frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fpls.2022.949470

OPEN ACCESS

EDITED BY
Chuanlei Zhang,
Tianjin University of Science and Technology, China

REVIEWED BY
Stephan Reinert,
University of Erlangen Nuremberg, Germany
Fazal Rehman,
South China Botanical Garden (CAS), China

*CORRESPONDENCE
Kousuke Seki
seki-kosuke@pref.nagano.lg.jp

SPECIALTY SECTION
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

RECEIVED 21 May 2022
ACCEPTED 26 September 2022
PUBLISHED 13 October 2022

CITATION
Seki K and Toda Y (2022) QTL mapping for seed morphology using the instance segmentation neural network in *Lactuca* spp. *Front. Plant Sci.* 13:949470. doi: 10.3389/fpls.2022.949470

QTL mapping for seed morphology using the instance segmentation neural network in *Lactuca* spp

Kousuke Seki[1]* and Yosuke Toda[2,3,4]

[1]Nagano Vegetable and Ornamental Crops Experiment Station, Shiojiri, Japan, [2]Phytometrics Co., Ltd., Shizuoka, Japan, [3]Bioscience and Biotechnology Center, Nagoya University, Nagoya, Japan, [4]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya, Japan

Wild species of lettuce (*Lactuca* sp.) are thought to have first been domesticated for oilseed contents to provide seed oil for human consumption. Although seed morphology is an important trait contributing to oilseed in lettuce, the underlying genetic mechanisms remain elusive. Since lettuce seeds are small, a manual phenotypic determination required for a genetic dissection of such traits is challenging. In this study, we built and applied an instance segmentation-based seed morphology quantification pipeline to measure traits in seeds generated from a cross between the domesticated oilseed type cultivar 'Oilseed' and the wild species 'UenoyamaMaruba' in an automated manner. Quantitative trait locus (QTL) mapping following ddRAD-seq revealed 11 QTLs linked to 7 seed traits (area, width, length, length-to-width ratio, eccentricity, perimeter length, and circularity). Remarkably, the three QTLs with the highest LOD scores, *qLWR-3.1*, *qECC-3.1*, and *qCIR-3.1*, for length-to-width ratio, eccentricity, and circularity, respectively, mapped to linkage group 3 (LG3) around 161.5 to 214.6 Mb, a region previously reported to be associated with domestication traits from wild species. These results suggest that the oilseed cultivar harbors genes acquired during domestication to control seed shape in this genomic region. This study also provides genetic evidence that domestication arose, at least in part, by selection for the oilseed type from wild species and demonstrates the effectiveness of image-based phenotyping to accelerate discoveries of the genetic basis for small morphological features such as seed size and shape.

KEYWORDS

deep learning, neural network, ddRAD-seq, QTL mapping, seed morphology, lettuce domestication

Introduction

Lettuce (*Lactuca sativa* L.) cultivars can be classified into several types based on differences in plant morphology, such as crisphead, butterhead, romaine, leaf, latin, stem, and oilseed (Ryder, 1999). Seed oil is also abundant in the seeds of wild lettuce species (Ramadan, 1976). The oilseed type of lettuce was proposed to have been domesticated from wild species in ancient Egypt (Wei et al., 2021) and is considered to be the earliest domesticated type. In support of this notion, there are indications of the use of oilseed type lettuce from as early as ca. 2500 B.C. in Egypt (De Vries, 1997). Lettuce cultivars from the oilseed type progress through the rosette stage very rapidly and bolt and flower early (Ryder, 1999). However, the leaves are bitter and are thus not eaten as salad (De Vries, 1997; Ryder, 1999). The seeds are more round and larger than those of wild species, contain more vitamins soluble in seed oil, and were used for human consumption in Egypt (De Vries, 1997). For oilseed-type cultivars, seed morphology is an important agronomic trait that has been altered over the course of domestication. The genetic basis of seed morphology has been characterized in some oil seed crops, and the 27 quantitative trait loci (QTLs) were reported for seed weight and size in peanut (Zhang et al., 2019). A total of 52 QTLs for 12 seed traits were detected in sunflower (Yue et al., 2009), and 152 QTLs for seven seed-shape traits were detected in oilseed rape (Sun et al., 2018). In general, it is considered that the traits of seed morphology are complicated. Little is known about the genetic basis for seed morphology in lettuce, except for previous research focused on domestication traits based on QTL analysis (Hartman et al., 2013). One of the main limitations comes from the small size of lettuce seeds, making it difficult to evaluate seed morphology traits by hand. In seed crops such as soybean (*Glycine max*), rice (*Oryza sativa*), and wheat (*Triticum aestivum*), seed morphology is also an important trait related to crop yield in which phenotyping has been performed *via* several approaches (Tanabata et al., 2012; Williams et al., 2013; Gao et al., 2019; Kumawat and Xu, 2021). For instance, seed dimensions have been measured manually with calipers or with image-processing software. These methods are labor-intensive and run the risk of bias due to human error. Automation is thus desirable for the precise phenotyping of seed morphology traits. Several software programs have been developed for plant phenotyping including seeds based on images captured with digital cameras, but the results are strongly influenced by the capture conditions. Indeed, overlapping seeds may be recognized as a single seed, resulting in an abnormal seed shape output. To address this issue, we implemented an image analysis pipeline using deep learning for phenotyping morphological traits of lettuce seeds. Specifically, we trained a neural network with a synthetic dataset to identify and segregate the respective seed area from densely oriented seed images, regardless of their orientation (Toda et al., 2020). In agriculture and plant science area, automation using deep learning has increasingly promoted the implementation of various high throughputs phenotyping such as weed detection, disease detection, and prediction of protein folding in recent years (Kamilaris and Prenafeta-Boldú, 2018; Cramer, 2021).

The major purpose of our study was to identify QTLs rapidly and reliably for seed morphology of oilseed-type cultivar by combining genotyping by double digest restriction-site associated DNA sequencing (ddRAD-seq) and phenotyping by automated image analysis. We expect to be able to conduct a more accurate genetic analysis since the seed shape can be efficiently grasped by automation. We characterized an F_2 population derived from a cross between an oilseed-type cultivar and a wild species for QTL mapping of seed morphology to elucidate the genetic mechanisms related to domestication from wild species to oilseed-type cultivars.

Materials and methods

Plant materials

All plant materials were grown at the Nagano Vegetable and Ornamental Crops Experiment Station (Shiojiri City, Nagano prefecture, Japan; 36° 10′ N, 137° 93′ E). The lettuce cultivar 'Oilseed' derived from upper Egypt was obtained from the Centre for Genetic Resources, the Netherlands (CGN), under the stock number 'CGN04769'. The wild *Lactuca* species 'UenoyamaMaruba' is the landrace of Nagano prefecture in Japan. A set of 173 F_2 individuals derived from the 'Oilseed' × 'UenoyamaMaruba' cross was used for linkage analysis and to produce self-pollinated F_3 lines that were used to evaluate the segregation patterns for genotypes and seed morphology of the F_2 individuals. F_3 seeds were spread onto white paper, and images were captured with a digital camera C5050Z (Olympus optical Co., Ltd, Japan) using transmission light. Each image size was 2560 × 1920 pixels at a resolution of 72 dpi.

Quantification of seed morphology by automated image analysis

An instance segmentation neural network (Mask R-CNN; He et al., 2017) was trained and applied to extract morphological parameters of lettuce seeds, as described previously (Toda et al., 2020). Briefly, 28 single seed images of 'UenoyamaMaruba' were used to generate a synthetic image. The generated image resolution was 700 × 700 pixels, and each image was center-cropped to the size of 512 × 512 pixels for model training. Mask R-CNN was implemented on a Keras backend (https://github.com/matterport/Mask_RCNN) with default parameters. The number of generated training and

validation synthetic data were 500 images, respectively, which were fed to the network model with a batch size of 8 images per graphics processing unit (GPU). Our training environment consisted of 4 GPUs, therefore the total batch size was 32. Since the total number of images in the real-world test data (images to be analyzed in this research) was only 175, visual inspection against the inference result was used to decide whether the training was adequate instead of depending on common machine learning metrics such as mean average precision or intersection-of-union. In our case, 40 epochs of training led to an acceptable result (Figure 1B).

After isolating the seed area, morphological parameters of the respective regions were calculated using the measure.resionprops module of the sci-kit image library. Seed length (major_axis_length), width, (minor_axis_length), area, perimeter, eccentricity, and bounding box coordinate (bbox) were directly obtained from the above module (Table 1), and the metrics were calculated by the following formulas; length_to_width_ratio = major_axis_length/minor_axis_length; circularity = (4 × Pi × area)/perimeter^2. However, the dataset may include data for incomplete seed shapes; e.g. when seeds are at the edge of the image, physically damaged, or partially covered and masked by other seeds. Therefore, such potential noise in the following analysis was excluded by the following criteria: bounding box coordinates that exceed a five-pixel margin from the edge and length_to_width_ratio that falls outside the 5% and 95% quantile range of the total population. The filtered dataset was normalized by a scaling factor of 43.833 (pixels/mm) before subsequent analysis.

Linkage map construction by ddRAD-seq analysis

Genomic DNA was extracted from leaves using the Nucleo-Spin Plant II Extract Kit (Machery-Nagel, Duren, Germany), and was digested with PacI (New England Biolabs, Beverly, MA, USA) and NlaIII (New England Biolabs). After ligation of the adapters to both ends of digested DNA fragments, they were amplified by indexed primers and pooled for Illumina sequencing (Matsumura et al., 2014; Seki et al., 2020). The ddRAD-seq libraries were sequenced on a HiSeqXten platform (Illumina, San Diego, CA, USA). Paired-end sequencing reads (150 bp × 2) were analyzed for quality control, adapter trimming, ddRAD-seq tag extraction, counting, and linkage map construction using RAD-R scripts (Seki, 2021). Briefly, the minimum PHRED score for quality control is set to 35 and the acceptable ratio for the low PHRED score is 0.1 (i.e., 10%). Adapter trimming is used by preprocessReads function in ShortReads package (Morgan et al., 2009). Reads were mapped with the RAD tags in each parent against the lettuce reference genome sequence (version8 from the crisphead cultivar 'Salinas', https://genomevolution.org/coge/GenomeInfo.pl?gid=28333) using Burrows-Wheeler Alignment tool (BWA) (Li and Durbin, 2009). The dataset with outliers of p values (3:1) by the chi-square test was excluded. The linkage map was drawn using R/qtl (Broman et al., 2003). The pipeline was run by the default setting. Raw sequence data (FASTQ) for this ddRAD-seq dataset were deposited in the DNA Data Bank of Japan (DDBJ) Sequence Read Archive (http://ddbj.nig.ac.jp/dra/index_e.html) under accession number DRA013652.

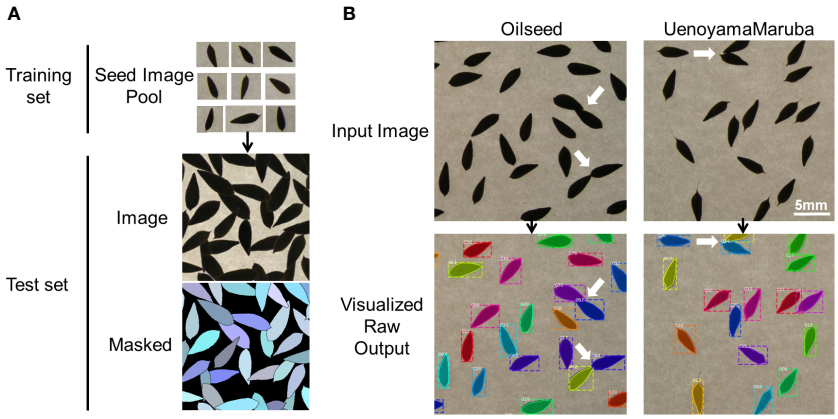

FIGURE 1
Overview of the image analysis pipeline using lettuce seeds. **(A)** Training pool with individual seed images to identify seeds and measure seed morphology traits. **(B)** Representative analyses of input images for the two parental cultivars. White arrows indicate overlapping seeds that are recognized as individual seeds by the image analysis pipeline.

TABLE 1 Traits measured on the parental cultivars and the F_2 population derived from the 'Oilseed' × 'UenoyamaMaruba' cross.

Trait name	Unit	Rating scale
Area	mm^2	Total projected area of the seed
Width	mm	Distance from side to side of the seed
Length	mm	Distance from top to bottom of the seed
Length-to-width ratio	arbitrary units (au)	major_axis_legnth/minor_axis_length
Eccentricity	arbitrary units (au)	Eccentricity of the ellipse which shares the second-moment of the seed
Perimeter length	mm	Perimeter length of the seed
Circularity	arbitrary units (au)	The roundness of the seed

See Methods for details of the respective parameters.

QTL detection by composite interval mapping (CIM)

QTL detection with CIM was conducted using the Haley–Knott regression of the R/qtl package in R (Broman et al., 2003). The genome-wide LOD threshold at the 1% and 5% significance levels was individually determined using a 10,000-permutation test for each trait. The proportion of phenotypic variance was calculated from the value at the peak, as indicated by CIM. A detailed script is described in CIM_script.R (https://github.com/KousukeSEKI/RAD-seq_scripts).

Candidate gene mining of QTL intervals

The high-density genetic map by ddRAD-seq was used to determine the genetic positions of the QTL intervals. Using the lettuce reference genome sequence (version8 from the crisphead cultivar 'Salinas'), the candidate genes in the QTL intervals between neighboring markers were identified. Candidate genes involved in the seed morphology were selected from the gene annotation data.

Results

Phenotyping seed morphology with instance segmentation

We applied our image analysis pipeline to phenotype seed morphology traits using seed images from 173 F_3 lines and those of the two parental lines. Because we trained the neural network with a training set of seed images, our pipeline recognized overlapping seeds as individual seeds (Figure 1). To analyze seed morphology traits, we performed a number of post-processing steps: We removed outliers falling outside of a 12.5% (lower limit) and 87.5% (upper limit) quantile threshold for all traits for each individual. Across the generated dataset, we measured at least 71 seeds per line, with a maximum number of 399 seeds and a mean of 243. We applied a Student's *t*-test to compare seed morphology traits between the two parents: We observed statistically significant differences in all seven traits (area, width, length, length-to-width ratio, eccentricity, perimeter length, and circularity) (Tables 1, 2). A principal component analysis (PCA) using all data of seven traits from 173 F3 lines revealed that the first two principal components

TABLE 2 Phenotypic traits measured in the parental cultivars.

Trait name	Oilseed				UenoyamaMaruba				*T*-test
	Min	Mean	Max	SD	Min	Mean	Max	SD	*P*-value
Area	2.52	3.51	4.14	0.39	2.91	3.02	3.15	0.08	< 0.01**
Width	1.04	1.28	1.43	0.09	1.14	1.18	1.20	0.02	< 0.01**
Length	3.00	3.54	3.85	0.20	3.27	3.36	3.47	0.06	0.003**
Length-to-Width ratio	2.50	2.76	3.06	0.15	2.76	2.86	2.96	0.07	0.04*
Eccentricity	0.92	0.93	0.95	0.01	0.93	0.94	0.94	0.003	0.03*
Perimeter length	7.53	8.84	9.52	0.48	8.21	8.53	8.87	0.18	0.03*
Circularity	0.52	0.56	0.60	0.02	0.50	0.52	0.55	0.01	< 0.01**

**,* significant at the level of 1% and 5%, respectively.

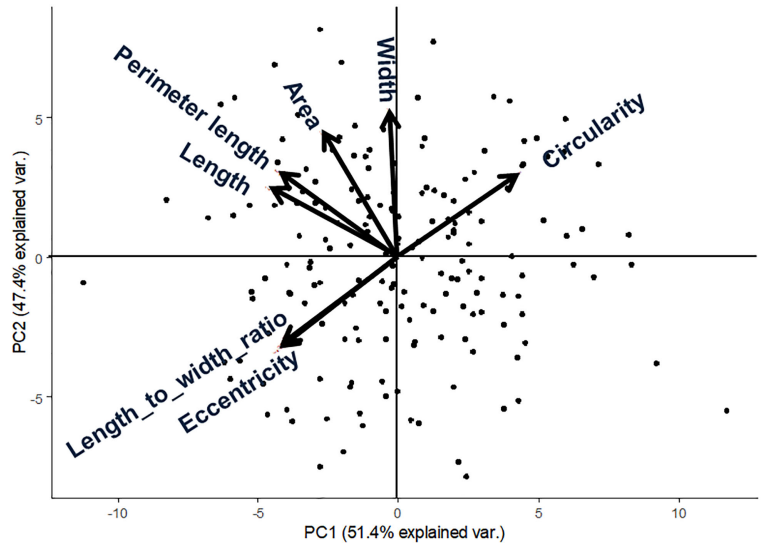

FIGURE 2
Principal component analysis from 173 F_3 lines of lettuce seed morphological parameters. Arrows indicate eigenvectors of each descriptor.

(PCs) explain 98.8% of the total standing variation (Figure 2). The length-to-width ratio, eccentricity, length, perimeter length, and circularity contributed to the first PC, with circularity being oriented in the direction opposite to that of length-to-width ratio and eccentricity. All traits contributed to the second PC. The direction of the length eigenvector was nearly identical to that of perimeter length. The direction of width differed from that of perimeter length and length. Notably, the area eigenvector was in between the direction of width and length. We calculated the Pearson's correlation coefficients between the traits using all measurements of seed morphology to explore relationships. We observed a range of correlation values, from weak correlations to significant correlations reaching up to 0.99 (Table 3). In addition, Pearson's correlation coefficients reflected the results of the PCA, with strong positive correlations between close eigenvectors (length-to-width ratio and eccentricity [0.99, $P < 0.01$], area and width [0.91, $P < 0.01$], between area and length [0.86, $P < 0.01$], and between area and perimeter length [0.92, $P< 0.01$]) and strong negative correlations oppositely pointing eigenvectors (length-to-width ratio and circularity [–0.96, $P < 0.01$]).

Genotyping by ddRAD-seq analysis

For genetic mapping of the loci controlling seed morphology, we conducted ddRAD-seq analysis for constructing a linkage map using an F_2 population derived from a cross between the accessions 'Oilseed' and 'UenoyamaMaruba'. Following sequencing of the ddRAD-seq libraries on an Illumina HiSeq instrument, we obtained 1,075,108 and 1,125,238 single reads (150 bp) for 'Oilseed' and 'UenoyamaMaruba', respectively. We converted the genotypes from 'Oilseed' and 'UenoyamaMaruba' genotypes to A and B, respectively. For the RAD-R scripts (Seki, 2021), we employed mem as BWA mode, select as construction method, and 7US as

TABLE 3 Pearson's correlation coefficients of seed morphology traits using the phenotyping data of F_3 seeds.

	Area	Width	Length	Length-to-width ratio	Eccentricity	Perimeterlength
Width	0.91**					
Length	0.86**	0.58**				
Length-to-width ratio	– 0.2**	– 0.58**	0.32**			
Eccentricity	– 0.2**	– 0.58**	0.32**	0.99**		
Perimeter length	0.92**	0.68**	0.99**	0.19**	0.19*	
Circularity	0.16*	0.54**	– 0.34**	– 0.96**	– 0.96**	– 0.24**

Significant at *, ** $P < 0.05$ and 0.01, respectively.

TABLE 4 Summary of integrated lettuce linkage groups and genetic map information.

Linkage groups	Total mapped tags	Common tags		Oilseed unique tags		UenoyamaMaruba unique tags		Linkage construct marker		
		No. RAD-tags	(%)	No. RAD-tags	(%)	No. RAD-tags	(%)	No. biallelic tags	Map length (cM)	Average interval between markers (cM)
LG1	35396	3142	8.9	15556	43.9	16698	47.2	173	133.1	0.8
LG2	35377	2971	8.4	16423	46.4	15983	45.2	214	140.6	0.7
LG3	49415	7519	15.2	21610	43.7	20286	41.1	308	169.2	0.5
LG4	67093	8632	12.9	29791	44.4	28670	42.7	371	236.1	0.6
LG5	64605	12409	19.2	26698	41.3	25498	39.5	227	203.1	0.9
LG6	42377	11312	26.7	15830	37.4	15235	36.0	127	134.1	1.1
LG7	46046	16001	34.8	15006	32.6	15039	32.7	120	122.2	1.0
LG8	64692	15570	24.1	25376	39.2	23746	36.7	221	179.6	0.8
LG9	51409	19153	37.3	16593	32.3	15663	30.5	109	187.4	1.7
Total	456410	96709	21.2	182883	40.1	176818	38.7	1870	1505.4	0.8

correction approach. We then used the 1,870 pairs of RAD tags extracted from the two parents as codominant markers for genetic mapping of seed morphology for linkage map construction based on the genotypes of the 173 F_2 individuals (Table 4). The total length of the resulting linkage map was 1,505.4 cM. The average inter-marker distance ranged from 0.5 (LG3) to 1.7 cM (LG9). The number of markers per linkage group ranged from 109 (LG9) to 371 (LG4).

QTL mapping of seed morphology

We used composite interval mapping (CIM) to detect QTLs, using the seed morphology trait values determined by the pipeline and the genotype data from the ddRAD-seq across 173 F_2 individuals. We detected 11 QTLs for 7 traits (Table 5). Importantly, the QTLs with the highest logarithm of odds (LOD) scores, *qLWR-3.1*, *qECC-3.1*, and *qCIR-3.1*, are associated with the length-to-width ratio, eccentricity, and circularity, respectively, mapped to LG3 between 161.5 and 214.6 Mb. These three QTLs each accounted for approximately 27.95%, 28.55%, and 29.84% of the phenotypic variation explained (PVE) for their respective trait. We observed significant differences in the phenotypic values of putative homozygote and heterozygote individuals for each of the three traits, suggesting that each trait is controlled by one semi-dominant locus (Figure 3). We also detected two QTLs linked to both length and perimeter length at almost the same position on LG5 and LG8. The QTLs *qWID-7.1* and *qWID-8.1* linked to width mapped to LG7 and LG8, while *qARE-8.1* was associated with area mapped almost to the same position as *qWID-8.1* (Table 5).

Several candidate genes in the locus associated with seed morphology

Each locus included several candidate genes annotated according to the lettuce reference genome sequence (version8 from crisphead cultivar 'Salinas') (Table 5). The number of candidate genes located on each locus ranged from 9 (*qCIR-4.1*) to 606 (*qLEN-5.1*). The three QTLs with the highest LOD, *qLWR-3.1*, *qECC-3.1*, and *qCIR-3.1*, have 517, 517, and 443 candidate genes, respectively.

Discussion

Lettuce, a diploid species (2n = 2x = 18), is a self-pollinated crop with a large genome (~ 2.3 Gb) (Reyes-Chin-Wo et al., 2017). Cultivated lettuce exhibits a low rate of intraspecific polymorphisms (Truco et al., 2007; Simko and Hu, 2008). In this study, we produced a segregating F_2 population derived from a cross between an oilseed-type cultivar and a wild species of lettuce to maximize the detection of polymorphisms. We thus developed 1,870 RAD markers and assigned them to 9 linkage groups by ddRAD-seq rapidly (Table 4). While our previous report detected the rate of common tags between parents of crisphead type to be as high as 60% (Seki et al., 2020), the two parental lines used here only shared 21.2% of common tags. This result suggests that polymorphisms between oilseed type and wild lettuce species are more numerous than between crisphead-type lettuce cultivars. With the help of our image analysis pipeline, we obtained phenotypic data for 42,036 seeds quickly and with high performance, which would be impossible to easily accomplish manually. The pipeline successfully recognized

TABLE 5 QTL detected by composite interval mapping in the F_2 population derived from the 'Oilseed' × 'UenoyamaMaruba' cross.

QTL name	Trait name	Linkage group	Marker	Genetic distance (cM)	Genetic position (cM)	Threshold LOD (1%)	Threshold LOD (5%)	LOD	Additive effect	Dominant effect	R^2 (%)	Number of candidate genes in QTL
qLWR-3.1	Length-to-Width ratio	3	LG3_v8_176.7Mbp to LG3_v8_214.6Mbp	16.88	138.0	4.57	3.88	12.46	0.1628	0.0167	27.95	517
qECC-3.1	Eccentricity	3	LG3_v8_176.7Mbp to LG3_v8_214.6Mbp	16.88	138.0	4.77	4.00	12.77	0.0051	0.0006	28.55	517
qCIR-3.1	Circularity	3	LG3_v8_161.5Mbp to LG3_v8_196.7Mbp	14.11	133.0	4.63	3.97	12.62	-0.0191	-0.0024	29.84	443
qCIR-4.1	Circularity	4	LG4_v8_73.0Mbp to LG4_v8_73.4Mbp	1.81	44.6	4.63	3.97	4.01	0.0032	0.0126	7.35	9
qLEN-5.1	Length	5	LG5_v8_230.7Mbp to LG5_v8_260.7Mbp	19.43	141.0	4.62	3.88	4.58	-0.0858	-0.0399	10.03	606
qLEN-8.1	Length	8	LG8_v8_293.1Mbp to LG8_v8_307.9Mbp	11.88	172.1	4.62	3.88	4.66	0.0070	-0.1280	10.49	92
qPER-5.1	Perimeter length	5	LG5_v8_253.7Mbp to LG5_v8_260.7Mbp	6.27	141.0	4.79	4.01	4.27	-0.1958	-0.1051	9.30	152
qPER-8.1	Perimeter length	8	LG8_v8_298.8Mbp to LG8_v8_305.4Mbp	3.46	172.1	4.79	4.01	5.01	0.0092	-0.3185	11.24	27
qWID-7.1	Width	7	LG7_v8_165.9Mbp to LG7_v8_192.4Mbp	20.64	106.5	4.37	3.78	4.54	-0.0336	0.0077	10.88	472
qWID-8.1	Width	8	LG8_v8_279.1Mbp to LG8_v8_293.1Mbp	4.01	166.0	4.37	3.78	3.80	-0.0258	-0.0205	8.44	80
qARE-8.1	Area	8	LG8_v8_279.1Mbp to LG8_v8_295.3Mbp	5.70	163.1	4.67	3.95	4.52	-0.0703	-0.1623	11.21	93

10.3389/fpls.2022.949470

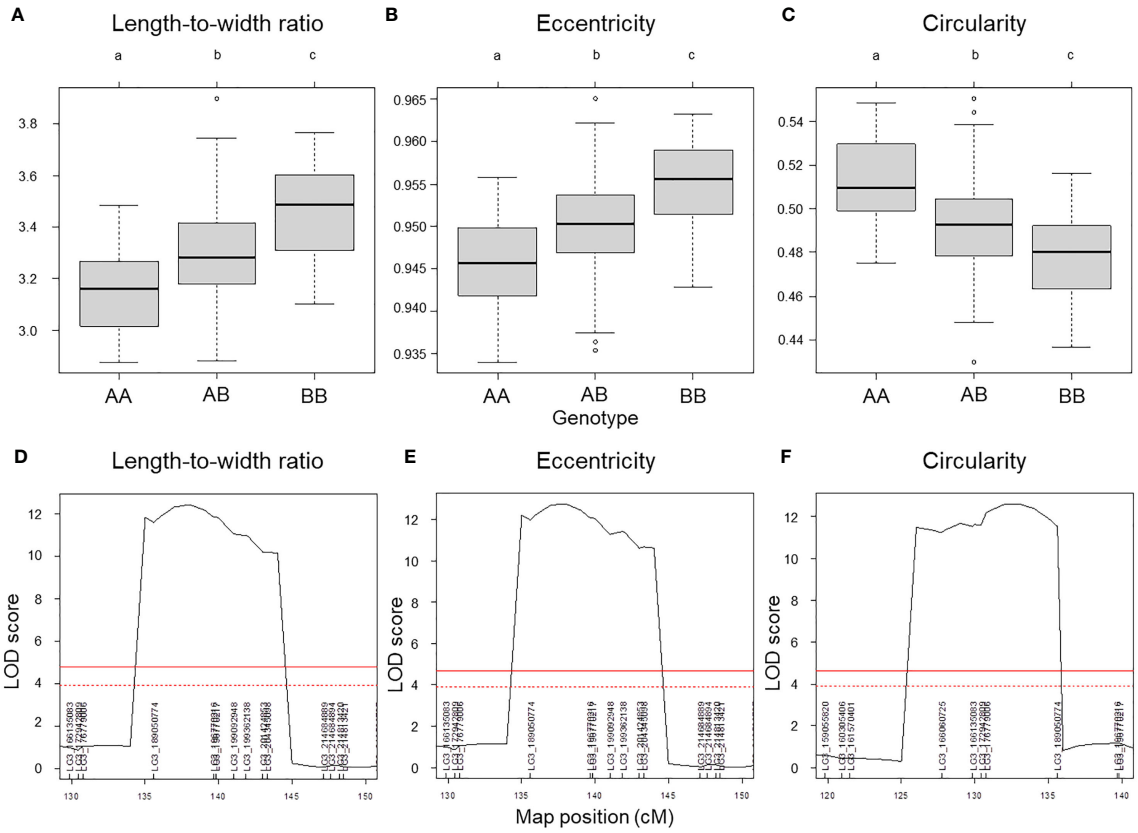

FIGURE 3
QTL mapping results for selected seed morphology traits. **(A–C)** Phenotypic difference for the traits length-to-width ratio **(A)**, eccentricity **(B)** and circularity **(C)** between the three possible genotypes at three major QTLs with high LOD scores. **(D–F)** QTL peaks with LOD scores at each locus. Different lowercase letters represent significant differences ($P < 0.01$).

overlapping seeds and separated them into individual seeds, which is instrumental in large-scale phenotyping without having to place seeds evenly apart manually. The data collected with our pipeline correctly identified the bigger and rounder seeds characteristic of oilseed-type lettuce compared to the wild species (Figure 1, Table 2). A more rounded seed shape would be advantageous to accumulate more seed oil. Perhaps the initial artificial selection during domestication was for uniform germination, for which seed morphology and size are important determinants. Using the genotyping data generated by ddRAD-seq and the above phenotyping data, we detected 11 QTLs for the 7 seed morphology traits (Table 5). Of these, three QTLs, *qLWR-3.1*, *qECC-3.1*, and *qCIR-3.1*, exhibited major effects on their respective trait (PVE > 25%). PCA illustrated the relationships between traits using the phenotyping data gathered from F_3 seeds (Figure 2 and Table 3). We measured a strong and positive correlation (R = 0.99) between perimeter length and seed length, while these two parameters were less strongly correlated with seed width (R = 0.68 and 0.58, respectively), reflecting the greater influence of seed length over seed width on seed perimeter length. In support of this observation, we detected two QTLs mapping to the same genomic regions on LG5 and LG8 for each of the two traits (Table 5). Similarly, seed area appeared to be positively and strongly correlated with seed width (R = 0.91), but not seed length (R = 0.86). We detected a common QTL associated with seed area and width on LG8 (Table 5), in the same location as a previously reported QTL for seed width (Hartman et al., 2013). The length-to-width ratio, eccentricity, and circularity associated with the characteristic round shape features of oilseed-type lettuce also displayed very strong correlations among them, prompting us to take a closer look at their underlying genetic basis (Table 3). Remarkably, the major QTLs associated with these three traits mapped to almost the same region on LG3, from 161.5 to 214.6 Mb (Figure 3). We hypothesize that this genomic region harbors a key causal locus associated with the domestication of seed morphology from wild-type to oilseed type lettuce. There are several possibilities as to why multiple QTLs for seed traits were detected in the same genomic region, such as whether linkage disequilibrium occurred or the same

gene was responsible. Domestication involves the genetic modification of wild species due to selection for a given phenotype that is beneficial to humans and is often determined by a few loci with major effects (Abbo et al., 2012). Although it is difficult to precisely compare genetic maps because of the limited number of markers, the same regions on LG3 and LG7 were reported to be associated with clusters of QTLs linked with domestication for the traits of seed output, leaf shape, and late flowering in lettuce (Hartman et al., 2013). Thus, our hypothesis that the genomic region on LG3 mainly reflects the domestication of seed morphology traits agrees with the results of previous genetic studies. In rice and *arabidopsis*, several QTLs for seed morphology have been molecularly identified and characterized. The molecular mechanisms involve secreted peptide hormones, transcription factors, phytohormones, proteasomal degradation, and G-proteins signaling (Fiume and Fletcher, 2012; Zuo and Li, 2014; Li and Li, 2016). Among others, brassinosteroid (BR) is known to play a crucial role in seed development and morphology (Jiang et al., 2013). In the three QTLs on LG3, a gene, encoding a homolog to BRI1 suppressor 1-like protein, was found within the common region as an intriguing candidate. In addition, secreted peptide hormones that could regulate seed morphology include members of the CLAVATA3/EMBRYO SURROUNDING REGION-RELATED (CLE) family (Fiume and Fletcher, 2012). Remarkably, a gene, encoding a homolog to CLE family, was found within the *qWID-8.1* and *qARE-8.1* loci on LG8. However, it is also possible that the Salinas genome sequence may have lost genetic information on seed morphology that was once present in wild species. Elucidations of the genetic mechanisms in these loci are a future concern.

Conclusion

This study demonstrated that accurate phenotyping by automation using deep learning, even if the plant size was too small, was highly effective for the QTL analysis. In plants with the announced genome sequence like lettuce, ddRAD-seq could rapidly show not only the loci with high LODs but also the results of candidate genes mining. In the future, it could be applied to analysis using images taken by microscopes or drones. However, deep learning is not good at processing exceptional data, and outlier removal would be necessary. Due to the recent advances in computation power, approaches based on deep learning are becoming accessible to agriculture science (de Medeiros et al., 2021; Nehoshtan et al., 2021). Our data demonstrate that deep learning could significantly contribute to developing the field of not only plant science but also agriculture for plant phenotyping.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: DDBJ Sequence Read Archive, DRA013652.

Author contributions

KS and YT planned the experiments. KS developed the mapping population, performed ddRAD-seq, and phenotyping data analysis. YT performed crop seed phenotyping. KS and YT wrote and approved the manuscript.

Funding

This work was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (21H05152 for YT).

Acknowledgments

We thank Dr. Ken Naito and all the organizers for hosting "Society of Post Youth Agronomists (SPY-A)", a research conference which not only provided the opportunity of the co-authors' acquaintance but also led to the collaboration of this research.

Conflict of interest

Author YT was employed by Phytometrics Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abbo, S., Lev-Yadun, S., and Gopher, A. (2012). Plant domestication and crop evolution in the near east: on events and processes. *CRC. Crit. Rev. Plant Sci.* 31, 241–257. doi: 10.1080/07352689.2011.645428

Broman, K. W., Wu, H., Sen, Ś., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112

Cramer, P. (2021). AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* 28, 704–705. doi: 10.1038/s41594-021-00650-1

de Medeiros, A. D., Bernardes, R. C., da Silva, L. J., de Freitas, B. A. L., Dias, D. C. F., dos, S., et al. (2021). Deep learning-based approach using X-ray images for classifying crambe abyssinica seed quality. *Ind. Crops Prod.* 164, 113378. doi: 10.1016/j.indcrop.2021.113378

De Vries, I. M. (1997). Origin and domestication of *Lactuca sativa* l. *Genet. Resour. Crop Evol.* 44, 165–174. doi: 10.1023/A:1008611200727

Fiume, E., and Fletcher, J. C. (2012). Regulation of *Arabidopsis* embryo and endosperm development by the polypeptide signaling molecule CLE8. *Plant Cell* 24, 1000–1012. doi: 10.1105/tpc.111.094839

Gao, H., Gu, Y., Jiang, H., Li, Y., Xin, D., Liu, C., et al. (2019). Mapping and analysis of QTLs related to seed length and seed width in *Glycine max*. *Plant Breed.* 138, 733–740. doi: 10.1111/pbr.12745

Hartman, Y., Hooftman, D. A. P., Eric Schranz, M., and van Tienderen, P. H. (2013). QTL analysis reveals the genetic architecture of domestication traits in crisphead lettuce. *Genet. Resour. Crop Evol.* 60, 1487–1500. doi: 10.1007/s10722-012-9937-0

He, K., Gkioxari, G., Dollar, P., and and Girshick, R.c (2017). Mask R-CNN. in *IEEE International Conference on Computer Vision (ICCV)* 2980–2988

Jiang, W. B., Huang, H. Y., Hu, Y. W., Zhu, S. W., Wang, Z. Y., and Lin, W. H. (2013). Brassinosteroid regulates seed size and shape in *Arabidopsis*. *Plant Physiol.* 162, 1965–1977. doi: 10.1104/pp.113.217703

Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Kumawat, G., and Xu, D. (2021). A major and stable quantitative trait locus qSS2 for seed size and shape traits in a soybean RIL population. *Front. Genet.* 12. doi: 10.3389/fgene.2021.646102

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008

Matsumura, H., Miyagi, N., Taniai, N., Fukushima, M., Tarora, K., Shudo, A., et al. (2014). Mapping of the gynoecy in bitter gourd (*Momordica charantia*) using RAD-seq analysis. *PloS One* 9, 1–10. doi: 10.1371/journal.pone.0087138

Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., and Gentleman, R. (2009). ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* 25, 2607–2608. doi: 10.1093/bioinformatics/btp450

Nehoshtan, Y., Carmon, E., Yaniv, O., Ayal, S., and Rotem, O. (2021). Robust seed germination prediction using deep learning and RGB image data. *Sci. Rep.* 11, 1–10. doi: 10.1038/s41598-021-01712-6

Ramadan, A. A. S. (1976). Characteristics of prickly lettuce seed oil in relation to methods of extraction. *Food / Nahrung* 20, 579–583. doi: 10.1002/food.19760200602

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., et al. (2017). Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* 8, 1–11. doi: 10.1038/ncomms14953

Ryder, E. J. (1999). *Lettuce, endive and chicory* (Wallingford: CABI Publishing).

Seki, K. (2021). RAD-r scripts: R pipeline for RAD-seq from FASTQ files to linkage maps construction and run R/QTL, operating only at copying and pasting scripts into r console. *Breed. Sci* 71, 426–434. doi: 10.1270/jsbbs.20159

Seki, K., Komatsu, K., Tanaka, K., Hiraga, M., Kajiya-Kanegae, H., Matsumura, H., et al. (2020). A CIN-like TCP transcription factor (*LsTCP4*) having retrotransposon insertion associates with a shift from salinas type to empire type in crisphead lettuce (*Lactuca sativa* l.). *Hortic. Res.* 7, 1–14. doi: 10.1038/s41438-020-0241-4

Simko, I., and Hu, J. (2008). Population structure in cultivated lettuce and its impact on association mapping. *J. Am. Soc Hortic. Sci.* 133, 61–68. doi: 10.21273/jashs.133.1.61

Sun, L., Wang, X., Yu, K., Li, W., Peng, Q., Chen, F., et al. (2018). Mapping of QTLs controlling seed weight and seed-shape traits in *Brassica napus* l. using a high-density SNP map. *Euphytica* 214, 1–12. doi: 10.1007/s10681-018-2303-3

Tanabata, T., Shibaya, T., Hori, K., Ebana, K., and 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

Toda, Y., Okura, F., Ito, J., Okada, S., Kinoshita, T., Tsuji, H., et al. (2020). Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Commun. Biol.* 3, 1–12. doi: 10.1038/s42003-020-0905-5

Truco, M. J., Antonise, R., Lavelle, D., Ochoa, O., Kozik, A., Witsenboer, H., et al. (2007). A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theor. Appl. Genet.* 115, 735–746. doi: 10.1007/s00122-007-0599-9

Wei, T., van Treuren, R., Liu, X., Zhang, Z., Chen, J., Liu, Y., et al. (2021). Whole-genome resequencing of 445 lactuca accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* 53, 752–760. doi: 10.1038/s41588-021-00831-0

Williams, K., Munkvold, J., and Sorrells, M. (2013). Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* l.). *Euphytica* 190, 99–116. doi: 10.1007/s10681-012-0783-0

Yue, B., Cai, X., Yuan-, W., Vick, B., and Hu, J. (2009). Mapping quantitative trait loci (QTL) controlling seed morphology and disk diameter in sunflower (*Helianthus annuus* l.). *Helia* 32, 17–36. doi: 10.2298/HEL0950017Y

Zhang, S., Hu, X., Miao, H., Chu, Y., Cui, F., Yang, W., et al. (2019). QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (*Arachis hypogaea* l.). *BMC Plant Biol.* 19, 1–15. doi: 10.1186/s12870-019-2164-5

Zuo, J., and Li, J. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* 48, 99–118. doi: 10.1146/annurev-genet-120213-092138