Utilization of variation to understand *Camelina sativa* genome evolution

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon

By

RAJU CHAUDHARY

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other use of the material in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences

College of Agriculture and Bioresources

University of Saskatchewan

51 Campus Drive

Saskatoon, Saskatchewan S7N 5A8, Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9, Canada

Abstract

Camelina sativa is an oilseed crop gaining interest for its oil content, protein content, and potential as a new oilseed for human consumption. The main disadvantages of this crop are its smaller seed size and lower total yield compared to other commercial oilseed crops; however, breeding efforts has been progressing to improve yield traits. A low level of genetic diversity and limited breeding efforts have been identified as impediments in *C. sativa* crop improvement. This study was designed to improve access to genetic diversity in *C. sativa* by developing genetic tools and identifying genetic mechanisms to accelerate *C. sativa* breeding. The objectives of this study were: to explore the genetic diversity in available *Camelina* germplasm using Genotyping-by-Sequencing (GBS), with a focus on close relatives of *C. sativa* and a collection from Ukraine and Russia; to develop segregating generations through intra- and interspecific hybridization; and to complete whole genome transcriptome analysis to observe gene expression patterns across subgenomes in hexaploid species of *Camelina*. Genetic markers in this study were developed using GBS, whereas whole transcriptome analysis was performed for subgenome dominance analysis. The genetic diversity study with 193 genotypes identified two subpopulations in *C. sativa,* where *C. microcarpa* was found to be a close relative of this species. Winter *C. sativa* species, such as *C. sativa* ssp. *pilosa* and *C. alyssum,* formed a separate clade and were closely-associated with *C. microcarpa*. Principal coordinate and phylogenetic analysis differentiated the genotypes based on species and subpopulations. Mapping of reads to the reference genome identified *C. neglecta* as a progenitor species for the first subgenome of *C. sativa*. Likewise, a tetraploid was identified that encompassed the first and second subgenomes, and a novel *C. microcarpa* species differing from *C. sativa* in terms of genome structure was also identified. Flow cytometry analysis and chromosome count validated the read mapping and confirmed that the novel *C. microcarpa* possessed 19 chromosomes (n, haploid number) with a different third subgenome not present in *C. sativa*. The inter- and intraspecific hybridizations enabled genetic linkage maps to be developed, where a common *C. sativa* genotype was hybridized with other related species. A mapping study identified four quantitative trait loci (QTL) associated with winter behaviour in *C. sativa*. The winter trait mapped to one locus on chromosome 8 (subgenome 1) in *C. sativa* ssp. *pilosa*, to two loci in *C. alyssum* on chromosomes 13 (subgenome 2) and 20 (subgenome 3), and to one locus on chromosome 13 (subgenome 2) in *C. microcarpa*. All of the QTL represented homologous segments in the *C. sativa* reference

genome and were proximate to a major flowering gene, *Flowering Locus C* (*FLC*). Differential gene expression analysis between the parents at the early seedling stage suggested *FLC* could be a candidate gene responsible for vernalization responses in winter *C. sativa* populations. In addition, interspecific hybridization identified a homoeologous recombination (HeR) event between subgenome 1 of *C. sativa* with subgenome 3 of *C. microcarpa* (n = 19), and a number of anueploids were identified, as expected. The nature of HeR could create challenges for the success of conventional breeding activities in *Camelina* species, as recombination could occur between any subgenomes due to the undifferentiated nature of the subgenomes. However, variation in morphology, such as leaf characteristics, days to flowering and fertility suggested a huge potential for increasing genetic variability in *C. sativa* by use of distantly-related *Camelina* species. Subgenome dominance has evolutionary significance and can play an important role in improving phenotypic diversity. Subgenome dominance analysis suggested the third subgenome was dominant in the case of *Camelina* species with n = 20, whereas the second subgenome was dominant for *Camelina* species with n = 19 and was correlated with the age of divergence of the subgenomes from *C. neglecta*. Overall, the results provided insight into the subgenome structure and a first step towards identifying the mechanism of a stepwise whole genome duplication process in polyploid *C. sativa*, which would be instrumental in developing genetic tools for *Camelina* breeding activities.

Acknowledgements

I would like to express my sincere gratitude and thanks to my co-supervisors Drs. Isobel Parkin and Kirstin Bett for this wonderful research opportunity to work in a new environment, support, motivation, and insightful guidance throughout my study and research. I am grateful to my advisory committee, Drs. Dwayne Hegedus, Aaron Beattie, Matthew Links and Tom Warkentin, for their valuable suggestions and constructive feedback on my research project and dissertation. I would also like to thank Dr. Mark Tepfer (Institut Jean-Pierre Bourgin, INRA-Versailles) for accepting to be my external examiner for this dissertation.

I would especially like to thank Lily Tang, Erin Higgins, Tina Bundrock, Myrtle Harrington and Siu Wah Wu for the technical support they provided for molecular and protein analysis. I am thankful to Dr. Andrew Sharpe for the guidance and support for this study. I appreciate the support from Dr. Christina Eynck and her team for green house and field studies. I also appreciate Drs. Sampath Perumal, Lingling Jin and Kevin Koh for their advice and support in bioinformatics analysis. I am thankful to Dr. Mark Smith for seed oil content analysis and Dr. Annaliese Mason for assistance with chromosome counts. I am also thankful to Drs. Sateesh Kagale, Raju Soolanayakanahally, and Yogendra Khedikar for their valuable suggestions during this study period. Special thanks to Dr. Kiran Baral and Sandesh Neupane for constant feedback during my study.

I am grateful for the financial support provided by the Global Institute of Food Security, Saskatoon from the projects "Developing *Camelina sativa* as a modern crop platform", and Plant Phenotyping and Imaging Research Centre, Flagship 1.

Finally, a special thanks to my family. Words can not express how grateful I am to my parents Ram Bahadur Chaudhary and Fulmati Chaudhary, brother Raj Kumar, sister Ranjana for continued support and encouragement. Special thanks to my beloved wife, Yashodhara Chaudhary who has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

ACK	Ancestral Crucifer Karyotype
AFLP	Amplified Fragment Length Polymorphism
BAM	Binary Alignment Map
bp	basepair
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	Cetyl trimethylammonium bromide
DAPI	4′, 6-diamidino-2-phenylindole
DNA	Deoxyribonucleic Acid
DTF	Days to Flower
EMS	Ethyl Methanesulfonate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GB	Genomic Blocks
GBS	Genotyping by Sequencing
GWAS	Genome-wide Association Study
hr	hour
kDa	kilodaltons
MCMC	Markov Chain Monte Carlo
mya	millions year ago
NCBI	National Center for Biotechnology Information
ndhF	NADH dehydrogenase F
^{o}C	Degree Celsius
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PIC	Polymorphic Information Content
PUFA	Polyunsaturated fatty acids
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RE	Repeat Elements
RNA	Ribonucleic Acid
SG	Subgenome
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeats
TE	Transposable Elements
TPM	Transcripts Per Kilobase Million
UTR	Untranslated region
UV	Ultraviolet
WGD	Whole Genome Duplication

Chapter 1. Introduction

1.1 Background

Camelina sativa (n=20) is an industrial oilseed crop from the family Brassicaceae. It is an ancient crop, hardy to drought and cold, and gaining prominence as a next generation biofuel. It is well suited to growth on the Canadian prairies with good germination capacity at low temperature and has no seed dormancy. The crop is primarily self-pollinated in nature (Francis and Warwick 2009) with low rates of out-crossing (Walsh et al. 2012b). Furthermore, it is a short duration crop and is resistant to a number of common pathogens and pests that commonly affect Brassicaceae species. Despite these benefits, there is interest in increasing seed size while maintaining oil content, and in improving stand establishment and overall yield for this crop (Gugel and Falk 2006). However, *Camelina* breeding is handicapped by low levels of genetic diversity in available germplasm (Zelt and Schoen 2016), thus further breeding could include hybridization of *C. sativa* with wild relatives to increase diversity (Séguin-Swartz et al. 2013). In order to allow further improvements to the crop it is essential to identify novel sources of allelic variation.

Analysis of genetic diversity in various *C. sativa* germplasm collections has been carried out using multiple marker types; Random Amplified Polymorphic DNA (RAPD) (Vollmann et al. 2005), Amplified Fragment Length Polymorphism (AFLP) (Ghamkhar et al. 2010), Simple Sequence Repeats (SSR) (Manca et al. 2013) and Single Nucleotide Polymorphism (SNP) marker analyses (Singh et al. 2015; Luo et al. 2019a), and all concluded that there were low levels of genetic diversity relative to other crops. Genetic diversity is vital for developing a robust breeding strategy and to incorporate the necessary variation for further crop improvement. Zelt and Schoen (2016) found some level of mid-parent heterosis, as well as best-parent heterosis in intraspecific crossing of this crop, which suggests there may be some value in defining the absolute level of relatedness among available germplasms. There are also reports of possibilities for wide hybridization to create variability in *Camelina* (Julié-Galau et al. 2014; Séguin-Swartz et al. 2013). A low level of genetic differentiation has been reported between the subgenomes of *C. sativa* (Kagale et al. 2014a), yet there has been no report of introgression between these

subgenomes, which suggests a high level of pairing control within this species. Therefore, interspecific hybridization in *Camelina* species could help to observe the behaviour of *C. sativa* chromosomes during meiosis upon hybridization with the related species. *Camelina sativa* has a number of related species, some with lower chromosome numbers and two (*C. microcarpa* and *C. alyssum)* that share the same autosomal chromosome number with *C. sativa* (Francis and Warwick 2009). As found in other crops, interspecific crossing could be a useful mechanism to capture additional variation (Zhang and Auer 2020). The availability of a genome sequence of *C. sativa* (Kagale et al. 2014a) has provided an opportunity to apply techniques such as Genotyping by Sequencing (GBS) to this crop, a highly useful tool for determining the genetic makeup of plants and developing markers associated with traits.

This thesis research has defined the level of variation among *C. sativa* and related species using modern marker techniques, and also utilized standard intra- and interspecific crosses to introduce novel allelic variation into current *C. sativa* breeding lines. In addition, this project studied subgenome dominance in the polyploid *C. sativa* and related species.

1.2 Hypothesis

1. The related species of *C. sativa* contain novel allelic variation not present within available *C. sativa* accessions.
2. Homology between chromosomes of *C. sativa* and related species will be sufficient to allow genetic introgression of useful variation.
3. Differential subgenome dominance and fractionation occurs in *Camelina* as the ploidy level increases.

1.3 Objectives

1. Elucidation of the degree of relatedness among *C. sativa* accessions and related species through GBS.
2. Development of *Camelina* populations through intra- and interspecific crosses and genetic analyses of developed populations.
3. Whole genome transcriptome analysis for the hexaploid *Camelina* species.

Chapter 2. Literature Review

2.1 Background

Camelina sativa or camelina is also known by the names 'false flax' and 'gold of pleasure'(Putnam et al. 1993). Camelina is economically important due to its unique long chain fatty acid and protein content in the seed, which could be utilized for human consumption as an alternative vegetable oil, for the feed industry as a source of protein, and in particular the aquaculture industry as an alternative to fish oil. Other applications include using *C. sativa* oil as a source of biofuel, for which a number of studies are ongoing with a successful test flight using *C. sativa* derived jet fuel (Shonnard et al. 2010). The center of origin of *C. sativa* is believed to be in regions of Russia and Ukraine (Ghamkhar et al. 2010); however, it is also distributed among European countries (Vollmann et al. 2005) with similar climatic conditions. This crop has gained renewed interest to develop it as a viable, contemporary crop (Faure and Tepfer 2016; Vollmann and Eynck 2015).

Camelina has a long history as revealed by evidence of ancient cultivation from charred and mineralized seed from the Republic of Armenia believed to be from the Neolithic period (Hovsepyan and Willcox 2008). It is also believed that camelina was popular during the Bronze age as a vegetable oil for human consumption, as well as for fuel in West European countries. Larsson (2013) claimed that camelina was popularly cultivated as a source of oil during the Roman Iron age in Sweden, but its cultivation collapsed after this time possibly due to different oilseed options becoming available. Despite such a long history, this crop has not been subjected to substantial amounts of breeding activities, as reflected by the low level of genetic diversity in the available *C. sativa* germplasm, or it could be that the current *C. sativa* is a different species from the camelina cultivated in the past (Čalasan et al. 2019). Although it seems a narrow genetic bottleneck has impacted the current level of genetic diversity, a number of attempts have been made to increase the genetic diversity and improve this crop. This review explores the current situation with respect to genetic diversity in *C. sativa* and the progress made in the context of genomics and the use of hybridization.

2.2 Taxonomy and classification of *Camelina* species

Camelina sativa is a member of the Brassicaceae family, which consists of 3709 species in 338 genera (Warwick and Al-Shehbaz 2006). Based on morphology, origin, and molecular methods, paricularly using the chloroplast *ndhF* gene, Al-Shehbaz et al. (2006) described 25 tribes of the Brassicaceae, later extended to 49 tribes (Al-Shehbaz 2012) classified into three lineages. Tribe Camelineae lies within Lineage I along with 14 other tribes. Camelineae includes eight genera viz. *Arabidopsis, Camelina, Capsella, Catolobus, Chrysochamela, Neslia, Noccidium* and *Pseudoarabidopsis* (Al-Shehbaz 2012) with the genus *Camelina* consisting of 11 species (a list can be found in Warwick and Al-Shehbaz (2006)) of which *C. sativa, C. alyssum, C. microcarpa, C. hispida, C. rumelica* and *C. neglecta* are commonly available species (Martin et al. 2017). *Camelina* species are usually distinguished from other species of mustard by their small seed size, pale yellow flowers and tear-shaped pods (Francis and Warwick 2009). The classification of *Camelina* species has been revised several times due to the lack of distinguishable characteristics, where higher plasticity and geographical growing conditions might have influenced the formation of particular morphological traits that differentiate the species (Angelini et al. 1997; Berti et al. 2011; Neupane et al. 2019). Based on highly conserved *ndhF* gene sequences, *Camelina* is more closely related with *Neslia* species within the Camelineae tribe and three species of *Camelina* viz. *sativa, microcarpa* and *alyssum* are extremely closely related (**Figure 2.1**). Among them, *C. alyssum* is almost morphologically indistinguishable from *C. sativa,* whereas *C. microcarpa* diverged earlier from *C. sativa* (Čalasan et al. 2019).

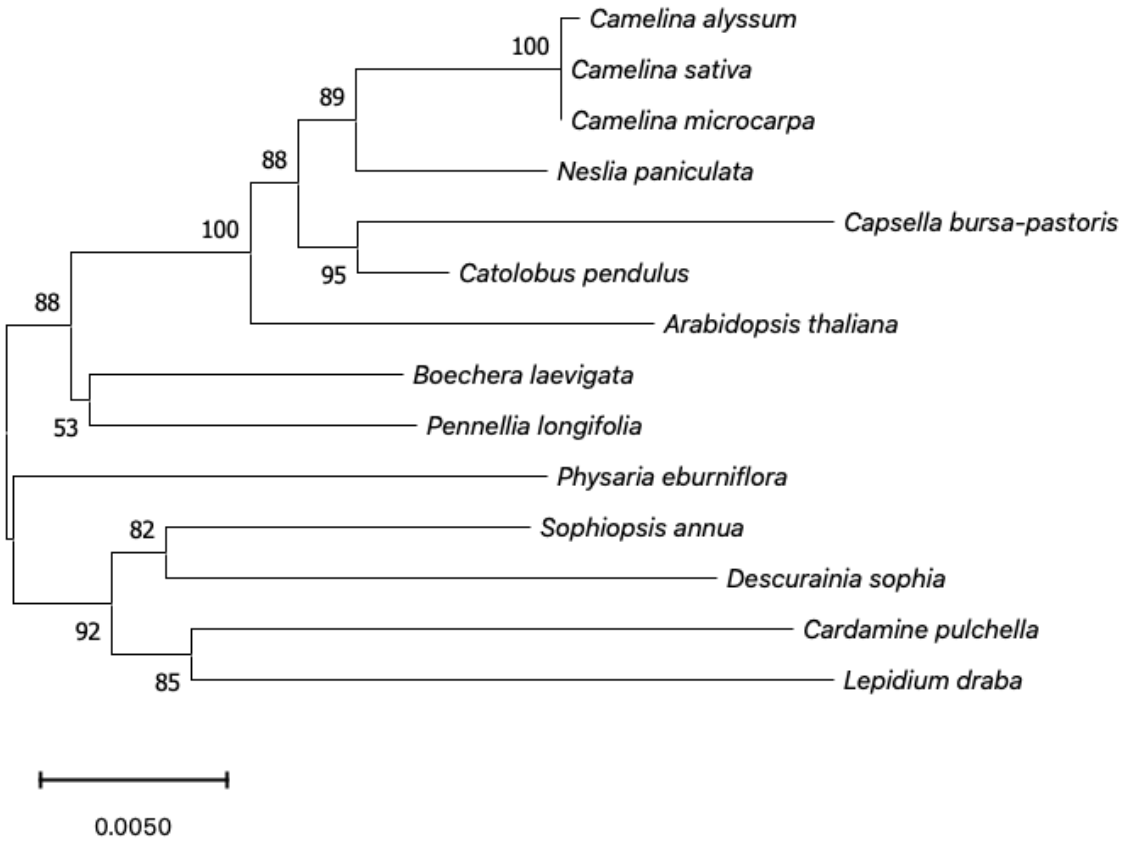


Figure 2.1 Phylogenetic relationship of different *Camelina* species and species of other tribes of Lineage I as inferred by *ndhF* sequences. The Maximum Likelihood tree was prepared from MEGA software (Stecher et al. 2020) using *ndhF* gene sequence from accessions HM120269.1 (*Camelina sativa*), DQ288746.1 (*Camelina microcarpa*), HM120265.1 (*Camelina alyssum*), HM120282.1 (*Neslia paniculate*), DQ288748.1 (*Capsella bursa-pastoris*), DQ288732.1 (*Catolobus pendulus*), HM120263.1 (*Arabidopsis thaliana*), DQ288739.1 (*Boechera laevigata*), JQ323089.1 (*Pennellia longifolia*), JQ323069.1 (*Physaria eburniflora*), DQ288831.1 (*Sophiopsis annua*), DQ288759.1 (*Descurainia Sophia*), DQ288749.1 (*Cardamine pulchella)* and DQ288790.1 (*Lepidium draba*).

Identifying *C. sativa* has been difficult, with two dominant ecotypes, where the annual (or spring-type) was considered as *C. sativa* and the biennial (or winter-type) as *C. microcarpa* (Pleesers et al. 1962). Both species share the same number of chromosomes and similar morphological characteristics; however, winter-types of *C. sativa* were identified and reclassified as a sub-species, *C. sativa* ssp. *pilosa,* leaving *C. microcarpa* as a separate species, mainly distinguished by a smaller pod and seed size. More recently, winter-type *C. sativa* lines that do

not cluster into one group or present as a subspecies have been found, and winter-type *C. sativa* is now assumed to be separate from *C. sativa* ssp. *pilosa* (Manca et al. 2013). Discrepancy has also arisen with the classification of *C. alyssum* which has both winter and spring ecotypes, yet has close resemblance to *C. sativa*, but possesses hairy stems unlike *C. sativa* (Čalasan et al. 2019). However, all these species have 20 chromosomes and are hexaploid. Most recently, identification of another *C. microcarpa*-like species indicated a need to reclassify the *C. microcarpa* species, with the new species having 19 chromosomes and sharing only 13 chromosomes with *C. sativa* (Chaudhary et al. 2020).

Chromosome number has played an important role in distinguishing *Camelina* germplasm and species. The chromosome number in these species varies based on geographical distribution, but the most common chromosome number for *C. sativa*, C. *microcarpa* and *C. alyssum* is 2n = 40. However, 2n = 12 (France), 28 (Spain), 40 (China), 26 (Bulgaria) and 40 (Argentina, Canada, Iceland, Poland) have also been reported for *C. sativa.* Similarly, 2n = 16 (United States), 26 (France, Morocco), 38 (Czech Rep.), 16 (Russia), 20 (Russia), 26 (France, Morocco, Spain), 32 (Russia) and 40 (Canada, Czech, Slovakia, Hungary, Iceland, Iran, Poland) have been reported for *C. microcarpa*; 2n = 12 (Hungary, United States), 24 (Iran), 16 (Greece), 26 (Iran), 32 (Russia) and 40 (Afghanistan) for *C. rumelica*; and 2n = 14 (Iran) for *C. hispida* (Warwick and Al-Shehbaz 2006; Galasso et al. 2015). This variation in chromosome number could result from different ploidy levels in *Camelina* species or these anomalies could be due to lack of proper identification of these species. The current understanding is that *C. sativa* has a hexaploid structure of 2n = 40 chromosomes, similarly hexaploid *C. alyssum* has 2n = 40, *C. microcarpa* is either tetraploid 2n = 26 or hexaploid with 2n = 38 and 40, *C. neglecta* is diploid with 2n = 12, *C. rumelica* is tetraploid with 2n = 26, *C. hispida* is diploid with 2n = 14, and *C. laxa* is diploid with 2n =12 (Chaudhary et al. 2020).

2.3 Biology of *Camelina sativa*

Camelina sativa is a herbaceous annual plant, growing to a height of 80-90 cm, with a tap root system. It is self-pollinating with a very low rate of outcrossing (0.09 to 0.78 percent) and maximum pollen-mediated gene flow through wind has been reported to be a distance of 0.2 m

(Walsh et al. 2015; Walsh et al. 2012b). Groeneveld and Klein (2014) reported that although insects are attracted to *C. sativa* flowers, there was no significant pollination by insects compared to self-pollination in *C. sativa*. This crop is hardy compared to other oilseed crops, such as mustard, rape or linseed, and can survive freezing temperatures and emerge well at 0 ºC (Allen et al. 2014). It has the potential to flourish within a wide range of sowing times (Schillinger et al. 2012) and is also highly adapted to multiple environments. *Camelina sativa* is preferentially cultivated on marginal land characterized by low soil moisture (Hunsaker et al. 2013). *Camelina sativa* produces pale yellow flowers about 5-7 mm in diameter, the flowers are mainly autogamous (Zubr 1997) with medium size pollen grains found in these species, viz. *C. sativa* 35.6 µm, *C. alyssum* 34.6 µm and *C. microcarpa* 37.6 µm (Sagun and Auer 2017). The seed of *C. sativa* is larger (1.6 gm per 1000 seed) in comparison to other species, such as *C. alyssum* (1.1 gm per 1000 seed), *C. microcarpa* (0.3 gm per 1000 seed) and *C. rumelica* (0.3 gm per 1000 seed) (Séguin-Swartz et al. 2013).

Camelina is a relatively short duration crop taking 39 to 44 days to flower (20% flowering) and 89 to 97 days to mature (95% maturity) in western Canada (Gugel and Falk 2006). Morphological traits of interest in *C. sativa* are plant height, 1000 seed weight, stem width and number of branches. High variation in these traits was reported in different geographical regions, where the plant height ranged from 49 to 57 cm (Gugel and Falk 2006), seed yield ranged from 1.15 to 2.20 t/ha, 1000 seed weight ranged from 1.32 to 1.76 gm (Gehringer et al. 2006) and branches per plant ranged from 10.44 to 15.92 (Lošák et al. 2011).

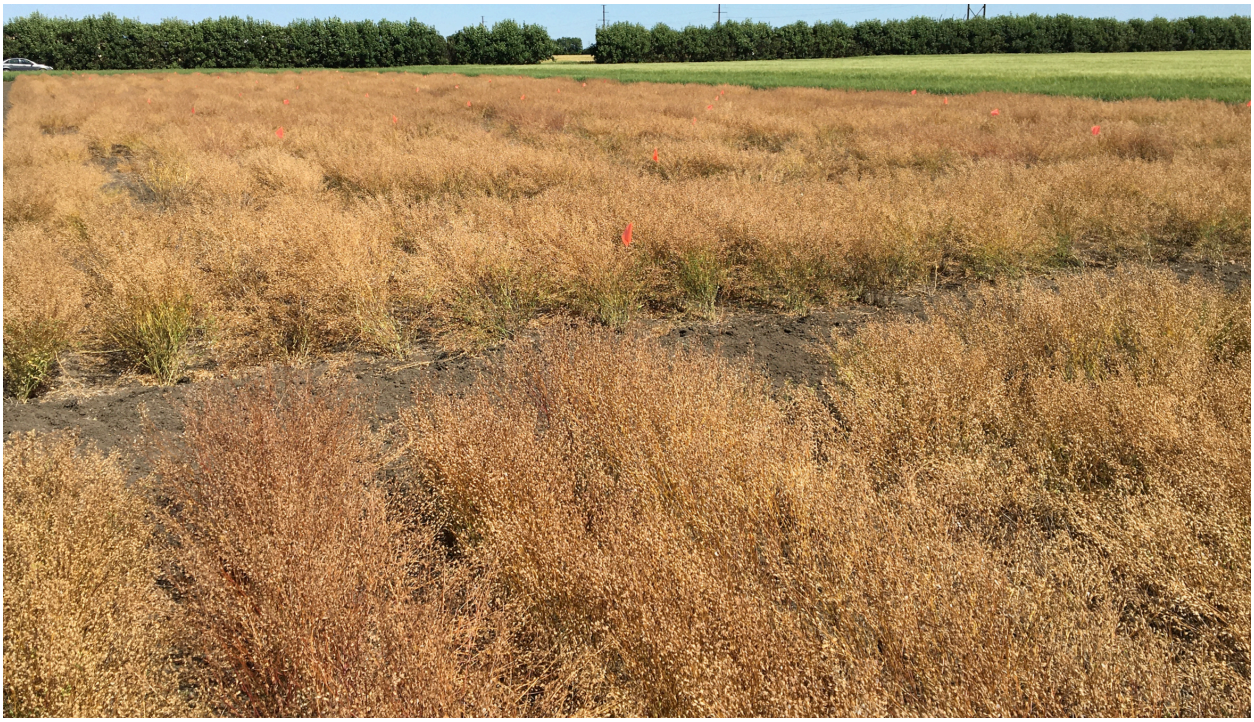


Figure 2.2 Fall seeded *Camelina sativa* $F_{3:4}$ populations in the field during 2019-20.

In *C. sativa*, plant height is positively associated with seed yield and negatively associated with 1000 seed weight (Gehringer et al. 2006). Likewise, the seed size is negatively associated with the oil content (Vollmann et al. 2005). Although oil yield per plant is positively associated with seed yield per plant, percentage oil content is negatively associated with oil yield per plant (Lošák et al. 2011) and there is no association of seed size with fatty acid content (Campbell et al. 2013). Seed size is a highly heritable characteristic which can directly influence yield, but is negatively correlated with oil content, so truncation selection (index selection) could be a possible selection method for *C. sativa* breeding, such as that applied in maize (Illinois long-term selection experiment) for generating variation in protein content.

Although *C. sativa* is considered a disease resistant crop in comparison to other *Brassica* crops (Séguin-Swartz et al. 2009), there are a number of diseases that are reported to impact yield. *Sclerotinia* stem rot caused by the fungus *Sclerotinia sclerotiorum* (Lib.) de Bary is a major disease in *C. sativa*; however, some resistant *C. sativa* genotypes have been identified suggesting that increased levels of stem lignification can reduce the spread of the fungus (Eynck et al.

2012). Other diseases commonly associated with this crop are clubroot (*Plasmodiophora brassicae*), downy mildew (*Peronospora camelinae* Gaum.), white rust (*Albugo candida*), and bacterial blight (*Pseudomonas syringae*) (Séguin-Swartz et al. 2009).

2.4 *Camelina sativa* Demography

Camelina sativa is distributed across Southwest Asia, Europe, Russia, Ukraine, as well as in North America as a naturalized weed (Chaudhary et al. 2020; Martin et al. 2017; Vollmann et al. 2005). Germplasm has been assessed for morphological variation, such as plant height, days to flowering, days to maturity, pod and seed characteristics, as well as biochemical variation, such as oil and protein profiles (Berti et al. 2011; Zubr 2003; Wiwart et al. 2019; Anderson et al. 2019; Vollmann et al. 2007). These analyses revealed a high degree of environmental plasticity among the genotypes and low levels of genetic variation.

Development of marker systems has enabled analysis of *C. sativa* germplasm at the molecular level. Vollmann et al. (2005) studied genetic diversity in 41 *C. sativa* lines using 24 RAPD primers; however, only 15 primers were found to be polymorphic, generating 30 loci and overall variation in the germplasm was found to be low. However, this marker system was not able to determine the population structure, unlike a codominant marker system. Manca et al. (2013) used 15 pairs of SSR markers to study genetic diversity in *C. sativa* where the observed heterozygosity was low (0.037) and the variation was higher between populations than within populations. In addition to this, Ghamkhar et al. (2010) analyzed the genetic diversity in 53 *C. sativa* samples with AFLP markers and suggested low levels of genetic diversity as the average Polymorphic Information Content (PIC) was only 0.24 and upon Principal Coordinate Analysis (PCoA) a maximum variation of only 7.5% was contributed by the first component. Although these methods generated polymorphisms that allowed for the differentiation of populations, they did not provide sufficient marker information for a comprehensive analysis, especially at the nucleotide base pair level, which can have a significant role in mutation, divergence and evolutionary studies.

The development of cost-effective sequencing systems has helped to identify SNPs in many crops and is widely used for population structure studies. Singh et al. (2015) used both pyrosequencing and Illumina sequencing in *C. sativa* to discover SNPs; they identified 493 SNPs, but these markers revealed low levels of diversity in 178 *C. sativa* accessions with PIC values ranging from 0.006 to 0.375 and low levels of genetic diversity (0.26). Luo et al. (2019a) also used SNPs to differentiate population structure among spring type *C. sativa* accessions, where two distinct populations were reported; however, these two populations had a very low level of genetic diversity. In the same study, a high proportion of genotypes collected from a similar geographical origin (Europe) represented an admixture of two populations, so it was difficult to infer that the population structure was due to geographical isolation.

All these findings suggested low levels of genetic diversity in current *C. sativa* populations, therefore researchers explored genetic diversity in related *Camelina* species to identify useful variation and to clarify the genetic relationship among these species. *Camelina microcarpa* was identified as a close relative of *C. sativa* and likely a progenitor species (Brock et al. 2018). The study conducted by Čalasan et al. (2019) suggested that, although *C. microcarpa* has diverse ribotypes (geographically isolated populations), *C. sativa* recently diverged (1.5 mya) from *C. microcarpa* and also supported the suggestion that *C. microcarpa* was a progenitor of *C. sativa,* with *C. sativa* as a more recently evolved species. *Camelina microcarpa* is distributed across Europe as a naturalized weed (Čalasan et al. 2019) and the diversity within its populations is somewhat higher than within populations of *C. sativa* (Chaudhary et al. 2020). Other wild relatives of *C. sativa* exist, such as *C. rumelica*, *C. hispida*, *C. laxa*; however, the genetic relationship between these species and *C. sativa* shows a high degree of divergence.

2.5 Insights into the *Camelina sativa* genome

Camelina sativa is an allohexaploid. The hexaploid nature of this crop was initially suggested by Hutcheon et al. (2010), by confirming the presence of three orthologs copies of two genes maintained as single copies in *A. thaliana*, *fatty acid desaturase (FAD2)* and *fatty acid elongase (FAE1)*. The reference genome sequence of *C. sativa* (Kagale et al. 2014a) confirmed this and revealed that *C. sativa* possesses three undifferentiated subgenomes having a total genome size

of 785 Mb. Identification of three copies of each of the 24 conserved Cruciferae genomic blocks (GBs) (Schranz et al. 2006) across the 20 chromosomes of *C. sativa* helped to determine the genome structure of *C. sativa* in comparison to the Ancestral Crucifer Karyotype (ACK) (Kagale et al. 2014a). The 24 GBs are duplicated and distributed in linear order in three homologous sets of chromosomes and are highly syntenic; however, the subgenomes (Cs-G1, Cs-G2 and Cs-G3) do not have equivalent numbers of chromosomes and have a reduced set of chromosomes (n = 6 or 7) compared to the ACK genome (Kagale et al. 2014a). Such a reduction in chromosome number from a base of n = 8 to n = 7 may be the result of loss or inactivation of one centromere through fusion of chromosomes; this can result from reciprocal translocation between an acrocentric chromosome and a metacentric chromosome (Mandakova and Lysak 2008). Based on homology of wild relatives with *C. sativa*, the initial subgenome organization has been updated for a few of the chromosomes (Chaudhary et al. 2020) with a final organization of 6, 7 and 7 chromosomes for the Cs-G1, Cs-G2 and Cs-G3 subgenomes of *C. sativa*, respectively. The 6 chromosomes of the Cs-G1 subgenome of *C. sativa* are the result of fusion of two chromosomes, yet this subgenome is highly syntenic with the Cs-G2 subgenome in terms of GB organization. In the case of the Cs-G3 subgenome, chromosome 2 and chromosome 20 show major reshuffling of GBs in comparison to the first and second subgenomes. The comparative order of GBs will play an important role in defining the possible genetic structural changes in *C. sativa*, which could have arisen as a consequence of the whole genome duplication process, and such changes can be identified and further characterized through comparison with lower ploidy progenitor species.

A comparison of the distribution of synonymous substitution rates across homologous genes from the three subgenomes with their orthologs from *A. thaliana* indicated that these subgenomes originated at a similar evolutionary time (Kagale et al. 2014a) and have retained a similar number of genes across all three subgenomes. The difference in the rate of gene loss at the sub-genomic level, referred to as genome fractionation (Murat et al. 2016), is essential for genetic novelty and is the basis for species evolution and new species formation (Liu et al. 2014). However, *C. sativa* has a comparable level of fractionated genes across its subgenomes, unlike other ancient duplicated genomes, such as *Brassica rapa,* where fractionation over time has reduced the number of genes across subgenomes (Cheng et al. 2012). This indicates recent

genome duplication in *C. sativa*; however, similar to other polyploid species, subgenome dominance has been reported (Kagale et al. 2016), which is considered a first step towards gene fractionation, suggesting fractionation is occurring during the ongoing course of evolution in this species.

It has been speculated that *C. sativa* was formed through an allopolyploidization event; however, the progenitors of *C. sativa* have not yet been identified. Identification of progenitor species is important to draw an evolutionary history of a plant and identify the genes that have played roles in shaping the current polyploid structure. These progenitors are also important in resynthesizing a polyploid to increase genetic diversity at the subgenome or species level. Initially, the lower chromosome number diploid *Camelina* species, such as *C. neglecta, C. laxa, C. hispida, C. rumelica*, were proposed to be extant progenitor species of *C. sativa*. Independent studies have shown that *C. neglecta* could be an immediate progenitor of the Cs-G1 subgenome of *C. sativa* (Mandáková and Lysak 2018; Chaudhary et al. 2020), where a high degree of similarity has been reported. Likewise, some level of similarity at the genome level has been found for *C. hispida* with the Cs-G3 subgenome of *C. sativa* suggesting that it could be an extant progenitor for the third subgenome. There are no reports of a progenitor for the Cs-G2 subgenome; however, it was speculated that a *C. neglecta*-like genome might have been this progenitor (Mandáková and Lysak 2018). Identification of tetraploid *C. microcarpa* with similarity with the Cs-G1 and Cs-G2 subgenomes has suggested step wise genome hybridization to form the hexaploid structure of *C. sativa* (Chaudhary et al. 2020). In addition, the dominance of the Cs-G3 subgenome over the other two subgenomes (Kagale et al. 2016) also supports that the Cs-G3 subgenome might have been the last to hybridize with a tetraploid in the formation of *C. sativa*. The whole genome sequence of these species will shed light on the actual structural changes before/after genome duplications.

2.6 Evolutionary significance of genome duplication

Whole genome duplication (WGD) events occurring multiple times in multiple lineages in the last 200 million years have contributed to the generation of many diverse plant species (Panchy et al. 2016). These events have been reported in almost all species, including plants (Adams and

Wendel 2005; Renny-Byfield and Wendel 2014). Genome duplication has been associated with the adaptation of plants to changes in environmental conditions during the course of evolution (Crow and Wagner 2005; Ramsey 2007). Duplication of the genome occurs either through interspecific hybridization and/or unreduced gamete formation during meiosis (Ramsey and Schemske 1998). These differences in WGD processes led to the formation of two different types of polyploid viz. allopolyploid and autopolyploid. Allopolyploids are formed from the fusion of related but distinct genomes, whereas autopolyploids are formed from the doubling of genomes with the same genomic background. These polyploids can be further categorized into three types, neopolyploid, mesopolyploid and paleopolyploid. Neopolyploid (e.g. *C. sativa*) refers to recently formed polyploids having limited differentiation of duplicated genome regions, mesopolyploid (e.g. *B. rapa*) refers to polyploids where the evidence of the WGD is obvious, yet more fragmented due to subsequent diploidization of the genomes and often biased subgenome fractionation occurs, and paleopolyploid refers to genomes where the evidence for ancient WGD is less visible due to high levels of chromosomal rearrangements and gene fractionation, such as in *A. thaliana* (Mandáková and Lysak 2018). Such whole genome duplication events can create barriers between species by limiting fertility between the new polyploid and its ancestral species, thus for the continued existence of the polyploid species by self-pollination there is a need for specific homologous pairing during meiosis, which has led to most young allopolyploid species exhibiting disomic inheritance. In contrast, a lack of stable homologous meiotic pairing hinders genome stability and affects plant fertility and fitness (Madlung et al. 2005; Tepfer et al. 2020).

Previously, WGD events were identified using karyotype studies; however, this is being replaced by genome sequencing which has become an indispensable tool and widely used to infer the composition of the genome, as well as to identify the nature of segregation in populations to identify diploidization events (Leal-Bertioli et al. 2015). Genome duplication can be easily distinguished from genome sequence information (Conant et al. 2014) and the genome sequences from several Brassicaceae species are now available including, *A. thaliana* (Kaul et al. 2000)*, B. rapa* (Wang et al. 2011), *Brassica napus* (Chalhoub et al. 2014), *Brassica oleracea* (Parkin et al. 2014), and *C. sativa* (Kagale et al. 2014a). The genome sequence also provides an idea of the fraction of genes retained by the polyploid, for example, the neopolyploid *C. sativa* contains a higher fraction of duplicated genes in each of its three subgenomes (~0.75 in all three

subgenomes) compared to the mesopolyploid *B. rapa* (0.48 in the least fractionated subgenome, 0.32 in most fractionated (MF) subgenome 1 and 0.28 in MF subgenome 2) (Kagale et al. 2014a). Reports in *Brassica* species also suggest that there are common patterns of whole genome triplication events followed by diploidization that have led to the evolution of multiple species (Lysak et al. 2005; Cheng et al. 2014).

Based on comparisons of sequence data and expression analysis in a number of plants, it has been suggested that some level of genome dominance across sub-genomes is prevalent among polyploids, which has been shown by higher expression of genes of one of the subgenomes in comparison to the others and is also reflected in a difference in the rate of gene loss (or fractionation) among subgenomes (Wang et al. 2011; Cheng et al. 2012; Chalhoub et al. 2014; Kagale et al. 2014a). Biased fractionation is mainly observed for duplicated genes with reduced expression levels (Schnable et al. 2012). Sub-functionalization, neofunctionalization and sub-neofunctionalization can result from genome and thus gene duplication (Freeling 2009; Wang et al. 2011; Tang et al. 2012). Sub-functionalization is the sharing of function by duplicate genes. In some cases, duplicate genes act additively to bring about the same level of gene expression as the original non-duplicated isoform, so there is a need to maintain both copies of the gene to maintain proper gene expression. Sub-functionalization may occur at the protein or gene expression level (Panchy et al. 2016). If the duplicate gene acquires a novel function, it is termed as neofunctionalization. If there is partitioning of function similar to sub-functionalization, as well as novel gene expression by either of the duplicated copies, this is referred to as sub-neofunctionalization (He and Zhang 2005). Understanding these phenomena is important in polyploid breeding for the selection of specific traits with a dominant expression pattern.

Realizing that polyploidization is a major event in plant evolution, there have been some attempts to synthesize polyploids by artificial interspecific hybridization, such as synthetic *Brassica napus* (Gaeta et al. 2007), and allohexaploid *Arabidopsis* (Matsushita et al. 2012) and wheat (Yang et al. 2009). These synthetic polyploids can be used to understand the nature of diploidization in polyploids through comparative study with natural polyploids; such analyses will help to reveal gene families that are under selective pressure during the evolution of the

polyploid. In *B. oleracea* and *B. rapa,* >40% of duplicate gene pairs are differentially expressed after whole genome duplication (Liu et al. 2014), which can affect particular traits. For example, in *B. rapa* there are three orthologs genes for *Flowering Locus C* (*FLC*) that have been retained after the triplication event (Wang et al. 2011). Similarly, three copies of *FLC* have also been retained in *C. sativa* (Kagale et al. 2014a); however, the function of these genes needs to be explored further to understand their exact role(s). Also, such changes in gene expression and function among the subgenomes may be associated with meiotic diploidization (Ma and Gustafson 2005). Likewise, there are also suggestions that chromosome stabilization can occur through decreases in chromosome number and additional chromosomal structural rearrangements, such as translocations and inversions (Mandáková and Lysak 2018).

2.7 Creation of variation in *Camelina* species

Genetic diversity is fundamental in plant breeding. During the course of evolution, interspecific hybridization has played an important role in creating the genetic diversity needed to domesticate plants from ancient to modern agriculture and is a continuing process. Some level of spontaneous natural mutation in plants also assists in creating variation, which undergoes natural selection for adaptive traits. With low genetic diversity found in *C. sativa* (Luo et al. 2019a), interspecific hybridization and mutation approaches can be exploited to increase genetic diversity. Besides these technologies, transgenic approaches for creating variation and gene editing are also gaining prominence. Attempts were made for interspecific hybridization of *C. sativa* with a close relative *C. microcarpa*, suggesting the potential to create diverse *Camelina* germplasm (Tepfer et al. 2020; Martin et al. 2019). Similarly, intergeneric and intertribal hybridization were also attempted (**Table 2.1**); however, pre-fertilization and post-fertilization barriers played a role in preventing successful outcomes in these wide hybridizations. Pre-fertilization barriers might be due to failure of pollen germination, abnormal growth of pollen tubes, or lack of pollen tube penetration of the ovule, whereas post-fertilization barriers may arise from degeneration of the endosperm, male and female sterility in the hybrid plants, or lethality in the hybrid progeny (Kuligowska et al. 2015). But reports did suggest that *C. sativa* is highly interfertile with *C. alyssum* and *C. microcarpa* (Séguin-Swartz et al. 2013), which all share similar genomic organizations.

In addition to conventional hybridization, protoplast fusion with other *Brassica* species has provided some hybridity, with the promise to improve some morphological traits, such as *Alternaria* blight resistance and modified fatty acid profiles (Narasimhulu et al. 1994; Hansen 1998; Jiang et al. 2009). Likewise, mutational approaches have also been used to improve herbicide resistance (Walsh et al. 2012a) and linolenic acid content (Büchsenschütz-Nothdurft et al. 1998). A number of transgenic approaches have been successful for modifying fatty acid composition in this crop (Lu and Kang 2008) (Ruiz-Lopez et al. 2015). Similarly, Betancor et al. (2015) also developed transgenic *Camelina* enhanced with eicosapentaenoic acid, which promotes growth in fish. Most fascinatingly, gene editing technology (CRISPR/Cas9) has also been applied in this crop where mutation of *FAD2* gene copies was carried out to increase oleic acid from 16% to 50% at the expense of other polyunsaturated fatty acids (Jiang et al. 2017). Similarly, modification of oil composition targeting *FAE1* (Ozseyhan et al. 2018) and seed protein profiles targeting *CRUCIFERIN C* (Lyzenga et al. 2019) have also been carried out to improve the seed composition in this crop.

Table 2.1 Different techniques of variability creation implemented in *C. sativa.*

Technique	**Description**	**Results**	**Reference**
Intergeneric crossing	Protoplast fusion of *B. carinata* and *C. sativa* by electrofusion	6.8% heterokaryons obtained. Plants failed to establish.	Narasimhulu et al. (1994)
Intergeneric crossing	Protoplast fusion of *B. oleracea* and *C. sativa* by PEG application	0.5% of shoots regenerated. Stand establishment weak. No plants survived	Hansen (1998)
Intergeneric crossing	Protoplast fusion of *B. napus* and *C. sativa* by electrofusion	6.5% of shoots regenerated. Intermediate fatty acid profile.	Jiang et al. (2009)
	C. alyssum × *C. sativa*	1.4 seeds per pollination.	

Interspecific hybridization	*C. microcarpa* × *C. sativa*	2.2 seeds per pollination. Backcross (*C. sativa* × F1) produced 0.24 seed per pollination.	Séguin-Swartz et al. (2013)
	C. rumelica ssp. *rumelica* × *C. sativa*	0.58 seeds per pollination.	
	C. sativa × *C. alyssum*	2.2 seeds per pollination	
	C. sativa × *C. microcarpa*	1.19 seed per pollination.	
	C. sativa × *C. rumelica* ssp. *rumelica*	1.02 seeds per pollination.	
Intertribal crossing	*A. thaliana* × *C. sativa*	No seed obtained.	Julié-Galau et al. (2014)
	Cardamine hirsuta × *C. sativa*	Few seeds recovered with intermediate morphology.	
	Capsella bursa-pastoris × *C. sativa*	Few seeds recovered with intermediate morphology.	
Intraspecific Hybridization	*C. sativa* × *C. sativa*	23% of the pollination were successful with 7.7 seeds per pod	
Intertribal crossing	*Capsella bursa-pastoris* × *C. sativa*	1.5 hybrids with intermediate phenotype) per 10000 pollinations. No self or backcross seed obtained	Martin et al. (2015)
Intraspecific hybridization	78 pairwise crosses with 13 *C. sativa* lines	27% of crosses with significant mid-parent heterosis and 13% with significant best-parent heterosis.	Zelt and Schoen (2016)

Mutagenesis	EMS	Allele conferring resistance to acetolactate synthase inhibitors developed by single base substitution.	Walsh et al. (2012a)
Mutagenesis	EMS	Variation in linolenic acid levels.	Büchsenschütz-Nothdurft et al. (1998)
Microspore Culture	NLN media with 12.5% sucrose and 12.5% PEG 4000.	38 embryos derived from 100,000 microspores.	Ferrie and Bethune (2011)
Transgenic Approach	*Agrobacterium*-mediated transfer of castor *Fatty Acid Hydroxylase* (*FAH12*) gene	1.3% transformation rate.	Lu and Kang (2008)
Transgenic Approach	Use of C18 Δ9 elongase to generate C20 PUFAs	Accumulation of up to 26.4% omega-3 eicosatetraenoic acid.	Ruiz-Lopez et al. (2015)
Transgenic Approach	Transformation of 5 microalgal genes into *C. sativa*	Accumulation of significant amounts of eicosapentaenoic acid.	Betancor et al. (2015)
CRISPR/Cas9 gene editing	Mutation in three copies of the *FAD2* gene	Increase in oleic content from 10% to 62% at the T_3 generation.	Morineau et al. (2017)
CRISPR/Cas9 gene editing	Mutation in *FAD2* gene	Increase in oleic content from 16% to 50% at the expense of other polyunsaturated fatty acids at the T_4 generation.	Jiang et al. (2017)

Broadly looking at the genetic diversity in *Camelina* germplasm and the hybridization success among species, *C. microcarpa* could have potential to improve the modern *C. sativa* genome through increasing genetic diversity. Variation in ploidy level in *C. microcarpa* has complicated the success of hybridization attempts (Martin et al. 2019), where only the hexaploid is interfertile with *C. sativa*. Diploid *C. neglecta,* previously classified as *C. microcarpa*, seems to have difficulty hybridizing with *C. sativa*, although its genome resembles the first subgenome of *C. sativa.* Likewise, hybridization of *C. microcarpa* with *C. sativa* produced a number of suspected aneuploid or meiotic abnormalities (Tepfer et al. 2020). Similarly, Zhang and Auer (2020) reported reduced pollen viability and reduced seed set in interspecific hybridizations performed between *C. sativa* and *C. microcarpa.* These studies have raised questions about the genetic organization of *C. microcarpa*, some of which were resolved by Chaudhary et al. (2020) (see Chapter 3). However, the results of Tepfer et al. (2020) are encouraging in terms of the variability generated, as they showed segregation of genes associated with fatty acid content such that in the F_3 population there was increase in the level of C18:2 and a decrease in C18:3 for a few lines, which can be exploited in *C. sativa* breeding programs. These reports suggest a need for careful selection and study of wild relatives for successful *Camelina* hybridization.

Prologue to chapter 3

This chapter focused on assessing genetic diversity in *C. sativa* with the aim to identify novel sources of genetic variation from species related to *C. sativa*, as well as from *C. sativa* accessions collected from Ukraine and Russia that were not previously examined. In addition, the findings from the thesis work corrected the subgenome assignment in hexaploid *C. sativa,* which is important in evolutionary studies, as well as in developing breeding tools for *Camelina* species.

This chapter has been published as:

Chaudhary, R., C.S. Koh, S. Kagale, L. Tang, S.W. Wu, Z. Lv, A.S. Mason, A.G. Sharpe, A. Diederichsen, and I.A.P. Parkin, 2020 Assessing Diversity in the *Camelina* Genus Provides Insights into the Genome Structure of *Camelina sativa*. *G3: Genes, Genomes, Genetics* 10 (4):1297-1308. https://doi.org/10.1534/g3.119.400957

For this article, Raju Chaudhary performed sample collection, flow cytometry analysis, GBS library preparation, GBS analysis, genetic diversity analysis, STRUCTURE analysis and manuscript writing; Chu Shin Koh helped in pipeline development for GBS analysis and circos plot design; Sateesh Kagale helped in subgenome dominance analysis; Lily Tang helped in data collection and GBS library preparation; Siu Wah Wu helped in flow cytometry analysis, Zhenling Lv and Annaliese S. Mason performed chromosome counts; Axel Diederichsen helped in sample collection; Andrew G. Sharpe and Isobel A.P. Parkin conceived of the project concept and supervised the work.

Chapter 3. Assessing diversity in the *Camelina* genus provides insights into the genome structure of *Camelina sativa*

3.1 Abstract

Camelina sativa (L.) Crantz an oilseed crop of the Brassicaceae family is gaining attention due to its potential as a source of high value oil for food, feed or fuel. The hexaploid domesticated *C. sativa* has limited genetic diversity, encouraging the exploration of related species for novel allelic variation for traits of interest. The current study utilized genotyping by sequencing to characterise 193 *Camelina* accessions belonging to seven different species collected primarily from the Ukrainian-Russian region and Eastern Europe. Population analyses among *Camelina* accessions with a 2n = 40 karyotype identified three subpopulations, two composed of domesticated *C. sativa* and one of *C. microcarpa* species. Winter type *Camelina* lines were identified as admixtures of *C. sativa* and *C. microcarpa.* Eighteen genotypes of related *C. microcarpa* unexpectedly shared only two subgenomes with *C. sativa*, suggesting a novel or cryptic sub-species of *C. microcarpa* with 19 haploid chromosomes. One *C. microcarpa* accession (2n = 26) was found to comprise the first two subgenomes of *C. sativa* suggesting a tetraploid structure. The defined chromosome series among *C. microcarpa* germplasm, including the newly designated *C. neglecta* diploid née *C. microcarpa*, suggested an evolutionary trajectory for the formation of the *C. sativa* hexaploid genome and re-defined the underlying subgenome structure of the reference genome.

Keywords: Camelina, Domestication, cryptic species, Reference genome, Subgenome, related species

3.2 Introduction

Camelina sativa (L.) Crantz is an ancient oilseed of the Brassicaceae family, that contributed to the human diet from the Bronze to the Middle Ages (Hjelmqvist 1979; Hovsepyan and Willcox 2008; Larsson 2013) before losing favour to higher yielding relatives. More recently it has shown potential to become a low-input high value oil crop for the food and feed industry (Faure and Tepfer 2016). Several advantages of this species have been reported (Brown et al. 2016; Ye et al. 2016) including the ability to yield well on dry and marginal lands and its unique seed

quality traits (Gugel and Falk 2006), particularly its balanced omega fatty acids (Simopoulos 2002). However, improvements can be made to the crop such as increasing seed size for improved harvestability and reducing the glucosinolate content, which is an anti-nutritional in animal feed (Schuster and Friedt 1998; Amyot et al. 2018). Biologically, *Camelina* species have two crop habits, annual spring and biennial winter types (Berti et al. 2016). Most of the domesticated *C. sativa* are spring type, whereas the majority of its wild relatives are winter type. Genetic diversity is vital for developing a robust breeding strategy to identify and incorporate the necessary variation for further crop improvement. Thus far, different molecular approaches have been explored to study a range of *Camelina* germplasm including, RAPD (Vollmann et al. 2005), AFLP (Ghamkhar et al. 2010), SSR (Manca et al. 2013), and SNP marker analyses (Singh et al. 2015); all the studies concluded that there were low levels of genetic diversity available within spring type *C. sativa* compared to other oilseed crop species.

The genus *Camelina* has been reported in the literature to contain anywhere from 6 to 11 species, suggesting some taxonomic confusion (Warwick and Al-Shehbaz 2006; Brock et al. 2019). Latterly there appear to be between six and seven commonly accepted species belonging to the genus which range in chromosome number and ploidy level; namely *C. sativa* (2n= 6x = 40), *Camelina microcarpa* Andrz. ex DC. (2n = 12, 2n= 4x = 26, 2n = 6x = 40) (Martin et al. 2017), *Camelina hispida* (Boiss.) Hedge (2n = 2x = 14), *Camelina rumelica* Velen. (2n = 4x = 26), *Camelina neglecta* (2n = 2x = 12) (Brock et al. 2019) and *Camelina laxa* C.A. Mey. (2n = 2x= 12) (Galasso et al. 2015). The seventh species *Camelina alyssum* is more contentious since current accessions available within genebanks appear indistinguishable from and are inter-fertile with *C. sativa*; therefore, it was suggested that *C. alyssum* is a synonym of *C. sativa*, although this has yet to be adopted by genebanks (Al-Shehbaz 1987; Martin et al. 2017). Although there was a well-documented chromosome series for *C. microcarpa* until recently there were no reported sub-species; however, Brock et al. (2019) suggested that the smallest *C. microcarpa* karyotype (2n = 12) should be re-classified as a new species, *Camelina neglecta*. Currently cultivated *C. sativa* is considered to be hexaploid with 20 chromosomes in a haploid set, while at least one of the related species (e.g. *C. microcarpa*) has the same chromosome number (Francis and Warwick 2009) most have lower numbers. The genome sequence of *C. sativa* suggested a neopolyploid that had evolved from three lower chromosome number species, specifically one n

= 6 and two n = 7 species (Kagale et al. 2014a). *Camelina* species such as *C. neglecta*, *C. laxa* and *C. hispida* possess the same haploid chromosome numbers as subgenomes of the hexaploid and recent work has proposed that *C. neglecta* and *C. hispida* could indeed be extant progenitors of *C. sativa* (Mandáková et al. 2019). The study of these lower ploidy species could be instrumental in defining the relationship among the species as well as uncovering the polyploidization history of *Camelina* (Brock et al. 2019). Defining the relationships between these species at the subgenome level may also help to identify those species that are potential novel sources of allelic variation for introgression into *C. sativa*.

Camelina microcarpa has been of interest in studies of *Camelina* diversity as it is believed to be the closest extant relative to domesticated *C. sativa* and could help in understanding the domestication process in *Camelina* species, as well as providing novel variation (Brock et al. 2018). The collections of *C. microcarpa* species in different genebanks suggest that it has a diverse range of origin including the Mediterranean region, Armenia (Brock et al. 2018), Germany, Poland, Czechia, Slovakia and Georgia (Smejkal 1971; Martin et al. 2017). Diversity studies, analyses of genome size and chromosome number along with the success of hybridization efforts between *C. microcarpa* and *C. sativa* (Séguin-Swartz et al. 2013; Martin et al. 2019) suggested the close relationship between these two species (Brock et al. 2018; Martin et al. 2017). However, not all the results were so encouraging with varying levels of hybridization success depending on the genotype (Séguin-Swartz et al. 2013). These results were likely due to confusion with the classification of *C. microcarpa* accessions, either due to disparities in chromosome number and/or crosses being attempted with completely different species such as *C. neglecta* (Brock et al. 2019; Martin et al. 2017). Such anomalies could have led to an assumption of higher diversity within *C. microcarpa* species, with the discovery of *C. neglecta* in particular there is a need to better understand the relationship between the different accessions of *C. microcarpa* and *C. sativa* for potential utilization of such germplasm in *Camelina* breeding programs.

Estimation of genome size using flow cytometry and chromosome counts are common tools to infer ploidy in a species (Johnston et al. 2005; Brock et al. 2018; Séguin-Swartz et al. 2013).

Complementary genomic tools can assist in clearly defining evolutionary relationships between species and in the case of *Camelina*, the available reference genome for *C. sativa* can facilitate such analyses (Kagale et al. 2014a). Here, we explored genetic diversity using predominantly genotyping by sequencing (GBS) in different *Camelina* species, with a focus on *C. microcarpa*. The analyses of these related species suggested a group of *C. microcarpa* lines could represent a novel cryptic species. In addition, the subgenome structure of the *C. sativa* reference genome was re-defined and will provide a basis for utilization of the related species in *C. sativa* breeding. For example, this study identified a range of potentially valuable minor alleles from *C. microcarpa*, including those in three flowering related genes which may have impacted the *Camelina* domestication process.

3.3 Materials and methods

3.3.1 Plant materials

This study included a collection of 160 *C. sativa*, 27 *C. microcarpa*, two *C. alyssum,* one *C. neglecta,* one *C. laxa,* one *C. hispida and* two *C. rumelica* to establish the genetic relationship among the accessions (**Table S1**). The accessions were mainly obtained from Plant Genetic Resources of Canada in Saskatoon (http://pgrc3.agr.gc.ca/). One accession, "Midas^{TM}", was a commercial Canadian variety and 12 accessions were commercial varieties from the United States and Europe. Five accessions are breeding lines from the Agriculture and Agri-Food Canada Saskatoon Research and Development Centre (provided by Dr. Christina Eynck) and the remainder of the lines were thought to originate from eastern Europe and the Russian-Ukraine region and were donated from the National Centre for Plant Genetic Resources of Ukraine in Kharkiv.

3.3.2 Flow cytometry analysis

The relative genome sizes of six different *Camelina* species were measured using flow cytometry according to the method described in Garcia et al. (2004) (**Table 3.1**). Approximately 1 cm^2 of leaf tissue of both sample and an internal standard was placed in a plastic petri dish with 2 ml of Galbraith buffer (Galbraith et al. 1983), the mixture was chopped up with a razor blade and the solution was supplemented with 200 µg of ribonuclease A, before being filtered through a filter

with a pore size of 30 µm. Propidium iodide was then added at a concentration of 60 µg/ml. The stained solution was kept at 4 ºC for 2 hr and allowed to incubate at room temperature for an hour before taking measurements. DNA content of the nuclei from each species was estimated using fluorescence measurements with a green laser (532 nm) in a CyFlow Space Flow Cytometer (Partec). *Camelina sativa* (TMP23992) having known ploidy level and genome size (Kagale et al. 2014a; Martin et al. 2017) was used as an internal standard to estimate the genome size of lower ploidy species. For all accessions three biological replicates were used.

3.3.3 Chromosome counts

For this study, seeds from six accessions (*C. sativa* TMP23992, *C. neglecta* PI650135, *C. hispida* PI650133, *C. microcarpa* CN119243, *C. microcarpa* TMP24026 and *C. microcarpa* TMP23999) were germinated on moist filter paper in Petri dishes at room temperature. Chromosome counts were carried out based on the protocols detailed in Harrison and Heslop-Harrison (1995) and Snowdon et al. (1997); Harrison and Heslop-Harrison (1995) with minor modifications. Growing root tips (1-2 cm) were collected into tubes containing 0.04% 8-hydroxyquinoline solution (290 mg 8- hydroxyquinoline powder dissolved in 1 L H_2O via treatment at 60 °C for 2 hours, then stored at -4 °C until use). The root-tip-containing solution was incubated in the dark for 2 hours at room temperature followed by incubation at 4 °C for 2 hours. Cells were fixed with Carnoy's I solution (3 parts ethanol to 1 part glacial acetic acid) for 2 days at room temperature. After fixation the root tips were stored in 70% ethanol at -20 °C. The fixed root tips were rinsed twice for 10 minutes with distilled water to remove the fixative and incubated in 0.1 M pH 4.5 citrate solution (1.47 g trisodium citrate-dihydrate ($Na_3C_6H_5O_7.2H_2O$) and 1.05 g citric acid monohydrate ($C_6H_8O_7.H_2O$) in 500 mL water) for 15 minutes at room temperature followed by incubation in enzyme solution (0.25 g (5%) Onozuka R-10 cellulase and 0.05 g (1%) pectinase in 5 mL citrate solution) for another 30-40 minutes at 37 °C. Root tips were washed with distilled water for 30 minutes and placed onto a slide with a few drops of Carnoy's I solution. On the slide, the root tissue was scrambled with a pin and left until the solution dried. Finally, a drop of DAPI staining solution VECTASHIELD® Antifade Mounting Medium with DAPI (4,6-diamidino-2-phenylindole; product number H-1200 from Vector Laboratories) was added and covered with a coverslip before observing under UV fluorescence using a Leica DRME microscope at 1000 × magnification.

3.3.4 DNA extraction

Immature leaf samples were collected for DNA extraction. Leaf tissue was stored at -80 °C prior to DNA extraction. All the samples were freeze-dried for at least 48 hrs before lysis. DNA extractions were performed using a CTAB method (2% CTAB, 100mM Tris-HCl, 20mM EDTA, 1.4M NaCl) (Murray and Thompson 1980). After DNA extraction, samples were treated with RNase at 37 °C to remove RNA contamination. Quantification of DNA was performed with Quant-iT™ PicoGreen® dsDNA Assay Kit (ThermoFisher Scientific) through fluorescence measured (485nm/535nm, 0.1s) using the Victor *X*Plate Reader (PerkinElmer).

3.3.5 Library preparation and DNA sequencing

Genotyping was performed by an established GBS method (Poland et al. 2012). After DNA normalization (20 ng/ul), 200 ng of DNA were digested with *PstI* and *MspI* at 37 °C for 2 hours. Next, adapters were ligated to the restriction digested DNA fragments using T4 DNA ligase at 22 °C for 2 hours. The products were inactivated before multiplexing and 96 samples were pooled into a single library. After pooling, the library was amplified with a short extension time (30 sec) and purified using a QIAquick PCR Purification Kit (Qiagen). The final libraries were quantified using a Bioanalyzer (Agilent Technologies) to confirm the fragment size and quality of the library. Sequencing of 35 *C. sativa*, 9 *C. microcarpa*, 1 *C. rumelica* and one *C. alyssum* were completed on an Illumina HiScan SQ module (paired-end 100 bp reads) and the remainder were sequenced on an Illumina HiSeq2500 platform (paired-end 125 bp reads).

3.3.6 DNA sequence analysis

An existing pipeline was used to demultiplex the reads and trim the reads for adapters, short reads and poor quality data using Trimmomatic (Bolger et al. 2014). Leading and trailing bases with quality below 15 and reads shorter than 55 bp were removed prior to mapping to the reference genome. The trimmed sequence reads were aligned with the reference genome of hexaploid *C. sativa* (Kagale et al. 2014a) using Bowtie2 (Langmead and Salzberg 2012). In bowtie2 mapping, *--local* with *-sensitive* parameters were used with –score-min of L,0,0.8. In addition, a custom perl script was used to extract the single best unique hits. Obtained binary files (BAM) were used for variant calling as well as mapping sequence distribution. BEDTools (Quinlan and Hall 2010) was used to extract mapped reads and calculate the frequency of

mapped reads along 100 Kb bins in the genome. Circos (Krzywinski et al. 2009) was used to plot the distribution of mapped reads along the *C. sativa* reference genome for the diploid, tetraploid and hexaploid *Camelina* genotypes. *UnifiedGenotyper* with standard parameters from the Genome Analysis Toolkit (McKenna et al. 2010) was used to call SNPs.

3.3.7 Population differentiation

Obtained SNPs were analyzed for average dissimilarity between genotypes and Principal Coordinate Analysis (PCoA) was performed utilizing AveDissR(Yang and Fu 2017) Package (Yang and Fu 2017) in the R program (R Core Team, 2017). Population structure was determined using Bayesian technique in STRUCTURE (Pritchard et al. 2000) with a burn-in period of 150,000 steps and 150,000 MCMC replicates where parallelization was performed with StrAuto tool(Chhatre and Emerson 2017b) (Chhatre and Emerson 2017a). To determine optimal K, three replications were run with each value of K from 1 to 10. The value of K was converted into LnP(K) to obtain the plateau of ΔK. The optimal K was determined using the online version of "Structure harvester" (Earl 2012). PowerMarker (Liu and Muse 2005) was used to calculate gene diversity, Polymorphic Information Content (PIC) and Nei's (1983) based genetic distance between the genotypes. MEGA 7 (Kumar et al. 2016) was used to construct the Neighbor Joining (NJ) tree among the genotypes. The phylogenetic tree was confirmed through the use of the maximum likelihood method (Tamura and Nei 1993) in MEGA 7 using bootstrap consensus tree (Felsenstein 1985) inferred from 1000 replicates, no significant differences were noted between the alternate tree structures (**Figure S5**). Analysis of Molecular Variance (AMOVA) and pairwise F_{ST} were calculated using GeneAlEx 6.5 (Peakall and Smouse 2006, 2012).

3.3.8 Subgenome dominance

Data previously published by Kagale et al. (2016) was re-analysed. The expression data from 12 tissues of *C. sativa* were arranged according to the re-defined subgenome structure and filtered for expression less than 0.01 TPM for all replicates. The 12 tissues were Germinating Seed (GS), Cotyledon (C), Young leaf (YL), Root (R), Stem (S), Senescing leaf (SL), Bud (BUD), Flower (F), Early seed development (ESD), Early mid seed development (EMSD), Late mid seed development (LMSD) and Late seed development (LSD). Filtering provided data for a range of expressed triplicated genes, from 9149 in LSD to 12634 triplets in Root (**Table S10**), which were

analysed for subgenome dominance in *C. sativa*. The analysis was performed using analysis of variance techniques where effects due to replication were kept as random. Genes that were expressed significantly (*P-value* <0.05) higher in any subgenome compared to the other two were considered dominant.

3.4 Results

3.4.1 Identification of ploidy series among *Camelina* species

GBS was performed for 193 *Camelina* accessions, high-quality sequence reads were aligned to the reference genome of *C. sativa*, DH55 (Kagale et al. 2014a). The number of reads per line and alignment rate is summarized in **Table S2.** As expected, consistent read coverage was found across all 20 linkage groups of the reference genome for all accessions of *C. sativa* and *C. alyssum*. However, for particular *Camelina* accessions the results showed biased read mapping across the reference linkage groups (**Figure 3.1, Table S2, Figure S6**). In particular the *C. neglecta* accession (PI650135) aligned significantly to six chromosomes; whereas, *C. microcarpa* accessions aligned to either thirteen or 20 chromosomes. For a proportion of the *C. microcarpa* lines showing read alignment to thirteen chromosomes it was observed that the read depth was somewhat higher for six of those chromosomes, which represented the first of the three sub-genomes of the *C. sativa* hexaploid (**Table S2**). In light of the observed bias in read mapping, flow cytometry and chromosome counts were performed to measure the relative size of the nuclear genome content as well as to infer the ploidy level for a subset of the different *Camelina* accessions (**Table 3.1, Figure 3.2, Figure S1**). *Camelina sativa* (TMP23992) a well-characterised hexaploid with a genome size estimated to be 1.50 pg/2C (Martin et al. 2017) was used as an internal standard to measure the absolute genome size of lower ploidy *Camelina* species.

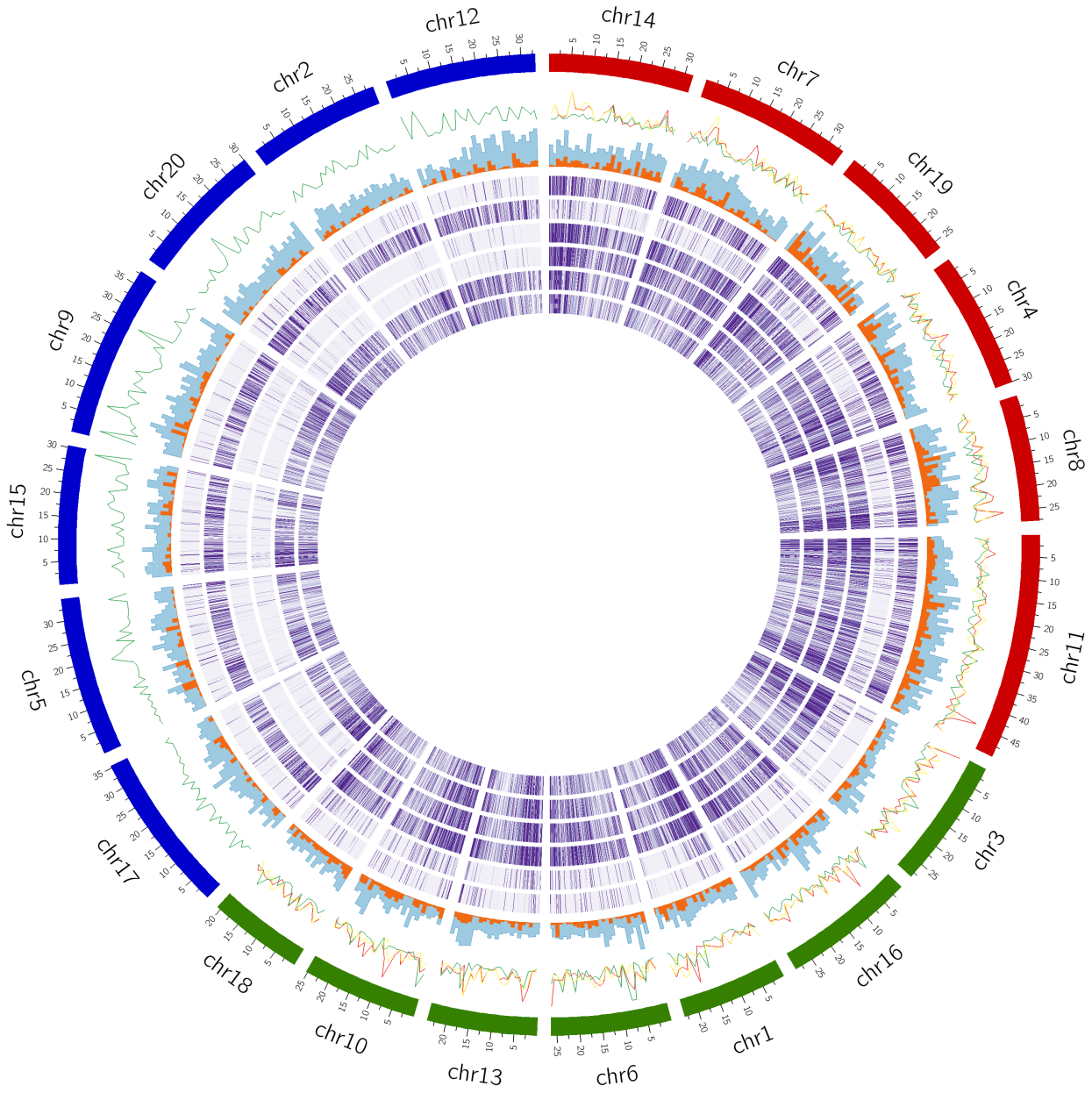


Figure 3.1 Identification of ploidy in *Camelina* species using genotyping by sequencing (GBS) data. From outer to inner track: 1) Clockwise three subgenomes of *C. sativa* reference genome in red, green and blue; 2) F_{ST} distribution across the genome: *C. sativa* vs *C. microcarpa* "Type 1" in green, *C. sativa* vs *C. microcarpa* "Type 2" in red and *C. microcarpa* "Type 1" vs *C. microcarpa* "Type 2" in yellow; 3) SNP distribution of *Camelina* species in 1 Mb bins in blue and filtered SNPs in orange; 4-9) Heat maps showing read alignment of diploid genotype *C. neglecta* (PI650135), *C. hispida* (PI650133), tetraploid *C. microcarpa* (CN119243), *C. microcarpa* "Type 2" (TMP23999), *C. microcarpa* "Type 1" (TMP26172) and *C. sativa* (TMP23992) to the reference genome.

For the known diploid *C. neglecta* (2n = 12) genotype (PI650135) (previously *C. microcarpa*) the GBS data mapped to only six chromosomes thus correlated well with the expected results. This line also had the lowest genome size (0.43 pg/2C) in comparison to *C. sativa* (1.50 pg/2C). Also as expected the diploid species, *C. hispida* was found to have 2n = 14 chromosomes with a relatively similar genome size of 0.59 pg/2C as of diploid *C. neglecta*. For the *C. hispida* GBS reads, there was a significant bias in mapping with just over 57% of the reads mapped to the third subgenome of the reference *C. sativa* genome (**Figure 3.1, Figure S6**). This might indicate an affinity of *C. hispida* with the third subgenome of reference *C. sativa* (Mandáková et al. 2019).

Table 3.1 Genome size estimation of different *Camelina* species using flow cytometry.

Species	Accession	2C DNA (pg)	Ploidy
C. neglecta	PI650135	0.43±0.01	2x
C. hispida	PI650133	0.59±0.02	2x
C. microcarpa "4x"	CN119243	0.95±0.02	4x
C. rumelica	TMP24027	1.26±0.02	4x
C. microcarpa "Type 2"	TMP23999	1.49±0.03	6x
C. sativa	TMP23992	1.50±0.03	6x

More interestingly, of the *C. microcarpa* lines where the GBS data aligned with 13 linkage groups from the reference genome, only one genotype (CN119243) possessed a lower genome size (0.95 pg/2C) in comparison to the hexaploids, and based on the read alignments as well as chromosome counts was inferred to be tetraploid (2n = 26) (**Figure 3.1** and **3.2**). Seven genotypes from *C. microcarpa* (hereafter referred to as "Type 1") showed consistent read coverage across all chromosomes from the reference genome of *C. sativa*, while GBS data from 18 *C. microcarpa* genotypes (hereafter referred to as "Type 2") aligned with only 13 linkage groups but with a somewhat higher read coverage in the first subgenome (**Table S2**). *Camelina microcarpa* (TMP24026*),* representing the "Type 1" group, had 2n = 40 chromosomes, as

expected. However, *C. microcarpa* (TMP23999), representing the "Type 2" group, had an estimated DNA content (1.49 pg/2C) similar to that of *C. sativa* yet was found to have 38-40 chromosomes, most likely 2n=38 (**Figure 3.2**). Estimates for this latter line were slightly confounded by the large variation in size between chromosomes and are hence presented with reasonable but not 100% certainty. Sub-genome 1 of *C. sativa*, with only six chromosomes possesses a larger "fusion" chromosome (Csa-11), it would seem likely that the unidentified six chromosome sub-genome of Type 2 *C. microcarpa* has a similar "fusion" chromosome which would interfere with accurate chromosome counts; see **Figure 3.3a**.

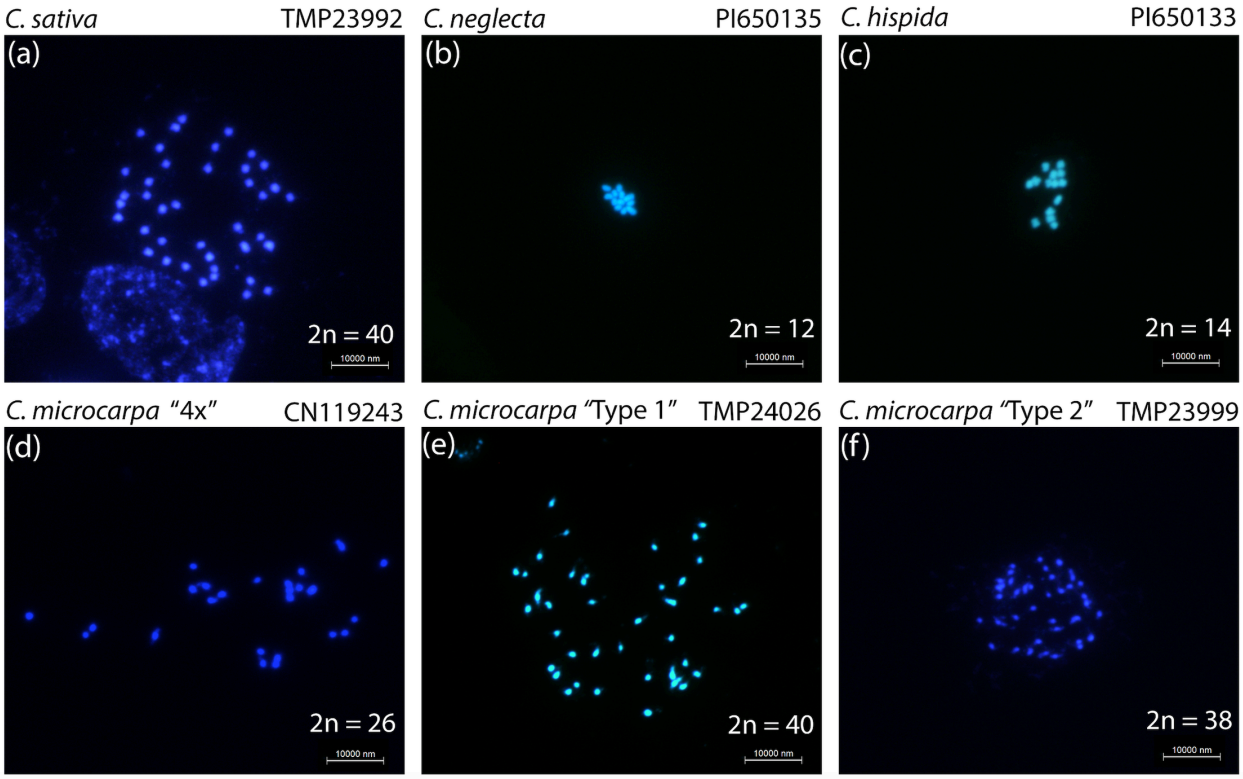


Figure 3.2 Chromosome counts for different *Camelina* species. a) *C. sativa* TMP23992 (2n = 40); b) *C. neglecta* PI650135 (2n = 12); c) *C. hispida* PI650133 (2n = 14); d) *C. microcarpa* "4x" CN119243 (2n = 26); e) *Camelina microcarpa* "Type 1" TMP24026 (2n = 40); and f) *C. microcarpa* "Type 2" TMP23999 (2n = 38).

Of the 13 chromosomes showing read alignment for the *C. microcarpa* "Type 2" group, six chromosomes were shared with the diploid species *C. neglecta* and seven with subgenome 2 of *C. sativa,* while the apparently missing chromosomes comprise subgenome 3, to which reads

from the diploid *C. hispida* also align. These results suggested two different types of higher chromosome number *C. microcarpa* accessions (Type 1: 2n = 40 and Type 2: 2n=38) with similar genome sizes; one which shares the genome organization as that of the reference *C. sativa* genome and the second which shares only two subgenomes with that of the reference. Thus, representatives of diploid, tetraploid and two different hexaploid *Camelina* "species" could be differentiated. The tetraploid *C. rumelica* (TMP24027) (Martin et al. 2017), previously suggested as a progenitor of *C. sativa* (Mandáková et al. 2019), had a higher nuclear genome content (1.26 pg/2C) than the tetraploid *C. microcarpa* (CN119243; 2n = 26). The read alignment data of *C. rumelica* mapped to all chromosomes with no observable pattern; this ambiguity with regards to its relationship to the subgenomes of *C. sativa* would not be expected if *C. rumelica* was indeed a progenitor genome (**Table S2**, **Figure S6**). Further accessions of this line would need to be tested.

3.4.2 A refined subgenome structure for *C. sativa*

The increase in ploidy level in *Camelina* species from 2n = 12 in *C. neglecta* to 2n = 26 and 2n = 40 in *C. microcarpa* might be expected to correspond to the three subgenomes of *C. sativa* as defined in the reference genome (Kagale et al. 2014a)*;* however, this was not the case. The original assignment of reference pseudo-molecules to each of the subgenomes used synteny analyses to identify the most parsimonious route, minimizing genome-restructuring events, from the ancestral karyotype of the Brassicaceae to the modern day *C. sativa* genome (Kagale et al. 2014a). However, it was recognized at the time that some linkage groups, for example Csa14 and Csa03, shared the same basic chromosome structure and their subgenome assignment was more difficult. Thus, based on the GBS read alignments and the assumption that the simplest path to the hexaploid genome is through the hybridization of identified lower chromosome number species the subgenome structure has been refined. More explicitly it was assumed that *C. neglecta* is an extant relative of subgenome 1, the tetraploid *C. microcarpa* CN119243 represents the second stage in the evolutionary path and is composed of subgenome 1 and 2, and finally hexaploid *C. microcarpa* (2n = 40) is a direct ascendant of *C. sativa*, comprised of all three subgenomes; where the origin of the third subgenome is still unclear, although likely a relative of *C. hispida*. Thus the new genome organisation is as follows Subgenome 1 (SG1) contains Csa14, Csa07, Csa19, Csa04, Csa08 and Csa11, which are shared with the diploid *C. neglecta* (formerly

C. microcarpa); SG2 is composed of Csa03, Csa16, Csa01, Csa06, Csa13, Csa10 and Csa18 that along with SG1 are in common with the tetraploid *C. microcarpa* CN119243; and finally SG3 that is found in all *C. sativa* lines consists of Csa17, Csa05, Csa15, Csa09, Csa20, Csa02 and Csa12, which are also shared with *C. hispida* (**Figure 3.1**, **Figure 3.3a**). As shown in **Figure 3.3a** the majority of the re-assignments were between SG1 and SG2, with four chromosomes changing in each instance, only two chromosomes from SG3 were re-assigned. There was no suggestion of chromosomal rearrangements, although this will have to be confirmed through either genetic mapping and/or genome sequencing of the lower ploidy species. It was noted that one scaffold assigned to SG3 was found to have a high read depth when reads were aligned from *C. microcarpa* "Type 2", which was an anomaly in the mapping pattern and could indicate a miss-assembly, which again will need to be confirmed through sequencing. The refined subgenome organization was used for all subsequent analyses.

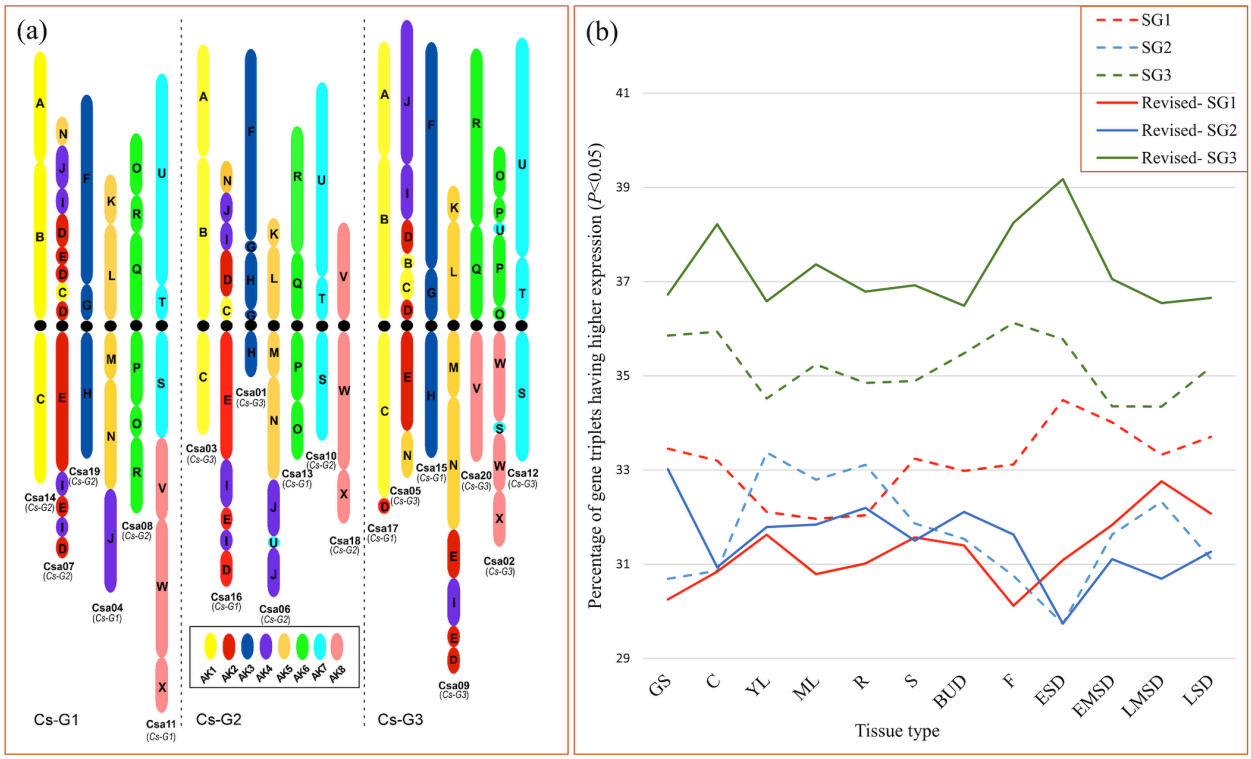


Figure 3.3 Re-defining the *Camelina sativa* subgenome composition. a) Newly defined subgenome architecture of *C. sativa;* b) Evidence of genome dominance based on refined subgenome structure and gene expression data (GS: Germinating Seed, C: Cotyledon, YL: Young Leaf, ML: Senescing Leaf, R: Root, S: Stem, BUD: Bud, F: Flower, ESD: Early Seed Development; EMSD: Early Mid Seed Development, LMSD: Late Mid Seed Development and LSD: Late Seed Development).

3.4.3 Population differentiation in *Camelina* species

Depending upon the distribution of the read alignments against the reference genome and corroborated by the chromosome counts and nuclear DNA content, only one genotype each belonged to *C. neglecta*, tetraploid *C. microcarpa*, *C. hispida* and *C. laxa*; two genotypes were classified as *C. rumelica*, and two as *C. alyssum;* seven genotypes were hexaploid *C. microcarpa* with 20 chromosomes, while, 18 genotypes belonged to *C. microcarpa* "Type 2" with putatively 19 chromosomes and a novel hexaploid structure compared to the *C. sativa* reference genome (e.g. TMP23999); the remaining 160 genotypes were classified as *C. sativa* with 20 chromosomes (**Table S1**).

Prior to filtering, variant calling in all 193 genotypes yielded 102,744 SNPs across the *C. sativa* reference genome where a significant proportion of SNPs were from the related species (**Table S3**). Due to the presence of these distant relatives and the presumption of novel alleles being captured, raw SNPs were filtered for a minor allele frequency of greater than 1% among all samples and after allowing varying levels of missing data points (**Figure S2**), SNPs with 20% of the genotypes with missing data were selected, providing 4803 variants including indels for all the *Camelina* species studied (**Figure 3.1**). These SNPs were further filtered for indels yielding 4268 SNPs which were used to study population structure and genetic diversity in *Camelina* species.

The SNP distribution across the subgenomes reflected the genome composition of the total collection of accessions; with the first subgenome having a greater number of SNPs in comparison to the second and third; and the third subgenome having the lowest number of SNPs (**Table 3.2**). Gene diversity was found to be low for all chromosomes, similarly the PIC values were low; however, the range for these parameters was high across all chromosomes (**Table 3.2**). These results were somewhat skewed due to the genotypes from *C. microcarpa* "Type 2" and other related species which led to lower coverage in the third subgenome therefore an independent analysis was performed with the 169 genotypes with the same 20 chromosomes as that of the reference genome (**Table S4**). Removing the related *Camelina* species reduced the overall number of SNPs but also filtered out less polymorphic loci leading to higher average gene diversity and average PIC values for each of the chromosomes. Likewise, the analysis

among the genotypes of domesticated *C. sativa* species (162 genotypes) including *C. alyssum* and *C. sativa* ssp. *pilosa* suggested an overall gene diversity of 0.181 and PIC value of 0.15 (**Table S5**).

Table 3.2 Genetic diversity parameters for 193 *Camelina* genotypes belonging to 8 species. The numbers in parenthesis indicate range.

Subgenome	Chromosome	Total SNP	Filtered SNP	Gene Diversity	PIC
SGI	Chr14	5754	263	0.117 (0.021-0.499)	0.103 (0.020-0.375)
	Chr7	6280	235	0.130 (0.021-0.499)	0.114 (0.021-0.374)
	Chr19	5209	298	0.111 (0.021-0.500)	0.098 (0.020-0.375)
	Chr4	5462	271	0.127 (0.021-0.500)	0.111 (0.021-0.375)
	Chr8	5535	309	0.101 (0.021-0.500)	0.091 (0.020-0.375)
	Chr11	9593	550	0.120 (0.021-0.500)	0.105 (0.021-0.410)
	Subtotal	37833	1926	0.118 (0.021-0.500)	0.104 (0.020-0.410)
SGII	Chr3	3642	166	0.117 (0.021-0.498)	0.102 (0.021-0.374)
	Chr16	4333	207	0.135 (0.021-0.500)	0.118 (0.021-0.375)
	Chr1	3406	195	0.112 (0.021-0.495)	0.101 (0.020-0.372)
	Chr6	3477	153	0.146 (0.021-0.500)	0.126 (0.021-0.375)
	Chr13	3337	146	0.110 (0.021-0.499)	0.097 (0.021-0.375)
	Chr10	3614	208	0.119 (0.021-0.500)	0.104 (0.021-0.375)
	Chr18	2740	167	0.111 (0.021-0.495)	0.099 (0.021-0.373)
	Subtotal	24549	1242	0.122 (0.021-0.498)	0.107 (0.021-0.374)
SGIII	Chr17	5200	139	0.102 (0.021-0.397)	0.094 (0.021-0.318)
	Chr5	4993	156	0.137 (0.021-0.500)	0.120 (0.021-0.375)
	Chr15	4726	152	0.082 (0.021-0.406)	0.075 (0.021-0.324)
	Chr9	6603	186	0.084 (0.022-0.499)	0.076 (0.022-0.374)
	Chr20	5031	105	0.089 (0.021-0.494)	0.079 (0.021-0.372)
	Chr2	4451	122	0.099 (0.021-0.498)	0.089 (0.021-0.374)
	Chr12	6450	188	0.106 (0.021-0.494)	0.093 (0.021-0.372)
	Subtotal	37454	1048	0.100 (0.021-0.470)	0.089 (0.021-0.359)
Scaffolds		2908	52		
Total SNPs		**102744**	**4268**	**0.114 (0.020-0.500)**	**0.101 (0.000-0.410)**

Principal coordinate analysis (PCoA) differentiated the related species from the *C. sativa* population including *C. alyssum* and *C. sativa* ssp. *pilosa* (**Figure 3.4**). The first coordinate explains 24.27% of the variation, which differentiated *C. sativa* from other *Camelina* relatives; the second coordinate explains 7.24% of variation, which differentiated more distant relatives

such as *C. rumelica*, *C. laxa* and *C. hispida* from *C. sativa* and *C. microcarpa*. The PCoA result suggested that *C. alyssum* followed by *C. microcarpa* "Type 1" genotypes were quite similar to domesticated *C. sativa*, while *C. microcarpa* "Type 2", *C. hispida, C. laxa* and *C. rumelica* species were clearly divergent. This analysis mainly differentiated between species; however, separate analysis of *Camelina* species with 20 chromosomes was used to differentiate among *C. sativa* genotypes, and to suggest some sub-population structure (**Figure S3**).

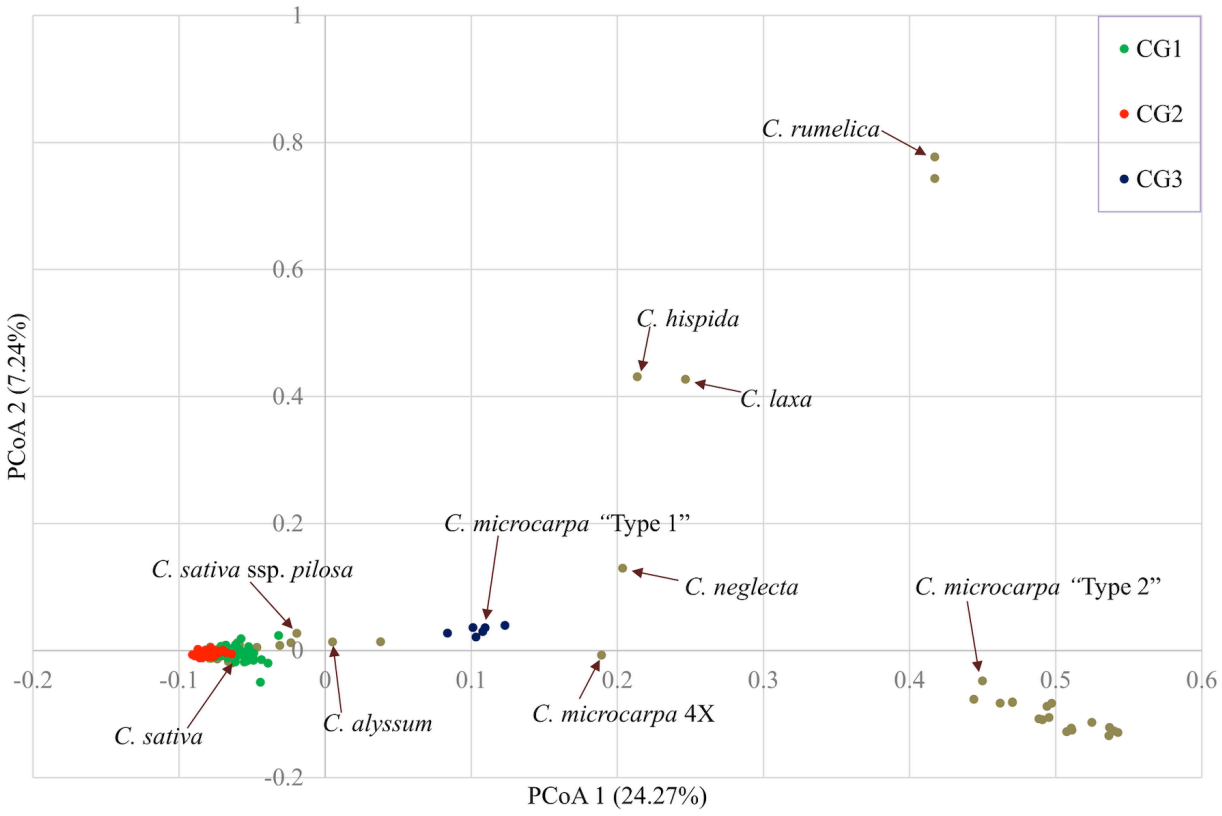


Figure 3.4 Principal coordinate analysis of 193 *Camelina* genotypes based on 4268 SNPs. The different colours represent three subpopulations defined by the STRUCTURE analysis.

The results from the PCoA were mirrored in the generation of a Neighbor Joining (NJ) tree showing the phylogenetic relationships among the 193 *Camelina* genotypes (**Figure 3.5**). All the domesticated *Camelina* genotypes were closely related to each other, forming a separate large cluster. The NJ tree showed that the related species, which all share a vernalisation requirement, were clustered next to a number of *Camelina* lines which were winter types, including *C. alyssum* (CAM176), *C. sativa* ssp. *pilosa* (CN113692) and the line Joelle (North Dakota State

University) (**Figure 3.5**). Tetraploid *C. microcarpa* CN119243 formed a separate cluster and was basal to the *C. sativa* sub-populations, the diploid *C. neglecta* (PI650135) was basal to all higher chromosome number accessions. One *C. microcarpa* genotype (TMP26168) had a very similar genomic organization as the reference genome; however, was categorized as *C. microcarpa* "Type 1" and formed a separate single cluster. *Camelina microcarpa* "Type 2" species formed their own separate cluster, but showed further sub-population structure, separating into two groups with 11 and 7 genotypes, respectively. Two genotypes belonging to *C. rumelica* formed a separate cluster along with *C. laxa* and *C. hispida* and suggesting these had diverged sometime earlier from the progenitors of domesticated *Camelina* species.

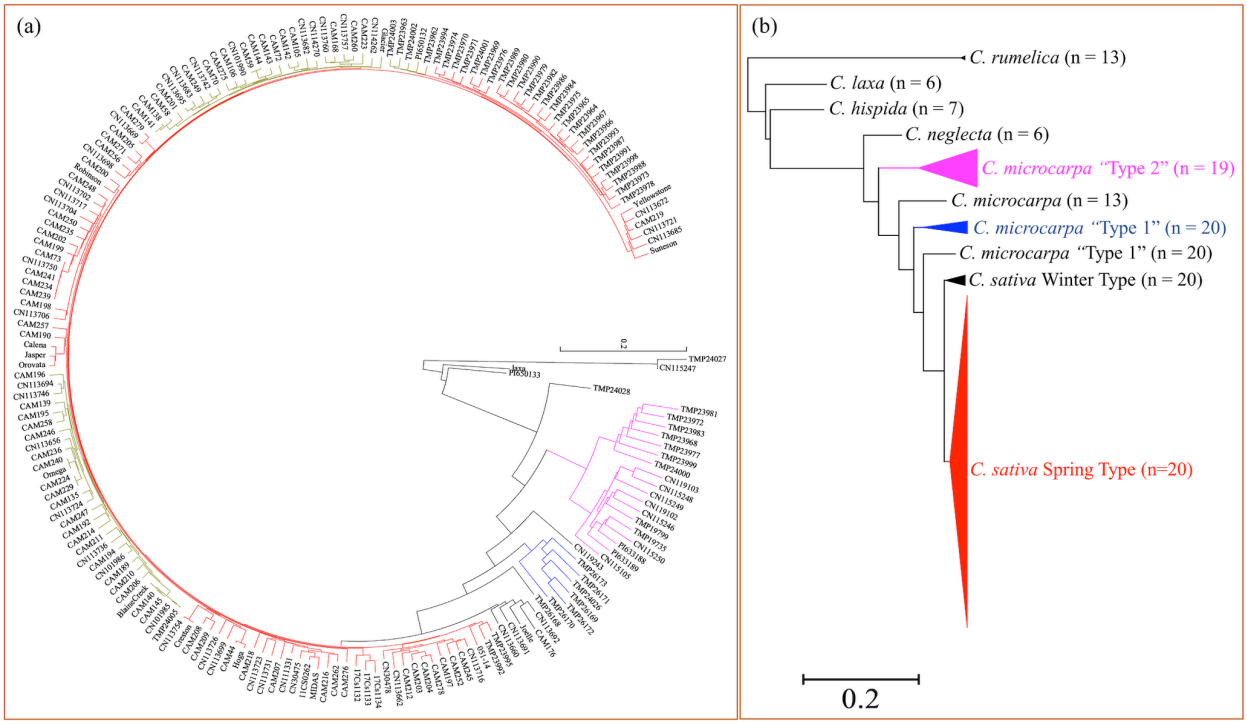


Figure 3.5 Genetic relationship among *Camelina* accessions as determined by NJ tree construction based on 4268 SNPs. a) Relationship among 193 *Camelina accessions;* b) Summary of the relationship among different species of *Camelina* (number in parenthesis indicate number of chromosomes in a haploid set).

The PCoA and NJ suggested some sub-structure among the domesticated *C. sativa* accessions, which was further assessed using the Bayesian clustering approach of STRUCTURE (Pritchard et al. 2000). This analysis was performed with the hexaploid *Camelina* accessions with 20

chromosomes only (n=169) and suggested two populations confirming the separation of *C. microcarpa* "Type 1" accessions from *C. sativa*. The peak of delta K also suggested further population differentiation at K=3, which identified two sub-populations among the *C. sativa* accessions. Assuming this three population structure and, based on a Q value cut-off of 70%, 124 genotypes were clustered into three subpopulations with 45 genotypes found to be an admixture of these subpopulations (**Table S6, Figure S4**). As shown in **Figure 3.6**, 162 *Camelina* genotypes were found in two sub-populations CG1 (red), CG2 (green) and *C. microcarpa* "Type 1" formed subpopulation CG3 (blue). The genotypes belonging to CG1 and CG2 were spring type whereas the genotypes belonging to CG3 were winter type. One genotype (TMP26168) belonging to *C. microcarpa* "Type 1" was found to be an admixture of CG3, CG2 and CG1, which confirmed its unique status, noted in the NJ tree analyses. The winter type *C. alyssum* (CAM176) was also an admixture of CG1, CG2 and CG3, with a higher contribution from subpopulation CG1. Other winter types such as *C. sativa* ssp. *pilosa* (CN113692) and *C. sativa* (Joelle) were grouped with CG1. All the winter type *Camelina* lines were found to have a contribution of alleles from subpopulation CG3, representing *C. microcarpa* "Type 1" (**Table S6**).

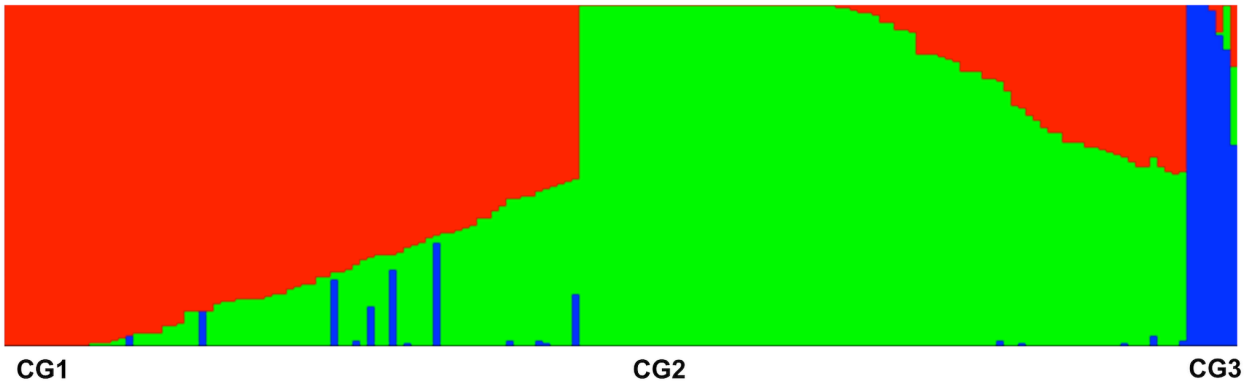


Figure 3.6 Population structure of *Camelina* species. CG1 (Red) and CG2 (Green) represent *C. sativa* genotypes, and CG3 (Blue) represents *C. microcarpa* "Type 1".

Pairwise F_{ST} values were calculated among the three subpopulations (124 genotypes), excluding the lines showing admixture. The results suggested that spring type *Camelina* species of subpopulations CG1 and CG2 were closely related with an F_{ST} of 0.065. F_{ST} values between the two spring *Camelina* sub-populations and *C. microcarpa* "Type 1" indicated greater

differentiation between the species, with values of 0.302 and 0.349, respectively (**Table 3.3**). However, a separate analysis of pairwise F_{ST} with all the genotypes irrespective of admixture suggested a lower F_{ST} value (0.263) (**Table S7d**). For all the subpopulation the third subgenome showed higher differentiation among subpopulations in comparison to the other subgenomes (**Table S7**). The F_{ST} analysis between *C. sativa* and *C. microcarpa* "Type 1" also suggested strong selection for alleles in *C. sativa* on chromosome Csa06 in a relatively small region (6Mb to 9 Mb region) (**Figure 1**).

Table 3.3 Pairwise F_{ST} among three subpopulations of *Camelina* species. CG1 (58 genotypes) and CG2 (60 genotypes) represent *C. sativa* genotypes and CG3 (6 genotypes) represents *C. microcarpa* "Type 1" accessions.

	CG1	**CG2**	**CG3**
CG1	0.000		
CG2	0.065	0.000	
CG3	0.302	0.349	0.000

3.4.4 Related *Camelina* species as a reservoir of minor alleles

Although, this study included a number of species, approximately 96% of the total samples were either classified as *C. sativa*, *C. microcarpa* "Type 1" or *C. microcarpa* "Type 2". Among the 4268 filtered SNPs, the number of minor alleles (less than 5% homozygous) were identified for each of the three species, to assess their potential as a source of novel alleles. Such minor alleles were found for 2300 SNPs; only 33 were shared by all three species (**Figure 3.7).** Of the minor alleles, 1111 were unique to *C. microcarpa* "Type 2", 433 were unique to *C. microcarpa* "Type 1" and 355 were unique to *C. sativa* species. The distribution of minor alleles along the subgenomes suggested the first subgenome of both *C. sativa* and *C. microcarpa* "Type 2" contained the highest number of minor alleles, while the third subgenome for *C. microcarpa* "Type 1" contained more minor alleles (**Table S8**).

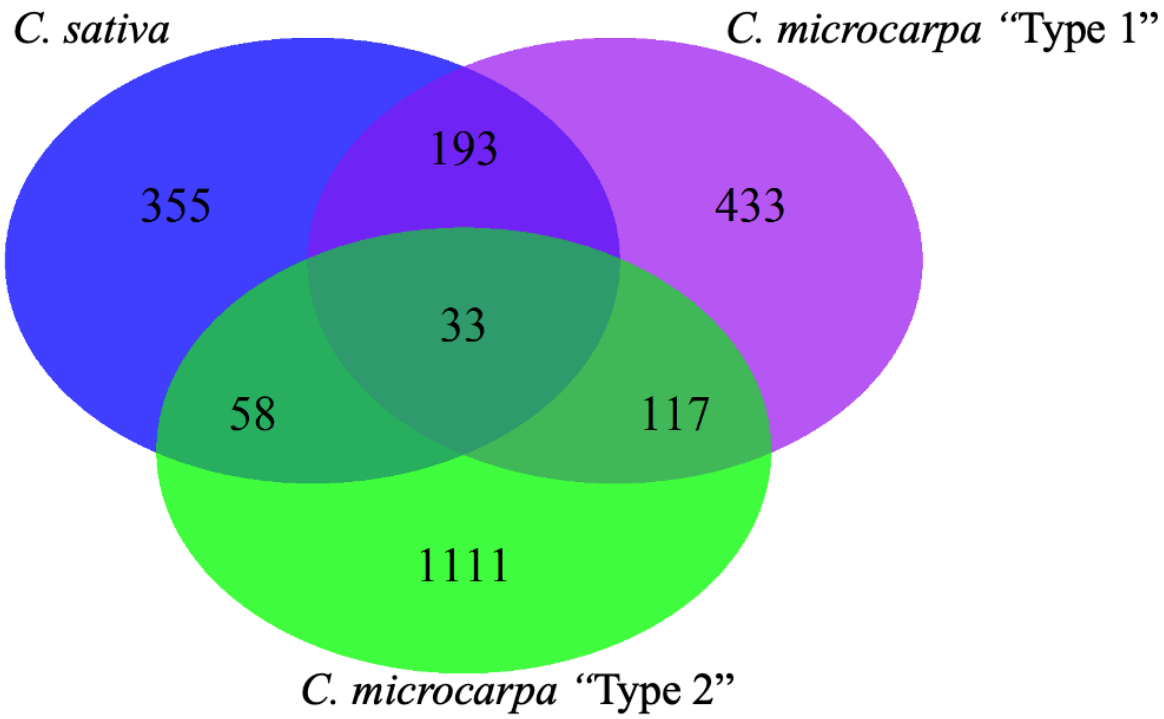


Figure 3.7 Venn diagram showing distribution of minor alleles in different species of *Camelina*.

Minor alleles not present in the domesticated *C. sativa* were explored to identify mutations that may have helped to shape the existing *C. sativa* accessions through selection for changes to particular genes. Of all the SNPs with minor alleles 536 were within the genic region of 355 genes. Of these, 275 genes had orthologs in *Arabidopsis thaliana* (**Table S8a**), although there was no apparent bias for particular functional category, three genes were found to have an influence on flowering time and photoperiod response and could be interesting candidates for manipulating phenology (**Table S8b**).

3.5 Discussion

The current study exploited GBS data and the reference genome of *C. sativa* to characterize variation among *Camelina* species, which not only identified a potentially novel *Camelina* species but also suggested refinements to the underlying subgenome structure of *C. sativa*. The hexaploid structure of *C. sativa* was clear from the genome assembly of Kagale et al. (2014a); however, the differentiation of the three subgenomes was complicated by the high degree of synteny between particular chromosomes. Phylogenetic analyses of a set of unanchored genome scaffolds of *C. neglecta* (PI650135) (Toro 2017) also suggested changes to the first subgenome of *C. sativa* genome, which concurred with the GBS data presented in this study. By alignment of GBS data from the diploid *C. neglecta* (2n = 12), a presumed tetraploid (*C. microcarpa*; 2n = 26) and multiple hexaploids (2n = 40) a step-wise hybridization path to the current *C. sativa* genome was suggested, implicating the diploid and tetraploid line as potential progenitor species of *C. sativa*. The third subgenome shares significant homology to *C. hispida,* implying this may represent an extant progenitor of the final subgenome, which is in agreement with the recent work of Mandáková et al. (2019).

After redefining the subgenome composition of *C. sativa*, there was a slight change in distribution of gene coverage, with a higher number of genes now present on the third subgenome (33.7% compared to 32.7% of total annotated genes) and a slight decrease in the number of genes for the second subgenome (30.2% compared to 31.1% of total genes) (**Table S9**). Although there was no change in number of genes retained in triplicate, in light of the re-definition of the karyotype, subgenome dominance was re-analysed based on the previously published gene expression data from Kagale et al. (2016). Depending on the tissue type between 9,188 (late seed development) and 12,688 (root) triplicated orthologs gene sets were analysed for evidence of genome dominance in *C. sativa* (**Table S10**). As found in Kagale et al. (2016) the results suggest dominance of the third subgenome over the other two; however, the impact was far more pronounced (**Figure 3.3b**). For all tissue types, the third subgenome had a greater number of genes with higher expression in comparison to both the first and second subgenome, deviating from a hypothetical 1:1:1 ratio of number of genes significantly expressing higher in any one subgenome (χ^2 test, *P-value*>0.05). There were some tissue specific patterns observed

with regards to SG1 and SG2: the second subgenome was found to dominate the first subgenome until flowering, after which the first subgenome dominated the second. However, the ratio of the total number of expressed genes for the third subgenome with either first or second subgenome was not particularly high (~1.11-1.27), suggesting limited gene silencing, and might reflect the young neopolyploid status of *Camelina* as suggested by Kagale et al. (Kagale et al. 2014a). The marked dominance of the third subgenome, or by inference the genome added last in the stepwise evolution of *C. sativa*, is in concordance with evidence from other polyploid species with similar evolutionary trajectories (Ramírez-González et al. 2018; Edger et al. 2019; Mandáková et al. 2019).

The chromosome numbers for *C. neglecta, C. hispida, C. sativa and C. microcarpa* "Type 1" were consistent with previous reports (Martin et al. 2017; Brock et al. 2018). However, *C. microcarpa* "Type 2" was suggested to have n = 19 chromosomes, noticeably the sequences from this genome mapped to only two of the *C. sativa* subgenomes, suggesting a hexaploid derived from progenitors with 6, 7 and 6 chromosomes. The available tetraploid (n = 13) which could be a progenitor of both "Type 1" and "Type 2" *C. microcarpa* suggests two different routes to the formation of the higher ploidy hexaploid genomes in the *Camelina* genus. The mapping of *C. hispida* (n = 7) to the third subgenome of *C. sativa* (**Figure 3.1**), also indicated by the results of (Mandáková et al. 2019) could suggest hybridization of the tetraploid with *C. hispida* in the formation of modern hexaploid *C. sativa.* As yet, the origin of the third subgenome for *C. microcarpa* "Type 2" remains elusive, although it shares some homology with subgenome 1, suggesting it could be a relative of *C. neglecta*. The current study did not find clear association of the tetraploid *C. rumelica* with specific subgenomes of the reference *C. sativa,* suggesting that greater genetic distance and possibly chromosomal rearrangement separate the two species (Čalasan et al. 2019).

The genetic characterization of the accessions confirmed the low level of differentiation among *C. sativa* lines (Vollmann et al. 2005; Singh et al. 2015; Luo et al. 2019a; Gehringer et al. 2006), yet there was some indication of sub-structure within the *C. sativa* population. A significant number of recently- collected accessions, which originated from the Russian/Ukraine border

populated CG1 and could provide a source of some limited variation in *C. sativa* breeding, but the related hexaploid species offer the potential of much more diversity. It appears that some of this variation may have begun to be captured, in particular with the generation of *C. sativa* types with a vernalisation requirement. Similarly, it was noted that one apparent *C. microcarpa* "Type 1" line showed evidence of shared alleles across the three defined sub-populations, including those seemingly specific to *C. sativa*. The evolutionary history of *Camelina* hexaploids may have played a role in limiting variation with a smaller number of SNPs found in the second subgenome, which may reflect a small number of hybridization events from which this subgenome was derived. Although *C. sativa* and *C. microcarpa* both evolved through polyploidy, *C. microcarpa* "Type 1" has maintained a greater collection of minor alleles, implicating the influence of selection on a crop which has been subjected to less intensive breeding than most, or again could result from a polyploidization bottleneck. The frequency of minor alleles was higher in the first subgenome of domesticated *C. sativa* in comparison to *C. microcarpa* "Type 1" (**Table S8**) and might indicate further differentiation of *C. sativa* subpopulations or relate to age of divergence of the subgenomes. The study of minor allele frequencies has been used to understand domestication and potential bottlenecks created during the process, enabling the identification of genes under selection that may underlie QTL controlling traits of interest (Ross-Ibarra et al. 2007). The current study identified a number of genes carrying minor alleles in the wild relative that may represent genes under selection in the crop, further comprehensive sequence analyses and trait association will determine the value of such variation.

Data Availability

Supplemental data (Tables S1-S10; Figures S1-S6) can be found at https://doi.org/10.25387/g3.11299280. The raw sequence data has been deposited at NCBI under the BioProject ID: PRJNA602698 (http://www.ncbi.nlm.nih.gov/bioproject/602698).

Prologue to chapter 4

After assessing the genetic diversity in the available *Camelina* germplasm, this study was designed to develop a segregating population to examine the phenotypic variability introduced from intraspecific hybridization and to facilitate breeding, specifically for the winter habit. Winter-type genotypes, such as C. *sativa* ssp. *pilosa* and *C. alyssum*, were distinct within the clade that included most of the *C. sativa* lines and were expected to produce suitable variation in the segregating population. Therefore, crosses were performed between *C. sativa* collected from Kaliningradskaya Oblast, Russia and two winter-type lines, *C. sativa* ssp. *pilosa* and *C. alyssum*. This study enabled QTL for flowering behaviour in winter-type *C. sativa* to be mapped. The identified QTL were the first found for *C. sativa* flowering behaviour. In addition, the genetic maps generated in this study will be used to map other traits of interest.

This chapter will be submitted to Molecular Breeding.

Chapter 4. Mapping QTL for vernalization requirement identified adaptive divergence of candidate gene *Flowering Locus C* in *Camelina sativa*

4.1 Abstract

Manipulating flowering behaviour is important for the improvement of plant architecture, adaptation, and yield. Vernalization requirement is an integral component of flowering in winter-type plants. The availability of winter ecotypes among *Camelina* species has facilitated the mapping of QTL for vernalization requirement in *C. sativa*. An intraspecific crossing scheme between related *Camelina* species, where two different sources of the winter-type habit were used, resulted in the development of two segregating populations. Linkage maps generated with GBS-based markers identified three QTL associated with vernalization requirement in *C. sativa*. All three QTL were found in proximity to the *FLOWERING LOCUS C* (*FLC*) gene, variants of which have been reported to affect the vernalization requirement in plants. However, the three loci were mapped to different homologous regions of the hexaploid *C. sativa* subgenomes in the two populations. Transcriptome analysis between winter-type *C. alyssum* and spring-type *C. sativa* confirmed as expected higher expression of *FLC* in the former, since *FLC* would be expected to supress floral initiation. However, the *FLC* gene on chromosome 8 showed higher expression in the spring-type parent relative to the *C. sativa* ssp. *pilosa* parent. A second gene, *G-box regulating factor 6,* in the QTL region may influence the photoperiod responses. The presence of three *FLC* QTL could suggest adaptive divergence of duplicate gene copies in *C. sativa,* which needs to be further explored. The three identified QTL provide opportunities for manipulating vernalization requirement in this young crop.

Keywords: *Flowering Locus C,* intraspecific hybridization**,** QTL, vernalization, winter-type *Camelina*

4.2 Background

The evolutionary path to *C. sativa* is believed to have created a narrow genetic bottleneck leading to low genetic diversity in spring *Camelina* germplasm, (Singh et al. 2015; Vollmann et al. 2005; Gehringer et al. 2006; Luo et al. 2019a) that has hindered the *Camelina* improvement program. In addition, hybridization of this crop with other species has had limited success

(Martin et al. 2015; Séguin-Swartz et al. 2013; Narasimhulu et al. 1994; Hansen 1998; Jiang et al. 2009; Julié-Galau et al. 2014). Interspecific hybridization was successful between *C. sativa* and *C. microcarpa* and produced plants of intermediate phenology, although there were low levels of pollen viability and reduced fitness in the hybrids (Martin et al. 2019). The use of wide crosses can be an important tool to increase the genetic diversity in a crop, as well as to identify QTL and associated candidate genes. However, challenges in recombination can exist due to several factors, such as asynchronised flowering behaviour, fertility issues and fundamental differences in the number of chromosomes between species (Chapter 3). Identification of wild relatives (Brock et al. 2018; Martin et al. 2017), which are closely related to the domesticated *C. sativa* have encouraged their use in *C. sativa* breeding. The extent of relatedness among the *Camelina* species almost certainly plays a role in the success of hybridization. As might be expected, *C. sativa* sub-species, such as *C. sativa* ssp. *pilosa* (DC.) N.W. Zinger, and the closely related *C. alyssum* (Mill.) Thell. show higher success in hybridization attempts relative to wild relatives, such as *C. microcarpa* (Séguin-Swartz et al. 2013; Martin et al. 2019).

Plants with winter growth habits usually require vernalization, exposure to a short period of low but non-freezing temperatures, to transition from the vegetative stage to the reproductive stage, often referred to as bolting. *Camelina sativa* is generally an annual species, but among close relatives a few, such as *C. sativa* ssp. *pilosa* and *C. alyssum* (also suggested to be a sub-species or even a synonym of *C. sativa*), have been characterized with a biennial growth habit (Galasso et al. 2015), yet they share the same number of chromosomes as hexaploid *C. sativa* (Chapter 3). A recent study of winter- and spring-types of *C. sativa* compared leaf morphology, growth behaviour and seed characteristics (Wittenberg et al. 2019), where marked reduction in leaf number, plant height and plant growth before vernalization were reported for winter-types. Winter-type *C. sativa* is hardy to adverse winter conditions and displays good crop establishment with higher yields than spring-types (Gesch et al. 2018), and is a suitable candidate for crop rotation on the Northern Great Plains (Berti et al. 2017). A number of experiments have reported a higher variation for the yield and fatty acid composition in *C. sativa* grown in different environmental conditions, as reviewed in Vollmann and Eynck (2015), where the benefits of higher linolenic acid, early flowering and avoidance of a number of biotic and abiotic factors were some of the noted added advantages of winter-type *C. sativa.* Also, a lower level of erucic

acid, an anti-nutritional compound, has been reported in winter-type *C. sativa* compared to spring-type *C. sativa* (Kurasiak-popowska et al. 2020). Thus far, there has been limited exploration of winter-type *C. sativa* germplasm that can survive prolonged harsh winters with similar yields as current spring-types.

A number of genes were identified as being responsible for the vernalization requirement in *A. thaliana* and other related *Brassica* species. Among them, *FLC*, a well-characterized gene, has been shown to control winter-type behaviour in *A. thaliana* (Michaels and Amasino 1999; Swiezewski et al. 2009). Orthologs of *FLC* have been reported in a number of *Brassica* species to affect vernalization requirement (Takada et al. 2019; Schiessl et al. 2019; Anderson et al. 2018), where higher expression of *FLC* suppresses bolting before vernalization. As such, the duration of vernalization is inversely correlated with the level of *FLC* expression over the course of the vernalization period (Sheldon et al. 2000), and *FLC* acts as a repressor for a number of genes associated with flowering responses (Deng et al. 2011). *Camelina sativa* is a hexaploid with three relatively undifferentiated subgenomes and syntenic analyses suggested the existence of three copies of *FLC* (Kagale et al. 2014a). One ortholog of *FLC* on chromosome 20 (*Csa20g015400*) was found to be differentially expressed in response to vernalization in the winter-type *C. sativa* variety Joelle in comparison to spring-type *C. sativa* (Anderson et al. 2018). This was confirmed by Chao et al. (2019) with an additional set of winter-type *C. sativa* lines, where expression differences for *FLC* on chromosome 20 could differentiate the two biotypes. It was speculated that the additional *FLC* orthologs might have succumbed to selection pressure that resulted in a change or loss of function and they may now have a role in seed and/or tissue development (Anderson et al. 2018). Similarly, potential sub-functionalization of the *FLC* orthologs/homoeologues has been reported in some *Brassica* species (Schiessl et al. 2019). However, other studies have suggested that the additional *FLC* genes are responsible for variation in flowering time in the absence of vernalization requirement (Zou et al. 2012; O'Neill et al. 2019; Xiao et al. 2013).

Various methods have been developed to detect QTL associated with a particular trait, among them, GWAS based QTL identification has become popular to capture variation present in

diverse populations where a number of allelic variants associated with a trait can be identified. However, it can be difficult to manage a large population, in particular, phenotyping can be cumbersome. The development of a biparental population to identify QTL is an establised approach where only prior knowledge for a quantitative difference in a trait of interest among parents is required. With advancements in sequencing technologies, the time and cost associated with marker generation has been reduced (Hall 2013). Likewise, availability of the *C. sativa* reference genome (Kagale et al. 2014a) offers the potential to identify candidate genes controlling traits of interest (King et al. 2019; Luo et al. 2019b). In this context, GBS is a valuable technique to generate genetic information at low cost (Poland et al. 2012) and can be used to create genetic linkage maps and further the mapping of traits of interest (Young and Tanksley 1989).

In this study, one spring type of *C. sativa* was crossed with two different winter biotypes of *Camelina* to study the genetic mechanisms underlying vernalization requirement in winter-type *C. sativa.* These studies identified multiple *FLC* orthologs as potential candidate genes controlling flowering in *Camelina* species. The original hypothesis was that the same QTL would control the vernalization requirement, irrespective of source; however, the results suggested that dependent upon the source of the winter phenotype, QTL originating from different subgenomes of the hexaploid act in determining the vernalization requirement in *C. sativa*.

4.3 Materials and methods

4.3.1 Plant materials

Three different species were used to generate F_2 and $F_{2:3}$ populations: viz. *C. sativa* (TMP23992), *C. alyssum* (CAM176), and *C. sativa* ssp. *pilosa* (CN113692) (**Figure 4.1**). TMP23992 is a spring-type line, while the other two are winter-types. TMP23992 produced flowers within 30 days of seeding, whereas the two winter-types required vernalization treatment to induce bolting. The two winter-type lines differed in morphology for winter behaviour. CN113692 was similar to the *C. sativa* spring-type in the early growth stages, but with increased vegetative branching and reduced height prior to cold treatment. CAM176 was characterized by

a reduced stem with profuse leaves where the vernalization treatment promoted stem elongation, as well as branching and flowering.

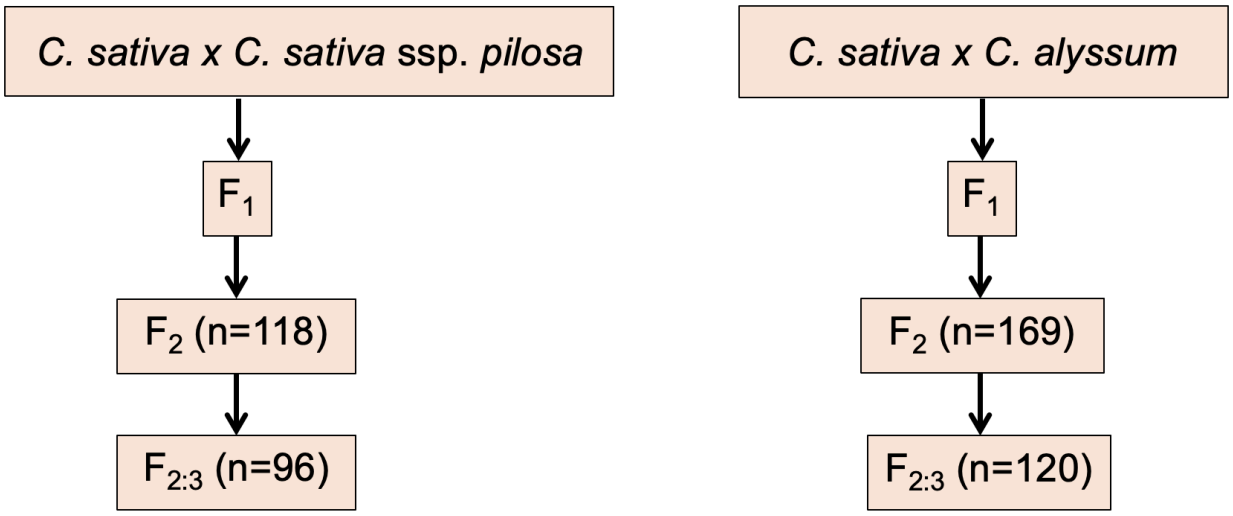


Figure 4.1 Intraspecific hybridization scheme adopted in this study with total number of plants for segregating populations. Accessions TMP23992 (*C. sativa*), CN113692 (*C. sativa* ssp. *pilosa*) and CAM176 (*C. alyssum*) were used in this study. F_1 was self-pollinated to produce F_2 plants and a single seed from individual F_2 plants were used to generate $F_{2:3}$ plants. The total number of plants phenotyped at each generation are shown in parenthesis.

According to **Figure 4.1**, manual crossing was performed with unopened fully developed buds, where TMP23992 (spring-type) was the maternal parent and the winter-types were pollen donors. After pollination, flowers were covered with a bag for 2 weeks. Seeds from mature pods were harvested and planted. The hybrids between TMP23992 and CAM176 produced a winter-type plant; whereas those between TMP23992 and CN113692 produced semi-winter type plants, which flowered in the absence of vernalization; however, with a lower number of reproductive branches relative to the parental lines. Self-seed of each hybrid were used to generate F_2 plants (**Figure 4.1**). F_2 plants showing winter-type morphology were vernalized at 4 °C for 30 days for the *C. sativa* × *C. alyssum* cross (Csa) and 15 days for *C. sativa* × *C. sativa* ssp. *pilosa* cross (Csp). All experiments were carried out in the greenhouse in a soil-less potting mixture with a 16/8 hr of light/dark conditions. Vernalization requirement was determined based on the growth habit 20 days after seeding, where reduced stems with profuse leaves were characteristic of winter-type behaviour. Single seed descent was adopted to generate $F_{2:3}$ plants for additional

confirmation of growth habit. $F_{2:3}$ plants were not subjected to vernalization; those plants either not flowering or late flowering with reduced flower numbers were assumed to have a winter habit. Days to first flower (DTF) for all the plants was recorded from the date of seeding. Plants not flowering 100 days after seeding were assigned a value of 100 for QTL mapping.

4.3.2 Genotyping of segregating populations

Young leaf tissue was harvested from all plants and kept at -80 °C until DNA extraction. DNA extraction was performed using the CTAB method as described in **Chapter 3** and library preparation was as described by Poland et al. (2012) using *PstI* and *MspI* for reduced representation. Paired-end 125 bp sequencing was performed with multiplexed libraries on a Hiseq platform (Illumina). The sequences were de-multiplexed followed by trimming of low quality bases and adapters using Trimmomatic version 0.33 (Bolger et al. 2014) where reads with a minimum length of 55 bp were retained. All high quality reads were mapped to the *C. sativa* reference genome (Kagale et al. 2014a) using BWA (Li et al. 2009) with *bwa-mem* tool with default parameters. From the aligned BAM files SNPs were called using the *UnifiedGenotyper* tool in GATK version 3.2-2 (McKenna et al. 2010) with default parameters.

4.3.3 Genetic analyses of segregating populations

For both populations, all markers polymorphic between the parents were considered, apart from those showing distorted segregation, i.e. deviation from 1:2:1 (χ^2 test, *P-value*< 0.05). Genetic linkage maps were prepared using MSTmap (Wu et al. 2007).

For the Csa population, SNPs for 96 F_2 plants with less than 5% missing genotypes were used to construct a genetic linkage map, where logarithm of odds ratio (LOD) score of 7, mapping threshold of 1 and mapping distance threshold of 1 cM settings were used to determine the number of linkage groups. Markers that failed to cluster with their presumed linkage group (LG) of origin, based on alignment to the reference genome, were forced to cluster with said LG using the single LG function in MSTmap.

For the Csp population, SNPs for 118 F_2 plants with less than 10% missing genotype data were used for map construction with a LOD score of 6, mapping threshold of 1 and mapping distance threshold of 1 cM. As before, further grouping of linkage groups was performed for those markers originating from the same physical chromosome, but separated by high genetic distances. The genetic maps were visualized using MapChart v2.32 (Voorrips 2002). The genetic maps were compared for contiguity using the online version of genetic map comparator (Holtz et al. 2017).

4.4.4 Identification of QTL

QTL analysis was performed with the R/qtl package (Broman et al. 2003) in R statistical software (R Core Team 2019). A single QTL model developed with the Haley-Knott regression method was used to identify QTL. The significance threshold (LOD value) was determined using 1000 permutations and α=0.05, above which QTL were assumed to be significant. The fitqtl method with the drop one term method was adopted for identifying phenotypic variation explained by the QTL, where the method analyzes sub-models to fit the best model, and the percent variance explained for the QTL was calculated by the formula $h^2 = 1\text{-}10^{-2(/n)\text{LOD}}$. The confidence interval of the QTL was identified using Bayesian Credible Interval in the R/qtl package and genes within the confidence interval of the QTL were identified from *C. sativa* annotated genes (Kagale et al. 2014a). Homoeologous chromosomes with QTL were further visualized using KaryoploteR package in the R software (Gel and Serra 2017).

4.4.5 RNA sequencing and sequence analysis

RNA sequencing of the parents, *C. sativa* (TMP23992) and *C. alyssum* (CAM176), with three biological replications of each, was performed to compare expression differences at early seedling growth stage. Seeds were grown in seed germination pouches (Mega international, Newport, MN 55055, USA) at room temperature for one week before total RNA extraction from the leaf samples. Total RNA was extracted using a standard RNeasy Plant Qiagen kit as described by the manufacturer with on-column DNA digestion. RNA was quantified using a Qubit (Invitrogen) and the quality determined using an RNA Nano labchip on a Bioanalyzer (Agilent Technologies). Paired-end RNAseq libraries were constructed using the TruSeq RNA preparation kit (Illumina), with 100 ng of RNA used for cDNA synthesis followed by RNA

library preparation. The final library quality was checked using a Bioanalyzer (Agilent Technologies). Sequencing was done on an Illumina HiSeq2000 platform (2 × 125 bp).

Sequence data were filtered for low quality reads, short reads and adapter contamination using Trimmomatic ver. 0.33 (Bolger et al., 2014). Leading and trailing bases with quality below 15 and reads shorter than 55 bp were removed. All trimmed reads were aligned with the annotated *C. sativa* reference genome (Kagale et al., 2014) using STAR (Dobin et al. 2013) using default parameters, except for --*alignIntronMax* set at 10000 and --*outFilterMismatchNmax* set at 4. *GeneCounts* in STAR provided read counts per annotated gene. Normalization of read counts was done using the Fragment Per Kilobase of transcripts per Million mapped reads (FPKM) method. Differential gene expression analysis was performed with the edgeR package (Robinson et al. 2010) in R statistical software (R Core Team 2019).

4.5 Results

4.5.1 Population development and determination of winter-type behaviour in *Camelina*

The cross between spring-type *C. sativa* (TMP23992) and winter-type *C. sativa* ssp. *pilosa* (CN113692) produced a semi-winter hybrid which took 54 days to flower without vernalization in comparison to 30 days for the maternal *C. sativa* (TMP23992) and 87-91 days for the paternal *C. sativa* ssp. *pilosa* (CN113692). Hybrids between *C. sativa* (TMP23992) and winter-type *C. alyssum* (CAM176) produced winter-type plants that required vernalization in order to flower.

Two F_2 populations (Csp and Csa) were developed from the F_1 hybrids (single hybrid plant for each population) derived from each cross and used to determine the segregation of winter-type behaviour. For both populations, segregation of winter-type habit (leaf morphology and early plant growth) was noted in the F_2 and $F_{2:3}$ plants. Segregation for days to flower and reduced stem growth was observed for both populations. In the case of the Csp population, 45 of 118 F_2 plants showed spring-type behaviour, while the remaining 73 plants showed semi-winter-type behaviour (**Figure 4.2A**). The latter were subjected to vernalization and all F_2 plants flowered within 70 days of seeding, including a vernalization period of 15 days. A total of 96 $F_{2:3}$ plants

were grown and phenotyped from the Csp population, where all plants flowered within a range of 27-55 days of seeding without vernalization. There were no typical winter-type plants among the $F_{2:3}$ plants; however, the plants segregated for days to first flower (**Figure 4.2C**).

For the Csa population, 169 F_2 plants were grown, of which 13 plants showed typical spring-type growth behaviour and the remaining 156 plants showed winter-type behaviour (based on reduced stem elongation) (**Figure 4.2B**). All 156 plants were subjected to vernalization treatment for 30 days; however, only 126 plants flowered within 100 days of seeding. From these F_2 plants, 120 were successfully established in the $F_{2:3}$ generation, which were tested for flowering behaviour in the absence of vernalization treatment. Among the 120 $F_{2:3}$ plants, 30 plants were identified as spring-type and produced flowers within 58 days of seeding, 18 plants transitioned to the flowering stage with few flowers after 70 days of seeding, while 72 plants did not flower until at least 100 days after seeding (**Figure 4.2D**).

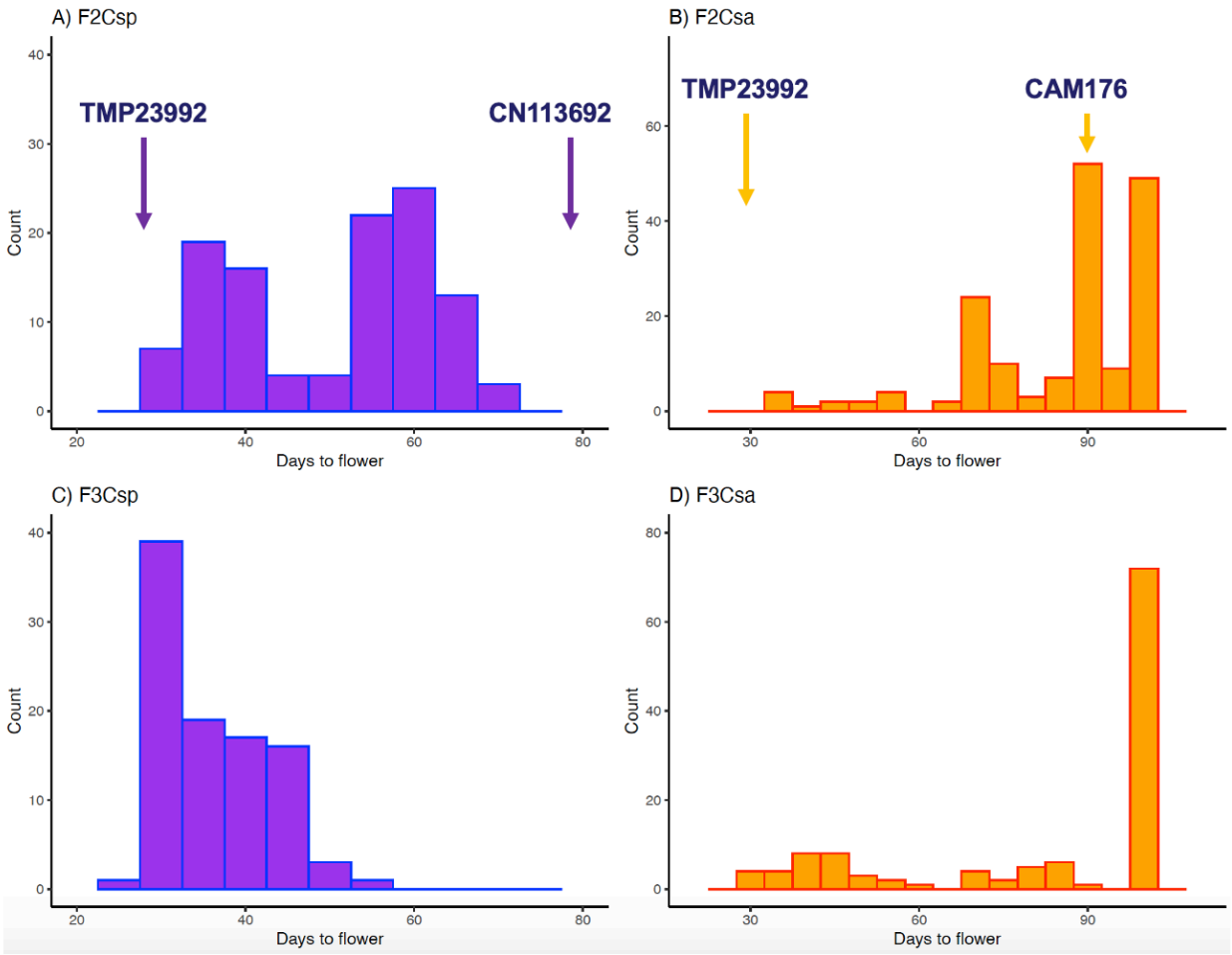


Figure 4.2 Flowering behaviour in segregating intra-specific *Camelina* populations. Frequency distribution of days to flowering in (A) F_2 developed from a TMP23992 × CN113692 (*C. sativa* × *C. sativa* ssp. *pilosa*) cross; (B) F_2 developed from a TMP23992 × CAM176 (*C. sativa* × *C. alyssum*) cross; (C) $F_{2:3}$ developed from a TMP23992 × CN113692 (*C. sativa* × *C. sativa* ssp. *pilosa*) cross; and (D) $F_{2:3}$ developed from a TMP23992 × CAM176 (*C. sativa* × *C. alyssum*) cross. Arrow indicates days to first flower for parental lines.

4.5.2 Genetic linkage maps of *Camelina sativa*

For genotyping, 118 F_2 plants from the Csp population and 169 F_2 plants from the Csa population were used. In the case of the Csp population, 84,346 SNPs were identified and after filtering for those with more than 10% missing genotypes and distorted segregation, 1550 SNPs were used for genetic linkage map construction (**Figure 4.3A**). SNPs were mainly filtered out due to missing data points, increasing the threshold for missing genotypes led to significant deviations from the expected segregation ratio, which could suggest errors in the genotype calls.

A map with a total length of 2193.8 cM was constructed where the number of markers per linkage group ranged from 16 on chromosome 2 to 158 on chromosome 20 with an average mapping interval of 1 marker per 1.42 cM (**Appendix A.1.1**). For the Csa population, 115,827 SNPs were identified for 169 genotypes; however, for the genetic linkage map only 96 genotypes with sufficient sequence coverage to confidently call SNPs were used. Upon filtering for distorted segregation and those with more than 5% missing genotypes, 3279 SNPs were identified and mapped across the 20 chromosomes of the reference *C. sativa* genome (**Figure 4.3B**). The map encompassed 2399.96 cM with an average of 0.73 cM/marker. The number of markers per linkage group ranged from 50 on chromosome 2 to 420 on chromosome 11 (**Appendix A.1.2**).

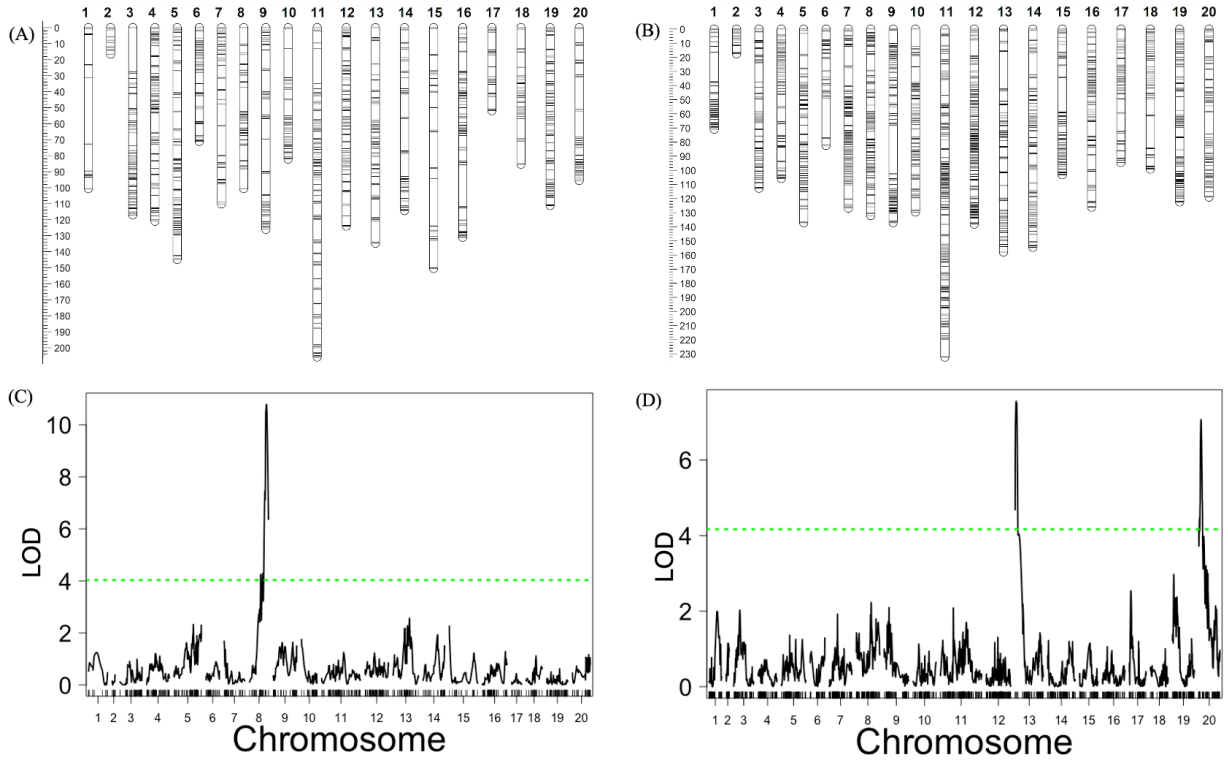


Figure 4.3 Mapping of QTL associated with vernalization requirement in *C. sativa.* (A) Genetic linkage map derived from the TMP23992 × CN113692 (*C. sativa* × *C. sativa* ssp. *pilosa*) F_2 population; (B) Genetic linkage map derived from the TMP23992 × CAM176 (*C. sativa* × *C. alyssum*) F_2 population; (C) QTL identified in the TMP23992 × CN113692 (*C. sativa* × *C. sativa* ssp. *pilosa*) F_2 population; and (D) QTL identified in the TMP23992 × CAM176 (*C. sativa* × *C. alyssum*) F_2 population. cM distance is shown to the left of the maps in panels (A) and (B). The significance threshold for identifying QTL is shown as a green line in panels (C) and (D).

The two genetic maps showed good collinearity along their length (**Appendix A.1.3**). In addition, the genetic maps identified a potential misassembly in the reference genome of *C. sativa* on chromosome 16 (~10 Mb region), where an insertion from the terminal region of chromosome 17 (34 Mb) was found for both maps. The inserted region represented a small fraction of ancestral genomic block D (Kagale et al. 2014a).

4.5.3 Mapping QTL for winter-type behaviour in *Camelina*

For both populations, DTF values measured from the F_2 plants, where the data represented variation in DTF in response to vernalization, were used to identify QTL in *C. sativa.* For the Csp population, the analysis identified a strong QTL correlated with winter-type behaviour on chromosome 8, base pair position 2,323,768 (LOD = 10.8), which explained 36.07% of the phenotypic variation (**Figure 4.3C**) (**Table 4.1**). The *C. sativa* ssp. *pilosa* allele was co-dominant, where heterozygosity at the linked SNP loci was associated with an intermediate late flowering phenotype (**Appendix A.1.4**). The range of the confidence interval for the identified QTL was 7 cM (0.18 Mb in the physical map). On the physical map, within the 95% confidence interval of the QTL, 37 annotated genes were identified. An ortholog of *FLC* (*Csa08g054450*) was identified, which was just 67 kb outside the QTL interval and no other flowering-related genes were identified within or close to the QTL region.

Table 4.1 QTL for vernalization requirement in *C. sativa* measured as a days to first flower in F_2 populations.

Populations	LG	Loci	LOD	Confidence interval	R^2
TMP23992 × CN113692	8	*FLC*	10.8	7 cM	36.07
TMP23992 × CAM176	13	*FLC*	7.50	8.5 cM	15.47
	20	*FLC*	7.07	7 cM	14.27

Two QTL were identified in the Csa population, one on chromosome 13 (LOD = 7.50) and the second on chromosome 20 (LOD = 7.07) (**Figure 4.3d**) (**Table 4.1**). The QTL map interval on chromosome 13 was 8.5 cM (3.07 Mb in the physical map), whereas it was 7 cM (3.3 Mb in the

physical map) on chromosome 20. In this population, QTL on chromosome 13 showed a dominant effect whereas on chromosome 20 showed an additive effect (**Appendix A.1.4**). These two QTL intervals represented homoeologous segments of the reference *C. sativa* genome, where the QTL interval on chromosome 13 comprised 867 genes and that on chromosome 20 comprised 1094 annotated genes in the physical map. The QTL interval on chromosome 13 represented the terminal region of the linkage group and encompassed 6 flowering related genes (*Csa13g001890*, *Csa13g002660*, *Csa13g003870*, *Csa13g003940*, *Csa13g006240* and *Csa13g008090*), among these *Csa13g006240* is related with the vernalization response, also known as *EARLY FLOWERING 6*, where mutation of the gene causes early flowering (Noh et al. 2004). Beside this, the peak of the QTL was 913 Kb away from *FLC.* In the case of the QTL on chromosome 20, 6 genes were identified as flowering-related genes (*Csa20g011780*, *Csa20g015400*, *Csa20g017070*, *Csa20g018140*, *Csa20g018850* and *Csa20g019190*) within the confidence interval of QTL, among these *FLC* (*Csa20g015400*), *EMF1* (*Csa20g017070*) and *FY* (*Csa20g018850*) were identified to have a role in the vernalization response (Michaels and Amasino 1999; Simpson et al. 2003; Aubert et al. 2001). These two QTLs together explained 29.74% of the phenotypic variation in the Csa population (**Table 4.1**). All three QTL represented homoeologous segments of the reference *C. sativa* genome, showing synteny with *A. thaliana* chromosome 5 ancestral genome block R, but with a difference in the absolute confidence interval. The one gene found in proximity to all three QTL was *FLC,* suggesting that it could be the probable candidate gene for the vernalization requirement in *C. sativa*.

4.5.4 Differential gene expression in winter-type *Camelina* during the seedling stage

Differential gene expression analysis was performed for two parental lines: *C. sativa* (TMP23992) spring-type and *C. alyssum* (CAM176) winter-type. Expression differences were compared for genes found only around the QTL on chromosomes 13 and 20 to identify genes showing differential expression between the winter- and spring-type parents at the early seedling establishment stage without vernalization.

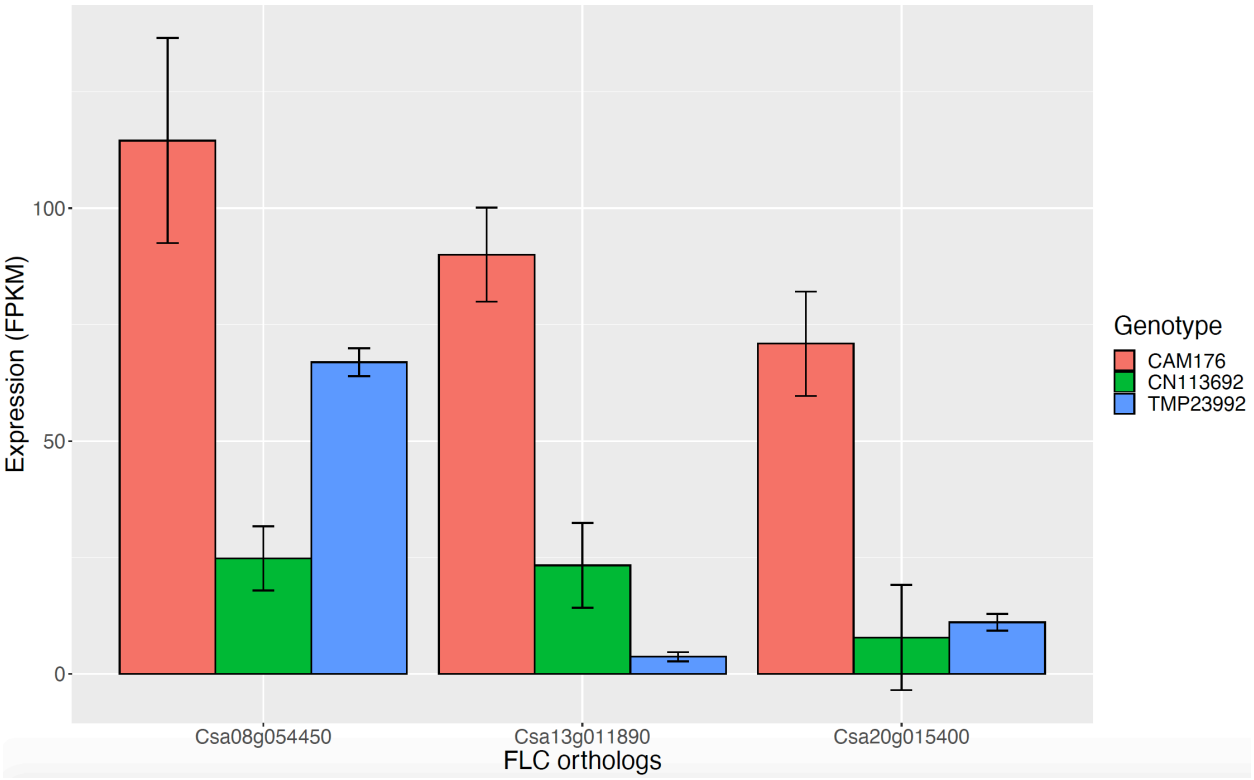


Figure 4.4 Comparison of the level of expression among *FLC* orthologs in spring- and winter-type parents. Genotype TMP23992 (*C. sativa*) is a spring-type whereas CAM176 (*C. alyssum*) and CN113692 (*C. sativa* ssp. *pilosa*) are winter-type. Gene expression for each ortholog is shown in the histogram as mean Fragments Per Kilobase of transcript per Million mapped reads (FPKM) calculated for replicated RNASeq data, error bars represent standard deviation of the mean value.

Differential gene expression analysis between spring-type *C. sativa* and the winter-type *C. alyssum* seedlings identified 2095 differentially expressed genes (*FDR* < 0.05). Among these, 37 genes were found in the QTL regions on chromosomes 13 and 20 (**Appendix A.1.5**). None of the genes showing higher expression in the spring-type parent compared to the winter-type parent were related to flowering behaviour. However, in the winter-type parent, *FLC* orthologs on chromosomes 13 and 20 showed higher expression, along with genes related to stress and reduced growth behaviour (*JAO2, ATST2A*, *GDH2*, *ATL55*, *THALIANA METHYL ESTERASE 5*) (**Table 4.2, Appendix A.1.5**). Duplicated orthologs of four *A. thaliana* genes (*AT5G07010*, *AT5G07440*, *AT5G09930*, *AT5G10140*) on chromosomes 13 and 20 showed higher expression for the winter-type parent, while the two genes (*AT5G03230* and *AT5G02840*) with higher expression in the spring-type parent were related to senescence and stress, respectively (**Table**

4.2). The *FLC* ortholog on subgenome 1 (chromosome 8) had a similar level of expression for both parents (**Figure 4.4**). *FLC* was the only gene found to display differential expression across both QTL.

Table 4.2 Comparison of differentially expressed genes in winter-type CAM176 and spring-type TMP23992 within QTL intervals for flowering time.

QTL	Gene	LogFC*	*A. thaliana* ortholog	Gene name
QTL13	**Csa13g003160**	**-0.33**	**AT5G03230**	***Senescence regulator***
	Csa13g009030	2.65	AT5G07010	*ATST2A, ST2A*
	Csa13g009470	2.88	AT5G07440	*GDH2*
	Csa13g011640	2.06	AT5G09930	GCN subfamily
	Csa13g011890	4.56	AT5G10140	*FLC*
	Csa13g012150	2.69	AT5G10380	*ATL55, ATRING1*
QTL20	**Csa20g002890**	**-0.57**	**AT5G02840**	***LCL1, LHY/CCA1-LIKE 1, REVEILLE 4***
	Csa20g006870	3.20	AT5G05600	*JAO2, JOX2*
	Csa20g009360	3.22	AT5G07010	*ATST2A, ST2A*
	Csa20g009840	3.50	AT5G07440	*GDH2*
	Csa20g013130	3.68	AT5G09930	GCN subfamily
	Csa20g015400	2.43	AT5G10140	*FLC*
	Csa20g015600	4.91	AT5G10300	*THALIANA METHYL ESTERASE 5*

*minimum logfold 2 change higher in CAM176 in compared to TMP23992 and the highlighted genes represents genes showing higher expression in TMP23992 compared to CAM176.

In the case of *C. sativa* ssp. *pilosa* expression data were not generated for this study; however, for a preliminary comparison, expression data generated previously from one-month old seedlings of *C. sativa* ssp. *pilosa* was used (Venkatesh Bollina, unpublished data). The expression data obtained from *C. sativa* ssp. *pilosa* and *C. sativa* were converted to a log2 scale

and were analysed for differential expression using eBayes function in limma package (Ritchie et al. 2015) in R statistical software.

This preliminary comparison between *C. sativa* and *C. sativa* ssp. *pilosa* identified 28,638 differentially expressed genes (adj. *P-value* <0.05). Among these, 212 were differentially expressed in the QTL region on chromosome 8, five of which were flowering-associated genes [*Csa08g054080* (*GRF6*), *Csa08g054450* (*FLC*), *Csa08g056180* (*NF-YB12*), *Csa08g058440* (*MYB33*) and *Csa08g058790* (*CPD*)]. The logFC in expression was higher for *GRF6* and *FLC* in comparison to other genes for spring-type in comparison to winter-type (**Appendix A.1.5**). *GRF6* (14-3-3 proteins) is mainly associated with promoting early flowering, whereas the mutant allele is responsible for late flowering (Mayfield et al. 2007; Higgins et al. 2010). Notably, the level of expression for *FLC* was higher for the spring-type in comparison to the winter-type parent (**Figure 4.4**).

4.6 Discussion

Flowering is a crucial stage in a plant's growth cycle and has a direct influence on adaptation, fitness and overall plant productivity. In nature, biennial and annual flowering behaviour has been reported in a number of *Brassica* species and cereal crops (Kim et al. 2009). The annual nature of flowering is often characterized as an important adaptive trait during the domestication process in crop species (Ågren et al. 2017). The study of vernalization in *C. sativ*a could provide insights in other *Brassica* crops due to the high degree of homology shared among these species, as well as the close relationship of this species with the model plant *A. thaliana*.

Winter-type *Camelina* species are represented by plants requiring prolonged cold treatment to promote bolting. In this study, two different types of winter-type *Camelina* plants were identified. The first type, represented by *C. alyssum*, had reduced stems characterized by profuse leaf production, where cold treatment promoted stem elongation and flowering. The second type, represented by *C. sativa* ssp. *pilosa*, was charaterized by longer stems and branching; however, the branches remained in the vegetative stage until vernalized for 2-3 weeks (**Appendix A.1.6**). The availability of these two forms of winter-type *Camelina* enabled potentially different

mechanisms controlling delayed flowering in *C. sativa* to be studied. All of the hybrids with winter *C. alyssum* produced a winter-type plant, which suggested that the winter-type behaviour was a dominant trait. A number of reports have shown a quantitative effect for duration of vernalization (Sheldon et al. 2000; Kemi et al. 2013). Similar observations were made in this study as the days to flowering was greater for hybrid plants (*C. sativa* × *C. alyssum*) vernalized for a shorter duration compared to those vernalized for a longer period (**Appendix A.1.7**). The extent of the variation in vernalization requirement for flowering, as reflected by days to flowering, as well as the difference in the number of reproductive branches in hybrids coming from the same parents (**Appendix A.1.8**), suggest that there is a quantitave variation for vernalization requirement in these species. The winter-type *C. sativa* ssp. *pilosa,* has early growth similar to spring *C. sativa*; however, it did not produce flowers unless exposed to cold treatment (**Appendix A.1.7**). The hybrid from *C. sativa* × *C. sativa* ssp. *pilosa* had a semi-winter type behviour that was more similar to spring-type *C. sativa,* but the number of days to flower was greater compared to the spring-type *C. sativa* parent. From this, it was speculated that a quantitative effect of alleles related to winter-type behaviour for *C. sativa* ssp. *pilosa* might be influencing the days to flowering as well as vernalization requirement.

Genetic maps developed for both populations have a Spearman correlation coefficient of 0.81 (**Appendix A.1.3**), suggesting high contiguity of the linkage maps. The contiguity was tested with 648 common markers between these populations. Since these populations were developed from crosses between one spring-type *C. sativa* parent with two different winter-type *Camelina* species/sub-species, the level of similarity shared among the winter-type parents influenced the number of common markers in the genetic maps. This study also identified a probable mis-assembly in the reference genome where a block representing chromosome 17 was linked together with chromosome 16 on both linkage maps. A revised subgenome structure of the *C. sativa* genome had revealed that chromosomes 16 and 17 should be in different subgenomes and a new syntelog table based on the revised subgenome structure suggested the genes belonging to the terminal region of chromosome 17 (subgenome 3) should be present in subgenome 2, which would be consistent with the results presented here (**Appendix A.1.9**).

Three major QTL affecting winter-type behaviour in *C. sativa* were identified. The QTL identified on chromosome 8 (subgenome 1) for the Csp population was in close proximity to *FLC,* which might suggest *FLC* affects flowering behaviour in *C. sativa* ssp. *pilosa.* The mapping used DTF data and for the $F_{2:3}$ Csp population no plants were kept in vernalization, yet all plants flowered within 55 days. QTL mapping using phenotype data from the $F_{2:3}$ plants and the Csp genetic map confirmed the same QTL controlled the variation in DTF as reflected by the winter-type behaviour of the plants (**Appendix A.1.10**). In contrast, two QTL were identified in the Csa population, where both QTL represented homologous regions in different subgenomes (subgenome 2 and subgenome 3); however, no QTL were detected in subgenome 1 as was found in the Csp population. Within the confidence interval of these two QTL, or in close proximity, a copy of *FLC* was identified as a major flowering gene. However, low linkage disequilibrium (LD) detected for the the markers around *FLC*, especially on chromosome 8, could suggest other genes might also be responsbile for affecting days to first flower (**Appendix A.1.11**). Additionally, these QTL were further confirmed through mapping of $F_{2:3}$ phenotypes, where the same QTLs were identified on chromosomes 13 and 20 as in the F_2 generation, but with a less significant p-value for the QTL on chromosome 20, which might suggest further segregation of codominant alelles (**Appendix A.1.12**). The low number of samples in both populations probably hindered the level of confidence for the identified QTL and in quantifying minor QTL; however, the study identified three major QTL in two populations that have a significant effect in causing variation for flowering time/vernalization requirement.

For the QTL on chromosome 8, the confidence interval was narrow in comparison to the other QTL, likely due to a higher number of markers around this QTL with a greater frequency of genetic recombination. The proximity of *FLC* to the QTL suggested that *FLC* could be a major flowering gene influencing vernalization requirement, as well as affecting days to flowering in winter-type *Camelina* similar to other crops (Deng et al. 2011; Okazaki et al. 2007; Xiao et al. 2013; Zhao et al. 2010). The QTL on chromosome 13 is an additional locus to that previously reported by Anderson et al. (2018) on chromosome 20 for vernalization requirement in *C. sativa*. The QTL on chromosome 8 (*FLC*) might have a role in flowering time variation in *C. sativa*, in addition to the vernalization response. Likewise, in the study by Chao et al. (2019) with additional winter-type *Camelina* lines, differential expression of *FLC* ortholog was observed on

chromosomes 8 and 20; however, a locus associated with *FLC* on chromosome 13 was not reported. The study by Anderson et al. (2018) suggested that *FLC* on chromosome 20 was a determinant for vernalization requirement where a one base deletion resulted in a non-funtional *FLC* protein in the spring-type *C. sativa.* Also, the *FLC* on chromosome 8 for winter-type *C. alysum* had a two base-pair insertion in comparison to *FLC* in spring-type *C. sativa;* however, a three basepair insertion was reported in another winter-type *C. sativa* variety Joelle (Anderson et al. 2018). The identified insertion would be reponsible for changes to the amino acid composition of *FLC* (**Appendix A.1.13**) in winter *C. alyssum*, which suggested a difference to the role of *FLC* on chromosome 8 in this winter line compared to the variety Joelle. The higher expression of *FLC* on chromosome 8 for both spring- and winter-type plants is of interest to identify the possible function of this homologue of *FLC*, especially due to the identification of a QTL on chromosome 8 from the Csp population.

For the Csa population, the higher expression of *FLC* on chromosomes 13 and 20 in winter-type lines suggested a similar role of *FLC* as that reported in a number of other species (Anderson et al. 2018; Sheldon et al. 2000; Okazaki et al. 2007). However, the level of expression of *FLC* on chromosome 8 (subgenome 1) was found to be similar for both parents, which suggests that *FLC* on chromosome 8 may have a different function other than controlling vernalization requirement for these parents. A similar constant expression for *FLC* ortholog (*Bna.FLC.CO2*) in winter- and spring-type *B. napus* has been reported, although the orthologs were found to be different in their 5′ UTR and first exons (Schiessl et al. 2019). For the *C. sativa* ssp. *pilosa* winter-type parent, the level of *FLC* expression on chromosome 8 (**Figure 4**) was lower compared to the other two parents suggesting the same locus has a role in vernalization, where a low amount of cold treatment might be able to surpass the threshold to initiate flowering. This might represent a case of *FLC* divergence during the course of evolution for these species as suggested for *Capsella rubella* (Yang et al. 2018), where a deletion in the 5′ UTR controls the amount of vernalization required for early flowering and variation for flowering time. Likewise, expression of another flowering gene, *Csa08054080* (*GRF6*), within the QTL range was also differentially expressed (*P-value*<0.05) being higher in the *C. sativa* parent compared to *C. sativa* ssp. *pilosa;* this particular gene is responsible for vegetative growth, leaf develoment and photoperiodism (Vercruyssen et al. 2014). In this study, beside the flowering gene *FLC,* other genes were also

identified with a higher expression in the winter-type parents and those genes might be responsible for inhibition of seedling growth, repression of leaf expansion and are mostly related with stress (Huang et al. 2017; Goel and Singh 2015; Tercé-Laforgue et al. 2013) (**Appendix A.1.14**).

Overall, the structured, biparental, mapping populations with different winter-type sources helped to identify QTL responsible for winter-type behaviour in *C. sativa* and, noticeably, they were independently identified on different subgenomes in two populations. The findings also increased the prospect of studying the structure of *FLC* orthologs as an evolutionary adaptation process for *Camelina* species, as the same orthologs have different levels of expression and, possibly different functions in controlling the flowering response. In addition, the approach used in this study can be used to identify additional forms of winter-type *Camelina* by developing markers associated with these QTL, where a combination of these QTL can be expected and will have a impact in the *Camelina* breeding program.

Prologue to chapter 5

After successful development of segregating plants from intraspecific hybridization, an attempt was made to produce segregating generations and observe the nature of recombination from interspecific hybridization. In this study, the parents differed with regards to the number of chromosomes and subgenome structure; however, they shared two subgenomes. Hybridization was poor; however, a few seeds were produced that were used to develop segregating populations. Analysis of the F_2 and the backcross-derived F_2 populations provided evidence for homoeologous recombination between subgenomes. In addition, the variability generated from this study can be used to further identify genetic mechanisms associated with traits of interest.

This chapter will be submitted to BMC Genomics.

Chapter 5. Evidence of homoeologous recombination in *Camelina* species from interspecific hybridization

5.1 Abstract

A recent whole genome triplication in *Camelina sativa* has led to an undifferentiated subgenome structure. The stability of such neopolyploid genomes depends on the nature of the recombination, where disomic inheritance plays an essential role in the fitness of the plant during the course of evolution. With the advancement of sequencing technologies, our understanding of the nature and distribution of genetic recombination is much clearer. Homoeologous recombination can occur where the subgenomes are less differentiated and share greater homology, as is the case for *C. sativa*. This study was carried out to assess whether homoeologous recombination in *Camelina* species was possible following interspecific hybridization between *C. sativa* and *C. microcarpa*. Although only a limited amount of segregating seed was obtained from crosses between these species, a high level of morphological variation for leaf characteristics, anthesis and sterility was observed. Homoeologous recombination in progeny of the F_2 and BC_1F_2 generations was noted, with possible evidence of aneuploidy. In addition, this study also mapped QTL associated with vernalization requirement in winter-type *C. microcarpa*, which fell in the same region as that found for winter-type *C. alyssum*, suggesting conserved function of vernalization requirement genes across *Camelina* species. Overall, the results of this study suggest the low level of genetic differentiation among subgenomes in *Camelina* species might facilitate non-homologous recombination.

Keywords: aneuploidy, homoeologous recombination, interspecific hybridization, QTL

5.2 Background

Camelina sativa (L.) Crantz is a hexaploid species of the Brassicaceae family (Al-Shehbaz et al. 2006). A number of wild relatives of this species have been reported (Martin et al. 2017) with some well-defined taxonomically, while others remain uncharacterised with evidence of cryptic species with indistinguishable morphologies. *Camelina microcarpa* Andrz. ex DC is one of the closest wild relatives of this crop (Brock et al. 2018) and two different hexaploid *C. microcarpa* species with distinct combinations of three subgenomes have been identified (Chaudhary et al.

2020). The first type of *C. microcarpa* has the same three subgenomes as *C. sativa*, whereas the second type, *C. microcarpa* "Type 2" shares only two subgenomes with *C. sativa*. The third subgenome from *C. microcarpa* "Type 2" is different from either of the subgenomes present in *C. sativa* and represents a fourth subgenome in *Camelina* species. A number of duplication events and combinations of subgenomes in the generation of the higher ploidy *Camelina* species seem to have occurred, similar to the triangle of U in *Brassica* (Nagaharu and Nagaharu 1935). The identification of the subgenome composition in *Camelina* species has provided a strategy for interspecific hybridization in *C. sativa* and could potentially help in broadening the narrow genetic base of *C. sativa* germplasm (Singh et al. 2015; Brock et al. 2018; Luo et al. 2019a).

Interspecific hybridizations in *Camelina* have been performed successfully (Zhang and Auer 2020; Tepfer et al. 2020; Martin et al. 2019; Séguin-Swartz et al. 2013). Interspecific hybridization between *C. sativa* and *C. microcarpa* "Type 2" conducted by Zhang and Auer (2020) resulted in a very low rate of seed set. Likewise, interspecific hybridization conducted by Tepfer et al. (2020) resulted in meiotic irregularities, likely caused by the difference in genomes shared between these two species, which led to the low levels of hybridization success resulting in reduced seed set. In the same context, interspecific hybridization of *Brassica* species have been performed where the parental polyploid species differed in their subgenome structure (Zhang et al. 2016; Wei et al. 2016). In such cases normal homologous recombination for only one of the contributing subgenomes commonly led to aneuploidy for the unpaired chromosomes. However, these processes are useful for increasing allelic diversity for particular subgenomes, as well as studying the homology between the chromosomes from different species. Such hybridization may also lead to homoeologous recombination due to higher homology shared among some chromosomes. Most of the interspecific hybridization where parents differ in subgenome structure result in chromosome abnormalities, such as aneuploidy, deletions and translocations, which leads to offspring that lack fitness. In order to stabilize such offspring, backcrossing has been popularly adopted (de Jong et al. 2018; Navabi et al. 2011), which helps to generate lines with a stable chromosome content, including those chromosomes with introgressed segments.

Vernalization of winter-type plants is required to induce bolting and flowering, and *FLC* is a major regulator of this process (Michaels and Amasino 1999). All three copies of *FLC* have been conserved across the subgenomes of *C. sativa* (Kagale et al. 2014a) and are suspected to have an effect on the vernalization response; however, the different subgenome copies appear to have independent roles across subspecies (**Chapter 4**) (Anderson et al. 2018). In this scenario, segregating plants developed from crosses with winter-type *C. microcarpa* could be a tool to study the source of the winter-type phenotype in *C. microcarpa*, a species closely related to *C. sativa*.

The aim of this study was to carry out interspecific hybridization among *Camelina* species to understand the nature of recombination between the contributed chromosomes and to follow the impact of such events on morphological variation in *C. sativa.* Since this study mimicked interspecific hybridization events that might happen in nature, the level of fitness was also explored to understand its potential role in the evolution of *Camelina* species. This study suggested that in the absence of homologous chromosomes, homoeologous recombination events are common in *Camelina* due to shared similarity among the subgenomes, but these could be detrimental to the fitness of the plant.

5.3 Materials and methods

5.3.1 Genetic material and population development

Two *Camelina* species were used in this study viz. *C. sativa* (spring-type) and *C. microcarpa* "Type 2", hereafter referred to as *C. microcarpa*. Hybrids were selfed to generate F_2 populations and backcrossed with *C. alyssum* (spring-type) to generate BC_1F_1 plants which were selfed to produce BC_1F_2 individuals (**Figure 5.1**). All three species were hexaploid, where *C. sativa* and *C. alyssum* share the same subgenomic structure (AABBCC) and *C. microcarpa* possessed a different subgenomic structure (AABBDD).

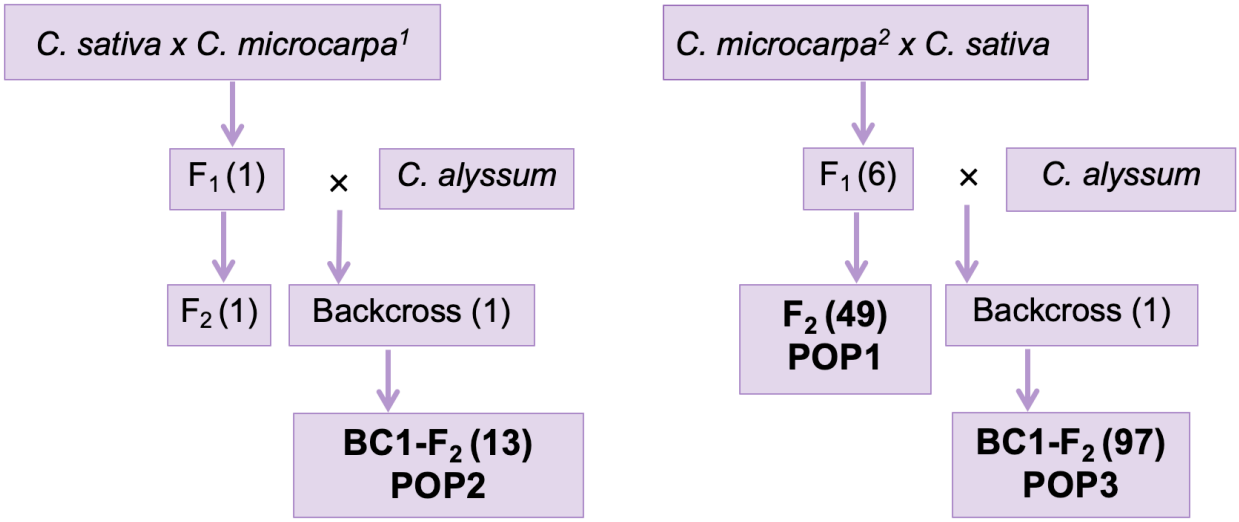


Figure 5.1 Interspecific hybridization scheme in *Camelina* species. Here, TMP23992 (*C. sativa*) and PI650132 (*C. alyssum*) were spring-type whereas [1]TMP23999 and [2]CN119102 (*C. microcarpa*) were winter-type. The number in parenthesis indicate number of plants successfully grown. Genetic analyses were performed for the highlighted populations.

5.3.2 Morphological data collection and protein analysis

Morphological data, such as leaf shape, leaf waxiness, vernalization requirement and days to flower were collected. Vernalization requirement was determined based on the growth habit 20 days after seeding; a profuse leaf phenotype characterized by reduced stem was indicative of winter-type *Camelina*, whereas stem elongation was characteristic of spring-type *Camelina* (also discussed in **Chapter 4**). Days to flowering (DTF) was defined as the number of days from seeding to the appearance of the first flower. Seed protein analysis was performed with a subset of POP2 (BC_1F_2) plants along with the parents to observe any changes in protein profile due to recombination with TMP23999 (*C. microcarpa*). The protein extraction and profiling were done as described by Lyzenga et al. (2019), where protein was extracted from 30 mg of seed and separated on an Experion Automated Electrophoresis system (Bio-Rad) using an Experion™ Pro 260 analysis kit. Two biological replications were performed for the parental lines; however, only single samples were available for the segregating plants.

5.3.3 Library preparation and sequence data processing

One week old leaf tissue was harvested and kept at -80 ºC until DNA extraction. DNA extractions were performed using the CTAB method and GBS library preparation was as described by Poland et al. (2012). Paired-end sequencing was done on the Hiseq platform. Sequences were trimmed for low base quality and adapters using Trimmomatic (Bolger et al. 2014). Reads were mapped to a pseudogenome (**Appendix A.2.1**), generated from combining the *C. sativa* reference genome (Kagale et al. 2014a) with the third subgenome from *C. microcarpa* Type 2 (Chaudhary et al, unpublished), using BWA (Li et al. 2009) with the *bwa-mem* program and default parameters. SNPs were called using GATK (McKenna et al. 2010) with default parameters from the aligned BAM file.

5.3.4 Genetic analyses of segregating populations and QTL mapping

Sequence reads were used to study possible aneuploidy and homoeologous recombination. BEDTools (Quinlan and Hall 2010) was used to extract mapped reads and to calculate the frequency of mapped reads along 100 Kb bins in the genome. Reads were plotted across the bin to confirm possible genetic events. Circos (Krzywinski et al. 2009) and karyotypeR (Gel and Serra 2017) were used to plot the distribution of mapped reads along the *Camelina* pseudogenome for visualizing aneuploidy and homoeologous recombination events. A genetic map could only be prepared for the BC_1F_2 population from the (*C. microcarpa* × *C. sativa*) × *C. alyssum* cross. The VCF file was filtered to remove SNPs with more than 15% missing genotypes and for those SNPs monomorphic between *C. alyssum* and either *C. sativa* or *C. microcarpa*. Alleles coming from *C. alyssum* were assigned as genotype ‘B’ and the alternate allele (A) was assumed to come from either *C. sativa* or *C. microcarpa*. Using the filtered SNPs, a genetic linkage map was prepared using MSTmap (Wu et al. 2007) with the Kosambi mapping function and maintaining a mapping distance threshold of 1 cM. QTL analysis for vernalization requirement in the segregating population was performed with the R/qtl package (Broman et al. 2003) in R statistical software (Team 2019). Haley-Knott regression method was used to identify QTL as described in **Chapter 4**.

5.4 Results

5.4.1 Development of segregating populations and observed morphological variation

Interspecific hybridization between *C. sativa* (TMP23992) × *C. microcarpa* (TMP23999) was performed and a limited number of seeds were obtained with *C. sativa* as the maternal parent. In these crosses, only 6 pods (1 seed) formed after manual pollination of 190 flowers. The reciprocal cross, with a different accession of *C. microcarpa* (CN119102), produced 71 seeds from 13 pods. Most of the hybrid seed obtained from these crosses were deformed and only 6 out of 71 seeds germinated. Both *C. microcarpa* genotypes were winter-type and required vernalization to reach the reproductive phase whereas TMP23992 (*C. sativa*) was a spring-type. The hybrids from these crosses produced a winter-type plant that had intermediate plant and leaf morphology from the parents (**Appendix A.2.2**) and showed some evidence of best parent heterosis for DTF (**Appendix A.2.3**). The hybrid from the TMP23992 × TMP23999 cross was effectively sterile and produced only one F_2 seed; whereas four F_1 plants from the CN119102 × TMP23992 cross, although partially sterile, produced 49 F_2 seeds (referred to as POP1) (**Figure 5.1**).

The hybrids from both combinations were also backcrossed with accession PI650132, a spring-type *C. alyssum* to generate BC_1F_1 plants. Only one BC_1F_1 plant was produced from the (*C. sativa* × *C. microcarpa*) × *C. alyssum* cross, whereas 8 seeds were obtained from the (*C. microcarpa* × *C. sativa*) × *C. alyssum* cross; however, only one plant was used to generate the BC_1F_2 (POP3) generation. BC_1F_2 plants derived from selfed-seed from the (TMP23992 × TMP23999) × PI650132 cross, referred to as POP2, and the (CN119102 × TMP23992) × PI650132 cross, referred to as POP3, were used for morphological and genetic analysis (**Figure 5.1**).

5.4.2 Characterization of the POP1 F_2 population

A total of 49 F_2 plants were obtained in POP1 of which four plants showed spring-type early seedling growth and flowered in 40, 54, 90 and 98 days after seeding, while the remainder of the plants were winter-type. Plants exhibiting winter growth habit after 20 days were then kept in

vernalization for 30 days. The winter-type plants flowered between 81 to 110 days after seeding. There was a marked difference in leaf shape and leaf waxiness in the segregating plants in comparison to the parents (**Figure 5.2**). These plants had 15 to 134 flowers on the main raceme; however, for some plants it was difficult to count the number of flowers due to stunted growth of the plant and reduced racemes. Among the 49 plants, only 7 were completely fertile, 3 were partially sterile and the rest were sterile. Nine plants displayed a waxy upper leaf surface, while the rest had a rough upper leaf surface (**Appendix A.2.4**). One of the plants, 83-8-28, showed leaf curling behaviour. The leaves from F_1 hybrid 83-8 were mostly horizontal, whereas those from the other hybrids were mostly upright (**Appendix A.2.5**).

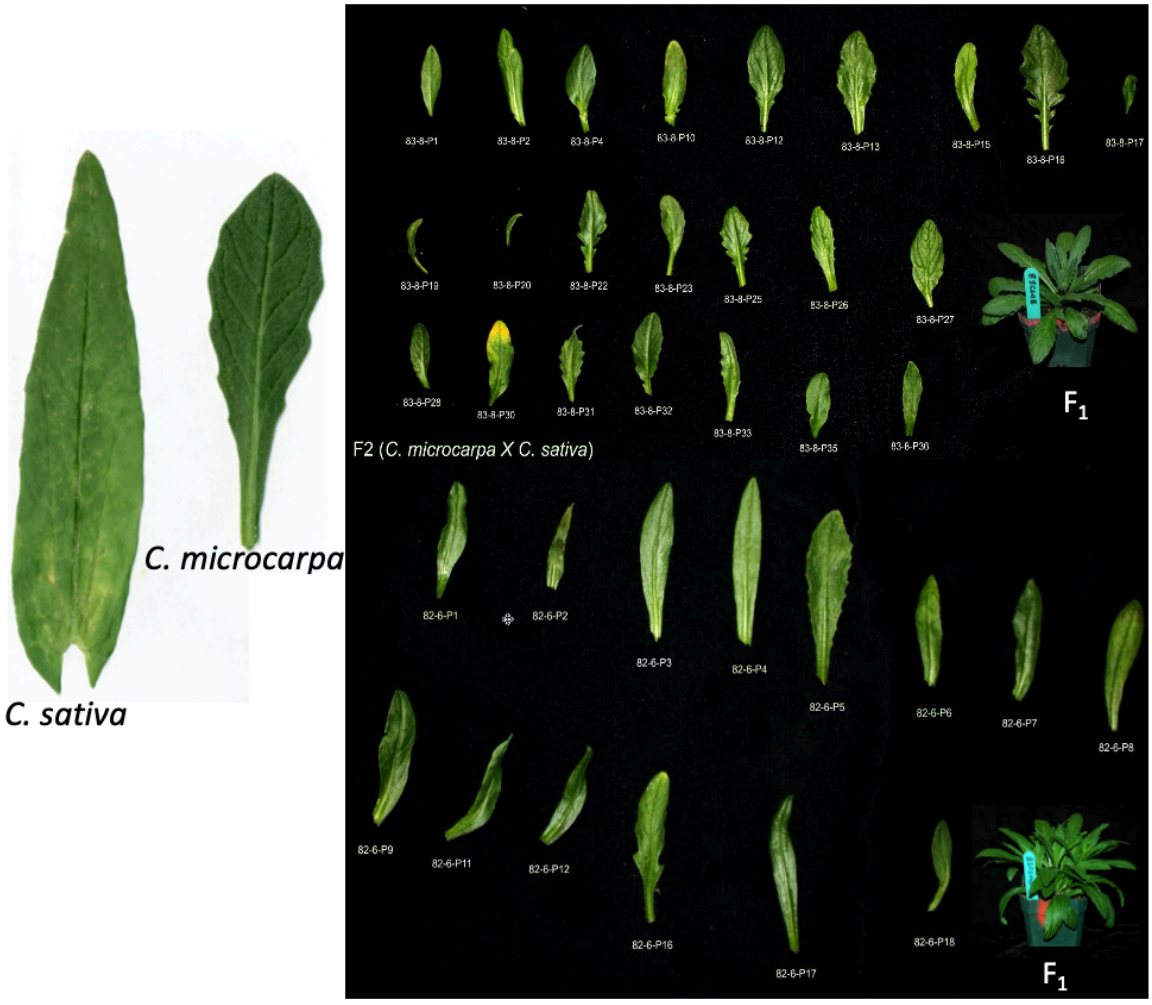


Figure 5.2 Leaf shape variation in two *C. microcarpa* × *C. sativa* F_2 plants. Leaves were observed after vernalization for 30 days. Leaves of F_2 plants descended from hybrid 83-8 were similar to the *C. microcarpa* parent (top), whereas leaves of F_2 plants descended from hybrid 83-6 were similar to the *C. sativa* parent (bottom).

It was anticipated that homologous chromosome pairing and recombination would occur between the common subgenomes 1 and 2, but the third subgenome, differentiating the parental lines, would result in some abnormal chromosome pairing. Upon genetic analysis of the POP1 plants, a peculiar mapping of sequence reads was observed where the segregating plants showed a particular bias according to the parental hybrid from which the progeny were derived (**Appendix A.2.6**). Plants in POP1 were bulked from four hybrids (82-6, 83-2, 83-8 and 83-9); among these, plants coming from 83-8 had a higher number of chromosomes missing from subgenome 3 of *C. sativa*, whereas F_2 plants coming from the remaining three hybrids had a higher number of missing chromosomes from subgenome 3 of *C. microcarpa.* Apart from this, a number of plants showed missing chromosomal segments either due to deletion or translocation and most probably due to homoeologous recombination.

The genetic events leading to the segregating plants were identified by analyzing the distribution of reads mapped to the pseudoreference genome. The level of homozygosity for SNPs from those subgenomes lacking homologous chromosomes from both parents (the third subgenome) were also studied to infer pairing of the non-homologous chromosome from the third subgenome belonging to both parents. Such non-homologous recombination events were identified in two F_2 plants, where pairing between chromosome 17 of *C. sativa* with chromosome 14 of *C. microcarpa* could be assumed to result in missing chromosome segments from chromosome 17 of *C. sativa*, where in the same plants the homoeologous chromosome segments (chromosome 14) belonging to *C. microcarpa* had heterozygous SNPs (**Figure 5.3**).

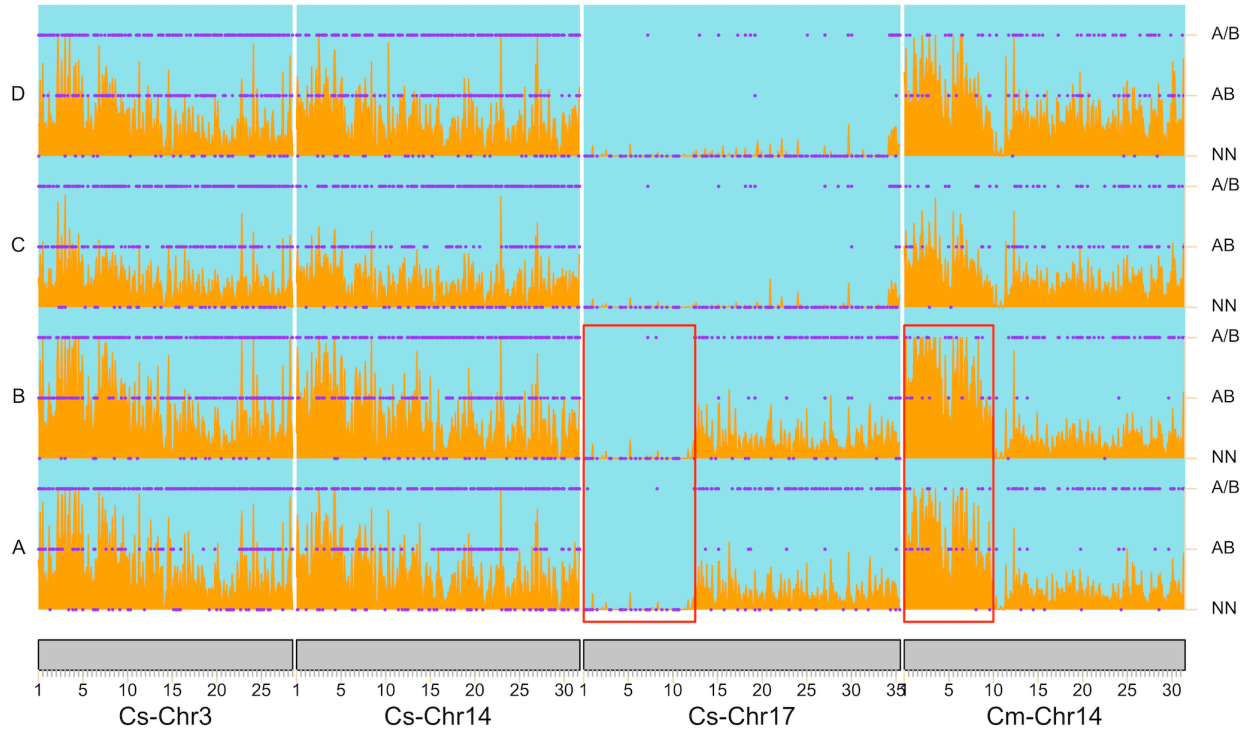


Figure 5.3 Non-homologous recombination in F_2 plants between chromosome 17 from *C. sativa* and chromosome 14 of *C. microcarpa*. Each track represents a different F_2 plant. Tracks A and B represent segmental recombination, highlighted in red box, whereas C and D show aneuploidy for the same homoeologous chromosomes, represented by loss of Cs-chr17. The purple dot shows SNP genotype [homozygotes (A/B), heterozygotes (AB) and missing (NN)] at the physical position in the genome, whereas the orange bar shows mapped read depth (in 100 Kb bins) across the four homoeologous chromosomes. Tracks represent plant 83-8-17 (A), 83-8-22 (B), 83-8-28 (C) and 83-8-32 (D).

5.4.3 Characterization of the POP2 backcross generation

The F_1 from TMP23992 (*C. sativa*) crossed with TMP23999 (*C. microcarpa*) was backcrossed with PI650132 (*C. alyssum*), a spring-type, to generate a backcross population. This backcross population was self-pollinated to get BC_1F_2 seeds. The BC_1F_1 was partially sterile and produced only 15 seeds of which 13 BC_1F_2 plants survived. After two weeks of growth, 4 plants showed winter-type behaviour and were kept in vernalization at 4 °C for 30 days. The leaf morphology of one of the winter-type plants was similar to *C. microcarpa*, while the other three were more similar to *C. sativa* (**Figure 5.4**). Spring-type plants flowered within 37 to 50 days, which was longer than the spring-type parent, and winter-type plants flowered within 97 to 100 days after seeding, which was less than the winter-type parent. The upper surface of the leaves was rough

for all plants (**Appendix A.2.7**). A greater variation in seed size, as reflected by seed weight, was found for the segregating plants; however, none of the plants surpassed the seed size of the spring parent (**Appendix A.2.7**).

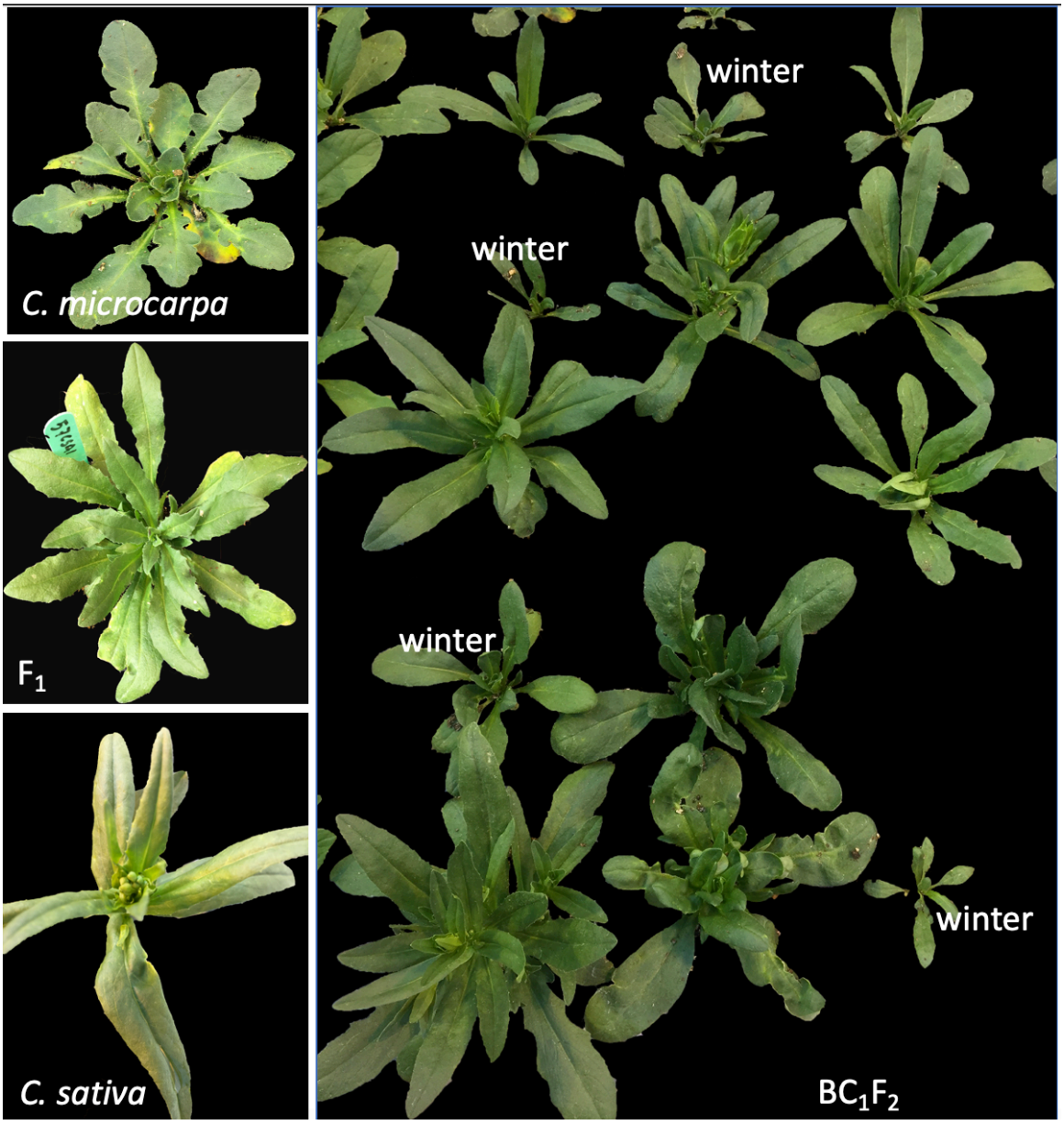


Figure 5.4 Leaf shape variation in BC_1F_2 (POP2) segregating plants (on right) developed from (TMP23992×TMP23999) ×PI650132. Here, TMP23992 (*C. sativa*) and PI650132 (*C. alyssum*) were spring-type, and TMP23999 (*C. microcarpa*) was winter-type. For these plants the maternal parent was TMP23992 (*C. sativa*). Winter-type plants are marked as 'winter' whereas the remainder of the plants were spring-type.

Protein profiling of these segregating plants also showed a biased towards TMP23992 (*C. sativa*). There were a number of marked differences between the protein profiles of TMP23999 (*C. microcarpa*) and TMP23992 (*C. sativa*) (**Figure 5.5A**) such as, higher molecular weight proteins were dominant in TMP23999, and the ratio of peaks for 22.22 kDa, 24.27 kDa and 29.13 kDa were nearly equal for this species compared to TMP23992. Noticeably the peaks at 49.69 kDa and 52.30 kDa were much lower in TMP23999. However, in the case of segregating BC_1F_2 plants most of the protein profile were biased towards spring-type TMP23992 (*C. sativa*) with a higher variation (**Figure 5.5B**) (Details are presented in **appendix A.2.8**).

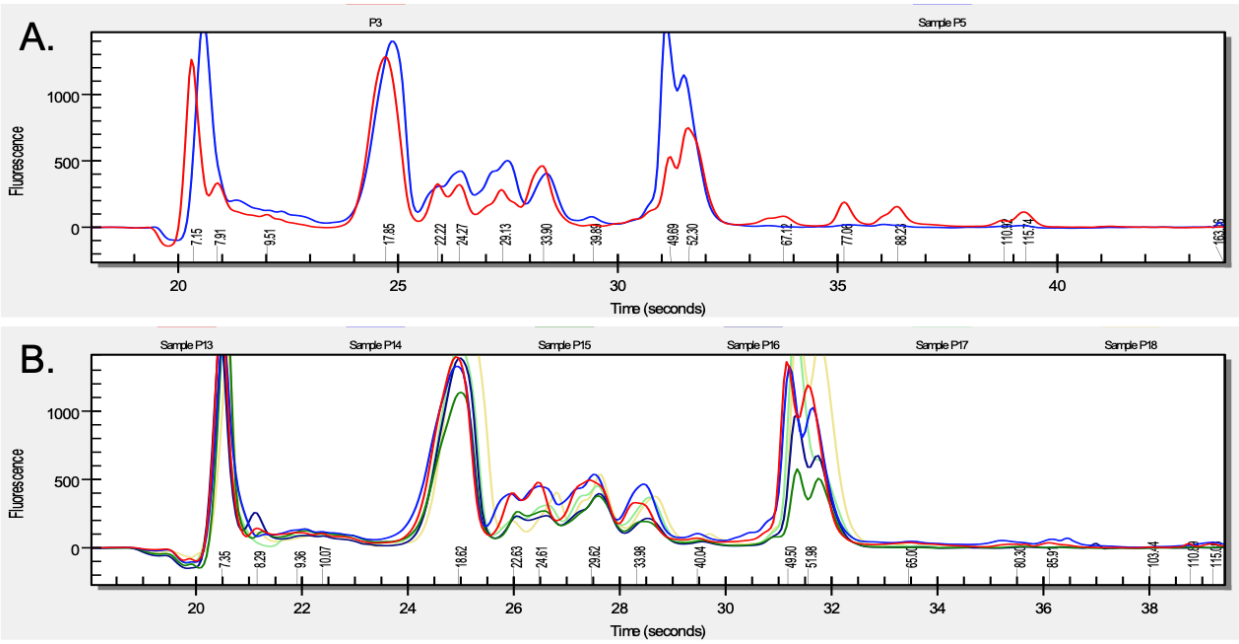


Figure 5.5 Protein profile of parents and segregating BC_1F_2 plants developed from (TMP23992×TMP23999) ×PI650132. A) Protein profile of parent TMP23999 (*C. microcarpa*) in red and TMP23992 (*C. sativa*) in blue; Peaks at 49.69 kD and 52.30 kDa show lower levels for *C. microcarpa*, and a prominent peak at 22.22 kDa was observed for TMP23999. B) Six individual samples represent different BC_1F_2 plants; differences at peaks 22.53 kDa and 8.29 kDa were found; however, most plants showed affinity with TMP23992 (*C. sativa*).

Genetic analysis for this population was also performed by mapping sequence reads to a pseudoreference. Results suggested that none of the segregating plants had missing reads for a whole chromosome for the first two subgenomes which indicated normal homologous pairing across these two subgenomes. However, there was variability in the distribution of reads across

the third subgenome of *C. microcarpa,* which lacks homologous chromosomes*,* since the backcross was performed with accession PI650132 (*C. alyssum*). All the plants showed loss of chromosome Cm15, chromosome Cm18 and chromosome Cm19 from the third subgenome of *C. microcarpa*; however, there were reads for segments of a few chromosomes belonging to the third subgenome of *C. microcarpa* (e.g. chromosomes Cm14, Cm16 and Cm17). There was an inconsistency in read depth across the third subgenome of *C. sativa*, specifically for chromosomes Cs17 and Cs5, and a terminal deletion of chromosome Cs11 of the first subgenome was also observed (**Figure 5.6A**). A variation in read mapping for the third subgenomes of *C. microcarpa* and *C. sativa*, for chromosome Cm14 of *C. microcarpa* and chromosome Cs17 of *C. sativa*, which are homoeologous chromosomes, might suggest a homoeologous recombination event between these chromosomes followed by segregation in the BC_1F_2 generation. For two plants deletion of chromosome Cs12 belonging to the third subgenome of *C. sativa* was found (**Figure 5.6B**).

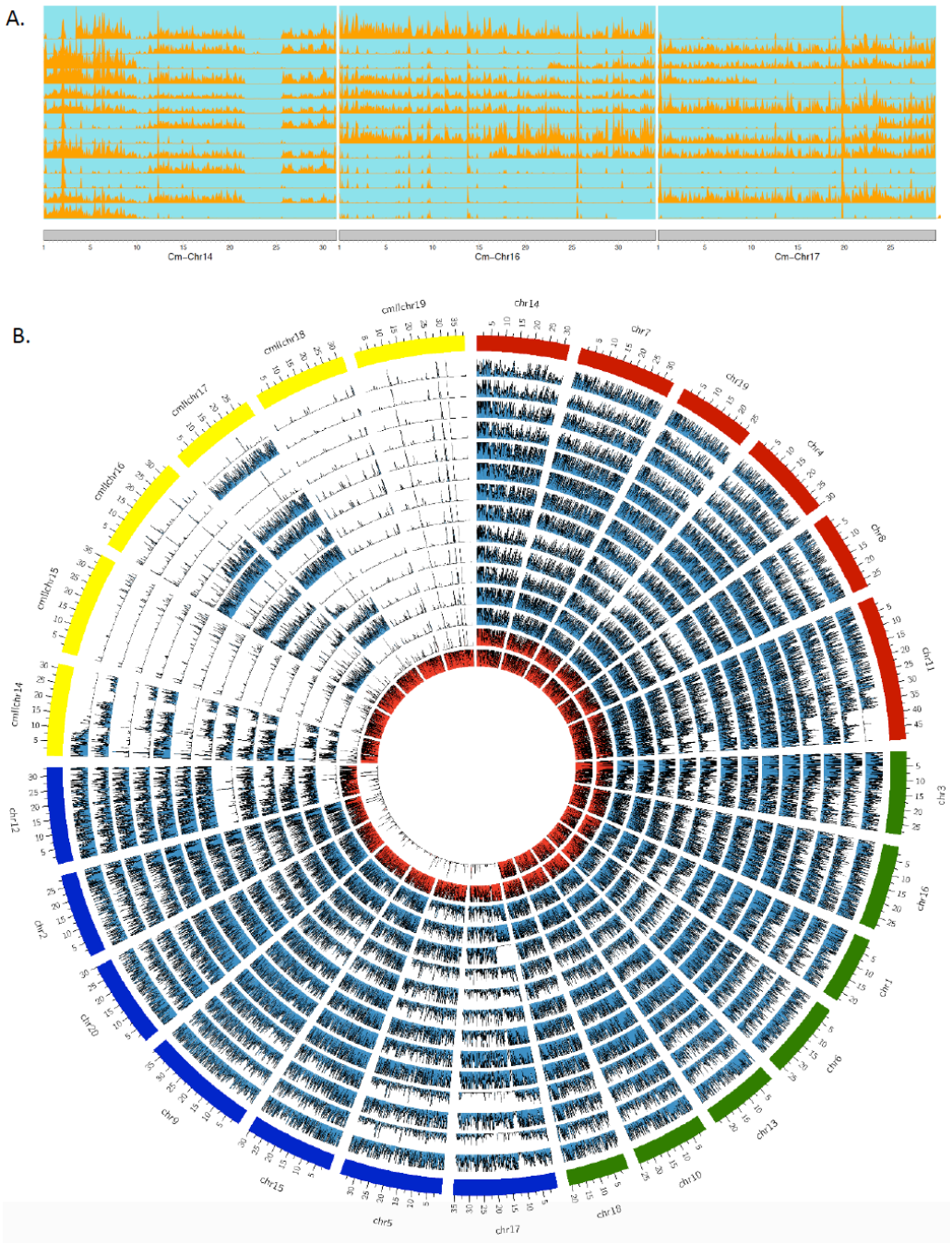


Figure 5.6 Genetic characterization of BC_1F_2 plants derived from TMP23999 (*C. microcarpa*). A) Distribution of mapped reads across three *C. microcarpa* chromosomes shows evidence of aberrant pairing between the chromosomes belonging to the third subgenome of *C. microcarpa.* Here, each row represents a different segregating plant. B) Circos plot showing distribution of reads from 13 different BC_1F_2 plants (*C. sativa* × *C. microcarpa*). Outer circle in red, green, blue represents *C. sativa* genomes, while yellow represents the third subgenome of *C. microcarpa*. Inner blue tracks represent distribution of mapped reads from different segregating plants, and red tracks represent read mapping of the parents, with *C. microcarpa* being the innermost circle.

5.4.4 QTL mapping for vernalization requirement

POP3, the BC_1F_2 derived from the cross (CN119102 ×TMP23992) × PI650132, was used for QTL mapping for vernalization requirement. Accession CN119102 (*C. microcarpa*) was the only source of winter habit for this population, and the BC_1F_1 also showed a winter-type phenotype. In POP3, the growth of most of the plants were similar to the *C. sativa* spring-type and no plants were vernalized to identify the winter-types. Among 97 plants, 54 plants flowered within 100 days of seeding, while the remainder did not flower and were assigned a value of 120 for days to first flower (DTF) for QTL mapping (**Figure 5.7A**). Most of the plants in this population were observed to be semi-sterile or sterile (based on observation, data not shown).

A genetic map was developed with 720 SNPs distributed across 12 chromosomes (**Figure 5.7B**) for 97 plants. Linkage maps could not be developed for chromosome 7 from subgenome 1, chromosome 6 from subgenome 2 and almost all chromosomes belonging to subgenome 3, except chromosome 12. This was likely due to the absence of homologous recombination for the third subgenome, as reflected by monomorphic SNPs identified for this subgenome inherited from PI650132 (*C. alyssum* spring type). Although, there appeared to be some level of recombination for chromosome 12 belonging to the third subgenome of *C. sativa* that resulted in segregation of markers, it was difficult to identify the homoeologous chromosome. The genetic map covers 230.31 Mb of the whole genome with a total genetic distance of 1352 cM (**Table 5.1**), which was nearly 65% of the estimated genome size represented by linkage groups (Total length was 350 Mb for the linkage groups in the physical map). A number of chromosomes lacked terminal portions of the linkage groups, for example chromosome 14, chromosome 19, chromosome 11, chromosome 3, chromosome 16, chromosome 1 and chromosome 12 that might correspond with a higher distal recombination frequency compounded by low polymorphism among the parents.

One major QTL was mapped for vernalization requirement in POP3 with a Logarithm of Odd (LOD) value of 9.03. The QTL was found on chromosome 13 (peak at position 3500655) with a confidence interval of 13 cM (**Figure 5.7C**). This QTL spanned 2.04 Mb in the physical map and represented the terminal region of the genetic map. Due to a lack of markers representing the

terminal part of the chromosomes in physical map, it is possible that the identified QTL is missing some of the genes present in the terminal region; therefore, it was difficult to identify candidate flowering related genes responsible for the vernalization response. However, a major flowering gene *FLC* (Michaels and Amasino 1999) was found 547 Kb away from the QTL peak. Beside this, within the QTL interval another photoperiod related gene, *CIRCADIAN CLOCK ASSOCIATED 1* HIKING EXPEDITION (*CHE)* (Pruneda-Paz et al. 2009) was also found. The identified QTL explained 34.86 % phenotypic variation for the DTF.

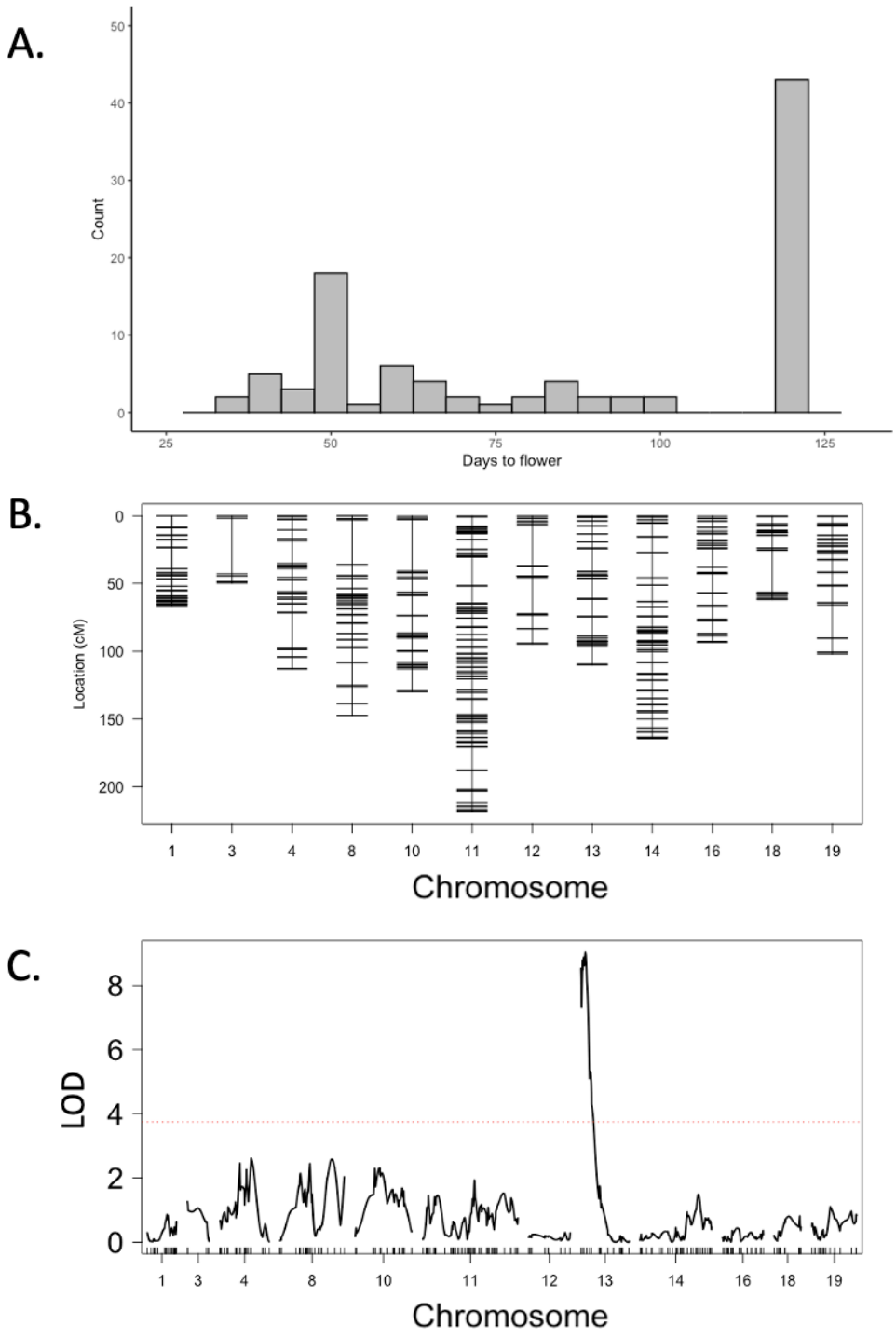


Figure 5.7 A) Distribution of days to flower in segregating population (POP3) generated from *C. microcarpa*; B) Genetic linkage map developed from BC_1F_2 derived from *C. microcarpa*; and C) QTL mapping of vernalization requirement in *C. microcarpa* derived segregating plants. cM distance is shown to the left of the map in Panel (B) and the significance threshold for identifying QTL is shown as a red line in panel (C).

Table 5.1 Details of linkage map construction in segregating BC_1F_2 population derived from *C. sativa* and *C. microcarpa.*

Subgenome	Chromosome	Genetic distance (cM)	No. of SNPs	Total length of chromosome (Mb)	Start position (bp)	End position (bp)	Physical distance (Mb)	Percentage Covered
SG1	Chr14	164.55	102	31.76	10358977	30387171	20.03	63.06
	Chr19	102.18	47	26.74	1648178	16521248	14.87	55.63
	Chr4	112.925	68	30.11	2572008	24367165	21.80	72.38
	Chr8	147.426	63	27.72	4270413	20974328	26.76	96.52
	Chr11	218.48	140	49.70	9545298	47905212	38.36	77.18
		745.56	**420**	166.03			**111.76**	67.31
SG2	Chr3	49.95	12	28.50	9309954	21731193	12.42	43.58
	Chr16	93.456	54	29.11	14600052	27896800	13.30	45.68
	Chr1	66.54	62	23.24	8310832	22833668	14.52	62.49
	Chr13	110.14	52	24.10	2254072	21615961	19.36	80.33
	Chr10	129.72	54	25.32	10566	24943841	24.93	98.48
	Chr18	61.76	35	20.87	6050402	17906794	11.86	56.81
		511.56	**269**	151.14			**96.39**	63.78
SG3	Chr12	94.90	31	33.04	9808946	31965463	22.16	67.05
		1352	720	350.22			230.31	65.76

Similar to the results generated from the previous study (**Chapter 4**), a QTL was found on chromosome 13 that might indicate the same genes are responsible for vernalization requirement in CAM176 (*C. alyssum* winter type) and CN119102 (*C. microcarpa*); however, no QTL were found on chromosome 8 or chromosome 20 for this experiment. Upon closer inspection of SNP haplotypes in the QTL region on chromosome 8 among the segregating individuals, no markers were found to be inherited from the accession CN119102 (*C. microcarpa*). Similarly, in the absence of markers from chromosome 20, it was impossible to link loci from this region with the phenotype data.

5.4.5 Identification of homoeologous recombination

In this study, both parents differed with regards to the third subgenome, introducing an extra subgenome to the genetics. Therefore, it was interesting to observe the nature of recombination for the extra subgenome, which would have no homologous chromosomes. Although there was some aberrant pairing observed in the POP1 F_2 population (**Figure 5.3**), it was difficult to identify homoeologous recombination for individual plants using the sequence reads, as the data represented independent recombination events inherited from both the male and female gametes for each individual plant. However, in the backcross generated population (POP3) homoeologous recombination could be observed by following the pattern of inheritance of alleles from the *C. microcarpa* parent, particularly from the third subgenome, which lacks homologous chromosomes in *C. sativa*.

Mapping of sequence reads from the POP3 individuals to the pseudogenome identified a uniform distribution of mapped reads across chromosome 15 of *C. microcarpa,* which originated from the third subgenome. Based on read depth, the reads that segregated for the absence (AA), presence with a single frequency (AB) and present with a double frequency (BB) were identified on chromosome 7 of *C. sativa* as well as on chromosome 15 of *C. microcarpa,* and since these two chromosomes are homoeologous, these data suggested a homoeologous recombination event between these chromosomes (**Figure 5.8**). This homoeologous recombination represented a cross over event between chromosome 7 of *C. sativa* with chromosome 15 of *C. microcarpa* where a

recombination event might have occurred at approximately 11.4 Mb in *C. sativa* Cs7 and at 7.2 Mb of *C. microcarpa* Cm15.

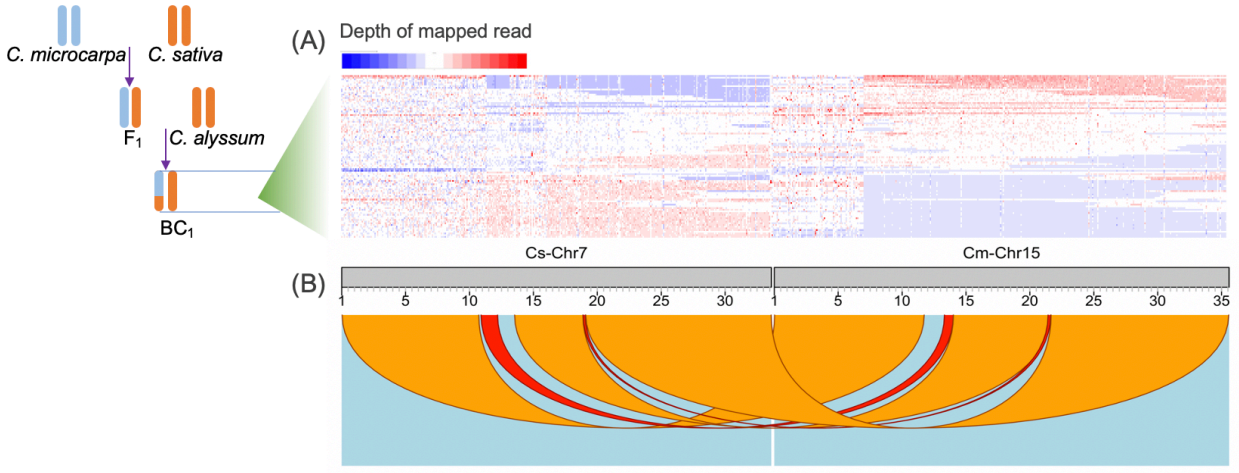


Figure 5.8 Homoeologous recombination between chromosome 7 of *C. sativa* and chromosome 15 of *C. microcarpa*. A) Distribution of scaled mapped reads across 100 kb bins in homoeologous chromosomes, where each row represents an individual segregating plant and left to right represents the length of chromosome; and B) Syntenic analysis between homoeologous chromosomes where red linked ribbons represent inverted regions. The crossing scheme and the expected nature of gametes is shown to the left of the figure.

5.5 Discussion

Disomic inheritance is common for diploid species as well as for allopolyploid species such as *C. sativa.* However, the low level of genetic differentiation among the subgenomes of *C. sativa* has raised the question as to whether disomic inheritance is constant, and also if there is any mechanism present which facilitates normal homologous pairing similar to other allopolyploids, such as wheat (Griffiths et al. 2006). A study by Higgins et al. (2018) has shown that homoeologous recombination is common in *Brassica napus* natural populations, which is also an allopolyploid; however this crop possess a strong pairing control mechanism on chromosome A9 (Higgins et al. 2020). Likewise, synthetic *Brassica* lines have also shown homoeologous recombination (Hurgobin et al. 2018; Zhang et al. 2016). In this context the present study was designed to observe possible homoeologous recombination between *C. sativa* and *C. microcarpa,* which differ by a unique third subgenome.

This study had low hybridization success between *C. sativa* and *C. microcarpa* followed by low fertility and seed set in the segregating generation, similar to previous studies (Tepfer et al. 2020; Zhang and Auer 2020). The hybridization success between these species was confirmed by the winter habit of the hybrid as well as segregation of winter habit in the subsequent generations. This interspecific hybridization was influential in broadening the genetic base of *Camelina* as reflected by variation in leaf morphology, days to flowering, growth habit, seed protein profile etc. However, as expected a high level of sterility was observed in the resulting F_2 population. Nonetheless, fertile plants from the F_2 population as well as backcross derived F_2 plants produced showed intermediate seed size (example in **Appendix A.2.9**) suggesting a higher level of fitness for some of the introgressed plants. Genetic analysis confirmed the aneuploidy status of these plants, where plants that inherited all twenty chromosomes from *C. sativa* with additional chromosomes from the third subgenome of *C. microcarpa* had higher fertility in comparison to partial sterility found in plants lacking chromosomes from the third subgenome of *C. sativa* (based on observations and partial results presented in **Appendix A.2.10**). However, stability of fitness for these plants needs to be tested in further generations. In this experiment the third subgenome of both parents was carefully analyzed for possible homoeologous recombination since they lack homologous chromosomes. In the F_2 segregating population, although reads mapped to the third subgenome for both parents, it was difficult to differentiate between homoeologous exchange (HEs) and deletion, but there were some evidence of aberrant pairing and also possibilities of aneuploidy (**Appendix A.2.10**). Interestingly, although all hybrids originated from the same parents and were grown in the same environmental conditions, the arrangement of chromosomes inherited from the third subgenome of either parental species was different for each segregating plant. For the retained chromosomes, evidence of non-homologous pairing was observed based on non-uniform coverage of mapped reads for parts of the chromosomes, followed by variation in level of heterozygosity (**Figure 5.3**).

To capture homoeologous recombination from interspecific hybridization, a backcross derived segregating population was generated, which also likely helped in the stabilization of recombination events, where segregation of HE segments can be observed, to confirm a homoeologous recombination event. In case of **Figure 5B**, on chromosome Cm-Chr14 there was a marked deletion in the 23-25 Mb region for all plants, which might suggest a HE event during

the generation of the backcross population (BC_1F_1). Similarly, for POP3, a recombination between chromosome 7 of *C. sativa* and chromosome 15 of *C. microcarpa* happened during the gamete formation of BC_1F_1 causing the loss of the terminal 7.2 Mb from *C. microcarpa* which subsequently segregated in the BC_1F_2 generation (**Figure 5.7**).

A good genetic linkage map represents disomic inheritance of markers in a biparental mapping population, whereas segregation distortion leads to spurious linkage in a genetic map. In this study, the genetic linkage map was developed from POP3, where deletion of the terminal region for most of the linkage groups was observed. Similarly, there was loss of linkage groups belonging to the third subgenome of *C. sativa* as well as *C. microcarpa*, except for chr12 belonging to *C. sativa* (**Table 5.2**). Such deletions and incompleteness in the linkage map could be due to higher genetic recombination in the terminal segments, compounded by low levels of polymorphism between the parents. Likewise, in the absence of homologous chromosomes for the third subgenome of both parents, this might have led to aneuploidy and further HEs in these chromosomes would made it difficult to incorporate them in the genetic linkage map. Despite these scenarios, the genetic linkage map enabled the mapping of a major QTL associated with vernalization requirement. The QTL was in close proximity to the major flowering gene *FLC* and overlapped with the result generated from another winter parent *C. alyssum* (**Chapter 4**), suggesting the function of the same genes controlling vernalization requirement has been conserved in *C. microcarpa* and *C. alyssum*. However, this study was unable to identify further previously reported QTLs, which might be due to the incomplete genetic linkage map or absence of variation for the QTL region where a segment from *C. sativa* "spring-type" was incorporated in the gamete formation during BC_1F_1 generation. The results generated showed that the gene responsible for vernalization response in *Camelina* species are conserved, where further functional analysis would be useful to confirm the gene or gene family impacting the vernalization requirement in *Camelina* species.

This study identified homoeologous recombination in *Camelina* species, which is a driving force to generate variability in different species. A mechanism responsible for controlling genetic recombination between orthologs, such as exists in wheat (Griffiths et al. 2006), could not be

inferred. The results suggested, however, a low level of genetic differentiation between subgenomes can result in genetic recombination hotspots which could facilitate homoeologous recombination in related *Camelina* species. The level of variation generated by this interspecific hybridization provides opportunities for the *Camelina* breeding program, as well as the ability to explore and identify genes underlying important traits present in the wild relative *C. microcarpa*. Some plants have shown stability for seed production in F_2 and BC_1F_2 populations and these could be interesting to study in the further generations.

Prologue to chapter 6

From chapters 3, 4, and 5, it was clear that two different hexaploid genomes were present among *Camelina* species, one with n = 20 and the other with n = 19. The nature of hybridization between these species as well as mapping of sequence reads, suggested a different subgenome structure for these species. Therefore, the following study was conducted to observe subgenome dominance in these hexaploid species, and to identify the subgenome with dominant expression pattern as a consequence of whole genome duplication.

This chapter will be a part of a genome paper describing *C. microcarpa* “Type 2”.

Chapter 6. Age of divergence among subgenomes determines gene expression between orthologs in *Camelina* species

6.1 Abstract

Almost every plant has undergone whole genome duplication and subsequently evolved to a diploidized state. The availability of different ploidy species in *Camelina* has facilitated an understanding of the evolutionary progression in *Camelina* and the fate of duplicated genes. Transcriptomic data were generated from *Camelina* species representing diploid, tetraploid, and two different types of hexaploid genomes to study subgenome (SG) dominance. Analysis of the rate of synonymous substitutions (Ks) was performed to infer the age of divergence of the subgenomes among the hexaploid *Camelina* species with respect to diploid *C. neglecta*. The results suggested a linear relationship of subgenome dominance with the age of divergence of subgenomes from the diploid *C. neglecta*. The level of fractionation among subgenomes was low in these relatively new polyploids; however, it was also found to be dependent on the age of divergence. In *C. sativa* the third subgenome has retained the most expressed genes and was also dominant among other subgenomes. The Ks analysis suggested a case of SG1, SG3′ and SG2 subgenome progression in hexaploid *C. microcarpa* "Type 2", in comparison to SG1, SG2 and SG3 subgenome progression in *C. sativa*. Overall, the subgenome dominance and the higher level of expression of duplicated genes was found to be associated with the adaptation process, and has a linear progression with the age of divergence for the subgenomes studied.

Keywords: age of divergence, subgenome dominance, fractionation

6.2 Background

Genome duplication has played a major role in the adaptation of plants (Crow and Wagner 2005; Renny-Byfield and Wendel 2014); however, after such events the function and maintenance of duplicate genes remains uncertain (Schnable et al. 2011). Duplicated genes tend to have a similar expression pattern as their ancestral genes; however, as reported for a number of different polyploid plants some gene copies may lose or gain expression (Birchler and Veitia 2012), and this can be biased leading to the overall dominance of one subgenome in the polyploid. The phenomenon of subgenome dominance has been reported in a number of polyploids including

Camelina sativa (Edger et al. 2019; Ramírez-González et al. 2018; Liu et al. 2014; Kagale et al. 2016). Genome dominance has also been observed in newly resynthesised polyploids (Edger et al. 2017; Wu et al. 2018), which indicated the biased expression of homologues occurs soon after genome duplication, and could be dependent on the dominant nature of either parent used in the hybridization (Bird et al. 2019). Therefore, evidence of subgenome dominance can be utilized for evolutionary studies to decipher the nature of the ancestral genomes and their relationship upon merger. As subgenome dominance is likely age-related and is a continous process, it could lead to important genes related to the domestication and adaptation process being identified. Since dominance can be inferred from gene expression data, exploration of such data at different ploidy levels within closely related species could help to understand the evolutionary processes determining the establishment of the species. The main cause of subgenome dominanace has yet to be revealed; relative methylation levels, amount of Transposable Elements (TE) across the genome and structural changes after genome merger, have all been suggested as having a role in shaping gene expression in polyploids (Bird et al. 2018). Among these factors, the role of TE has often been discussed as potentially impacting subgenome dominance in allopolyploids (Alger and Edger 2020; Hollister and Gaut 2009). But there are exceptions, with a few species which have not undergone fractionation after genome duplication such as Cucurbits, Soybean and *Capsella bursa-pastoris* (Zhao et al. 2017; Sun et al. 2017; Douglas et al. 2015).

Camelina sativa is an allopolyploid which has been formed by the hybridization of similar species generating a hexaploid with three undifferentiated subgenomes (Kagale et al. 2014a). The identification of *C. neglecta* as a progenitor of the first subgenome of *C. sativa*, and the suggestion that *C. hispida* is the potential progenitor for the third subgenome, has shed some light on the evoultionary history of *C. sativa* (Chaudhary et al. 2020; Mandáková et al. 2019). Likewise, a *C. microcarpa* with a tetraploid structure comprising the first and second subgenomes has also been reported (Chaudhary et al. 2020). This could suggest a step wise merger of species from diploid (n = 6) to tetraploid (n = 13) and than hexaploid (n = 20), as found in *C. sativa* and *C. microcarpa*. The identification of *C. microcarpa* "Type 2" (Chaudhary et al. 2020) with n = 19 has suggested an alternate route for the formation of higher ploidy *Camelina* species where three subgenomes with 6, 7 and finally 6 chromosomes were merged. The *C. microcarpa* "Type 2" shares two subgenomes with *C. sativa*; however, the third

subgenome is different, suggesting different progenitors for the third subgenome. The third subgenome of *C. microcarpa* "Type 2" shares some homology with the first subgenome, which was supported by evidence of homoeologous recombination between chromosomes from the first subgenome of *C. sativa* with chromosomes from the third subgenome of *C. microcarpa* "Type 2" (**Chapter 5**).

After correcting the subgenome structure of *C. sativa*, the reported dominance of the third subgenome of *C. sativa,* with the highest gene expression among the three subgenomes in *C. sativa,* was more pronounced (Kagale et al. 2016; Chaudhary et al. 2020). Likewise, the availability of diploid, tetraploid and hexaploid(s) species in *C. sativa* has increased the opportunities for studying subgenome dominance in *Camelina.* Such analyses could help to elucidate the impact of adding genomes to the nucleus, moving from the diploid to the tetraploid and finally the hexaploid state. The current transcriptomic study of different ploidy levels in *Camelina* has shed light on the evolutionary progression in *Camelina,* where addition of subgenomes was associated with a linear progression in increased levels of fitness, as shown by reduced gene loss, and correlated with the level of dominance governed by the subgenomes. In addition, subgenome gene fractionation was found to be continous and dependent on the age of divergence of the subgenomes.

6.3 Materials and methods

6.3.1 Plant materials

Ten genotypes consisting of diploid, tetraploid and hexaploid species of *Camelina* were utilized in this study. Among these, one genotype was hexaploid *C. microcarpa* "Type 1" (TMP24026), three genotypes were hexaploid *C. microcarpa* "Type 2" (CN119103, CN115248 and TMP23999), three genotypes were hexaploid *C. sativa* (DH55, TMP23992 and TMP23986), one genotype was hexaploid *C. alyssum* (CAM176)*,* one was tetraploid *C. microcarpa* (CN119243) and one was diploid *C. neglecta* (TMP24028) (**Appendix A.3.1 and A.3.2**). The three *C. sativa* genotypes were spring-types and the remainder were winter type. For one winter-type genotype, CN115248, vernalized leaf samples were collected after 30 days of cold treatment (4 ^{o}C). For all

other samples, RNA extraction was performed with leaf samples from seedlings grown in cyg seed germination pouches (Mega international, Newport, MN 55055, USA) for one week. Three biological replications for each sample were included in the RNA extraction process.

6.3.2 RNA sequencing and sequence analysis

RNA extraction, RNA library preparation and RNA sequencing were done as described in **Chapter 4**. Raw sequences were filtered prior to mapping to the genome using Trimmomatic (Bolger et al., 2014) as described in **Chapter 4**. All read mapping was performed with STAR aligner (Dobin et al., 2013) using default parameters except for *–alignIntronMax* 10000 and *–outFilterMismatchNmax* 4 using annotated genomes of each *Camelina* species. Trimmed reads belonging to *C. sativa*, *C. alyssum* (CAM176) and *C. microcarpa* "Type 1" (TMP24026) were aligned with the annotated DH55 *C. sativa* reference genome (Kagale et al., 2014). Diploid *C. neglecta* (TMP24028) and tetraploid *C. microcarpa* (CN119243) were also aligned with the annotated DH55 *C. sativa* reference genome (Kagale et al., 2014); however, only the first subgenome was used for the diploid, while the first and second subgenomes formed the reference for the tetraploid. Similarly, hexaploid *C. microcarpa* "Type 2" (TMP23999, CN119103 and CN115248) were aligned with the annotated TMP23999 reference genome (unpublished). The *GeneCounts* feature in STAR provided a count of the number of transcripts per annotated gene, which was converted into Fragment Per Kilobase of transcripts per Million mapped reads (FPKM) for downstream analysis. Differential gene expression analysis was performed with count data using "DESeq2" (Love et al. 2014) combined with the log fold change shrinkage function using the "apeglm" package (Zhu et al. 2019) in R statistical software (Team 2019). Functional enrichment analysis and identification of gene families for the highly expressed genes were performed using the online version of GenFam (Bedre and Mandadi 2019).

6.3.3 Estimation of age of divergence

Syntenic genes from reference genomes belonging to hexaploid *C. sativa*, hexaploid *C. microcarpa*, and diploid *C. neglecta* were first identified by sequence homology to genes from *Arabidopsis thaliana* identified using BLAST (Altschul et al. 1990). Those showing conserved synteny were curated using DAGchainer (Haas et al. 2004). After developing a synteny table (**Appendix A.3.3**) across all genomes, the age of divergence for the subgenomes of hexaploid *C.*

sativa and *C. microcarpa* were calculated based on orthologs from the *C. neglecta* diploid genome. Rate of synonymous substitutions (Ks value) for each gene-pair was calculated using GenoDup pipeline (Mao 2019). The Gaussian mixture model was used from R package mclust (Scrucca et al. 2016) to plot distribution of Ks and identify number of components. Based on these components, the mean value of Ks distribution was identified and used to calculate the age of divergence as described by Kagale et al. (2014a).

6.3.4 Subgenome dominance analysis

Subgenome dominance was analyzed for genotypes belonging to hexaploid *C. sativa* (reference DH55) separately from hexaploid *C. microcarpa* "Type 2" (reference TMP23999). Based on syntenic analysis with *A. thaliana,* orthologs in the subgenomes were identified for comparison. Genes expressed at less than 0.01 FPKM in any of the replications or for any of the orthologs were discarded. Using the three replications, for each set of orthologs genes in the hexaploids, analysis of variance was performed with a custom script to observe differences in expression patterns between the subgenomes. The genes were placed in three categories: 1) Balanced, showing no difference in expression among orthologs; 2) Dominant, showing higher expression in comparison to the other two orthologs; and 3) Suppressed, showing lower expression in comparison to the other two orthologs. More than 11,000 genes were studied for all the samples belonging to hexaploid *Camelina* species, and the difference in the number of genes representing each category was determined using an analysis of variance test with a significant threshold of *P-value*<0.05.

6.4 Results

6.4.1 Expression of genes across different ploidy levels

Since gene expression varies with tissue type and growth stage, all leaf samples were collected from the early seedling stage, and the analyses were performed with genes with a minimum expression of at least 0.01 FPKM, which suggested functional genes. The results showed that the hexaploid genotypes had around 50,000 expressed genes (56-69% of annotated genes), whereas there were 38,290 expressed genes in the case of the tetraploid species and 18,410 in the diploid species (**Table 6.1**). The results also suggested that a higher number of expressed genes were

found in *Camelina* species with 20 chromosomes in comparison to those with 19 chromosomes. The number of expressed genes in *C. neglecta* was comparable to the first subgenome of *C. sativa*; however, was higher when compared to the first subgenome of *C. microcarpa* "Type 2", suggesting a number of genes are not expressed from the first subgenome of *C. microcarpa* "Type 2". In the case of *Camelina* species with 20 chromosomes, a significantly higher number of genes were expressed/retained in the third subgenome, whereas a lower number of genes were expressed in the second subgenome (**Table 6.1**). However, in the case of *C. microcarpa* "Type 2", all the subgenomes had a similar number of expressed genes. In the case of tetraploid *C. microcarpa*, there was a significantly lower number of expressed genes in the second subgenome compared to the first subgenome. These results suggested that *Camelina* species with 20 chromosomes had an unequal number of expressed genes among the subgenomes, whereas *Camelina* species with 19 chromosomes had a stable number of expressed genes across the subgenomes (**Table 6.1**).

Table 6.1 Total number of expressed genes (≥0.01 FPKM) in different genotypes and different subgenomes, for all biological replicates.

Reference	Species	Genotype	Total genes[1]	Subgenome 1	Subgenome 2	Subgenome 3	Chi-Square test[2]
DH55	*C. sativa*	DH55	56665 (66.5)	18807 (66.8)	18412 (68.2)	19446 (64.6)	28.83***
		TMP243986	58911 (69.1)	19519 (69.3)	19086 (70.7)	20306 (67.4)	38.96***
		TMP23992	58091 (68.1)	19251 (68.4)	18899 (70.0)	19941 (66.2)	29.02***
	C. microcarpa	TMP24026	52116 (61.1)	17274 (61.3)	17040 (63.1)	17802 (59.1)	17.54***
	C. alyssum	CAM176	52939 (62.1)	17523 (62.2)	17284 (64.0)	18132 (60.2)	21.67***
TMP23999	*C. microcarpa* "Type 2"	CN119103	52430 (59.1)	17500 (58.4)	17340 (60.4)	17590 (58.6)	1.83[ns]
		TMP23999	50396 (56.8)	16810 (62.0)	16746 (58.4)	16840 (56.1)	0.27 [ns]
		CN115248	53633 (60.5)	17908 (70.1)	17782 (62.0)	17943 (59.7)	0.80 [ns]
DH55-SG1-SG2	*C. microcarpa*	CN119243	38290 (69.4)	19364 (68.8)	18926 (70.1)		
DH55-SG1	*C. neglecta*	TMP24028	18410 (65.4)	18410 (65.4)			

[1]The number in parantheses indicates percentage of total annotated genes.

[2]Chi-Square test for equivalent number of expressed genes in each subgenome.

*** represents *P-value*<0.001 and [ns] represents non-significant.

The expressed genes representing different genotypes from the same species were further analysed to identify the proportion of expressed genes common to all genotypes, which could remove potential bias in the generated results. More than 46,648 expressed genes were common among the 5 genotypes belonging to the reference genome DH55-type (n=20), and 942 genes or less were unique to individual genotypes. Similarly, 45,586 expressed genes were common among the three genotypes belonging to *C. microcarpa* "Type 2", with slightly higher numbers of genes unique to each genotype (1,288 - 2,627) (**Figure 6.1**).

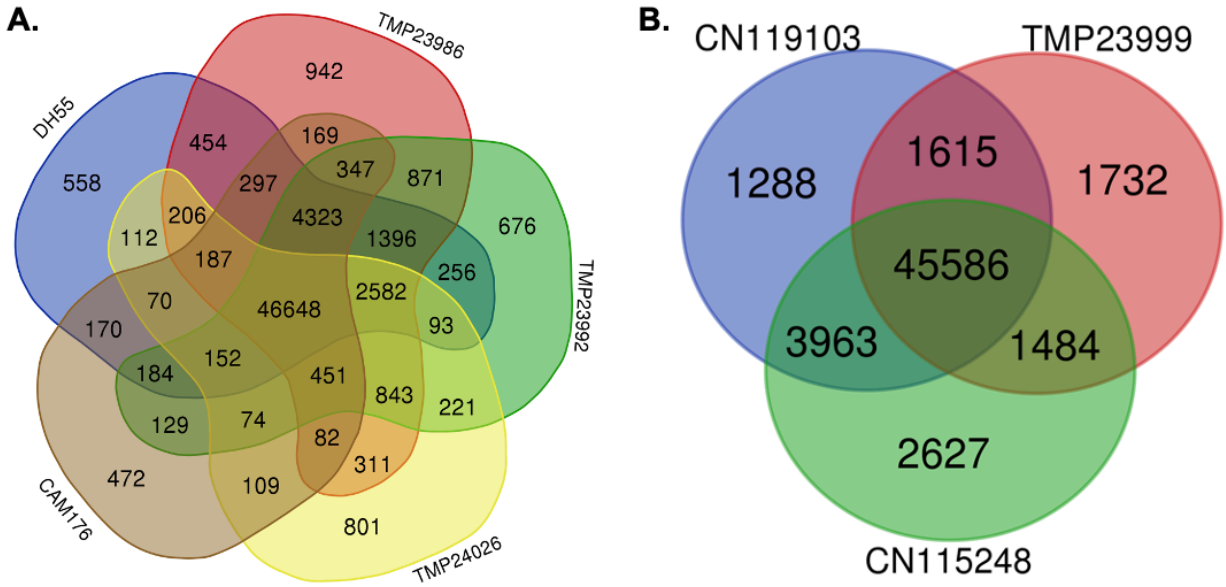


Figure 6.1 Venn diagram showing numbers of shared and specific expressed genes across the different genotypes belonging to: A) hexaploid *C. sativa* (n=20); and B) hexaploid *C. microcarpa* "Type 2" (n=19).

6.4.2 Comparison of gene expression among orthologs of different ploidy *Camelina* species

This analysis was performed to identify those genes which potentially increased their expression levels during or post genome duplication. Samples from diploid *C. neglecta*, tetraploid *C. microcarpa* and hexaploid *C. sativa* were mapped to a common reference genome, *C. sativa* DH55. The analysis suggested that a slightly higher number of genes were more significantly expressed (*FDR*<0.05) in the diploid compared with the tetraploid first subgenome, whereas this

number was similar when compared with the hexaploid first subgenome. Likewise, a slightly higher number of genes were more significantly expressed (*FDR*<0.05) in the hexaploid compared to the tetraploid for the first and second subgenomes (**Table 6.2**). The number of more highly expressed genes (log2fold change) was greater for the tetraploid and hexaploid, in comparison to the diploid for the first subgenome of *C. sativa* (**Table 6.2**). The number was also higher for the hexaploid compared to the tetraploid, which suggested that levels of gene expression in the first subgenome of *C. sativa* has increased as the ploidy level increased.

Table 6.2 Number of differentially expressed genes across *Camelina* species with different ploidy levels: diploid TMP24028, tetraploid CN119243 and hexaploid DH55.

Sub-genome	Diploid vs Tetraploid		Diploid vs Hexaploid		Tetraploid vs Hexaploid	
	TMP24028	**CN119243**	**TMP24028**	**DH55**	**CN119243**	**DH55**
SG1	6401 (3754)*	6083 (3777)	3763 (1730)	3771 (2329)	6424 (2775)	6825 (3012)
SG2					6318 (2732)	6710 (2865)

*Indicates number of genes with ≥2-fold higher difference

Genes which had higher expression after whole genome duplication in the tetraploid and hexaploid, which could have a role in the adaptation process were identified. The tetraploid possessed 3777 genes in the first subgenome, which were more highly expressed (*FDR*<0.05) in comparison to the diploid, this was reduced to 2329 (*FDR*<0.05) when the hexaploid was compared with the diploid genotype. Among these genes 1358 were common, whereas 2419 were unique to the tetraploid, and 971 were unique to the hexaploid (**Figure 6.2A**). The genes showing higher expression in the hexaploid were enriched (*P-value*<0.05) with: protein kinase, IQD, papain-like cysteine proteases, NB-LRR, cysteine-rich receptor like kinases, C2H2 zinc finger, dirigent protein, glycosyltransferase 58, GDPD, NPH3, Expansin, CAF1 etc. gene families (**Figure 6.2B, Appendix A.3.4**). These gene families are responsible for plant development, response to stress, defense, development during seedling stage, leaf expansion and growth, cell wall proliferation, and transcription factors related; all of which could play important roles in overall plant growth and development.

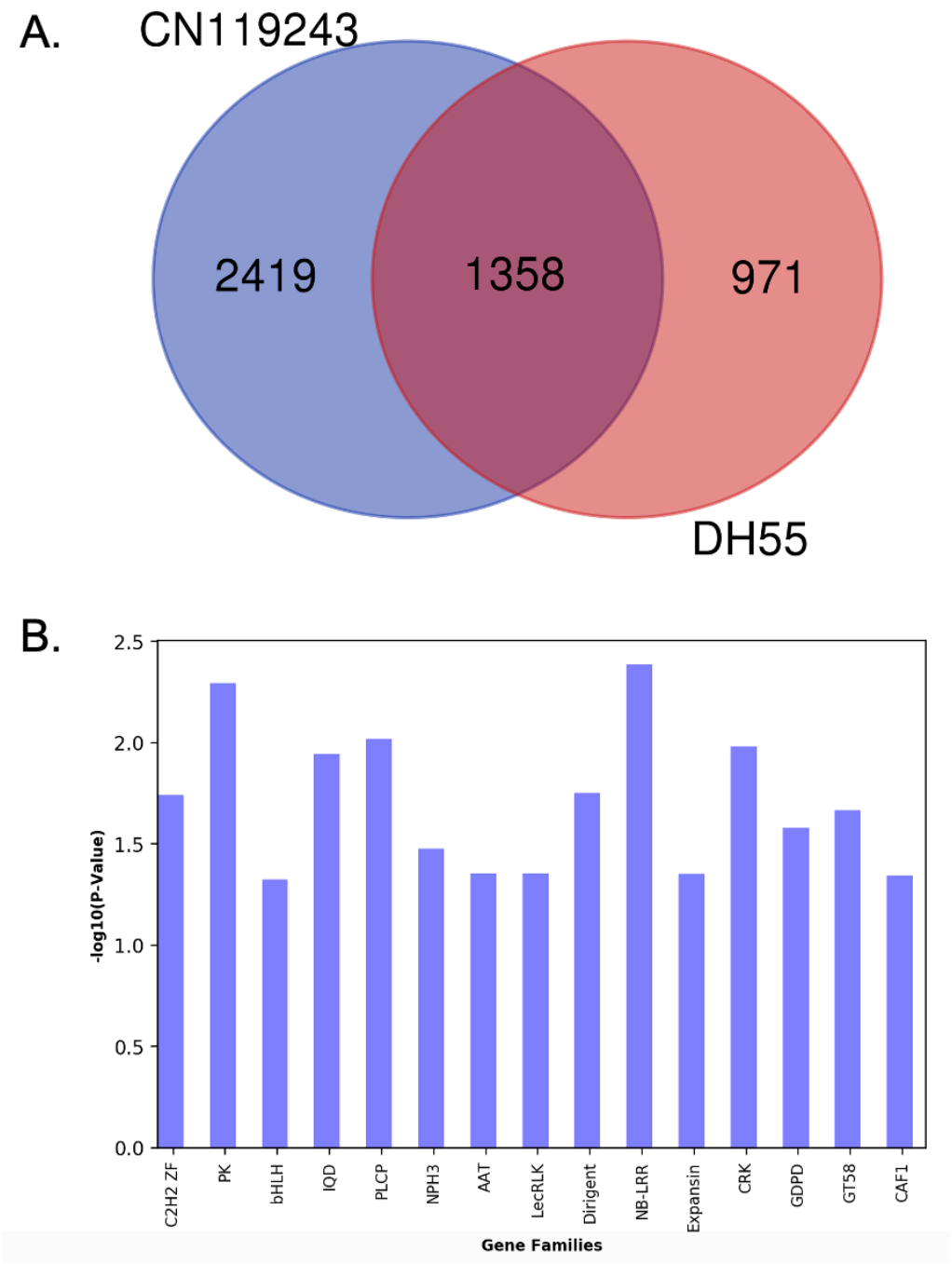


Figure 6.2 Comparison of differentially expressed genes across different ploidy *Camelina* species. A) A Venn diagram of the number of unique and common differentially expressed genes in the tetraploid (CN119243) and hexaploid (DH55) compared with diploid *Camelina* species; and B) Gene families showing higher expression in the first subgenome of hexaploid *C. sativa* in comparison to tetraploid and diploid species. The higher the -log10(P-value), the greater the confidence in enrichment of the particular gene family.

6.4.3 Age of divergence of the subgenomes in *Camelina* species

Only one diploid (*C. neglecta*) species of *Camelina* has been identified. The diploid species shared similar structure with the first subgenome of higher ploidy *Camelina* species; therefore, in this study the diploid species was utilized to identify the age of divergence of the subgenomes encompassed by hexaploid *C. sativa* and *C. microcarpa* "Type 2". Such allopolyploids are expected to form after hybridization of lower ploidy species; this analysis suggests the age of divergence of the progenitor species, representing subgenomes in the hexaploids. Ortholog gene pairs were identified from the hexaploid *C. sativa* and hexaploid *C. microcarpa* "Type 2" with reference to *C. neglecta,* using syntelog tables developed based on *A. thaliana* orthologs (**Appendix A.3.3**). The rate of synonymous substitutions per site (Ks) were calculated between each gene pair and used to estimate the age of divergence of each subgenome. The results suggested that the third subgenome (SG3) from *C. sativa* had diverged 5.1 million year ago (mya) from diploid *C. neglecta*, and was the earliest among the subgenomes of *C. sativa*. In the case of *C. microcarpa* "Type 2", the second subgenome (SG2) diverged earlier (~4.3 mya) from *C. neglecta* in comparison to the third subgenome (SG3'). The age of divergence for the second subgenome, representing both hexaploid *Camelina* species, was similar which might imply that the progenitor for the second subgenome was the same for both species. Likewise, the first subgenome for both hexaploids did not show any peak in the Ks distribution, when compared with *C. neglecta.* This indicated little differentiation between *C. neglecta* and the first subgenome of both hexaploid *Camelina* species, which might be expected if *C. neglecta* was the progenitor species for the first subgenome of both species. The third subgenome (SG3') from *C. microcarpa* that has only 6 chromosomes appeared to have diverged around 3.4 mya from *C. neglecta,* which was latest in comparison to second and third subgenome from *C. sativa* (**Figure 6.3**). Consequently, the order of divergence of subgenomes from the diploid progenitor species, with most similar first, would be SG1, SG3', SG2 and SG3.

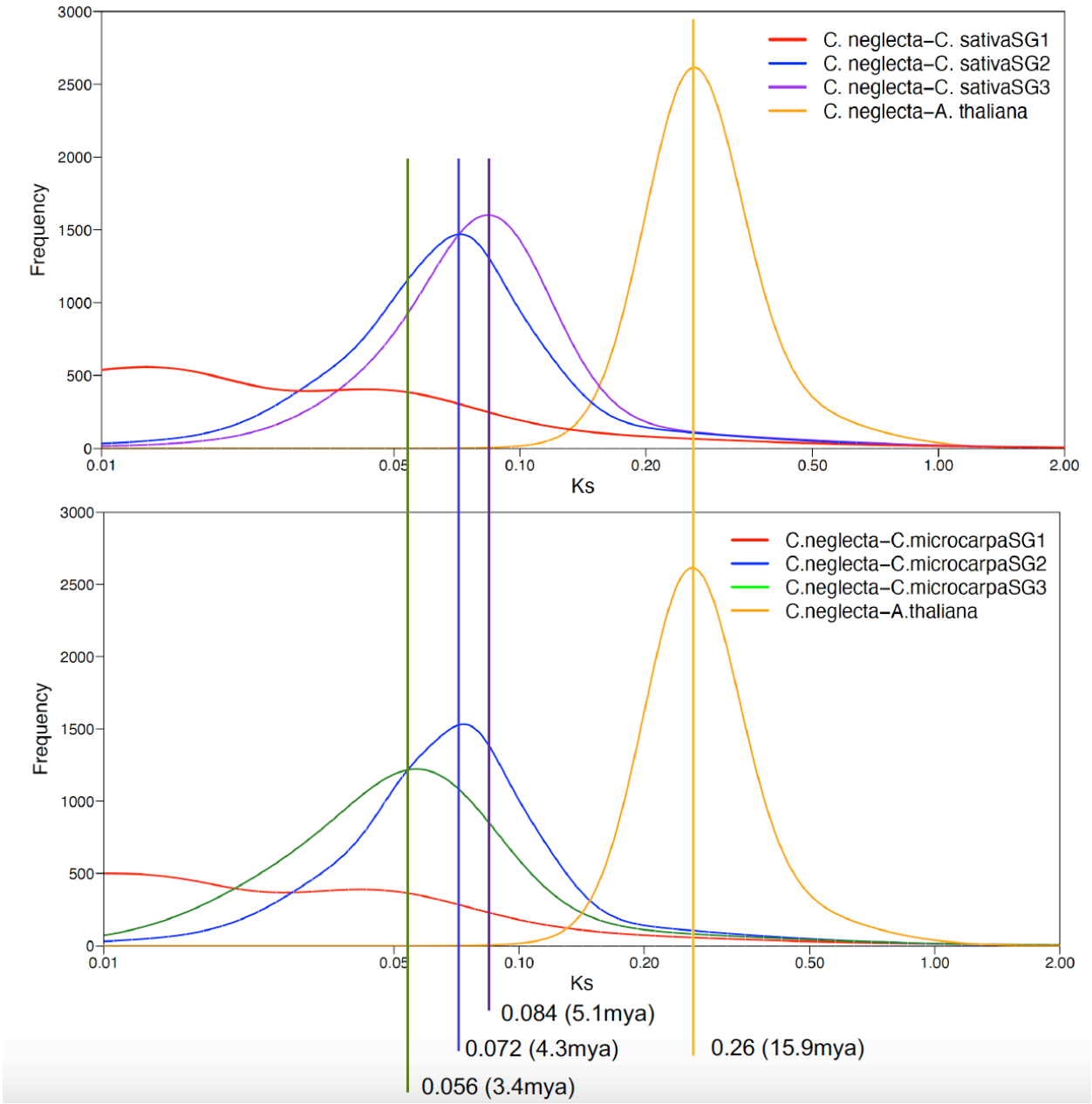


Figure 6.3 Age of divergence for subgenomes in *Camelina* species. The plots represent distribution of Ks for the three subgenomes of *C. microcarpa* "Type 2" (top) and *C. sativa* (bottom). The distribution of Ks from *A. thaliana* is represented by orange line in both plots. The estimated age(s) of divergence calculated from the peaks in the distribution are shown at the bottom of the figure.

6.4.4 Subgenome dominance among *Camelina* species

Hexaploid *Camelina* species were analyzed for subgenome dominance where triplicates of *A. thaliana* orthologs were identified (**Appendix A.3.3**) and tested for variance in gene expression. More than 10,812 triplicates had expression levels of at least 0.01 FPKM for all hexaploid accessions and were analysed for subgenome dominance. The third subgenome was found to be dominant for all genotypes with a *C. sativa* genome structure, whereas the second subgenome was dominant in genotypes with a *C. microcarpa* "Type 2" genome structure (**Figure 6.4**). For all genotypes a lower proportion of genes were suppressed for the third subgenome (<5.3%) (**Appendix A.3.5**). Here, the first and second subgenomes were similar in structure for both type of hexaploid *Camelina* species; however, the third subgenome differs in genome structure as well as chromosome number.

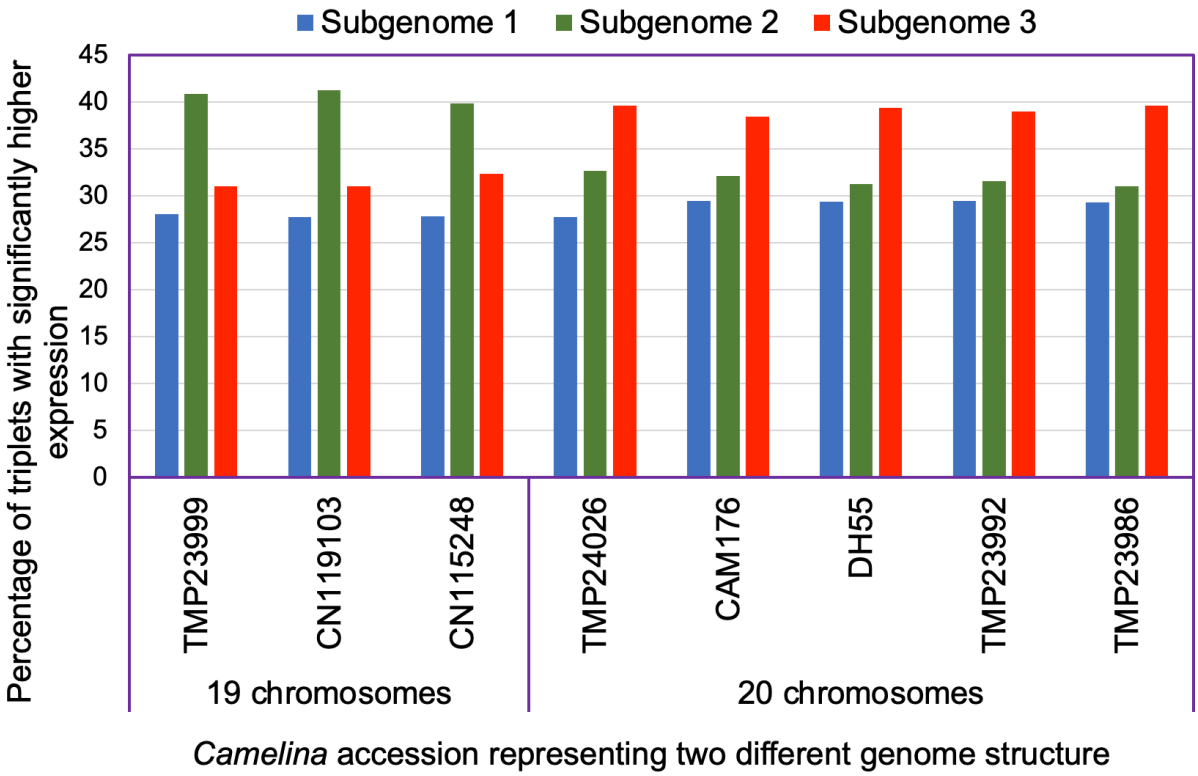


Figure 6.4 Subgenome expression dominance in different *Camelina* species. The number of genes showing higher expression (F-test) in one subgenome in comparison to the other two subgenomes are presented as a percentage of all genes tested. The 19 chromosome accessions have *C. microcarpa* "Type 2" genome structure whereas the 20 chromosome accessions represent *C. sativa* genome structure.

6.5 Discussion

Normal chromosome pairing behaviour in *C. sativa* confirmed the species was a stable allopolyploid (**Chapter 3**), and suggested *C. sativa* resulted from hybridization events of lower ploidy *Camelina* species, potentially with progenitors with low levels of genomic differentiation as assumed previously (Kagale et al. 2014a). Diploid *C. neglecta* is the only species identified to date that could be the progenitor of the first subgenome of *C. sativa* (Chaudhary et al. 2020; Brock et al. 2019; Mandáková et al. 2019), and there are no reports of progenitors for the other subgenomes. The similarity shared by diploid *C. neglecta* and tetraploid *C. microcarpa* with the first two subgenomes of *C. sativa* have suggested the potential step-wise hybridization events that occurred during *C. sativa* evolution (Chaudhary et al. 2020). This suggested evolutionary pathway was assumed for this study to identify genes which have shaped the evolution of *C. sativa*, and to study subgenome dominance upon the merger of subgenomes, as found in other polyploid species (Edger et al. 2019; Schnable et al. 2011; Douglas et al. 2015; Bird et al. 2019).

Genome duplication through hybridization doubles the number of genes; however, due to various events post-duplication changes to gene function can lead to loss of gene expression and finally fractionation, with a number of genes lost over time (Freeling et al. 2015). The current study identified changes in gene expression among subgenomes of accessions representing the *C. sativa* genome structure. However, a wild relative of *C. sativa*, termed *C. microcarpa* "Type 2" did not show the expected pattern, instead the number of expressed genes across subgenomes was stable. The number of expressed genes among subgenomes were tested for equivalence (1:1:1 ratio) using the chi-square test. It should be noted that although this tested if the same number of genes were expressed across all three subgenomes, it did not test if ortholog genes were expressed across the subgenomes. In the case of *C. sativa*, the differences in number of expressed genes suggested that the orthologs across the subgenomes were not expressed in a similar manner. For example, the third subgenome of *C. sativa* tends to have a higher number of expressed genes in comprison to diploid *C. neglecta* which might suggest neofunctionalizaton in the third subgenome. The mapping was relatively low for *C. neglecta* in the first subgenome of *C. sativa* (**Appendix A.3.1)** in comparison to the tetraploid and hexaploid species, which could be related to homoeologous exchanges between *C. neglecta* and the other subgenomes, or more

likely loss of genes from the first subgenome of *C. sativa* upon merger of genomes in the formation of hexaploid *C. sativa*. In the other hexaploid *C. microcarpa* "Type 2" genome, the number of expressed genes were similar across the subgenomes, which might suggest low levels of fractionation in the subgenomes and by inference the orthologs in all three subgenomes might be expected to have maintained a similar function. The reference genome of these species confirmed this result; for instance, the total number of genes identified in the first subgenome of *C. sativa* compared to the diploid *C. neglecta* were low; however, the number of genes identified were comparable between *C. neglecta* and the first subgenome of *C. microcarpa* "Type 2". Also, the subgenomes of *C. microcarpa* "Type 2" have lower variation for number of genes compared to the subgenomes belonging to *C. sativa* (**Appendix A.3.6**).

Analysis of synonymous substitution rate per site has widely been used to infer the age of genome divergence. In *C. sativa,* comparison of the subgenomes with diploid *C. neglecta* suggested the third subgenome (SG3) diverged earlier than the second subgenome (SG2), while the first subgenome (SG1) showed virtually no divergence from *C. neglecta*. This study also suggested the second subgenome (SG2) of *C. microcarpa* "Type 2" diverged earlier than the third subgenome (SG3'). In the context of gene expression, the earliest diverged third subgenome from *C. sativa* had a higher number of expressed genes, while the other two subgenomes possessed a similar number of expressed genes. This suggested that the age of divergence might have shaped the pattern of gene expression in the progenitors of these subgenomes. However, this analysis did not answer which genome had hybridized first in the formation of hexaploid *Camelina* species. Although, the higher number of expressed genes in the third subgenome suggested it might have been added later to the tetraploid structure formed by the combination of the first and second subgenome (as of tetraploid *C. microcarpa* in this study), as suggested also by Mandáková et al. (2019). This path failed to describe the hybridization steps that formed *C. microcarpa* "Type 2", which has a completely different subgenome, that more recently diverged from the *C. neglecta* genome (**Figure 6.3**). *Camelina microcarpa* "Type 2" also possessed similar numbers of expressed gene across the subgenomes, where subgenome SG3' appeared to be inbetween SG1 and SG2 with regards to age of divergence from *C. neglecta.*

Subgenome dominance has been a common phenomenon in allopolyploid evolution, and has a major role in the diploidization process of plants. The current study was designed to compare subgenome dominance in hexaploid *Camelina* species differing by subgenome structure. The level of dominance of genes has been reported to vary with tissue type in *C. sativa* (Kagale et al. 2016), which was not included in this study, instead comparisons were made for a single tissue/stage. Study of subgenome dominance demonstrated that the third subgenome was dominant in the case of the *C. sativa* genome structure, as in previous reports (Chaudhary et al. 2020; Kagale et al. 2016); whereas the second subgenome was dominant in the case of the *C. microcarpa* "Type 2" genome structure. The fact that in both hexaploid genomes the first and second subgenomes were the same in terms of genome structure, the dominant nature of second subgenome in *C. microcarpa* "Type 2" but not in *C. sativa*, suggested the dynamic nature of subgenome dominance. Thus, the assumption that the dominant nature of the subgenome was already determined by the progenitor species, as described by Edger et al. (2017), could not be justified as reflected by the disparity shown by the second subgenome in these two hexaploid genomes. In the case of the *C. sativa* genome, the level of fractionation was associated with the subgenome dominance. For example, the third subgenome had retained a higher number of expressed genes and was also dominant. The dominance of the third subgenome in *C. sativa* might be responsible for retaining higher number of expressed genes due to higher phenotypic expression, as a result of masking/deletion of genes from the lower expressed subgenome (Renny-Byfield et al. 2017), which was also shown by the nature of differentially expressed genes from the *C. sativa* in comparion to diploid and tetraploid (**Appendix A.3.4**). A linear relationship between the age of divergence and the dominance of the subgenome was discovered in this study that could suggest the early speciation led to the retention of higher expression during the evolution process.

Chapter 7. General discussion, conclusion and future directions

In this thesis, I have conducted experiments in *C. sativa* with the intent of improving the current genetic diversity to facilitate *C. sativa* breeding. In doing so the results have led to novel discoveries and provided insights into genome evolution of hexaploid *C. sativa*. The activities were focused on: i) surveying the available diversity and identifying the genetic relationships among different genotypes and species from a world collection; ii) understanding the nature of genetic recombination between species and subspecies with a focus on identifying useful loci controlling flower initiation; and iii) studying the impact of subgenome dominance which has evolutionary significance in the adaptation of the polyploid *C. sativa*.

This study started with the analyses of genetic diversity among 193 accessions of camelina, representing a number of different species. Since this crop is self-pollinated, all the accessions were assumed to be naturalized in their place of collection. The results obtained did not find any sub-population differentiation as a result of geographical isolation, rather there was a higher differentiation among type of species. For example, *C. microcarpa* was distantly related with *C. sativa,* whereas two subpopulations of *C. sativa* were closely related, which was also reported previously (Luo et al. 2019a; Vollmann et al. 2005; Singh et al. 2015). The low level of polymorphism found among the *C. sativa* lines suggested that the subpopulations might be the result of breeding activities performed some time before their distribution across the collection points. The most novel aspect of this study was an enhanced understanding of the taxonomic variability within the genus based on the inclusion of related species, which provided insights into the potential evolution of the *C. sativa* genome. The genome structure of polyploids formed from highly similar lower ploidy genomes can often make the resolution of evolutionary relationships and the identification of progenitor species, which may no longer exist, difficult and hence confound the assignment of the subgenomes in the genome assembly. However, this study facilitated differentiation of the subgenome structure of *C. sativa* with the help of lower ploidy species. The correction of the subgenome assignment in the reference genome of *C. sativa* will further facilitate the identification of diploid and tetraploid progenitors, and will provide a foundation for the study of polyploid formation, adaptation and functionalization. These findings also helped to resolve disparity in the taxonomic nomenclature of *Camelina* species, which was

common, particularly for close relatives and lower ploidy species. Such ambiguity has been misleading, not only in genetic diversity studies but also in interspecific hybridization studies (Tepfer et al. 2020). The genetic analyses of related species also identified a unique subgenome structure in *C. microcarpa* "Type 2", which was different to that of *C. sativa*, suggesting a potential fourth progenitor for this species. These findings indicated taxonomic confusion, where the *C. microcarpa* accessions in the plant genebank, although annotated as hexaploids, were not differentiated despite disparities in the underlying subgenome structure, as well as chromosome number (2n = 19 and 2n = 20).

The identified low level of genetic diversity among camelina accessions limits the development of robust breeding lines/varieties, due to the lack of sufficient allelic variation to select for improved traits and to increase possible heterosis (Mackay et al. 2020). The data did however suggest some diversity could be acquired from winter-type accessions, as well as more distantly related species. Thus, a crossing scheme was developed to exploit this variation, using the winter-type accessions CAM176 (*C. alyssum*) and CN113692 (*C. sativa* ssp. *pilosa*). In addition, interspecific crosses with more distantly related accessions, TMP23999 (*C. microcarpa*) and CN119102 (*C. microcarpa*), were carried out to observe the nature of recombination with the newly identified 'fourth' genome and potentially capture novel sources of variation for the winter habit.

In *C. sativa,* both annual and biennial behaviours are found, although most of the close relatives are biennial in nature. Biennial behaviour (or winter-type) alters the developmental process of a plant to allow it to survive winter freezing temperatures and facilitate flowering at the onset of spring. This survival mechanism is of particular importance in areas such as western Canada where winter temperatures drop well below freezing. Two different types of winter behaviour were noticed while growing camelina species without vernalization treatment, one where only a few branches on the plant transition to the reproductive stage, and the other where the plant maintains a vegetative state. This might indicate the quantitative nature of similar genes responsible for the winter-type, or the action of different genes among different *Camelina* species. The identified homologous QTL regions for vernalization requirement in *C. sativa* in all

crosses suggested the conservation of genes controlling this trait across the species. This study also supported previous reports from *C. sativa* and related species that *FLC* is the major determinant of this trait (Anderson et al. 2018; Sheldon et al. 2000; Michaels and Amasino 1999). Of note, study of this trait emphasised the ability of polyploid genomes to exploit the presence of multiple duplicated loci to create variation. From the backcross derived from *C. microcarpa,* a QTL was mapped on chromosome 13 at the same locus as that from CAM176 (*C. alyssum*), which not only suggested conservation of QTL across species, but a shared common ancestor for the second subgenome. In addition, the higher expression of *FLC* on chromosome 8 (subgenome 1) for *C. sativa* might be evidence of functional divergence. The developed material will allow further study of the impact of polyploidy on the function of *FLC* and its resulting effect on the vernalization response. Further, the identified QTL can be utilized in the winter *C. sativa* breeding programs, by developing markers to allow selection of genomic regions controlling winter behaviour. Likewise, the developed genetic maps will be further used to identify QTL associated with other yield traits.

The success of interspecific hybridization with *C. microcarpa* was relatively low; however, variation for a number of traits was observed among the segregating populations. This variation might be expected based on the genetic distance between these species; however, evidence of non-homologous recombination, including homoeologous recombination, was found. In most allopolyploids a genetic control mechanism has been suggested or confirmed, as in wheat (Griffiths et al. 2006) and *B. napus* (Higgins et al. 2020). It might be inferred from the highly undifferentiated subgenome structure that *C. sativa* also has such a control mechanism. The fact that homoeologous recombination can be induced in wide crosses suggests that the control of pairing could be manipulated, which opens up novel avenues for creating new variation in *C. sativa*. Homoeologous recombination might provide benefits such as transferring genes from wild species to increase the allelic richness, but further, inducing homoeologous recombination within *C. sativa* itself could create copy number variation, which can be very powerful in generating new and improved traits, as shown in other crops (Gabur et al. 2019).

A number of the plants generated from this study will be useful in the *Camelina* breeding program, as well as in the genetic dissection of yield related traits that were observed to be segregating, such as: seed size, pod size, number of seeds per pod, and main raceme length. Further, the differences in subgenome recombination patterns shown among segregating plants derived from the same parents could provide useful material to further understand subgenome dominance, particularly in the material with recombined subgenomes (Edger et al. 2017). Likewise, for the inter-specific segregating plants which recovered their fertility, these materials should allow insights into the genome changes and chromosomal compositions which provide stability in the adaptation of *Camelina* species.

The fact that the winter habit could be controlled by different combinations of three homologous loci, and the strong inter-relationship among homologous chromosomes within *Camelina* species, as stressed by the evidence of homoeologous recombination, led to questions related to the influence and expression balance of the duplicated genes. Further, the identification of closely related diploid, tetraploid and hexaploids among the *Camelina* species from the genetic diversity study provided an avenue to begin the study of dominance for each particular subgenome (Bottani et al. 2018).

Whole genome duplication has been associated with speciation and the generation of phenotypic novelty (Nieto Feliner et al. 2020) and the phenomenon has been evident for most plants (Renny-Byfield and Wendel 2014). The comparison of mutation rates between homologous genes in related species can be utilized to estimate timeframes for probable divergence events between the species. For example among Brassicaceae species, most diversification events are believed to have taken place within the last 23 mya (Kagale et al. 2014b). Based on chloroplast DNA, *C. sativa* was suggested to have diverged from *A. thaliana* about 8.16 mya (Hohmann et al. 2015); however, based on nuclear DNA this timeframe was suggested to be 14.6 to 17.2 mya (Huang et al. 2016). The comparison of the *C. neglecta* (diploid) genome with the triplicated subgenomes from hexaploid *C. sativa* and *C. microcarpa* “Type 2” showed interesting differences. The third subgenome of *C. sativa* was found to have diverged earliest, in contrast, the third subgenome from *C. microcarpa* “Type 2” diverged more recently and possessed 6 chromosomes similar to

C. neglecta. If we assume a concept of chromosome number reduction in evolution of a species, such as from 8 to 5 as in *A. thaliana* (Mandakova and Lysak 2008)*, C. neglecta* would be a more recent form of a higher chromosome number species. Therefore, the results could suggest the splitting/merging of chromosomes took place in the time period between the differentiation of the third subgenome of *C. microcarpa* "Type 2" and the second subgenome of *C. sativa*. For gene expression analysis, in each case the subgenome of the hexaploid with the highest age of divergence from *C. neglecta* was found to have dominant expression. Interestingly, the third subgenome was dominant in *C. sativa*, whereas the second subgenome was dominant in *C. microcarpa* "Type 2". Thus, although the age of divergence among the subgenomes was fairly low in case of *Camelina,* it suggests that the differentiation of the expression pattern of homologous gene copies has started, which might influence the stability of the polyploid genome. In previous studies, evidence of genome dominance has been used to infer the steps in the evolutionary path leading to the current genome structure, that is the last genome hybridized to form the polyploid would be the dominant genome (Cheng et al. 2012). If this assertion is true, then this suggests a different path led to the generation of the *C. sativa* and *C. microcarpa* "Type 2" genomes (a representative diagram is presented in **appendix A.4.1**).

The results presented from this study have an impact on understanding subgenome dominance in a relatively recent polyploid and demonstrated a linear relationship between age of divergence and subgenome dominance. However, further analyses to identify the basal subgenome, as well as progenitors of the second and the third subgenomes (if they still exist), would provide better understanding of how gene expression is impacted by the addition of subgenomes in the formation of polyploids.

Hypotheses tested in the thesis:

The identified differences in the subpopulation structure among *Camelina* species and the identification of subpopulation specific minor alleles provided evidence for acceptance of the first hypothesis proposed in the thesis, namely that "the related species of *C. sativa* contain novel allelic variation not present within available *C. sativa* accessions".

The use of intra- and inter-species hybridization to develop segregating populations created the variation necessary to map QTL associated with winter behaviour. Therefore, the second hypothesis, "homology between chromosomes of *C. sativa* and related species will be sufficient to allow genetic introgression of useful variation", was accepted.

The identified subgenome dominance in the *Camelina* species and the variation in gene fractionation for the subgenomes of *C. sativa* were associated with the age of divergence of the subgenomes. Thus, the third hypothesis, "Differential subgenome dominance and fractionation occurs in *Camelina* as the ploidy level increases", is partially accepted; since although subgenome dominance was found in both *Camelina* species, the level of fractionation was low among subgenomes belonging to *C. microcarpa* "Type 2".

Future directions

- Quantification of morphological traits among *C. microcarpa* “Type 1” and *C. microcarpa* “Type 2” to estimate their utility to the *C. sativa* breeding program.
- Identifying bridging species/genotypes between independently naturalized *C. microcarpa* “Type 1” and *C. sativa* populations to facilitate *Camelina* breeding program.
- Further study of the mechanism for diploidization in allohexaploid *C. sativa* by utilizing segregating plants (probable aneuploids) developed in this study, and the manipulation of genes identified as potentially influential from related species.
- Expression analysis of the second subgenome of *C. microcarpa* “Type 2” introgressed into *C. sativa*, for a better understanding of the dominant or suppressive nature of traits of interest upon genome merger. In addition, extending the analysis of subgenome dominance to understand the role of epigenetic phenomenon.
- Utilization of ancient recombination from *C. microcarpa* populations to explore genetic mechanisms of important traits through phenotypic association studies.

Reference

Adams, K.L., and J.F. Wendel, 2005 Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8 (2):135-141.

Ågren, J., C.G. Oakley, S. Lundemo, and D.W. Schemske, 2017 Adaptive divergence in flowering time among natural populations of *Arabidopsis thaliana*: Estimates of selection and QTL mapping. *Evolution* 71 (3):550-564.

Al-Shehbaz, I.A., 1987 Camelina. *Journal of the Arnold Arboretum* 68:234-240.

Al-Shehbaz, I.A., 2012 A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61 (5):931-954.

Al-Shehbaz, I.A., M.A. Beilstein, and E.A. Kellogg, 2006 Systematics and phylogeny of the Brassicaceae (Cruciferae): An overview. *Plant Systematics and Evolution* 259:89-120.

Alger, E.I., and P.P. Edger, 2020 One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current Opinion in Plant Biology* 54:108-113.

Allen, B.L., M.F. Vigil, and J.D. Jabro, 2014 Camelina growing degree hour and base temperature requirements. *Agronomy Journal* 106:940-944.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, 1990 Basic local alignment search tool. *Journal of molecular biology* 215 (3):403-410.

Amyot, L., T. McDowell, S.L. Martin, J. Renaud, M.Y. Gruber *et al.*, 2018 Assessment of Antinutritional Compounds and Chemotaxonomic Relationships between *Camelina sativa* and Its Wild Relatives. *Journal of agricultural and food chemistry* 67 (3):796-806.

Anderson, J.V., D.P. Horvath, M. Doğramaci, K.M. Dorn, W.S. Chao *et al.*, 2018 Expression of *FLOWERING LOCUS C* and a frameshift mutation of this gene on chromosome 20 differentiate a summer and winter annual biotype of *Camelina sativa*. *Plant Direct* 2 (7):e00060.

Anderson, J.V., A. Wittenberg, H. Li, and M.T. Berti, 2019 High throughput phenotyping of *Camelina sativa* seeds for crude protein, total oil, and fatty acids profile by near infrared spectroscopy. *Industrial Crops and Products* 137:501-507.

Angelini, L.G., E. Moscheni, G. Colonna, P. Belloni, and E. Bonari, 1997 Variation in agronomic characteristics and seed oil composition of new oilseed crops in central Italy. *Industrial Crops and Products* 6 (3-4):313-323.

Aubert, D., L. Chen, Y.-H. Moon, D. Martin, L.A. Castle *et al.*, 2001 *EMF1*, a novel protein involved in the control of shoot architecture and flowering in *Arabidopsis*. *The Plant Cell* 13 (8):1865-1875.

Bedre, R., and K. Mandadi, 2019 GenFam: A web application and database for gene family-based classification and functional enrichment analysis. *Plant Direct* 3 (12):e00191.

Berti, M., R. Gesch, C. Eynck, J. Anderson, and S. Cermak, 2016 Camelina uses, genetics, genomics, production, and management. *Industrial Crops and Products* 94:690-710.

Berti, M., D. Samarappuli, B.L. Johnson, and R.W. Gesch, 2017 Integrating winter camelina into maize and soybean cropping systems. *Industrial crops and Products* 107:595-601.

Berti, M., R. Wilckens, S. Fischer, A. Solis, and B. Johnson, 2011 Seeding date influence on camelina seed yield, yield components, and oil content in Chile. *Industrial Crops and Products* 34 (2):1358-1365.

Betancor, M.B., M. Sprague, S. Usher, O. Sayanova, P.J. Campbell *et al.*, 2015 A nutritionally-enhanced oil from transgenic *Camelina sativa* effectively replaces fish oil as a source of eicosapentaenoic acid for fish. *Scientific Reports* 5:8104.

Birchler, J.A., and R.A. Veitia, 2012 Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences* 109 (37):14746-14753.

Bird, K.A., C. Niederhuth, S. Ou, M. Gehan, J.C. Pires *et al.*, 2019 Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *BioRxiv*:814491.

Bird, K.A., R. VanBuren, J.R. Puzey, and P.P. Edger, 2018 The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* 220 (1):87-93.

Bolger, A.M., M. Lohse, and B. Usadel, 2014 Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

Bottani, S., N.R. Zabet, J.F. Wendel, and R.A. Veitia, 2018 Gene expression dominance in allopolyploids: hypotheses and models. *Trends iin Plant Science* 23 (5):393-402.

Brock, J.R., A.A. Donmez, M.A. Beilstein, and K.M. Olsen, 2018 Phylogenetics of *Camelina* Crantz. (Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*). *Molecular Phylogenetics and Evolution* 127:834-842.

Brock, J.R., T. Mandakova, M.A. Lysak, and I.A. Al-Shehbaz, 2019 *Camelina neglecta* (Brassicaceae, Camelineae), a new diploid species from Europe. *PhytoKeys* (115):51-57.

Broman, K.W., H. Wu, Ś. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7):889-890.

Brown, T.D., T.S. Hori, X. Xue, C.L. Ye, D.M. Anderson *et al.*, 2016 Functional Genomic Analysis of the Impact of Camelina (*Camelina sativa*) Meal on Atlantic Salmon (*Salmo salar*) Distal Intestine Gene Expression and Physiology. *Marine Biotechnology* 18:418-435.

Büchsenschütz-Nothdurft, A., A. Schuster, and W. Friedt, 1998 Breeding for modified fatty acid composition via experimental mutagenesis in *Camelina sativa* (L.) Crtz. *Industrial Crops and Products* 7:291-295.

Čalasan, A.Ž., A.P. Seregin, H. Hurka, N.P. Hofford, and B. Neuffer, 2019 The Eurasian steppe belt in time and space: Phylogeny and historical biogeography of the false flax (*Camelina* Crantz, Camelineae, Brassicaceae). *Flora* 260:151477.

Campbell, M.C., A.F. Rossi, and W. Erskine, 2013 Camelina (*Camelina sativa* (L.) Crantz): Agronomic potential in Mediterranean environments and diversity for biofuel and food uses. *Crop and Pasture Science* 64:388-398.

Chalhoub, B., F. Denoeud, S. Liu, I.A. Parkin, H. Tang *et al.*, 2014 Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345 (6199):950-953.

Chao, W.S., H. Wang, D.P. Horvath, and J.V. Anderson, 2019 Selection of endogenous reference genes for qRT-PCR analysis in *Camelina sativa* and identification of *FLOWERING LOCUS C* allele-specific markers to differentiate summer-and winter-biotypes. *Industrial crops and Products* 129:495-502.

Chaudhary, R., C.S. Koh, S. Kagale, L. Tang, S.W. Wu *et al.*, 2020 Assessing Diversity in the *Camelina* Genus Provides Insights into the Genome Structure of *Camelina sativa*. *G3: Genes, Genomes, Genetics* 10 (4):1297-1308.

Cheng, F., J. Wu, L. Fang, S. Sun, B. Liu *et al.*, 2012 Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PloS one* 7 (5):e36442.

Cheng, F., J. Wu, and X. Wang, 2014 Genome triplication drove the diversification of *Brassica* plants. *Horticulture research* 1:14024.

Chhatre, V.E., and K.J. Emerson, 2017a StrAuto: automation and parallelization of STRUCTURE analysis. *BMC bioinformatics* 18 (1):192.

Chhatre, V.E., and K.J. Emerson, 2017b StrAuto: automation and parallelization of STRUCTURE analysis. *BMC bioinformatics* 18 (1):1-5.

Conant, G.C., J.A. Birchler, and J.C. Pires, 2014 Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current opinion in plant biology* 19:91-98.

Crow, K.D., and G.P. Wagner, 2005 What is the role of genome duplication in the evolution of complexity and diversity? *Molecular biology and evolution* 23 (5):887-892.

de Jong, T., K. Escobedo Quevedo, C. van der Veen-van Wijk, and M. Moshgani, 2018 Performance of aneuploid backcross hybrids between the crop *Brassica napus* and its wild relative B. rapa. *Plant Biology* 20 (1):67-74.

Deng, W., H. Ying, C.A. Helliwell, J.M. Taylor, W.J. Peacock *et al.*, 2011 *FLOWERING LOCUS C* (*FLC*) regulates development pathways throughout the life cycle of Arabidopsis. *Proceedings of the National Academy of Sciences* 108 (16):6680-6685.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21.

Douglas, G.M., G. Gos, K.A. Steige, A. Salcedo, K. Holm *et al.*, 2015 Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences* 112 (9):2806-2811.

Earl, D.A., 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4 (2):359-361.

Edger, P.P., T.J. Poorten, R. VanBuren, M.A. Hardigan, M. Colle *et al.*, 2019 Origin and evolution of the octoploid strawberry genome. *Nature genetics* 51 (3):541-547.

Edger, P.P., R. Smith, M.R. McKain, A.M. Cooley, M. Vallejo-Marin *et al.*, 2017 Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell* 29 (9):2150-2167.

Eynck, C., G. Séguin-Swartz, W.E. Clarke, and I.A.P. Parkin, 2012 Monolignol biosynthesis is associated with resistance to Sclerotinia sclerotiorum in *Camelina sativa*. *Molecular Plant Pathology* 13:887-899.

Faure, J.-D., and M. Tepfer, 2016 *Camelina*, a Swiss knife for plant lipid biotechnology. *Ocl* 23:D503.

Felsenstein, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39 (4):783-791.

Ferrie, A.M.R., and T.D. Bethune, 2011 A microspore embryogenesis protocol for *Camelina sativa*, a multi-use crop. *Plant Cell, Tissue and Organ Culture* 106:495-501.

Francis, a., and S.I. Warwick, 2009 The Biology of Canadian Weeds. 142. *Camelina alyssum* (Mill.) Thell.; *C. microcarpa* Andrz. ex DC.; *C. sativa* (L.) Crantz. *Canadian Journal of Plant Science* 89:791-810.

Freeling, M., 2009 Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual review of plant biology* 60:433-453.

Freeling, M., M.J. Scanlon, and J.E. Fowler, 2015 Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* 35:110-118.

Gabur, I., H.S. Chawla, R.J. Snowdon, and I.A. Parkin, 2019 Connecting genome structural variation with complex traits in crop plants. *Theoretical and Applied Genetics* 132 (3):733-750.

Gaeta, R.T., J.C. Pires, F. Iniguez-Luy, E. Leon, and T.C. Osborn, 2007 Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant Cell* 19 (11):3403-3417.

Galasso, I., A. Manca, L. Braglia, E. Ponzoni, and D. Breviario, 2015 Genomic fingerprinting of *Camelina* species using cTBP as molecular marker. *American Journal of Plant Sciences* 6:1184-1200.

Galbraith, D.W., K.R. Harkins, J.M. Maddox, N.M. Ayres, D.P. Sharma *et al.*, 1983 Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220 (4601):1049-1051.

Garcia, S., M. Sanz, T. Garnatje, A. Kreitschitz, E.D. McArthur *et al.*, 2004 Variation of DNA amount in 47 populations of the subtribe Artemisiinae and related taxa (Asteraceae, Anthemideae): karyological, ecological, and systematic implications. *Genome* 47 (6):1004-1014.

Gehringer, A., W. Friedt, W. Lühs, and R.J. Snowdon, 2006 Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp. *sativa*). *Genome* 49:1555-1563.

Gel, B., and E. Serra, 2017 karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33 (19):3088-3090.

Gesch, R.W., H.L. Matthees, A.L. Alvarez, and R.D. Gardner, 2018 Winter camelina: Crop growth, seed yield, and quality response to cultivar and seeding rate. *Crop Science* 58 (5):2089-2098.

Ghamkhar, K., J. Croser, N. Aryamanesh, M. Campbell, N. Kon'kova *et al.*, 2010 Camelina (*Camelina sativa* (L.) Crantz) as an alternative oilseed: molecular and ecogeographic analyses. *Genome* 53 (7):558-567.

Goel, P., and A.K. Singh, 2015 Abiotic stresses downregulate key genes involved in nitrogen uptake and assimilation in *Brassica juncea* L. *PloS one* 10 (11).

Griffiths, S., R. Sharp, T.N. Foote, I. Bertin, M. Wanous *et al.*, 2006 Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature* 439 (7077):749-752.

Groeneveld, J.H., and A.M. Klein, 2014 Pollination of two oil-producing plant species: Camelina (*Camelina sativa* L. Crantz) and pennycress (*Thlaspi arvense* L.) double-cropping in Germany. *GCB Bioenergy* 6:242-251.

Gugel, R.K., and K.C. Falk, 2006 Agronomic and seed quality evaluation of *Camelina sativa* in western Canada. *Canadian Journal of Plant Science* 86:1047-1058.

Haas, B.J., A.L. Delcher, J.R. Wortman, and S.L. Salzberg, 2004 DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20 (18):3643-3646.

Hall, N., 2013 After the gold rush. *Genome Biology* 14 (5):115.

Hansen, L.N., 1998 Intertribal somatic hybridization between rapid cycling *Brassica oleracea* L. and *Camelina sativa* (L.) Crantz. *Euphytica* 104:173-179.

Harrison, G., and J. Heslop-Harrison, 1995 Centromeric repetitive DNA sequences in the genus *Brassica*. *Theoretical Applied Genetics* 90 (2):157-165.

He, X., and J. Zhang, 2005 Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169 (2):1157-1164.

Higgins, E.E., W.E. Clarke, E.C. Howell, S.J. Armstrong, and I.A.P. Parkin, 2018 Detecting *de Novo* Homoeologous Recombination Events in Cultivated *Brassica napus* Using a Genome-Wide SNP Array. *G3: Genes, Genomes, Genetics* 8 (8):2673-2683.

Higgins, E.E., E.C. Howell, S.J. Armstrong, and I.A. Parkin, 2020 A major quantitative trait locus on chromosome A9, *BnaPh1*, controls homoeologous recombination in *Brassica napus*. *New Phytologist*.

Higgins, J.A., P.C. Bailey, and D.A. Laurie, 2010 Comparative genomics of flowering time pathways using Brachypodium distachyon as a model for the temperate grasses. *PloS one* 5 (4):e10065.

Hjelmqvist, H., 1979 *Beiträge zur Kenntnis der prähistorischen Nutzpflanzen in Schweden*: Verlag nicht ermittelbar.

Hohmann, N., E.M. Wolf, M.A. Lysak, and M.A. Koch, 2015 A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *The Plant Cell* 27 (10):2770-2784.

Hollister, J.D., and B.S. Gaut, 2009 Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* 19 (8):1419-1428.

Holtz, Y., J.L. David, and V. Ranwez, 2017 The genetic map comparator: a user-friendly application to display and compare genetic maps. *Bioinformatics* 33 (9).

Hovsepyan, R., and G. Willcox, 2008 The earliest finds of cultivated plants in Armenia: evidence from charred remains and crop processing residues in pisé from the Neolithic settlements of Aratashen and Aknashen. *Vegetation History and Archaeobotany* 17 (1):63-71.

Huang, C.-H., R. Sun, Y. Hu, L. Zeng, N. Zhang *et al.*, 2016 Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular biology and evolution* 33 (2):394-412.

Huang, H., B. Liu, L. Liu, and S. Song, 2017 Jasmonate action in plant growth and development. *Journal of Experimental Botany* 68 (6):1349-1359.

Hunsaker, D.J., A.N. French, and K.R. Thorp, 2013 Camelina water use and seed yield response to irrigation scheduling in an arid environment. *Irrigation Science* 31:911-929.

Hurgobin, B., A.A. Golicz, P.E. Bayer, C.K.K. Chan, S. Tirnaz *et al.*, 2018 Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnology Journal* 16 (7):1265-1274.

Hutcheon, C., R.F. Ditt, M. Beilstein, L. Comai, J. Schroeder *et al.*, 2010 Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. *BMC Plant Biology* 10:233.

Jiang, J.J., X.X. Zhao, W. Tian, T.B. Li, and Y.P. Wang, 2009 Intertribal somatic hybrids between *Brassica napus* and *Camelina sativa* with high linolenic acid content. *Plant Cell, Tissue and Organ Culture* 99:91-95.

Jiang, W.Z., I.M. Henry, P.G. Lynagh, L. Comai, E.B. Cahoon *et al.*, 2017 Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnology Journal* 15:648-657.

Johnston, J.S., A.E. Pepper, A.E. Hall, Z.J. Chen, G. Hodnett *et al.*, 2005 Evolution of genome size in Brassicaceae. *Annals of Botany* 95:229-235.

Julié-Galau, S., Y. Bellec, J.D. Faure, and M. Tepfer, 2014 Evaluation of the potential for interspecific hybridization between *Camelina sativa* and related wild Brassicaceae in anticipation of field trials of GM camelina. *Transgenic Research* 23:67-74.

Kagale, S., C. Koh, J. Nixon, V. Bollina, W.E. Clarke *et al.*, 2014a The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature Communications* 5:1-11.

Kagale, S., J. Nixon, Y. Khedikar, A. Pasha, N.J. Provart *et al.*, 2016 The developmental transcriptome atlas of the biofuel crop *Camelina sativa*. *Plant Journal* 88:879-894.

Kagale, S., S.J. Robinson, J. Nixon, R. Xiao, T. Huebert *et al.*, 2014b Polyploid evolution of the Brassicaceae during the Cenozoic era. *The Plant Cell* 26 (7):2777-2791.

Kaul, S., H.L. Koo, J. Jenkins, M. Rizzo, T. Rooney *et al.*, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 (6814):796-815.

Kemi, U., A. Niittyvuopio, T. Toivainen, A. Pasanen, B. Quilot-Turion *et al.*, 2013 Role of vernalization and of duplicated *FLOWERING LOCUS C* in the perennial *Arabidopsis lyrata*. *New Phytologist* 197 (1):323-335.

Kim, D.-H., M.R. Doyle, S. Sung, and R.M. Amasino, 2009 Vernalization: winter and the timing of flowering in plants. *Annual Review of Cell and Developmental* 25:277-299.

King, K., H. Li, J. Kang, and C. Lu, 2019 Mapping quantitative trait loci for seed traits in *Camelina sativa*. *Theoretical Applied Genetics* 132:2567–2577.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: an information aesthetic for comparative genomics. *Genome Research* 19 (9):1639-1645.

Kuligowska, K., H. Lütken, B. Christensen, I. Skovgaard, M. Linde *et al.*, 2015 Evaluation of reproductive barriers contributes to the development of novel interspecific hybrids in the *Kalanchoë* genus. *BMC Plant Biology* 15.

Kumar, S., G. Stecher, and K. Tamura, 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33 (7):1870-1874.

Kurasiak-popowska, D., M. Graczyk, and K. Stuper-szablewska, 2020 Winter Camelina seeds as a raw material for the production of erucic acid-free oil. *Food Chemistry*:127265.

Langmead, B., and S.L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357-359.

Larsson, M., 2013 Cultivation and processing of *Linum usitatissimum* and *Camelina sativa* in southern Scandinavia during the Roman Iron Age. *Vegetation History and Archaeobotany* 22 (6):509-520.

Leal-Bertioli, S., K. Shirasawa, B. Abernathy, M. Moretzsohn, C. Chavarro *et al.*, 2015 Tetrasomic recombination is surprisingly frequent in allotetraploid *Arachis*. *Genetics* 199 (4):1093-1105.

Li, H., H. Li, R. Durbin, and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25:1754-1760.

Liu, K., and S.V. Muse, 2005 PowerMaker: An integrated analysis environment for genetic maker analysis. *Bioinformatics* 21:2128-2129.

Liu, S., Y. Liu, X. Yang, C. Tong, D. Edwards *et al.*, 2014 The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* 5:1-11.

Lošák, T., J. Hlusek, J. Martinec, J. Vollmann, J. Peterka *et al.*, 2011 Effect of combined nitrogen and sulphur fertilization on yield and qualitative parameters of *Camelina sativa* [L.] Crtz. (false flax). *Acta Agriculturae Scandinavica Section B: Soil and Plant Science* 61:313-321.

Love, M., S. Anders, and M. Huber, 2014 Differential gene expression analysis based on the negative binomial distribution. *Genome Biology* 15:550.

Lu, C., and J. Kang, 2008 Generation of transgenic plants of a potential oilseed crop *Camelina sativa* by Agrobacterium-mediated transformation. *Plant Cell Reports* 27:273-278.

Luo, Z., J. Brock, J.M. Dyer, T.M. Kutchan, M. Augustin *et al.*, 2019a Genetic diversity and population structure of a *Camelina sativa* spring panel. *Frontiers in Plant Science* 10:184.

Luo, Z., P. Tomasi, N. Fahlgren, and H. Abdel-Haleem, 2019b Genome-wide association study (GWAS) of leaf cuticular wax components in *Camelina sativa* identifies genetic loci related to intracellular wax transport. *BMC Plant Biology* 19 (1):187.

Lysak, M.A., M.A. Koch, A. Pecinka, and I. Schubert, 2005 Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15 (4):516-525.

Lyzenga, W.J., M. Harrington, D. Bekkaoui, M. Wigness, D.D. Hegedus *et al.*, 2019 CRISPR/Cas9 editing of three *CRUCIFERIN C* homoeologues alters the seed protein profile in *Camelina sativa*. *BMC Plant Biology* 19 (1):292.

Ma, X.-F., and J. Gustafson, 2005 Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenetic and genome research* 109 (1-3):236-249.

Mackay, I.J., J. Cockram, P. Howell, and W. Powell, 2020 Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. *Plant Biotechnology Journal*.

Madlung, A., A.P. Tyagi, B. Watson, H. Jiang, T. Kagochi *et al.*, 2005 Genomic changes in synthetic *Arabidopsis* polyploids. *The Plant Journal* 41 (2):221-230.

Manca, A., P. Pecchia, S. Mapelli, P. Masella, and I. Galasso, 2013 Evaluation of genetic diversity in a *Camelina sativa* (L.) Crantz collection using microsatellite markers and biochemical traits. *Genetic Resources and Crop Evolution* 60:1223-1236.

Mandakova, T., and M.A. Lysak, 2008 Chromosomal Phylogeny and Karyotype Evolution in x=7 Crucifer Species (Brassicaceae). *The Plant Cell* 20:2559-2570.

Mandáková, T., and M.A. Lysak, 2018 Post-polyploid diploidization and diversification through dysploid changes. *Current opinion in plant biology* 42:55-65.

Mandáková, T., M. Pouch, J.R. Brock, I.A. Al-Shehbaz, and M.A. Lysak, 2019 Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. *The Plant Cell* 31 (11):2596-2612.

Mao, Y., 2019 GenoDup Pipeline: a tool to detect genome duplication using the dS-based method. *PeerJ* 7:e6303.

Martin, S.L., B.E. Lujan-Toro, C.A. Sauder, T. James, S. Ohadi *et al.*, 2019 Hybridization rate and hybrid fitness for *Camelina microcarpa* Andrz. ex DC (♀) and *Camelina sativa* (L.) Crantz (Brassicaceae) (♂). *Evolutionary Applications* 12 (3):443-455.

Martin, S.L., C.A. Sauder, T. James, K.W. Cheung, F.M. Razeq *et al.*, 2015 Sexual hybridization between *Capsella bursa-pastoris* (L.) Medik (♀) and *Camelina sativa* (L.) Crantz (♂) (Brassicaceae). *Plant Breeding* 134:212-220.

Martin, S.L., T.W. Smith, T. James, F. Shalabi, P. Kron *et al.*, 2017 An update to the Canadian range, abundance, and ploidy of *Camelina* spp.(Brassicaceae) east of the Rocky Mountains. *Botany* 95 (4):405-417.

Matsushita, S.C., A.P. Tyagi, G.M. Thornton, J.C. Pires, and A. Madlung, 2012 Allopolyploidization lays the foundation for evolution of distinct populations: evidence from analysis of synthetic *Arabidopsis* allohexaploids. *Genetics* 191 (2):535-547.

Mayfield, J.D., K.M. Folta, A.-L. Paul, and R.J. Ferl, 2007 The 14-3-3 proteins μ and υ influence transition to flowering and early phytochrome response. *Plant physiology* 145 (4):1692-1702.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297-1303.

Michaels, S.D., and R.M. Amasino, 1999 *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell* 11 (5):949-956.

Morineau, C., Y. Bellec, F. Tellier, L. Gissot, Z. Kelemen *et al.*, 2017 Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnology Journal* 15:729-739.

Murat, F., A. Louis, F. Maumus, A. Armero, R. Cook *et al.*, 2016 Erratum to: Understanding Brassicaceae evolution through ancestral genome reconstruction [Genome Biol. 16, (2015), (262)]. *Genome Biology* 17:1-17.

Murray, M.G., and W.F. Thompson, 1980 Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8 (19):4321-4325.

Nagaharu, U., and N. Nagaharu, 1935 Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japan J Bot*.

Narasimhulu, S.B., P.B. Kirti, S.R. Bhatt, S. Prakash, and V.L. Chopra, 1994 Intergeneric protoplast fusion between *Brassica carinata* and *Camelina sativa*. *Plant Cell Reports* 13:657-660.

Navabi, Z.K., K.E. Stead, J.C. Pires, Z. Xiong, A.G. Sharpe *et al.*, 2011 Analysis of B-genome chromosome introgression in interspecific hybrids of *Brassica napus*× *B. carinata*. *Genetics* 187 (3):659-673.

Neupane, D., J.K. Solomon, E. Mclennon, J. Davison, and T. Lawry, 2019 Sowing date and sowing method influence on camelina cultivars grain yield, oil concentration, and biodiesel production. *Food and Energy Security* 8 (3):e00166.

Nieto Feliner, G., J.M. Casacuberta, and J.F. Wendel, 2020 Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics* 11:792.

Noh, B., S.-H. Lee, H.-J. Kim, G. Yi, E.-A. Shin *et al.*, 2004 Divergent roles of a pair of homologous jumonji/zinc-finger–class transcription factor proteins in the regulation of *Arabidopsis* flowering time. *The Plant Cell* 16 (10):2601-2613.

O'Neill, C.M., X. Lu, A. Calderwood, E.H. Tudor, P. Robinson *et al.*, 2019 Vernalization and floral transition in autumn drive winter annual life history in oilseed rape. *Current Biology* 29 (24):4300-4306. e4302.

Okazaki, K., K. Sakamoto, R. Kikuchi, A. Saito, E. Togashi *et al.*, 2007 Mapping and characterization of *FLC* homologs and QTL analysis of flowering time in *Brassica oleracea*. *Theoretical Applied Genetics* 114 (4):595-608.

Ozseyhan, M.E., J. Kang, X. Mu, and C. Lu, 2018 Mutagenesis of the *FAE1* genes significantly changes fatty acid composition in seeds of *Camelina sativa*. *Plant Physiology and Biochemistry* 123:1-7.

Panchy, N., M. Lehti-Shiu, and S.-H. Shiu, 2016 Evolution of gene duplication in plants. *Plant physiology* 171 (4):2294-2316.

Parkin, I.A., C. Koh, H. Tang, S.J. Robinson, S. Kagale *et al.*, 2014 Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology* 15 (6):1-18.

Peakall, R., and P.E. Smouse, 2006 GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Peakall, R., and P.E. Smouse, 2012 GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537-2539.

Pleesers, A.G., W.G. McGregor, R.B. Carson, and W. Nakoneshny, 1962 Species Trials With Oilseed Plants: II.Camelina. *Canadian Journal of Plant Science* 9:452-459.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink, 2012 Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one* 7 (2):e32253.

Pritchard, J.K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Pruneda-Paz, J.L., G. Breton, A. Para, and S.A. Kay, 2009 A functional genomics approach reveals *CHE* as a component of the Arabidopsis circadian clock. *science* 323 (5920):1481-1485.

Putnam, D.H., J.T. Budin, L.A. Field, and W.M. Breene, 1993 *Camelina : A Promising Low-Input Oilseed*. New York: Wiley.

Quinlan, A.R., and I.M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6):841-842.

Ramírez-González, R., P. Borrill, D. Lang, S. Harrington, J. Brinton *et al.*, 2018 The transcriptional landscape of polyploid wheat. *Science* 361 (6403):eaar6089.

Ramsey, J., 2007 Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). *Heredity* 98 (3):143-150.

Ramsey, J., and D.W. Schemske, 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual review of ecology and systematics* 29 (1):467-501.

Renny-Byfield, S., E. Rodgers-Melnick, and J. Ross-Ibarra, 2017 Gene fractionation and function in the ancient subgenomes of maize. *Molecular biology and evolution* 34 (8):1825-1832.

Renny-Byfield, S., and J.F. Wendel, 2014 Doubling down on genomes: polyploidy and crop plants. *American journal of botany* 101 (10):1711-1725.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law *et al.*, 2015 limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43 (7):e47-e47.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1):139-140.

Ross-Ibarra, J., P.L. Morrell, and B.S. Gaut, 2007 Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences* 104 (suppl 1):8641-8648.

Ruiz-Lopez, N., R.P. Haslam, S. Usher, J.A. Napier, and O. Sayanova, 2015 An alternative pathway for the effective production of the omega-3 long-chain polyunsaturates EPA and ETA in transgenic oilseeds. *Plant Biotechnology Journal* 13:1264-1275.

Sagun, V.G., and C. Auer, 2017 Pollen morphology of selected Camelineae (Brassicaceae). *Palynology* 41:255-266.

Schiessl, S.V., D. Quezada-Martinez, E. Tebartz, R.J. Snowdon, and L. Qian, 2019 The vernalisation regulator *FLOWERING LOCUS C* is differentially expressed in biennial and annual *Brassica napus*. *Scientific Reports* 9 (1):1-15.

Schillinger, W.F., D.J. Wysocki, T.G. Chastain, S.O. Guy, and R.S. Karow, 2012 Camelina: Planting date and method effects on stand establishment and seed yield. *Field Crops Research* 130:138-144.

Schnable, J.C., N.M. Springer, and M. Freeling, 2011 Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences* 108 (10):4069-4074.

Schnable, J.C., X. Wang, J.C. Pires, and M. Freeling, 2012 Escape from preferential retention following repeated whole genome duplications in plants. *Frontiers in Plant Science* 3:94.

Schranz, M.E., M.A. Lysak, and T. Mitchell-Olds, 2006 The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in plant science* 11 (11):535-542.

Schuster, A., and W. Friedt, 1998 Glucosinolate content and composition as parameters of quality of *Camelina* seed. *Industrial Crops and Products* 7:297-302.

Scrucca, L., M. Fop, T.B. Murphy, and A.E. Raftery, 2016 mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal* 8 (1):289.

Séguin-Swartz, G., C. Eynck, R.K. Gugel, S.E. Strelkov, C.Y. Olivier *et al.*, 2009 Diseases of *Camelina sativa* (false flax). *Canadian Journal of Plant Pathology* 31:375-386.

Séguin-Swartz, G., J.A. Nettleton, C. Sauder, S.I. Warwick, and R.K. Gugel, 2013 Hybridization between *Camelina sativa* (L.) Crantz (false flax) and North American *Camelina* species. *Plant Breeding* 132:390-396.

Sheldon, C.C., D.T. Rouse, E.J. Finnegan, W.J. Peacock, and E.S. Dennis, 2000 The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC)*. *Proceedings of the National Academy of Sciences* 97 (7):3753-3758.

Shonnard, D.R., L. Williams, and T.N. Kalnes, 2010 Camelina-derived jet fuel and diesel: Sustainable advanced biofuels. *Environmental Progress & Sustainable Energy* 29 (3):382-392.

Simopoulos, A.P., 2002 The importance of the ratio of omega-6/omega-3 essential fatty acids. *Biomedicine and Pharmacotherapy* 56:365-379.

Simpson, G.G., P.P. Dijkwel, V. Quesada, I. Henderson, and C. Dean, 2003 *FY* is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* 113 (6):777-787.

Singh, R., V. Bollina, E.E. Higgins, W.E. Clarke, C. Eynck *et al.*, 2015 Single-nucleotide polymorphism identification and genotyping in *Camelina sativa*. *Molecular Breeding* 35 (1):35.

Smejkal, M., 1971 Revision der tschechoslowakischen Arten der Gattung *Camelina*-Crantz (Cruciferae). *Preslia*.

Snowdon, R., W. Köhler, W. Friedt, and A. Köhler, 1997 Genomic in situ hybridization in *Brassica* amphidiploids and interspecific hybrids. *Theoretical and Applied Genetics* 95 (8):1320-1324.

Stecher, G., K. Tamura, and S. Kumar, 2020 Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular biology and evolution* 37 (4):1237-1239.

Sun, H., S. Wu, G. Zhang, C. Jiao, S. Guo *et al.*, 2017 Karyotype stability and unbiased fractionation in the paleo-allotetraploid Cucurbita genomes. *Molecular plant* 10 (10):1293-1306.

Swiezewski, S., F. Liu, A. Magusin, and C. Dean, 2009 Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature genetics* 462 (7274):799.

Takada, S., A. Akter, E. Itabashi, N. Nishida, D.J. Shea *et al.*, 2019 The role of *FRIGIDA* and *FLOWERING LOCUS C* genes in flowering time of *Brassica rapa* leafy vegetables. *Scientific Reports* 9 (1):1-11.

Tamura, K., and M. Nei, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution* 10 (3):512-526.

Tang, H., M.R. Woodhouse, F. Cheng, J.C. Schnable, B.S. Pedersen *et al.*, 2012 Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190 (4):1563-1574.

Team, R.C., 2017 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www. R-project. org.

Team, R.C., 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www. R-project. org.

Tepfer, M., A. Hurel, F. Tellier, and E. Jenczewski, 2020 Evaluation of the progeny produced by interspecific hybridization between *Camelina sativa* and *C. microcarpa*. *Annals of Botany*.

Tercé-Laforgue, T., M. Bedu, C. Dargel-Grafin, F. Dubois, Y. Gibon *et al.*, 2013 Resolving the role of plant glutamate dehydrogenase: II. Physiological characterization of plants overexpressing the two enzyme subunits individually or simultaneously. *Plant Cell Physiology* 54 (10):1635-1647.

Toro, B.E.L., 2017 Genome Assembly of *Camelina microcarpa* Andrz. Ex DC, A step towards understanding genome evolution in Camelina. Carleton University, Ottawa, Ontario.

Vercruyssen, L., A. Verkest, N. Gonzalez, K.S. Heyndrickx, D. Eeckhout *et al.*, 2014 *ANGUSTIFOLIA3* binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. *The Plant Cell* 26 (1):210-229.

Vollmann, J., and C. Eynck, 2015 *Camelina* as a sustainable oilseed crop: Contributions of plant breeding and genetic engineering. *Biotechnology Journal* 10 (4):525-535.

Vollmann, J., H. Grausgruber, G. Stift, V. Dryzhyruk, and T. Lelley, 2005 Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. *Plant Breeding* 124:446-453.

Vollmann, J., T. Moritz, C. Kargl, S. Baumgartner, and H. Wagentristl, 2007 Agronomic evaluation of camelina genotypes selected for seed quality characteristics. *Industrial Crops and Products* 26 (3):270-277.

Voorrips, R., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity* 93 (1):77-78.

Walsh, D.T., E.M. Babiker, I.C. Burke, and S.H. Hulbert, 2012a *Camelina* mutants resistant to acetolactate synthase inhibitor herbicides. *Molecular Breeding* 30:1053-1063.

Walsh, K.D., M.J. Hills, S.L. Martin, and L.M. Hall, 2015 Pollen-mediated Gene Flow in *Camelina sativa* (L.) Crantz. *Crop Science* 55:196.

Walsh, K.D., D.M. Puttick, M.J. Hills, R.-C. Yang, K.C. Topinka *et al.*, 2012b Short Communication: First report of outcrossing rates in camelina [*Camelina sativa* (L.) Crantz], a potential platform for bioindustrial oils. *Canadian Journal of Plant Science* 92:681-685.

Wang, X., H. Wang, J. Wang, R. Sun, J. Wu *et al.*, 2011 The genome of the mesopolyploid crop species *Brassica rapa. Nature genetics* 43 (10):1035.

Warwick, S.I., and I.A. Al-Shehbaz, 2006 Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution* 259:237-248.

Wei, Z., M. Wang, S. Chang, C. Wu, P. Liu *et al.*, 2016 Introgressing subgenome components from *Brassica rapa* and *B. carinata* to *B. juncea* for broadening its genetic base and exploring intersubgenomic heterosis. *Frontiers in Plant Science* 7:1677.

Wittenberg, A., J.V. Anderson, and M.T. Berti, 2019 Winter and summer annual biotypes of *camelina* have different morphology and seed characteristics. *Industrial Crops and Products* 135:230-237.

Wiwart, M., D. Kurasiak-Popowska, E. Suchowilska, U. Wachowska, and K. Stuper-Szablewska, 2019 Variation in the morphometric parameters of seeds of spring and

winter genotypes of *Camelina sativa* (L.) Crantz. *Industrial Crops and Products* 139:111571.

Wu, J., L. Lin, M. Xu, P. Chen, D. Liu *et al.*, 2018 Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genomics* 19 (1):586.

Wu, Y., P. Bhat, T.J. Close, and S. Lonardi, 2007 Efficient and accurate construction of genetic linkage maps from noisy and missing genotyping data, pp. 395-406 in *International Workshop on Algorithms in Bioinformatics*. Springer.

Xiao, D., J.J. Zhao, X.L. Hou, R.K. Basnet, D.P. Carpio *et al.*, 2013 The *Brassica rapa FLC* homologue *FLC2* is a key regulator of flowering time, identified through transcriptional co-expression networks. *Journal of Experimental Botany* 64 (14):4503-4516.

Yang, L., H.-N. Wang, X.-H. Hou, Y.-P. Zou, T.-S. Han *et al.*, 2018 Parallel evolution of common allelic variants confers flowering diversity in *Capsella rubella. The Plant Cell* 30 (6):1322-1336.

Yang, M.H., and Y.B. Fu, 2017 AveDissR: An R function for assessing genetic distinctness and genetic redundancy. *Applications in plant sciences* 5 (7):1700018.

Yang, W., D. Liu, J. Li, L. Zhang, H. Wei *et al.*, 2009 Synthetic hexaploid wheat and its utilization for wheat genetic improvement in China. *Journal of Genetics and Genomics* 36 (9):539-546.

Ye, C.L., D.M. Anderson, and S.P. Lall, 2016 The effects of camelina oil and solvent extracted camelina meal on the growth, carcass composition and hindgut histology of Atlantic salmon (*Salmo salar*) parr in freshwater. *Aquaculture* 450:397-404.

Zelt, N.H., and D.J. Schoen, 2016 Testing for heterosis in traits associated with seed yield in *Camelina sativa*. *Canadian Journal of Plant Science* 529:525-529.

Zhang, C.-J., and C. Auer, 2020 Hybridization between *Camelina sativa* (L.) Crantz and common *Brassica* weeds. *Industrial Crops and Products* 147:112240.

Zhang, X., T. Liu, X. Li, M. Duan, J. Wang *et al.*, 2016 Interspecific hybridization, polyploidization and backcross of *Brassica oleracea* var. *alboglabra* with *B. rapa* var. *purpurea* morphologically recapitulate the evolution of *Brassica* vegetables. *Scientific Reports* 6 (1):18618.

Zhao, J., V. Kulkarni, N. Liu, D. Pino Del Carpio, J. Bucher *et al.*, 2010 *BrFLC2* (*FLOWERING LOCUS C*) as a candidate gene for a vernalization response QTL in *Brassica rapa*. *Journal of Experimental Botany* 61 (6):1817-1825.

Zhao, M., B. Zhang, D. Lisch, and J. Ma, 2017 Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *The Plant Cell* 29 (12):2974-2994.

Zhu, A., J.G. Ibrahim, and M.I. Love, 2019 Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* 35 (12):2084-2092.

Zou, X., I. Suppanz, H. Raman, J. Hou, J. Wang *et al.*, 2012 Comparative analysis of *FLC* homologues in Brassicaceae provides insight into their role in the evolution of oilseed rape. *PloS one* 7 (9):e45751.

Zubr, J., 1997 Oil-seed crop: *Camelina sativa*. *Industrial Crops and Products* 6:113-119.

Zubr, J., 2003 Qualitative variation of *Camelina sativa* seed from different locations. *Industrial Crops and Products* 17 (3):161-169.

Appendices

A.1 Chapter 4 Supplementary Tables and Figures

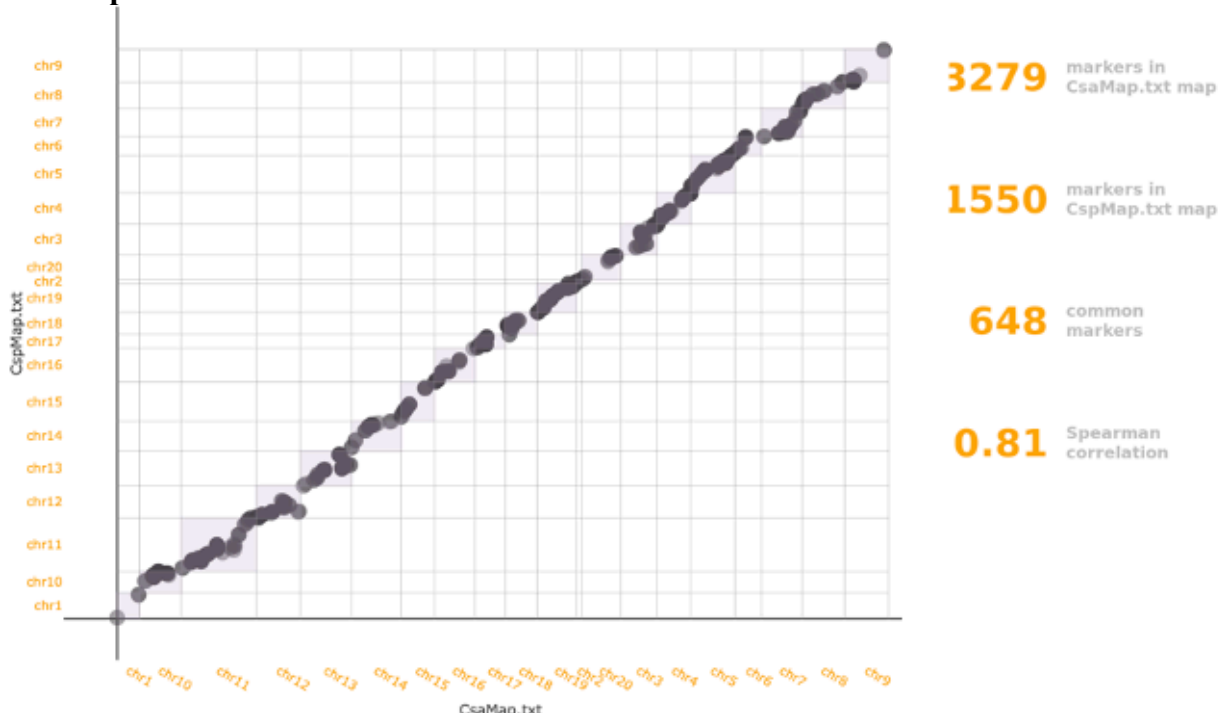
A.1.1 Distribution of markers on linkage map from TMP23992 × CN113692 (*C. sativa* × *C. sativa* ssp. *pilosa*) F_2 populations.

Subgenome	Chromosome	Genetic distance (cM)	Number of SNP	Average genetic distance per SNP
SG1	Chr14	114.35	53	2.16
	Chr7	110.11	61	1.81
	Chr19	111.15	97	1.15
	Chr4	120.92	98	1.23
	Chr8	100.48	67	1.50
	Chr11	205.82	158	1.30
		762.83	534	1.43
SG2	Chr3	116.99	91	1.29
	Chr16	130.97	106	1.24
	Chr1	100.60	22	4.57
	Chr6	71.12	63	1.13
	Chr13	134.70	93	1.45
	Chr10	82.09	71	1.16
	Chr18	85.28	39	2.19
		721.75	485	1.49
SG3	Chr17	51.91	36	1.44
	Chr5	144.79	139	1.04
	Chr15	150.52	39	3.86
	Chr9	126.04	82	1.54
	Chr20	95.53	53	1.80
	Chr2	16.46	16	1.03
	Chr12	123.98	166	0.75
		709.23	531	1.34

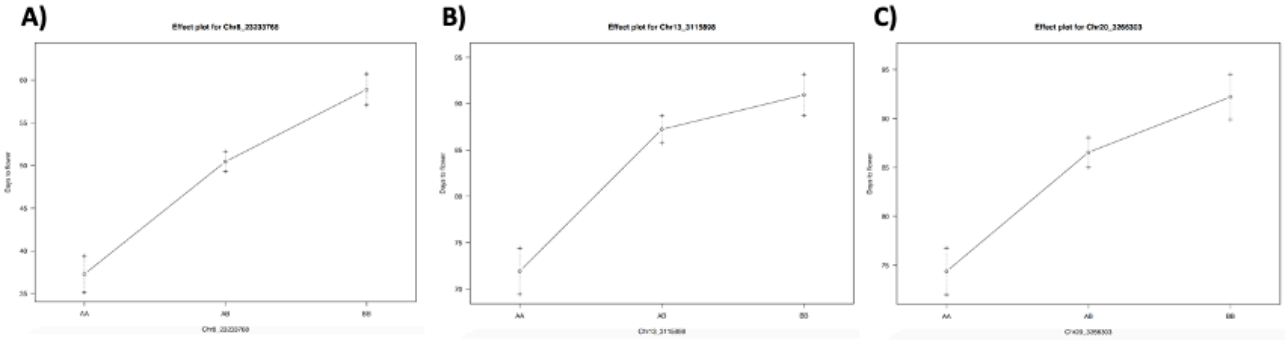
A.1.2 Distribution of markers on linkage map from TMP23992 × CAM176 (*C. sativa* × *C. alyssum*) F_2 populations.

Subgenome	Chromosome	Genetic distance (cM)	Number of SNP	Average genetic distance per SNP
SG1	Chr14	154.63	207	0.75
	Chr7	126.80	195	0.65
	Chr19	122.48	197	0.62
	Chr4	105.80	169	0.63
	Chr8	132.15	201	0.66
	Chr11	232.29	420	0.55
		874.15	1389	0.63
SG2	Chr3	112.69	125	0.90
	Chr16	126.22	143	0.88
	Chr1	70.85	100	0.71
	Chr6	82.43	62	1.33
	Chr13	157.88	136	1.16
	Chr10	129.59	173	0.75
	Chr18	99.19	86	1.15
		778.85	825	0.94
SG3	Chr17	94.45	93	1.02
	Chr5	137.22	190	0.72
	Chr15	103.23	104	0.99
	Chr9	137.19	206	0.67
	Chr20	119.04	170	0.70
	Chr2	17.62	50	0.35
	Chr12	138.22	252	0.55
		746.97	1065	0.70

A.1.3 Testing collinearity between genetic map Csa and Csp using online tool genetic map comparator.



A.1.4 Effect of markers on days to flower. A) Marker from QTL on chr 8 from Csp population. B) Marker from QTL on chr 13 from Csa population, and C) Marker from QTL on chr 20 from Csa population. Here, AA allele represented spring-type whereas BB allele represented winter-type.



A.1.5 Differential gene expression between TMP23992 (*C. sativa*) and CN113692 (*C. sativa* ssp. *pilosa*) in the QTL region on chromosome 8.

Arabidopsis thaliana identifier	geneid	logFC	AveExpr	t	P.Value	adj.P.Val
AT5G10780	Csa08g053730	0.923	3.988	8.719	0.000	0.003
	Csa08g053740	-1.320	2.183	-5.721	0.003	0.012
AT5G10760	Csa08g053750	3.793	2.573	14.537	0.000	0.001
	Csa08g053760	-2.047	1.993	-5.678	0.003	0.013
AT5G10730	Csa08g053790	1.590	1.738	6.149	0.002	0.010
AT5G10660	Csa08g053860	-1.307	-2.667	-8.065	0.001	0.004
AT5G10650	Csa08g053870	-3.157	2.502	-23.431	0.000	0.000
AT5G10630	Csa08g053880	1.013	0.840	4.602	0.006	0.025
AT5G10550	Csa08g053960	1.033	4.250	6.289	0.002	0.009
	Csa08g053970	1.280	3.117	10.896	0.000	0.002
AT5G10520	Csa08g054000	-2.310	-1.185	-3.967	0.012	0.039
AT5G10480	Csa08g054040	-3.223	2.592	-7.145	0.001	0.006
AT5G10470	Csa08g054050	-0.833	2.937	-4.813	0.005	0.021
	Csa08g054070	-2.840	-0.673	-3.950	0.012	0.039
AT5G10450	Csa08g054080	-3.050	7.965	-18.317	0.000	0.000
AT5G10440	Csa08g054090	-4.547	-0.503	-7.661	0.001	0.005
AT5G10380	Csa08g054140	2.960	1.490	6.674	0.001	0.007
AT5G10370	Csa08g054150	-1.037	1.145	-6.779	0.001	0.007
AT5G10360	Csa08g054160	1.753	5.720	16.192	0.000	0.001
	Csa08g054210	2.103	1.828	4.327	0.008	0.030
AT5G10320	Csa08g054230	-1.667	1.193	-6.628	0.001	0.008
AT5G10280	Csa08g054280	-3.877	-1.382	-19.088	0.000	0.000
	Csa08g054290	-8.540	0.950	-65.802	0.000	0.000
	Csa08g054300	-0.940	2.577	-5.204	0.004	0.017
AT5G10170	Csa08g054420	2.233	0.113	4.964	0.005	0.019
AT5G10160	Csa08g054430	0.823	2.772	5.391	0.003	0.015
AT5G10140	Csa08g054450	-1.463	5.335	-6.942	0.001	0.007
	Csa08g054460	-6.843	0.102	-25.592	0.000	0.000
AT5G10110	Csa08g054490	-0.570	2.078	-3.798	0.014	0.044
AT5G10100	Csa08g054500	-3.403	-1.055	-5.750	0.003	0.012
AT5G10090	Csa08g054510	-2.000	-2.320	-10.956	0.000	0.002
AT5G10080	Csa08g054520	-3.327	1.557	-13.052	0.000	0.001
AT5G10060	Csa08g054540	-2.017	2.258	-4.228	0.009	0.032
AT5G10050	Csa08g054550	1.533	-0.193	4.003	0.011	0.038
	Csa08g054560	-3.770	-1.435	-6.388	0.002	0.009
AT5G10030	Csa08g054570	-1.403	2.138	-5.045	0.004	0.018
AT5G10010	Csa08g054590	1.437	2.092	6.787	0.001	0.007

AT5G09980	Csa08g054650	-5.593	-0.523	-17.898	0.000	0.000
AT5G09930	Csa08g054710	-4.953	-0.843	-14.929	0.000	0.001
AT5G09900	Csa08g054740	1.270	3.172	7.679	0.001	0.005
AT5G09880	Csa08g054760	-1.580	3.197	-17.840	0.000	0.000
AT5G09870	Csa08g054780	1.927	2.690	14.436	0.000	0.001
AT5G09850	Csa08g054800	-0.543	2.775	-4.440	0.007	0.027
AT5G09830	Csa08g054820	1.683	3.728	23.489	0.000	0.000
AT5G09820	Csa08g054830	1.580	3.713	8.868	0.000	0.003
	Csa08g054900	1.437	1.958	6.458	0.002	0.008
AT5G09670	Csa08g054970	-1.337	2.732	-17.100	0.000	0.000
AT5G09640	Csa08g054990	-1.397	-2.622	-7.445	0.001	0.005
	Csa08g055020	-1.477	-2.582	-5.487	0.003	0.014
AT5G09530	Csa08g055120	-3.743	-0.835	-5.791	0.002	0.012
AT5G09460	Csa08g055180	-1.653	3.753	-5.273	0.004	0.016
AT5G09450	Csa08g055190	1.517	2.388	6.953	0.001	0.007
AT5G09440	Csa08g055200	-3.190	3.988	-5.733	0.003	0.012
AT5G09390	Csa08g055250	-1.157	3.145	-6.860	0.001	0.007
AT5G09300	Csa08g055350	1.207	1.520	4.651	0.006	0.024
AT5G09225	Csa08g055420	-2.933	-0.260	-3.677	0.015	0.049
AT5G09220	Csa08g055430	-3.433	3.673	-8.439	0.000	0.004
AT5G08780	Csa08g055450	-0.933	1.180	-3.753	0.014	0.046
AT5G08690	Csa08g055540	3.413	3.967	16.495	0.000	0.000
AT5G08680	Csa08g055550	1.627	2.280	5.061	0.004	0.018
	Csa08g055560	2.353	4.077	10.111	0.000	0.002
AT5G08650	Csa08g055580	0.613	4.550	19.908	0.000	0.000
AT5G08640	Csa08g055590	3.373	1.593	6.545	0.001	0.008
AT5G08630	Csa08g055600	1.407	1.387	12.622	0.000	0.001
	Csa08g055610	0.670	3.102	4.609	0.006	0.024
AT5G08590	Csa08g055640	-1.317	4.715	-11.060	0.000	0.001
AT5G08570	Csa08g055660	1.120	3.490	9.373	0.000	0.003
AT5G08560	Csa08g055680	-1.273	3.263	-18.456	0.000	0.000
AT5G08540	Csa08g055700	1.723	3.122	16.326	0.000	0.001
AT5G08535	Csa08g055710	-2.430	2.402	-6.728	0.001	0.007
AT5G08530	Csa08g055720	1.587	4.043	10.889	0.000	0.002
AT5G08520	Csa08g055730	-0.943	3.532	-6.818	0.001	0.007
AT5G08510	Csa08g055770	2.107	-1.627	7.624	0.001	0.005
AT5G08470	Csa08g055820	-1.860	2.137	-9.884	0.000	0.002
	Csa08g055850	-3.713	-1.463	-21.060	0.000	0.000
	Csa08g055860	-5.733	-0.453	-17.426	0.000	0.000
AT5G08430	Csa08g055890	-1.233	1.790	-4.885	0.005	0.020
AT5G08420	Csa08g055900	1.343	2.912	10.852	0.000	0.002
AT5G08410	Csa08g055920	2.140	5.290	17.089	0.000	0.000
AT5G08400	Csa08g055930	2.613	1.597	12.180	0.000	0.001

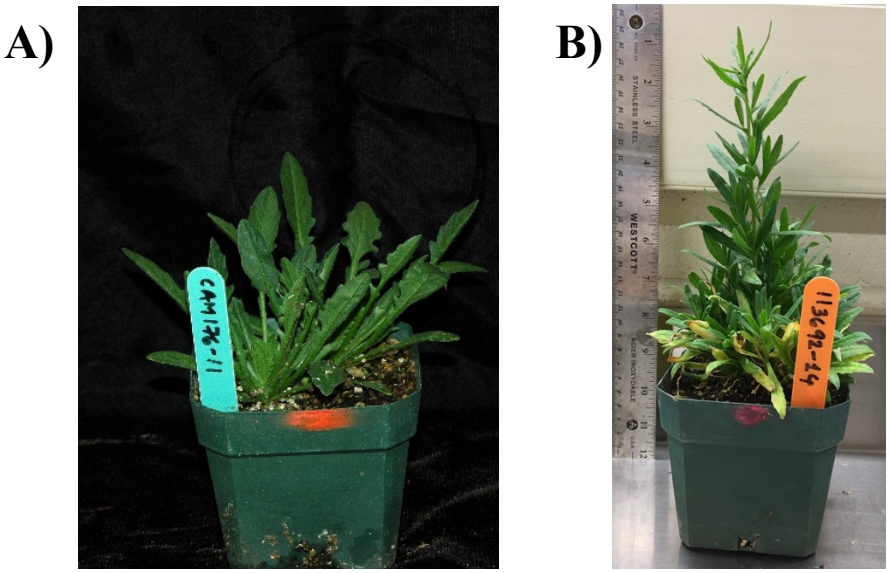
AT5G08390	Csa08g055940	-1.190	1.525	-11.371	0.000	0.001
AT5G08370	Csa08g055970	-1.310	2.512	-4.169	0.010	0.033
	Csa08g055990	-5.147	-0.747	-15.811	0.000	0.001
AT5G08335	Csa08g056020	4.243	-1.198	8.815	0.000	0.003
AT5G08280	Csa08g056110	2.723	4.745	26.364	0.000	0.000
	Csa08g056130	-2.913	-1.863	-8.462	0.000	0.003
AT5G08240	Csa08g056150	-4.687	-0.977	-13.816	0.000	0.001
	Csa08g056160	-1.033	0.607	-4.230	0.009	0.032
AT5G08190	Csa08g056180	-1.323	1.648	-5.799	0.002	0.012
AT5G08180	Csa08g056190	2.903	2.895	8.340	0.000	0.004
AT5G08160	Csa08g056210	1.083	2.998	7.724	0.001	0.005
AT5G08139	Csa08g056230	-1.853	1.407	-5.949	0.002	0.011
AT5G08130	Csa08g056240	-1.643	2.712	-8.835	0.000	0.003
AT5G08100	Csa08g056270	-1.787	3.017	-14.819	0.000	0.001
AT5G08070	Csa08g056300	-1.820	-2.410	-5.576	0.003	0.013
	Csa08g056320	-2.690	1.402	-3.770	0.014	0.045
	Csa08g056340	-2.873	-0.350	-3.650	0.016	0.050
AT5G08050	Csa08g056370	2.203	5.422	14.230	0.000	0.001
AT5G08040	Csa08g056380	1.383	4.358	5.802	0.002	0.012
AT5G08020	Csa08g056400	3.153	0.560	12.705	0.000	0.001
AT5G07990	Csa08g056450	4.050	-0.138	5.137	0.004	0.017
AT5G07970	Csa08g056460	-1.137	2.122	-14.889	0.000	0.001
AT5G07920	Csa08g056520	-1.820	2.927	-11.864	0.000	0.001
AT5G07900	Csa08g056540	1.917	2.495	7.746	0.001	0.005
AT5G07880	Csa08g056560	-2.003	-2.318	-8.138	0.001	0.004
AT5G07830	Csa08g056600	-1.370	4.528	-9.202	0.000	0.003
	Csa08g056650	-0.630	1.672	-4.864	0.005	0.021
AT5G07710	Csa08g056680	-1.320	0.810	-6.571	0.001	0.008
	Csa08g056700	-2.787	-1.340	-4.862	0.005	0.021
AT5G07680	Csa08g056710	-1.727	-2.457	-6.934	0.001	0.007
AT5G07670	Csa08g056720	-1.117	3.125	-6.362	0.002	0.009
AT5G07660	Csa08g056730	-3.720	0.457	-4.065	0.011	0.036
AT5G07580	Csa08g056810	2.047	3.067	7.824	0.001	0.004
AT5G07470	Csa08g056930	-1.587	1.980	-11.434	0.000	0.001
AT5G07440	Csa08g056970	-2.613	3.760	-7.246	0.001	0.006
AT5G07380	Csa08g057040	-1.753	-2.443	-8.585	0.000	0.003
AT5G07330	Csa08g057080	-5.427	-0.607	-17.715	0.000	0.000
AT5G07310	Csa08g057110	-1.667	-2.487	-5.903	0.002	0.011
AT5G07290	Csa08g057130	-2.347	2.160	-14.628	0.000	0.001
AT5G07240	Csa08g057190	3.537	0.342	12.437	0.000	0.001
AT5G07180	Csa08g057260	-2.200	-1.420	-4.848	0.005	0.021
AT5G07170	Csa08g057270	-0.930	-2.855	-6.320	0.002	0.009
AT5G07090	Csa08g057350	2.303	5.698	29.488	0.000	0.000

	Csa08g057380	-4.173	1.023	-3.801	0.014	0.044
AT5G07030	Csa08g057430	1.903	1.652	7.653	0.001	0.005
AT5G07020	Csa08g057440	3.300	5.200	56.939	0.000	0.000
	Csa08g057450	-6.713	0.567	-12.345	0.000	0.001
AT5G06980	Csa08g057470	-6.100	2.737	-12.212	0.000	0.001
	Csa08g057510	-0.833	1.613	-4.739	0.006	0.022
AT5G06940	Csa08g057520	2.583	-0.762	6.445	0.002	0.008
AT5G06905	Csa08g057570	-0.927	-2.857	-4.803	0.005	0.021
AT5G06839	Csa08g057630	-0.800	-2.920	-5.052	0.004	0.018
AT5G06830	Csa08g057640	1.770	0.948	5.116	0.004	0.018
AT5G06800	Csa08g057670	-4.217	-0.825	-7.895	0.001	0.004
AT5G06760	Csa08g057710	-6.367	-0.137	-18.321	0.000	0.000
AT5G06750	Csa08g057730	0.747	1.973	4.281	0.009	0.031
AT5G06710	Csa08g057770	-2.747	0.323	-8.022	0.001	0.004
AT5G06700	Csa08g057780	0.670	3.298	5.044	0.004	0.018
AT5G06690	Csa08g057790	1.857	2.502	4.952	0.005	0.019
AT5G06660	Csa08g057820	-1.170	3.035	-7.137	0.001	0.006
AT5G06600	Csa08g057860	1.567	2.827	19.390	0.000	0.000
AT5G06560	Csa08g057900	-2.007	1.350	-9.764	0.000	0.002
AT5G06550	Csa08g057910	3.220	-0.930	7.642	0.001	0.005
AT5G06530	Csa08g057930	-4.043	3.932	-17.478	0.000	0.000
	Csa08g057940	-3.123	-0.585	-5.030	0.005	0.018
AT5G06480	Csa08g057990	-3.360	-1.640	-24.507	0.000	0.000
AT5G06460	Csa08g058010	1.103	1.152	4.206	0.009	0.032
AT5G06450	Csa08g058020	-1.100	3.003	-3.834	0.013	0.043
AT5G06440	Csa08g058030	-1.237	3.532	-4.688	0.006	0.023
	Csa08g058050	-1.087	-2.777	-5.700	0.003	0.012
AT5G06390	Csa08g058090	2.053	1.763	8.638	0.000	0.003
AT5G06370	Csa08g058120	-2.010	4.445	-12.626	0.000	0.001
AT5G06360	Csa08g058130	-0.503	4.715	-6.849	0.001	0.007
AT5G06350	Csa08g058140	0.930	1.382	4.696	0.006	0.023
AT5G06300	Csa08g058180	-4.320	-0.067	-6.704	0.001	0.007
AT5G06280	Csa08g058190	-0.583	2.178	-3.750	0.014	0.046
	Csa08g058200	-5.263	-0.688	-15.792	0.000	0.001
	Csa08g058220	-6.493	-0.073	-24.170	0.000	0.000
	Csa08g058230	-1.537	5.012	-6.536	0.001	0.008
AT5G06260	Csa08g058260	0.323	4.162	5.410	0.003	0.015
	Csa08g058270	9.380	1.370	30.857	0.000	0.000
AT5G06230	Csa08g058300	-4.030	-0.102	-6.372	0.002	0.009
AT5G06220	Csa08g058310	-1.160	4.107	-11.252	0.000	0.001
AT5G06200	Csa08g058330	-4.547	0.047	-4.315	0.008	0.030
AT5G06130	Csa08g058400	-1.023	4.405	-11.093	0.000	0.001
AT5G06120	Csa08g058420	-1.197	1.345	-5.865	0.002	0.011

AT5G06100	Csa08g058440	-3.307	0.697	-9.507	0.000	0.002
AT5G05990	Csa08g058510	2.460	0.663	6.253	0.002	0.009
AT5G05987	Csa08g058520	-1.383	2.455	-4.237	0.009	0.032
AT5G05980	Csa08g058530	-1.040	2.247	-15.715	0.000	0.001
AT5G05930	Csa08g058580	-4.273	2.333	-6.732	0.001	0.007
	Csa08g058610	-1.080	1.313	-5.945	0.002	0.011
AT5G05840	Csa08g058630	3.737	-0.145	8.447	0.000	0.004
AT5G05800	Csa08g058680	-1.440	2.067	-5.082	0.004	0.018
AT5G05750	Csa08g058730	-2.440	3.940	-9.493	0.000	0.002
	Csa08g058740	3.240	3.333	31.709	0.000	0.000
AT5G05700	Csa08g058780	-1.147	1.503	-4.850	0.005	0.021
AT5G05690	Csa08g058790	-1.707	3.380	-9.001	0.000	0.003
AT5G05580	Csa08g058890	3.943	3.138	25.039	0.000	0.000
AT5G05520	Csa08g058970	0.510	3.188	3.996	0.011	0.038
AT5G05460	Csa08g059040	3.867	0.390	6.264	0.002	0.009
AT5G05440	Csa08g059060	-2.377	0.315	-3.797	0.014	0.044
	Csa08g059100	3.137	4.855	8.061	0.001	0.004
AT5G05340	Csa08g059130	-3.323	-1.658	-22.328	0.000	0.000
AT5G05230	Csa08g059230	-1.347	1.333	-5.662	0.003	0.013
AT5G05170	Csa08g059290	0.647	4.467	5.392	0.003	0.015
AT5G05140	Csa08g059320	-3.143	2.632	-8.343	0.000	0.004
AT5G05120	Csa08g059330	-5.577	-0.532	-17.583	0.000	0.000
AT5G05110	Csa08g059340	-3.870	4.512	-12.450	0.000	0.001
AT5G05080	Csa08g059370	-1.997	5.115	-13.133	0.000	0.001
AT5G05070	Csa08g059380	-1.613	-2.513	-4.086	0.010	0.035
AT5G05060	Csa08g059390	-3.857	-1.392	-11.577	0.000	0.001
AT5G05010	Csa08g059410	2.513	2.937	20.978	0.000	0.000
	Csa08g059430	2.807	-0.027	7.979	0.001	0.004
AT5G04990	Csa08g059450	4.457	-1.092	17.263	0.000	0.000
AT5G04980	Csa08g059460	-1.650	0.782	-5.931	0.002	0.011
AT5G04940	Csa08g059500	1.470	1.182	8.102	0.001	0.004
AT5G04930	Csa08g059510	1.253	1.817	5.384	0.003	0.015
AT5G04910	Csa08g059530	-2.220	2.637	-4.168	0.010	0.033
	Csa08g059560	-4.220	-1.210	-15.414	0.000	0.001
	Csa08g059570	-3.217	-1.712	-14.409	0.000	0.001
AT5G04885	Csa08g059600	2.747	1.900	8.825	0.000	0.003
AT5G04870	Csa08g059610	1.313	3.060	5.204	0.004	0.017
AT5G04860	Csa08g059620	-0.887	1.553	-6.005	0.002	0.011
AT5G04830	Csa08g059650	-1.840	2.877	-6.297	0.002	0.009
AT5G04800	Csa08g059680	2.437	5.798	21.832	0.000	0.000
AT5G04770	Csa08g059720	-3.163	-1.292	-6.019	0.002	0.010
AT5G04740	Csa08g059750	1.243	4.735	9.808	0.000	0.002
	Csa08g059760	-7.233	1.263	-8.104	0.001	0.004

AT5G04720	Csa08g059780	1.240	3.267	5.177	0.004	0.017
AT5G04710	Csa08g059790	-1.030	1.782	-5.257	0.004	0.016
AT5G04590	Csa08g059880	1.537	5.755	14.122	0.000	0.001

A.1.6 Nature of plant growth in two different winter type. A) Plant growth of CAM176 (*C. alyssum)* with reduced stem and profuse leaves and B) Plant growth of CN113692 (*C. sativa* ssp. *pilosa*) characterise by stem elongation and profuse branching in the absence of vernalization.



A.1.7 Effect of duration of vernalization on flowering on hybrids coming from same parents. Plant kept in vernalization for 1 week has few flowers, whereas for another plant kept for 3 weeks into vernalization possess many reproductive branches with flowers (95 days after seeding)

Kept for 1 week in vernalization

Kept for 3 weeks in vernalization

A.1.8 Days to first flowering for the hybrids developed from *C. sativa* × *C. alyssum* winter type (TMP23992 × CAM176).

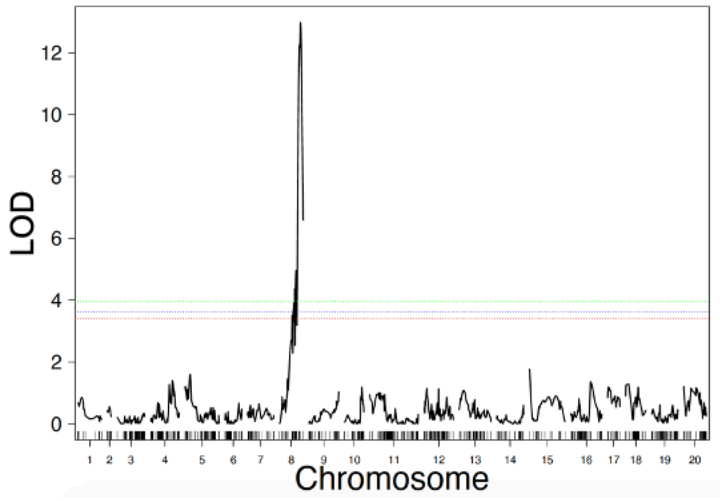
(DTG: Days to germination; VER: vernalization period; DTF: days to first flower from seeding; DTFv: days to first flower after vernalization; PH: Plant height, and Branch: number of primary branches).

Hybrid	DTG	VER	DTF	DTFv	PH	Branch
61CSA1	4	7	76	49	63	32
61CSA2	4	7	78	54	66	39
61CSA3	4	7	69	45	59	40
62CSA1	7	7	80	56	53	28
62CSA2	4	7	67	43	45	36
62CSA3	4	7	75	51	58	44
63CSA1	3	7	75	51	56	44
63CSA2	4	7	78	54	47	35
56CSA1	3	14	58	30	65	23
56CSA2	3	14	57	29	67	16
56CSA3	3	14	56	28	68	17
56CSA4	5	14	58	30	67	16
56CSA5	4	14	59	31	72	12
56CSA6	4	14	58	30	68	18
56CSA8	7	14	57	29	72	14
56CSA9	3	14	56	28	73	15
79CSA1	5	21	61	23	67	20
79CSA2	5	21	64	26	64	22
79CSA3	3	21	61	23	63	17
80CSA1	6	21	64	26	63	18
80CSA2	5	21	61	23	67	19
80CSA3	6	21	63	25	63	17
81CSA1	4	21	60	22	56	10
81CSA2	4	21	61	23	53	10
81CSA3	4	21	60	22	57	27
81CSA4	6	21	62	24	56	14

A.1.9 Revised syntelog matrix adopted from Kagale et al. 2014 and revised based on Chapter 3.

https://1drv.ms/x/s!Ah1gWPis3PD6gpxCrux4a_4sML8t7g?e=cK2zTf

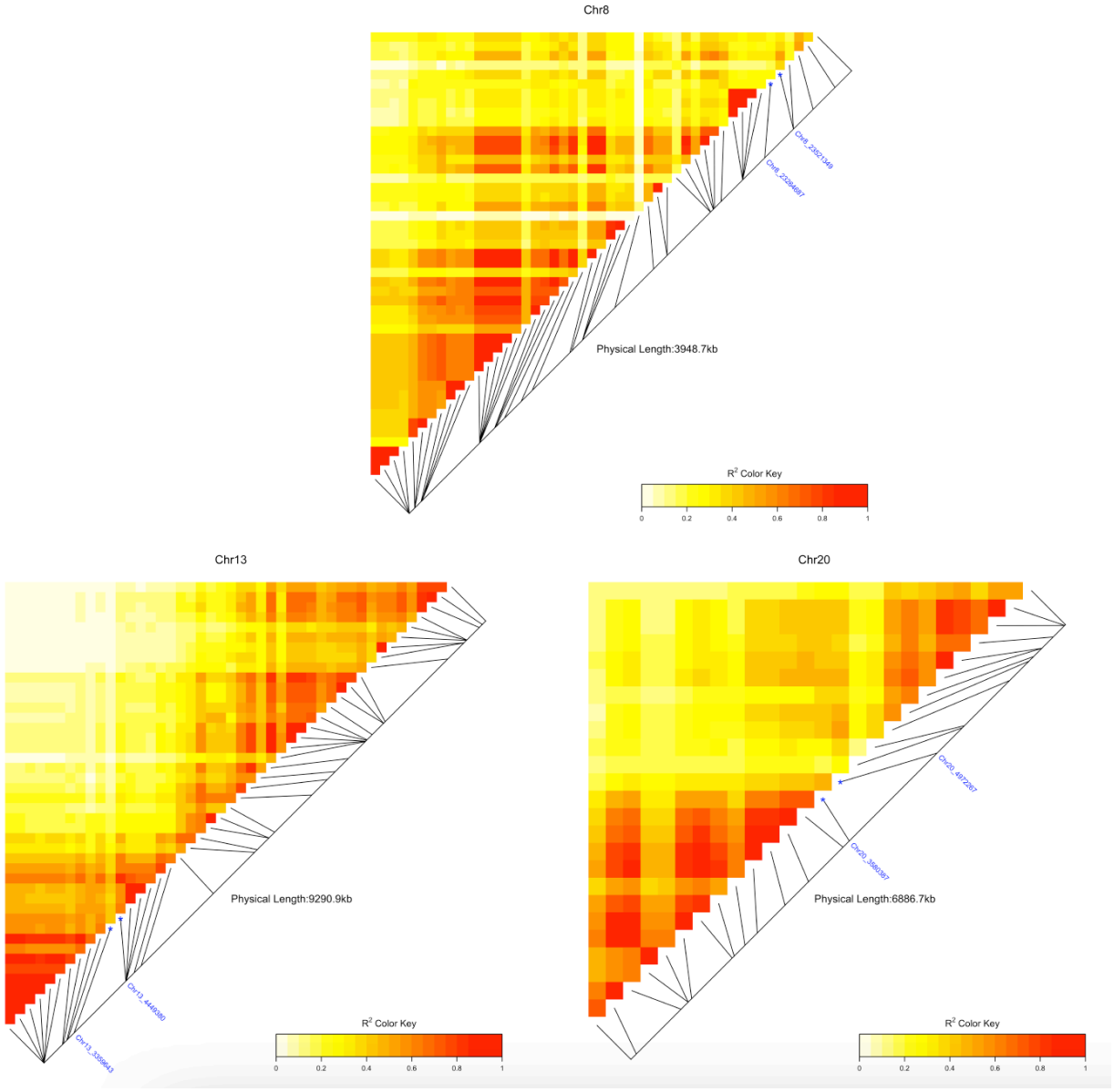
A.1.10 QTL mapping of days to flowering from $F_{2:3}$ derived from Csp population.



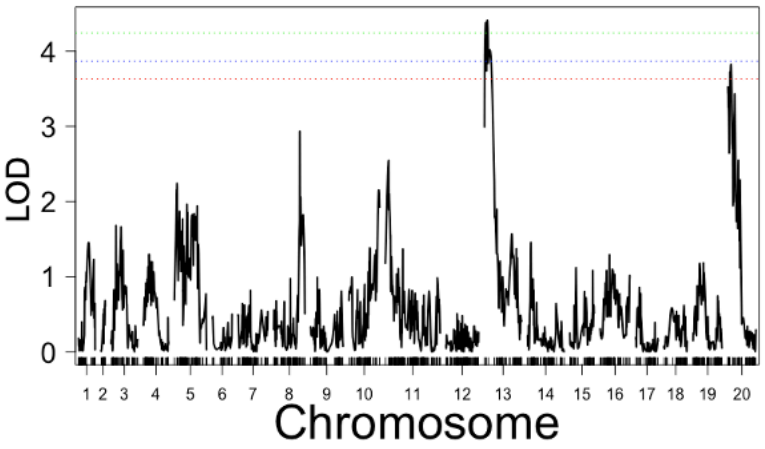
fitqtl summary

```
Method: Haley-Knott regression
Model:  normal phenotype
Number of observations : 96

Full model result
----------------------------------
Model formula: y ~ Q1

      df       SS        MS     LOD     %var Pvalue(Chi2)    Pvalue(F)
Model  2 1849.249 924.62455 12.9555 46.28523 1.108003e-13 2.814415e-13
Error 93 2146.084  23.07617
Total 95 3995.333
```

A.1.11 Linkage disequilibrium heatmap showing relationship of markers around the QTL regions on chromosome 8, chromosome 13 and chromosome 20. The marker label on blue represents flanking markers around the *Flowering Locus C*.



A.1.12 QTL mapping of days to flowering from $F_{2:3}$ derived from Csa population.




```
                fitqtl summary

Method: Haley-Knott regression
Model:  normal phenotype
Number of observations : 161

Full model result
----------------------------------
Model formula: y ~ Q1 + Q2

       df       SS        MS      LOD     %var Pvalue(Chi2)    Pvalue(F)
Model   4 19214.46 4803.6144 8.286923 21.10366 1.037217e-07 1.631174e-07
Error 156 71833.53  460.4713
Total 160 91047.99


Drop one QTL at a time ANOVA table:
----------------------------------
        df Type III SS   LOD   %var F value Pvalue(Chi2) Pvalue(F)
13@12.0  2        9830 4.484 10.796  10.674            0  4.52e-05 ***
20@12.0  2        8462 3.893  9.294   9.188            0  0.000169 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A.1.13 Comparison of *Flowering Locus C* sequence from on chromosome 8 for DH55 (reference genome), TMP23992 (*C. sativa*) and CAM176 (*C. alyssum*) from subgenome 1.

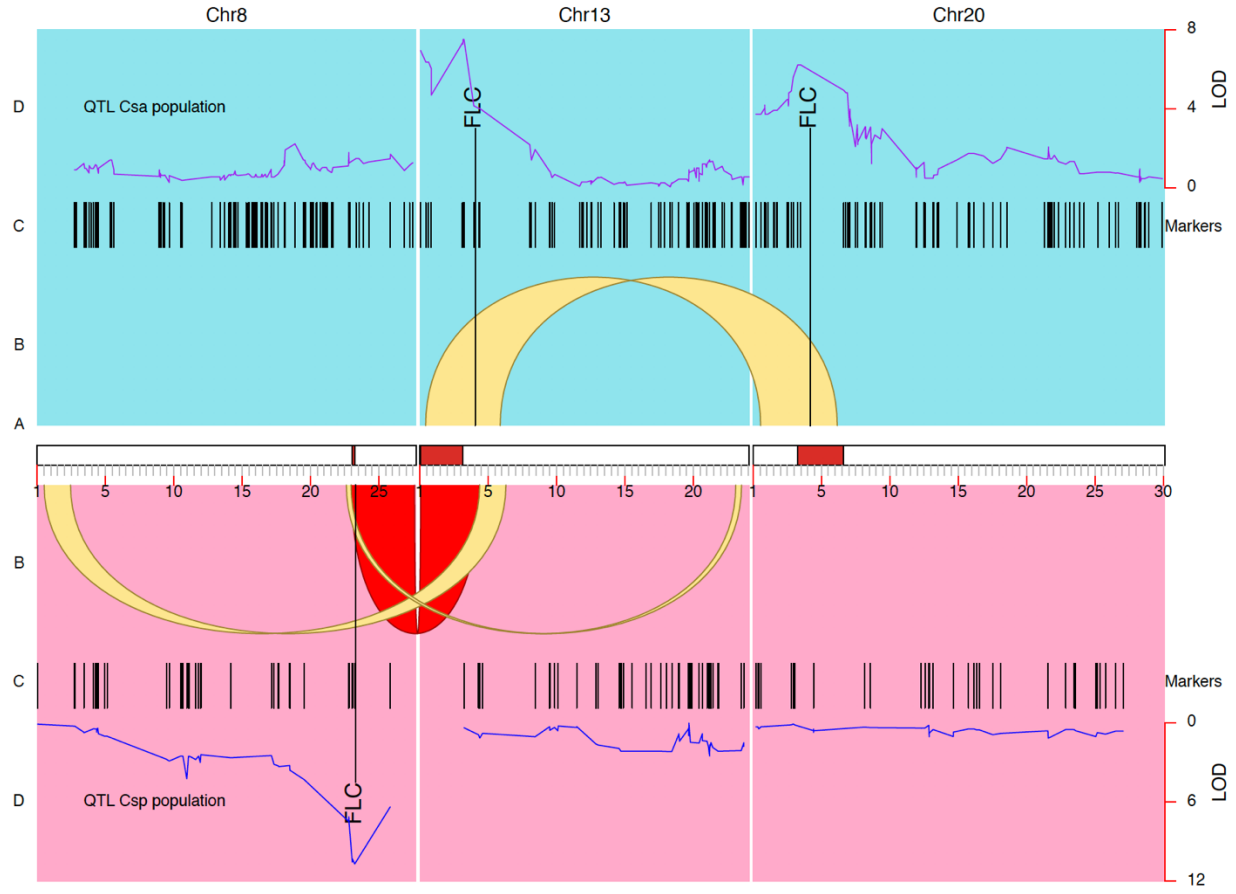
```
CLUSTAL O(1.2.4) multiple sequence alignment


Csa08g054450_DH55         ATGGGAAGGAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACC  60
Csa08g054450TMP23992      ATGGGAAGGAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACC  60
Csa08g054450CAM176        ATGGGAAGGAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACC  60
                          ************************************************************

Csa08g054450_DH55         TTCTCCAAACGTCGCAATGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGAC  120
Csa08g054450TMP23992      TTCTCCAAACGTCGCAATGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGAC  120
Csa08g054450CAM176        TTCTCCAAACGTCGCAATGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGAC  120
                          ************************************************************

Csa08g054450_DH55         GCATCCGTCGCTCTTCTCGTCGTCTCCGCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGT  180
Csa08g054450TMP23992      GCATCCGTCGCTCTTCTCGTCGTCTCCGCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGT  180
Csa08g054450CAM176        GCATCCGTCGCTCTTCTCGTCGTCTCCGCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGT  180
                          ************************************************************

Csa08g054450_DH55         GATAACCTGGTCAAGATCCTTGATCGATATGGGAAACAACATGCTGATGATCTCAAAGCC  240
Csa08g054450TMP23992      GATAACCTGGTCAAGATCCTTGATCGATATGGGAAACAACATGCTGATGATCTCAAAGCC  240
Csa08g054450CAM176        GATAACCTGGTCAAGATCCTTGATCGATATGGGAAACAACATGCTGATGATCTCAAAGCC  240
                          ************************************************************

Csa08g054450_DH55         TTGGATCTTCAGTCAAAAGCTCTGAACTATGGTTCGCACCATGAGCTACTAGAACTCGTG  300
Csa08g054450TMP23992      TTGGATCTTCAGTCAAAAGCTCTGAACTATGGTTCGCACCATGAGCTACTAGAACTCGTG  300
Csa08g054450CAM176        TTGGATCTTCAGTCAAAAGCTCTGAACTATGGTTCTCACCATGAGCTACTAGAACTCGTG  300
                          ********************************** ************************

Csa08g054450_DH55         GAAAGCAATCTTGTGGAATCAAATGTCAATAATGTAAGTGTCGATGCCCTCGTTC--TGG  358
Csa08g054450TMP23992      GAAAGCAATCTTGTGGAATCAAATGTCAATAATGTAAGTGTCGATGCCCTCGTTC--TGG  358
Csa08g054450CAM176        GAAAGCAATCTTGTGGAATCAAATGTCAATAATGTAAGTGTCGATGCCCTCGTTCAATGG  360
                          ************************************************  ***

Csa08g054450_DH55         AGGAACACCTTGAGACCGCCCTCTCCGTAACTAGTGCCAAGAAGACAGAACTAATGTTGA  418
Csa08g054450TMP23992      AGGAACACCTTGAGACCGCCCTCTCCGTAACTAGTGCCAAGAAGACAGAACTAATGTTGA  418
Csa08g054450CAM176        AGGAACACCTTGAGACCGCCCTCTCCGTAACGAGTGCCAAGAAGACAGAACTAATGTTGA  420
                          ****************************** ****************************

Csa08g054450_DH55         AGCTTGTTGAGAACCTCAAAGAAAAGGAAAAATTGCTGAAAGAAGAGAACCAGGTTTTGG  478
Csa08g054450TMP23992      AGCTTGTTGAGAACCTCAAAGAAAAGGAAAAATTGCTGAAAGAAGAGAACCAGGTTTTGG  478
Csa08g054450CAM176        AGCTTGTTGAGAACCTCAAAGAAAAGGAAAAATTGCTGAAAGAAGAGAACCAGGTTTTGG  480
                          ************************************************************

Csa08g054450_DH55         CTAGCCAGATGGAGACGAATCATGTTGTTGGAGCAGAAGCTGATATGGAGATGGAGATGT  538
Csa08g054450TMP23992      CTAGCCAGATGGAGACGAATCATGTTGTTGGAGCAGAAGCTGATATGGAGATGGAGATGT  538
Csa08g054450CAM176        CTAGCCAGATGGAGACGAATCATGTTGTTGGAGCAGAAGCTGATATGGAGATGGAGATGT  540
                          ************************************************************

Csa08g054450_DH55         CACCTGTTGGACAAATCTCCGACAATCTTCCGGTGACTCTCCCGCTGCTCAATTAG      594
Csa08g054450TMP23992      CACCTGTTGGACAAATCTCCGACAATCTTCCGGTGACTCTCCCGCTGCTCAATTAG      594
Csa08g054450CAM176        CACCTGTTGGACAAATCTCCGACAATCTTCCGGTGACTCTCCCGCTGCTCAATTAG      596
                          ************************************************************
```

A.1.14 Differential gene expression between TMP23992 (*C. sativa*) and CAM176 (*C. alyssum*) around QTL regions on chromosome 13 and chromosome 20.

geneid	*A. thaliana* identifier	Log fold change in CAM176	Gene function
Csa13g002260	AT5G02370	-0.351	
Csa13g002440	AT5G02540	-0.331	
Csa13g002650	AT5G02800	1.261	
Csa13g003160	AT5G03230	-0.329	senescence regulator
Csa13g003380	AT5G03430	-0.938	
Csa13g003750	AT5G03700	1.002	
Csa13g003960	AT5G03860	0.613	
Csa13g006680	AT5G04720	0.863	*ADR1-L2, ADR1-LIKE 2, PHOENIX 21, PHX21*
Csa13g009030	AT5G07010	2.653	*ARABIDOPSIS THALIANA SULFOTRANSFERASE 2A*
Csa13g009470	AT5G07440	2.877	
Csa13g010260	AT5G08260	-0.626	
Csa13g011100	AT5G09440	0.243	
Csa13g011640	AT5G09930	2.059	
Csa13g011710	AT5G09990	-0.608	
Csa13g011890	AT5G10140	4.562	*FLOWERING LOCUS C*
Csa13g012010	AT5G10250	-0.240	
Csa13g012150	AT5G10380	2.694	*ARABIDOPSIS TÃ³XICOS EN LEVADURA 55*
Csa20g002890	AT5G02840	-0.575	*LCL1*
Csa20g003000	AT5G02940	-0.757	
Csa20g003240	AT5G03230	1.263	
Csa20g005420	AT5G04220	2.739	
Csa20g005920	AT5G04720	0.286	
Csa20g006630	AT5G05365	1.076	
Csa20g006870	AT5G05600	3.198	*JASMONATE-INDUCED OXYGENASE2, JASMONIC ACID OXIDASE 2*
Csa20g008010	AT5G05750	1.873	
Csa20g008370	AT5G06160	1.979	
Csa20g008600	AT5G06360	-0.858	
Csa20g008610	AT5G06370	0.466	
Csa20g009440	AT5G07080	-0.459	

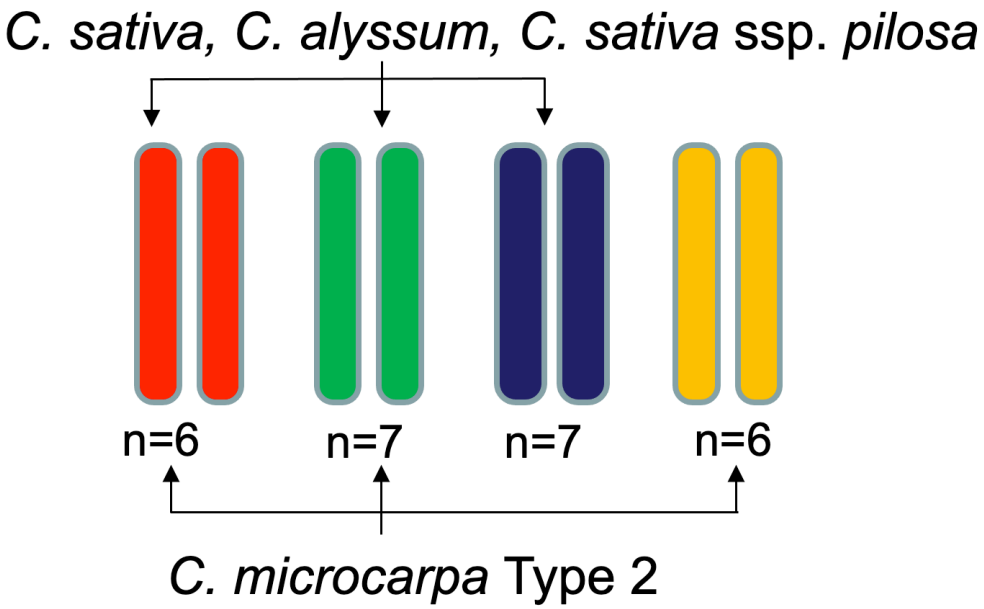
Csa20g009840	AT5G07440	3.501	*GLUTAMATE DEHYDROGENASE 2*
Csa20g011660	AT5G08280	0.429	
Csa20g012380	AT5G09220	-0.231	
Csa20g012590	AT5G09440	1.167	*EXORDIUM LIKE 4*
Csa20g013130	AT5G09930	3.676	*ABCF2, ATP-BINDING CASSETTE F2*
Csa20g015400	AT5G10140	2.426	
Csa20g015600	AT5G10300	4.910	*THALIANA METHYL ESTERASE 5*
Csa20g015640	AT5G10320	-0.257	

A.1.15 Summary of QTL mapping for vernalization requirement in *C. sativa.* A) Three homoeologous chromosomes were represented with the QTL as well as confidence interval of QTL at the center in red and position of *FLC* gene, B) Links represented relation of syntenic genes around the QTL among homoeologous chromosomes; red link represent inversion, C) Distribution of markers from Csa (top) and Csp (bottom), and D) Distribution of *P-value* for identified QTL in Csa (top) and Csp (bottom) populations.



A.2 Chapter 5 Supplementary Tables and Figures

A.2.1 Subgenome combinations in *Camelina* species.



A.2.2 Morphology of hybrids develop from *C. sativa* × *C. microcarpa.*

(TMP23992)
Female Parent

(TMP23999)
Male Parent

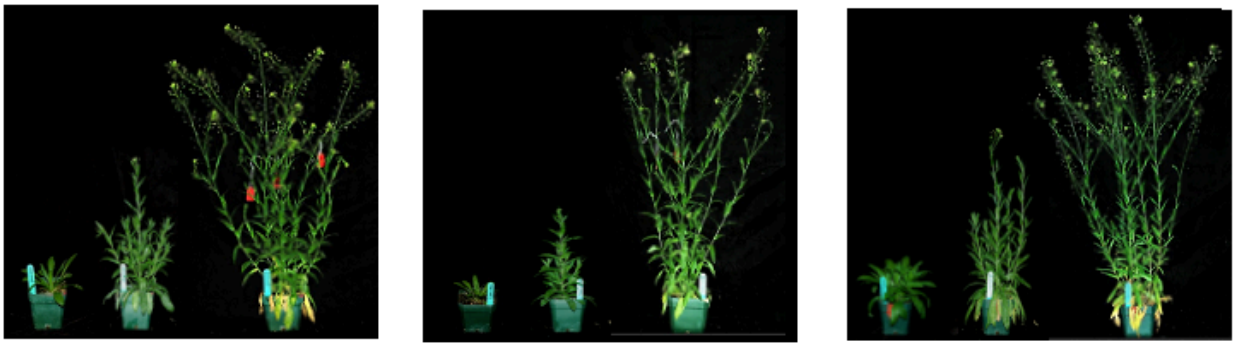
F_1

Morphology of hybrids develop from *C. microcarpa* × *C. sativa*

CN119102 (Female Parent)

X

TMP23992 (Male Parent)



F_1s at 11, 25 and 40 days after Vernalization

A.2.3 Best Parent Heterosis shown by F_1 of *C. microcarpa* × *C. sativa.*

S.N.	Genotype	Vernalization Period	Days to Flower after vernalization	BPH (%)
1	TMP23992xTMP23999	14	39	30
2	CN119102xTMP23992	30	25	-16.7
3	CN119102xTMP23992	30	21	-30
4	CN119102xTMP23992	30	23	-23.3
5	CN119102xTMP23992	30	22	-26.7
6	CN119102xTMP23992	30	22	-26.7
7	CN119102xTMP23992	30	21	-30
	TMP23999 (Parent)	31	30	
	CN119102 (Parent)	37	30	

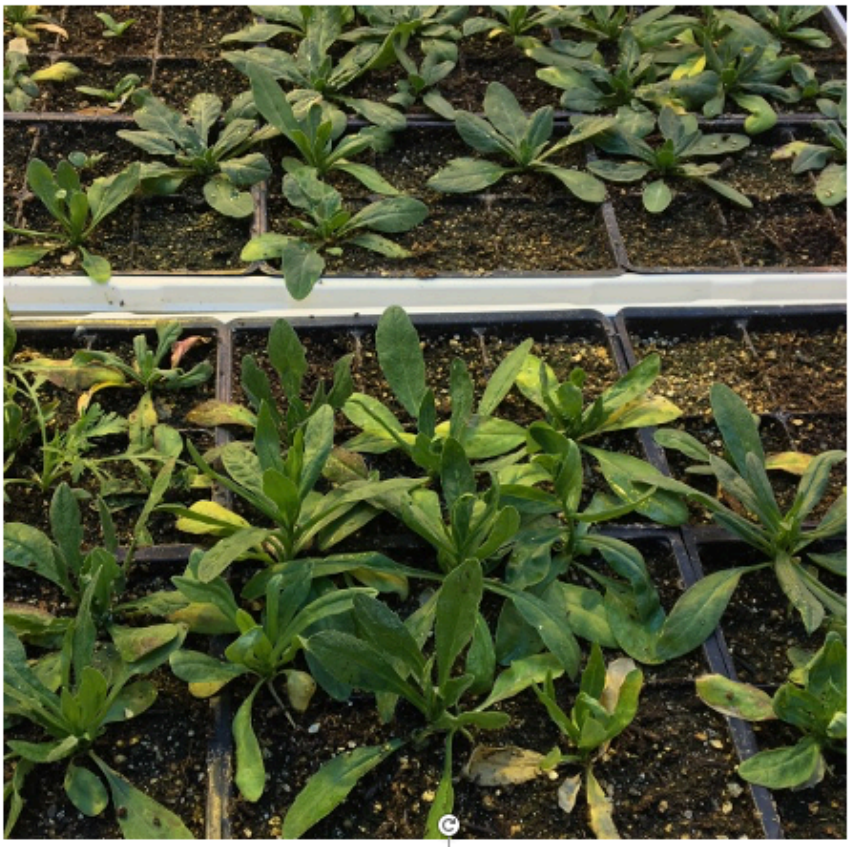
A.2.4 Morphological characteristics of F_2 plants developed from CN119102 × TMP23992 (*C. microcarpa* × *C. sativa*) (POP1).

(VER: Vernalization period; DTF: Days to first flower after seeding; LS, Leaf surface texture (W=Waxy, R= Rough (Upper/lower)); SL=Size of leaf in comparison to C. sativa parent (N= narrow, I= Intermediate, B= Broad); NF: number of flower in main raceme; and SW= weight of 200 seeds).

Plants	VER	DTF	LS	SL	NF	Notes	SW
82-6F2-P1	30	96	W/R	N/N	49	sterile	
82-6F2-P2	spring	98				sterile	
82-6F2-P3	30					sterile	
82-6F2-P4	30					sterile	
82-6F2-P5	30	95	R/R	N/N	34	sterile	
82-6F2-P6	spring	90	W/W	N/N	16	Fertile	0.2452
82-6F2-P7	30	87	W/R	N/N	31	Fertile	0.2363
82-6F2-P8	30	96	W/W	N/N	42	sterile	
82-6F2-P9	30	81	W/R	N/N	34	Fertile	0.2447
82-6F2-P11	30	84	W/W	N/N	24	semi winter, powdery mildew	
82-6F2-P12	30	84	W/R	N/N	39	semi winter, Powdery Mildew	
82-6F2-P14	spring	54			24	sterile	
82-6F2-P16	30	92	R/R	N/I	54	Fertile	0.1769
82-6F2-P17	30					sterile	
82-6F2-P18	30	89	R/R	N/N	57	No seeds	
83-2F2-P1	30	96	R/R	N/N	28	sterile	
83-2F2-P2	30	97	R/R	N/I	34	Fertile	0.1432
83-2F2-P3	30						
83-2F2-P4	30						
83-2F2-P5	30	89	R/R	N/I		No seeds	

83-2F2-P6	30						
83-2F2-P7	30	95	W/R	N/I	42		
83-8F2-P1	30	103	R/R	N/I	65	No seeds	
83-8F2-P2	30						
83-8F2-P4	30	95	R/R	N/N	57		
83-8F2-P10	30						
83-8F2-P12	30	101	R/R	N/I	68		
83-8F2-P13	30	93	R/R	N/N	62		
83-8F2-P15	30	107	R/R	N/N	32	No seeds	
83-8F2-P16	30	93	R/R	I/I	119	2 seeds	
83-8F2-P17	30	95	R/R	I/I	38	No seeds	
83-8F2-P18	30	107	R/R	I/B	134	sterile	
83-8F2-P19	30						
83-8F2-P20	30	110	R/R	N/N	43	No seeds	
83-8F2-P22	30	89	R/R	N/I	78		
83-8F2-P23	30	98	R/R	N/N	96	70 seeds	0.1256
83-8F2-P25	30	98	R/R	N/N		stunted growth	
83-8F2-P26	30	101	R/R	N/N		stunted growth, no seed	
83-8F2-P27	30	101	R/R	I/I	92		
83-8F2-P28	30	98	R/R	I/N	124	LEAF ROLLING BEHAVIOUR	
83-8F2-P29	spring type	40			96		
83-8F2-P30	30	96	R/R	N/N	71	sterile	
83-8F2-P31	30	95	R/R	N/N	114	4 seeds	
83-8F2-P32	30	85	R/R	N/I	108	sterile	
83-8F2-P33	30	98	R/R	N/N	15	sterile	
83-8F2-P34	30	109	R/R	B/B	110	Fertile	0.125
83-8F2-P35	30	101	R/R	I/I	99	Fertile	0.1715
83-8F2-P36	30	92	R/R	N/N	27	dwarf plant	
83-9F2-P3	30	89	W/R	N/N	59	sterile	

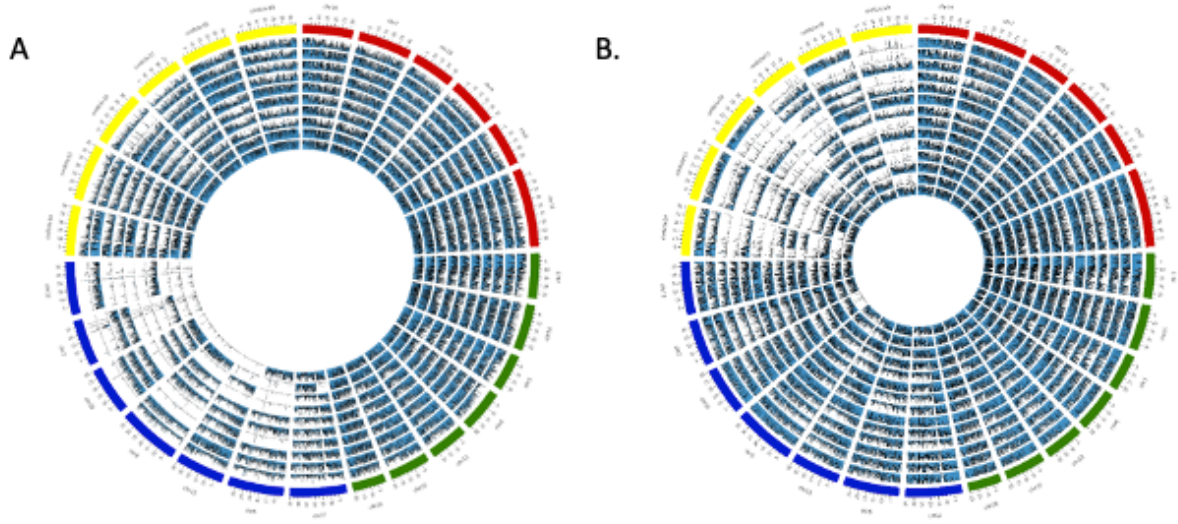
A.2.5 Differences in a leaf orientation of F_2 Plants developed from same parents.



F_2 from hybrid 83-8

F_2 from hybrid 82-6

A.2.6 Mapping of F_2 reads to the pseudogenome suggests biasness of reads for particular subgenome. A) F_2 Plants-83-8 and B) F_2 Plants-82-6.



A.2.7 Morphological characteristics of BC_1F_2 plants developed from (TMP23992 × TMP23999) × PI650132 (POP2). Here, VER: Vernalization period; DTF: Days to first flower after seeding; LS, Leaf surface texture (W=Waxy, R= Rough (Upper/lower)); SL=Size of leaf in comparison to *C. sativa* parent (N= narrow, I= Intermediate, B= Broad); NF: number of flowers in main raceme; F= Fertility; and SW= weight of 200 seeds.

Plants	VER	DTF	LS	SL	NF	Fertility	SW	Chromosome numbers
95-1F2-P1	30	100	R/W	I/I		Fertile		2n
95-1F2-P2		44	R/W	I/B	78	Fertile	0.1873	2n+1
95-1F2-P3		37	R/W	I/B	30	Fertile	0.2014	2n
95-1F2-P4	30	97	R/W	I/B		Fertile		2n
95-1F2-P5		49	R/W	I/B	52	Fertile	0.143	2n+1
95-1F2-P7		44	R/W	I/I	63	Sterile		2n+2
95-1F2-P9		40	R/W	I/I	73	Fertile	0.1919	2n
95-1F2-P10	30	97	R/W	I/I		Sterile		(2n-1)+2
95-1F2-P11		38	R/W	I/I	75	Fertile	0.21	2n
95-1F2-P12		50	R/W	I/B	107	Sterile		2n
95-1F2-P13		50	R/W	I/B	56	Sterile		2n
95-1F2-P14	30	99	R/W	I/I		Sterile		2n+1
95-1F2-P15		47	R/W	I/I	85	Fertile	0.2004	(2n-1)+1

A.2.8 Protein profile of BC_1F_2 developed from (TMP23992 × TMP23999) × PI650132 (POP2).

Mol. Wt. (kDa)	*C. microcarpa* Winter		*C. sativa* Spring		*C. alyssum* Spring		Backcross lines ((TMP23992 × TMP23999) × PI650132)															
	TMP23999		TMP23992		PI650132		95F2-2		95F2-3		95F2-5		95F2-9		95F2-11		95F2-15		95F2-1		95F2-4	
	Sample P3		Sample P5		Sample P10		Sample P13		Sample P14		Sample P15		Sample P16		Sample P17		Sample P18		Sample 37		Sample 38	
	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%	kDa	%
9.51	9.5	2.6			9.5	7.2	9.4	2.6	9.6	6.3	9.6	4.2	9.7	5.4	9.4	2.2						
10.1			10.1	2.9			10.1	2.9			10.2	4.5			10.2	1.1	10.5	1.6	10.1	3.6	10.2	6.7
11.2															11.2	1.8						
14.2															14.2	0.4						
17.9	17.9	39.9	18.4	35.9	18.6	40.1	18.6	34.9	18.7	33.0	18.8	43.4	18.8	45.4	18.8	40.5	19.1	41.3	18.5	44.8	18.4	40.5
22.2	22.2	5.1	22.4	4.1			22.6	5.1	22.4	5.4	23.1	4.6	22.9	4.2	22.9	3.1	22.6	2.1	22.2	0.8	21.8	1.2
24.3	24.3	5.1	24.5	7.1	24.7	7.1	24.6	6.7	24.6	7.2	25.2	6.2	25.5	4.9	25.5	5.8	26.2	5.5	24.9	4.4	25.3	4.5
29.1	29.1	7.0	29.7	11.9	30.1	10.9	29.6	12.0	30.0	10.8	30.3	12.5	30.4	11.3	30.4	10.8	30.6	8.8	27.3	1.8	27.2	2.1
33.9	33.9	10.2	34.3	7.2	33.3	2.9	34.0	6.3	34.7	7.9	34.6	5.6	35.0	5.1	35.2	7.2	35.8	6.5	29.4	7.3	29.9	7.2
34.9					34.9	2.8													34.7	7.5	34.7	8.6
39.9	39.9	0.1	39.8	1.1	40.1	2.0	40.0	1.0	40.2	1.4	40.7	1.3	40.7	0.8	41.0	0.8	41.5	1.5	40.2	1.1	40.4	1.3
48.4											48.4	1.1							47.4	1.2	47.6	1.5
49.7	49.7	7.6	49.2	15.5	49.9	16.5	49.5	12.2	49.8	13.2	50.8	6.9	50.6	12.6	50.6	26.2	50.6	14.5	49.2	18.1	49.5	12.4
52.3	52.3	14.4	51.6	13.5	52.4	10.2	52.0	14.1	52.8	11.5	53.5	9.0	53.3	9.8			53.7	17.5	52.3	8.6	52.0	13.6
62.6																	62.6	0.3	61.3	0.4		
67.1	67.1	1.9	64.9	0.2			65.0	1.0	65.5	1.0			66.4	0.0			67.1	0.2				
77.1	77.1	2.2	77.8	0.3	78.6	0.4	80.3	0.4	77.8	0.9									72.5	0.2	73.0	0.2
81.3											81.3	0.2	82.0	0.1								
88.2	88.2	2.5	85.6	0.3			85.9	0.4	86.2	0.5												
89.3									89.3	0.6												

A.2.9 Segregating seeds from BC_1F_2 developed from *C. microcarpa* (POP2).

A.2.10 Distribution of missing SNPs across the chromosomes compared to population in the third subgenome of *C. sativa* and *C. microcarpa* for the parents and segregating plants generated from interspecific hybridization.

Plants	Third Subgenome of *C. sativa*							Third subgenome of *C. microcarpa*						Probable chromosomes numbers
	Chr 17	Chr5	Chr 15	Chr9	Chr 20	Chr2	Chr12	CmII-Chr14	CmII-Chr15	CmII-Chr16	CmII-Chr17	CmII-Chr18	CmII-Chr19	
TMP23992 (*C. sativa*)	30	21	35	29	25	26	25	99	99	95	98	98	99	20
CN119102 (*C. microcarpa*)	93	96	97	96	97	88	96	67	59	61	60	63	59	19
F2 populations														
F2-82-6-1	19	22	27	23	15	25	30	40	37	39	95	32	97	24
F2-82-6-11	38	33	44	32	39	26	28	66	96	96	49	45	98	23
F2-82-6-12	33	24	45	46	47	33	41	96	48	50	52	47	97	24
F2-82-6-14	40	38	29	47	41	30	30	55	48	95	49	92	98	23
F2-82-6-16	33	33	35	34	27	27	39	47	46	47	45	91	54	25
F2-82-6-17	34	29	50	50	45	41	47	97	59	54	48	54	98	24
F2-82-6-18	36	19	30	40	98	28	31	42	43	45	47	45	45	25
F2-82-6-2	35	27	48	41	34	30	34	56	95	96	54	93	99	22
F2-82-6-3	28	19	31	45	39	34	39	98	46	96	45	44	48	24
F2-82-6-4	49	32	45	43	39	45	32	99	94	96	76	55	98	21
F2-82-6-5	39	31	52	36	33	37	32	58	50	53	98	51	58	25
F2-82-6-6	20	14	24	24	32	23	20	98	94	94	97	38	41	22
F2-82-6-7	42	46	45	44	33	35	38	99	53	96	55	58	64	24
F2-82-6-8	40	34	29	33	37	41	39	62	46	96	70	92	96	22
F2-82-6-9	34	15	25	40	20	24	31	59	95	41	41	39	43	25
F2-83-2-1	38	22	25	27	39	26	34	48	45	92	98	40	54	24
F2-83-2-2	32	24	33	35	27	26	26	97	94	94	48	94	98	21
F2-83-2-3	45	23	30	33	41	37	30	91	47	51	96	48	52	24
F2-83-2-4	32	23	31	33	33	28	44	51	50	51	93	44	96	24
F2-83-2-5	30	24	35	34	42	30	28	52	94	50	50	52	57	25
F2-83-2-6	30	23	28	28	40	35	39	97	94	94	43	45	48	23
F2-83-2-7	25	16	22	36	99	32	30	94	96	42	40	40	42	23
F2-83-8-1	49	45	53	57	98	90	96	58	57	58	58	41	47	23

F2-83-8-10	45	95	53	89	42	50	97	54	50	37	37	33	56	23
F2-83-8-12	36	31	38	43	99	88	95	29	44	45	41	39	31	23
F2-83-8-13	51	33	51	50	46	88	95	50	51	48	50	44	40	24
F2-83-8-15	43	93	94	49	40	81	44	50	32	51	40	49	55	23
F2-83-8-16	55	97	53	95	51	54	96	56	40	55	47	55	61	23
F2-83-8-17	47	94	93	40	32	90	95	36	38	27	42	35	29	22
F2-83-8-P18	41	34	98	45	99	80	38	37	30	28	42	42	33	23
F2-83-8-19	40	31	96	93	99	88	95	38	41	27	26	24	27	22
F2-83-8-2	40	32	29	47	40	82	40	32	42	94	50	43	50	24
F2-83-8-20	85	95	95	41	45	56	46	40	32	34	52	36	45	23
F2-83-8-22	43	94	33	32	32	36	96	34	39	22	33	30	40	24
F2-83-8-23	26	91	97	87	98	82	24	26	26	30	26	12	16	21
F2-83-8-25	38	95	41	94	98	46	96	40	40	44	42	42	30	22
F2-83-8-26	45	37	54	51	99	41	65	50	45	40	54	49	38	25
F2-83-8-27	29	25	32	31	27	85	94	34	34	35	35	31	32	24
F2-83-8-28	86	35	43	94	100	81	38	29	54	43	32	30	31	22
F2-83-8-29	50	26	49	49	47	37	34	52	96	93	58	93	57	23
F2-83-8-30	26	20	31	26	98	81	95	30	30	30	27	20	26	23
F2-83-8-31	30	22	96	36	30	35	94	34	34	22	34	23	25	24
F2-83-8-32	84	67	33	87	30	33	96	23	37	21	27	35	40	23
F2-83-8-33	30	94	33	26	98	78	28	32	19	29	18	14	20	23
F2-83-8-P34	40	94	98	45	37	41	95	43	27	39	44	42	32	23
F2-83-8-35	65	98	98	67	59	89	57	69	66	49	64	62	54	23
F2-83-8-36	46	93	57	49	99	45	42	54	38	45	41	36	41	24
F2-83-8-4	34	26	42	91	33	40	95	38	41	44	44	24	48	24
F2-83-9-3	43	37	55	54	54	55	95	99	56	45	60	45	63	24
Backcross derived F_2 population														
TMP23999 (*C. microcarpa*)	92	97	98	95	99	89	97	22	15	21	27	17	24	
TMP23992 (*C. sativa*)	30	21	35	29	25	26	25	99	99	95	98	98	99	
95-F2-1	37	30	31	40	35	29	30	84	98	94	98	97	99	2n
95-F2-2	63	51	34	38	34	36	32	54	99	96	35	96	98	2n+1
95-F2-3	34	27	37	35	31	30	31	98	100	99	99	100	99	2n
95-F2-4	81	45	39	41	35	41	36	70	99	95	98	92	99	2n
95-F2-5	20	20	20	23	16	22	30	44	98	55	33	89	99	2n+1
95-F2-7	26	29	27	28	21	24	33	96	99	25	36	90	99	2n+2
95-F2-9	38	28	32	35	28	25	32	71	99	94	86	88	99	2n

95-F2-10	44	40	34	34	27	36	94	52	98	38	31	91	99	(2n-1)+2
95-F2-11	56	42	42	45	39	45	42	68	99	56	100	94	99	2n
95-F2-12	59	60	49	52	42	48	61	57	99	56	77	92	99	2n
95-F2-13	55	38	48	50	41	47	49	82	99	80	59	94	99	2n
95-F2-14	34	35	31	33	28	23	35	64	96	93	36	88	99	2n+1
95-F2-15	43	41	35	40	30	42	96	64	98	43	98	94	99	(2n-1)+1

A.3 Chapter 6 Supplementary Tables

A.3.1 Mapping percentage of reads across the reference genome DH55.

Genome	Library	Input Reads	Uniquely mapped reads %	Multi-mapped reads %	Unmapped reads %
SG1	S50_TMP24028	26160815	76.15	5.39	18.46
	S51_TMP24028	10340019	79.65	5.98	14.37
	S52_TMP24028	20486489	77.28	5.15	17.57
SG1-SG2	CN119243_L1	23389547	77.7	17.61	4.67
	CN119243_L2	25660738	78.44	17.35	4.19
	CN119243_L3	21716827	76.52	18.76	4.71
SG1-SG2-SG3	S10_TMP23986	19616009	81.11	15.35	3.53
	S11_TMP23986	16552642	80.64	15.59	3.76
	S12_TMP23986	20309322	81.08	15.52	3.4
	S20_TMP23992	15883235	80.01	15.48	4.51
	S21_TMP23992	17348342	79.12	15.27	5.6
	S22_TMP23992	18251535	78.51	17.13	4.36
	S30_DH55	12443945	81.81	14.5	3.69
	S31_DH55	15716292	80.43	14.5	5.07
	S32_DH55	10197178	80.82	14.5	4.68
	S40_CAM176	6750450	67.49	12.84	19.66
	S41_CAM176	14079090	80.07	15.72	4.21
	S42_CAM176	15796764	80.15	16.7	3.14
	S83_TMP24026	13466366	77.11	18.01	4.88
	S84_TMP24026	13559546	9.91	2.24	87.84
	S85_TMP24026	10361461	76.07	17.26	6.67

A.3.2 Mapping percentage of reads across the reference genome TMP23999.

Genome	Library	Input Reads	Uniquely mapped reads %	Multi-mapped reads %	Unmapped reads %
C. microcarpa "Type 2" (TMP23999)	S60_TMP23999	13653150	82.5	13.05	4.45
	S61_TMP23999	13687913	81.34	13.98	4.67
	S62_TMP23999	9308554	43.49	24.88	31.63
	S1_CN119103	19155394	82.69	13.18	4.13
	S2_CN119103	21255147	82.65	12.68	4.67
	S3_CN119103	14512574	81.48	14.78	3.74
	S90_CN115248	14753235	79.92	16.04	4.04
	S91_CN115248	14685238	81.84	14.74	3.42
	S92_CN115248	14142294	81.32	14.54	4.14

A.3.3 Synteny table between *C. neglecta*, *C. sativa* and *C. microcarpa* "Type 2".

https://1drv.ms/x/s!Ah1gWPis3PD6gpxCrux4a_4sML8t7g?e=cK2zTf

A.3.4 Enrichment analysis of genes showing higher expression in the first subgenome of *C. sativa* in comparison to *C. neglecta* and first subgenome of tetraploid *C. microcarpa.*

Gene Family	Short Name	Annotated query per family	Background annotated per family	P-value
C2H2 zinc finger gene family	C2H2 ZF	6	97	0.018
Protein kinase (PK) gene superfamily	PK	29	813	0.005
Basic helix-loop-helix (bHLH) gene family gene family	bHLH	7	153	0.047
IQD gene family	IQD	4	41	0.011
Papain-like cysteine proteases (PLCP) gene family	PLCP	4	39	0.010
NPH3 gene family	NPH3	3	33	0.033
Amino acid transporters (AAT) gene family	AAT	4	62	0.044
Lectin-like Receptor Kinase (LecRLK) gene family	LecRLK	4	62	0.044
Dirigent gene family	Dirigent	3	26	0.018
NB-LRR Gene Family	NB-LRR	9	145	0.004
Expansin gene family	Expansin	3	37	0.045
Cysteine-rich receptor-like kinases (CRK) gene family	CRK	4	40	0.010
Glycerophosphodiester phosphodiesterase (GDPD) gene family	GDPD	2	12	0.026
Glycosyltransferase 58 (GT58) gene family	GT58	1	1	0.021
CAF1 gene family	CAF1	2	16	0.045

A.3.5 Subgenome dominance analysis in hexaploid *Camelina* species.

Life form	**Winter type**					**Spring type**		
Subgenome	**(6+7+6)**					**(6+7+7)**		
Genotype	**TMP 23999**	**CN11 9103**	**CN115 248**	**TMP24 026**	**CAM 176**	**DH55**	**TMP2 3992**	**TMP 23986**
Balanced	5542	7597	4792	6222	4325	4209	2833	2783
Dominant- SG1	1068	751	1446	994	1561	1827	2270	2296
Dominant- SG2	1556	1117	2069	1173	1703	1943	2433	2432
Dominant- SG3	1182	840	1681	1421	2036	2449	3007	3103
Suppressed- SG1	533	459	726	666	750	760	867	896
Suppressed- SG2	496	419	686	664	696	751	809	842
Suppressed- SG3	435	359	633	515	588	617	673	676
	10812	11542	12033	11655	11659	12556	12892	13028

A.3.6 Number of genes across subgenomes in reference genome of different *Camelina* species. Genes assigned to unanchored scaffolds were not presented in this table.

Subgenome	***C. neglecta***	***C. microcarpa* "Type 2"**	***C. sativa***
SG1	29896	29945	28158
SG2		28694	26996
SG3			30120
Cm-SG3		30031	

A.4.1 A simplified pathway of genome evolution of *Camelina sativa*. Four diploid progenitors evolved from a single ancestral diploid, and the age of divergence between the subgenomes are presented in million years ago (mya). The colour indicates the dominance of respective diploid species in the related hexaploid. The dominance of SG2 in *C. microcarpa* (n=19) could suggest an alternate pathway for the formation of this species, with a novel tetraploid formed between *C. neglecta* and *C. macrocarpa*-SG3, not shown in this diagram.