agronomy

MDPI

Article

Deep Learning-Assisted Image Phenotyping for Genetic Dissection of Pod-Related Traits in Soybean

Liping Wei [1], Shunchang Su [1], Tuanjie Zhao [2] and Fangguo Chang [3,*]

[1] College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China; weilp@gsau.edu.cn (L.W.); susc@gsau.edu.cn (S.S.)
[2] Key Laboratory of Biology and Genetics Improvement of Soybean, Ministry of Agriculture, Zhongshan Biological Breeding Laboratory (ZSBBL), National Innovation Platform for Soybean Breeding and Industry-Education Integration, State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, College of Agriculture, Nanjing Agricultural University, Nanjing 211800, China; tjzhao@njau.edu.cn
[3] College of Agronomy, Gansu Agricultural University, Lanzhou 730070, China
* Correspondence: changfg@gsau.edu.cn

Abstract

Soybean pod-related traits are important for seed development, yield formation, and cultivar evaluation. However, conventional measurements are inefficient and have limited ability to characterize complex pod features such as curvature, local enlargement, and continuous color variation. In this study, mature pod images of 187 cultivated soybean accessions collected across two successive years were analyzed using a deep learning-assisted image phenotyping approach. Eight pod-related traits related to size, morphology, and color were extracted from pod images. A genome-wide association study (GWAS) was performed using 61,541 high-quality SNP markers to dissect the genetic basis of these image-derived pod traits. A total of 16 stable loci associated with pod size, morphology, and color traits were identified across 11 chromosomes. Among these loci, eight were not reported in the previous image-based soybean pod GWAS study. Based on SoyBase gene annotation and Gene Ontology biological process information, 32 biologically relevant candidate gene records were prioritized within the corresponding candidate genomic intervals, while pod-related expression profiles and SoyBase association information were used as supporting evidence for candidate gene evaluation. These findings indicate that refined image-derived traits can provide complementary genetic information beyond conventional pod measurements and offer additional opportunities for dissecting soybean pod development, morphology, and mature pod color variation.

Keywords: soybean; deep learning; image phenotyping; pod-related traits; GWAS

Academic Editor: Paul Kwan

Received: 19 June 2026
Revised: 25 July 2026
Accepted: 29 July 2026
Published: 31 July 2026

1. Introduction

Soybean (*Glycine max* (L.) Merr.) is an important food and oilseed crop worldwide, and its seeds provide major sources of plant protein and vegetable oil. During seed development, the pod protects the developing seeds and contributes to seed filling and yield formation [1]. Pod- and seed-related traits are commonly considered in soybean germplasm characterization and cultivar evaluation [2]. However, conventional phenotyping of these traits generally relies on manual measurement or visual assessment, which is time-consuming, labor-intensive, and potentially affected by measurement position and operator judgment. Moreover, commonly used measurements, such as straight-line

pod length, maximum width, and visually assigned color classes, have limited ability to characterize curved axial paths, local width variation, irregular margins, and continuous color differences.

Image-based phenotyping provides an efficient and reproducible approach for quantifying plant organ size, shape, and color under standardized imaging conditions. High-throughput imaging has been used to characterize plant growth, development, and stress responses [3–5], and image-derived traits have also been incorporated into genetic analyses of complex crop phenotypes [6–8]. In soybean, Baek et al. developed an image-analysis method for high-throughput seed phenotyping [9], while Uzal et al. and Yang et al. applied deep learning and image segmentation to estimate seed number per pod [10,11]. Liu et al. further developed a keypoint-based method for fine-grained phenotyping of mature soybean pods [12]. In our previous study, mature pod image traits were integrated with GWAS and linkage mapping to identify stable loci associated with pod size and mature pod color [13]. Nevertheless, most commonly used image-derived pod traits mainly describe overall dimensions and may not fully capture variation in axial geometry, local width distribution, and curvature-related morphology.

GWAS and linkage mapping have been widely used to dissect the genetic basis of complex quantitative traits in crops [14–17]. In soybean, loci associated with pod number, pod wall characteristics, maturity, pod shattering, and mature pod color have been reported [18–24]. Functional studies have also identified genes involved in specific pod-related traits, such as the NAC gene associated with resistance to pod shattering [25]. However, many reported associations remain confined to broad genomic intervals, and the genetic basis of complex pod morphology is still insufficiently understood. The successful application of morphology-specific image descriptors in peanut pod QTL analysis further indicates that refined phenotypic definitions can facilitate the detection of genetic variation associated with complex organ morphology [26]. Therefore, integrating refined image-derived pod traits with GWAS may provide complementary information for identifying and interpreting loci associated with soybean pod variation.

Accordingly, this study aimed to (1) develop a deep learning-assisted image phenotyping approach to extract pod size-, morphology-, and color-related traits from mature soybean pods of 187 cultivated accessions; (2) evaluate the phenotypic variation and heritability of these refined traits and conduct GWAS to identify stable trait-associated loci; and (3) prioritize putative candidate genes within the corresponding candidate genomic intervals using Gene Ontology biological process information and SoyBase annotations, with pod-related expression profiles and previously reported association records used as supporting evidence.

2. Materials and Methods

2.1. Plant Materials and Experimental Design

A total of 187 cultivated soybean accessions were used in this study. The detailed construction of the association panel was described by Chang et al. [13]. All plant materials were obtained from the National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China. Field trials were conducted at the Jiangpu Experimental Station of Nanjing Agricultural University (31°02′ N, 118°04′ E) during the summer growing seasons of 2017 and 2018 using a completely randomized design with three replications. Each accession was grown in a 1-m-long single-row plot with 0.5 m spacing between adjacent rows. Field management followed standard agronomic practices under normal growing conditions.

2.2. Pod Image Acquisition and Preprocessing

The original pod images analyzed in this study were previously acquired by Chang et al. [13]. Five mature three-seeded pods from each genotype were photographed using an industrial camera (MERCURY, MER-310-12UC, Daheng Group Inc., Beijing, China) equipped with a 50-mm lens (M5018-MP2, F1.8–F16C, Computer, Inc., Tokyo, Japan). Images were acquired on a white background under a controllable lighting system (OPT-LI38037-w/AP1024-2, OPT Inc., Dongguan, China). The camera was mounted on a fixed stand, and a 7-cm ruler was included in each image for scale calibration. Pods were placed horizontally along their longitudinal axis in the center of the imaging field. Backlight and LED illumination were used during image acquisition. White balance correction was performed by setting the initial RGB values to 255, 255, and 255. The final image resolution was 2048 × 1536 pixels.

2.3. Pod Image Segmentation and Image-Derived Trait Extraction

To extract refined morphological and color traits from the acquired pod images, a binary semantic segmentation model, DeepLabV3Binary, was applied to delineate the pod foreground, and image-derived traits were subsequently calculated from the resulting segmentation masks. The overall workflow comprised image segmentation, binary mask post-processing, skeletonization, width sampling along the main axis, and color feature extraction (Figure 1). DeepLabV3Binary was implemented by adapting the DeepLabV3 semantic segmentation framework [27], with ResNet50 as the backbone network [28]. The original multi-class classifier and auxiliary classifier were replaced with single-channel convolutional layers to generate pod foreground probability maps. Multi-scale atrous convolution modules were used to extract image features at different receptive fields. The resulting foreground masks were used for subsequent skeletonization, width sampling, and color feature extraction.

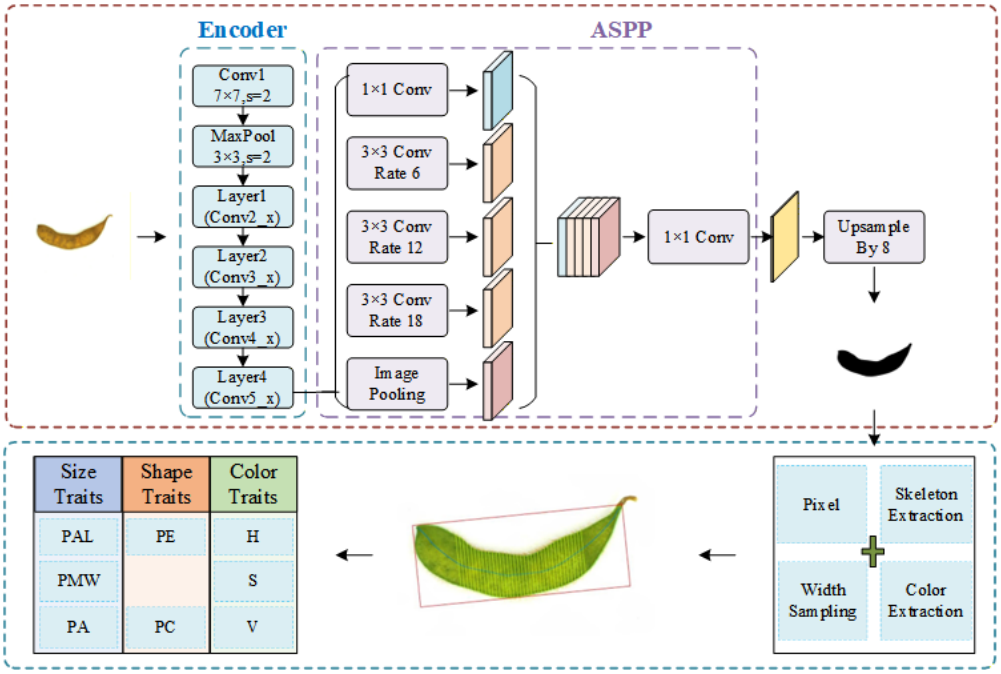

Figure 1. Workflow of DeepLabV3Binary segmentation and soybean pod image-derived trait extraction. DeepLabV3Binary was implemented by adapting DeepLabV3-ResNet50 for binary pod foreground segmentation. Based on the resulting mask, eight image-derived traits were extracted, including size-related traits (PAL, PMW, and PA), morphology-related traits (PE and PC), and color-related traits (H, S, and V). The red rectangle represents the bounding rectangle of the segmented pod. The abbreviations and detailed definitions of all image-derived traits are provided in Table 1.

The model training data were obtained from four image datasets generated using the same image acquisition procedure as that used for subsequent phenotypic extraction, comprising a total of 188 pod images, each derived from a distinct genotype and individual plant. All images were manually annotated with polygon labels using LabelMe software, and binary masks were generated, with the pod region defined as the foreground and the remaining region as the background. Model training was performed using the AdamW optimizer, and the best-performing model was selected according to validation IoU. The annotated images were randomly divided at the image level into training and validation sets at a ratio of 9:1. A combined loss function consisting of Focal Loss [29] and Dice Loss [30] was used for model optimization:

$$L = L_{\text{focal}} + 2L_{\text{dice}}. \tag{1}$$

The Focal Loss was defined as follows:

$$L_{\text{focal}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \alpha \cdot (1-p_i)^{\gamma} \cdot \log(p_i) + (1-y_i) \cdot (1-\alpha) \cdot p_i^{\gamma} \cdot \log(1-p_i)\right], \tag{2}$$

which was used to reduce the dominance of background regions and easily classified pixels during model optimization. Here, N denotes the total number of pixels in a single input image; $y_i \in 0, 1$ represents the ground-truth label of the i-th pixel, with 1 indicating the pod foreground and 0 indicating the background; γ is the focusing parameter used to adjust the weight assigned to hard and easy samples; and α is the balancing parameter. The Dice Loss was defined as follows:

$$L_{\text{dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i p_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \epsilon}. \tag{3}$$

This term was used to improve the consistency of the foreground boundary and the overall segmentation mask. In this equation, $\sum_{i=1}^{N} y_i p_i$ represents the intersection between the predicted mask and the ground-truth region, whereas $\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i$ represents the sum of their areas. The parameter ϵ is a smoothing term, which was set to 1×10^{-6} in this study.

Model performance was evaluated using IoU, the Dice coefficient, precision, and recall, and the model with the highest validation IoU was selected for subsequent phenotypic extraction. The best validation IoU was 0.9774, with a Dice coefficient of 0.9885, precision of 0.9880, and recall of 0.9892 (Table S1), indicating that the segmentation performance was sufficient for extracting the main pod region and for supporting subsequent image-based phenotypic extraction.

During model inference, input images were first subjected to image enhancement and size normalization consistent with the training stage, and pod foreground probability maps were then obtained using the DeepLabV3Binary model. After being mapped back to the original image size, each probability map was binarized using a fixed threshold. Connected-component analysis was then performed to retain only the largest pod body region, thereby removing background noise and small incorrectly segmented regions. Morphological processing was subsequently applied to the binary mask to obtain a continuous and complete pod region. This post-processing procedure helped reduce the influence of local segmentation errors on the calculation of image-derived traits such as skeleton, width, and curvature. Based on the final binary mask, eight pod image-derived traits were extracted for genetic analysis, including PAL, PMW, PA, PE, PC, H, S, and V. PAL, PMW, PA, PE, and PC were derived from mask geometry, skeleton structure, and width-sampling information, whereas H, S, and V were calculated from HSV color values within the segmented pod region. The definitions of these traits are shown in Table 1.

Table 1. Definitions of soybean pod image-derived traits used for genetic analysis.

Trait	Image-Based Calculation Definition	Image Phenotypic Interpretation
PAL	Length of the longest main skeleton path	Main-axis scale of the pod based on the skeleton path
PMW	Mean vertical width sampled from the central region of the skeleton	Main lateral scale based on middle-section skeleton sampling
PA	Mask pixel area converted using the scale ruler	Two-dimensional projected area of the pod
PE	PAL/PMW	Degree of pod elongation
PC	Mean skeleton curvature	Overall bending degree of the pod
H	Mean HSV hue value within the mask region	Surface hue of the pod
S	Mean HSV saturation value within the mask region	Color saturation of the pod
V	Mean HSV value within the mask region	Color brightness of the pod

For each genotype, measurements from multiple replicated samples were averaged to reduce random errors caused by differences among individual samples. The phenotypic values for 2017, 2018, and multi-environment best linear unbiased predictors (BLUPs) were then organized for subsequent trait distribution analysis, heritability estimation, and genome-wide association analysis.

2.4. Statistical Analysis

Descriptive statistics and Pearson correlation analyses were performed for the phenotypic traits using R version 4.5.3. Variance components were estimated by fitting linear mixed models with the R package lme4. The field plot of each genotype within each replication and year was considered the experimental unit, and the mean value of five sampled pods from each plot was used for statistical analysis. Broad-sense heritability (H^2) across environments was calculated as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GE}^2}{n_E} + \frac{\sigma_e^2}{n_E \times n_r}}. \quad (4)$$

Broad-sense heritability within each individual environment was calculated as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2}, \quad (5)$$

where σ_G^2, σ_{GE}^2, and σ_e^2 represent the genotypic variance, genotype-by-environment interaction variance, and residual error variance, respectively; n_E denotes the number of environments; and n_r denotes the number of replications. Following Chang et al. [13], environment, replication nested within environment, genotype, and genotype-by-environment interaction were treated as random effects in the across-environment analysis. For the individual-environment analysis, genotype and replication were treated as random effects. Residuals were assumed to be independent and normally distributed with homogeneous variance.

2.5. Genotypic Data

The soybean association population was genotyped using restriction site-associated DNA sequencing (RAD-Seq). The sequencing procedure, SNP calling, and quality-control pipeline followed the methods described by Li et al. [31]. Briefly, sequencing reads were aligned to the Williams 82 reference genome, SNPs were called, heterozygous genotypes were treated as missing data before imputation with fastPHASE, and SNPs with a minor allele frequency (MAF) below 0.05 were removed. After completion of the quality-control procedure, 61,541 high-quality SNP markers from 187 genotypes were retained for GWAS.

2.6. Genome-Wide Association Analysis of Pod-Related Traits

Genome-wide association analysis was performed using GAPIT 4.1. For the eight pod image-derived traits, including PAL, PMW, PA, PE, PC, H, S, and V, phenotypic values from 2017, 2018, and the multi-environment BLUPs were used as inputs, and four models, namely MLM, CMLM, FarmCPU, and BLINK, were applied. Compared with the previous MLM-based GWAS, four models were applied here to evaluate the robustness of associations for the refined image-derived traits.

A significance threshold of 1×10^{-4} was used to screen marker trait associations, and the GWAS results were visualized using Manhattan plots. This threshold was used for the initial detection of potentially associated SNPs, which were subsequently evaluated based on their reproducibility across phenotypic datasets. For multiple SNPs located at nearby physical positions for the same trait, adjacent signals were consolidated into the same association locus, and the SNP with the smallest p-value was selected as the lead SNP.

Reproducible association signals were evaluated within each GWAS model. Because the multi-environment BLUPs were estimated by integrating phenotypic data from 2017 and 2018, they were not considered independent environmental evidence. Therefore, the results from 2017 and 2018 were used as the primary independent evidence, whereas the BLUP-based results were used as auxiliary evidence from the integrated phenotype. An association locus was considered stable when it reached the significance threshold of $p < 1 \times 10^{-4}$ in both 2017 and 2018, or in either single-year phenotype together with the BLUP phenotype. Nearby SNPs representing the same association region were consolidated, and the SNP with the smallest p-value was selected as the lead SNP of the corresponding stable locus.

2.7. Identification of Putative Candidate Genes for Pod-Related Traits

Candidate genomic intervals were defined according to the physical positions of lead SNPs at stable associated loci. Following our previous image-based GWAS of soybean pod traits, in which the candidate-gene interval was selected with reference to the LD pattern of the mapping population [13], annotated genes located within 500 kb upstream and downstream of each lead SNP were retrieved from the Williams 82 reference genome. This interval was used as a standardized candidate-gene search window for subsequent functional annotation and candidate gene prioritization.

Putative candidate genes were further prioritized by integrating SoyBase gene annotation and Gene Ontology (GO) biological process information. Gene annotation and reported soybean trait-associated gene or locus information were obtained from SoyBase-related resources. Because positional genes were extracted using Wm82.a1.v1 gene IDs, whereas GO annotations were based on Wm82.a2.v1 gene IDs, gene identifiers were converted using the reference genome correspondence table from SoyBase/GlycineMine. GO biological process terms related to development, cell expansion, cell elongation, cell division, cell proliferation, cell wall organization, cellulose and polysaccharide metabolism, plant hormone response, pigment metabolism, and oxidation–reduction processes were considered potentially relevant to soybean pod-related traits.

Pod-related expression profiles were used as supporting evidence for candidate gene evaluation. Expression data from one cm pod, pod shell at 10 DAF, and pod shell at 14 DAF were examined. For each candidate gene, the mean and maximum expression values across the three pod-related tissues were calculated. The genome-wide mean expression value across these tissues was also calculated, and twice this value was used as a reference threshold to indicate relatively strong pod-related expression support. In addition, SoyBase association records overlapping the candidate regions were checked as additional supporting evidence.

3. Results

3.1. Phenotypic Distribution and Descriptive Statistics of Pod-Related Traits

A total of eight image-derived phenotypic traits were obtained from mature pod images for genetic analysis, including size-related traits PAL, PMW, and PA; morphological traits PE and PC; and color-related traits H, S, and V. All traits showed continuous distributions in the 2017, 2018, and multi-environment BLUP datasets, indicating typical quantitative trait characteristics (Table 2; Figure S1). Descriptive statistics showed that the degree of variation differed among traits. Among the size-related traits, PA showed the greatest population variation, with coefficients of variation of 21.95% and 23.38% in 2017 and 2018, respectively. The coefficients of variation for PAL were 13.26% and 13.49%, whereas those for PMW were 10.47% and 10.71%, respectively. In contrast, the morphological traits showed relatively lower variation. The coefficients of variation for PE were 9.54% and 8.62% in 2017 and 2018, respectively, and those for PC were 9.86% and 10.31%, indicating that pod elongation and curvature exhibited continuous variation in this population, although their overall dispersion was lower than that of size-related traits such as PA. The color-related traits showed certain differences between years. The annual mean of H decreased from 42.23 in 2017 to 39.82 in 2018, and that of S decreased from 0.71 to 0.54, whereas the annual mean of V increased from 0.72 to 0.83, suggesting that color-related phenotypes fluctuated markedly across years.

Table 2. Descriptive statistics and heritability of soybean pod image-derived traits under different environments.

Trait	Env.	Mean ± SD	Range	CV (%)	H^2_{Single}	H^2_{Comb}
PAL	2017	5.95 ± 0.79	5.70	13.26	0.80	0.90
	2018	6.09 ± 0.82	4.97	13.49	0.88	
PMW	2017	1.00 ± 0.11	0.63	10.47	0.81	0.92
	2018	1.06 ± 0.11	0.70	10.71	0.84	
PA	2017	4.82 ± 1.06	6.17	21.95	0.85	0.92
	2018	5.20 ± 1.22	7.14	23.38	0.89	
PE	2017	5.96 ± 0.57	4.29	9.54	0.62	0.83
	2018	5.77 ± 0.50	3.42	8.62	0.73	
PC	2017	2.76 ± 0.27	1.69	9.86	0.30	0.53
	2018	2.73 ± 0.28	1.75	10.31	0.33	
H	2017	42.23 ± 3.01	19.77	7.13	0.70	0.44
	2018	39.82 ± 5.46	58.64	13.71	0.73	
S	2017	0.71 ± 0.07	0.51	10.50	0.85	0.76
	2018	0.54 ± 0.10	0.61	17.67	0.84	
V	2017	0.72 ± 0.12	0.65	16.98	0.77	0.76
	2018	0.83 ± 0.12	0.52	14.29	0.80	

Broad-sense heritability estimates revealed clear differences in the stability of different types of image-derived traits (Table 2). In trait order, the across-environment H^2 values of PAL, PMW, and PA were 0.90, 0.92, and 0.92, respectively, indicating high heritability and suggesting that size-related traits were strongly influenced by genotypic effects across the two years. Among the morphological traits, PE also showed relatively high genetic stability, with an across-environment H^2 of 0.83. In contrast, the across-environment H^2 of PC was 0.53, which was markedly lower than those of the size-related traits and PE, indicating that pod curvature was less stable under across-environment conditions. For the color-related traits, the across-environment H^2 values of H, S, and V were 0.44,

0.76, and 0.76, respectively. H showed the lowest heritability, whereas S and V exhibited moderate to relatively high heritability, suggesting that the genetic stability of different color components was inconsistent.

Further correlation analysis showed varying degrees of association among the eight image-derived traits (Figure 2). PAL, PMW, and PA showed stable positive correlations in the 2017, 2018, and BLUP datasets, indicating that these size-related traits jointly reflected variation in pod size. The correlation pattern between PE and the size-related traits differed from those among PAL, PMW, and PA, suggesting that pod elongation was not determined solely by length or area. PC showed generally weak or low correlations with size-related traits, indicating that pod curvature could provide relatively independent morphological information. The color-related traits H, S, and V generally showed low correlations with size- and morphology-related traits. Among the color traits, S and V exhibited a certain positive correlation, whereas the correlations between H and the other color components were relatively unstable. The phenotypic distribution, heritability estimates, and correlation patterns showed that the size-related traits were more stable and more closely correlated than most morphology- and color-related traits.

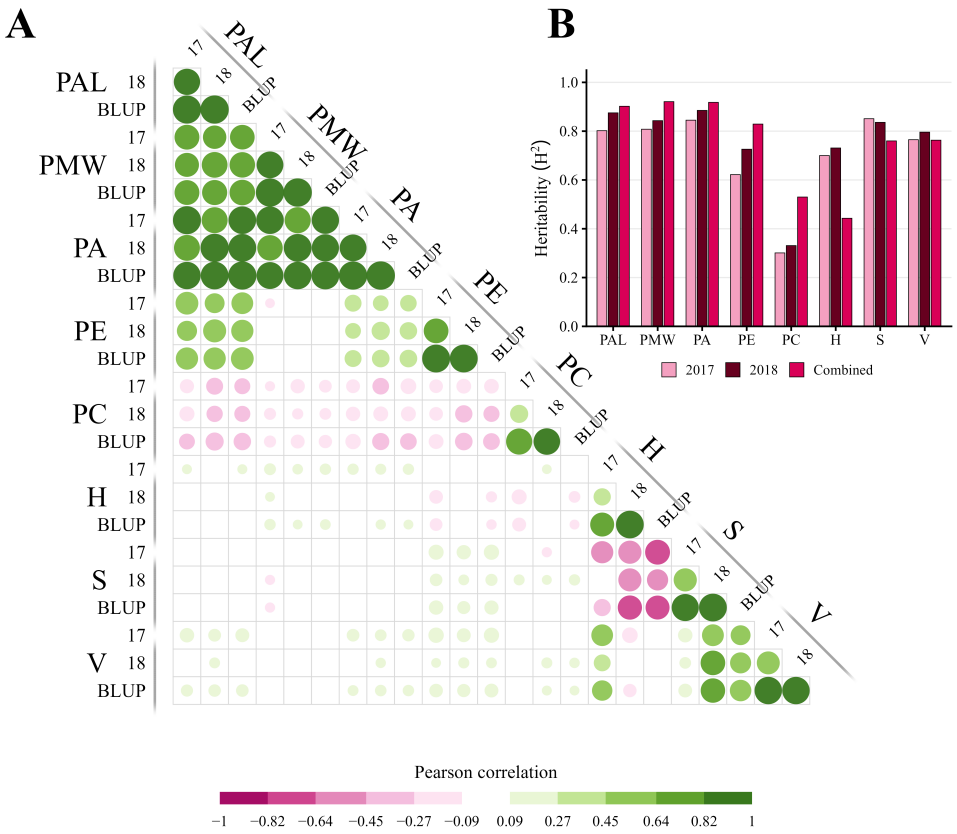

Figure 2. Correlation and heritability analyses of soybean pod image-derived traits. (**A**) Pearson correlation matrices of the eight pod image-derived traits across different phenotypic datasets. (**B**) Broad-sense heritability of each trait in 2017, 2018, and across environments. PAL, PMW, and PA represent size-related traits; PE and PC represent morphological traits; and H, S, and V represent color-related traits.

3.2. Genome-Wide Association Analysis

Based on the phenotypic data from 2017, 2018, and the BLUP dataset, four models, namely MLM, CMLM, FarmCPU, and BLINK, were used to perform GWAS for the eight pod image-derived traits. A significance threshold of $p < 1 \times 10^{-4}$ was applied to identify association signals. The four models differed in both the number and distribution of

detected associations across traits and datasets (Figure 3). MLM, CMLM, FarmCPU, and BLINK yielded 264, 306, 219, and 242 significant SNP detections, respectively, while PAL, PMW, PA, PE, PC, H, S, and V accounted for 157, 105, 94, 105, 63, 116, 235, and 156 detections, respectively. These values represent cumulative detections across the 2017, 2018, and BLUP datasets and may include the same SNP detected in more than one dataset. Although CMLM and MLM detected larger numbers of SNPs overall, inspection of the corresponding Manhattan and QQ plots showed that BLINK produced clearer association peaks while maintaining good agreement between the observed and expected *p*-value distributions. This performance is consistent with previous reports indicating that BLINK can improve statistical power and reduce false-positive associations through linkage disequilibrium-based marker selection and Bayesian information criterion optimization [32]. Therefore, subsequent analyses focused on BLINK.

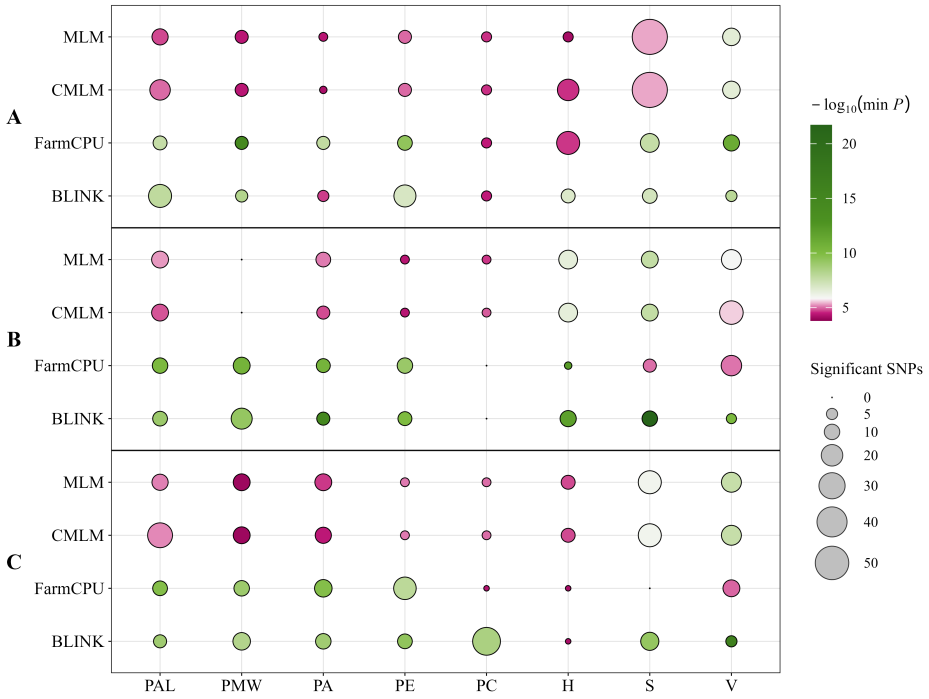

Figure 3. Overview of association signals detected using different GWAS models and phenotypic datasets. (**A**–**C**) Association signals detected using the 2017, 2018, and BLUP phenotypic datasets, respectively. Each circle represents a trait–model–phenotypic dataset combination; circle size indicates the number of SNPs reaching the significance threshold, and color indicates the $-\log_{10}(P)$ value corresponding to the minimum *p*-value in that combination. Darker colors indicate stronger peak signals. This figure was used to compare both the number of detected signals and the peak signal intensity across different models.

To characterize the genomic distribution of association signals, the Manhattan and QQ plots for the eight image-derived traits were examined based on the BLINK model and BLUP phenotypes (Figure 4). A total of 93 SNPs reached the significance threshold in the BLINK-BLUP analysis, including 7, 13, 10, 9, 34, 1, 14, and 5 SNPs for PAL, PMW, PA, PE, PC, H, S, and V, respectively. The strongest association signals for PAL, PMW, PA, and PE were located at Gm19_45415096 on Chr19, Gm14_1830770 on Chr14, Gm16_36289049 on Chr16, and Gm06_44525409 on Chr06, respectively. Among the color-related traits, the strongest signals for S and V were located at Gm08_9303841 on Chr08 and Gm03_743679 on Chr03, respectively, with Gm03_743679 representing the strongest association peak in the BLINK-BLUP analysis. Only one significant SNP was detected for H, whereas 34 SNPs reached the significance threshold for PC. However, the relatively large number of

significant SNPs detected for PC did not necessarily correspond to an equivalent number of independent or stable loci; these signals were therefore further evaluated according to their genomic proximity and reproducibility across the 2017, 2018, and BLUP datasets. Detailed information on the SNPs detected by BLINK across phenotypic datasets is provided in Table S3.

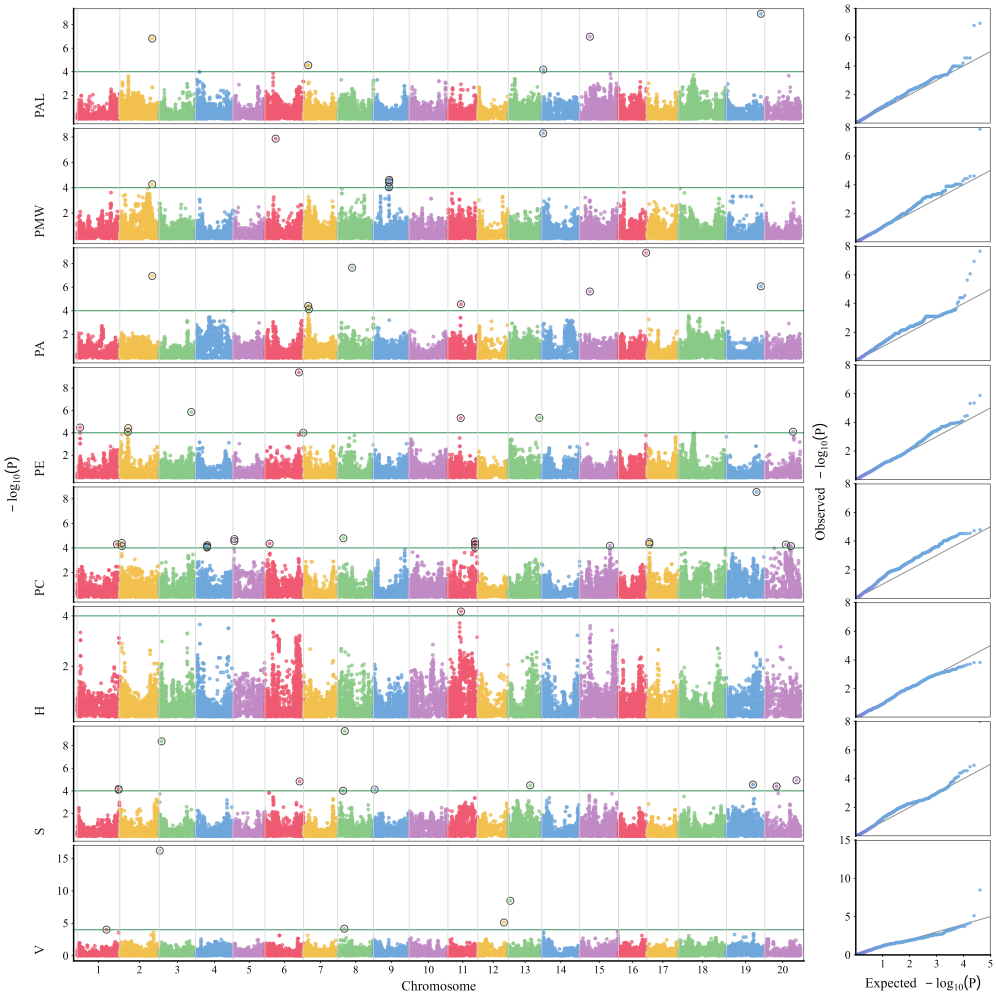

Figure 4. Manhattan plots and QQ plots of the eight soybean pod image-derived traits based on BLUP phenotypes and the BLINK model. Manhattan plots are shown on the left, and the corresponding QQ plots are shown on the right. The eight traits are PAL, PMW, PA, PE, PC, H, S, and V. In the Manhattan plots, the x-axis represents chromosomal position, and the y-axis represents $-\log_{10}(P)$; the green horizontal line indicates the significance threshold, and hollow circles mark associated SNPs reaching this threshold. The QQ plots were used to evaluate the consistency between observed and expected *p*-value in the GWAS results for different traits.

To reduce redundancy among closely linked significant SNPs and improve the reliability of the detected associations, nearby SNPs identified by BLINK were consolidated into genomic loci, and their reproducibility across the 2017, 2018, and BLUP datasets was subsequently evaluated. After locus consolidation and stability screening, 16 stable loci associated with PAL, PMW, PA, PE, S, and V were retained for candidate gene analysis, whereas no stable loci were retained for PC or H.

3.3. Definition of Candidate Intervals and Candidate Gene Prioritization

Based on the 16 stable associated loci identified from the BLINK results, candidate genomic intervals were defined as the ±500 kb regions surrounding the corresponding lead SNPs (Table 3). Table 3 also summarizes the allele composition, minor allele frequency

(MAF), and estimated marker effect of each peak SNP across the phenotypic datasets. Several loci were associated with more than one pod size-related trait. For example, the locus represented by Gm02_42534148 on Chr02 was detected for both PAL and PA, whereas the Chr14 locus represented by Gm14_1830770 was shared by PAL and PMW. Similarly, Gm19_45415096 on Chr19 was associated with both PAL and PA and was detected in both years and in the BLUP dataset for PAL. These shared signals suggest that some stable genomic regions may contribute to multiple pod size-related traits. Among the retained loci, PAL19 and V3 showed particularly consistent support across phenotypic datasets. PAL19 was detected in 2017, 2018, and BLUP, with Gm19_45415096 as the peak SNP in all three datasets. The V3 locus on Chr03 was also detected in 2017, 2018, and BLUP, with Gm03_743679 showing the strongest association in 2018 and BLUP. These loci were subsequently used to extract positional candidate genes, and function-related candidate genes were prioritized based on GO biological process information and SoyBase gene annotation, with pod-related expression profiles used as supporting evidence.

A comparison with the loci reported by Chang et al. [13] showed that eight of the 16 stable BLINK loci overlapped previously reported pod-related loci or genomic regions, whereas the other eight loci were not reported in that study. Among the overlapping loci, Gm14_1830770, Gm16_36289049, and Gm19_45415096 corresponded to the same lead SNPs reported previously, while Gm02_42534148 was located within the same Chr02 genomic region reported by Chang et al. In addition, Gm03_743679 was located within the previously reported Chr03 pod color region. The remaining loci, including Gm06_13926598, Gm09_20961247, Gm07_6071137, Gm08_18944573, Gm01_4181515, Gm06_44525409, Gm11_16816801, and Gm06_45392146, were considered previously unreported relative to Chang et al. [13].

Annotated genes located within the ±500 kb regions surrounding the lead SNPs of the 16 stable BLINK loci were used for putative candidate gene prediction. Based on GO biological process annotations and SoyBase gene annotations, 32 putative candidate gene records were identified from these intervals (Table 4). These candidate records were associated with biological processes related to pod development and growth, cell expansion or elongation, cell division or proliferation, cell wall and polysaccharide metabolism, plant hormone response, pigment-related processes, and oxidation–reduction processes.

Among the 32 putative candidate gene records, those associated with PAL, PMW, and PA were mainly related to oxidative stress, redox homeostasis, cell wall or polysaccharide metabolism, DNA replication, cytokinesis, embryo development, and hormone response. Several gene IDs appeared in more than one size-related trait. *Glyma02g37160* and *Glyma02g36720* were detected in both the PAL2 and PA2 intervals, whereas *Glyma19g38820* and *Glyma19g37820* were shared by PAL19 and PA19. The Chr14 region shared by PAL14 and PMW14 contained genes annotated with DNA replication, cellulose biosynthesis, lipopolysaccharide biosynthesis, and cytokinesis. *Glyma06g17320* in the PMW6 interval was annotated with response to hormone and showed relatively high expression in the examined pod-related tissues, with a Pod mean of 41.00 and a Pod max of 53.00.

For PE, six candidate gene records were obtained, mainly involving cell redox homeostasis, hormone response, and cell cycle regulation. *Glyma11g19780* in the PE11 interval showed the highest expression among all candidate records, with a Pod mean of 59.33 and a Pod max of 79.00, and was annotated with cell cycle. *Glyma11g20490*, another gene in the same interval, was annotated with response to hormone. For the color-related traits, the S6 and V3 intervals mainly pointed to redox-related processes. *Glyma06g42130* in S6 was associated with cell redox homeostasis and showed relatively high expression in pod-related tissues, whereas *Glyma03g01010* and *Glyma03g01020* in V3 were annotated with response to oxidative stress.

Table 3. Stable associated loci, peak SNPs, and marker statistics detected across phenotypic datasets.

Trait	Loci	Chr	Peak SNP	Position (bp)	Alleles	MAF	$-\log_{10}(P)$	Effect	Env.
PAL	*qPAL2*	2	Gm02_42781688	42,781,688	T/A	0.31	5.87	−0.16	2017
			Gm02_42534148	42,534,148	A/G	0.08	9.05	−0.36	2018
			Gm02_42534148	42,534,148	A/G	0.08	6.82	−0.26	BLUP
	qPAL14	14	Gm14_1830770	1,830,770	G/A	0.10	8.42	0.34	2018
			Gm14_1830770	1,830,770	G/A	0.10	4.18	0.19	BLUP
	qPAL19	19	Gm19_45415096	45,415,096	C/A	0.23	7.88	0.22	2017
			Gm19_45415096	45,415,096	C/A	0.23	8.23	0.22	2018
			Gm19_45415096	45,415,096	C/A	0.23	8.92	0.24	BLUP
PMW	*qPMW6*	6	Gm06_13880026	13,880,026	C/G	0.48	5.30	0.02	2017
			Gm06_13926598	13,926,598	G/A	0.43	9.27	0.03	2018
			Gm06_13926598	13,926,598	G/A	0.43	7.87	0.02	BLUP
	qPMW9	9	Gm09_20961247	20,961,247	G/A	0.23	7.49	−0.02	2017
			Gm09_20800904	20,800,904	T/C	0.21	4.61	−0.01	BLUP
	qPMW14	14	Gm14_2686637	2,686,637	C/T	0.14	8.27	0.03	2017
			Gm14_2820006	2,820,006	A/C	0.33	4.16	0.02	2018
			Gm14_1830770	1,830,770	G/A	0.10	8.29	0.04	BLUP
PA	*qPA2*	2	Gm02_42534148	42,534,148	A/G	0.08	8.57	−0.44	2018
			Gm02_42534148	42,534,148	A/G	0.08	6.94	−0.34	BLUP
	qPA7	7	Gm07_6986309	6,986,309	T/A	0.45	4.17	−0.18	2017
			Gm07_6071137	6,071,137	G/A	0.28	4.40	0.19	BLUP
	qPA8	8	Gm08_18944573	18,944,573	C/T	0.05	14.66	0.74	2018
			Gm08_18944573	18,944,573	C/T	0.05	7.65	0.41	BLUP
	qPA16	16	Gm16_36289049	36,289,049	T/A	0.21	9.07	0.36	2018
			Gm16_36289049	36,289,049	T/A	0.21	8.91	0.32	BLUP
	qPA19	19	Gm19_45415096	45,415,096	C/A	0.23	4.30	0.15	2017
			Gm19_45415096	45,415,096	C/A	0.23	6.07	0.24	BLUP
PE	*qPE1*	1	Gm01_4181515	4,181,515	G/A	0.07	6.75	0.25	2018
			Gm01_3777842	3,777,842	C/T	0.48	4.47	0.11	BLUP
	qPE6	6	Gm06_44525409	44,525,409	G/A	0.48	4.89	−0.13	2017
			Gm06_44525409	44,525,409	G/A	0.48	9.40	−0.15	BLUP
	qPE11	11	Gm11_16816801	16,816,801	T/A	0.22	4.33	−0.14	2018
			Gm11_16816801	16,816,801	T/A	0.22	5.32	−0.15	BLUP
S	*qS6*	6	Gm06_45392146	45,392,146	T/C	0.31	6.95	−0.02	2017
			Gm06_45391795	45,391,795	C/T	0.27	4.84	−0.02	BLUP
V	*qV3*	3	Gm03_843004	843,004	G/C	0.22	6.28	−0.03	2017
			Gm03_743679	743,679	A/G	0.17	10.20	−0.05	2018
			Gm03_743679	743,679	A/G	0.17	16.21	−0.06	BLUP

Note: Candidate genomic intervals were defined as ±500 kb around the lead SNPs of stable associated loci. The same lead SNP detected for multiple traits was used once when defining candidate intervals. MAF, minor allele frequency.

Expression profiles from one cm pod, pod shell at 10 DAF, and pod shell at 14 DAF showed that 19 of the 32 records were expressed in at least one pod-related tissue, seven had Pod max values of at least 5, and two exceeded twice the genome-wide mean expression level across the three tissues. Seven records overlapped SoyBase region-level association evidence, whereas no direct gene-level evidence was detected in the reported gene field.

Table 4. Function-related candidate gene records within the candidate intervals of the 16 stable associated loci.

Trait	Loci	Lead SNP	Gene ID	Trait-Related GO Biological Process
PA	*qPA2*	Gm02_42534148	*Glyma02g36720*	cellulose biosynthetic process
	qPA2	Gm02_42534148	*Glyma02g37160*	response to oxidative stress
	qPA7	Gm07_6071137	*Glyma07g07270*	lipopolysaccharide biosynthetic process
	qPA16	Gm16_36289049	*Glyma16g33250*	response to oxidative stress
	qPA19	Gm19_45415096	*Glyma19g37820*	cell wall macromolecule catabolic process
	qPA19	Gm19_45415096	*Glyma19g38820*	cell redox homeostasis
PAL	*qPAL2*	Gm02_42534148	*Glyma02g36720*	cellulose biosynthetic process
	qPAL2	Gm02_42534148	*Glyma02g37160*	response to oxidative stress
	qPAL14	Gm14_1830770	*Glyma14g02200*	cytokinesis
	qPAL14	Gm14_1830770	*Glyma14g02860*	DNA replication
	qPAL14	Gm14_1830770	*Glyma14g03190*	lipopolysaccharide biosynthetic process
	qPAL14	Gm14_1830770	*Glyma14g03200*	DNA replication, synthesis of primer
	qPAL14	Gm14_1830770	*Glyma14g03310*	cellulose biosynthetic process
	qPAL19	Gm19_45415096	*Glyma19g37820*	cell wall macromolecule catabolic process
	qPAL19	Gm19_45415096	*Glyma19g38820*	cell redox homeostasis
PE	*qPE1*	Gm01_4181515	*Glyma01g04760*	cell redox homeostasis
	qPE1	Gm01_4181515	*Glyma01g04800*	cell redox homeostasis
	qPE1	Gm01_4181515	*Glyma01g04810*	cell redox homeostasis
	qPE11	Gm11_16816801	*Glyma11g19780*	cell cycle
	qPE11	Gm11_16816801	*Glyma11g20490*	response to hormone
	qPE6	Gm06_44525409	*Glyma06g41610*	cell redox homeostasis
PMW	*qPMW6*	Gm06_13926598	*Glyma06g17320*	response to hormone
	qPMW9	Gm09_20961247	*Glyma09g17200*	embryo development
	qPMW14	Gm14_1830770	*Glyma14g02200*	cytokinesis
	qPMW14	Gm14_1830770	*Glyma14g02860*	DNA replication
	qPMW14	Gm14_1830770	*Glyma14g03190*	lipopolysaccharide biosynthetic process
	qPMW14	Gm14_1830770	*Glyma14g03200*	DNA replication, synthesis of primer
	qPMW14	Gm14_1830770	*Glyma14g03310*	cellulose biosynthetic process
S	*qS6*	Gm06_45392146	*Glyma06g41610*	cell redox homeostasis
	qS6	Gm06_45392146	*Glyma06g42130*	cell redox homeostasis
V	*qV3*	Gm03_743679	*Glyma03g01010*	response to oxidative stress
	qV3	Gm03_743679	*Glyma03g01020*	response to oxidative stress

Note: Candidate gene records were selected from the ±500 kb intervals surrounding the lead SNPs of the 16 stable associated loci. Function-related gene records were identified based on GO biological process annotations related to development or growth, cell expansion or elongation, cell division or proliferation, cell wall and polysaccharide metabolism, plant hormone response, pigment-related processes, and oxidation–reduction processes.

4. Discussion

4.1. Image-Based Phenotyping Expands Soybean Pod-Related Trait Characterization

High-throughput image-based phenotyping has become an important approach for quantitative plant trait analysis because it enables rapid, objective, and reproducible measurements that are often difficult to obtain manually [3–5]. In soybean, previous studies have mainly focused on seed phenotyping or on extracting relatively basic pod characteristics. Baek et al. [9], for example, developed an image-analysis approach for high-throughput measurement of soybean seed traits. Our previous study established an image-based framework for mature three-seeded soybean pods and quantified basic descriptors of pod size and color, including pod length, width, area, length-to-width ratio, and color-related variables [13]. More recent studies have further improved soybean pod segmentation and seed-per-pod estimation [11] or developed keypoint-detection

methods for fine-grained pod phenotyping [12]. Building on these studies, the present work focused on extracting refined mask- and skeleton-based descriptors of mature pod morphology.

Compared with the basic pod measurements used in previous studies, the image-derived traits examined here represent different structural dimensions of pod shape. PAL was calculated along the pod skeleton rather than as a straight-line distance and therefore retains information on the curved axial path of the pod. PMW describes the mean width of the central pod region rather than only the maximum width, while PE reflects proportional morphology and PC characterizes pod curvature. These measurements therefore extend phenotypic characterization beyond conventional pod length, width, area, and aspect ratio. In particular, skeleton- and mask-based analysis makes it possible to quantify local and nonlinear structural features that are difficult to define consistently using routine manual measurements. Thus, the contribution of the present framework lies not simply in automating existing measurements, but also in expanding the range of pod characteristics that can be quantitatively described.

The different image-derived traits also showed distinct levels of stability under the present experimental conditions. PAL, PMW, PA, and PE exhibited relatively high heritability, suggesting that pod size, axial structure, and proportional morphology were captured consistently by the imaging workflow. In contrast, PC showed lower heritability, indicating that curvature may be more sensitive to developmental heterogeneity, local segmentation variation, or environmental effects. The HSV components also differed in their stability. S and V showed greater across-environment stability than H, whereas hue may be more sensitive to variation in maturity, drying conditions, illumination, and image preprocessing. Therefore, the biological interpretation of image-derived traits should consider not only their geometric or color definitions, but also the potential sources of variation introduced during sample preparation and image acquisition.

DeepLabV3Binary was used in this study as a practical segmentation tool under standardized imaging conditions. The present study did not include a direct comparison with manual phenotyping or conventional image-processing methods, particularly because skeleton-based axial length, centrally sampled width, and curvature-related descriptors do not yet have widely accepted manual reference protocols. Accordingly, the current results demonstrate the feasibility of automatically extracting refined pod traits and incorporating them into quantitative genetic analysis, rather than establishing the superiority of the proposed workflow over existing approaches. Further evaluation using independent soybean populations, diverse pod types, developmental stages, and imaging conditions will be required to assess the general applicability and reproducibility of the framework.

4.2. Genetic Insights from Image-Derived Soybean Pod-Related Traits

Previous genetic studies of soybean pod-related traits have mainly focused on pod number, maturity, pod wall characteristics, pod dehiscence, and pod color [18–23,33], whereas the genetic basis of quantitative pod size and shape has received comparatively less attention. Chang et al. [13] addressed this gap by combining image-based phenotyping with GWAS and linkage mapping and identified loci for pod length, width, area, length-to-width ratio, and color, demonstrating that basic image-derived pod traits are suitable for genetic mapping. In the present study, 16 stable loci were retained for PAL, PMW, PA, PE, S, and V after genomic consolidation and reproducibility screening across the 2017, 2018, and BLUP datasets. The reproducible associations detected for PAL, PMW, PA, and PE were consistent with their relatively high heritability, indicating that these refined image-derived traits captured genetically stable variation in pod axial scale, lateral scale,

projected area, and proportional morphology. In contrast, no stable loci were retained for PC or H, probably because curvature and hue were more sensitive to pod deformation, maturity, drying conditions, illumination, and image preprocessing under the present experimental conditions.

Comparison with Chang et al. [13] revealed both corresponding and distinct association signals. The stable regions represented by Gm02_42534148, Gm14_1830770, Gm16_36289049, and Gm19_45415096 corresponded to pod-related regions reported previously, while the V3 interval on chromosome 3 coincided with a known pod-color region. These correspondences support the reliability of the present mapping results because refined image-derived traits recovered genomic regions previously associated with conventional pod measurements. Gm19_45415096 was particularly notable because it was repeatedly associated with PAL across annual and BLUP datasets and was also located within a previously reported pod-related region. However, several loci associated with PMW, PE, S, and part of PA were not reported by Chang et al. [13]. This difference is likely related to phenotype definition rather than inconsistency between studies. Chang et al. mainly analyzed conventional descriptors such as total pod length, maximum width, area, length-to-width ratio, and overall color, whereas PMW represents the mean width sampled from the central pod region and PE reflects the proportional relationship between axial and lateral growth. Alternative measurements of the same underlying phenotype can produce overlapping but nonidentical GWAS signals [8]; therefore, the additional loci detected here may represent complementary genetic information revealed by increased phenotypic resolution.

Several stable regions were shared by more than one size-related trait. The loci represented by Gm02_42534148, Gm14_1830770, and Gm19_45415096 were associated with PAL, PMW, or PA, consistent with the strong correlations among pod axial length, width, and projected area. Chang et al. [13] similarly reported colocalization between pod-area and pod-length loci, including PA2/PL2-2, PA7/PL7, PA14/PL14, PA15/PL15, and PA19/PL19-1. Such shared associations may result from pleiotropic genes affecting several dimensions of pod growth or from tightly linked genes independently controlling correlated traits. In addition, the qPAL19/qPA19 region overlapped previously reported QTL intervals for seed width, seed height, and seed weight, qPA7 coincided with a seed-length QTL, and qPA16 overlapped intervals for seed length and seed length-to-width ratio [19,20]. Similar pod-seed QTL correspondence was also observed by Chang et al. [13], suggesting that some genomic regions may participate in the coordinated development of pod and seed dimensions. Nevertheless, these overlaps represent regional correspondence rather than direct evidence that the same causal genes control both organs, and further fine mapping, haplotype analysis, and functional validation are required.

4.3. Putative Functions of Candidate Genes for Soybean Pod-Related Traits

Candidate genes within the stable loci were interpreted by integrating GO biological process annotations, SoyBase functional records, pod-related expression profiles, and previously reported association evidence. Previous studies of plant organ development have shown that variation in organ size and shape commonly results from coordinated regulation of cell proliferation, cell expansion, cell wall remodeling, and hormone signaling [34–41]. Consistent with this general framework, the candidate records identified here repeatedly involved cell wall metabolism, cellulose biosynthesis, DNA replication, cytokinesis, hormone response, growth coordination, and redox regulation. However, because most of these genes have not been functionally validated for soybean pod traits,

their biological relevance should be interpreted as positional and functional support rather than direct evidence of causality.

For the size- and morphology-related traits, the candidate records were mainly associated with processes controlling cell division, tissue expansion, and coordinated organ growth. The loci shared by PAL, PMW, and PA contained genes related to cell wall modification and redox homeostasis, which is consistent with the strong phenotypic relationships among pod axial length, width, and projected area. Plant organ dimensions are determined by the balance between cell proliferation and subsequent cell expansion, while auxin- and brassinosteroid-related pathways contribute to cell wall loosening, tissue differentiation, and directional growth [34–36]. *Glyma06g17320* in the PMW6 interval was annotated as hormone-responsive and showed relatively high expression in pod-related tissues, making it a plausible candidate for pod width variation. For PE, which represents proportional morphology rather than absolute size, genes involved in cell-cycle regulation and growth coordination may be particularly relevant. *Glyma11g19780* showed the highest pod-related expression among the candidate records and was associated with cell-cycle-related functions, whereas *Glyma11g20490* was linked to hormone response. These annotations are consistent with a possible role in coordinating axial and lateral pod growth, although they do not yet demonstrate direct regulation of PE.

The candidate genes associated with the color-related traits S and V were predominantly linked to redox homeostasis and oxidative-stress responses. *Glyma06g42130* in the S6 interval was annotated with cell redox homeostasis and showed relatively high pod-related expression, whereas *Glyma03g01010* and *Glyma03g01020* in the V3 interval were associated with responses to oxidative stress. Previous studies have shown that mature plant tissue color can be influenced not only by pigment biosynthesis but also by senescence, oxidation, tissue dehydration, and browning processes. Therefore, the redox-related annotations in the S6 and V3 intervals provide a biologically plausible explanation for variation in mature pod saturation and brightness. The correspondence between the V3 interval and the previously reported pod-color region on chromosome 3 further supports the relevance of this genomic region. Nevertheless, the present evidence is still based on genomic position, annotation, expression profiles, and previous association records.

5. Conclusions

This study developed a deep learning-assisted image phenotyping framework to characterize and genetically dissect soybean pod-related traits. Eight image-derived traits representing pod size, morphology, and color captured refined phenotypic dimensions beyond conventional pod measurements, among which PAL, PMW, PA, and PE showed relatively high heritability. GWAS identified 16 stable loci associated with PAL, PMW, PA, PE, S, and V. Several loci overlapped previously reported pod- or seed-related regions, whereas additional loci detected for the refined traits suggested complementary genetic information. The integration of functional annotation, SoyBase information, and pod-related expression profiles enabled the prioritization of candidate genes potentially involved in pod trait variation. These findings support the use of refined image-derived phenotypes for dissecting the genetic basis of soybean pod-related traits, although the identified loci and candidate genes require validation in independent populations and functional studies.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/agronomy16151454/s1, Figure S1: Phenotypic distributions of the eight soybean pod image-derived traits across the 2017, 2018, and BLUP datasets; Table S1: Segmentation performance of the DeepLabV3Binary model; Table S2: Significant trait–SNP associations identified by GWAS; Table S3: BLINK-detected trait-associated loci across phenotypic datasets.

Author Contributions: L.W. analyzed the data and drafted the manuscript; S.S. participated in data analysis and result interpretation; F.C. designed the overall project and revised the manuscript; T.Z. provided scientific supervision, contributed to study design and interpretation of results, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Doctoral Research Foundation of Gansu Agricultural University, grant number GAU-KYQD-2022-32.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available within the article and Supplementary Materials. Additional data and code are available from the corresponding author upon reasonable request.

Acknowledgments: The work was supported by the High Performance Computing Platform of Gansu Computing Center.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, Y.; Jing, C.; Wang, X.; Lin, C.; Wang, X.; Zhou, Z.S. Photosynthetic characteristics of soybean pod and its contribution to yield. *J. Northeast Agric. Univ.* **2008**, *39*, 51–56.
2. Qiu, L.J.; Chang, R.Z. *Descriptors and Data Standard for Soybean (Glycine spp.)*; China Agriculture Press: Beijing, China, 2006.
3. Chen, D.; Neumann, K.; Friedel, S.; Kilian, B.; Chen, M.; Altmann, T.; Klukas, C. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* **2014**, *26*, 4636–4655. https://doi.org/10.1105/tpc.114.129601.
4. Fahlgren, N.; Gehan, M.A.; Baxter, I. Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* **2015**, *24*, 93–99. https://doi.org/10.1016/j.pbi.2015.02.006.
5. Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.; Yan, J. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* **2020**, *13*, 187–214. https://doi.org/10.1016/j.molp.2020.01.008.
6. Diaz-Garcia, L.; Schlautman, B.; Covarrubias-Pazaran, G.; Maule, A.; Johnson-Cicalese, J.; Grygleski, E.; Vorsa, N.; Zalapa, J. Massive phenotyping of multiple cranberry populations reveals novel QTLs for fruit anthocyanin content and other important chemical traits. *Mol. Genet. Genom.* **2018**, *293*, 1379–1392. https://doi.org/10.1007/s00438-018-1464-z.
7. Guo, Z.; Yang, W.; Chang, Y.; Ma, X.; Tu, H.; Xiong, F.; Jiang, N.; Feng, H.; Huang, C.; Yang, P.; et al. Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. *Mol. Plant* **2018**, *11*, 789–805. https://doi.org/10.1016/j.molp.2018.03.018.
8. Gage, J.L.; de Leon, N.; Clayton, M.K. Comparing genome-wide association study results from different measurements of an underlying phenotype. *G3 Genes Genomes Genet.* **2018**, *8*, 3715–3722. https://doi.org/10.1534/g3.118.200700.
9. Baek, J.; Lee, E.; Kim, N.; Kim, S.L.; Choi, I.; Ji, H.; Chung, Y.S.; Choi, M.S.; Moon, J.K.; Kim, K.H. High throughput phenotyping for various traits on soybean seeds using image analysis. *Sensors* **2020**, *20*, 248. https://doi.org/10.3390/s20010248.
10. Uzal, L.C.; Grinblat, G.L.; Namías, R.; Larese, M.G.; Bianchi, J.S.; Morandi, E.N.; Granitto, P.M. Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agric.* **2018**, *150*, 196–204. https://doi.org/10.1016/j.compag.2018.04.024.
11. Yang, S.; Zheng, L.; Wu, T.; Sun, S.; Zhang, M.; Li, M.; Wang, M. High-throughput soybean pods high-quality segmentation and seed-per-pod estimation for soybean plant breeding. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107580.
12. Liu, F.; Liu, H.; Wu, Q.; Guo, S.; Yu, J.; Liu, Y. Pod-pose: An efficient top-down keypoint detection model for fine-grained pod phenotyping in mature soybean. *Plant Methods* **2025**, *21*, 82.
13. Chang, F.; Lv, W.; Lv, P.; Xiao, Y.; Yan, W.; Chen, S.; Zheng, L.; Xie, P.; Wang, L.; Karikari, B.; et al. Exploring genetic architecture for pod-related traits in soybean using image-based phenotyping. *Mol. Breed.* **2021**, *41*, 28.
14. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 29. https://doi.org/10.1186/1746-4811-9-29.
15. Tian, D.; Wang, P.; Tang, B.; Teng, X.; Li, C.; Liu, X.; Zou, D.; Song, S.; Zhang, Z. GWAS Atlas: A curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic Acids Res.* **2020**, *48*, D927–D932. https://doi.org/10.1093/nar/gkz828.

16. Würschum, T. Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* **2012**, *125*, 201–210. https://doi.org/10.1007/s00122-012-1887-6.
17. Yano, M.; Tuberosa, R. Genome studies and molecular genetics-from sequence to crops: Genomics comes of age. *Curr. Opin. Plant Biol.* **2009**, *12*, 103–106. https://doi.org/10.1016/j.pbi.2009.01.001.
18. Guo, G.Y.; Sun, R.; Hou, M.; Guo, Y.X.; Xin, D.W.; Jiang, H.W.; Liu, C.Y.; Hu, G.H.; Chen, Q.S. Quantitative trait locus (QTL) analysis of pod related traits in different environments in soybean. *Afr. J. Biotechnol.* **2011**, *10*, 11848–11854.
19. Fang, C.; Ma, Y.; Wu, S.; Liu, Z.; Wang, Z.; Yang, R.; Hu, G.; Zhou, Z.; Yu, H.; Zhang, M.; et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **2017**, *18*, 161. https://doi.org/10.1186/s13059-017-1289-9.
20. Yang, Z.; Xin, D.; Liu, C.; Jiang, H.; Han, X.; Sun, Y.; Qi, Z.; Hu, G.; Chen, Q. Identification of QTLs for seed and pod traits in soybean and analysis for additive effects and epistatic effects of QTLs among multiple environments. *Mol. Genet. Genom.* **2013**, *288*, 651–667. https://doi.org/10.1007/s00438-013-0779-z.
21. Zhang, D.; Cheng, H.; Wang, H.; Zhang, H.; Liu, C.; Yu, D. Identification of genomic regions determining flower and pod numbers development in soybean (Glycine max L.). *J. Genet. Genom.* **2010**, *37*, 545–556. https://doi.org/10.1016/S1673-8527(09)60074-6.
22. Hu, D.; Kan, G.; Hu, W.; Li, Y.; Hao, D.; Li, X.; Yang, H.; Yang, Z.; He, X.; Huang, F.; et al. Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. *Front. Plant Sci.* **2019**, *10*, 811. https://doi.org/10.3389/fpls.2019.00811.
23. Bandillo, N.B.; Lorenz, A.J.; Graef, G.L.; Jarquin, D.; Hyten, D.L.; Nelson, R.L.; Specht, J.E. Genome-wide association mapping of qualitatively inherited traits in a germplasm collection. *Plant Genome* **2017**, *10*, 18. https://doi.org/10.3835/plantgenome2016.06.0054.
24. He, Q.; Yang, H.; Xiang, S.; Tian, D.; Wang, W.; Zhao, T.; Gai, J.; Singh, R. Fine mapping of the genetic locus L1 conferring black pods using a chromosome segment substitution line population of soybean. *Plant Breed.* **2015**, *134*, 437–445. https://doi.org/10.1111/pbr.12272.
25. Dong, Y.; Yang, X.; Liu, J.; Wang, B.H.; Liu, B.L.; Wang, Y.Z. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nat. Commun.* **2014**, *5*, 3352. https://doi.org/10.1038/ncomms4352.
26. Zhang, S.; Wang, F.; Hu, X.; Miao, H.; Hong, J.; Shan, S.; Chi, X.; Chen, J.; Zhang, X. Development of innovative image descriptors for phenotyping peanut pod constriction and discovery of QTLs underlying this trait. *J. Integr. Agric.* **2026** . https://doi.org/10.1016/j.jia.2026.02.004
27. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
29. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE: New York, NY, USA; pp. 2980–2988.
30. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*; IEEE: New York, NY, USA, 2016; pp. 565–571.
31. Li, L.; Guo, N.; Niu, J.; Wang, Z.; Cui, X.; Sun, J.; Zhao, T.; Xing, H. Loci and candidate gene identification for resistance to Phytophthora sojae via association analysis in soybean [*Glycine max* (L.) Merr.]. *Mol. Genet. Genom.* **2016**, *291*, 1095–1103. https://doi.org/10.1007/s00438-015-1164-x.
32. Wang, J.; Zhang, Z. GAPIT version 4: Integration of GWAS into genomic prediction. *Mol. Biol. Evol.* **2026**, *43*, 1–8.
33. Song, Q.J.; Marek, L.F.; Shoemaker, R.C.; Lark, K.G.; Concibido, V.C.; Delannay, X.; Specht, J.E.; Cregan, P.B. A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* **2004**, *109*, 122–128. https://doi.org/10.1007/s00122-004-1602-3.
34. Majda, M.; Robert, S. The role of auxin in cell wall expansion. *Int. J. Mol. Sci.* **2018**, *19*, 951.
35. Wang, S.; Chang, Y.; Guo, J.; Chen, J.G. Arabidopsis Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation. *Plant J.* **2007**, *50*, 858–872. https://doi.org/10.1111/j.1365-313X.2007.03096.x.
36. Yang, C.; Ma, Y.; He, Y.; Tian, Z.; Li, J. OsOFP19 modulates plant architecture by integrating the cell division pattern and brassinosteroid signaling. *Plant J.* **2018**, *93*, 489–501. https://doi.org/10.1111/tpj.13793.
37. Duan, P.; Xu, J.; Zeng, D.; Zhang, B.; Geng, M.; Zhang, G.; Huang, K.; Huang, L.; Xu, R.; Ge, S.; et al. Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. *Mol. Plant* **2017**, *10*, 685–694. https://doi.org/10.1016/j.molp.2017.03.009.
38. Liu, J.; Chen, J.; Zheng, X.; Wu, F.; Lin, Q.; Heng, Y.; Tian, P.; Cheng, Z.; Yu, X.; Zhou, K.; et al. GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants* **2017**, *3*, 17043. https://doi.org/10.1038/nplants.2017.43.
39. Snouffer, A.; Kraus, C.; van der Knaap, E. The shape of things to come: Ovate family proteins regulate plant organ shape. *Curr. Opin. Plant Biol.* **2020**, *53*, 98–105. https://doi.org/10.1016/j.pbi.2019.10.005.

40. Wu, S.; Zhang, B.; Keyhaninejad, N.; Rodriguez, G.R.; Kim, H.J.; Chakrabarti, M.; Illa-Berenguer, E.; Taitano, N.K.; Gonzalo, M.J.; Diaz, A.; et al. A common genetic mechanism underlies morphological diversity in fruits and other plant organs. *Nat. Commun.* **2018**, *9*, 4734. https://doi.org/10.1038/s41467-018-07216-8.
41. Xiao, H.; Jiang, N.; Schaffner, E.; Stockinger, E.J.; van der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **2008**, *319*, 1527–1530. https://doi.org/10.1126/science.1153040.