Mapping quantitative trait loci for seed traits in Camelina sativa

Authors: Kevin King, Huang Li, Jinling Kang, and Chaofu Lu

The final publication is available at Springer via http://dx.doi.org/10.1007/s00122-019-03371-8.

King, Kevin, Huang Li, Jinling Kang, and Chaofu Lu. "Mapping quantitative trait loci for seed traits in Camelina sativa." Theoretical and Applied Genetics Volume 132, no. Issue 9(September 2019): 2567-2577. DOI:10.1007/s00122-019-03371-8.

Mapping quantitative trait loci for seed traits in *Camelina sativa*

Kevin King*, Huang Li*, Jinling Kang, and Chaofu Lu

Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT 59717-3150 USA

* These authors contributed equally to this work.

Correspondence: Chaofu Lu

Email: clu@montana.edu

ORCID ID: 0000-0001-6901-7490

Abstract

Camelina (*Camelina sativa* L. Crantz) is an oilseed crop that has great potential to provide sustainable feedstock for biofuel production and to improve dryland agriculture. A major breeding objective for camelina is to increase seed size and oil content. Understanding the genetics behind variation of seed size and associated traits such as oil content would help breeders develop varieties of increased oil yield that are more robust, easier to plant and harvest, better for oil processing. In this study, we developed a recombinant inbred population derived from the two *Camelina* accessions, Suneson and Pryzeth, with contrasting traits especially seed size and oil content. Using 189 lines, a genetic map was constructed containing 2376 single nucleotide polymorphism (SNP) markers spanning 2034.6 cM of 20 linkage groups with an average density of 1.5 cM per locus. Field trials were conducted for two years (2017 and 2018) in two environments (dryland and irrigated) in Bozeman, Montana. The results revealed important correlations of seed size with other associated traits such as oil content, pod size and seed number per pod. Significant QTL were also discovered for these traits. The results of this study are the first step to isolate genes controlling seed development and oil accumulation, and to develop advanced varieties of camelina better adapted to modern agriculture by marker-assisted breeding.

Keywords

Camelina sativa, genetic linkage map, quantitative trait loci, oil content, seed size

Key Message

Genetic dissection of oil content and seed size in *Camelina sativa* was conducted by QTL mapping using a SNP-based linkage map and a recombinant inbred population.

Introduction

Camelina (*Camelina sativa* L. Crantz) is a member of the *Brassicaceae* family. It is a potentially productive oilseed crop that is highly adaptable to a wide range of environmental conditions (Berti et al. 2016). The main interest in camelina oil is its use as biofuels such as a hydroprocessed renewable jet fuel (Moore et al. 2017). Camelina cultivation may improve sustainability of modern agriculture through crop rotation in the cereal-based cropping systems (Obour et al. 2018), and in double-cropping with soybean to enhance oil production (Gesch and Johnson 2015). However, several agronomic traits need to be improved to make the production of camelina economically viable. One of the most important breeding objectives is to increase seed size, an important yield component. The small seed size (~1.5 mm x 0.8 mm, or 1 mg/seed) of camelina (Fleenor 2011) may limit its yield potential. Producers could benefit with having larger seeded varieties especially during harvest. Though camelina can be harvested using a standard combine used for canola harvest, due to its small size, a large amount of seeds can be lost in the field during harvesting. In addition to yield loss, this can result in volunteers for the following crops (Hunter and Roth 2010). Most importantly, small seeds may pose a high risk of germination and seedling establishment, especially in dryland conditions when periods of no rainfall after planting dry out the soil surface (McVay and Khan 2011; Sintim et al. 2016). It has been shown that increasing seed size may enhance seedling survival in Arabidopsis (Krannitz et al. 1991) and accelerate seedling emergence to avoid imminent stress conditions in *Brassica oleracae* (Finch-Savage et al. 2010). Higher seedling rates and greater biomass associated with large seed sizes also increased crop competition with weeds and hasten flowering in canola (Harker et al. 2015).

Seed size is a complex trait that is predominantly controlled by genetic factors in maternal (e.g., seed coat) and zygotic (embryo and endosperm) tissues resulted from the event of double-fertilization (Belmonte et al. 2013). Several molecular signaling pathways have been identified that influence seed size in Arabidopsis and crops (Li and Li 2016; Orozco-Arroyo et al. 2015; Li et al. 2019). Our recent studies indicated that some of these orthologous genetic pathways might also be involved in controlling seed size in camelina (Na et al. 2019). In addition, a great variation in seed weight, which is related to seed size,

in the camelina germplasm (Vollmann et al. 2005) also suggested that multiple genes are involved in regulating this important trait, and quantitative trait loci (QTL) were identified controlling thousand seed mass (Gehringer et al. 2006). In this study, we created a recombinant inbred line (RIL) population from the cross between two camelina varieties of contrasting traits especially in seed size. The distribution of seed size in the RIL population clearly indicated that seed size is quantitatively inherited. We constructed a high-density genetic linkage map using single nucleotide polymorphism (SNP) markers, and localized several QTL that control seed size and its associated traits including oil content, pod size and seed number per pod of camelina.

Materials and Methods

Plant materials and field experiments

A population consisting of 189 recombinant inbred lines (RILs) was developed by single seed descent (SSD) from a cross between a big-seeded accession Pryzeth and an elite cultivar Suneson. SSD was conducted as described previously by growing plants in the greenhouse to the F_5 generation (Gehringer et al. 2006). The RIL population was evaluated in four field-trial environments: The $F_{5\text{-}6}$ RILs and two parental lines were arranged in an augmented randomized complete block design and planted at the Arthur H. Post Research Farm (Bozeman, MT, USA) in 2017 and 2018. Field trials were conducted under rain-fed, dryland (D17, D18) and irrigated conditions (I17, I18) following standard agricultural practices. In the irrigation treatment, approximately 12cm of supplemental water was applied during the period of seed filling by overhead sprinklers. Five plants in the middle of each plot were randomly selected and harvested for further analysis.

Phenotypic measurement

Twenty mature pods were harvested from the main stems by hand. Pods and seeds were scanned by an Epson Perfection V600 Photo Scanner at 600 dpi (0.024 mm/pixel) to calculate their sizes using SmartGrain (Tanabata et al. 2012), a high-throughput seed measurement software. A set of coordinates representing seed morphological perimeters (seed area, seed length and seed width) were obtained as raw phenotypic data. Seeds from

these 20 pods were counted using a Seedburo 801 Count-A-Pak (Seedburo Equipment, Des Plaines, IL, USA) seed counter and weighted with a balance (AE100, Mettler). Morphological pod traits (pod area, pod length and pod width) were measured by the same method as above. Approximately 20 mg dry seeds from each line were used for oil content measurement. Seed oil content was determined by a bench-top NMR seed analyzer (MQC23, Oxford Instruments, Concord, MA, USA) and presented as percentages of seed mass. The NMR was calibrated using seed samples of known oil contents and results from gas chromatography (GC) analyses, similar to the procedure described previously (Zhu et al. 2018). Best linear unbiased prediction (BLUP) values of each trait of each line were obtained and used as phenotypic data for mapping analysis. The broad-sense heritability of each trait was calculated according to (Holland et al. 2003). All descriptive statistics including Pearson's correlation coefficient of all traits were calculated in Rstudio (v1.1.383).

Genotyping by sequencing (GBS) and SNP calling

Total genomic DNA was isolated from young leaves of all RILs and two parents using the FastPrep DNA extraction and purification kit (MP biomedicals). GBS was performed at the LGC Genomics (http://www.lgcgroup.com). Specifically, a commercially licensed normalized GBS (nGBS), using MslI as the restriction enzyme to efficiently exclude conserved repetitive elements in the genome, was performed on Illumina NextSeq 500 sequencer. Sequencing reads were pre-processed and aligned by using bcl2fastq and Bowtie2 version 2.2.3 software (Langmead and Salzberg 2012). Variant detection and progeny genotyping were performed with Freebayes v1.0.2-16 pineline (Garrison and Marth 2012). Raw SNP calls with minor allele frequency (MAF) >0.05 in the biparental population were concatenated into a single Variant Call File (VCF) and proceeded for linkage map construction.

Linkage map construction

First, SNPs with more than 15% missing rate and those that were monomorphic between two parental lines were eliminated. The remaining SNPs with significant segregation distortion ($p<0.05$) were removed through chi-square test combined with sequential

Bonferroni correction. A genetic linkage map was constructed using MapDisto 2.0 (Lorieux 2012). In MapDisto, all markers were grouped with a recombination fraction of 0.3 and logarithm of the odds ratio (LOD) score of 7 using the Kosambi mapping function (Kosambi 1944). The Seriation algorithm was used for ordering the linkage groups. The "Automap" command which successively runs multiple steps including "Ripple order" and "Check inversions" was used to refine the marker order of each linkage group. Linkage groups were assigned to chromosomes by blasting the tag sequences of their allocated GBS makers against the *C. sativa* genome database (http://www.camelinadb.ca/). An iterative strategy was employed to obtain the final linkage map which agrees with the number of camelina chromosomes. By blasting the sequences of markers located on the ends of each linkage group, the physical coverage was calculated and average interval length was estimated. The graphical representation of the genetic map was generated using MapChart version 2.2 (Voorrips 2002).

QTL analysis and candidate gene selection

Phenotypic data of each trait from individual environments, along with the mean values of all four trials were integrated with the genotypic data and used for QTL mapping. All mapping analysis was performed using QGene 4.4.0v (Joehanes and Nelson 2008) with the composite interval mapping (CIM) function. LOD significance thresholds of linkage groups were determined independently for each trait using the permutation test (n=1000, a=0.05). Cofactors of each group were selected by using the default parameters. The walking speed was set to 1cM to scan intervals and the coefficient of determination (R^2) was calculated to estimate the amount of phenotypic variation explained (PVE) by each QTL.

Flanking markers of major QTL detected in multiple environments were used to delineate the physical intervals. Their tag sequences in GBS dataset were blasted against the *C. sativa* genome V2.0 (http://www.camelinadb.ca/) using BLASTn. All annotated genes residing within the QTL intervals were subject to further screening for candidates implicated in seed/pod development (Li et al. 2019) or oil accumulation (Li-Beisson et al. 2013).

Results

Seed traits evaluations

Two parental lines, Suneson and Pryzeth, have contrasting sizes in seed and pod. Field−harvested seeds from Pryzeth were, on average, 66% larger and 71% heavier than seeds from Suneson. Similarly, pods of Pryzeth were also 80% larger than those of Suneson (Fig. 1, Table S1). Oil contents also differ in Pryzeth (33%) and Suneson (37%). To more accurately evaluate seed and pod sizes, their length and width were measured along with area by scanning and imaging analyses. In addition, 1000-seed weight (TSW), seed number per pod (SPP) and total oil content (TOC) were also obtained. Frequency distributions of all these measurements did not show significant deviation from normal distributions in the RIL population (Fig. 2), and all traits exhibited high broad-sense heritability (Table S1). These results suggest that oil content and all seed and pod traits in this study are quantitatively controlled and their genetic effects are the major determinant of the phenotypic variations. Interestingly, among above traits, only SPP showed a pattern of transgressive segregation in our population.

Similar to results observed in Arabidopsis (Gnan et al. 2014), pair-wise comparison indicated that seed size determined by scanned seed area (SA) was highly correlated with TSW (r = 0.89). We then analyzed seed size for its separate parameters, seed length (SL) and seed width (SW). SA was shown to be highly positively correlated with both SL and SW (r = 0.92 and 0.88, respectively), and the latter two were also clearly correlated (r = 0.65) (Table S2). Pod area (PA) was similarly measured, which showed weak correlation with SA (r = 0.39), indicating different genetic mechanisms for these two quantitative traits. The width-length correlation was also less significant in pods (r = 0.31). SPP was positively correlated with PA (r = 0.39) but had a negative correlation with SA (r = -0.30). The correlations between total oil content (TOC) and all these morphometric traits were slightly negative, with the highest observed between TOC and SL (r = -0.32).

Genotyping and linkage map construction

We obtained approximately 604 million sequencing reads for the RIL population; each line had 608,286 to 8,148,564, an average of nearly 3.2 million number of reads. A total

of 8,087 SNP markers with >5% minimum allele frequency (MAF) were identified through the Freebayes pipeline (Garrison and Marth 2012). Among them, 3702 homozygous SNPs are polymorphic between two parents and were used for constructing the linkage map. We assigned a total of 2376 SNPs with <15% missing data points and normal segregation onto 20 linkage groups, which correspond to 20 chromosomes in camelina (Kagale et al. 2014). After placing co-segregating markers into bins, the linkage map contains 1345 informative loci (Fig. 3). The final map spans 2034.6 cM with an average density of 1.5 cM per locus. The total genetic distance corresponds to a physical length of 501.48 Mb which covers 78.2% of the total assembled camelina genome size (641.45 Mb). The average interval length between two adjacent unique loci is estimated to be 0.37 Mb. SNP distribution and characteristics of our linkage map are presented in detail in Table 1.

QTL for seed size and 1000-seed weight

We detected a total of 40 QTL across the four environments for the following traits: seed area (SA), seed length (SL), seed width (SW), pod area (PA), pod length (PL), pod width (PW), 1000-seed weight (TSW), seed number per pod (SPP) and total oil content (TOC). The QTL were named SA1.1, SL1.1, SW4.1 and such that the letters indicate the trait associated, the following two digits represent the chromosome on which the QTL was identified and the occurrence number. For the morphometric traits (SA, SL, SW, PA, PL, PW), a total of 28 QTL consistently identified in at least two environments were described in Table 2, along with six QTL for TSW, two for SPP and four for TOC.

For seed size, seven QTL were identified for SA and each of them explained a moderate percentage of the phenotypic variation (PVE), varying from 8.3% to 13.1%. Four of these QTL for SA were shared with other traits including SL, SW, PL and TSW. Among them, a seed area QTL on chromosome 17 (SA17.1) with significant LOD and PVE was identified in nearly all environments and overlapped with QTL for three other morphometric parameters (SL, SW, PL) and TSW. The peaks of this QTL detected in all field trials were located around 40.0 cM on chromosome 17, flanked by two adjacent markers MC00670835 and MC00824085. This QTL region spans 6.0 cM and corresponds to a physical distance of 1.7 Mb on camelina chromosome 17 (Fig. 4).

Besides SA17.1, we detected two more QTL for SA (SA4.1 and SA8.1), which also overlapped with TSW QTL. While multiple QTL were colocalizd between SA and TSW, overall seed area (SA) shared fewer QTL with its related morphometric components (SW and SL). Except the major QTL on chromosome 17 mentioned above, only one QTL on chromosome 3 was shared between SA and SL while none was shared between SA and SW. Similarly, SL17.1 and SW17.1 are the only QTL shared between SL and SW.

Genetically, overlapping QTL for multiple traits could be explained by either tight linkage or pleiotropy. To disentangle the genetic causation of the QTL co-localized on chromosome 17 for SA and TSW (Fig. 4), we performed conditional QTL analysis on seed size and weight in a reciprocal manner. Specifically, we obtained conditional phenotypic values of SA and TSW using QGAStation1.0 (http://ibi.zju.edu.cn/software/qga/index.htm) and compared the results with their corresponding unconditional analysis. Both QTL peaks detected by the unconditional method disappeared in our conditional analysis, indicating that QTL SA17.1 and TSW17.1 are not independent of each other. In other words, pleiotropy rather than tight linkage resulted in the co-localization of SA17.1 and TSW17.1 (Fig. S1).

QTL for pod size, seed number per pod and oil content

Compared to seed morphometric traits, fewer QTL were observed in pod traits. A colocalized QTL on chromosome 5 was detected among PA, PL and PW. It explained up to 13.9% PVE for PA. Interestingly, this shared QTL was also observed for SL (SL5.1) and TSW (TSW5.1), though with less significant LOD and PVE values. The additive effects of all QTL for seed and pod morphological traits were positive, suggesting that the alleles for increasing these traits were contributed by the female parent Pryzeth.

Two QTL were identified on chromosome 2 and chromosome 9 for SPP with smaller PVE value (7.2%) compared to the QTL for other traits. One of them (SPP9.1) colocalized with PW (PW9.1). The two QTL (SPP2.1 and SPP9.1) had opposite additive effects, suggesting that alleles from both parents control seed numbers within a pod. This result was in line with our observation that SPP showed a pattern of transgressive segregation in the RIL population (Fig. 2).

For total oil content, we identified four QTL (TOC2.1, TOC9.1, TOC16.1, and TOC20.1) with PVE values ranging from 9.1% to 16.7%. Among them, TOC2.1, TOC 9.1 and TOC20.1 were detected in multiple environments with the highest PVE values in this study. Only one (TOC20.1) was colocalized with seed size traits (SW20.1). This result as well as a slightly negative correlation (Table S2) suggest that there was not a tight correlation between oil content and seed size.

Prediction of candidate genes

To indirectly validate our QTL mapping results, we attempted to identify candidate genes that may be responsible for the observed trait variations. The sequences of the QTL regions were used to anchor the QTL to the camelina physical map (Kagale et al. 2014). The sequence of the major QTL for seed size (Fig. 4) was found to be syntenic to a segment on Arabidopsis chromosome 1. This region contains approximately 180 putative genes, however many of these genes, of which 21 candidates are listed in Table S3, have been annotated with putative functions that may be involved in seed development, such as auxin or gibberellin mediated signaling pathways, sugar metabolism and transport, and metabolic activities related to embryo and seed coat development. Likewise, we tried to identify genes for oil content. In the regions of QTL, we found several lipid metabolism genes, including an acyl-CoA-binding domain-containing protein, a glycerol-3-phosphate acyltransferase, and a WSD1-like acyltransferase. These are known to be involved in surface lipid biosynthesis (Li-Beisson et al. 2013). However, candidate genes encoding a sucrose transporter (SUC5) (Baud et al. 2005) and a GDSL1 lipase residing in the TOC9.1 may contribute to oil content. Interestingly, repression of *BnGDSL1* in canola leads to a 5% increase in seed oil content, while overexpression results in 9% lower oil (Ding et al. 2019). Future fine mapping and characterization of these candidate genes are needed to determine their roles in controlling seed size and oil accumulation in camelina.

Discussion

To develop camelina as a viable bioenergy crop, genetic resources in camelina have been made available in recent years, including genome (Kagale et al. 2014) and transcriptome (Kagale et al. 2016; Abdullah et al. 2016; Nguyen et al. 2013) sequences. An efficient

genetic transformation method (Lu and Kang 2008) has also been developed. However, the lack of knowledge of genes controlling important traits in camelina limits the ability to efficiently improve camelina by genetic and biotechnology approaches. Like most crops, many important traits in camelina are quantitatively inherited, therefore identification of molecular markers and genes associated with these traits are critical for efficient breeding efforts. Compared to extensive gene mapping studies in other oilseed crops such as rapeseed (Chao et al. 2017), few such studies have been conducted in camelina. Genetic maps were constructed with amplified fragment length polymorphism (AFLP) or SNP markers using a recombinant inbred (RI) population derived from a cross between two varieties Lindo and Licalla (Singh et al. 2015; Gehringer et al. 2006). QTL for a limited number of traits including oil content, 1000-seed-weight and plant height were identified in the same study (Gehringer et al. 2006). These earlier studies demonstrated the power of quantitative genetics combined with molecular markers in dissecting complex traits in camelina. Our study using a different RIL population derived from two distinct parental lines allowed us to specifically analyze traits related to seed size and oil content. Compared to previous studies using two cultivated German varieties (Lindo and Licalla), we used an accession (Pryzeth) selected from a natural germplasm collection that has significantly larger seed size and lower oil content than the cultivated variety Suneson (Fig.1, Table S1), and differences in other traits. Such diverse parental lines created vast genetic diversities in the progeny population and allowed us to generate a large number of polymorphic SNP markers for linkage mapping. The greater phenotypic diversities in the RIL population also provided increased power of QTL localization through repeated field trials and trait evaluations.

Seed area is usually estimated by digital collection of seed area or by weight (Herridge et al. 2011; Gnan et al. 2014). Besides these, we also measured seed length and seed width by a document scanner. Distribution frequencies of all these traits did not show significance from normal distributions, indicating their quantitative inheritance. Phenotypically, SA had high correlations with SL and SW, and all these three morphometric traits (SA, SL and SW) are highly correlated with TSW. However, the numbers of QTL for these traits differed, and few QTL were co-localized although SA has more overlapping QTL with SL than SW. These results suggest that evaluation of seed size

by single measurements may be oversimplified. As SA, SL and SW may be captured by a single scan, all these measurements along with TSW should be used to map QTL. Further fine mapping of QTL and characterization of candidate genes, such as those listed in Table S3, will allow for their contributions to the final determination of seed size and seed yield.

Pods not only protect immature seeds, they are also photosynthetically active and play important roles in seed development and storage filling (Bennett et al. 2011). Multiple QTL were detected for pod size. Our observations indicate that sizes of pod and seed are correlated, and we found that one QTL may have pleiotropic effects on seed and pod sizes. The sizes of pod and seed affect the number of seeds per pod. As expected, larger pods contain more seeds per pod, but SPP decreases with increased seed size. These relationships suggest that increasing pod size may be beneficial for seed enlargement, but may limit the number of seeds per pod. Seed number and seed size/weight are all components of seed yield, therefore it is important to find the optimum seed size and seed numbers. Unfortunately, due to severe shattering by hail storms, we were not able to investigate the total numbers of pods and seed per plant, their correlations and effects on seed yield.

Since seed size/weight, seed number and oil content are all important traits to determine seed and oil yield, understanding their genetic basis is necessary to camelina improvement. Potential trade-offs between these traits may be observed in population studies. The negative correlation between seed size (measured as 1000-seed weight) and oil content was observed in camelina (Gehringer et al. 2006; Vollmann et al. 2007). Our results also showed a negative correlation between TOC and TSW, however was less significant than previous studies (r=-0.18 compared to -0.37 and -0.92). There was no overlap of QTL for seed size traits and TOC. This discrepancy suggested that the correlation between oil content and seed size might be dependent on genotypes and environments. The production of seed and its contents depends on the allocation of finite resources and is not discrete from the whole plant carbon budget (Venable 1992). As exampled in our recent study, seed size was increased by suppression of starch biosynthesis in camelina, however oil content did not change, and the increased seed weight was contributed by increased protein content (Na et al. 2018). Therefore, a

systematic approach is possible to simultaneously increase seed size and oil content by orchestrating carbon allocation, transport and sink activities.

Author contributions

JK and CL developed the RIL population; KK, HL collected and analyzed phenotypic data, constructed the linkage map and conducted QTL analysis. CL conceived the project and guided the analyses. CL and HL wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgement

We thank Drs. John Martin and Jamie Sherman for their help on statistical analyses. This work was supported by the Office of Science (BER), US Department of Energy, grant no. DE-SC0016400, and the USDA National Institute of Food and Agriculture Hatch project no. 1015638.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

Abdullah HM, Akbari P, Paulose B, Schnell D, Qi W, Park Y, Pareek A, Dhankher OP (2016) Transcriptome profiling of *Camelina sativa* to identify genes involved in triacylglycerol biosynthesis and accumulation in the developing seeds. Biotechnol Biofuels 9: 136. doi:10.1186/s13068-016-0555-5

Baud S, Wuillème S, Lemoine R, Kronenberger J, Caboche M, Lepiniec L, Rochat C (2005) The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in Arabidopsis. Plant J 43: 824-836. doi:10.1111/j.1365-313X.2005.02496.x

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci USA 110: E435-E444

Bennett EJ, Roberts JA, Wagstaff C (2011) The role of the pod in seed development: strategies for manipulating yield. New Phytol 190: 838-853
Berti M, Gesch R, Eynck C, Anderson J, Cermak S (2016) Camelina uses, genetics, genomics, production, and management. Ind Crop Prod 94: 690-710. doi:https://doi.org/10.1016/j.indcrop.2016.09.034
Chao H, Wang H, Wang X, Guo L, Gu J, Zhao W, Li B, Chen D, Raboanatahiry N, Li M (2017) Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. Sci Rep 7:46295. doi:10.1038/srep46295
Ding LN, Guo XJ, Li M, Fu ZL, Yan SZ, Zhu KM, Wang Z, Tan XL (2019) Improving seed germination and oil contents by regulating the GDSL transcriptional level in *Brassica napus*. Plant Cell Rep 38: 243-253. doi:10.1007/s00299-018-2365-7
Finch-Savage WE, Clay HA, Lynn JR, Morris K (2010) Towards a genetic understanding of seed vigour in small-seeded crops using natural variation in *Brassica oleracea*. Plant Sci 179: 582-589. doi:https://doi.org/10.1016/j.plantsci.2010.06.005
Fleenor RA (2011) Plant Guide for Camelina (Camelina sativa). USDA-Natural Resources Conservation Service, Spokane, WA 99201,
Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv:12073907:
Gehringer A, Friedt W, Luhs W, Snowdon RJ (2006) Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp. *sativa*). Genome 49: 1555-1563. doi:10.1139/g06-117
Gesch RW, Johnson JMF (2015) Water use in camelina–soybean dual cropping systems. Agron J 107: 1098. doi:10.2134/agronj14.0626
Gnan S, Priest A, Kover PX (2014) The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. Genetics 198: 1751-1758. doi:10.1534/genetics.114.170746
Harker K, O'Donovan J, Smith E, Johnson E, Peng G, Willenborg C, Gulden R, Mohr R, Gill K, Grenkow L (2015) Seed size and seeding rate effects on canola emergence, development, yield and seed weight. Can J Plant Sci 95: 1-8
Herridge R, Day R, Baldwin S, Macknight R (2011) Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. Plant Methods 7: 3
Holland JB, Nyquist WE, Cervantes-Martínez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22: 9-112
Hunter J, Roth G (2010) Camelina production and potential in Pennsylvania. Agronomy Facts 72. College of Agricultural Sciences, Pennsylvania State University
Joehanes R, Nelson JC (2008) QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24: 2788-2789
Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, Higgins EE, Huebert T, Sharpe AG, Parkin IA (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat Commun 5: 3706. doi:10.1038/ncomms4706
Kagale S, Nixon J, Khedikar Y, Pasha A, Provart NJ, Clarke WE, Bollina V, Robinson SJ, Coutu C, Hegedus DD, Sharpe AG, Parkin IA (2016) The developmental

transcriptome atlas of the biofuel crop *Camelina sativa*. Plant J 88: 879-894. doi:10.1111/tpj.13302
Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12: 172-175
Krannitz PG, Aarssen LW, Dow JM (1991) The effect of genetically based differences in seed size on seedling survival in *Arabidopsis thaliana* (Brassicaceae). Amer J Bot 78: 446-450
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359. doi:10.1038/nmeth.1923
Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, Franke RB, Graham IA, Katayama K, Kelly AA, Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L, Schmid KM, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J (2013) Acyl-lipid metabolism. Arabidopsis Book 11: e0161. doi: 0110.1199/tab.0161. doi:10.1199/tab.0161
Li N, Li Y (2016) Signaling pathways of seed size control in plants. Curr Opin Plant Biol 33: 23-32. doi:https://doi.org/10.1016/j.pbi.2016.05.008
Li N, Xu R, Li Y (2019) Molecular networks of seed size control in plants. Annu Rev Plant Biol 70: 11.11-11.30. doi:10.1146/annurev-arplant-050718-095851
Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. Mol Breeding 30: 1231-1235
Lu C, Kang J (2008) Generation of transgenic plants of a potential oilseed crop *Camelina sativa* by Agrobacterium-mediated transformation. Plant Cell Rep 27: 273-278. doi:10.1007/s00299-007-0454-0
McVay KA, Khan QA (2011) Camelina yield response to different plant populations under dryland conditions. Agron J 103: 1265-1269. doi:10.2134/agronj2011.0057
Moore RH, Thornhill KL, Weinzierl B, Sauer D, D'Ascoli E, Kim J, Lichtenstern M, Scheibe M, Beaton B, Beyersdorf AJ, Barrick J, Bulzan D, Corr CA, Crosbie E, Jurkat T, Martin R, Riddick D, Shook M, Slover G, Voigt C, White R, Winstead E, Yasky R, Ziemba LD, Brown A, Schlager H, Anderson BE (2017) Biofuel blending reduces particle emissions from aircraft engines at cruise conditions. Nature 543: 411-415. doi:10.1038/nature21420
Na G, Aryal N, Fatihi A, Kang J, Lu C (2018) Seed-specific suppression of ADP-glucose pyrophosphorylase in *Camelina sativa* increases seed size and weight. Biotechnol Biofuels 11: 330. doi:10.1186/s13068-018-1334-2
Na G, Mu X, Grabowski P, Schmutz J, Lu C (2019) Enhancing microRNA167A expression in seed decreases the a-linolenic acid content and increases seed size in *Camelina sativa*. Plant J 98: 346-358. doi:10.1111/tpj.14223
Nguyen HT, Silva JE, Podicheti R, Macrander J, Yang W, Nazarenus TJ, Nam J-W, Jaworski JG, Lu C, Scheffler BE, Mockaitis K, Cahoon EB (2013) Camelina seed transcriptome: a tool for meal and oil improvement and translational research. Plant Biotechnol J 11: 759-769. doi:10.1111/pbi.12068
Obour AK, Chen C, Sintim HY, McVay K, Lamb P, Obeng E, Mohammed YA, Khan Q, Afshar RK, Zheljazkov VD (2018) *Camelina sativa* as a fallow replacement crop in wheat-based crop production systems in the US Great Plains. Ind Crop Prod 111: 22-29. doi:https://doi.org/10.1016/j.indcrop.2017.10.001

Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L (2015) Networks controlling seed size in Arabidopsis. Plant Reprod 28: 17-32. doi:10.1007/s00497-015-0255-5

Singh R, Bollina V, Higgins EE, Clarke WE, Eynck C, Sidebottom C, Gugel R, Snowdon R, Parkin IA (2015) Single-nucleotide polymorphism identification and genotyping in *Camelina sativa*. Mol Breed 35: 35. doi:10.1007/s11032-015-0224-6

Sintim HY, Zheljazkov VD, Obour AK, Garcia AGY, Foulke TK (2016) Evaluating agronomic responses of Camelina to seeding date under rain-fed conditions. Agron J 108: 349-357. doi:10.2134/agronj2015.0153

Tanabata T, Shibaya T, Hori K, Ebana K, Yano M (2012) SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol 160: 1871-1880. doi:10.1104/pp.112.205120

Venable DL (1992) Size-number trade-offs and the variation of seed size with plant resource status. Amer Nat 140: 287-304

Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T (2005) Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. Plant Breed 124: 446-453. doi:DOI 10.1111/j.1439-0523.2005.01134.x

Vollmann J, Moritz T, Kargl C, Baumgartner S, Wagentristl H (2007) Agronomic evaluation of camelina genotypes selected for seed quality characteristics. Ind Crop Prod 26: 270-277. doi:10.1016/j.indcrop.2007.03.017

Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93: 77-78

Zhu Y, Xie L, Chen GQ, Lee MY, Loque D, Scheller HV (2018) A transgene design for enhancing oil content in Arabidopsis and Camelina seeds. Biotechnol Biofuels 11: 46. doi:10.1186/s13068-018-1049-4

Figure legends

Fig. 1. Comparison of sizes of dry seeds (above) and pods (below) between camelina varieties ‘Suneson’ and ‘Pryzeth’.

Fig. 2. Phenotypic variations of seed traits in the RIL population. Histograms represent the trait frequency data of the 189 RILs over both years and treatments (n=760). The red line represents the mean of that trait in the population. The two box-plots under the histograms represent the combined data for the two parents over both years and treatments (n=30). Bold lines in the boxes represent mean values and dotted lines show data distribution range of each trait in two parents. Suneson= Red, Pryzeth= Blue.

Fig. 3. A genetic linkage map of camelina constructed using the Pryzeth x Suneson population. Each bar represents one group of cosegregating SNP markers showing in bands.

Fig. 4. A major QTL was colocalized for SA, SL, SW, PL and TSW on chromosome 17. The horizontal line indicates the average LOD threshold for significance at the 1% level. Physical locations of flanking SNP markers are indicated by arrows.

Table 1. SNP marker distribution and characteristics of a linkage map.

Chr.	SNP number	Unique loci number	Genetic distance (cM)	Marker density (cM/locus)	Corresponding physical length (Mb）	Average interval length (Mb/locus)
1	77	42	85.2	2.1	19.44	0.47
2	143	70	93.0	1.3	28.45	0.41
3	182	95	140.3	1.5	27.52	0.29
4	149	59	86.4	1.5	28.65	0.49
5	64	52	98.3	1.9	31.15	0.61
6	133	63	80.2	1.3	19.65	0.32
7	163	110	123.2	1.1	32.66	0.30
8	113	71	94.8	1.4	27.45	0.39
9	68	56	112.1	2.0	37.44	0.68
10	15	13	40.0	3.3	11.21	0.93
11	235	123	154.9	1.3	47.97	0.39
12	138	66	94.8	1.5	18.84	0.29
13	137	62	99.2	1.6	10.34	0.17
14	98	55	103.0	1.9	26.75	0.50
15	113	69	112.1	1.6	15.64	0.23
16	168	92	127.4	1.4	24.80	0.27
17	56	45	89.7	2.0	32.55	0.74
18	101	63	89.5	1.4	16.19	0.26
19	120	78	101.7	1.3	19.54	0.25
20	103	61	108.8	1.8	25.22	0.42
Total	2376	1345	2034.6	1.5	501.48	0.37

Table 2. QTL for seed area (SA), seed length (SL), seed width (SW), pod area (PA), pod length (PL), pod width (PW), 1000-seed weight (TSW), seed number per pod (SPP) and seed oil content (TOC).

Trait	QTL	Chr.	Peak (cM)	Marker interval	Environments	LOD range	PVE range	Additive effect
Seed area	SA1.1	1	24.0	MC00234299-MC00990463	D17, D18, I17, M	3.4-4.6	7.9-10.6	0.03
	SA3.1	3	52.0	MC01633505-MC01061047	D17, D18, I17, M	3.0-3.9	6.9-9.0	0.02
	SA4.1	4	60.1	MC00535890-MC00230778	I17, I18, M	2.6-3.6	6.2-8.3	0.02
	SA8.1	8	8.0	MC01364407-MC00124954	D17, I17, M	2.8-4.1	6.6-9.4	0.02
	SA16.1	16	24.0	MC00823230-MC00111882	D17, D18, I17, M	2.7-4.1	6.5-9.5	0.02
	SA16.2	16	110.0	MC01575409-MC00200166	D17, D18, I18, M	3.0-6.0	7.1-13.5	0.03
	SA17.1	17	39.9	MC00670835-MC00824085	D17, D18, I17, I18, M	4.1-5.8	9.6-13.1	0.03
Seed length	SL1.1	1	8.4	MC00352045-MC00799081	D17, D18, I17, M	2.7-4.1	6.5-9.5	0.02
	SL2.1	2	55.0	MC01397441-MC00000806	D18, I18, M	3.0-3.3	7.0-7.7	0.02
	SL3.1	3	30.0	MC00588500-MC01316809	D17, D18, I17, I18, M	3.2-5.4	7.6-12.3	0.02
	SL3.2	3	52.0	MC01633505-MC01061047	D17, D18, I17, I18, M	3.0-4.7	7.0-10.8	0.02
	SL5.1	5	58.0	MC00628341-MC00006273	D17, I18, M	2.9	6.7-6.9	0.02
	SL17.1	17	39.8	MC00670835-MC00824085	D17, D18, I17, I18, M	3.5-5.1	8.1-11.7	0.03
Seed width	SW4.1	4	72.3	MC01298997-MC01336929	D17, D18, I17, I18, M	3.5-6.2	8.2-14.0	0.01
	SW8.1	8	2.1	MC00622177-MC00154746	D17, D18, I17, M	2.8-4.0	6.6-9.4	0.01
	SW17.1	17	40.3	MC00670835-MC00824085	D18, I17, I18, M	3.3-3.9	7.8-9.1	0.01

	SW20.1	20	94.0	MC01446056-MC00178584	D18, I18, M	2.8-3.3	6.7-7.7	0.01
Pod area	PA5.1	5	58.0	MC00628341-MC00006273	D17, D18, I17, I18, M	4.3-6.1	9.9-13.9	0.70
	PA17.1	17	35.2	MC00219016-MC01347180	D17, I18, M	2.7-2.8	6.3-6.7	0.50
	PA18.1	18	68.9	MC00721630-MC00785854	D17, I17, M	3.4-3.8	8.0-8.8	0.53
	PA19.1	19	84.0	MC01085171-MC01480861	D17, D18, M	2.6-4.7	6.2-10.9	0.56
Pod length	PL5.1	5	58.0	MC00628341-MC00006273	D17, D18, I17, I18, M	3.7-5.5	8.7-12.4	0.14
	PL11.1	11	97.9	MC00780322-MC00000240	D17, I17, M	2.8-4.2	6.7-9.7	0.10
	PL17.1	17	39.6	MC00670835-MC00824085	D17, D18, I17, I18, M	3.0-5.1	7.2-11.7	0.14
Pod width	PW1.1	1	44.1	MC00511717-MC00397943	D17, I17, I18, M	2.8-4.4	6.6-10.2	0.06
	PW4.1	4	70.7	MC01298997-MC01336929	D17, D18, I17, I18, M	2.6-5.8	6.0-13.2	0.09
	PW5.1	5	59.0	MC00006273-MC00445645	D18, I17, M	2.9-3.6	6.9-8.4	0.06
	PW9.1	9	54.1	MC00308400-MC01578045	D17, D18, I18, M	3.4-4.0	7.9-9.3	0.06
1000-seed weight	TSW1.1	1	6.9	MC00352045-MC00799081	D17, D18, I17, M	3.6-4.8	8.5-11.1	0.03
	TSW2.1	2	29.0	MC00219687-MC00971906	D18, I18, M	3.2-3.5	7.4-8.1	0.02
	TSW4.1	4	60.1	MC00535890-MC00230778	D18, I17, M	2.7-3.1	6.3-7.2	0.02
	TSW5.1	5	58.0	MC00628341-MC00006273	I18	3.9	9.0	0.02
	TSW8.1	8	7.2	MC01364407-MC00124954	D17, I17, M	3.0-5.1	7.0-11.6	0.02
	TSW17.1	17	40.1	MC00670835-MC00824085	D17, D18, I17, I18, M	4.0-6.8	9.3-15.4	0.03
Seeds per pod	SPP2.1	2	33.0	MC00971906-MC00202412	D17, D18, I18, M	2.5-3.1	5.8-7.2	-0.30

	SPP9.1	9	54.0	MC00592527-MC01578045	I18, M	3.0-3.1	7.1-7.2	0.30
Oil content	TOC2.1	2	17.9	MC00615304-MC01451818	D17, D18, I17, I18, M	2.9-6.4	6.7-14.4	-0.42
	TOC9.1	9	42.0	MC01266417-MC00047724	D17, D18, I17, I18, M	2.5-7.5	6.0-16.7	0.42
	TOC16.1	16	59.5	MC00197595-MC01022622	D17, I17	2.5-3.9	5.9-9.1	0.26
	TOC20.1	20	94.8	MC00767095-MC00178584	D17, D18, I17, M	4.0-7.3	9.3-16.4	-0.41

D17, dryland 2017; D18, dryland 2018; I17, irrigated 2017; I18, irrigated 2018; M, mean of all four environments. LOD, logarithm of odds; PVE, phenotypic variation explained.