutrition security.

Chapter 9. Reference material

9.1 Literature cited

Aachary, A.A., and Thiyam, U. 2012. A pursuit of the functional nutritional and bioactive properties of canola proteins and peptides. Crit. Rev. Food Sci. Nutr. **52**: 965–979. doi:10.1080/10408398.2010.516033.

Abdallah, H. Ben, and Bauer, P. 2016. Quantitative Reverse Transcription-qPCR-based gene expression analysis in plants. Pages 9–24 *in* Methods in molecular biology (Clifton, N.J.). doi:10.1007/978-1-4939-3115-6_2.

Adem, H.N., Tressel, R.P., Pudel, F., Slawski, H., and Schulz, C. 2014. Rapeseed use in aquaculture. OCL - Oilseeds fats **21**: 1–9. doi:10.1051/ocl/2013041.

Adewole, D.I., Rogiewicz, A., Dyck, B., and Slominski, B.A. 2017. Effects of canola meal source on the standardized ileal digestible amino acids and apparent metabolizable energy contents for broiler chickens. Poult. Sci. **96**: 4298–4306. Poultry Science Association Inc. doi:10.3382/ps/pex245.

Aider, M., and Barbana, C. 2011. Canola proteins: Composition, extraction, functional properties, bioactivity, applications as a food ingredient and allergenicity - A practical and critical review. Trends Food Sci. Technol. **22**: 21–39. Elsevier Ltd. doi:10.1016/j.tifs.2010.11.002.

Almeida, F.N., Htoo, J.K., Thomson, J., and Stein, H.H. 2014. Effects of heat treatment on the apparent and standardized ileal digestibility of amino acids in canola meal fed to growing pigs. Anim. Feed Sci. Technol. **187**: 44–52. Elsevier B.V. doi:10.1016/j.anifeedsci.2013.09.009.

Altenbach, S.B., Kuo, C.C., Staraci, L.C., Pearson, K.W., Wainwright, C., Georgescu, A., and Townsend, J. 1992. Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol. Biol. **18**: 235–245. doi:10.1007/BF00034952.

Altenbach, S.B., Tanaka, C.K., Pineau, F., Lupi, R., Drouet, M., Beaudouin, E., Morisset, M., and Denery-Papini, S. 2015. Assessment of the allergenic potential of transgenic wheat (*Triticum aestivum*) with reduced levels of ω5-gliadins, the major sensitizing allergen in wheat-dependent exercise-induced anaphylaxis. J. Agric. Food Chem. **63**: 9323–9332. doi:10.1021/acs.jafc.5b03557.

Altenbach, S.B., Tanaka, C.K., and Seabourn, B.W. 2014. Silencing of omega-5 gliadins in transgenic wheat eliminates a major source of environmental variability and improves dough mixing properties of flour. BMC Plant Biol. **14**. doi:10.1186/s12870-014-0393-1.

Altpeter, F., Vasil, V., Srivastava, V., and Vasil, I.K. 1996. Integration and expression of the high-molecular-weight glutenin subunit 1Ax1 gene in wheat. Nat. Biotechnol. **14**: 1246–1251. doi:10.1038/nbt0898-773.

Aluko, R.E., and McIntosh, T. 2001. Polypeptide profile and functional properties of defatted meals and protein isolates of canola seeds. J. Sci. Food Agric. **81**: 391–396. doi:10.1002/1097-0010(200103)81:4<391::AID-JSFA823>3.0.CO;2-S.

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. [Online] Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Angeles-Núñez, J.G., and Tiessen, A. 2011. Mutation of the transcription factor LEAFY COTYLEDON 2 alters the chemical composition of *Arabidopsis* seeds, decreasing oil and protein content, while maintaining high levels of starch and sucrose in mature seeds. J. Plant Physiol. **168**: 1891–1900. doi:10.1016/j.jplph.2011.05.003.

Angelovici, R., Lipka, A.E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., Buell, R., Gore, M.A., and DellaPenna, D. 2013. Genome-wide analysis of branched-chain amino acid levels in Arabidopsis Seeds. Plant Cell **25**: 4827–4843. doi:10.1105/tpc.113.119370.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. **25**: 25–29. doi:10.1038/75556.

Bahadur KC, K., Dias, G.M., Veeramani, A., Swanton, C.J., Fraser, D., Steinke, D., Lee, E., Wittman, H., Farber, J.M., Dunfield, K., McCann, K., Anand, M., Campbell, M., Rooney, N., Raine, N.E., Van Acker, R., Hanner, R., Pascoal, S., Sharif, S., Benton, T.G., and Fraser, E.D.G. 2018. When too much isn't enough: Does current food production meet global nutritional needs? PLoS One **13**: 1–16. doi:10.1371/journal.pone.0205683.

Bailey, H.M., and Stein, H.H. 2019. Can the digestible indispensable amino acid score methodology decrease protein malnutrition. Anim. Front. **9**: 18–23. doi:10.1093/af/vfz038.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One **3**: 1–7. doi:10.1371/journal.pone.0003376.

Bao, B., Chao, H., Wang, H., Zhao, W., Zhang, L., Raboanatahiry, N., Wang, X., Wang, B., Jia, H., and Li, M. 2018. Stable, environmental specific and novel QTL identification as well as genetic dissection of fatty acid metabolism in *Brassica napus*. Front. Plant Sci. **9**: 1–25. doi:10.3389/fpls.2018.01018.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics **21**: 263–265. doi:10.1093/bioinformatics/bth457.

Barro, F., Iehisa, J.C.M., Giménez, M.J., García-Molina, M.D., Ozuna, C. V., Comino, I., Sousa,

C., and Gil-Humanes, J. 2016. Targeting of prolamins by RNAi in bread wheat: Effectiveness of seven silencing-fragment combinations for obtaining lines devoid of coeliac disease epitopes from highly immunogenic gliadins. Plant Biotechnol. J. **14**: 986–996. doi:10.1111/pbi.12455.

Barro, F., Rooke, L., Bekes, F., Gras, P., Tatham, A.S., Fido, R., Lazzeri, P.A., Shrewy, P.R., Barcelo, P., Tatha, A.S., Fido, R., Lazzeri, P.A., Shrewy, P.R., Barcelo, P., Tatham, A.S., Fido, R., Lazzeri, P.A., Shrewy, P.R., Barcelo, P., Tatha, A.S., Fido, R., Lazzeri, P.A., Shrewy, P.R., Barcelo, P., Békés, F., Gras, P., Tatham, A.S., Fido, R., Lazzeri, P.A., Shewry, P.R., and Barceló, P. 1997. Transformation of wheat with high molecular weight subunit genes results in improved functional properties. Nat. Biotechnol. **15**: 871–875. doi:10.1038/nbt1197-1295.

Barthet, V.J. 2016. Comparison between Canadian canola harvest and export surveys. Plants **5**: 1209–1214. doi:10.3390/plants5030030.

Barthet, V.J. 2018a. Quality of western Canadian Canola 2018. Winnipeg. [Online] Available: www.grainscanada.gc.ca.

Barthet, V.J. 2018b. Quality of Canadian oilseed-type soybeans 2018. Winnipeg. [Online] Available: https://www.grainscanada.gc.ca/en/grain-research/export-quality/oilseeds/soybean-oil/2018/pdf/report18.pdf.

Barthet, V.J. 2019. Quality of western Canadian Canola 2019. Winnipeg. [Online] Available: https://www.grainscanada.gc.ca/en/grain-research/export-quality/oilseeds/canola/2019/index.html.

Barthet, V.J., Petryk, M.W.P., and Siemens, B. 2020. Rapid nondestructive analysis of intact canola seed using a handheld near-infrared spetrometer. JAOCS, J. Am. Oil Chem. Soc. **97**: 577–589. doi:10.1002/aocs.12335.

Bass, J.J., Wilkinson, D.J., Rankin, D., Phillips, B.E., Szewczyk, N.J., Smith, K., and Atherton, P.J. 2017. An overview of technical considerations for Western blotting applications to physiological research. Scand. J. Med. Sci. Sport. **27**: 4–25. doi:10.1111/sms.12702.

Bates, D., Mächler, M., Bolker, B.M., and Walker, S.C. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. **67**. doi:10.18637/jss.v067.i01.

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

Behnke, N., Suprianto, E., and Möllers, C. 2018. A major QTL on chromosome C05 significantly reduces acid detergent lignin (ADL) content and increases seed oil and protein content in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. **131**: 2477–2492. Springer Berlin Heidelberg. doi:10.1007/s00122-018-3167-6.

Bell, J.M. 1993. Factors affecting the nutritional value of canola meal: A review. Can. J. Anim. Sci. **73**: 689–697. doi:10.4141/cjas93-075.

Bentham, J., Singh, G.M., Danaei, G., Green, R., Lin, J.K., Stevens, G.A., Farzadfar, F., Bennett, J.E., Di Cesare, M., Dangour, A.D., and Ezzati, M. 2020. Multidimensional characterization of global food supply from 1961 to 2013. Nat. Food **1**: 70–75. Springer US. doi:10.1038/s43016-019-0012-2.

Van den Berge, K., Hembach, K.M., Soneson, C., Tiberi, S., Clement, L., Love, M.I., Patro, R., and Robinson, M.D. 2019. RNA sequencing data: Hitchhiker's guide to expression analysis. Annu. Rev. Biomed. Data Sci. **2**: 139–173. doi:10.1146/annurev-biodatasci-072018-021255.

Bernardo, R. 2020. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. Heredity (Edinb). Springer US. doi:10.1038/s41437-020-0312-1.

Bérot, S., Compoint, J.P., Larré, C., Malabat, C., and Guéguen, J. 2005. Large scale purification of rapeseed proteins (*Brassica napus* L.). J. Chromatogr. B Anal. Technol. Biomed. Life Sci. **818**: 35–42. doi:10.1016/j.jchromb.2004.08.001.

Bieker, S., Riester, L., Doll, J., Franzaring, J., Fangmeier, A., and Zentgraf, U. 2019. Nitrogen supply drives senescence-related seed storage protein expression in rapeseed leaves. Genes (Basel). **10**. doi:10.3390/genes10020072.

Blackstone, N.T., El-Abbadi, N.H., McCabe, M.S., Griffin, T.S., and Nelson, M.E. 2018. Linking sustainability to the healthy eating patterns of the Dietary Guidelines for Americans: a modelling study. Lancet Planet. Heal. **2**: e344–e352. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY NC-ND 4.0 license. doi:10.1016/S2542-5196(18)30167-0.

Blechl, A.E., and Anderson, O.D. 1996. Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat. Nat. Biotechnol. **14**: 875–879. doi:10.1038/nbt0796-875.

Boehm, J.D., Nguyen, V., Tashiro, R.M., Anderson, D., Shi, C., Wu, X., Woodrow, L., Yu, K., Cui, Y., and Li, Z. 2018. Genetic mapping and validation of the loci controlling 7S α′ and 11S A-type storage protein subunits in soybean [*Glycine max* (L.) Merr.]. Theor. Appl. Genet. **131**: 659–671. Springer Berlin Heidelberg. doi:10.1007/s00122-017-3027-9.

Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120. doi:10.1093/bioinformatics/btu170.

Bonds, R.S., Midoro-Horiuti, T., and Goldblum, R. 2008. A structural basis for food allergy: The role of cross-reactivity. Curr. Opin. Allergy Clin. Immunol. **8**: 82–86. doi:10.1097/ACI.0b013e3282f4177e.

Borisjuk, L., Neuberger, T., Schwender, J., Heinzel, N., Sunderhaus, S., Fuchs, J., Hay, J.O., Tschiersch, H., Braun, H.-P., Denolf, P., Lambert, B., Jakob, P.M., and Rolletschek, H.

2013. Seed architecture shapes embryo metabolism in oilseed rape. Plant Cell **25**: 1625–1640. doi:10.1105/tpc.113.111740.

Borpatragohain, P., Rose, T.J., and King, G.J. 2016. Fire and brimstone: Molecular interactions between sulfur and glucosinolate biosynthesis in model and crop brassicaceae. Front. Plant Sci. **7**: 1–18. doi:10.3389/fpls.2016.01735.

Bos, C., Airinei, G., Mariotti, F., Benamouzig, R., Bérot, S., Evrard, J., Fénart, E., Tomé, D., and Gaudichon, C. 2007. The poor digestibility of rapeseed protein is balanced by its very high metabolic utilization in humans. J. Nutr. **137**: 594–600. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/17311946.

Boutilier, K., Hattori, J., Baum, B.R., and Miki, B.L. 1999. Evolution of 2S albumin seed storage protein genes in the Brassicaceae. Biochem. Syst. Ecol. **27**: 223–234. doi:10.1016/S0305-1978(98)00103-3.

Braatz, J., Harloff, H.-J.J., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. 2017. CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). Plant Physiol. **174**: 935–942. doi:10.1104/pp.17.00426.

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

Brunei-Muguet, S., D'Hooghe, P., Bataillé, M.P., Larré, C., Kim, T.H., Trouverie, J., Avice, J.C., Etienne, P., and Dürr, C. 2015. Heat stress during seed filling interferes with sulfur restriction on grain composition and seed germination in oilseed rape (</i>Brassica napus</i> L.). Front. Plant Sci. **6**: 1–12. doi:10.3389/fpls.2015.00213.

Bryan, D.D.S.L., Abbott, D.A., Van Kessel, A.G., and Classen, H.L. 2019. In vivo digestion characteristics of protein sources fed to broilers. Poult. Sci. **98**: 3313–3325. Poultry Science Association Inc. doi:10.3382/ps/pez067.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., and Wittwer, C.T. 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. **55**: 611–622. doi:10.1373/clinchem.2008.112797.

Byczyńska, A., and Barciszewski, J. 1999. The biosynthesis, structure and properties of napin - The storage protein from rape seeds. J. Plant Physiol. **154**: 417–425. doi:10.1016/S0176-1617(99)80277-6.

Calhoun, W., Crane, J.M., and Stamp, D.L. 1975. Development of a low glucosinolate, high erucic acid rapeseed breeding program. J. Am. Oil Chem. Soc. **52**: 363–365. doi:10.1007/BF02639197.

Callaway, E. 2018. Gene Editing. EU law deals blow to CRISPR crops. Nature **560**: 2016.

doi:10.1038/d41586-018-05814-6.

Capron, A., Chatfield, S., Provart, N., and Berleth, T. 2009. Embryogenesis: Pattern formation from a single cell. Arab. B. **7**: e0126. doi:10.1199/tab.0126.

Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E., Fey, P., Thomas, P.D., Albou, L.-P., Ebert, D., Kesling, M.J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., LaBonte, S.A., Siegele, D.A., Antonazzo, G., Attrill, H., Brown, N.H., Garapati, P., Marygold, S.J., Trovisco, V., dos Santos, G., Falls, K., Tabone, C., Zhou, P., Goodman, J.L., Strelets, V.B., Thurmond, J., Garmiri, P., Ishtiaq, R., Rodríguez-López, M., Acencio, M.L., Kuiper, M., Lægreid, A., Logie, C., Lovering, R.C., Kramarz, B., Saverimuttu, S.C.C., Pinheiro, S.M., Gunn, H., Su, R., Thurlow, K.E., Chibucos, M., Giglio, M., Nadendla, S., Munro, J., Jackson, R., Duesbury, M.J., Del-Toro, N., Meldal, B.H.M., Paneerselvam, K., Perfetto, L., Porras, P., Orchard, S., Shrivastava, A., Chang, H.-Y., Finn, R.D., Mitchell, A.L., Rawlings, N.D., Richardson, L., Sangrador-Vegas, A., Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D.M., Harris, M.A., Oliver, S.G., Rutherford, K., Wood, V., Hayles, J., Bähler, J., Bolton, E.R., De Pons, J.L., Dwinell, M.R., Hayman, G.T., Kaldunski, M.L., Kwitek, A.E., Laulederkind, S.J.F., Plasterer, C., Tutaj, M.A., Vedi, M., Wang, S.-J., D'Eustachio, P., Matthews, L., Balhoff, J.P., Aleksander, S.A., Alexander, M.J., Cherry, J.M., Engel, S.R., Gondwe, F., Karra, K., Miyasato, S.R., Nash, R.S., Simison, M., Skrzypek, M.S., Weng, S., Wong, E.D., Feuermann, M., Gaudet, P., Morgat, A., Bakker, E., Berardini, T.Z., Reiser, L., Subramaniam, S., Huala, E., Arighi, C.N., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Bateman, A., Blatter, M.-C., Boutet, E., Bowler, E., Breuza, L., Bridge, A., Britto, R., Bye-A-Jee, H., Casas, C.C., Coudert, E., Denny, P., Estreicher, A., Famiglietti, M.L., Georghiou, G., Gos, A., Gruaz-Gumowski, N., Hatton-Ellis, E., Hulo, C., Ignatchenko, A., Jungo, F., Laiho, K., Le Mercier, P., Lieberherr, D., Lock, A., Lussi, Y., MacDougall, A., Magrane, M., Martin, M.J., Masson, P., Natale, D.A., Hyka-Nouspikel, N., Orchard, S., Pedruzzi, I., Pourcel, L., Poux, S., Pundir, S., Rivoire, C., Speretta, E., Sundaram, S., Tyagi, N., Warner, K., Zaru, R., Wu, C.H., Diehl, A.D., Chan, J.N., Grove, C., Lee, R.Y.N., Muller, H.-M., Raciti, D., Van Auken, K., Sternberg, P.W., Berriman, M., Paulini, M., Howe, K., Gao, S., Wright, A., Stein, L., Howe, D.G., Toro, S., Westerfield, M., Jaiswal, P., Cooper, L., and Elser, J. 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. **49**: D325–D334. doi:10.1093/nar/gkaa1113.

Carmona, S., Alvarez, J.B., and Caballero, L. 2010. Genetic diversity for morphological traits and seed storage proteins in Spanish rivet wheat. Biol. Plant. **54**: 69–75. doi:10.1007/s10535-010-0010-6.

Cassin-Ross, G., and Hu, J. 2014. Systematic phenotypic screen of arabidopsis peroxisomal mutants identifies proteins involved in β-oxidation. Plant Physiol. **166**: 1546–1559. doi:10.1104/pp.114.250183.

Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-Ponce, B., Sánchez, M. de la P., Álvarez-Buylla, E.R., and Garay-Arroyo, A. 2019. MADS-box genes are key components of genetic regulatory networks involved in abiotic stress and

plastic developmental responses in plants. Front. Plant Sci. **10**. doi:10.3389/fpls.2019.00853.

Cernac, A., and Benning, C. 2004. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. Plant J. **40**: 575–585. doi:10.1111/j.1365-313X.2004.02235.x.

Chai, S.J., Cole, D., Nisler, A., and Mahon, B.E. 2017. Poultry: The most common food in outbreaks with known pathogens, United States, 1998-2012. Epidemiol. Infect. **145**: 316–325. doi:10.1017/S0950268816002375.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.-C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.-H., Thi, V.H.D., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H.D.D., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.-M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., and Wincker, P. 2014. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science (80-.). **345**: 950–953. doi:10.1126/science.1253435.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience **4**: 1–16. doi:10.1186/s13742-015-0047-8.

Chao, H., Wang, H., Wang, X., Guo, L., Gu, J., Zhao, W., Li, B., Chen, D., Raboanatahiry, N., and Li, M. 2017. Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. Sci. Rep. **7**: 1–16. Nature Publishing Group. doi:10.1038/srep46295.

Chardigny, J.-M.M., and Walrand, S. 2016. Plant protein for food: Opportunities and bottlenecks. OCL - Oilseeds fats, Crop. Lipids **23**: D404. doi:10.1051/ocl/2016019.

Chen, D., Shao, Q., Yin, L., Younis, A., and Zheng, B. 2019a. Polyamine function in plants: Metabolism, regulation on development, and roles in abiotic stress responses. Front. Plant Sci. **9**: 1–13. doi:10.3389/fpls.2018.01945.

Chen, F., Zhang, W., Yu, K., Sun, L., Gao, J., Zhou, X., Peng, Q., Fu, S., Hu, M., Long, W., Pu, H., Chen, S., Wang, X., and Zhang, J. 2018a. Unconditional and conditional QTL analyses of seed fatty acid composition in *Brassica napus* L. BMC Plant Biol. **18**: 1–14. BMC Plant Biology. doi:10.1186/s12870-018-1268-7.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. 2019b. CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. Annu. Rev. Plant Biol. **70**: 667–697.

doi:10.1146/annurev-arplant-050718-100049.

Chen, L., Wan, H., Qian, J., Guo, J., Sun, C., Wen, J., Yi, B., Ma, C., Tu, J., Song, L., Fu, T., and Shen, J. 2018b. Genome-wide association study of cadmium accumulation at the seedling stage in rapeseed (Brassica napus L.). Front. Plant Sci. **9**: 1–15. doi:10.3389/fpls.2018.00375.

Chen, P., Shen, Z., Ming, L., Li, Y., Dan, W., Lou, G., Peng, B., Wu, B., Li, Y., Zhao, D., Gao, G., Zhang, Q., Xiao, J., Li, X., Wang, G., and He, Y. 2018c. Genetic basis of variation in rice seed storage protein (Albumin, globulin, prolamin, and glutelin) content revealed by genome-wide association analysis. Front. Plant Sci. **9**: 1–15. doi:10.3389/fpls.2018.00612.

Chen, X., Snyder, C.L., Truksa, M., Shah, S., and Weselake, R.J. 2011a. Sn-glycerol-3-phosphate acyltransferases in plants. Plant Signal. Behav. **6**: 1695–1699. doi:10.4161/psb.6.11.17777.

Chen, X., Truksa, M., Shah, S., and Weselake, R.J. 2010. A survey of quantitative real-time polymerase chain reaction internal reference genes for expression studies in Brassica napus. Anal. Biochem. **405**: 138–140. Elsevier Inc. doi:10.1016/j.ab.2010.05.032.

Chen, X., Truksa, M., Snyder, C.L., El-Mezawy, A., Shah, S., and Weselake, R.J. 2011b. Three homologous genes encoding sn-glycerol-3-phosphate acyltransferase 4 exhibit different expression patterns and functional divergence in brassica napus. Plant Physiol. **155**: 851–865. doi:10.1104/pp.110.169482.

Cheung, L., Wanasundara, J., and Nickerson, M.T. 2014a. The Effect of pH and NaCl Levels on the Physicochemical and Emulsifying Properties of a Cruciferin Protein Isolate. Food Biophys. **9**: 105–113. doi:10.1007/s11483-013-9323-2.

Cheung, L., Wanasundara, J., and Nickerson, M.T. 2014b. Effect of pH and NaCl on the Emulsifying Properties of a Napin Protein Isolate. Food Biophys. **10**: 30–38. doi:10.1007/s11483-014-9350-7.

Chung, J., Babka, H.L., Graef, G.L., Staswick, P.E., Lee, D.J., Cregan, P.B., Shoemaker, R.C., and Specht, J.E. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. (Cell Biology & Molecular Genetics).PDF. Crop Sci. **43**: 1053–1067. doi:10.2135/cropsci2003.1053.

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., Sharpe, A.G., Ganal, M.W., and Parkin, I.A.P. 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. Theor. Appl. Genet. **129**: 1887–1899. Springer Berlin Heidelberg. doi:10.1007/s00122-016-2746-7.

Cobb, J.N., Biswas, P.S., and Platten, J.D. 2019. Back to the future: revisiting MAS as a tool for modern plant breeding. Theor. Appl. Genet. **132**: 647–667. Springer Berlin Heidelberg.

doi:10.1007/s00122-018-3266-4.

Cockram, J., and Mackay, I. 2018. Genetic Mapping Populations for Conducting High-Resolution Trait Mapping in Plants. Pages 109–138 *in* Advances in biochemical engineering/biotechnology. doi:10.1007/10_2017_48.

Conesa, A., and Götz, S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics **2008**. doi:10.1155/2008/619832.

Congdon, R.W., Muth, G.W., and Splittgerber, A.G. 1993. The binding interaction of Coomassie blue with proteins. Anal. Biochem. **213**: 407–413.

Constantine, N.T., Bansal, J., Zhang, X., Hyams, K.C., and Hayes, C. 1994. Enhanced chemiluminescence as a means of increasing the sensitivity of Western blot assays for HIV antibody. J. Virol. Methods **47**: 153–164. doi:10.1016/0166-0934(94)90074-4.

Cox, J., and Mann, M. 2011. Quantitative, high-resolution proteomics for data-driven systems biology. Annu. Rev. Biochem. **80**: 273–299. doi:10.1146/annurev-biochem-061308-093216.

Craig, B.M. 1971. Production and utilization of rapeseed in Canada. J. Am. Oil Chem. Soc. **48**: 737–739. doi:10.1007/BF02638532.

Crichton, M., Craven, D., Mackay, H., Marx, W., De Van Der Schueren, M., and Marshall, S. 2019. A systematic review, meta-analysis and meta-regression of the prevalence of protein-energy malnutrition: Associations with geographical region and sex. Age Ageing **48**: 38–48. doi:10.1093/ageing/afy144.

Crouch, M.L. 1982. Non-zygotic embryos of *Brassica napus* L. contain embryo-specific storage proteins. Planta **156**: 520–524. doi:10.1007/BF00392774.

Crouch, M.L., and Sussex, I.M. 1981. Development and storage-protein synthesis in Brassica napus L. embryos in vivo and in vitro. Planta **153**: 64–74.

Cruz-Suarez, L.E., Ricque-Marie, D., Tapia-Salazar, M., McCallum, I.M., and Hickling, D. 2001. Assessment of differently processed feed pea (*Pisum sativum*) meals and canola meal (*Brassica sp.*) in diets for blue shrimp (*Litopenaeus stylirostris*). Aquaculture **196**: 87–104. doi:10.1016/S0044-8486(00)00572-X.

Dalgalarrondo, M., Robin, J.-M., and Azanza, J.-L. 1986. Subunit composition of the globulin fraction of rapeseed (Brassica napus L.). Plant Sci. **43**: 115–124. doi:10.1016/0168-9452(86)90151-2.

Datta, S., Hettiarachchi, G.H.C.M., Deng, X.W., and Holm, M. 2006. Arabidopsis CONSTANS-LIKE3 is a positive regulator of red light signaling and root growth. Plant Cell **18**: 70–84. doi:10.1105/tpc.105.038182.

Daun, J.K. 1986. Glucosinolate levels in western Canadian rapeseed and canola. J. Am. Oil

Chem. Soc. **63**: 639–643. doi:10.1007/BF02638228.

Daun, J.K., Clear, K.M., and Williams, P. 1994. Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. J. Am. Oil Chem. Soc. **71**: 1063–1068. doi:10.1007/BF02675897.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A. V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., and Turnbaugh, P.J. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature **505**: 559–563. Nature Publishing Group. doi:10.1038/nature12820.

Day, L. 2013. Proteins from land plants - Potential resources for human nutrition and food security. Trends Food Sci. Technol. **32**: 25–42. Elsevier Ltd. doi:10.1016/j.tifs.2013.05.005.

Dekkers, B.L., Boom, R.M., and van der Goot, A.J. 2018. Structuring processes for meat analogues. Trends Food Sci. Technol. **81**: 25–36. Elsevier. doi:10.1016/j.tifs.2018.08.011.

Delisle, A.J., and Crouch, M.L. 1989. Seed Storage Protein Transcription and mRNA Levels in Brassica napus during Development and in Response to Exogenous Abscisic Acid . Plant Physiol. **91**: 617–623. doi:10.1104/pp.91.2.617.

Demirer, G.S., Zhang, H., Matos, J.L., Goh, N.S., Cunningham, F.J., Sung, Y., Chang, R., Aditham, A.J., Chio, L., Cho, M.-J., Staskawicz, B., and Landry, M.P. 2019. High aspect ratio nanomaterials enable delivery of functional genetic material without DNA integration in mature plants. Nat. Nanotechnol. Springer US. doi:10.1038/s41565-019-0382-5.

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

Earl, D.A., and VonHoldt, B.M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4**: 359–361. doi:10.1007/s12686-011-9548-7.

Edwards, D., Batley, J., and Snowdon, R.J. 2013. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. **126**: 1–11. doi:10.1007/s00122-012-1964-x.

EFSA Panel on Contaminants in the Food Chain 2016. Erucic acid in feed and food. EFSA J. **14**. doi:10.2903/j.efsa.2016.4593.

EFSA Panel on Dietetic Products Nutrition and Allergies 2013. Scientific Opinion on the safety of “rapeseed protein isolate” as a Novel Food ingredient. EFSA J. **11**: 1–23. doi:10.2903/j.efsa.2013.3420.

Eldakak, M., Milad, S.I.M., Nawar, A.I., and Rohila, J.S. 2013. Proteomics: A biotechnology tool for crop improvement. Front. Plant Sci. **4**: 1–12. doi:10.3389/fpls.2013.00035.

Elferjani, R., and Soolanayakanahally, R. 2018. Canola responses to drought, heat, and combined

stress: shared and specific effects on carbon assimilation, seed yield, and oil composition. Front. Plant Sci. **9**: 1–17. doi:10.3389/fpls.2018.01224.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6**: 1–10. doi:10.1371/journal.pone.0019379.

Environment Directorate, and Organisation for Economic Co-operation and Development 2001. Consensus document on key nutrient and key toxicants in low erucic acid rapeseed (canola). doi:ENV/JM/MONO(2007)10.

Ereken-Tumer, N., Richter, J.D., and Nielsen, N.C. 1982. Structural characterization of the glycinin precursors. J. Biol. Chem. **257**: 4016–4018. doi:10.1016/S0021-9258(18)34675-1.

Ericson, M.L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L.G., Rask, L., Rödin, J., Lenman, M., Glimelius, K., Josefsson, L.G., and Rask, L. 1986. Structure of the rapeseed 1.7 S storage protein, napin, and its precursor. J. Biol. Chem. **261**: 14576–14581. doi:10.1016/S0021-9258(18)66909-1.

Ersoz, E.S., Yu, J., and Buckler, E.S. 2009. Applications of Linkage Disequilibrium and Association Mapping in Maize. Pages 173–195 *in* Molecular Genetic Approaches to Maize Improvement. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-68922-5_13.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. **14**: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

Everaert, C., Luypaert, M., Maag, J.L.V., Cheng, Q.X., DInger, M.E., Hellemans, J., and Mestdagh, P. 2017. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci. Rep. **7**: 1–11. Springer US. doi:10.1038/s41598-017-01617-3.

Ezcurra, I., Wycliffe, P., Nehlin, L., Ellerström, M., and Rask, L. 2000. Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. Plant J. **24**: 57–66. doi:10.1046/j.1365-313X.2000.00857.x.

Falco, S.C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R.T., and Webber, P. 1995. Transgenic Canola and Soyban Seeds with Increased Lysine. Nat. Biotechnol. **13**: 577–582. doi:10.1038/nbt0695-577.

Fauteux, F., and Strömvik, M. V. 2009. Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. BMC Plant Biol. **9**: 1–11. doi:10.1186/1471-2229-9-126.

Fedoroff, N. V., Battisti, D.S., Beachy, R.N., Cooper, P.J.M., Fischhoff, D.A., Hodges, C.N.,

Knauf, V.C., Lobell, D., Mazur, B.J., Molden, D., Reynolds, M.P., Ronald, P.C., Rosegrant, M.W., Sanchez, P.A., Vonshak, A., and Zhu, J.-K. 2010. Radically rethinking agriculture for the 21st century. Science (80-.). **327**: 833–834. doi:10.1126/science.1186834.

Fernandez, D.E., Turner, F.R., and Crouch, M.L. 1991. In situ localization of storage protein mRNAs in developing meristems of *Brassica napus* embryos. Development **111**: 299–313.

Fetzer, A., Herfellner, T., Stäbler, A., Menner, M., and Eisner, P. 2018. Influence of process conditions during aqueous protein extraction upon yield from pre-pressed and cold-pressed rapeseed press cake. Ind. Crops Prod. **112**: 236–246. Elsevier. doi:10.1016/j.indcrop.2017.12.011.

Finkelstein, R.R., Tenbarge, K.M., Shumway, J.E., and Crouch, M.L. 1985. Role of ABA in maturation of rapeseed embryos. Plant Physiol. **78**: 630–636. doi:10.1104/pp.78.3.630.

Finlayson, A.J., Krzymanski, J., and Downey, R.K. 1973. Comparison of chemical and agronomic characteristics of two *Brassica napus* L. cultivars, Bronowski and Target. J. Am. Oil Chem. Soc. **50**: 407–410. doi:10.1007/BF02641819.

Fiocchi, A., Dahdah, L., Riccardi, C., Mazzina, O., and Fierro, V. 2016. Preacutionary labelling of cross-reactive foods: The case of rapeseed. Asthma Res. Pract. **2**: 1–8. Asthma Research and Practice. doi:10.1186/s40733-016-0028-4.

Fitton, N., Alexander, P., Arnell, N., Bajzelj, B., Calvin, K., Doelman, J., Gerber, J.S., Havlik, P., Hasegawa, T., Herrero, M., Krisztin, T., van Meijl, H., Powell, T., Sands, R., Stehfest, E., West, P.C., and Smith, P. 2019. The vulnerabilities of agricultural land and food production to future water scarcity. Glob. Environ. Chang. **58**: 101944. Elsevier Ltd. doi:10.1016/j.gloenvcha.2019.101944.

Fleddermann, M., Fechner, A., Rößler, A., Bähr, M., Pastor, A., Liebert, F., and Jahreis, G. 2013. Nutritional evaluation of rapeseed protein compared to soy protein for quality, plasma amino acids, and nitrogen balance - A randomized cross-over intervention study in humans. Clin. Nutr. **32**: 519–526. Elsevier Ltd. doi:10.1016/j.clnu.2012.11.005.

Foegeding, E.A., and Davis, J.P. 2011. Food protein functionality: A comprehensive approach. Food Hydrocoll. **25**: 1853–1864. Elsevier Ltd. doi:10.1016/j.foodhyd.2011.05.008.

Food and Agriculture Organization 2016. Pulses Nutritious Seeds for a Sustainable Future. United Nations. [Online] Available: https://www.fao.org/3/i5528e/i5528e.pdf.

Food and Agriculture Organization of the United Nations 2003. Trade reforms and food security - Conceptualizing the linkages. Rome.

Food and Agriculture Organization of the United Nations 2020. FAOSTAT Database. [Online] Available: http://www.fao.org/faostat/en/#data/QC.

Food and Agriculture Organization of the United Nations, and FAO 2018. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. Rome.

doi:10.1093/japr/3.1.101.

Food Security Information Network 2018. Global report on food crises 2018. [Online] Available: http://www.fao.org/3/a-br323e.pdf.

Fredua-Agyeman, R., Yu, Z., Hwang, S.F., and Strelkov, S.E. 2020. Genome-Wide Mapping of Loci Associated With Resistance to Clubroot in Brassica napus ssp. napobrassica (Rutabaga) Accessions From Nordic Countries. Front. Plant Sci. **11**: 1–17. doi:10.3389/fpls.2020.00742.

Fu, Y.B. 2015. Understanding crop genetic diversity under modern plant breeding. Theor. Appl. Genet. **128**: 2131–2142. Springer Berlin Heidelberg. doi:10.1007/s00122-015-2585-y.

Fu, Y.B., and Gugel, R.K. 2010. Genetic diversity of Canadian elite summer rape (Brassica napus L.) cultivars from the pre- to post-canola quality era. Can. J. Plant Sci. **90**: 23–33. doi:10.4141/CJPS09073.

Fujiwara, T., Nambara, E., Yamagishi, K., Goto, D.B., and Naito, S. 2002. Storage Proteins. Arab. B. **1**: e0020. doi:10.1199/tab.0020.

Fukushima, D. 1969. Denaturation of Soybean proteins by organic solvents.

Gabur, I., Chawla, H.S., Liu, X., Kumar, V., Faure, S., von Tiedemann, A., Jestin, C., Dryzska, E., Volkmann, S., Breuer, F., Delourme, R., Snowdon, R., and Obermeier, C. 2018. Finding invisible quantitative trait loci with missing data. Plant Biotechnol. J. **16**: 2102–2112. doi:10.1111/pbi.12942.

Gabur, I., Chawla, H.S., Lopisso, D.T., von Tiedemann, A., Snowdon, R.J., and Obermeier, C. 2020. Gene presence-absence variation associates with quantitative Verticillium longisporum disease resistance in Brassica napus. Sci. Rep. **10**: 1–11. doi:10.1038/s41598-020-61228-3.

Gabur, I., Chawla, H.S., Snowdon, R.J., and Parkin, I.A.P. 2019. Connecting genome structural variation with complex traits in crop plants. Theor. Appl. Genet. **132**: 733–750. Springer Berlin Heidelberg. doi:10.1007/s00122-018-3233-0.

Gajardo, H.A., Wittkop, B., Soto-Cerda, B., Higgins, E.E., Parkin, I.A.P., Snowdon, R.J., Federico, M.L., and Iniguez-Luy, F.L. 2015. Association mapping of seed quality traits in Brassica napus L. using GWAS and candidate QTL approaches. Mol. Breed. **35**: 1–19. Springer Netherlands. doi:10.1007/s11032-015-0340-3.

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., and Rajjou, L. 2014. Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. Mol. Cell. Proteomics **13**: 252–268. doi:10.1074/mcp.M113.032227.

Gao, C., Qi, S., Liu, K., Li, D., Jin, C., Li, Z., Huang, G., Hai, J., Zhang, M., and Chen, M. 2016a. MYC2, MYC3, and MYC4 function redundantly in seed storage protein

accumulation in *Arabidopsis*. Plant Physiol. Biochem. **108**: 63–70. Elsevier Masson SAS. doi:10.1016/j.plaphy.2016.07.004.

Gao, X., Chen, J., Dai, X., Zhang, D., and Zhao, Y. 2016b. An effective strategy for reliably isolating heritable and Cas9-free arabidopsis mutants generated by CRISPR/Cas9-mediated genome editing. Plant Physiol. **171**: 1794–1800. doi:10.1104/pp.16.00663.

de Gavelle, E., Huneau, J.-F., Bianchi, C., Verger, E., and Mariotti, F. 2017. Protein adequacy is primarily a matter of protein quantity, not quality: modeling an increase in plant:animal protein ratio in French adults. Nutrients **9**: 1333. doi:10.3390/nu9121333.

Gazave, E., Tassone, E.E., Ilut, D.C., Wingerson, M., Datema, E., Witsenboer, H.M.A., Davis, J.B., Grant, D., Dyer, J.M., Jenks, M.A., Brown, J., and Gore, M.A. 2016. Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. Front. Plant Sci. **7**: 1–16. doi:10.3389/fpls.2016.00525.

Ge, S.X., Jung, D., and Yao, R. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics **36**: 2628–2629. doi:10.1093/bioinformatics/btz931.

Geng, X., Dong, N., Wang, Y., Li, G., Wang, L., Guo, X., Li, J., Wen, Z., and Wei, W. 2018. RNA-seq transcriptome analysis of the immature seeds of two *Brassica napus* lines with extremely different thousand-seed weight to identify the candidate genes related to seed weight. PLoS One **13**: e0191297. doi:10.1371/journal.pone.0191297.

Gerbens-Leenes, P.W., Nonhebel, S., and Krol, M.S. 2010. Food consumption patterns and economic growth. Increasing affluence and the use of natural resources. Appetite **55**: 597–608. Elsevier Ltd. doi:10.1016/j.appet.2010.09.013.

Gerten, D., Heck, V., Jägermeyr, J., Bodirsky, B.L., Fetzer, I., Jalava, M., Kummu, M., Lucht, W., Rockström, J., Schaphoff, S., and Schellnhuber, H.J. 2020. Feeding ten billion people is possible within four terrestrial planetary boundaries. Nat. Sustain. **3**: 200–208. doi:10.1038/s41893-019-0465-1.

Gerzhova, A., Mondor, M., Benali, M., and Aider, M. 2016. Study of total dry matter and protein extraction from canola meal as affected by the pH, salt addition and use of zeta-potential/turbidimetry analysis to optimize the extraction conditions. Food Chem. **201**: 243–252. Elsevier Ltd. doi:10.1016/j.foodchem.2016.01.074.

Gil-Humanes, J., Pistón, F., Giménez, M.J., Martín, A., and Barro, F. 2012. The Introgression of RNAi Silencing of γ-Gliadins into Commercial Lines of Bread Wheat Changes the Mixing and Technological Properties of the Dough. PLoS One **7**. doi:10.1371/journal.pone.0045937.

Gil-Humanes, J., Piston, F., Tollefsen, S., Sollid, L.M., and Barro, F. 2010. Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. Proc. Natl. Acad. Sci. **107**: 17023–17028. doi:10.1073/pnas.1007773107.

Gilani, G.S., Cockell, K.A., and Sepehr, E. 2005. Effects of antinutritional factors on protein digestibility and amino acid availability in foods. J. AOAC Int. **88**: 967–987. doi:10.1093/jaoac/88.3.967.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., and Toulmin, C. 2010. Food security: The challenge of feeding 9 billion people. Science (80-.). **327**: 812–818. doi:10.1126/science.1185383.

Goering, K.J., Eslick, R., and Brelsford, D.L. 1965. A search for high erucic acid containing oils in the cruciferae. Econ. Bot. **19**: 251–256. doi:10.1007/BF02914311.

Goossens, A., Van Montagu, M., and Angenon, G. 1999. Co-introduction of an antisense gene for an endogenous seed storage protein can increase expression of a transgene in *Arabidopsis thaliana* seeds. FEBS Lett. **456**: 160–164. doi:10.1016/S0014-5793(99)00943-6.

Gould, H.J., and Sutton, B.J. 2008. IgE in allergy and asthma today. Nat. Rev. Immunol. **8**: 205–217. doi:10.1038/nri2273.

Government of Canada 2011. Regulations amending the food and drug regulations (1220 - Enhanced labeling for food allergen and gluten sources and added sulphites). Canada Gaz. Part II **145**: 325.

Graham, F., Russell, J., Holdsworth, M., Menon, M., and Barker, M. 2019. Exploring the relationship between environmental impact and nutrient content of sandwiches and beverages available in cafés in a UK university. Sustain. **11**. doi:10.3390/su11113190.

Graham, I.A. 2008. Seed storage oil mobilization. Annu. Rev. Plant Biol. **59**: 115–142. doi:10.1146/annurev.arplant.59.032607.092938.

Grant, C.A., Clayton, G.W., and Johnston, A.M. 2003. Sulphur fertilizer and tillage effects on canola seed quality in the Black soil zone of western Canada. Can. J. Plant Sci. **83**: 745–758. doi:10.4141/P02-107.

Griffith, M., Griffith, O.L., Mwenifumbo, J., Goya, R., Morrissy, A.S., Morin, R.D., Corbett, R., Tang, M.J., Hou, Y.C., Pugh, T.J., Robertson, G., Chittaranjan, S., Ally, A., Asano, J.K., Chan, S.Y., Li, H.I., McDonald, H., Teague, K., Zhao, Y., Zeng, T., Delaney, A., Hirst, M., Morin, G.B., Jones, S.J.M., Tai, I.T., and Marra, M.A. 2010. Alternative expression analysis by RNA sequencing. Nat. Methods **7**: 843–847. doi:10.1038/nmeth.1503.

Gruis, D., Selinger, D.A., Curran, J.M., and Jung, R. 2002. Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell **14**: 2863–2882. doi:10.1105/tpc.005009.

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., and Qu, C. 2019. Association mapping analysis of fatty acid content in different ecotypic rapeseed using mrMLM. Front. Plant Sci. **9**.

doi:10.3389/fpls.2018.01872.

Guerche, P., De Almeida, E.R.P., Schwarztein, M.A., Gander, E., Krebbers, E., and Pelletier, G. 1990. Expression of the 2S albumin from Bertholletia excelsa in Brassica napus. MGG Mol. Gen. Genet. **221**: 306–314. doi:10.1007/BF00259393.

Gyawali, S., Hegedus, D.D., Parkin, I.A.P., Poon, J., Higgins, E., Horner, K., Bekkaoui, D., Coutu, C., and Buchwaldt, L. 2013. Genetic Diversity and Population Structure in a World Collection of Brassica napus Accessions with Emphasis on South Korea, Japan, and Pakistan. Crop Sci. **53**: 1537–1545. doi:10.2135/cropsci2012.10.0614.

Gylling, H. 2006. Rapeseed oil does not cause allergic reactions. Allergy Eur. J. Allergy Clin. Immunol. **61**: 895. doi:10.1111/j.1398-9995.2006.01119.x.

Hajduch, M., Casteel, J.E., Hurrelmeyer, K.E., Song, Z., Agrawal, G.K., and Thelen, J.J. 2006. Proteomic analysis of seed filling in *Brassica napus*. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol. **141**: 32–46. doi:10.1104/pp.105.075390.

Hannoufa, A., Pillai, B.V.S., and Chellamma, S. 2014. Genetic enhancement of *Brassica napus* seed quality. Transgenic Res. **23**: 39–52. doi:10.1007/s11248-013-9742-3.

Harloff, H.J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J.G., Dreyer, F., Leckband, G., and Jung, C. 2012. A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. Theor. Appl. Genet. **124**: 957–969. doi:10.1007/s00122-011-1760-z.

Harvey, B.L., and Downey, R.K. 1964. The Inheritance of Erucic Acid Content in Rapeseed (Brassica napus). Can. J. Plant Sci. **44**: 104–111. doi:10.4141/cjps64-019.

Hatzig, S., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G., Abbadi, A., and Snowdon, R.J. 2018. Hidden effects of seed quality breeding on germination in oilseed rape (Brassica napus L.). Front. Plant Sci. **9**: 1–13. doi:10.3389/fpls.2018.00419.

Hatzig, S. V., Frisch, M., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G., Abbadi, A., and Snowdon, R.J. 2015. Genome-wide association mapping unravels the genetic control of seed germination and vigor in Brassica napus. Front. Plant Sci. **6**: 1–13. doi:10.3389/fpls.2015.00221.

He, M.L., Gibb, D., McKinnon, J.J., and McAllister, T.A. 2013. Effect of high dietary levels of canola meal on growth performance, carcass quality and meat fatty acid profiles of feedlot cattle. Can. J. Anim. Sci. **93**: 269–280. doi:10.4141/cjas2012-090.

He, R., He, H.Y., Chao, D., Ju, X., and Aluko, R. 2014. Effects of High Pressure and Heat Treatments on Physicochemical and Gelation Properties of Rapeseed Protein Isolate. Food Bioprocess Technol. **7**: 1344–1353. doi:10.1007/s11947-013-1139-z.

He, Y., Hu, D., You, J., Wu, D., Cui, Y., Dong, H., Li, J., and Qian, W. 2019. Genome-wide

association study and protein network analysis for understanding candidate genes involved in root development at the rapeseed seedling stage. Plant Physiol. Biochem. **137**: 42–52. Elsevier. doi:10.1016/j.plaphy.2019.01.028.

He, Y., Wu, D., Wei, D., Fu, Y., Cui, Y., Dong, H., Tan, C., and Qian, W. 2017. GWAS, QTL mapping and gene expression analyses in Brassica napus reveal genetic control of branching morphogenesis. Sci. Rep. **7**: 1–9. Springer US. doi:10.1038/s41598-017-15976-4.

He, Y., Yang, X., Xia, J., Zhao, L., and Yang, Y. 2016. Consumption of meat and dairy products in China: A review. Proc. Nutr. Soc. **75**: 385–391. doi:10.1017/S0029665116000641.

Health Canada 2009. Mustard : A Priority Food Allergen in Canada.

Health Canada 2019. Canada's dietary guidelines for health professionals and policy makers. Health Canada. Ottawa. [Online] Available: https://food-guide.canada.ca/static/assets/pdf/CDG-EN-2018.pdf.

van Heerwaarden, J., Hufford, M.B., and Ross-Ibarra, J. 2012. Historical genomics of North American maize. Proc. Natl. Acad. Sci. U. S. A. **109**: 12420–12425. doi:10.1073/pnas.1209275109.

Hegedus, D.D., Baron, M., Labbe, N., Coutu, C., Lydiate, D., Lui, H., and Rozwadowski, K. 2014. A strategy for targeting recombinant proteins to protein storage vacuoles by fusion to *Brassica napus* napin in napin-depleted seeds. Protein Expr. Purif. **95**: 162–168. Elsevier Inc. doi:10.1016/j.pep.2013.12.009.

Henderson, I.R., and Salt, D.E. 2017. Natural genetic variation and hybridization in plants. J. Exp. Bot. **68**: 5415–5417. doi:10.1093/jxb/erx377.

Herman, El.M., and Larkins, B.A. 1999. Protein storage bodies and vacoules. Plant Cell **11**: 601–613. doi:10.1105/tpc.11.4.601.

Hoek, A.C., Luning, P.A., Weijzen, P., Engels, W., Kok, F.J., and de Graaf, C. 2011. Replacement of meat by meat substitutes. A survey on person- and product-related factors in consumer acceptance. Appetite **56**: 662–673. Elsevier Ltd. doi:10.1016/j.appet.2011.02.001.

Hoglund, A.-S., Rodin, J., Larsson, E., Rask, L., Höglund, A.S., Rödin, J., Larsson, E., and Rask, L. 1992. Distribution of napin and cruciferin in developing rape seed embryos. Plant Physiol. **98**: 509–515. doi:10.1104/pp.98.2.509.

Hossain, M.A.A., Islam, A.F.F., and Iji, P.A.A. 2012. Energy utilization and performance of broiler chickens raised on diets with vegetable proteins or conventional feeds. Asian J. Poult. Sci. **6**: 117–128. doi:10.3923/ajpsaj.2012.117.128.

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

Hu, L., Zhang, H., Yang, Q., Meng, Q., Han, S., Nwafor, C.C., Khan, M.H.U., Fan, C., and Zhou, Y. 2018. Promoter variations in a homeobox gene, *BnA10.LMI1*, determine lobed leaves in rapeseed (*Brassica napus* L.). Theor. Appl. Genet. **131**: 2699–2708. Springer Berlin Heidelberg. doi:10.1007/s00122-018-3184-5.

Hu, Y., Wu, G., Cao, Y., Wu, Y., Xiao, L., Li, X., and Lu, C. 2009. Breeding response of transcript profiling in developing seeds of *Brassica napus*. BMC Mol. Biol. **10**. doi:10.1186/1471-2199-10-49.

Hu, Z.Y., Hua, W., Zhang, L., Deng, L. Bin, Wang, X.F., Liu, G.H., Hao, W.J., and Wang, H.Z. 2013. Seed structure characteristics to form ultrahigh oil content in rapeseed. PLoS One **8**: 2–11. doi:10.1371/journal.pone.0062099.

Huang, S., Adams, W.R., Zhou, Q., Malloy, K.P., Voyles, D.A., Anthony, J., Kriz, A.L., and Luethy, M.H. 2004. Improving Nutritional Quality of Maize Proteins by Expressing Sense and Antisense Zein Genes. J. Agric. Food Chem. **52**: 1958–1964. doi:10.1021/jf0342223.

Huang, X., and Han, B. 2014. Natural Variations and Genome-Wide Association Studies in Crop Plants. Annu. Rev. Plant Biol. **65**: 531–551. doi:10.1146/annurev-arplant-050213-035715.

Hughes, G.J., Ryan, D.J., Mukherjea, R., and Schasteen, C.S. 2011. Protein digestibility-corrected amino acid scores (PDCAAS) for soy protein isolates and concentrate: Criteria for evaluation. J. Agric. Food Chem. **59**: 12707–12712. doi:10.1021/jf203220v.

Huhtanen, P., Hetta, M., and Swensson, C. 2012. Evaluation of canola meal as a protein supplement for dairy cows: A review and a meta-analysis. Can. J. Anim. Sci. **91**: 529–543. doi:10.4141/cjas2011-029.

Hundertmark, M., and Hincha, D.K. 2008. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics **9**: 1–22. doi:10.1186/1471-2164-9-118.

Hwalla, N., El Labban, S., and Bahn, R.A. 2016. Nutrition security is an integral component of food security. Front. Life Sci. **9**: 167–172. doi:10.1080/21553769.2016.1209133.

Ilić-Grubor, K., Attree, S.M., and Fowke, L.C. 1998. Comparative morphological study of zygotic and microspore-derived embryos of *Brassica napus* L. as revealed by scanning electron microscopy. Ann. Bot. **82**: 157–165. doi:10.1006/anbo.1998.0661.

International Rapeseed Congress 2019. 15th International Rapeseed Congress Congress Guide. Pages 1–62 *in*. International Rapeseed Congress, Berlin. [Online] Available: https://www.irc2019-berlin.com/application/files/4915/6059/7572/IRC_Congress_Guide_150619_1630.pdf.

J., B., and Cloutier, S. 2012. Association Mapping in Plant Genomes. Page *in* Genetic Diversity in Plants. InTech, London. doi:10.5772/33005.

Jan, H.U., Guan, M., Yao, M., Liu, W., Wei, D., Abbadi, A., Zheng, M., He, X., Chen, H., Guan,

C., Nichols, R.A., Snowdon, R.J., Hua, W., and Qian, L. 2019. Genome-wide haplotype analysis improves trait predictions in Brassica napus hybrids. Plant Sci. **283**: 157–164. Elsevier. doi:10.1016/j.plantsci.2019.02.007.

Janes, K.A. 2015. An analysis of critical factors for quantitative immunoblotting. Sci. Signal. **8**: rs2–rs2. doi:10.1126/scisignal.2005966.

Jasinski, S., Chardon, F., Nesi, N., Lécureuil, A., and Guerche, P. 2018. Improving seed oil and protein content in Brassicaceae: Some new genetic insights from Arabidopsis thaliana. OCL - Oilseeds fats, Crop. Lipids **25**: 0–6. doi:10.1051/ocl/2018047.

Jensen, S.K., Liu, Y.-G., and Eggum, B.O. 1995. The effect of heat treatment on glucosinolates and nutritional value of rapeseed meal in rats. Anim. Feed Sci. Technol. **53**: 17–28. doi:10.1016/0377-8401(94)00740-Z.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive BActerial Immunity. Science (80-.). **337**: 816–822. doi:10.1126/science.1225829.

Job, C., Rajjou, L., Lovigny, Y., Belghazi, M., Job, D., Loïc, R., Yoann, L., Maya Belghazi, Dominique, J., Lovigny, Y., Belghazi, M., Job, D., and Job, C. 2005. Patterns of protein oxidation in Arabidopsis seeds and during germination. Plant Physiol. **138**: 790–802. doi:10.1104/pp.105.062778.ing.

Jofuku, K.D., Omidyar, P.K., Gee, Z., and Okamuro, J.K. 2005. Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proc. Natl. Acad. Sci. U. S. A. **102**: 3117–3122. doi:10.1073/pnas.0409893102.

Jolivet, P., Boulard, C., Bellamy, A., Larré, C., Barre, M., Rogniaux, H., D'Andréa, S., Chardot, T., and Nesi, N. 2009. Protein composition of oil bodies from mature *Brassica napus* seeds. Proteomics **9**: 3268–3284. doi:10.1002/pmic.200800449.

Jolivet, P., Deruyffelaere, C., Boulard, C., Quinsac, A., Savoire, R., Nesi, N., and Chardot, T. 2013. Deciphering the structural organization of the oil bodies in the Brassica napus seed as a mean to improve the oil extraction yield. Ind. Crops Prod. **44**: 549–557. Elsevier B.V. doi:10.1016/j.indcrop.2012.09.024.

Jones, O.G. 2016. Recent advances in the functionality of non-animal-sourced proteins contributing to their use in meat analogs. Curr. Opin. Food Sci. **7**: 7–13. Elsevier Ltd. doi:10.1016/j.cofs.2015.08.002.

José-Estanyol, M., Gomis-Rüth, F.X., and Puigdomènech, P. 2004. The eight-cysteine motif, a versatile structure in plant proteins. Plant Physiol. Biochem. **42**: 355–365. doi:10.1016/j.plaphy.2004.03.009.

Josefsson, L.G., Lenman, M., Ericson, M.L., and Rask, L. 1987. Structure of a gene encoding the 1.7 S storage protein, napin, from *Brassica napus*. J. Biol. Chem. **262**: 12196–12201. doi:10.1016/s0021-9258(18)45336-7.

Jung, R., Nam, Y.-W., Saalbach, I., MÜntz, L., and Nielsen, N.C. 1997. Role of the sulfhydryl redox state and disulfide bonds in processing and assembly of 11S seed globulins. Plant Cell **9**: 2037–2050. doi:10.1105/tpc.9.11.2037.

Jung, R., Paul Scott, M., Nam, Y.W., Beaman, T.W., Bassüner, R., Saalbach, I., Müntz, K., and Nielsen, N.C. 1998. The role of proteolysis in the processing and assembly of 11S seed globulins. Plant Cell **10**: 343–357. doi:10.1105/tpc.10.3.343.

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., and Hattori, T. 2005. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of *FUSCA3* and *ABSCISIC ACID INSENSITIVE3*. Plant Cell Physiol. **46**: 399–406. doi:10.1093/pcp/pci048.

Kaler, A.S., Gillman, J.D., Beissinger, T., and Purcell, L.C. 2020. Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. Front. Plant Sci. **10**: 1–13. doi:10.3389/fpls.2019.01794.

Kambhampati, S., Aznar-Moreno, J.A., Hostetler, C., Caso, T., Bailey, S.R., Hubbard, A.H., Durrett, T.P., and Allen, D.K. 2020. On the inverse correlation of protein and oil: Examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants. Metabolites **10**: 1–15. doi:10.3390/metabo10010018.

Kasprzak, M.M., Houdijk, J.G.M., Liddell, S., Davis, K., Olukosi, O.A., Kightley, S., White, G.A., and Wiseman, J. 2017. Rapeseed napin and cruciferin are readily digested by poultry. J. Anim. Physiol. Anim. Nutr. (Berl). **101**: 658–666. doi:10.1111/jpn.12576.

Kasprzak, M.M., Houdijk, J.G.M., Olukosi, O.A., Appleyard, H., Kightley, S.P.J., Carré, P., Sutton, T., and Wiseman, J. 2018. The content and standardized ileal digestibility of crude protein and amino acids in rapeseed co-products fed to pigs. Livest. Sci. **208**: 22–27. Elsevier B.V. doi:10.1016/j.livsci.2017.11.024.

Kawagoe, Y., Suzuki, K., Tasaki, M., Yasuda, H., Akagi, K., Katoh, E., Nishizawa, N.K., Ogawa, M., and Takaiwa, F. 2005. The critical role of disulfide bond formation in protein sorting in the endosperm of rice. Plant Cell **17**: 1141–1153. doi:10.1105/tpc.105.030668.

Kawakatsu, T., and Takaiwa, F. 2010. Cereal seed storage protein synthesis: Fundamental processes for recombinant protein production in cereal grains. Plant Biotechnol. J. **8**: 939–953. doi:10.1111/j.1467-7652.2010.00559.x.

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

Khattab, R.Y., and Arntfield, S.D. 2009. Functional properties of raw and processed canola meal. LWT - Food Sci. Technol. **42**: 1119–1124. Elsevier Ltd. doi:10.1016/j.lwt.2009.02.009.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. 2019a. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**: 907–

915. Springer US. doi:10.1038/s41587-019-0201-4.

Kim, H.J., Lee, J.Y., Yoon, U.H., Lim, S.H., and Kim, Y.M. 2013a. Effects of reduced prolamin on seed storage protein composition and the nutritional quality of rice. Int. J. Mol. Sci. **14**: 17073–17084. doi:10.3390/ijms140817073.

Kim, J.-S., Cho, S.W., Kim, S.-T., Kim, H., Woo, J.W., Choe, S., Kwon, S. Il, Kim, S.-G., Corvalán, C., and Kim, J. 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat. Biotechnol. **33**: 1162–1164. doi:10.1038/nbt.3389.

Kim, J.H., Varankovich, N. V., Stone, A.K., and Nickerson, M.T. 2016. Nature of protein-protein interactions during the gelation of canola protein isolate networks. Food Res. Int. **89**: 408–414. Elsevier Ltd. doi:10.1016/j.foodres.2016.08.018.

Kim, S.W., Less, J.F., Wang, L., Yan, T., Kiron, V., Kaushik, S.J., and Lei, X.G. 2019b. Meetin global feed protein demand: Challenge, opportunity, and strategy. Annu. Rev. Anim. Biosci. **7**: 221–243. doi:10.1146/annurev-animal-030117-014838.

Kim, W.-S., Jez, J.M., and Krishnan, H.B. 2014. Effects of proteome rebalancing and sulfur nutrition on the accumulation of methionine rich δ-zein in transgenic soybeans. Front. Plant Sci. **5**: 1–12. doi:10.3389/fpls.2014.00633.

Kim, W.S., Gillman, J.D., and Krishnan, H.B. 2013b. Identification of a plant introduction soybean line with genetic lesions affecting two distinct glycinin subunits and evaluation of impacts on protein content and composition. Mol. Breed. **32**: 291–298. doi:10.1007/s11032-013-9870-8.

Kim, W.S., and Krishnan, H.B. 2004. Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. Plant Biotechnol. J. **2**: 199–210. doi:10.1111/j.1467-7652.2004.00063.x.

Klassen, A.J., Downey, R.K., and Capcara, J.J. 1987. Westar summer rape. Can. J. Plant Sci. **67**: 491–493. doi:10.4141/cjps87-068.

Kohno-Murase, J., Murase, M., Ichikawa, H., and Imamura, J. 1994. Effects of an antisense napin gene on seed storage compounds in transgenic *Brassica napus* seeds. Plant Mol. Biol. **26**: 1115–1124. doi:10.1007/BF00040693.

Kohno-Murase, J., Murase, M., Ichikawa, H., and Imamura, J. 1995. Improvement in the quality of seed storage protein by transformation of *Brassica napus* with an antisense gene for cruciferin. Theor. Appl. Genet. **91**: 627–631. doi:10.1007/BF00223289.

Kolde, R. 2019. Package 'pheatmap.' : 1–8.

Kovaka, S., Zimin, A. V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. **20**: 278. Genome Biology. doi:10.1186/s13059-019-1910-1.

Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. 2003. Regulation of storage protein gene expression in Arabidopsis. Development **130**: 6065–6073. doi:10.1242/dev.00814.

Kubala, S., Garnczarska, M., Wojtyla, Ł., Clippe, A., Kosmala, A., Żmieńko, A., Lutts, S., and Quinet, M. 2015. Deciphering priming-induced improvement of rapeseed (*Brassica napus* L.) germination through an integrated transcriptomic and proteomic approach. Plant Sci. **231**: 94–113. doi:10.1016/j.plantsci.2014.11.008.

Kumar, P., Chatli, M.K., Mehta, N., Singh, P., Malav, O.P., and Verma, A.K. 2017. Meat analogues: Health promising sustainable meat substitutes. Crit. Rev. Food Sci. Nutr. **57**: 923–932. Taylor & Francis. doi:10.1080/10408398.2014.939739.

L'Hocine, L., Pitre, Achouri, A., Pitre, M., Achouri, A., Pitre, and Achouri, A. 2019. Detection and identification of allergens from Canadian mustard varieties of *Sinapis alba* and *Brassica juncea*. Biomolecules **9**: 489. doi:10.3390/biom9090489.

Lancashire, P.D., Bleiholder, H., Boom, T. Van Den, Langelüddeke, P., Stauss, R., Weber, E., and Witzenberger, A. 1991. A uniform decimal code for growth stages of crops and weeds. Ann. Appl. Biol. **119**: 561–601. doi:10.1111/j.1744-7348.1991.tb04895.x.

Lander, E.S., and Schork, N.J. 1996. Genetic dissection of complex traits. Science (80-.).: 2037–2048.

Landero, J.L., Beltranena, E., and Zijlstra, R.T. 2012. Growth performance and preference studies to evaluate solvent-extracted *Brassica napus* or *Brassica juncea* canola meal fed to weaned pigs. J. Anim. Sci. **90**: 406–408. doi:10.2527/jas.53955.

Langner, T., Kamoun, S., and Belhaj, K. 2018. CRISPR crops: Plant genome editing toward disease resistance. Annu. Rev. Phytopathol. **56**: 479–512. doi:10.1146/annurev-phyto-080417-050158.

Lara, P., Oñate-Sánchez, L., Abraham, Z., Ferrándiz, C., Díaz, I., Carbonero, P., and Vicente-Carbajosa, J. 2003. Synergistic activation of seed storage protein gene expression in *Arabidopsis* by ABI3 and two bZIPs related to OPAQUE2. J. Biol. Chem. **278**: 21003–21011. doi:10.1074/jbc.M210538200.

Lau, S., Slane, D., Herud, O., Kong, J., and Jürgens, G. 2012. Early embryogenesis in flowering plants: setting up the basic body pattern. Annu. Rev. Plant Biol. **63**: 483–506. doi:10.1146/annurev-arplant-042811-105507.

Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K., and Ritchie, M.E. 2018. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research **5**: 1408. doi:10.12688/f1000research.9005.3.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. 2014. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. **15**: 1–17. doi:10.1186/gb-2014-15-2-r29.

Ledford, H. 2013. US regulation misses some GM crops. Nature **500**: 389–390. doi:10.1038/500389a.

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L.K., and Mian, M.A.R. 2019. Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. Theor. Appl. Genet. **132**: 1639–1659. Springer Berlin Heidelberg. doi:10.1007/s00122-019-03304-5.

Li, C., Hao, M., Wang, W., Wang, H., Chen, F., Chu, W., Zhang, B., Mei, D., Cheng, H., and Hu, Q. 2018. An Efficient CRISPR/Cas9 Platform for Rapidly Generating Simultaneous Mutagenesis of Multiple Gene Homoeologs in Allotetraploid Oilseed Rape. Front. Plant Sci. **9**: 1–9. doi:10.3389/fpls.2018.00442.

Li, F., Chen, B., Xu, K., Gao, G., Yan, G., Qiao, J., Li, J., Li, H., Li, L., Xiao, X., Zhang, T., Nishio, T., and Wu, X. 2016a. A genome-wide association study of plant height and primary branch number in rapeseed (Brassica napus). Plant Sci. **242**: 169–177. Elsevier Ireland Ltd. doi:10.1016/j.plantsci.2015.05.012.

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., and Wu, X. 2014. Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (Brassica napus L.). DNA Res. **21**: 355–367. doi:10.1093/dnares/dsu002.

Li, H., Zhang, L., Hu, J., Zhang, F., Chen, B., Xu, K., Gao, G., Li, H., Zhang, T., Li, Z., and Wu, X. 2017a. Genome-wide association mapping reveals the genetic control underlying branch angle in rapeseed (Brassica napus L.). Front. Plant Sci. **8**. doi:10.3389/fpls.2017.01054.

Li, J., Bus, A., Spamer, V., and Stich, B. 2016b. Comparison of statistical models for nested association mapping in rapeseed (Brassica napus L.) through computer simulations. BMC Plant Biol. **16**. BMC Plant Biology. doi:10.1186/s12870-016-0707-6.

Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., Huang, Q., Xiao, X., Zhang, T., Hu, J., Li, F., and Wu, X. 2016c. A genome-wide association study reveals new loci for resistance to clubroot disease in Brassica napus. Front. Plant Sci. **7**: 1–12. doi:10.3389/fpls.2016.01483.

Li, S., Zhu, Y., Varshney, R.K., Zhan, J., Zheng, X., Shi, J., Wang, X., Liu, G., and Wang, H. 2020. A systematic dissection of the mechanisms underlying the natural variation of silique number in rapeseed (Brassica napus L.) germplasm. Plant Biotechnol. J. **18**: 568–580. doi:10.1111/pbi.13224.

Li, W.P., Shi, H. Bin, Zhu, K., Zheng, Q., and Xu, Z. 2017b. The quality of sunflower seed oil changes in response to nitrogen fertilizer. Agron. J. **109**: 2499–2507. doi:10.2134/agronj2017.01.0046.

Li, X., Shao, Z., Tian, R., Zhang, H., Du, H., Kong, Y., Li, W., and Zhang, C. 2019. Mining QTLs and candidate genes for seed protein and oil contents across multiple environments and backgrounds in soybean. Mol. Breed. **39**. Molecular Breeding. doi:10.1007/s11032-019-1055-7.

Liao, P., Woodfield, H.K., Harwood, J.L., Chye, M.L., and Scofield, S. 2019. Comparative Transcriptomics Analysis of *Brassica napus* L. During Seed Maturation Reveals Dynamic Changes in Gene Expression between Embryos and Seed Coats and Distinct Expression Profiles of Acyl-CoA-Binding Proteins for Lipid Accumulation. Plant Cell Physiol. **60**: 2812–2825. doi:10.1093/pcp/pcz169.

Liu, H.J., and Yan, J. 2019. Crop genome-wide association study: a harvest of biological relevance. Plant J. **97**: 8–18. doi:10.1111/tpj.14139.

Liu, J., Wang, J., Wang, H., Wang, W., Zhou, R., Mei, D., Cheng, H., Yang, J., Raman, H., and Hu, Q. 2016a. Multigenic control of pod shattering resistance in Chinese rapeseed germplasm revealed by genome-wide association and linkage analyses. Front. Plant Sci. **7**: 1–14. doi:10.3389/fpls.2016.01058.

Liu, J., Wang, W., Mei, D., Wang, H., Fu, L., Liu, D., Li, Y., and Hu, Q. 2016b. Characterizing variation of branch angle and genome-wide association mapping in rapeseed (Brassica napus L.). Front. Plant Sci. **7**: 1–10. doi:10.3389/fpls.2016.00021.

Liu, S., Fan, C., Li, J., Cai, G., Yang, Q., Wu, J., Yi, X., Zhang, C., and Zhou, Y. 2016c. A genome-wide association study reveals novel elite allelic variations in seed oil content of Brassica napus. Theor. Appl. Genet. **129**: 1203–1215. Springer Berlin Heidelberg. doi:10.1007/s00122-016-2697-z.

Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C., and Zhou, Y. 2020. Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of Brassica napus by genome-wide association study. Plant Biotechnol. J. **18**: 1472–1484. doi:10.1111/pbi.13314.

Liu, S., Ohta, K., Dong, C., Thanh, V.C., Ishimoto, M., Qin, Z., and Hirata, Y. 2006. Genetic diversity of soybean (*Glycine max* (L.) Merrill) 7S globulin protein subunits. Genet. Resour. Crop Evol. **53**: 1209–1219. doi:10.1007/s10722-005-2434-y.

Liu, Y., Beyer, A., and Aebersold, R. 2016d. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell **165**: 535–550. Elsevier Inc. doi:10.1016/j.cell.2016.03.014.

Livak, K.J., and Schmittgen, T.D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods **25**: 402–408. doi:10.1006/meth.2001.1262.

LMC International 2020. The Economic Impact of Canola on the Canadian Economy : 2020 Update. New York. [Online] Available: https://www.canolacouncil.org/download/131/economic-impact/17818/economic-impact-report-canada_december-2020.

Lo, M., and Bell, J.M. 1972. Effects of various dietary glucosinolate on growth, feed intake, and thyroid function of rats. Can. J. Anim. Sci. **52**: 295–302.

Lomascolo, A., Uzan-Boukhris, E., Sigoillot, J.C., and Fine, F. 2012. Rapeseed and sunflower

meal: A review on biotechnology status and challenges. Appl. Microbiol. Biotechnol. **95**: 1105–1114. doi:10.1007/s00253-012-4250-6.

Lönnerdal, B., and Janson, J.-C. 1972. Studies on Brassica seed proteins. Biochim. Biophys. Acta - Protein Struct. **278**: 175–183. doi:10.1016/0005-2795(72)90119-5.

Lord, E.M., and Russell, S.D. 2002. The mechanisms of pollination and fertilization in plants. Annu. Rev. Cell Dev. Biol. **18**: 81–105. doi:10.1146/annurev.cellbio.18.012502.083438.

Lorieux, M. 2012. MapDisto: Fast and efficient computation of genetic linkage maps. Mol. Breed. **30**: 1231–1235. doi:10.1007/s11032-012-9706-y.

Louwaars, N.P. 2018. Plant breeding and diversity: A troubled relationship? Euphytica **214**: 1–9. Springer Netherlands. doi:10.1007/s10681-018-2192-5.

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., Liang, Y., Xu, X., Qu, C., Zhang, K., Liu, L., Zhu, Q., Fu, M., Yuan, X., and Li, J. 2017. Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in brassica napus. Front. Plant Sci. **8**: 1–15. doi:10.3389/fpls.2017.00206.

Lu, K., Xiao, Z., Jian, H., Peng, L., Qu, C., Fu, M., He, B., Tie, L., Liang, Y., Xu, X., and Li, J. 2016. A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in Brassica napus. Sci. Rep. **6**: 1–13. Nature Publishing Group. doi:10.1038/srep36452.

Lund, M.N., and Ray, C.A. 2017. Control of Maillard reactions in foods: strategies and chemical mechanisms. J. Agric. Food Chem. **65**: 4537–4552. doi:10.1021/acs.jafc.7b00882.

Luo, X., Ma, C., Yue, Y., Hu, K., Li, Y., Duan, Z., Wu, M., Tu, J., Shen, J., Yi, B., and Fu, T. 2015. Unravelling the complex trait of harvest index in rapeseed (Brassica napus L.) with association mapping. BMC Genomics **16**: 1–10. ??? doi:10.1186/s12864-015-1607-0.

Luo, X., Xue, Z., Ma, C., Hu, K., Zeng, Z., Dou, S., Tu, J., Shen, J., Yi, B., and Fu, T. 2017. Joint genome-wide association and transcriptome sequencing reveals a complex polygenic network underlying hypocotyl elongation in rapeseed (Brassica napus L.). Sci. Rep. **7**: 1–12. Nature Publishing Group. doi:10.1038/srep41561.

Luo, Z., Li, X.D., Wang, W.M., Tan, X.Y., and Liu, X. 2011. Partial replacement of fish meal by a mixture of soybean meal and rapeseed meal in practical diets for juvenile Chinese mitten crab *Eriocheir sinensis*: Effects on growth performance and in vivo digestibility. Aquac. Res. **42**: 1615–1622. doi:10.1111/j.1365-2109.2010.02751.x.

Mag, T.K. 1983. Canola oil processing in Canada. J. Am. Oil Chem. Soc. **60**: 380–384. doi:10.1007/BF02543522.

Malabat, C., Atterby, H., Chaudhry, Q., Renard, M., Guéguen, J., Recherche, U. De, Végétales, P., Kingdom, U., and Rheu, L. 2003. Genetic variability of rapeseed protein composition . Proc. 11th Int. Rapeseed Congr.: 205–208.

Malabat, C., nchez-Vioque, R.I.S., Rabiller, C., and Gu guen, J. 2001. Emulsifying and foaming properties of native and chemically modified peptides from the 2S and 12S proteins of rapeseed (Brassica napus L.). J. Am. Oil Chem. Soc. **78**: 235–242. doi:10.1007/s11746-001-0251-x.

Manan, S., Chen, B., She, G., Wan, X., and Zhao, J. 2017. Transport and transcriptional regulation of oil production in plants. Crit. Rev. Biotechnol. **37**: 641–655. doi:10.1080/07388551.2016.1212185.

Marambe, H.K., McIntosh, T.C., Cheng, B., and Wanasundara, J.P.D. 2015. Structural stability and Sin a 1 anti-epitope antibody binding ability of yellow mustard (Sinapis alba L.) napin during industrial-scale myrosinase inactivation process. Food Funct. **6**: 2384–2395. Royal Society of Chemistry. doi:10.1039/C4FO00806E.

Marambe, H.K., McIntosh, T.C., Cheng, B., and Wanasundara, J.P.D.D. 2014. Quantification of major 2S allergen protein of yellow mustard using anti-Sin a 1 epitope antibody. Food Control **44**: 233–241. Elsevier Ltd. doi:10.1016/j.foodcont.2014.03.053.

Masilamani, M., Commins, S., and Shreffler, W. 2012. Determinants of food allergy. Immunol. Allergy Clin. North Am. **32**: 11–33. Elsevier Inc. doi:10.1016/j.iac.2011.12.003.

Mason, A.S., Higgins, E.E., Snowdon, R.J., Batley, J., Stein, A., Werner, C., and Parkin, I.A.P. 2017. A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array. Theor. Appl. Genet. **130**: 621–633. Springer Berlin Heidelberg. doi:10.1007/s00122-016-2849-1.

Mauricio, R. 2001. Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. Nat. Rev. Genet. **2**: 370–381. doi:10.1038/35072085.

McCarthy, D.J., Chen, Y., and Smyth, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. **40**: 4288–4297. doi:10.1093/nar/gks042.

McCurdy, S.M. 1990. Effects of processing on the functional properties of canola/rapeseed protein. J. Am. Oil Chem. Soc. **67**: 281–284. doi:10.1007/BF02539677.

Measures, M., Weinberger, P., and Baer, H. 1973. Variability of plant growth within controlled-environment chambers as related to temperature and light distribution. Can. J. Plant Sci. **53**: 215–220. doi:10.4141/cjps73-038.

Mejia, L.A., Korgaonkar, C.K., Schweizer, M., Chengelis, C., Marit, G., Ziemer, E., Grabiel, R., and Empie, M. 2009a. A 13-week sub-chronic dietary toxicity study of a cruciferin-rich canola protein isolate in rats. Food Chem. Toxicol. **47**: 2645–2654. Elsevier Ltd. doi:10.1016/j.fct.2009.07.029.

Mejia, L.A., Korgaonkar, C.K., Schweizer, M., Chengelis, C., Novilla, M., Ziemer, E., Williamson-Hughes, P.S., Grabiel, R., and Empie, M. 2009b. A 13-week dietary toxicity study in rats of a napin-rich canola protein isolate. Regul. Toxicol. Pharmacol. **55**: 394–402. Elsevier Inc. doi:10.1016/j.yrtph.2009.09.008.

Mejicanos, G., Sanjayan, N., Kim, I.H., and Nyachoti, C.M. 2016. Recent advances in canola meal utilization in swine nutrition. J. Anim. Sci. Technol. **58**: 1–13. Journal of Animal Science and Technology. doi:10.1186/s40781-016-0085-5.

Melina, V., Craig, W., and Levin, S. 2016. Position of the Academy of Nutrition and Dietetics: Vegetarian Diets. J. Acad. Nutr. Diet. **116**: 1970–1980. doi:10.1016/j.jand.2016.09.025.

Members of the Complex Trait Consortium 2003. The nature and identification of quantitative trait loci: a community's view. Nat. Rev. Genet. **4**: 911–916. doi:10.1038/nrg1206.

Meng, L., Li, H., Zhang, L., and Wang, J. 2015. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J. **3**: 269–283. Crop Science Society of China and Institute of Crop Science, CAAS. doi:10.1016/j.cj.2015.01.001.

Metje-Sprink, J., Menz, J., Modrzejewski, D., and Sprink, T. 2019. DNA-Free genome editing: Past, present and future. Front. Plant Sci. **9**: 1–9. doi:10.3389/fpls.2018.01957.

Millward, D.J. 2012. Amino acid scoring patterns for protein quality assessment. Br. J. Nutr. **108**. doi:10.1017/S0007114512002462.

Monsalve, R.I., Gonzalez de la Pena, M.A., Lopez-Otiin, C., Fiandor, A., Fernandez, C., Villalba, M., and Rodriguez, R. 1997. Detection, isolation and complete amino acid sequence of an aeroallergenic protein from rapeseed flour. Clin Exp Allergy **27**: 833–841. doi:10.1046/j.1365-2222.1997.660843.x.

Moreno, F.J., and Clemente, A. 2008. 2S Albumin Storage Proteins: What Makes them Food Allergens? Open Biochem. J. **2**: 16–28. doi:10.2174/1874091x00802010016.

Moritz, C.P. 2020. 40 years Western blotting: A scientific birthday toast. J. Proteomics **212**. Elsevier. doi:10.1016/j.jprot.2019.103575.

Mosenthin, R., Messerschmidt, U., Sauer, N., Carré, P., Quinsac, A., and Schöne, F. 2016. Effect of the desolventizing/toasting process on chemical composition and protein quality of rapeseed meal. J. Anim. Sci. Biotechnol. **7**: 1–12. Journal of Animal Science and Biotechnology. doi:10.1186/s40104-016-0095-7.

Mottet, A., de Haan, C., Falcucci, A., Tempio, G., Opio, C., and Gerber, P. 2017. Livestock: On our plates or eating at our table? A new analysis of the feed/food debate. Glob. Food Sec. **14**: 1–8. Elsevier B.V. doi:10.1016/j.gfs.2017.01.001.

Moyers, B.T., Morrell, P.L., and McKay, J.K. 2018. Genetic costs of domestication and improvement. J. Hered. **109**: 103–116. doi:10.1093/jhered/esx069.

Müntz, K. 1998. Deposition of storage proteins. Plant Mol. Biol. **38**: 77–99. doi:10.1007/978-94-011-5298-3_4.

Müntz, K. 2007. Protein dynamics and proteolysis in plant vacuoles. J. Exp. Bot. **58**: 2391–2407.

doi:10.1093/jxb/erm089.

Mupondwa, E., Li, X., and Wanasundara, J.P.D. 2018. Technoeconomic Prospects for Commercialization of Brassica (Cruciferous) Plant Proteins. JAOCS, J. Am. Oil Chem. Soc. **95**: 903–922. doi:10.1002/aocs.12057.

Murovec, J., Guček, K., Bohanec, B., Avbelj, M., and Jerala, R. 2018. DNA-free genome editing of *Brassica oleracea* and *B. rapa* protoplasts using CRISPR-Cas9 ribonucleoprotein complexes. Front. Plant Sci. **9**: 1–9. doi:10.3389/fpls.2018.01594.

Murphy, D.J., and Cummins, I. 1989. Biosynthesis of Seed Storage Products during Embryogenesis in Rapeseed, Brassica napus. J. Plant Physiol. **135**: 63–69. doi:10.1016/S0176-1617(89)80225-1.

Murphy, D.J., Cummins, I., and Kang, A.S. 1989. Synthesis of the major oil-body membrane protein in developing rapeseed (*Brassica napus*) embryos. Integration with storage-lipid and storage-protein synthesis and implications for the mechanism of oil-body formation. Biochem. J. **258**: 285–293. doi:10.1042/bj2580285.

Myers, S.S., Smith, M.R., Guth, S., Golden, C.D., Vaitla, B., Mueller, N.D., Dangour, A.D., and Huybers, P. 2017. Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition. Annu. Rev. Public Health **38**: 259–277. doi:10.1146/annurev-publhealth-031816-044356.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., and Buckler, E. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell **21**: 2194–2202. doi:10.1105/tpc.109.068437.

Mylne, J.S., Hara-Nishimura, I., and Rosengren, K.J. 2014. Seed storage albumins: Biosynthesis, trafficking and structures. Funct. Plant Biol. **41**: 671–677. doi:10.1071/FP14035.

Nagel, F., Slawski, H., Adem, H., Tressel, R.P., Wysujack, K., and Schulz, C. 2012. Albumin and globulin rapeseed protein fractions as fish meal alternative in diets fed to rainbow trout (Oncorhynchus mykiss W.). Aquaculture **354–355**: 121–127. Elsevier B.V. doi:10.1016/j.aquaculture.2012.03.024.

Nelson, G., Bogard, J., Lividini, K., Arsenault, J., Riley, M., Sulser, T.B., Mason-D'Croz, D., Power, B., Gustafson, D., Herrero, M., Wiebe, K., Cooper, K., Remans, R., and Rosegrant, M. 2018. Income growth and climate change effects on global nutrition security to mid-century. Nat. Sustain. **1**: 773–781. Springer US. doi:10.1038/s41893-018-0192-z.

Newkirk, R.W., and Classen, H.L. 2002. The effects of toasting canola meal on body weight, feed conversion efficiency, and mortality in broiler chickens. Poult. Sci. **81**: 815–825. doi:10.1093/ps/81.6.815.

Nieschlag, H.J., and Wolff, I.A. 1971. Industrial uses of high erucic oils. J. Am. Oil Chem. Soc. **48**: 723–727. doi:10.1007/BF02638529.

Nietzel, T., Dudkina, N. V., Haase, C., Denolf, P., Semchonok, D.A., Boekema, E.J., Braun, H.P., and Sunderhaus, S. 2013. The native structure and composition of the cruciferin complex in *Brassica napus*. J. Biol. Chem. **288**: 2238–2245. doi:10.1074/jbc.M112.356089.

Nishinari, K., Fang, Y., Guo, S., and Phillips, G.O. 2014. Soy proteins: A review on composition, aggregation and emulsification. Food Hydrocoll. **39**: 301–318. Elsevier Ltd. doi:10.1016/j.foodhyd.2014.01.013.

Nishinari, K., Fang, Y., Nagano, T., Guo, S., and Wang, R. 2018. Soy as a food ingredient. *Edited By*R.Y. Yada Proteins in Food Processing: Second Edition, 2nd edition. Elsevier Science and Technology, Sawstons. doi:10.1016/B978-0-08-100722-8.00007-3.

Niu, Y., Wu, G.Z., Ye, R., Lin, W.H., Shi, Q.M., Xue, L.J., Xu, X.D., Li, Y., Du, Y.G., and Xue, H.W. 2009. Global analysis of gene expression profiles in *Brassica napus* developing seeds reveals a conserved lipid metabolism regulation with *Arabidopsis thaliana*. Mol. Plant **2**: 1107–1122. The Authors. All rights reserved. doi:10.1093/mp/ssp042.

Norton, G., and Harris, J.F. 1975. Compositional changes in developing rape seed (*Brassica napus* L.). Planta **123**: 163–174. doi:10.1007/BF00383865.

Ohlson, R. 1992. Modern processing of rapeseed. J. Am. Oil Chem. Soc. **69**: 195–198. doi:10.1007/BF02635885.

Ohto, M. aki, Floyd, S.K., Fischer, R.L., Goldberg, R.B., and Harada, J.J. 2009. Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. Sex. Plant Reprod. **22**: 277–289. doi:10.1007/s00497-009-0116-1.

Okuzaki, A., Ogawa, T., Koizuka, C., Kaneko, K., Inaba, M., Imamura, J., and Koizuka, N. 2018. CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. Plant Physiol. Biochem. **131**: 63–69. Elsevier. doi:10.1016/j.plaphy.2018.04.025.

Osborne, T.B., and Osbornb, T.B. 1924. The vegetable proteins. *Edited By*R.H.A. Plimmer and F.G. Hopkins 2nd edition. Longmans, Green and Co., London. doi:10.1002/jctb.5000431704.

Paek, N.C., Sexton, P.J., Naeve, S.L., and Shibles, R. 2000. Differential accumulation of soybean seed storage protein subunits in response to sulfur and nitrogen nutritional sources. Plant Prod. Sci. **3**: 268–274. doi:10.1626/pps.3.268.

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., and Giraudat, J. 1994. Regulation of gene expression programs during Arabidopsis seed development: Roles of the *ABI3* locus and of endogenous abscisic acid. Plant Cell **6**: 1567–1582. doi:10.1105/tpc.6.11.1567.

Patil, G., Vuong, T.D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., Wu, X., Bai, Y., Yungbluth, D., Lu, F., Kumpatla, S., Shannon, J.G., Varshney, R.K., and Nguyen, H.T. 2018. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an

interspecific mapping population of soybean using high-density linkage mapping. Plant Biotechnol. J. **16**: 1939–1953. doi:10.1111/pbi.12929.

Paul, R.H., Frestedt, J., and Magurany, K. 2017. GRAS from the ground up: Review of the Interim Pilot Program for GRAS notification. Food Chem. Toxicol. **105**: 140–150. Elsevier Ltd. doi:10.1016/j.fct.2017.03.064.

Pelletier, J.M., Kwong, R.W., Park, S., Le, B.H., Baden, R., Cagliari, A., Hashimoto, M., Munoz, M.D., Fischer, R.L., Goldberg, R.B., and Harada, J.J. 2017. LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. Proc. Natl. Acad. Sci. U. S. A. **114**: E6710–E6719. doi:10.1073/pnas.1707957114.

Peng, Y., Liu, H., Chen, J., Shi, T., Zhang, C., Sun, D., He, Z., Hao, Y., and Chen, W. 2018. Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. Front. Plant Sci. **9**: 1–9. doi:10.3389/fpls.2018.01196.

Perera, S.P., McIntosh, T.C., and Wanasundara, J.P.D. 2016. Structural properties of cruciferin and napin of *Brassica napus* (canola) show distinct responses to changes in pH and temperature. Plants **5**: 36. doi:10.3390/plants5030036.

Perignon, M., Vieux, F., Soler, L.-G.G., Masset, G., and Darmon, N. 2017. Improving diet sustainability through evolution of food choices: Review of epidemiological studies on the environmental impact of diets. Nutr. Rev. **75**: 2–17. doi:10.1093/nutrit/nuw043.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. **11**: 1650–1667. Nature Publishing Group. doi:10.1038/nprot.2016.095.

Pilorgé, E., Flénet, F., Quinsac, A., and Pinochet, X. 2019. From one rapeseed congress to another: what research for which issues, A review of the Saskatoon 2015 conference in the perspective of Berlin 2019. Ocl **26**: 30. doi:10.1051/ocl/2019026.

Pimentel, D., and Pimentel, M. 2003. Sustainability of meat-based and plant-based diets and the environment. Am. J. Clin. Nutr. **78**: 660–663.

Pistón, F., Gil-Humanes, J., Rodríguez-Quijano, M., and Barro, F. 2011. Down-Regulating γ-Gliadins in bread wheat leads to Non-Specific increases in other gluten proteins and has no major effect on dough gluten strength. PLoS One **6**: 1–10. doi:10.1371/journal.pone.0024754.

Poisson, E., Trouverie, J., Brunel-Muguet, S., Akmouche, Y., Pontet, C., Pinochet, X., and Avice, J.C. 2019. Seed yield components and seed quality of oilseed rape are impacted by sulfur fertilization and its interactions with nitrogen fertilization. Front. Plant Sci. **10**: 1–14. doi:10.3389/fpls.2019.00458.

Poore, J., and Nemecek, T. 2018. Reducing food's environmental impacts through producers and consumers. Science (80-.). **360**: 987–992. doi:10.1126/science.aaq0216.

Porebski, S., Bailey, L.G., Baum, B.R., and Bum, B.R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rep. **15**: 8–15. doi:10.1210/en.2002-220569.

Porter, A.S., Evans-Fitz.Gerald, C., McElwain, J.C., Yiotis, C., and Elliott-Kingston, C. 2015. How well do you know your growth chambers? Testing for chamber effect using plant traits. Plant Methods **11**: 1–10. BioMed Central. doi:10.1186/s13007-015-0088-0.

Poysa, V., Woodrow, L., and Yu, K. 2006. Effect of soy protein subunit composition on tofu quality. Food Res. Int. **39**: 309–317. doi:10.1016/j.foodres.2005.08.003.

Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. **11**: 459–463. Nature Publishing Group. doi:10.1038/nrg2813.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**: 945–59. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/10835412.

Proudfoot, F.C., Hulan, H.W., and McRae, K.B. 1983. Effect of feeding poultry diets supplemented with rapeseed meal as a primary protein source to juvenile and adult meat breeder genotypes. Can. J. Anim. Sci. **63**: 957–965. doi:10.4141/cjas83-110.

Pudel, F., Tressel, R.P., and Düring, K. 2015. Production and properties of rapeseed albumin. Lipid Technol. **27**: 112–114. doi:10.1002/lite.201500023.

Das Purkayastha, M., Das, S., Manhar, A.K., Deka, D., Mandal, M., and Mahanta, C.L. 2013. Removing antinutrients from rapeseed press-cake and their benevolent role in waste cooking oil-derived biodiesel: Conjoining the valorization of two disparate industrial wastes. J. Agric. Food Chem. **61**: 10746–10756. doi:10.1021/jf403657c.

Puumalainen, T.J., Puustinen, A., Poikonen, S., Turjanmaa, K., Palosuo, T., and Vaali, K. 2015. Proteomic identification of allergenic seed proteins, napin and cruciferin, from cold-pressed rapeseed oils. Food Chem. **175**: 381–385. Elsevier Ltd. doi:10.1016/j.foodchem.2014.11.084.

Qian, L., Qian, W., and Snowdon, R.J. 2014. Sub-genomic selection patterns as a signature of breeding in the allopolyploid Brassica napus genome. BMC Genomics **15**: 1–17. doi:10.1186/1471-2164-15-1170.

Qian, L., Qian, W., and Snowdon, R.J. 2016a. Haplotype hitchhiking promotes trait coselection in Brassica napus. Plant Biotechnol. J. **14**: 1578–1588. doi:10.1111/pbi.12521.

Qian, L., Voss-Fels, K., Cui, Y., Jan, H.U., Samans, B., Obermeier, C., Qian, W., and Snowdon, R.J. 2016b. Deletion of a Stay-Green Gene Associates with Adaptive Selection in Brassica napus. Mol. Plant **9**: 1559–1569. Elsevier Ltd. doi:10.1016/j.molp.2016.10.017.

Qiao, Z., Qi, W., Wang, Q., Feng, Y., Yang, Q., Zhang, N., Wang, S., Tang, Y., and Song, R.

2016. ZmMADS47 Regulates Zein Gene Transcription through Interaction with Opaque2. PLoS Genet. **12**: 1–26. doi:10.1371/journal.pgen.1005991.

Qu, C., Jia, L., Fu, F., Zhao, H., Lu, K., Wei, L., Xu, X., Liang, Y., Li, S., Wang, R., and 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**: 1–17. BMC Genomics. doi:10.1186/s12864-017-3607-8.

Qu, C.M., Li, S.M., Duan, X.J., Fan, J.H., Jia, L.D., Zhao, H.Y., Lu, K., Li, J.N., Xu, X.F., and Wang, R. 2015. Identification of candidate genes for seed glucosinolate content using association mapping in brassica napus L. Genes (Basel). **6**: 1215–1229. doi:10.3390/genes6041215.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Online] Available: https://www.r-project.org/.

Raab, B., Leman, H., Schwenke, K.D., and Kozlowska, H. 1992. Comparative study of the protein patterns of some rapeseed (Brassica napus L.) varieties by means of polyacrylamide gel electrophoresis and high-performance liquid chromatography. Food / Nahrung **36**: 239–247. doi:10.1002/food.19920360304.

Rahman, M., Hoque, A., and Roy, J. 2021. Linkage disequilibrium and population structure in a core collection of. bioRxiv. doi:10.1101/2021.04.06.438572.

Rahman, M., Sun, Z., McVetty, P.B.E., and 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. Theor. Appl. Genet. **117**: 895–904. doi:10.1007/s00122-008-0829-9.

Rakow, G. 2011. Classical genetics and traditional breeding. *Edited By*D. Edwards, J. Batley, I. Parkin, and C. Kole Genetics, Genomics and Breeding of Oilseed Brassicas. CRC Press, Boca Raton. doi:10.1201/b11406.

Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., and Rieseberg, L.H. 2018. Trends in Global Agricultural Land Use: Implications for Environmental Health and Food Security. Annu. Rev. Plant Biol. **69**: 789–815. doi:10.1146/annurev-arplant-042817-040256.

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R.K., and He, Z. 2017. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. Mol. Plant **10**: 1047–1064. Elsevier Ltd. doi:10.1016/j.molp.2017.06.008.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. 2019. g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. **47**: W191–W198. Oxford University Press. doi:10.1093/nar/gkz369.

Ravindran, V., Abdollahi, M.R., and Bootwalla, S.M. 2014. Nutrient analysis, metabolizable

energy, and digestible amino acids of soybean meals of different origins for broilers. Poult. Sci. **93**: 2567–2577. doi:10.3382/ps.2014-04068.

Risch, N., and Merikangas, K. 1996. The future of genetic studies of complex human diseases. Science (80-.). **273**: 1516–1517. doi:10.1097/00001648-199805000-00023.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. **43**: e47. doi:10.1093/nar/gkv007.

Rizzo, G., and Baroni, L. 2018. Soy, Soy Foods and Their Role in Vegetarian Diets. Nutrients **10**: 43. doi:10.3390/nu10010043.

Robin, J.M.M., Inquello, V., Mimouni, B., and Azanza, J.L. 1991. Relationship between immunological properties and structural model of 11S rapeseed globulin. Phytochemistry **30**: 3511–3513. doi:10.1016/0031-9422(91)80056-7.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140. doi:10.1093/bioinformatics/btp616.

Robinson, M.D., and Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. **11**: R25. doi:10.1186/gb-2010-11-3-r25.

Rödiger, A., and Baginsky, S. 2018. Tailored use of targeted proteomics in plant-specific applications. Front. Plant Sci. **9**: 1–4. doi:10.3389/fpls.2018.01204.

Rödin, J., and Rask, L. 1990. Characterization of the 12S storage protein of *Brassica napus* (cruciferin): Disulfide bonding between subunits. Physiol. Plant. **79**: 421–426. doi:10.1111/j.1399-3054.1990.tb02097.x.

Rolletschek, H., Schwender, J., König, C., Chapman, K.D., Romsdahl, T., Lorenz, C., Braun, H.P., Denolf, P., van Audenhove, K., Munz, E., Heinzel, N., Ortleb, S., Rutten, T., McCorkle, S., Borysyuk, T., Guendel, A., Shi, H., Auwermeulen, M. Vander, Bourot, S., and Borisjuk, L. 2020. Cellular plasticity in response to suppression of storage proteins in the *Brassica napus* embryo. Plant Cell **32**: 2383–2401. doi:10.1105/tpc.19.00879.

Rong, H., Yang, W., Zhu, H., Jiang, B., Jiang, J., and Wang, Y. 2021. Genomic imprinted genes in reciprocal hybrid endosperm of *Brassica napus*. BMC Plant Biol. **21**: 1–17. BMC Plant Biology. doi:10.1186/s12870-021-02908-8.

Röös, E., Bajželj, B., Smith, P., Patel, M., Little, D., and Garnett, T. 2017a. Protein futures for Western Europe: potential land use and climate impacts in 2050. Reg. Environ. Chang. **17**: 367–377. doi:10.1007/s10113-016-1013-4.

Röös, E., Bajželj, B., Smith, P., Patel, M., Little, D., and Garnett, T. 2017b. Greedy or needy? Land use and climate impacts of food in 2050 under different livestock futures. Glob.

Environ. Chang. **47**: 1–12. doi:10.1016/j.gloenvcha.2017.09.001.

RStudio Team 2020. RStudio: Integrated Development Environment for R. RStudio PBC, Boston, MA. [Online] Available: http://www.rstudio.com/.

Sabaté, J., and Soret, S. 2014. Sustainability of plant-based diets: Back to the future. Am. J. Clin. Nutr. **100**: 476–482. doi:10.3945/ajcn.113.071522.

Sabaté, J., Sranacharoenpong, K., Harwatt, H., Wien, M., and Soret, S. 2015. The environmental cost of protein food choices. Public Health Nutr. **18**: 2067–2073. doi:10.1017/S1368980014002377.

Sah, S.K., Reddy, K.R., and Li, J. 2016. Abscisic acid and abiotic stress tolerance in crop plants. Front. Plant Sci. **7**: 1–26. doi:10.3389/fpls.2016.00571.

Saio, K., Kamiya, M., Watanabe, T., Koyama, E., Yamazaki, S., Chem, A.B., Wolf, W.J., Babcock, T.E., Smith, A.K., Teramachi, Y., Abe, K., Suruga, S., Miyanaga, S., Kamiya, M., and Watanabe, T. 1969. Food processing characteristics of soybean 11s and 7s proteins: Part I. Effect of difference of protein components among soybean varieties on formation of tofu-gel. Agric. Biol. Chem. **33**: 1301–1308. doi:10.1271/bbb1961.33.1301.

Salazar-Villanea, S., Bruininx, E.M.A.M., Gruppen, H., Carré, P., Quinsac, A., and van der Poel, A.F.B. 2017. Effects of toasting time on digestive hydrolysis of soluble and insoluble 00-rapeseed meal proteins. JAOCS, J. Am. Oil Chem. Soc. **94**: 619–630. Springer Berlin Heidelberg. doi:10.1007/s11746-017-2960-8.

Salazar-Villanea, S., Bruininx, E.M.A.M., Gruppen, H., Hendriks, W.H., Carré, P., Quinsac, A., and van der Poel, A.F.B. 2016. Physical and chemical changes of rapeseed meal proteins during toasting and their effects on in vitro digestibility. J. Anim. Sci. Biotechnol. **7**: 1–11. Journal of Animal Science and Biotechnology. doi:10.1186/s40104-016-0120-x.

Salleh, M.R.B.M., Maruyama, N., Adachi, M., Hontani, N., Saka, S., Kato, N., Ohkawa, Y., and Utsumi, S. 2002. Comparison of Protein Chemical and Physicochemical Properties of Rapeseed Cruciferin with Those of Soybean Glycinin. J. Agric. Food Chem. **50**: 7380–7385. doi:10.1021/jf0202537.

Sashidhar, N., Harloff, H., and Jung, C. 2019. Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large scale screening of mutant populations through amplicon sequencing. New Phytol.: nph.16281. doi:10.1111/nph.16281.

Sashidhar, N., Harloff, H.J., and Jung, C. 2020. Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large-scale screening of mutant populations through amplicon sequencing. New Phytol. **225**: 2022–2034. doi:10.1111/nph.16281.

Scarth, R., McVetty, P.B.E., Rimmer, S.R., and Stefansson, B.R. 1988. Stellar low linolenic-high linoleic acid summer rape. Can. J. Plant Sci. **68**: 509–511. doi:10.4141/cjps88-061.

Scarth, R., and Tang, J. 2006. Modification of *Brassica* oil using conventional and transgenic

approaches. Crop Sci. **46**: 1225–1236. doi:10.2135/cropsci2005.08-0245.

Schaafsma, G. 2000. The protein digestibility–corrected amino acid score. J. Nutr. **130**: 1865S-1867S. doi:10.1093/jn/130.7.1865S.

Schatzki, J., Ecke, W., Becker, H.C., and Möllers, C. 2014. Mapping of QTL for the seed storage proteins cruciferin and napin in a winter oilseed rape doubled haploid population and their inheritance in relation to other seed traits. Theor. Appl. Genet. **127**: 1213–1222. doi:10.1007/s00122-014-2292-0.

Scheben, A., Batley, J., and Edwards, D. 2017. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol. J. **15**: 149–161. doi:10.1111/pbi.12645.

Schiessl, S., Iniguez-Luy, F., Qian, W., and Snowdon, R.J. 2015. Diverse regulatory factors associate with flowering time and yield responses in winter-type Brassica napus. BMC Genomics **16**: 1–20. BMC Genomics. doi:10.1186/s12864-015-1950-1.

Schmidt, I., Renard, D., Rondeau, D., Richomme, P., Popineau, Y., and Axelos, M.A.V. 2004. Detailed physicochemical characterization of the 2S storage protein from rape (*Brassica napus* L.). J. Agric. Food Chem. **52**: 5995–6001. doi:10.1021/jf0307954.

Schmidt, M.A., Barbazuk, W.B., Sandford, M., May, G., Song, Z., Zhou, W., Nikolau, B.J., and Herman, E.M. 2011. Silencing of soybean storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. Plant Physiol. **156**: 330–345. doi:10.1104/pp.111.173807.

Schmidt, M.A., and Herman, E.M. 2008. Proteome rebalancing in soybean seeds can be exploited to enhance foreign protein accumulation. Plant Biotechnol. J. **6**: 832–842. doi:10.1111/j.1467-7652.2008.00364.x.

Schwartz, J.M., Solé, V., Guéguen, J., Ropers, M.H., Riaublanc, A., and Anton, M. 2015. Partial replacement of β-casein by napin, a rapeseed protein, as ingredient for processed foods: Thermoreversible aggregation. LWT - Food Sci. Technol. **63**: 562–568. doi:10.1016/j.lwt.2015.03.084.

Schwender, J., and Hay, J.O. 2012. Predictive modeling of biomass component tradeoffs in Brassica napus developing oilseeds based on in silico manipulation of storage metabolism. Plant Physiol. **160**: 1218–1236. doi:10.1104/pp.112.203927.

Scofield, S.R., and Crouch, M.L. 1987. Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. J. Biol. Chem. **262**: 12202–12208. doi:10.1016/s0021-9258(18)45337-9.

Scott, M.F., Ladejobi, O., Amer, S., Bentley, A.R., Biernaskie, J., Boden, S.A., Clark, M., Dell'Acqua, M., Dixon, L.E., Filippi, C. V., Fradgley, N., Gardner, K.A., Mackay, I.J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R.K., Thudi, M., Varshney, R.K., Venturini, L., Whan, A., Cockram, J., and Mott, R. 2020. Multi-parent populations in

crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity (Edinb). **125**: 396–416. Springer US. doi:10.1038/s41437-020-0336-6.

Sela, A., Piskurewicz, U., Megies, C., Mène-Saffrané, L., Finazzi, G., and Lopez-Molina, L. 2020. Embryonic photosynthesis affects post-germination plant growth. Plant Physiol. **182**: 2166–2181. doi:10.1104/PP.20.00043.

Semba, R.D. 2016. The rise and fall of protein malnutrition in global health. Ann. Nutr. Metab. **69**: 79–88. doi:10.1159/000449175.

Shahid, M., Cai, G., Zu, F., Zhao, Q., Qasim, M.U., Hong, Y., Fan, C., and Zhou, Y. 2019. Comparative transcriptome analysis of developing seeds and silique wall reveals dynamic transcription networks for effective oil production in *Brassica napus* L. Int. J. Mol. Sci. **20**: 1–23. doi:10.3390/ijms20081982.

Shepon, A., Eshel, G., Noor, E., and Milo, R. 2016. Energy and protein feed-to-food conversion efficiencies in the US and potential food security gains from dietary changes. Environ. Res. Lett. **11**: 105002. IOP Publishing. doi:10.1088/1748-9326/11/10/105002.

Shepon, A., Eshel, G., Noor, E., and Milo, R. 2018. The opportunity cost of animal based diets exceeds all food losses. Proc. Natl. Acad. Sci. U. S. A. **115**: 3804–3809. doi:10.1073/pnas.1713820115.

Shewry, P.R., and Halford, N.G. 2002. Cereal seed storage proteins: Structures, properties and role in grain utilization. J. Exp. Bot. **53**: 947–958. doi:10.1093/jexbot/53.370.947.

Shewry, P.R., Napier, J.A.J.A.J.A., and Tatham, A.S. 1995. Seed storage proteins: Structures and biosynthesis. Plant Cell Online **7**: 945–956. doi:10.1105/tpc.7.7.945.

Shu, K., Zhou, W., and Yang, W. 2018. APETALA 2-domain-containing transcription factors: focusing on abscisic acid and gibberellins antagonism. New Phytol. **217**: 977–983. doi:10.1111/nph.14880.

Sicherer, S.H., and Sampson, H.A. 2018. Food allergy: A review and update on epidemiology, pathogenesis, diagnosis, prevention, and management. J. Allergy Clin. Immunol. **141**: 41–58. Elsevier Inc. doi:10.1016/j.jaci.2017.11.003.

Le Signor, C., Aimé, D., Bordat, A., Belghazi, M., Labas, V., Gouzy, J., Young, N.D., Prosperi, J.M., Leprince, O., Thompson, R.D., Buitink, J., Burstin, J., and Gallardo, K. 2017. Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. New Phytol. **214**: 1597–1613. doi:10.1111/nph.14500.

Sikdar, M.S.I., Bowra, S., Schmidt, D., Dionisio, G., Holm, P.B., and Vincze, E. 2016. Targeted modification of storage protein content resulting in improved amino acid composition of barley grain. Transgenic Res. **25**: 19–31. Springer International Publishing. doi:10.1007/s11248-015-9911-7.

Sim, S.C., Robbins, M.D., Van Deynze, A., Michel, A.P., and Francis, D.M. 2011. Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). Heredity (Edinb). **106**: 927–935. Nature Publishing Group. doi:10.1038/hdy.2010.139.

Singh, P., Kumar, R., Sabapathy, S.N., and Bawa, A.S. 2008. Functional and edible uses of soy protein products. Compr. Rev. Food Sci. Food Saf. **7**: 14–28. [Online] Available: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-4337.2007.00025.x.

Sjödahl, S., Gustavsson, H.O., Rödin, J., Lenman, M., Höglund, A.S., and Rask, L. 1993. Cruciferin gene families are expressed coordinately but with tissue-specific differences during *Brassica napus* seed development. Plant Mol. Biol. **23**: 1165–1176. doi:10.1007/BF00042350.

Sjödahl, S., Gustavsson, H.O., Rödin, J., and Rask, L. 1995. Deletion analysis of the Brassica napus cruciferin gene *cru 1* promoter in transformed tobacco: promoter activity during early and late stages of embryogenesis is influenced by *cis*-acting elements in partially separate regions. Planta **197**: 264–271. doi:10.1007/BF00202646.

Sjödahl, S., RÖDIN, J., and Rask, L. 1991. Characterization of the 12S globulin complex of *Brassica napus* Evolutionary relationship to other 11–12S storage globulins. Eur. J. Biochem. **196**: 617–621. doi:10.1111/j.1432-1033.1991.tb15857.x.

Slawski, H., Adem, H., Tressel, R.P., Wysujack, K., Koops, U., Kotzamanis, Y., Wuertz, S., and Schulz, C. 2012. Total fish meal replacement with rapeseed protein concentrate in diets fed to rainbow trout (*Oncorhynchus mykiss* Walbaum). Aquac. Int. **20**: 443–453. doi:10.1007/s10499-011-9476-2.

Smaczniak, C., Immink, R.G.H., Angenent, G.C., and Kaufmann, K. 2012. Developmental and evolutionary diversity of plant MADS-domain factors: Insights from recent studies. Dev. **139**: 3081–3098. doi:10.1242/dev.074674.

Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. **3**. doi:10.2202/1544-6115.1027.

Soneson, C., Love, M.I., and Robinson, M.D. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research **4**: 1521. doi:10.12688/f1000research.7563.2.

Song, J., Liu, D., Xie, W., Yang, Z., Guo, L., Liu, K., Yang, Q., and Chen, L. 2021. BnPIR: Brassica napus pan-genome information resource for 1689 accessions. Plant Biotechnol. J. **19**: 412–414. doi:10.1111/pbi.13491.

van der Spiegel, M., Noordam, M.Y., and van der Fels-Klerx, H.J. 2013. Safety of novel protein sources (insects, microalgae, seaweed, duckweed, and rapeseed) and legislative aspects for their application in food and feed production. Compr. Rev. Food Sci. Food Saf. **12**: 662–678. doi:10.1111/1541-4337.12032.

Springmann, M., Clark, M., Mason-D'Croz, D., Wiebe, K., Bodirsky, B.L., Lassaletta, L., de Vries, W., Vermeulen, S.J., Herrero, M., Carlson, K.M., Jonell, M., Troell, M., DeClerck, F., Gordon, L.J., Zurayk, R., Scarborough, P., Rayner, M., Loken, B., Fanzo, J., Godfray, H.C.J., Tilman, D., Rockström, J., and Willett, W. 2018. Options for keeping the food system within environmental limits. Nature **562**: 519–525. Springer US. doi:10.1038/s41586-018-0594-0.

Stålberg, K., Ellerstöm, M., Ezcurra, I., Ablov, S., and Rask, L. 1996. Disruption of an overlapping E-box/ABRE motif abolished high transcription of the *napA* storage-protein promoter in transgenic *Brassica napus* seeds. Planta **199**: 515–519. doi:10.1007/bf00195181.

Stark, R., Grzelak, M., and Hadfield, J. 2019. RNA sequencing: the teenage years. Nat. Rev. Genet. **20**: 631–656. Springer US. doi:10.1038/s41576-019-0150-2.

Staswick, P.E., Huang, J.F., and Rhee, Y. 1991. Nitrogen and methyl jasmonate induction of soybean vegetative storage protein genes. Plant Physiol. **96**: 130–136. doi:10.1104/pp.96.1.130.

Statistics Canada 2021. Crushing statistics of major oilseeds. doi:https://doi.org/10.25318/3210035201-eng.

Steeves, T.A. 1983. The evolution and biological significance of seeds. Can. J. Bot. **61**: 3550–3560. doi:10.1139/b83-404.

Stefansson, B.R., Hougen, F.W., and Downey, R.K. 1961. Note on the isolation of rape plants with seed oil free from erucic acid. Can. J. Plant Sci. **41**: 218–219. doi:10.4141/cjps61-028.

Stefansson, B.R., and Kondra, Z.P. 1975. Tower summer rape. Can. J. Plant Sci. **344**: 343–344.

Steffen, J.G., Kang, I.H., Portereiko, M.F., Lloyd, A., and Drews, G.N. 2008. AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. Plant Physiol. **148**: 259–268. doi:10.1104/pp.108.119404.

Suárez, J.A., Gaxiola, G., Mendoza, R., Cadavid, S., Garcia, G., Alanis, G., Suárez, A., Faillace, J., and Cuzon, G. 2009. Substitution of fish meal with plant protein sources and energy budget for white shrimp *Litopenaeus vannamei* (Boone, 1931). Aquaculture **289**: 118–123. Elsevier B.V. doi:10.1016/j.aquaculture.2009.01.001.

Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., Li, S., Yan, L., Guan, C., Zhang, J., Zhang, Z., Chen, S., Wen, J., Tu, J., Shen, J., Fu, T., and Yi, B. 2016a. Genome-Wide Association Study Dissecting the Genetic Architecture Underlying the Branch Angle Trait in Rapeseed (Brassica napus L.). Sci. Rep. **6**: 1–11. Nature Publishing Group. doi:10.1038/srep33673.

Sun, F., Liu, J., Hua, W., Sun, X., Wang, X., and Wang, H. 2016b. Identification of stable QTLs for seed oil content by combined linkage and association mapping in Brassica napus. Plant Sci. **252**: 388–399. Elsevier Ireland Ltd. doi:10.1016/j.plantsci.2016.09.001.

Sun, Q., Lin, L., Liu, D., Wu, D., Fang, Y., Wu, J., and Wang, Y. 2018. CRISPR/Cas9-mediated multiplex genome editing of the *BnWRKY11* and *BnWRKY70* genes in *Brassica napus* L. Int. J. Mol. Sci. **19**: 1–19. doi:10.3390/ijms19092716.

Swaminathan, K., Peterson, K., and Jack, T. 2008. The plant B3 superfamily. Trends Plant Sci. **13**: 647–655. doi:10.1016/j.tplants.2008.09.006.

Swanepoel, N., Robinson, P.H., and Erasmus, L.J. 2014. Determining the optimal ratio of canola meal and high protein dried distillers grain protein in diets of high producing Holstein dairy cows. Anim. Feed Sci. Technol. **189**: 41–53. Elsevier B.V. doi:10.1016/j.anifeedsci.2013.12.007.

Takahashi, M., Uematsu, Y., Kashiwaba, K., Yagasaki, K., Hajika, M., Matsunaga, R., Komatsu, K., and Ishimoto, M. 2003. Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. Planta **217**: 577–586. doi:10.1007/s00425-003-1026-3.

Tal, M., Silberstein, A., and Nusser, E. 1985. Why does Coomassie Brilliant Blue R interact differently with different proteins? A partial answer. J. Biol. Chem. **260**: 9976–9980.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. **20**: 467–484. Springer US. doi:10.1038/s41576-019-0127-1.

Tamburino, L., Bravo, G., Clough, Y., and Nicholas, K.A. 2020. From population to production: 50 years of scientific literature on how to feed the world. Glob. Food Sec. **24**. doi:10.1016/j.gfs.2019.100346.

Tan, H., Qi, X., Li, Y., Wang, X., Zhou, J., Liu, X., Shi, X., Ye, W., and Xiang, X. 2020. Light induces gene expression to enhance the synthesis of storage reserves in *Brassica napus* L. embryos. Plant Mol. Biol. **103**: 457–471. Springer Netherlands. doi:10.1007/s11103-020-01003-0.

Tan, S.H., Mailer, R.J., Blanchard, C.L., and Agboola, S.O. 2011. Canola Proteins for Human Consumption: Extraction, Profile, and Functional Properties. J. Food Sci. **76**. doi:10.1111/j.1750-3841.2010.01930.x.

Tandang-Silvas, M.R.G., Fukuda, T., Fukuda, C., Prak, K., Cabanos, C., Kimura, A., Itoh, T., Mikami, B., Utsumi, S., and Maruyama, N. 2010. Conservation and divergence on plant seed 11S globulins based on crystal structures. Biochim. Biophys. Acta - Proteins Proteomics **1804**: 1432–1442. Elsevier B.V. doi:10.1016/j.bbapap.2010.02.016.

Tandang-Silvas, M.R.G., Tecson-Mendoza, E.M., Mikami, B., Utsumi, S., and Maruyama, N. 2011. Molecular design of seed storage proteins for enhanced food physiochemical properties. Annu. Rev. Food Sci. Technol. **2**: 59–73. doi:10.1146/annurev-food-022510-133718.

Tang, M., Zhang, Y., Liu, Y., Tong, C., Cheng, X., Zhu, W., Li, Z., Huang, J., and Liu, S. 2019.

Mapping loci controlling fatty acid profiles, oil and protein content by genome-wide association study in *Brassica napus*. Crop J. **7**: 217–226. Elsevier B.V. doi:10.1016/j.cj.2018.10.007.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Su, Z., Pan, Y., Liu, D., Lipka, A.E., Buckler, E.S., and Zhang, Z. 2016. GAPIT version 2: An enhanced integrated tool for genomic association and prediction. Plant Genome **9**: plantgenome2015.11.0120. doi:10.3835/plantgenome2015.11.0120.

Teh, L., and Möllers, C. 2016. Genetic variation and inheritance of phytosterol and oil content in a doubled haploid population derived from the winter oilseed rape Sansibar × Oase cross. Theor. Appl. Genet. **129**: 181–199. Springer Berlin Heidelberg. doi:10.1007/s00122-015-2621-y.

Thacker, P.A., and Newkirk, R.W. 2011. Performance of growing-finishing pigs fed barley-based diets containing toasted or non-toasted canola meal. Can. J. Anim. Sci. **85**: 53–59. doi:10.4141/a04-056.

Le Thanh, B. V., Beltranena, E., Zhou, X., Wang, L.F., and Zijlstra, R.T. 2019. Amino acid and energy digestibility of *Brassica napus* canola meal from different crushing plants fed to ileal-cannulated grower pigs. Anim. Feed Sci. Technol. **252**: 83–91. Elsevier. doi:10.1016/j.anifeedsci.2019.04.008.

The UniProt Consortium, Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., and Zhang, J. 2021. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res. **49**: D480–D489. Oxford University Press. doi:10.1093/nar/gkaa1100.

Thomke, S. 1981. Review of rapeseed meal in animal nutrition: Ruminant animals. J. Am. Oil Chem. Soc. **58**: 805–810. doi:10.1007/BF02665585.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E.S. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. **28**: 286–289. doi:10.1038/90135.

Tian, D., Wang, P., Tang, B., Teng, X., Li, C., Liu, X., Zou, D., Song, S., and Zhang, Z. 2020. GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. Nucleic Acids Res. **48**: D927–D932. Oxford University Press. doi:10.1093/nar/gkz828.

Tilton, G.B., Shockey, J.M., and Browse, J. 2004. Biochemical and Molecular Characterization of ACH2, an Acyl-CoA Thioesterase from Arabidopsis thaliana. J. Biol. Chem. **279**: 7487–7494. doi:10.1074/jbc.M309532200.

Toroser, D., Thormann, C.E., Osborn, T.C., and Mithen, R. 1995. RFLP mapping of quantitative trait loci controlling seed aliphatic-glucosinolate content in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. **91**: 802–808. doi:10.1007/BF00220963.

Towbin, H., Staehelin, T., and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. **76**: 4350–4354. doi:10.1073/pnas.76.9.4350.

Triboï, E., Martre, P., Triboï-Blondel, A.M., and Croue, S. De 2003. Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. J. Exp. Bot. **54**: 1731–1742. doi:10.1093/jxb/erg183.

Tripathi, M.K., and Mishra, A.S. 2007. Glucosinolates in animal nutrition: A review. Anim. Feed Sci. Technol. **132**: 1–27. doi:10.1016/j.anifeedsci.2006.03.003.

Tripathi, P., Carvallo, M., Hamilton, E.E., Preuss, S., and Kay, S.A. 2017. Arabidopsis B-BOX32 interacts with CONSTANS-LIKE3 to regulate flowering. Proc. Natl. Acad. Sci. U. S. A. **114**: 172–177. doi:10.1073/pnas.1616459114.

Troise, A.D., Wilkin, J.D., and Fiore, A. 2018. Impact of rapeseed press-cake on Maillard reaction in a cookie model system. Food Chem. **243**: 365–372. Elsevier. doi:10.1016/j.foodchem.2017.09.153.

Tuomisto, H.L. 2018. Importance of considering environmental sustainability in dietary guidelines. Lancet Planet. Heal. **2**: e331–e332. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license. doi:10.1016/S2542-5196(18)30174-8.

Turchini, G.M., Trushenski, J.T., and Glencross, B.D. 2019. Throughts for the future of aquaculture nutrition: REaligning perspectives to reflect contemporary issues related to judicious use of marine resources in aquafeeds. N. Am. J. Aquac. **81**: 13–39. doi:10.1002/naaq.10067.

Tykarska, T. 1976. Rape embryogenesis I. The proembryo development. Acta Soc. Bot. Pol. **45**: 3–16. doi:10.5586/asbp.1976.001.

Tykarska, T. 1979. Rape embryogenesis. II. Development of embryo proper. Acta Soc. Bot. Pol. **48**: 391–421. doi:10.5586/asbp.1979.033.

Tykarska, T. 1980. Rape embryogenesis. III. Embryo development in time. Acta Soc. Bot. Pol. **49**: 369–385. doi:10.5586/asbp.1980.033.

Udvardi, M.K., Czechowski, T., and Scheible, W.R. 2008. Eleven golden rules of quantitative RT-PCR. Plant Cell **20**: 1736–1737. doi:10.1105/tpc.108.061143.

Unger, E.H. 2011. Crushing and Extraction. Canola Chem. Prod. Process. Util.: 163–188. AOCS Press.

United Nations 1975. Report of the World Food Conference. New York.

United Nations Department of Economic and Social Affairs Population Division 2019. World Population Prospects 2019. World Population Prospects 2019: Highlights. New York. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/12283219.

Uzunova, M., Ecke, W., Weissleder, K., and Röbbelen, G. 1995. Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor. Appl. Genet. **90**: 194–204. doi:10.1007/BF00222202.

Verdier, J., and Thompson, R.D. 2008. Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. Plant Cell Physiol. **49**: 1263–1271. doi:10.1093/pcp/pcn116.

Vermorel, M., Davicco, M.-J.J., and Evrard, J. 1987. Valorization of rapeseed meal. 3. Effects of glucosinolate content on food intake, weight gain, liver weight and plasma thyroid hormone levels in growing rats. Reprod. Nutr. Dev. **27**: 57–66. doi:10.1051/rnd:19870106.

Vicient, C.M., Bies-Etheve, N., and Delseny, M. 2000. Changes in gene expression in the *leafy cotyledon1* (*lec1*) and *fusca3* (*fus3*) mutants of *Arabidopsis thaliana* L. J. Exp. Bot. **51**: 995–1003. doi:10.1093/jexbot/51.347.995.

Vieira, S.L., and Lima, I.L. 2005. Live performance, water intake and excreta characteristics of broilers fed all vegetable diets based on corn and soybean meal. Int. J. Poult. Sci. **4**: 365–368. doi:10.3923/ijps.2005.365.368.

Walling, L., Drews, G.N., and Goldberg, R.B. 1986. Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. Proc. Natl. Acad. Sci. **83**: 2123–2127. doi:10.1073/pnas.83.7.2123.

Waltz, E. 2015. A face-lift for biotech rules begins. Nat. Biotechnol. **33**: 1221–1222. Nature Publishing Group. doi:10.1038/nbt1215-1221.

Waltz, E. 2018. With a free pass, CRISPR-edited plants reach market in record time. Nat. Biotechnol. **36**: 6–7. doi:10.1038/nbt0118-6b.

Wan, H., Chen, L., Guo, J., Li, Q., Wen, J., Yi, B., Ma, C., Tu, J., Fu, T., and Shen, J. 2017a. Genome-wide association study reveals the genetic architecture underlying salt tolerance-related traits in rapeseed (Brassica napus L.). Front. Plant Sci. **8**: 1–15. doi:10.3389/fpls.2017.00593.

Wan, H., Cui, Y., Ding, Y., Mei, J., Dong, H., Zhang, W., Wu, S., Liang, Y., Zhang, C., Li, J., Xiong, Q., and Qian, W. 2017b. Time-series analyses of transcriptomes and proteomes reveal molecular networks underlying oil accumulation in canola. Front. Plant Sci. **7**: 1–17. doi:10.3389/fpls.2016.02007.

Wan, H., Wei, Y., Qian, J., Gao, Y., Wen, J., Yi, B., Ma, C., Tu, J., Fu, T., and Shen, J. 2018. Association mapping of salt tolerance traits at germination stage of rapeseed (Brassica napus L.). Euphytica **214**: 1–16. Springer Netherlands. doi:10.1007/s10681-018-2272-6.

Wan, L., Ross, A.R.S., Yang, J., Hegedus, D.D., and Kermode, A.R. 2007. Phosphorylation of the 12 S globulin cruciferin in wild-type and *abi1-1* mutant *Arabidopsis thaliana* (thale cress) seeds. Biochem. J. **404**: 247–256. doi:10.1042/BJ20061569.

Wanasundara, J.P.D. 2011. Proteins of brassicaceae oilseeds and their potential as a plant protein source. Crit. Rev. Food Sci. Nutr. **51**: 635–677. doi:10.1080/10408391003749942.

Wanasundara, J.P.D., Abeysekara, S.J., McIntosh, T.C., and Falk, K.C. 2012. Solubility differences of major storage proteins of brassicaceae oilseeds. JAOCS, J. Am. Oil Chem. Soc. **89**: 869–881. doi:10.1007/s11746-011-1975-9.

Wanasundara, J.P.D., MacIntosh, T.C., Perera, S.P., Withana-Gamage, T.S., Mitra, P., McIntosh, T.C., Perera, S.P., Withana-Gamage, T.S., and Mitra, P. 2016a. How might oilseeds help meet the protein challenge? OCL - Ol. Corps Gras Lipides **23**: D407. doi:10.1051/ocl/2016028.

Wanasundara, J.P.D., McIntosh, T.C., Perera, S.P., Withana-Gamage, T.S., and Mitra, P. 2016b. Canola/rapeseed protein-functionality and nutrition. OCL **23**: D407. doi:10.1051/ocl/2016028.

Wang, H., Guo, J., Lambert, K.N., and Lin, Y. 2007. Developmental control of Arabidopsis seed oil biosynthesis. Planta **226**: 773–783. doi:10.1007/s00425-007-0524-0.

Wang, J., Jian, H., Wei, L., Qu, C., Xu, X., Lu, K., Qian, W., Li, J., Li, M., and Liu, L. 2015. Genome-Wide Analysis of Seed Acid Detergent Lignin (ADL) and hull content in rapeseed (Brassica napus L.). PLoS One **10**: 1–18. doi:10.1371/journal.pone.0145045.

Wang, J., Singh, S.K., Du, C., Li, C., Fan, J., Pattanaik, S., and Yuan, L. 2016a. Comparative transcriptomic analysis of two *Brassica napus* near-isogenic lines reveals a network of genes that influences seed oil accumulation. Front. Plant Sci. **7**: 1–14. doi:10.3389/fpls.2016.01498.

Wang, J., Xian, X., Xu, X., Qu, C., Lu, K., Li, J., and Liu, L. 2017. Genome-Wide Association Mapping of Seed Coat Color in Brassica napus. J. Agric. Food Chem. **65**: 5229–5237.

doi:10.1021/acs.jafc.7b01226.

Wang, N., Chen, B., Xu, K., Gao, G., Li, F., Qiao, J., Yan, G., Li, J., Li, H., and Wu, X. 2016b. Association mapping of flowering time QTLs and insight into their contributions to rapeseed growth habits. Front. Plant Sci. **7**: 1–11. doi:10.3389/fpls.2016.00338.

Wang, X., Sun, L., Li, W., Peng, M., Chen, F., Zhang, W., Sun, C., Chen, S., Hua, W., and Zhang, J. 2020. Dissecting the genetic mechanisms of waterlogging tolerance in Brassica napus through linkage mapping and a genome-wide association study. Ind. Crops Prod. **147**: 112269. Elsevier. doi:10.1016/j.indcrop.2020.112269.

Wang, X., Zhang, C., Li, L., Fritsche, S., Endrigkeit, J., Zhang, W., Long, Y., Jung, C., and Meng, J. 2012. Unraveling the genetic basis of seed tocopherol content and composition in rapeseed (*Brassica napus* L.). PLoS One **7**: e50038. doi:10.1371/journal.pone.0050038.

Wang, Z.-H., Li, S.-X., and Malhi, S. 2008. Effects of fertilization and other agronomic measures on nutritional quality of crops. J. Sci. Food Agric. **88**: 7–23. doi:10.1002/jsfa.3084.

Wang, Z., Gerstein, M., and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. **10**: 57–63. doi:10.1038/nrg2484.

Waserman, S., Bégin, P., and Watson, W. 2019. IgE-mediated food allergy. Clin. Rev. Allergy Immunol. **57**: 244–260. BioMed Central. doi:10.1007/s12016-018-8710-3.

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

Wen, J., Xu, J.F., Long, Y., Wu, J.G., Xu, H.M., Meng, J.L., and Shi, C.H. 2016. QTL mapping based on the embryo and maternal genetic systems for non-essential amino acids in rapeseed (*Brassica napus* L.) meal. J. Sci. Food Agric. **96**: 465–473. doi:10.1002/jsfa.7112.

Werner, C.R., Voss-Fels, K.P., Miller, C.N., Qian, W., Hua, W., Guan, C.-Y., Snowdon, R.J., and Qian, L. 2018. Effective Genomic Selection in a Narrow-Genepool Crop with Low-Density Markers: Asian Rapeseed as an Example. Plant Genome **11**: 170084. doi:10.3835/plantgenome2017.09.0084.

Wieser, H. 2007. Chemistry of gluten proteins. Food Microbiol. **24**: 115–119. doi:10.1016/j.fm.2006.07.004.

de Wildt, D.J., and Speijers, G.J.A. 1984. Influence of dietary rapeseed oil and erucic acid upon myocardial performance and hemodynamics in rats. Toxicol. Appl. Pharmacol. **74**: 99–108. doi:10.1016/0041-008X(84)90275-8.

Withana-Gamage, T.S., Hegedus, D.D., Qiu, X., Yu, P., May, T., Lydiate, D., and Wanasundara, J.P.D. 2013. Characterization of <>Arabidopsis thaliana</i> lines with altered seed storage

protein profiles using synchrotron-powered FT-IR spectromicroscopy. J. Agric. Food Chem. **61**: 901–912. doi:10.1021/jf304328n.

Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F., and Quake, S.R. 2014a. Quantitative assessment of single-cell RNA-sequencing methods. Nat. Methods **11**: 41–46. doi:10.1038/nmeth.2694.

Wu, G., Bazer, F.W., and Cross, H.R. 2014b. Land-based production of animal protein: Impacts, efficiency, and sustainability. Ann. N. Y. Acad. Sci. **1328**: 18–28. doi:10.1111/nyas.12566.

Wu, J., and Muir, A.D. 2008. Comparative structural, emulsifying, and biological properties of 2 major canola proteins, cruciferin and napin. J. Food Sci. **73**. doi:10.1111/j.1750-3841.2008.00675.x.

Wu, J., Zhao, Q., Liu, S., Shahid, M., Lan, L., Cai, G., Zhang, C., Fan, C., Wang, Y., and Zhou, Y. 2016. Genome-wide association study identifies new loci for resistance to sclerotinia stem rot in Brassica napus. Front. Plant Sci. **7**: 1–12. doi:10.3389/fpls.2016.01418.

Wu, Y., and Messing, J. 2014. Proteome balancing of the maize seed for higher nutritional value. Front. Plant Sci. **5**: 1–5. doi:10.3389/fpls.2014.00240.

Würschum, T., Liu, W., Maurer, H.P., Abel, S., and Reif, J.C. 2012. Dissecting the genetic architecture of agronomic traits in multiple segregating populations in rapeseed (Brassica napus L.). Theor. Appl. Genet. **124**: 153–161. doi:10.1007/s00122-011-1694-5.

Xu, L., Hu, K., Zhang, Z., Guan, C., Chen, S., Hua, W., Li, J., Wen, J., Yi, B., Shen, J., Ma, C., Tu, J., and Fu, T. 2015. Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (Brassica napus L.). DNA Res. **23**: 43–52. doi:10.1093/dnares/dsv035.

Xu, Y., Li, P., Yang, Z., and Xu, C. 2017. Genetic mapping of quantitative trait loci in crops. Crop J. **5**: 175–184. Elsevier B.V. doi:10.1016/j.cj.2016.06.003.

Yang, C., Wang, Y., Vasanthan, T., and Chen, L. 2014. Impacts of pH and heating temperature on formation mechanisms and properties of thermally induced canola protein gels. Food Hydrocoll. **40**: 225–236. Elsevier Ltd. doi:10.1016/j.foodhyd.2014.03.011.

Yang, H., Wu, J.J., Tang, T., Liu, K. De, and Dai, C. 2017. CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in *Brassica napus*. Sci. Rep. **7**: 1–13. Springer US. doi:10.1038/s41598-017-07871-9.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J.H., Batchelor, W.D., Xiong, L., and Yan, J. 2020. Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. Mol. Plant **13**: 187–214. Elsevier Ltd. doi:10.1016/j.molp.2020.01.008.

Yang, W., Pollard, M., Li-Beisson, Y., Beisson, F., Feig, M., and Ohlrogge, J. 2010. A distinct type of glycerol-3-phosphate acyltransferase with sn-2 preference and phosphatase activity

producing 2-monoacylglycerol. Proc. Natl. Acad. Sci. U. S. A. **107**: 12040–12045. doi:10.1073/pnas.0914149107.

Yang, W., Simpson, J.P., Li-Beisson, Y., Beisson, F., Pollard, M., and Ohlrogge, J.B. 2012. A land-plant-specific glycerol-3-phosphate acyltransferase family in arabidopsis: Substrate specificity, *sn*-2 preference, and evolution. Plant Physiol. **160**: 638–652. doi:10.1104/pp.112.201996.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J.C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Lavidas, I., Le, T., Lemos, D., Martinez, J.G., Maurel, T., McDowall, M., McMahon, A., Mohanan, S., Moore, B., Nuhn, M., Oheh, D.N., Parker, A., Parton, A., Patricio, M., Sakthivel, M.P., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sycheva, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Flint, B., Frankish, A., Hunt, S.E., Iisley, G., Kostadima, M., Langridge, N., Loveland, J.E., Martin, F.J., Morales, J., Mudge, J.M., Muffato, M., Perry, E., Ruffier, M., Trevanion, S.J., Cunningham, F., Howe, K.L., Zerbino, D.R., and Flicek, P. 2020. Ensembl 2020. Nucleic Acids Res. **48**: D682–D688. doi:10.1093/nar/gkz966.

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

Yeung, E.C., Rahman, M.H., and Thorpe, T.A. 1996. Comparative development of zygotic and microspore-derived embryos in *Brassica napus* L. cv Topas. I. Histodifferentiation. Int. J. Plant Sci. **157**: 27–39. doi:10.1086/297317.

Yin, L. 2020. Package ‘ CMplot .’ [Online] Available: https://cran.r-project.org/web/packages/CMplot/index.html.

Yin, X., He, D., Gupta, R., and Yang, P. 2015. Physiological and proteomic analyses on artificially aged *Brassica napus* seed. Front. Plant Sci. **6**: 1–11. doi:10.3389/fpls.2015.00112.

You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. 2018. Development and applications of a high throughput genotyping tool for polyploid crops: Single nucleotide polymorphism (SNP) array. Front. Plant Sci. **9**. doi:10.3389/fpls.2018.00104.

Yu, J., and Buckler, E.S. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. **17**: 155–160. doi:10.1016/j.copbio.2006.02.003.

Zarkadas, C.G., Gagnon, C., Poysa, V., Khanizadeh, S., Cober, E.R., Chang, V., and Gleddie, S. 2007. Protein quality and identification of the storage protein subunits of tofu and null soybean genotypes, using amino acid analysis, one- and two-dimensional gel electrophoresis, and tandem mass spectrometry. Food Res. Int. **40**: 111–128.

doi:10.1016/j.foodres.2006.08.005.

Zhai, Y., Yu, K., Cai, S., Hu, L., Amoo, O., Xu, L., Yang, Y., Ma, B., Jiao, Y., Zhang, C., Khan, M.H.U., Khan, S.U., Fan, C., and Zhou, Y. 2019. Targeted mutagenesis of BnTT8 homologs controls yellow seed coat development for effective oil production in Brassica napus L. Plant Biotechnol. J. **18**: pbi.13281. doi:10.1111/pbi.13281.

Zhang, F., Xiao, X., Xu, K., Cheng, X., Xie, T., Hu, J., and Wu, X. 2020. Genome-wide association study (GWAS) reveals genetic loci of lead (Pb) tolerance during seedling establishment in rapeseed (Brassica napus L.). BMC Genomics **21**: 1–12. BMC Genomics. doi:10.1186/s12864-020-6558-4.

Zhang, F., Xiao, X., Yan, G., Hu, J., Cheng, X., Li, L., Li, H., and Wu, X. 2018. Association mapping of cadmium-tolerant QTLs in Brassica napus L. and insight into their contributions to phytoremediation. Environ. Exp. Bot. **155**: 420–428. Elsevier. doi:10.1016/j.envexpbot.2018.07.014.

Zhang, J., Mason, A.S., Wu, J., Liu, S., Zhang, X., Luo, T., Redden, R., Batley, J., Hu, L., and Yan, G. 2015. Identification of putative candidate genes for water stress tolerance in canola (Brassica napus). Front. Plant Sci. **6**: 1–13. doi:10.3389/fpls.2015.01058.

Zhang, Y.M., Jia, Z., and Dunwell, J.M. 2019. Editorial: The applications of new multi-locus gwas methodologies in the genetic dissection of complex traits. Front. Plant Sci. **10**: 1–6. doi:10.3389/fpls.2019.00100.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **42**: 355–360. Nature Publishing Group. doi:10.1038/ng.546.

Zhao, J., Becker, H.C., Zhang, D., Zhang, Y., and Ecke, W. 2006. Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor. Appl. Genet. **113**: 33–38. doi:10.1007/s00122-006-0267-5.

Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., and Nordborg, M. 2007. An Arabidopsis example of association mapping in structured samples. PLoS Genet. **3**: 0071–0082. doi:10.1371/journal.pgen.0030004.

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., Liu, J., Sun, X., Wang, X., Xu, J., Hua, W., and Wang, H. 2017. Genome-wide association study reveals candidate genes for control of plant height, branch initiation height and branch number in rapeseed (Brassica napus L.). Front. Plant Sci. **8**: 1–13. doi:10.3389/fpls.2017.01246.

Zhu, C., Gore, M., Buckler, E.S., and Yu, J. 2008. Status and Prospects of Association Mapping in Plants. Plant Genome **1**: 5–20. doi:10.3835/plantgenome2008.02.0089.

9.2 Appendix

Appendix Table S1. All genes in linkage disequilibrium with single nucleotide polymorphism markers showing suggestive association to seed protein content, oil content, and cruciferin content, respectively, in a population of 43 spring *Brassica napus* L. genotypes grown under greenhouse conditions in 2019.

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
Protein	MLM +K+Q	9	Bn-chrC05_random-p460689	BnaC05g00860D	endoribonuclease Dicer homolog 1
				BnaC05g00870D	B3 domain-containing transcription factor NGA3
				BnaC05g00880D	protein arv1 homolog
		54	Bn-chrC05-p6839299	BnaC05g11790D	serine-threonine kinase receptor-associated protein
			Bn-chrC05-p6841152	BnaC05g11800D	probable disease resistance protein At5g63020
			Bn-chrC05-p6847199	BnaC05g11810D	probable disease resistance protein At5g63020

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g11820D	probable disease resistance protein At5g63020
	All MLM	flanking	Bn-chrA04-p12364246	BnaA04g14620D	glutamate receptor 2.2
				BnaA04g14630D	ADP-ribosylation factor 1
				BnaA04g14640D	zinc finger protein CONSTANS-LIKE 3
				BnaA04g14650D	UPF0329 protein ECU05_1680/ECU11_0050-like
				BnaA04g14660D	eukaryotic translation initiation factor 3 subunit B
				BnaA04g14670D	F-box/kelch-repeat protein At3g27150-like
				BnaA04g14680D	cysteine-rich receptor-like protein kinase 18 isoform X2

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaA04g14690D	RPP27 protein
				BnaA04g14700D	agamous-like MADS-box protein AGL61
				BnaA04g14710D	agamous-like MADS-box protein AGL61
				BnaA04g14720D	DnaJ/Hsp40 cysteine-rich domain superfamily protein
Oil	MLM +K	flanking	Bn-chrC05_random-p196753	BnaC05g00340D	probable serine/threonine-protein kinase At5g41260 isoform X1
				BnaC05g00350D	UvrABC system protein A
				BnaC05g00360D	uncharacterized protein LOC103844674
				BnaC05g00370D	NAC domain-containing protein 2

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g00380D	acyl-CoA thioesterase II
				BnaC05g00390D	protein PAIR1-like isoform X1
				BnaC05g00400D	U-box domain-containing protein 54
				BnaC05g00410D	putative U-box domain-containing protein 55
				BnaC05g00420D	SIGNAL PEPTIDE PEPTIDASE-LIKE 4
				BnaC05g00430D	random slug protein 5
				BnaC05g00440D	aquaporin PIP1-3-like
				BnaC05g00450D	glycerol-3-phosphate 2-O-acyltransferase 4-like
				BnaC05g00460D	---NA---

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g00470D	transferring glycosyl group transferase (DUF604)
				BnaC05g00480D	protein BPS1, chloroplastic
		51	Bn-chrC05-p4943192	BnaC05g09160D	ethylene-responsive transcription factor ESR1-like
			Bn-chrC05-p4945701	BnaC05g09170D	transportin MOS14 isoform X1
				BnaC05g09180D	putative F-box/kelch-repeat protein At3g24610
				BnaC05g09190D	eukaryotic peptide chain release factor subunit 1-2
				BnaC05g09200D	wd repeat-containing protein lwd1
				BnaC05g09210D	glyceraldehyde-3-phosphate dehydrogenase GAPA2, chloroplastic isoform X1

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g09220D	nudix hydrolase 12, mitochondrial-like isoform X1
				BnaC05g09230D	nudix hydrolase 12, mitochondrial-like
				BnaC05g09240D	uncharacterized protein LOC106400590
				BnaC05g09250D	transcription factor SCREAM2-like
				BnaC05g09260D	transcription factor SCREAM2-like
				BnaC05g09270D	transcription factor SCREAM2-like
				BnaC05g09280D	Phosphoglycerate mutase family protein
				BnaC05g09290D	uncharacterized protein BNAC05G09290D
				BnaC05g09300D	V-type proton ATPase subunit C

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g09310D	nucleotide binding protein
				BnaC05g09320D	unnamed protein product
				BnaC05g09330D	malate dehydrogenase, chloroplastic
				BnaC05g09340D	uncharacterized protein LOC106448014
				BnaC05g09350D	probable sugar phosphate/phosphate translocator At1g12500
				BnaC05g09360D	guard cell S-type anion channel SLAC1
				BnaC05g09370D	vacuolar protein sorting-associated protein 18 homolog

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g09380D	probable LRR receptor-like serine/threonine-protein kinase At1g12460
				BnaC05g09390D	transmembrane protein 64
				BnaC05g09400D	kinesin-like protein KIN-UA isoform X2
				BnaC05g09410D	ACT domain-containing protein ACR8
				BnaC05g09420D	ATP-dependent Clp protease proteolytic subunit-related protein 2, chloroplastic
				BnaC05g09430D	RNA polymerase II transcription factor B subunit 5
				BnaC05g09440D	protein cornichon homolog 4

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaC05g09450D	uncharacterized protein BNAC05G09450D isoform X1
	ALL C05	5	Bn-chrC05_random-p161562	BnaC05g00300D	K(+) efflux antiporter 1, chloroplastic-like
				BnaC05g00310D	LIM domain-containing protein PLIM2b
Cruciferin	MLM +K +PCA	94	Bn-chrA06-p5218016	BnaA06g09770D	sulfhydryl oxidase 1
			Bn-chrA06-p5250207	BnaA06g09780D	unnamed protein product
			Bn-chrA06-p5258210	BnaA06g09790D	putative glutamine amidotransferase GAT1_2.1

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
				BnaA06g09800D	auxin-responsive protein IAA32-like isoform X1
				BnaA06g09810D	E3 ubiquitin-protein ligase RHA2A-like
				BnaA06g09820D	zinc finger BED domain-containing protein DAYSLEEPER-like
				BnaA06g09830D	fruit protein pKIWI502
				BnaA06g09840D	protein DETOXIFICATION 12
				BnaA06g09850D	fasciclin-like arabinogalactan protein 19
				BnaA06g09860D	protein-protein interaction regulator family protein
	MLM +K+Q	129	Bn-chrA07-p11076585	BnaA07g12030D	probable inactive receptor kinase At5g67200

Appendix Table S1 continued

Trait	Model	Linkage block	Markers in block	Genes in Block	Description
			Bn-chrA07-p11107078	BnaA07g12040D	ethylene-responsive transcription factor
				BnaA07g12050D	AP2-like ethylene-responsive transcription factor TOE3
				BnaA07g12060D	protein ENHANCED PSEUDOMONAS SUSCEPTIBILTY 1-like

Appendix Table S2. Primers used for quantitative reverse transcription PCR to validate a comparative transcriptomic experiment of developing seeds at 25 days after pollination of the high-cruciferin *Brassica napus* L. cultivar 'Red River 1852' and the low-cruciferin cultivar 'Wizzard.'

Target Gene identifier	Direction	Sequence (5' to 3')
BnaA10g02240D	Forward	TGACCCGTCAAGTGCTGATGT
	Reverse	GGCAGCACCATAGCGTTGTTA
BnaA08g21340D	Forward	GGTATCGCAGCCACTGTGTTC
	Reverse	AAACCCATTTGGGATTGCAACAA
BnaA05g05070D	Forward	AGATCCCGACATGGCTTGGTAT
	Reverse	TCACCGTTGTTTACCCTAGGAGG
BnaC09g21610D	Forward	TGAAAAACAAAATGCTGGTGAATGT
	Reverse	GGGGATGGAGATGCCCTAACAA
BnaCnng01660D	Forward	TCGGAGGTTCTACACCTGGA
	Reverse	ATCACAACGGGAGGACCAGC
BnaA07g20190D	Forward	ATGGCCACCGAGGGTAACTT
	Reverse	TGATGGAAAACACGGAGTTGTCTG

Appendix Table S2 continued

Target Gene identifier	Direction	Sequence (5’ to 3’)
BnaC01g36680D	Forward	ACATGCTCAAGACTCATGGGAGA
	Reverse	GCGAGTCCCGTGGCTATGAA
BnaA03g05550D	Forward	CAGGAGAGTGACAACAGGGGG
	Reverse	CCCCTCTAACGCCCTGAACC
BnaC08g17990D	Forward	GCGGCTCGAGTCTGTTTAT
	Reverse	TCAAAGTTTTCACTGACTCTTTTGT
BnaA03g52820D	Forward	TCTGCAGAAACTACATCCATGACA
	Reverse	ACCTGGCCGACTTTATTCCGA