Hortic. Environ. Biotechnol. 58(2):187-194. 2017.
DOI 10.1007/s13580-017-0264-4

ISSN (print) : 2211-3452
ISSN (online) : 2211-3460

Research Report

Characterization of Twenty Camelina spp. Accessions Using Single Nucleotide Polymorphism Genotyping

Changsoo Kim[1†], Jeong Hwan Lee[4,5†], Yong Suk Chung[1], Sang Chul Choi[1], Hui Guo[2], Tae-Ho Lee[3], and Sanghyeob Lee[4*]

[1]Department of Crop Science, Chungnam National University, College of Agriculture and Life Science, Daejeon 34134, Republic of Korea
[2]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA
[3]Genomics Division, National Academy of Agricultural Science, Rural Development Administration, Jeonju 54875, Republic of Korea
[4]Department of Bioindustry and Bioresource Engineering, Sejong University, Seoul 05006, Republic of Korea
[5]Department of Life Sciences, Chonbuk National University, 567 Baekje-daero, deokjin-gu, Jeonju, Jeollabuk-do 54896, Republic of Korea

*Corresponding author: sanglee@sejong.ac.kr
†These authors contributed equally to this work.

Received October 18, 2016 / Revised October 18, 2016 / Accepted November 3, 2016

Abstract. Sequencing the complete genome of *Camelina sativa* will facilitate studies to improve this oilseed crop. We analysed 20 accessions of *Camelina* spp. using genotyping-by-sequence technology. After stringent screening, 35,783 single nucleotide polymorphisms (SNPs) were generated, and basic genetic studies were performed to check the diversity of these SNPs. STRUCTURE and phylogenetic analyses revealed five subgroups in these 20 *Camelina* accessions. Winter-types may have diverged from summer-types. Some genomic regions were negatively selected, and most of these were gene-rich regions. As expected, the most ancient subgroup was less affected by negative selection. Marker-trait associations with plant height, leaf length, and pod size generated 154 SNPs, and 72 adjacent genes were significantly associated with these phenotypes. Further large-scale analysis and gene expression studies with these SNPs and genes are needed to develop potentially valuable resources for improving *C. sativa*.

Additional key words: genotype-by-sequencing, marker-trait association, genetic variation

Introduction

Camelina sativa (L.) Crantz (false flax) is an annual oilseed crop, cultivated in Europe since the Bronze Age (Gehringer et al., 2006). Cultivation in Europe and some parts of North America has gradually decreased since the middle of the 20th century, and has been replaced by more productive species such as oilseed rape and sunflower. However, its short growing period (100-120 days) (Agegnehu and Honermeier, 1997) and flexible seeding times (spring or winter types) mean the species can be used in mixed cropping systems in Europe. It is highly regarded as a food, forage, and biofuel crop.

C. sativa belongs to the mustard family (Brassicaceae); 11 wild relatives of *C. sativa* are still found in Eurasia (Akeroyd, 1993), and *C. sativa, C. rumelica, C. microcarpa, C. hispida,* and *C. alyssum* are widely distributed. Chromosome counts include 2n = 12 for *C. rumelica* (Brooks, 1985); 2n = 14 for *C. hispida* (Maassoumi, 1980); and 2n = 40 for *C. sativa, C. microcarpa,* and *C. alyssum* (Francis and Warwick, 2009). Found in three species, 2n = 40 is the most common chromosome number, but regional differences in chromosomal counts, including different ploidy numbers or aneuploidy, indicate that natural variation among populations is prevalent. Gehringer et al. (2006) reported disomic inheritance and multiple amplifications of *Brassica*-derived single sequence repeat markers in *C. sativa*, suggesting that it could be an allopolyploid. More recently, Hutcheon et al. (2010) reported that two important genes, *FATTY ACID DESATURASE 2* and *FATTY ACID ELONGASE 1*, have three copies in *C. sativa* but a single copy in *Arabidopsis*, suggesting that the *C. sativa* genome might be allohexaploid. The genome of a homozygous doubled-haploid *C. sativa* line with 20 pseudomolecules was recently published

(Kagale et al., 2014), the phylogeny of which indicates its likely allopolyploid origin.

Understanding genetic variation in *C. sativa* is essential to improve its economically valuable traits. Important agronomic traits have been identified using amplified fragment length polymorphism (AFLP) markers (Gehringer et al., 2006). Using AFLP fingerprints of 53 *C. sativa* accessions, Ghamkhar et al. (2010) showed that geographical origin is strongly associated with genetic variation and fatty acid content. Comparative genomics studies on the sequenced transcriptomes of seeds (Mudalkar et al., 2014; Nguyen et al., 2013) and leaf tissue (Liang et al., 2013) have resulted in the identification of numerous genes related to fatty acid metabolism.

In this study, we explored the complete genome of *C. sativa* using genotyping-by-sequencing (GBS) technology to investigate the genetic relationships and phenotypic variation in 20 *Camelina* spp., mainly *C. sativa* accessions collected from multiple locations in Europe. The results of this study will facilitate future genomics-assisted breeding in this oilseed crop.

Materials and Methods

Camelina Accessions and Phenotyping

Nineteen accessions of *Camelina* spp. were obtained from the United States Department of Agriculture Germplasm Resources Information Network (USDA GRIN). Accessions were either spring or winter types, with wide phenotypic variation. Table 1 summarizes the profiles of these 20 *Camelina* accessions. Seeds were germinated in August 2011, on a seedbed with daily irrigation. Ten-day-old seedlings were transplanted in a random block design at a spacing of 75 cm × 60 cm, with three replications. Observation of basic phenotypes, namely plant height, leaf length, and pod size, were recorded using ten randomly selected plants of each accession at different growth stages (Table 1). Plant height 1 and leaf length 1 were measured when 5 or 6 leaves emerged, and plant height 2 and leaf length 2 were measured four weeks after the first measurement. Pod sizes were measured three

Table 1. Twenty *Camelina* spp. accessions and their phenotypes

Accession (ID)	Scientific name	Origin[#]	Height1[$] (cm)	Height2[$]	Leaf1[$]	Leaf2[$]	Pod1[$]	Pod2[$]	Pod3[$]
CAM93* (sa006)	*C. sativa*	Unknown	NA**	NA	NA	NA	NA	NA	NA
CAM268 (sa012)	*C. sativa*	BGR	8.10	56.33	4.84	9.67	8.86	8.89	9.26
CAM200 (sa014)	*C. sativa* subsp. *sativa*	DEU	6.40	57.00	4.36	9.00	7.54	7.29	7.62
CAM208 (sa016)	*C. sativa* subsp. *sativa*	DEU	6.70	68.67	4.30	8.33	7.22	7.59	8.93
CAM165 (sa025)	*C. sativa* subsp. *sativa*	DEU	5.10	65.67	4.10	11.00	7.49	7.37	7.44
CAM29 (sa027)	*C. sativa* subsp. *sativa*	UKR	6.80	46.00	3.86	9.00	7.87	7.97	7.97
CAM273 (sa032)	*C. sativa* subsp. pilosa	SWE	5.50	54.00	3.90	7.33	6.73	7.25	7.04
CAM160 (sa067)	*C. sativa* subsp. *sativa*	DEU	5.80	73.00	4.20	11.33	7.87	7.41	7.20
CAM96 (sa068)	*C. sativa* subsp. *sativa*	DEU	6.76	76.00	6.00	12.00	7.06	7.46	7.40
CAM232 (sa076)	*C. sativa* subsp. *sativa*	DEU	4.60	74.00	6.00	12.00	8.70	8.33	7.79
CAM255 (sa091)	*C. sativa* subsp. *pilosa*	DEU	4.75	85.33	6.00	10.33	7.44	7.84	7.22
CAM240** (sa101)	*C. sativa* subsp. *sativa*	DEU	6.50	84.00	7.00	11.00	8.53	7.63	NA
CAM21 (sa102)	*C. alyssum*	DEU	6.00	81.00	7.10	11.33	7.50	7.17	7.98
CAM81 (sa106)	*C. sativa* subsp. *sativa*	DEU	4.00	52.33	6.00	8.33	9.15	8.91	8.61
CAM132* (sa135)	*C. sativa* subsp. *pilosa*	Unknown	1.00	4.00	6.30	9.33	NA	NA	NA
CAM110 (sa136)	*C. sativa* subsp. *sativa*	POL	9.60	80.67	5.70	9.67	6.80	6.64	6.84
CAM100* (sa178)	*C. sativa* subsp. *pilosa*	Unknown	1.00	4.67	6.40	6.00	NA	NA	NA
sa201***	*C. sativa* polkie 99 25	POL	10.4	40.00	4.80	3.00	6.68	7.01	7.35
sa205***	*C. sativa* svnkor 2006-44	SVN	7.80	47.33	4.60	4.33	8.01	7.64	7.70
sa227***	*C. sativa* bgr 150	BGR	9.20	50.67	5.20	5.00	8.86	7.99	6.66

The times of phenotyping described in the materials and methods.
All measurements shown are in centimeters (cm).
*Winter-type accession.
**Phenotypes not available because the plant died (CAM240), did not germinate (CAM93)showed .
***Variety names are shown for plant materials for which accession numbers were not available.
[#] BGR, Bulgaria; DEU, Germany; UKR, Ukraine; SWE, Sweden; POL, Poland; SVN, Slovenija.
[$] Height, Height of plant; Leaf, length of leaf; Pod, length of pod.

times every 20 days from the emergence of pods, and were scored as pod size 1, pod size 2, and pod size 3. As expected, winter-type accessions (sa006, sa135, and sa178) did not flower during the experiment because of the seeding season, thus pod sizes could not be estimated. Mean values of the observed data for each accession in each replication were calculated and used in this study. Statistical analysis was done using XLSTAT-Pro 7.5 software (XLStat, NY, USA).

Reduced Representation Library Construction and Single Nucleotide Polymorphism Identification

A reduced representation library (RRL) was constructed using the protocol described by Poland et al. (2012) with slight modifications. First, genomic DNA samples from seeds used for phenotypic analysis were double-digested using ApeKI (GCWGC, where W = A or T) and MspI (CCGG) restriction enzymes to reduce genome complexity. Forward (P7) adapters with 20 barcodes were matched to ApeKI restriction sites, whereas reverse (P5) adapters were matched to MspI overhangs. Restriction fragments with adapters, which were pooled via polymerase chain reaction (PCR) amplification, were attached to polymerase chain reaction (). The library was sequenced on a HiSeq 2000 (Illumina) using 100-base pair (bp) paired-end reactions. Sequences were parsed with in-house scripts according to the 20 barcode sequences, and were further aligned to the recently published reference genome of C. sativa (Kagale et al., 2014) using the bwasw function of the Burrows-Wheeler Aligner (Li and Durbin, 2009). SNPs were identified and genotyped using a reference-based SNP discovery pipeline included in TASSEL (Bradbury et al., 2007). Genotype data were again filtered with a minor allele frequency (MAF) of greater than 0.2, and no missing values were allowed in the 20 accessions. In all, 35 783 filtered SNPs were used for further genetic analysis. Our data were deposited in the Genbank/NCBI database with reference numbers as follows: SAMN05898476 (Biosample), PRJNA347894 (Bioproject), and SRR4427937 (Sequence Read Archive).

Genetic Data Analysis

The population structure of *Camelina* spp. was analyzed using STRUCTURE software (version 2.3.4) (Pritchard et al., 2000). The lengths of burn-in period and Markov Chain Monte Carlo replications were set to 50 000 and 100 000, respectively. K values (the number of populations) were pre-set from 2 to 10 with five iterations each. To determine the optimal number of subpopulations, STRUCTURE results were subjected to an *ad hoc* statistic delta K, the rate of change in the log probability of data between successive K values (Evanno et al., 2005). Principal component analysis (PCA) and phylogenetic analysis (maximum parsimony) were performed using TASSEL (Bradbury et al., 2007) and SNPhylo (Lee et al., 2014), respectively, to support the results from STRUCTURE. Linkage disequilibrium (LD) was estimated using TASSEL (Bradbury et al., 2007) for each chromosome with five subpopulations each. LD decay was evaluated based on the r^2 value of the pairwise distance betweeen two SNPs. For each subpopulation and accessions, nucleotide diversity and Tajima's D were estimated using DnaSP version 5.10.01 (Librado and Rozas, 2009).

Preliminary Association Analysis

Although 20 accessions are certainly insufficient for an adequate association study, marker-trait associations (MTAs) were tentatively analyzed to obtain a list of candidate SNP markers that might be evaluated in future, if and when additional genetic information is accumulated. A General linear models result in many spurious associations since they only take into account a fixed effect based on admixture percentages (Q matrix). Thus, using TASSEL version 4.3 (Bradbury et al., 2007) and STRUCTURE version 2.3.4 (Pritchard et al., 2000), two different association mapping models were run for each trait: a mixed linear model with a kinship (K) matrix as a random effect, and with a Q matrix as a fixed effect. MTAs with $p < 10^{-4}$ were considered significant. Manhattan and QQ plots were generated using SNPEVG version 3.2 (Wang et al., 2012).

Results

Phenotypic variation of *Camelina* accessions

Based on a preliminary phenotype screen, remarkable variety was observed among our 20 *Camelina* spp. accessions (Supplementary Fig. 1). Table 1 lists the accessions, their origins, and traits used for this study. Plants belonging to *Camelina* spp. have two different growth types: spring and winter types. In this study, we included three winter-type *C. sativa* accessions (sa006, sa135, and sa178). Since we planted these in August, they did not flower (hence, no pod size measurements were obtained). In particular, sa006 did not germinate at all; thus, this accession was excluded from the preliminary association study.

All 20 accessions were subjected to Tukey's multiple comparison for each phenotype; all phenotypes were statistically different from each other ($p < 0.0001$). In addition, the Pearson's correlation test (Table 2) showed that height 1 was negatively correlated with leaf 2, whereas height 2 was positively correlated with leaves 1 and 2 ($p < 0.05$). In other words, early stem elongation was negatively correlated with leaf expansion, whereas late elongation was positively correlated with leaf expansion. Pod size was not significantly statistically related to plant height or leaf length.

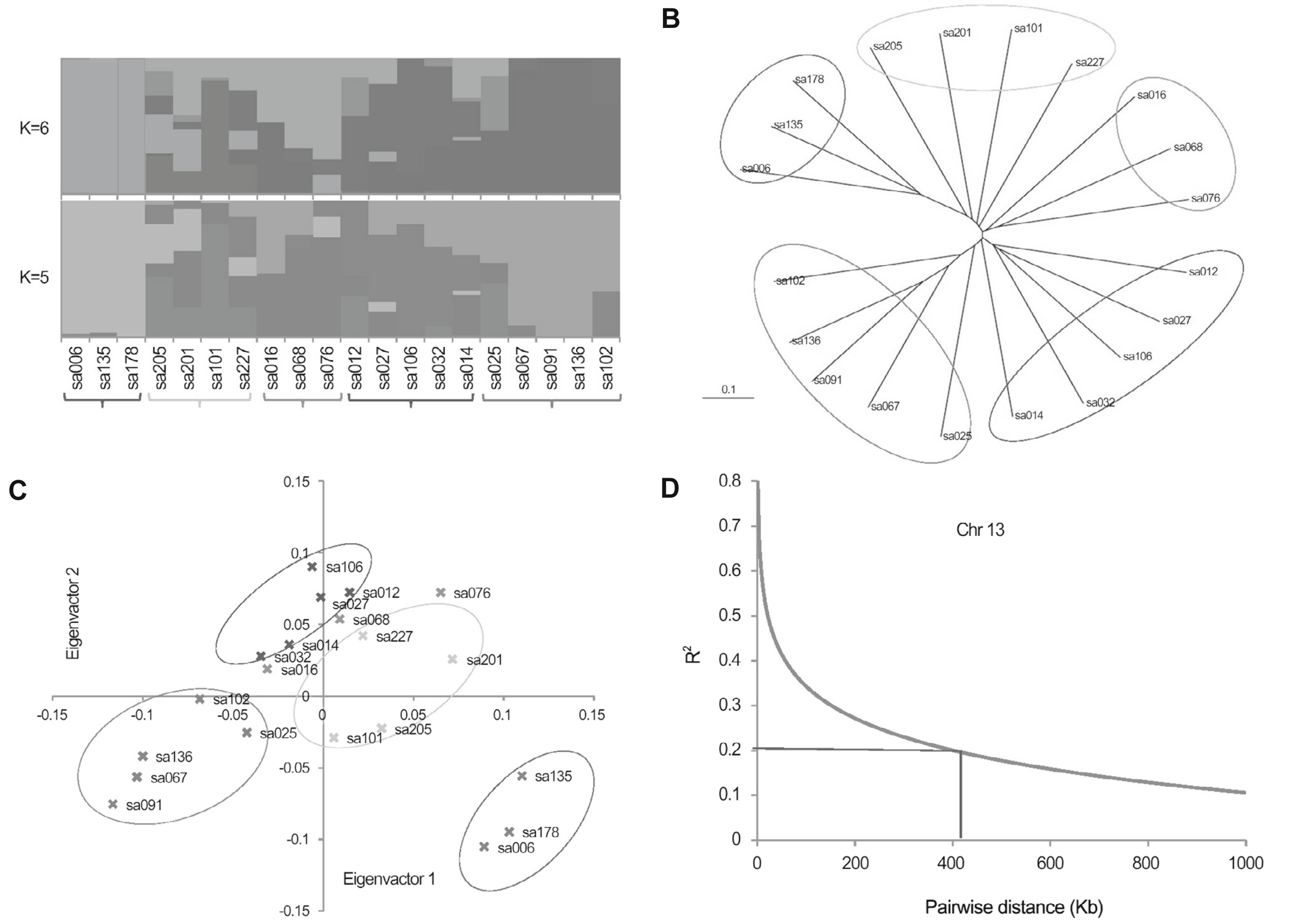

Fig. 1. STRUCTURE (A), principal component (B), phylogenetic (C), and linkage disequilibrium (D) analyses of 20 *Camelina* accessions. Brackets (STRUCTURE) and circles (principal component and phylogenetic analyses) in the same color indicate the same subgroups.

Table 2. Pearson's correlation of each phenotypic value

Trait	Height1	Height2	Leaf1	Leaf2	Pod1	Pod2	Pod3
Height1	1.000	−0.358	−0.129	**−0.602**	−0.247	−0.358	−0.216
Height2		1.000	**0.552**	**0.782**	−0.213	−0.234	−0.126
Leaf1			1.000	0.327	0.170	0.108	0.007
Leaf2				1.000	0.044	0.081	0.136
Pod1					1.000	**0.875**	0.397
Pod2						1.000	**0.611**
Pod3							1.000

Values calculated to be significant at α = 0.05 (two-tailed test) are marked in bold.

Population structure of *Camelina* accessions

The RRL sequenced using a single lane of the Illumina HiSeq 2000 generated about 384 million reads at 100 cycles. TASSEL-GBS software (Glaubitz et al., 2014) uses the first 64 bp (out of 100 bp) for SNP discovery and genotyping, securing high-quality sequences to avoid potential sequencing errors. Despite the hexaploid nature of *C. sativa*, a recently released reference genome (Kagale et al., 2014) allows easier detection of SNPs with accurate locus information. However, so as to analyze only those SNPs with high credibility, SNPs from any of the 20 accessions with any missing values, and which had a MAF of 0.2 or lower, were excluded. Thus, a total of 35 783 SNPs were used for diversity and association analyses (Table 3).

Filtered SNPs were used to estimate LD (Fig. 1D) in each chromosome; LD decay (the point at which R2 < 0.2) was at about 420 kb, almost twice as high as other closely related species such as Arabidopsis thaliana (in which LD = 250 kb)

Table 3. Genetic information on the subgroups within 20 *Camelina* accessions

Groups	Number of SNPs	Heterozygosity	Diversity (π)	Tajima's D*
All 20 accessions	35 783	0.451	0.438	3.248
Subgroup I CAM93, 132, 100 (sa006, 135, 178)	19 899	0.704	0.563	1.869
Subgroup II svnkor, polkie, CAM240, bgr (sa205, 201, 101, 227)	28 460	0.582	0.509	1.759
Subgroup III CAM208, 96, 232 (sa016, 068, 076)	21 746	0.535	0.554	1.729
Subgroup IV CAM268, 29, 81, 273, 200 (sa012, 027, 106, 032, 014)	27 440	0.527	0.470	1.663
Subgroup V CAM165, 160, 255, 110, 21 (sa025, 067, 091, 136, 102)	28 426	0.574	0.487	1.897

*Tajima's D significant at p = 0.05.

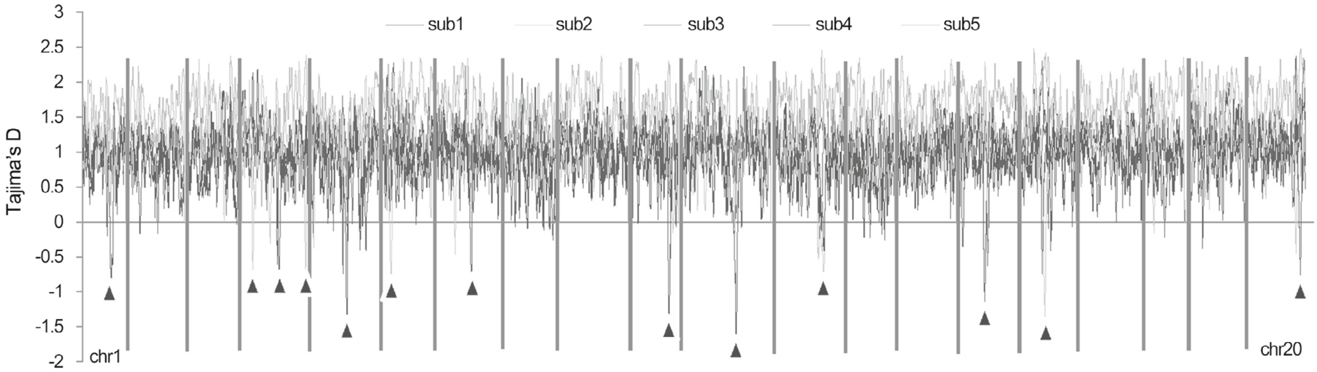

Fig. 2. Genome-wide distribution of Tajima's D values. Window and step sizes are 50 and 10, respectively. Different chromosomes are partitioned by gray bars. Red triangles indicate genomic regions with negative values (D < −0.5).

(Nordborg et al., 2002). Using STRUCTURE, phylogenetic analysis and PCA (Fig. 1A, 1B, and 1C, respectively), five subpopulations were confirmed. In STRUCTURE analysis, an optimal K value obtained using Evanno's test indicated five substructures, consistent with the results of PCA and phylogenetic analyses. Table 3 shows the numbers of SNPs and diversity statistics of each subgroup.

Although the 20 accessions were sampled from a wide European geographical range (Table 1), the subpopulation structure did not reflect their geographical distances. Genome-wide Tajima's D values are shown in Fig. 2. Although most genomic regions showed positive values, 13 genomic regions (except subgroup V) showed negative Tajima's D values (D < -0.5) (Table 4). A list of locations, features, and annotations involved in those regions is given in Supplementary Data 1.

MTA Analysis

Although a study of 20 genotypes is insufficient for reliable association analysis, we conducted preliminary association analysis to test the power of the SNPs identified, and to lay the foundation for larger-scale analysis in the near future. We initially phenotyped plant height, leaf length, and pod size at different stages of growth. A total of 154 SNPs and 72 adjacent genes were identified from this preliminary study. All associated SNPs are shown in Fig. 3, and listed in Supplementary Data 2 with the adjacent genes, where available. For height 1 and leaf 1, 36 and 6 SNPs, respectively, were associated ($p < 10^{-4}$), and 17 genes located near to those SNPs were also identified. Five and 59 SNPs were significantly associated with height 2 and leaf 2, respectively, and 33 genes were located close to one another. Based on Pearson's correlation at p = 0.05, height 1 was negatively correlated with leaf 2 (Table 2). Our study revealed 10 SNPs that were commonly associated with these two phenotypes (Table S1). Despite no phenotypic correlation between heights 1 and 2, all associated SNPs in height 2 were also included in height 1 (Table S1). However, the SNPs (excluding the five overlapped SNPs in heights 1 and

Table 4. Genomic regions with negative Tajima's D values (D < −0.5) and their genomic components

Chromosomes	Physical regions (bp)	Subgroups	Interval (kb)	Gene	CDS	Exon	mRNA	5′ UTR	3′ UTR	GD
1	14889453-16002129	IV	1112	208	619	648	113	91	90	5
4	4465687-4481533	II	16	0	0	0	0	0	0	0
	15556156-16147365	I, III, IV	591	182	427	446	94	59	55	3
	29041661-29127726	II	86	34	118	122	17	15	12	3
5	16822553-18925100	III, IV	2102	332	955	1001	175	128	108	6
6	4187141-4545861	II	359	64	139	146	33	24	21	6
7	17371116-17430899	IV	60	24	95	117	13	8	7	3
10	21198074-21532954	I, IV	335	54	114	120	29	20	16	6
11	32857603-33274447	IV	417	62	193	199	35	27	29	7
12	24286706-25416951	II	1130	88	269	280	48	39	39	13
15	12686841-13962582	IV	1276	262	530	556	136	102	88	5
16	12089188-12767846	I, II	679	102	183	200	53	42	32	7
20	28188875-28289474	IV	101	18	42	42	9	3	4	6

Abbreviations: CDS, coding sequence; GD, gene density (one gene per χ kb); UTR, untranslated region.

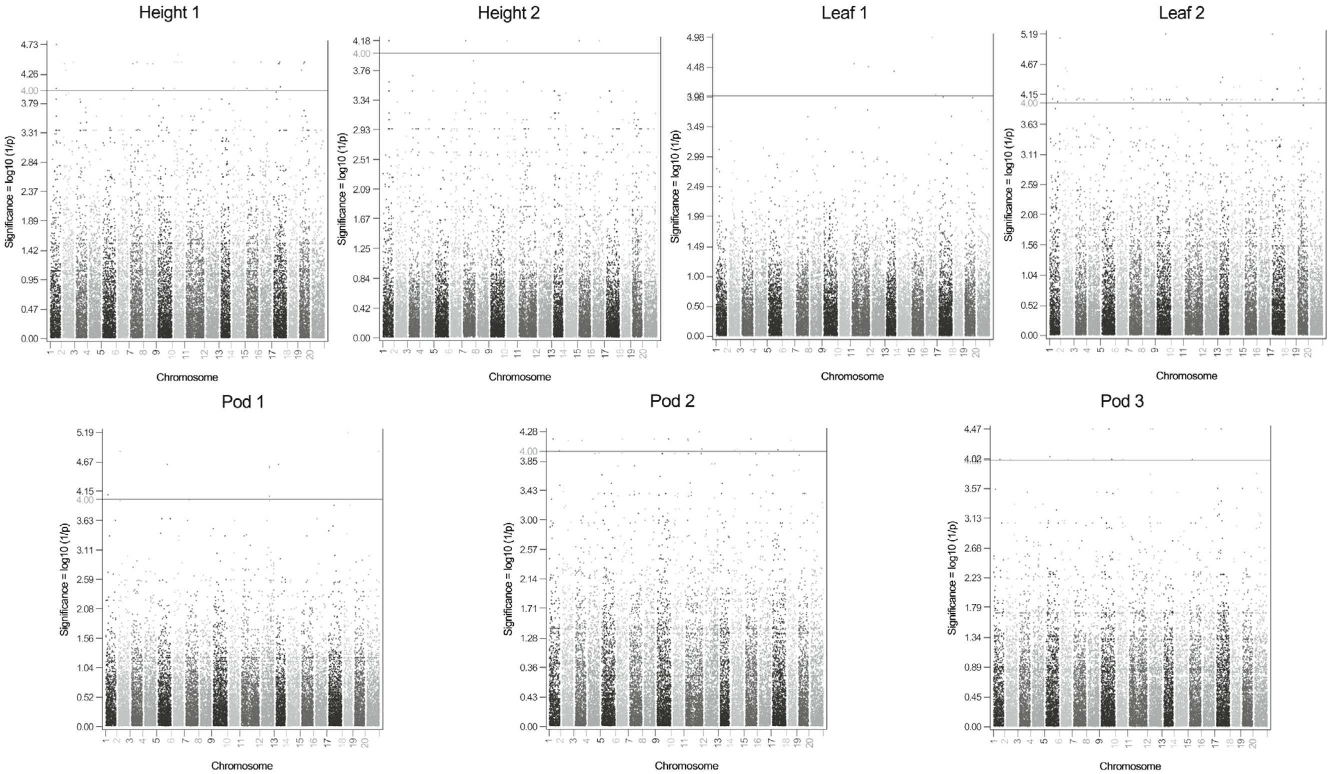

Fig. 3. Manhattan plots of associated single nucleotide polymorphisms and tested phenotypes. Green bars indicate cut-off thresholds of P-values [log10 (1/p) = 4.00].

2) associated with height 1 were not represented in height 2 (the first 26 SNPs out of 31 in height 1; Supplementary Data 2). For pod sizes 1, 2, and 3, 17, 17, and 14 SNPs were significantly associated, respectively; however, no overlapping loci were found between these three different measurements despite their strong phenotypic correlations (Table 2). The three most significant SNPs with adjacent genes in each phenotypic category are listed in Table S2. The SNPs and genes listed in Table S2 are a primary target for further analysis; however, all candidate SNPs must be re-validated by large-scale analysis. Adjacent genes might be of interest for gene expression studies in different organs.

Discussion

C. sativa is an oil crop that can be utilized both as biodiesel and food. In particular, its high unsaturated fatty acid content increases its value for the production of healthy oil. Despite its small size, the structure of the C. sativa genome might have been formed by multiple allopolyploidisation events, complicating genetic or genomic studies. In fact, genotyping of SNPs using sequence data is difficult because true nucleotide variations might remain undetected without a well-established reference genome. Without a reference genome, SNPs derived from different parents or accessions are often confused with those from subgenomes (hemi-SNPs), thus compromising genotyping data sets with false-positives. We conducted accurate SNP genotyping by aligned SNPs to the recently released reference genome (Kim et al., 2015; Kagale et al., 2014). At the genomic level, use of the reference genome might facilitate C. sativa research and contribute to the development of new and/or improved approaches in breeding.

Using the SNP markers, we divided the 20 Camelina accessions into five subgroups. Little remarkable phenotypic variation was found between these subgroups, except for subgroup I (winter types), and that the plant height of subgroup II was notably shorter than that of subgroup V (mostly wild accessions). Although the sampled Camelina spp. show a high level of both winter and summer type accessions showed relatively high heterozygosity, suggesting that the winter types might have further diverged from summer-type accessions. Our results showed a high level of nucleotide diversity (π) and positive Tajima's D values across the subgroups, implying that Camelina might have experienced a population bottleneck or over-dominant selection. However, considering Tajima's D across the genome (Fig. 2), 13 genomic regions that were absent in all subgroup V accessions, had negative Tajima's D values ($D < -0.5$), indicating that these regions might have undergone purifying selection. (Larsson et al., 2013). Some genomic regions may have been affected by selective sweep because of purifying selection during domestication. A C-value paradox means that gene density is often scarce in plants with large genome sizes. Previous studies have shown the average gene density to be 1/4.5 kb in Arabidopsis, 1/20 kb in rice, and 1/30 kb in sorghum (Keller and Feuillet, 2000; Lin et al., 1999). Hutcheon et al. found the genome size of C. sativa to 750 Mb, similar to that of sorghum, therefore the gene densities (last column in Table 4) in the genomic region with negative Tajima's D values were thought to be gene-rich regions, In addition, the LD of C. sativa (about 420 kb) may be have been caused by its high level of self-compatibility, implying that the genome was under high selective pressure during its evolution.

Although this MTA was small-scale, the candidate SNPs generated are potentially very useful in future breeding efforts.; Tthe genomic regions covered by SNPs (Table S1) are particularly worthy of further functional investigation. For example, 10 SNPs covering five chromosomal regions were commonly associated with height 1 and leaf 2 (Table S1), but no SNPs overlapped between leaf 1 and any of the traits. These five chromosomal regions might function differently during stem and leaf elongation, perhaps because of the differential expression of candidate genes or source-sink changes during different stages of growth. Conversely, five SNPs found in five chromosomal regions were significantly associated with heights 1 and 2 (Table S1), indicating that these regions might play a key role in stem elongation in *Camelina* spp. However, 31 SNPs associated with height 1 were not associated with height 2 (Supplementary Data 2), which could have limited affects at the early stage of stem elongation. In fact, height 2 was evaluated when vegetative growth was almost completed; therefore, the possibility that these 31 SNPs might be actively involved in stem elongation cannot be excluded. All candidate SNPs must be re-evaluated and confirmed with additional accessions in the near future.

Few resources for comparative studies are currently available for *C. sativa*. However, comparing these SNPs with previously published quantitative trait locus resources (Gehringer et al., 2006) might provide new insights into the key markers that are tightly linked to agriculturally important traits. In future work we will analyze more accessions, as well as key phenotypic data such as yield, oil content, and disease resistance. These data sets will provide additional valuable insights for *C. sativa* breeding. Productive alternatives such as oilseed rape and sunflower mean that *C. sativa* has been overlooked as an oil seed crop; however since our final genome assembly covers 82% of the estimated genome size (Kagale et al., 2014) it is hopeful that our results might have a significant impact on the research to improve *C. sativa* as a producer of edible oil and biodiesel. Furthermore, our data might form the basis for improved molecular breeding efforts, and expand our current knowledge of *C. sativa*.

Acknowledgements: This work was supported by the Bioindustry Technology Development Program (grant number 111057-5, 312033-5) of iPET (Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry) to SL, and Chungnam National University grant number 2015-1730-01 to CK.

Literature Cited

Agegnehu M, Honermeier B (1997) Effects of seeding rates and nitrogen fertilization on seed yield, seed quality and yield components of false flax (*Camelina sativa* Crtz). Bodenkultur 48:15-21

Akeroyd JR (1993) Camelina Crantz. Cambridge University Press, Cambridge

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Brooks RE (1985) Chromosome number reports. LXXXVII. Taxon 34:346-351

Evanno G, S. R, J. G (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620

Francis A, Warwick SI (2009) The biology of Canadian weeds. 142. *Camelina alyssum* (Mill.) Thell.; *C. microcarpa* Andrz. ex DC.; *C. sativa* (L.) Crantz. Can J Plant Sci 89:791-810

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

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

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346

Hutcheon C, Ditt RF, Beilstein M, Comai L, Schroeder J, Goldstein E, Shewmaker CK, Nguyen T, De Rocher J, et al. (2010) Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. BMC Plant Biol 10:233

Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, et al. (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat Commun 5:3706

Keller B, Feuillet C (2000) Colinearity and gene density in grass genomes. Trends Plant Sci 5:246-251

Kim B, Kim N, Kang J, Choi JY, Sim S-C, Min SR, Park Y (2015) Single Nucleotide Polymorphisms linked to the SlMYB12 Gene that Controls Fruit Peel Color in Domesticated Tomatoes (*Solanum lycopersicum* L.). Korean J Hortic Sci Technol 33:566-574

Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. PLoS Genet 9:e1003246

Lee TH, Guo H, Wang X, Kim C, Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15:162

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

Liang C, Liu X, Yiu SM, Lim BL (2013) De novo assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. BMC Genomics 14:146

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, et al. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-768

Maassoumi A (1980) Cruciferes de la flore d'Iran: etude caryosystematique, Strasbourg, France.

Mudalkar S, Golla R, Ghatty S, Reddy AR (2014) De novo transcriptome analysis of an imminent biofuel crop, *Camelina sativa* L. using Illumina GAIIX sequencing platform and identification of SSR markers. Plant Mol Biol 84:159-171

Nguyen HT, Silva JE, Podicheti R, Macrander J, Yang W, Nazarenus TJ, Nam JW, Jaworski JG, Lu C, et al. (2013) Camelina seed transcriptome: a tool for meal and oil improvement and translational research. Plant Biotechnol J 11:759-769

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, et al. (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190-193

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Wang S, Dvorkin D, Da Y (2012) SNPEVG: a graphical tool for GWAS graphing with mouse clicks. BMC Bioinformatics 13:319