

[Click here to view linked References](#)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

# 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](mailto:clu@montana.edu)

ORCID ID: 0000-0001-6901-7490

1  
2  
3  
4 **Abstract**  
5

6 Camelina (*Camelina sativa* L. Crantz) is an oilseed crop that has great potential to  
7 provide sustainable feedstock for biofuel production and to improve dryland agriculture.  
8 A major breeding objective for camelina is to increase seed size. Understanding the  
9 genetics behind variation of seed size and associated traits such as oil content would help  
10 breeders develop varieties of increased oil yield that are more robust, easier to plant and  
11 harvest, better for oil processing. In this study, we developed a recombinant inbred  
12 population derived from the two *Camelina* accessions, Suneson and Pryzeth, with  
13 contrasting traits especially seed size and oil content. Using 189 lines, a genetic map was  
14 constructed containing 2376 single nucleotide polymorphism (SNP) markers spanning  
15 2034.6 cM of 20 linkage groups with an average density of 1.5 cM per locus. Field trials  
16 were conducted for two years (2017 and 2018) in two environments (dryland and  
17 irrigated) in Bozeman, Montana. The results revealed important correlations of seed size  
18 with other associated traits such as oil content, pod size and seed number per pod.  
19 Significant QTL were also discovered for these traits. The results of this study are the  
20 first step to isolate genes controlling seed development and oil accumulation, and to  
21 develop advanced varieties of camelina better adapted to modern agriculture by marker-  
22 assisted breeding.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

44 **Keywords**

45 *Camelina sativa*, genetic linkage map, quantitative trait loci, oil content, seed size  
46  
47  
48  
49

50 **Key Message**

51  
52  
53 Genetic dissection of oil content and seed size in *Camelina sativa* was conducted by QTL  
54 mapping using a SNP-based linkage map and a recombinant inbred population.  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Introduction

Camelina (*Camelina sativa* L. Crantz) is a member of the *Brassicaceae* family. It is a potentially productive oilseed crop that is highly adaptable in a wide 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 which result in volunteers for the following crops besides yield loss (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 oleracea* (Finch-Savage et al. 2010). Higher seedling rates and greater biomass (e.g., oil content) associated with large seed sizes also increased crop competition with weeds and hastened 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,

1  
2  
3  
4 in the camelina germplasm (Vollmann et al. 2005) also suggested that multiple genes are  
5 involved in regulating this important trait, and quantitative trait loci (QTL) were identified  
6 controlling thousand seed mass (Gehring et al. 2006). In this study, we created a  
7 recombinant inbred line (RIL) population from the cross between two camelina varieties of  
8 contrasting traits especially in seed size. The distribution of seed size in the RIL population  
9 clearly indicated that seed size is quantitatively inherited. We constructed a high-density  
10 genetic linkage map using single nucleotide polymorphism (SNP) markers, and localized  
11 several QTL that control seed size and its associated traits including oil content, pod size  
12 and seed per pod of camelina.  
13  
14  
15  
16  
17  
18  
19  
20  
21

## 22 **Materials and Methods**

### 23 **Plant materials and field experiments**

24  
25 A population consisting of 189 recombinant inbred lines (RILs) was developed by single  
26 seed descent from a cross between a big-seeded accession Pryzeth and an elite cultivar  
27 Suneson. This population was evaluated in four field-trial environments: The F<sub>5-6</sub> RILs  
28 and two parental lines were arranged in an augmented randomized complete block design  
29 and planted at the Arthur H. Post Research Farm (Bozeman, MT, USA) in 2017 and  
30 2018. Field trials were conducted under rain-fed, dryland (D17, D18) and irrigated  
31 conditions (I17, I18) following standard agricultural practices. In the irrigation treatment,  
32 approximately 12cm of supplemental water was applied during the period of seed filling  
33 by overhead sprinklers. Five plants in the middle of each plot were randomly selected and  
34 harvested for further analysis.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

### 46 **Phenotypic measurement**

47  
48 Twenty mature pods were harvested from the main stems by hand. Pods and seeds were  
49 scanned by an Epson Perfection V600 Photo Scanner at 600 dpi (0.024 mm/pixel) to  
50 calculate their sizes using SmartGrain (Tanabata et al. 2012), a high-throughput seed  
51 measurement software. A set of coordinates representing seed morphological perimeters  
52 (seed area, seed length and seed width) were obtained as raw phenotypic data. Seeds from  
53 these 20 pods were counted using a seed counter machine and weighted to calculate seed  
54 weight and seed number per pod. Morphological pod traits (pod area, pod length and pod  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 width) were measured by the same method as above. Approximately 20 mg dry seeds  
5 from each line were used for oil content measurement. Seed oil content was determined  
6 by a bench-top NMR seed analyzer (MQC23, Oxford Instruments, Concord, MA, United  
7 States) and presented as percentages of seed mass. Best linear unbiased prediction  
8 (BLUP) values of each trait of each line were obtained and used as phenotypic data for  
9 mapping analysis. The broad-sense heritability of each trait was calculated according to  
10 (Holland et al. 2003). All descriptive statistics including Pearson's correlation coefficient  
11 of all traits were calculated in Rstudio (v1.1.383).  
12  
13  
14  
15  
16  
17  
18  
19

### 20 **Genotyping by sequencing (GBS) and SNP calling**

21  
22 Total genomic DNA was isolated from young leaves of all RILs and two parents using  
23 the FastPrep DNA extraction and purification kit (MP biomedical). GBS was performed  
24 at the LGC Genomics (<http://www.lgcgroup.com>). Specifically, a commercially licensed  
25 normalized GBS (nGBS) using MspI as the restriction enzyme, which efficiently excludes  
26 the conserved repetitive elements in the genome, was performed on Illumina NextSeq  
27 500 sequencer. Sequencing reads were pre-processed and aligned by using bcl2fastq and  
28 Bowtie2 version 2.2.3 software (Langmead and Salzberg 2012). Variant detection and  
29 progeny genotyping were performed with Freebayes v1.0.2-16 pipeline (Garrison and  
30 Marth 2012). Raw SNP calls with minor allele frequency (MAF) >0.05 in the biparental  
31 population were concatenated into a single Variant Call File (VCF) and proceeded for  
32 linkage map construction.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

### 44 **Linkage map construction**

45  
46 First, SNPs with more than 15% missing rate and those that were monomorphic between  
47 two parental lines were eliminated. The remaining SNPs with significant segregation  
48 distortion ( $p < 0.05$ ) were removed through chi-square test combined with sequential  
49 Bonferroni correction. A genetic linkage map was constructed using MapDisto 2.0  
50 (Lorieux 2012). In MapDisto, all markers were grouped with a recombination fraction of  
51 0.3 and logarithm of the odds ratio (LOD) score of 7 using the Kosambi mapping  
52 function (Kosambi 1944). The Seriation algorithm was used for ordering the linkage  
53 groups. The "Automap" command which successively runs multiple steps including  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 “Ripple order” and “Check inversions” was used to refine the marker order of each  
5 linkage group. Linkage groups were assigned to chromosomes by blasting the tag  
6 sequences of their allocated GBS makers against the *Camelina sativa* genome database  
7 (<http://www.camelinadb.ca/>). An iterative strategy was employed to obtain the final  
8 linkage map which agrees with the number of camelina chromosomes. By blasting the  
9 sequences of markers located on the ends of each linkage group, the physical coverage  
10 was calculated and average interval length was estimated. The graphical representation of  
11 the genetic map was generated using MapChart version 2.2 (Voorrips 2002).  
12  
13  
14  
15  
16  
17  
18  
19  
20

### 21 **QTL analysis and candidate gene selection**

22  
23 Phenotypic data of each trait from individual environments, along with the mean values of  
24 all four trials were integrated with the genotypic data and used for QTL mapping. All  
25 mapping analysis was performed using QGene 4.4.0v (Joehanes and Nelson 2008) with the  
26 composite interval mapping (CIM) function. LOD significance thresholds of linkage  
27 groups were determined independently for each trait using the permutation test (n=1000,  
28 a=0.05). Cofactors of each group were selected by using the default parameters. The  
29 walking speed was set to 1cM to scan intervals and the coefficient of determination ( $R^2$ )  
30 was calculated to estimate the amount of phenotypic variation explained (PVE) by each  
31 QTL.  
32  
33  
34  
35  
36  
37  
38  
39

40 Flanking markers of major QTL detected in multiple environments were used to  
41 delineate the physical intervals. Their tag sequences in GBS dataset were blasted against  
42 the *Camelina sativa* genome V2.0 (<http://www.camelinadb.ca/>) using BLASTn. All  
43 annotated genes residing within the QTL intervals were subject to further screening for  
44 candidates implicated in seed/pod development, and their orthologous to the Arabidopsis  
45 genes (<http://www.arabidopsis.org/>).  
46  
47  
48  
49  
50  
51

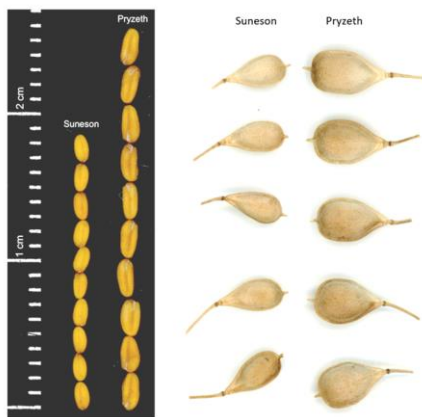
## 52 **Results**

### 53 **Seed traits evaluations**

54  
55 Two parental lines, Suneson and Pryzeth, have contrasting sizes in seed and pod. Field-  
56 harvested seeds from Pryzeth were, on average, 66% larger and 71% heavier than seeds  
57  
58  
59  
60  
61  
62  
63  
64  
65

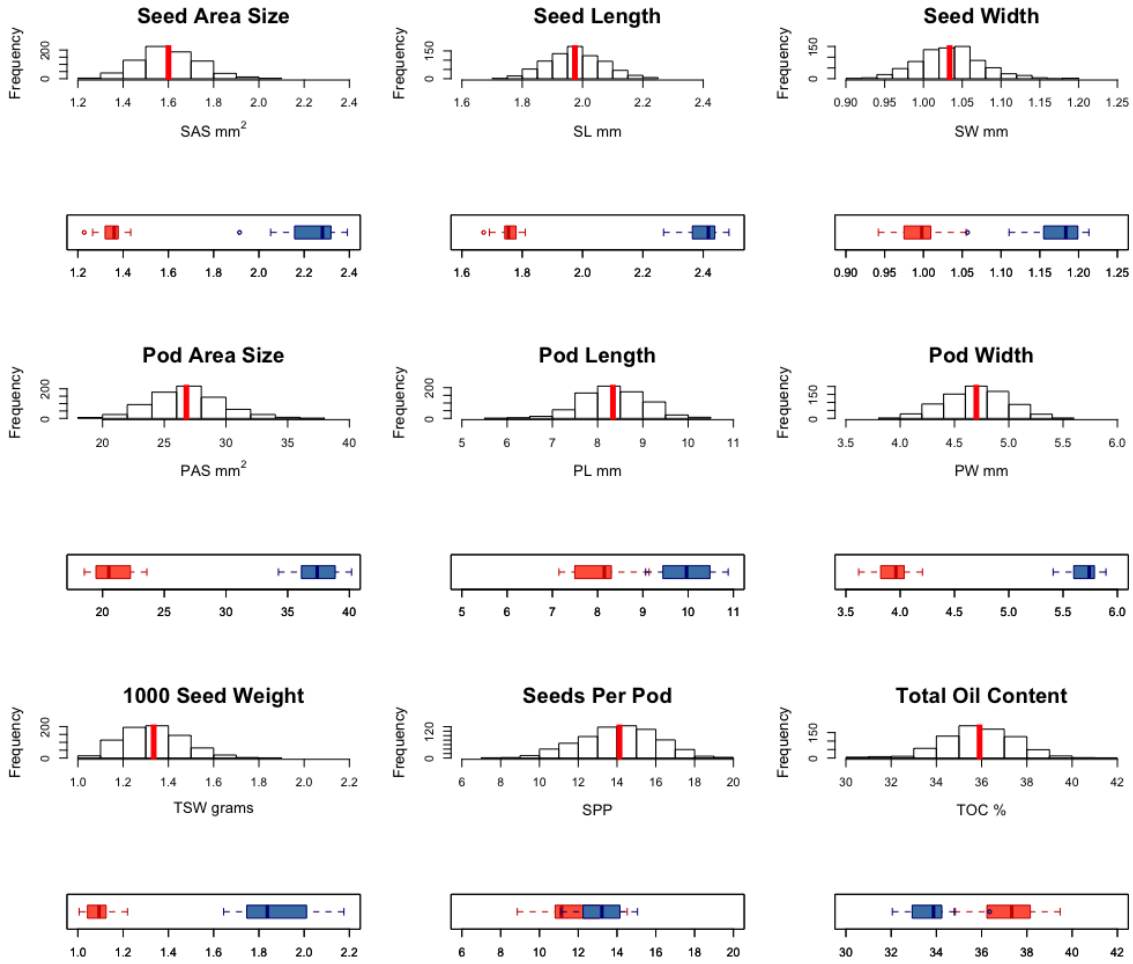
1  
2  
3  
4 from Suneson. Similarly, pods of Pryzeth were also 80% larger than those of Suneson  
5 (Fig. 1, Table S1). Oil contents also differ in Pryzeth (33%) and Suneson (37%). To more  
6 accurately evaluate seed and pod sizes, their length and width were measured along with  
7 area by scanning and imaging analyses. In addition, 1000-seed weight (TSW), seed  
8 number per pod (SPP) and total oil content (TOC) were also obtained. Frequency  
9 distributions of all these measurements did not show significant deviation from normal  
10 distributions in the RIL population (Fig. 2), and all traits exhibited high broad-sense  
11 heritability (Table S1). These results suggest that oil content and all seed and pod traits in  
12 this study are quantitatively controlled and their genetic effects are the major determinant  
13 of the phenotypic variations. Interestingly, among above traits, only SPP showed a  
14 pattern of transgressive segregation in our population.  
15  
16  
17  
18  
19  
20  
21  
22  
23

24 Similar to results observed in *Arabidopsis* (Gnan et al. 2014), pair-wise  
25 comparison indicated that seed size determined by scanned seed area (SA) was highly  
26 correlated with TSW ( $r = 0.89$ ). We then analyzed seed size for its separate parameters,  
27 seed length (SL) and seed width (SW). SA was shown to be highly positively correlated  
28 with both SL and SW ( $r = 0.92$  and  $0.88$ , respectively), and the latter two were also  
29 clearly correlated ( $r = 0.65$ ) (Table S2). Pod area (PA) was similarly measured, which  
30 showed weak correlation with SA ( $r = 0.39$ ), indicating different genetic mechanisms for  
31 these two quantitative traits. The width-length correlation was also less significant in  
32 pods ( $r = 0.31$ ). SPP was positively correlated with PA ( $r = 0.39$ ) but had a negative  
33 correlation with SA ( $r = -0.30$ ). The correlations between total oil content (TOC) and all  
34 these morphometric traits were slightly negative, with the highest observed between TOC  
35 and SL ( $r = -0.32$ ).  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45



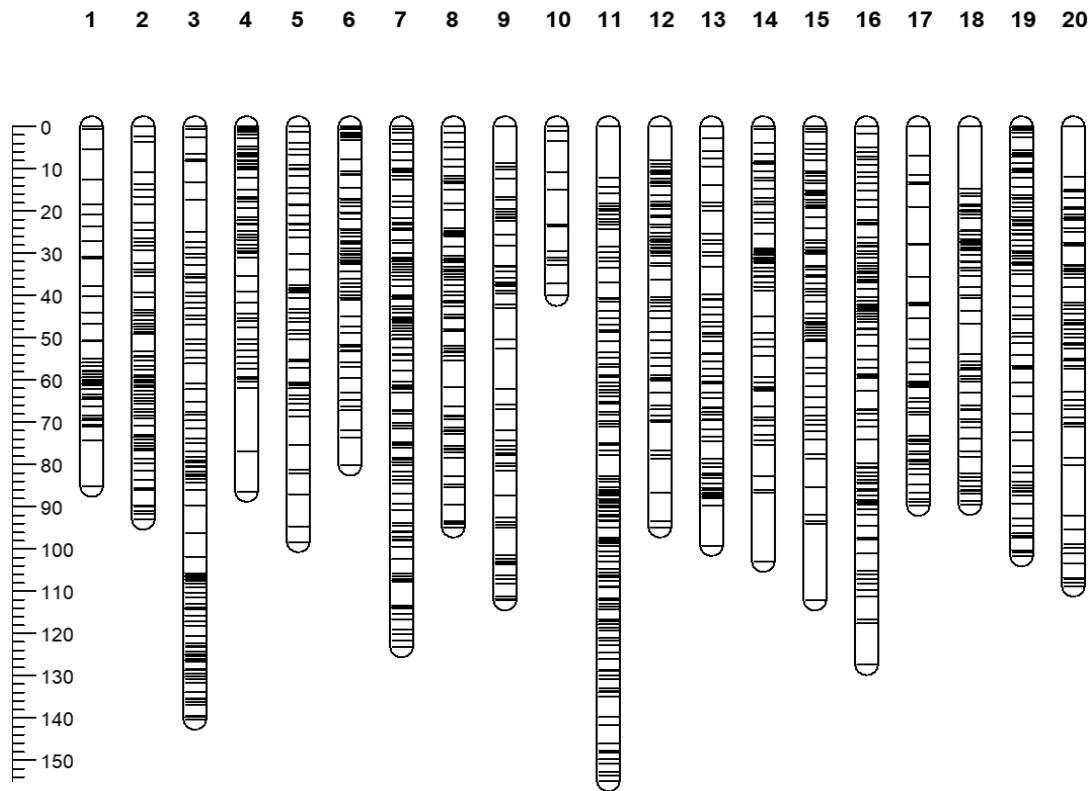
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Fig. 1.** Comparison of sizes of dry seeds and pods between camelina varieties ‘Suneson’ (left) and ‘Pryzeth’ (right).



**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.

### 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 3.

**Table 3.** 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

## QTL for seed size

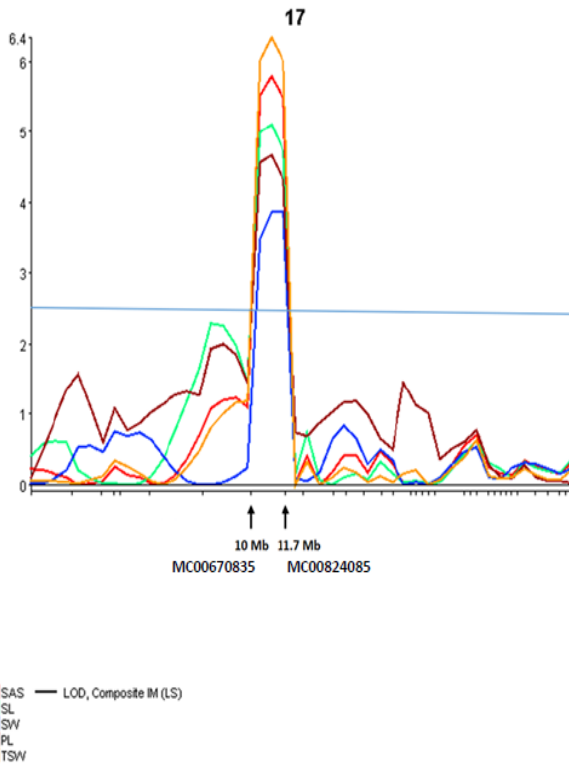
We detected a total of 40 QTL across the four environments for the following traits: seed area (SA), seed length (SL), seed width (SW), pod area (PA), pod length (PL), pod width (PW), 1000-seed weight (TSW), seed number per pod (SPP) and total oil content (TOC). The QTL were named SA1.1, SL1.1, SW4.1 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, PAS, PL, PW), a total of 28 QTL consistently identified in at least two environments were described in Table 4, 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 colocalized 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 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 and compared the results with their corresponding unconditional

1  
2  
3  
4 analysis. Both QTL peaks detected by the unconditional method disappeared in our  
5  
6 conditional analysis, indicating that QTL SA17.1 and TSW17.1 are not independent of  
7  
8 each other. In other words, pleiotropy rather than tight linkage resulted in the co-  
9  
10 localization of SA17.1 and TSW17.1 (Fig. S1).  
11  
12  
13  
14  
15  
16



17  
18  
19  
20  
21  
22 **Fig. 4.** A major QTL was  
23 colocalized for SA, SL, SW, PL  
24 and TSW on chromosome 17.  
25 The horizontal line indicates the  
26 average LOD threshold for  
27 significance at the 1% level.  
28 Physical locations of flanking  
29 SNP markers are indicated by  
30 arrows.  
31  
32

**Table 4.** QTL for seed area (SA), seed length (SL), seed width (SW), pod area size (PAS), 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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

	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.

### QTL for pod size, seed 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. Among the four TOC QTL, only one (TOC20.1) was colocalized with seed size traits (SW20.1), suggesting 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, in the regions of QTL for oil content, we found candidate genes that are known to be involved in lipid biosynthesis (Li-Beisson et al. 2013), including an acyl-CoA-binding domain-containing protein, a glycerol-3-phosphate acyltransferase, and a WSD1-like acyltransferase. 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



1  
2  
3  
4 number of traits including oil content, 1000-seed-weight and plant height in the same study  
5  
6 (Gehring et al. 2006).

7  
8 In this study, we focused on identification of QTL controlling the most important  
9  
10 seed traits in camelina including seed size and oil content. We generated a RIL population  
11  
12 from two accessions of camelina, Suneson and Prezyth, which are significantly different in  
13  
14 seed sizes and other traits. Such diverse parental lines created vast genetic and phenotypic  
15  
16 diversities in the progeny population, and allowed us to generate a large number of  
17  
18 polymorphic SNP markers for linkage mapping. The RIL population also allowed for  
19  
20 repeated field trials for trait evaluations and QTL localization.

21 Seed area is usually estimated by digital collection of seed area or by weight  
22  
23 (Herridge et al. 2011; Gnan et al. 2014). Besides these, we also measured seed length and  
24  
25 seed width by a document scanner. Distribution frequencies of all these traits did not show  
26  
27 significance from normal distributions, indicating their quantitative inheritance.  
28  
29 Phenotypically, Seed area (SA) showed high correlation with both seed length (SL) and  
30  
31 seed width (SW), and all these seed morphometric traits (SA, SL and SW) are highly  
32  
33 correlated with TSW. However, the numbers of QTL for these traits differed, and few QTL  
34  
35 were co-localized although SA has more overlapping QTL with SL than SW. These results  
36  
37 suggest that evaluation of seed size by single measurements may be oversimplified. As SA,  
38  
39 SL and SW may be captured by a single scan, all these measurements along with TSW  
40  
41 should be used to map QTL. Further fine mapping of QTL and characterization of  
42  
43 candidate genes, such as those listed in Table S3, will allow for their contributions to the  
44  
45 final determination of seed size and seed yield.

46  
47 Pods not only protect immature seeds, they are also photosynthetically active and  
48  
49 play important roles in seed development and storage filling (Bennett et al. 2011). Multiple  
50  
51 QTL were detected for pod size. Our observations indicate that sizes of pod and seed are  
52  
53 correlated, and we found that one QTL may have pleiotropic effects on seed and pod sizes.  
54  
55 The sizes of pod and seed affect the number of seeds per pod. As expected, larger pods  
56  
57 contain more seeds per pod, but SPP decreases with increased seed size. These  
58  
59 relationships suggest that increasing pod size may be beneficial for seed enlargement, but  
60  
61 may limit the number of seeds per pod. Seed number and seed size/weight are all  
62  
63 components of seed yield, therefore it is important to find the optimum seed size and seed  
64  
65

1  
2  
3  
4 numbers. Unfortunately due to severe shattering by hail storms, we were not able to  
5 investigate the total numbers of pods and seed per plant, their correlations and effects on  
6 seed yield.  
7  
8

9  
10 Since seed size/weight, seed number and oil content are all important traits to  
11 determine seed and oil yield, understanding their genetic basis is necessary to camelina  
12 improvement. Potential trade-offs between these traits may be observed in population  
13 studies. The negative correlation between seed size (measured as 1000-seed weight) and  
14 oil content was observed in camelina (Gehring et al. 2006; Vollmann et al. 2007). Our  
15 results also showed a negative correlation between TOC and TSW, however was less  
16 significant than previous studies ( $r=-0.18$  compared to  $-0.37$  and  $-0.92$ ). There was no  
17 overlap of QTL for seed size traits and TOC. This discrepancy suggested that the  
18 correlation between oil content and seed size might be dependent on genotypes and  
19 environments. The production of seed and its contents depends on the allocation of finite  
20 resources and is not discrete from the whole plant carbon budget (Venable 1992). As  
21 exemplified in our recent study, seed size was increased by suppression of starch  
22 biosynthesis in camelina, however oil content did not change, and the increased seed  
23 weight was contributed by increased protein content (Na et al. 2018). Therefore, a  
24 systematic approach is possible to simultaneously increase seed size and oil content by  
25 orchestrating carbon allocation, transport and sink activities.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

#### 40 **Author contributions**

41  
42 JK and CL developed the RIL population; KK, HL collected and analyzed phenotypic data  
43 and constructed the linkage map. CL conceived the project and guided the analyses. CL  
44 and HL wrote the manuscript. All authors read and approved the final manuscript.  
45  
46  
47  
48

#### 49 **Acknowledgement**

50  
51  
52 We thank Drs. John Martin and Jamie Sherman for their help on statistical analyses. This  
53 work was supported by the Office of Science (BER), US Department of Energy, grant no.  
54 DE-SC0016400, and the USDA National Institute of Food and Agriculture Hatch project  
55 no. 1015638.  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5 **Compliance with ethical standards**  
6  
7  
8  
9

10  
11 **Conflict of interest**  
12  
13

14 The authors declare that they have no conflict of interest.  
15  
16  
17

18 **References**  
19  
20

- 21 Abdullah HM, Akbari P, Paulose B, Schnell D, Qi W, Park Y, Pareek A, Dhankher OP  
22 (2016) Transcriptome profiling of *Camelina sativa* to identify genes involved in  
23 triacylglycerol biosynthesis and accumulation in the developing seeds. *Biotechnol*  
24 *Biofuels* 9: 136. doi:10.1186/s13068-016-0555-5  
25  
26 Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M,  
27 Fei J, Harada CM, Munoz MD (2013) Comprehensive developmental profiles of  
28 gene activity in regions and subregions of the *Arabidopsis* seed. *Proc Natl Acad*  
29 *Sci USA* 110: E435-E444  
30  
31 Bennett EJ, Roberts JA, Wagstaff C (2011) The role of the pod in seed development:  
32 strategies for manipulating yield. *New Phytol* 190: 838-853  
33  
34 Berti M, Gesch R, Eynck C, Anderson J, Cermak S (2016) *Camelina* uses, genetics,  
35 genomics, production, and management. *Ind Crop Prod* 94: 690-710.  
36 doi:<https://doi.org/10.1016/j.indcrop.2016.09.034>  
37  
38 Chao H, Wang H, Wang X, Guo L, Gu J, Zhao W, Li B, Chen D, Raboanatahiry N, Li M  
39 (2017) Genetic dissection of seed oil and protein content and identification of  
40 networks associated with oil content in *Brassica napus*. *Sci Rep* 7:46295.  
41 doi:10.1038/srep46295  
42  
43 Finch-Savage WE, Clay HA, Lynn JR, Morris K (2010) Towards a genetic understanding  
44 of seed vigour in small-seeded crops using natural variation in *Brassica oleracea*.  
45 *Plant Sci* 179: 582-589. doi:<https://doi.org/10.1016/j.plantsci.2010.06.005>  
46  
47 Fleenor RA (2011) *Plant Guide for Camelina (Camelina sativa)*. USDA-Natural  
48 Resources Conservation Service, Spokane, WA 99201,  
49  
50 Garrison E, Marth G (2012) Haplotype-based variant detection from short-read  
51 sequencing. arXiv:12073907:  
52  
53 Gehringer A, Friedt W, Luhs W, Snowdon RJ (2006) Genetic mapping of agronomic  
54 traits in false flax (*Camelina sativa* subsp. *sativa*). *Genome* 49: 1555-1563.  
55 doi:10.1139/g06-117  
56  
57 Gesch RW, Johnson JMF (2015) Water use in camelina–soybean dual cropping systems.  
58 *Agron J* 107: 1098. doi:10.2134/agronj14.0626  
59  
60 Gnan S, Priest A, Kover PX (2014) The genetic basis of natural variation in seed size and  
61 seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines.  
62 *Genetics* 198: 1751-1758. doi:10.1534/genetics.114.170746  
63  
64  
65

- 1  
2  
3  
4 Harker K, O'Donovan J, Smith E, Johnson E, Peng G, Willenborg C, Gulden R, Mohr R,  
5 Gill K, Grenkow L (2015) Seed size and seeding rate effects on canola  
6 emergence, development, yield and seed weight. *Can J Plant Sci* 95: 1-8  
7  
8 Herridge R, Day R, Baldwin S, Macknight R (2011) Rapid analysis of seed size in  
9 *Arabidopsis* for mutant and QTL discovery. *Plant Methods* 7: 3  
10  
11 Holland JB, Nyquist WE, Cervantes-Martínez CT (2003) Estimating and interpreting  
12 heritability for plant breeding: an update. *Plant Breed Rev* 22: 9-112  
13  
14 Hunter J, Roth G (2010) *Camelina* production and potential in Pennsylvania, *Agronomy*  
15 *Facts* 72. College of Agricultural Sciences, Crop and Soil Sciences, Pennsylvania  
16 State University:  
17  
18 Joehanes R, Nelson JC (2008) QGene 4.0, an extensible Java QTL-analysis platform.  
19 *Bioinformatics* 24: 2788-2789  
20  
21 Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ,  
22 Links MG, Clarke C, Higgins EE, Huebert T, Sharpe AG, Parkin IA (2014) The  
23 emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid  
24 genome structure. *Nat Commun* 5: 3706. doi:10.1038/ncomms4706  
25  
26 Kagale S, Nixon J, Khedikar Y, Pasha A, Provart NJ, Clarke WE, Bollina V, Robinson  
27 SJ, Coutu C, Hegedus DD, Sharpe AG, Parkin IA (2016) The developmental  
28 transcriptome atlas of the biofuel crop *Camelina sativa*. *Plant J* 88: 879-894.  
29 doi:10.1111/tbj.13302  
30  
31 Kosambi DD (1944) The estimation of map distances from recombination values. *Ann*  
32 *Eugen* 12: 172-175  
33  
34 Krannitz PG, Aarssen LW, Dow JM (1991) The effect of genetically based differences in  
35 seed size on seedling survival in *Arabidopsis thaliana* (Brassicaceae). *Amer J Bot*  
36 78: 446-450  
37  
38 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat*  
39 *Methods* 9: 357-359. doi:10.1038/nmeth.1923  
40  
41 Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S,  
42 Bird D, Debono A, Durrett TP, Franke RB, Graham IA, Katayama K, Kelly AA,  
43 Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L,  
44 Schmid KM, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J (2013) Acyl-lipid  
45 metabolism. *Arabidopsis Book* 11: e0161. doi:10.1199/tab.0161  
46  
47 Li N, Li Y (2016) Signaling pathways of seed size control in plants. *Curr Opin Plant Biol*  
48 33: 23-32. doi:<https://doi.org/10.1016/j.pbi.2016.05.008>  
49  
50 Li N, Xu R, Li Y (2019) Molecular Networks of Seed Size Control in Plants. *Annu Rev*  
51 *Plant Biol* 70: 11.11-11.30. doi:10.1146/annurev-arplant-050718-095851  
52  
53 Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. *Mol*  
54 *Breeding* 30: 1231-1235  
55  
56 Lu C, Kang J (2008) Generation of transgenic plants of a potential oilseed crop *Camelina*  
57 *sativa* by *Agrobacterium*-mediated transformation. *Plant Cell Rep* 27: 273-278.  
58 doi:10.1007/s00299-007-0454-0  
59  
60 McVay KA, Khan QA (2011) *Camelina* yield response to different plant populations  
61 under dryland conditions. *Agron J* 103: 1265-1269. doi:10.2134/agronj2011.0057  
62  
63 Moore RH, Thornhill KL, Weinzierl B, Sauer D, D'Ascoli E, Kim J, Lichtenstern M,  
64 Scheibe M, Beaton B, Beyersdorf AJ, Barrick J, Bulzan D, Corr CA, Crosbie E,  
65 Jurkat T, Martin R, Riddick D, Shook M, Slover G, Voigt C, White R, Winstead

- 1  
2  
3  
4 E, Yasky R, Ziemba LD, Brown A, Schlager H, Anderson BE (2017) Biofuel  
5 blending reduces particle emissions from aircraft engines at cruise conditions.  
6 Nature 543: 411-415. doi:10.1038/nature21420  
7
- 8 Na G, Aryal N, Fatihi A, Kang J, Lu C (2018) Seed-specific suppression of ADP-glucose  
9 pyrophosphorylase in *Camelina sativa* increases seed size and weight. Biotechnol  
10 Biofuels 11: 330. doi:10.1186/s13068-018-1334-2  
11
- 12 Na G, Mu X, Grabowski P, Schmutz J, Lu C (2019) Enhancing microRNA167A  
13 expression in seed decreases the alpha-linolenic acid content and increases seed  
14 size in *Camelina sativa*. Plant J: doi:10.1111/tpj.14223. doi:10.1111/tpj.14223  
15
- 16 Nguyen HT, Silva JE, Podicheti R, Macrander J, Yang W, Nazareus TJ, Nam J-W,  
17 Jaworski JG, Lu C, Scheffler BE, Mockaitis K, Cahoon EB (2013) Camelina seed  
18 transcriptome: a tool for meal and oil improvement and translational research.  
19 Plant Biotechnol J 11: 759-769. doi:10.1111/pbi.12068  
20
- 21 Obour AK, Chen C, Sintim HY, McVay K, Lamb P, Obeng E, Mohammed YA, Khan Q,  
22 Afshar RK, Zheljzkov VD (2018) *Camelina sativa* as a fallow replacement crop  
23 in wheat-based crop production systems in the US Great Plains. Ind Crop Prod  
24 111: 22-29. doi:<https://doi.org/10.1016/j.indcrop.2017.10.001>  
25
- 26 Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L (2015) Networks controlling seed size  
27 in Arabidopsis. Plant Reprod 28: 17-32. doi:10.1007/s00497-015-0255-5  
28
- 29 Singh R, Bollina V, Higgins EE, Clarke WE, Eynck C, Sidebottom C, Gugel R, Snowdon  
30 R, Parkin IA (2015) Single-nucleotide polymorphism identification and  
31 genotyping in *Camelina sativa*. Mol Breed 35: 35. doi:10.1007/s11032-015-0224-  
32 6  
33
- 34 Sintim HY, Zheljzkov VD, Obour AK, Garcia AGY, Foulke TK (2016) Evaluating  
35 agronomic responses of Camelina to seeding date under rain-fed conditions.  
36 Agron J 108: 349-357. doi:10.2134/agronj2015.0153  
37
- 38 Tanabata T, Shibaya T, Hori K, Ebana K, Yano M (2012) SmartGrain: High-throughput  
39 phenotyping software for measuring seed shape through image analysis. Plant  
40 Physiol 160: 1871-1880. doi:10.1104/pp.112.205120  
41
- 42 Venable DL (1992) Size-number trade-offs and the variation of seed size with plant  
43 resource status. Amer Nat 140: 287-304  
44
- 45 Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T (2005) Genetic diversity in  
46 camelina germplasm as revealed by seed quality characteristics and RAPD  
47 polymorphism. Plant Breed 124: 446-453. doi:DOI 10.1111/j.1439-  
48 0523.2005.01134.x  
49
- 50 Vollmann J, Moritz T, Kargl C, Baumgartner S, Wagentristl H (2007) Agronomic  
51 evaluation of camelina genotypes selected for seed quality characteristics. Ind  
52 Crop Prod 26: 270-277. doi:10.1016/j.indcrop.2007.03.017  
53
- 54 Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and  
55 QTLs. J Hered 93: 77-78  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8 **Supplementary Tables and Figure**  
9

10 **Table S1.** Phenotype-based EBLUP and heritability for the RIL population.  
11  
12

Trait	Pryzeth	Suneson	RIL 2017	RIL 2018	Range	Heritability
SA	2.24 ±0.12	1.35 ±0.04	1.56 ±0.004	1.64 ±0.004*	1.33-1.94	0.85 ±0.017
SL	2.4 ±0.05	1.75 ±0.03	1.96 ± 0.004	1.99 ±0.003*	1.76-2.19	0.88 ±0.014
SW	1.17 ±0.04	1.00 ±0.03	1.02 ± 0.001	1.04 ±0.001*	0.95-1.14	0.79 ±0.023
PA	37.43 ±1.69	20.77 ±1.57	26.41 ± 0.105	27.2 ±0.109*	21.10-35.66	0.80 ±0.022
PL	9.97 ±0.61	7.98 ±0.51	8.14 ± 0.024	8.54 ±0.02*	6.51-9.77	0.83 ±0.019
PW	5.69 ±0.13	3.94 ±0.15	4.68 ± 0.01	4.72 ±0.012*	4.00-5.41	0.80 ±0.022
TS	1.86 ±0.15	1.09 ±0.06	1.29 ±0.004	1.38 ±0.004*	1.09-1.65	0.80±0.023
SPP	13.19 ±1.14	11.7 ±1.53	14.24 ±0.064	14.03 ±0.071*	9.77-17.95	0.72 ±0.031
TOC	33.69±0.87	37.15±1.37	35.85±0.064	35.97±0.071	30.83-40.74	0.66±0.036

29 SA: seed area (mm<sup>2</sup>); SL: seed length (mm); SW: seed width (mm); PA: pod area (mm<sup>2</sup>); PL: pod length (mm);  
30 PW: pod width (mm); TSW: thousand seed weight (g); SPP: seed number per pod. TOC: total oil content (%)  
31 # Means ± SE

32 \* = P < 0.001 from *t*-Test showing year to year significance  
33  
34  
35  
36

37 **Table S2.** Pearson's correlation coefficients among seed area (SA), seed length (SL), seed  
38 width (SW), pod area (PA), pod length (PL), pod width (PW), thousand seed weight (TSW)  
39 and seed number per pod (SPP).  
40

Trait	SA	SL	SW	PA	PL	PW	TSW	SPP
SL	0.92***							
SW	0.88***	0.65***						
PA	0.39***	0.43***	0.22**					
PL	0.34***	0.39***	0.16*	0.84***				
PW	0.35***	0.31***	0.30***	0.70***	0.31***			
TSW	0.89***	0.79***	0.82***	0.35***	0.34***	0.30***		
SPP	-0.30***	-0.24**	-0.31***	0.39***	0.31***	0.25**	-0.32**	
TOC	-0.20***	-0.32***	-0.04	-0.14***	-0.16***	-0.07	-0.18***	-0.32

59 \*\*\* Significance level: p<0.001  
60  
61

1  
2  
3  
4 \*\* Significance level:  $p < 0.01$

5 \* Significance level:  $p < 0.1$   
6  
7

8 **Table S3.** Summary of candidate genes associated with seed and/or pod development in the  
9 region of colocalized QTL on chromosome 17.  
10

Candidate gene	Arabidopsis ortholog	Annotation	Biological significance	Reference
Csa17g028400	At1g77850	auxin response factor 17-like	embryo-proper development	Mallory, A. C., et al. (2005).
Csa17g028440	At1g21690	replication factor C subunit 4	endosperm proliferation; embryo maturation	Qian, J., et al. (2018)
Csa17g028530	At1g77540	acetyltransferase -like	cell proliferation, seed filling	Song, X. J., et al. (2015).
Csa17g028640	AT1g35540	putative auxin response factor 14	auxin signaling; embryo development	Mallory, A. C., et al. (2005).
Csa17g029060	AT1g22140	zinc finger CCCH domain-containing protein 15	flowering, Seed development	Wang, D., et al. (2008)
Csa17g029440	AT1g22500	E3 ubiquitin-protein ligase ATL15	seed development, seed size regulator	Aoyama, S., et al. (2017).
Csa17g029450	AT1g22510	E3 ubiquitin-protein ligase RNF170	Seed development, seed size regulator	Aoyama, S., et al. (2017).
Csa17g029550	AT1g22640	transcription factor MYB3	Seed coat development, pigmentation	Wan, L., et al. (2016)
Csa17g029620	AT1g22690	gibberellin-regulated protein 9	seed development	Roxrud, I., et al. (2007)
Csa17g029640	AT1g22710	sucrose transport protein SUC2	early seed development	Wang, L., et al. (2015)
Csa17g030160	AT1g23080	auxin efflux carrier component 7	early embryogenesis	Wang, B., et al. (2018)
Csa17g030300	AT1g23190	probable phosphoglucomutase 1	carbohydrate partitioning; gametophyte development	Egli, B., et al. (2010)
Csa17g030320	AT1g23200	probable pectinesterase inhibitor 6	silique development	Louvet, R., et al. (2006)
Csa17g030330	AT1g17150	exopolysaccharuronase-like	Cell elongation, Early embryo development	Babu, Y., et al. (2013)

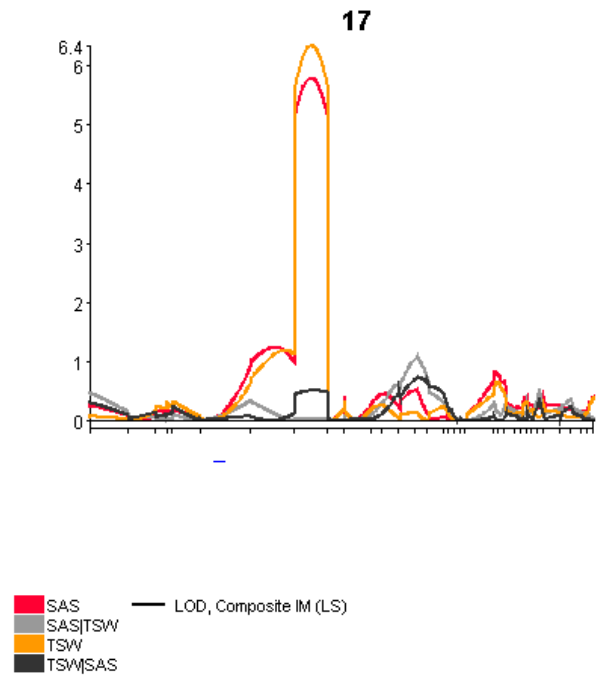
Csa17g030550	AT1g23380	homeobox protein knotted-1-like 6	Embryogenesis, lateral organ development	Hay, A. and M. Tsiantis (2010)
Csa17g030560	AT1g23380	homeobox protein knotted-1-like 6	Embryogenesis, lateral organ development	Hay, A. and M. Tsiantis (2010)
Csa17g030620	AT1g23460	probable polygalacturonase	cell elongation, suspensor development	Babu, Y., et al. (2013)
Csa17g030630	AT1g23480	cellulose synthase like A3	seed coat development, mucilage attachment	Goubet, F., et al. (2009)
Csa17g030920	AT1g23800	aldehyde dehydrogenase family	seed maturation and viability	Shin, J. H., et al. (2009)
Csa17g030930	AT1g23800	aldehyde dehydrogenase family	seed maturation and viability	Shin, J. H., et al. (2009)
Csa17g030940	AT1g23820	spermidine synthase 1	embryo development; seed morphology	Tao, Y., et al. (2018)

## References

- Aoyama, S., et al. (2017). Membrane-localized ubiquitin ligase ATL15 functions in sugar-responsive growth regulation in Arabidopsis. *Biochem Biophys Res Commun* 491: 33-39.
- Babu, Y., et al. (2013). Suspensor length determines developmental progression of the embryo in Arabidopsis. *Plant Physiol* 162: 1448-1458.
- Egli, B., et al. (2010). Loss of cytosolic phosphoglucomutase compromises gametophyte development in Arabidopsis. *Plant Physiol* 154: 1659-1671.
- Goubet, F., et al. (2009). Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant J* 60: 527-538.
- Hay, A. and M. Tsiantis (2010). KNOX genes: versatile regulators of plant development and diversity. *Development* 137: 3153-3165.
- Louvet, R., et al. (2006). Comprehensive expression profiling of the pectin methylesterase gene family during silique development in Arabidopsis thaliana. *Planta* 224: 782-791.
- Mallory, A. C., et al. (2005). MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* 17: 1360-1375.
- Okushima, Y., et al. (2005). Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17: 444-463.
- Qian, J., et al. (2018). Arabidopsis replication factor C4 is critical for DNA replication during the mitotic cell cycle. *Plant J* 94: 288-303.
- Roxrud, I., et al. (2007). GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. *Plant Cell Physiol* 48: 471-483.
- Shin, J. H., et al. (2009). Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol* 149: 905-915.
- Song, X. J., et al. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proc Natl Acad Sci U S A* 112: 76-81.



1  
2  
3  
4 Tao, Y., et al. (2018). The Spermine Synthase OsSPMS1 Regulates Seed Germination, Grain  
5 Size, and Yield. *Plant Physiol* 178: 1522-1536.  
6 Wang, B., et al. (2018). The RAF-like mitogen-activated protein kinase kinases RAF22  
7 and RAF28 are required for the regulation of embryogenesis in Arabidopsis. *Plant J* 96: 734-  
8 747.  
9 Wang, D., et al. (2008). Genome-wide analysis of CCCH zinc finger family in Arabidopsis and  
10 rice. *BMC Genomics* 9: 44.  
11 Wan, L., et al. (2016). Transcriptome analysis of a new peanut seed coat mutant for the  
12 physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Front*  
13 *Plant Sci* 7: 1491.  
14 Wang, L., et al. (2015). Enhanced Sucrose Loading Improves Rice Yield by Increasing Grain  
15 Size. *Plant Physiol* 169: 2848-2862  
16  
17  
18  
19  
20  
21  
22



**Fig. S1.** Conditional analysis of SA and TSW at the major QTL region on chromosome 17. SA/TSW, SA Conditioned by TSW; TSW/SA, TSW Conditioned by SA.