plants

MDPI

Article

Genetic Diversity and Evaluation of Agro-Morphological Traits in Lettuce Core Collection

Suyun Moon [1,†], Onsook Hur [2,†], Seong-Hoon Kim [1], Yoonjung Lee [1], Hyeonseok Oh [1], Jungyoon Yi [1], Ho-Cheol Ko [1], Hee-Jong Woo [1], Nayoung Ro [1,*] and Young-Wang Na [1,*]

[1] National Agrobiodiversity Center, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Republic of Korea; sooym21@korea.kr (S.M.); shkim0819@korea.kr (S.-H.K.); yoon112@korea.kr (Y.L.); zzjiy@korea.kr (H.O.); naaeskr@korea.kr (J.Y.); hchko@korea.kr (H.-C.K.); woo001@korea.kr (H.-J.W.)
[2] Department of Crop Breeding, National Institute of Crop Science, Rural Development Administration, Wanju 55365, Republic of Korea; oshur09@korea.kr
* Correspondence: nonanona@korea.kr (N.R.); ywna@korea.kr (Y.-W.N.)
† These authors contributed equally to this work.

Abstract: Lettuce (*Lactuca sativa*) is a globally significant leafy vegetable, valued for both its economic and nutritional contributions. The efficient conservation and use of the lettuce germplasm are crucial for breeding and genetic improvement. This study examined the genetic diversity and population structure of a core collection of the lettuce germplasm using genotyping by sequencing (GBS). A total of 7136 high-quality single-nucleotide polymorphisms (SNPs) were identified across nine chromosomes. Population analysis through Bayesian clustering and discriminant analysis of principal components (DAPC) revealed three distinct genetic clusters. Cluster 2 exhibited the greatest genetic diversity (*He* = 0.29, *I* = 0.44), while Cluster 3 had high levels of inbreeding (*F* = 0.79). Agro-morphological trait evaluation further identified significant differences in leaf length, plant weight, and head height across clusters. These findings provide valuable insights into the genetic and phenotypic diversity of lettuce, facilitating the development of more robust breeding programs. Additionally, the core collection established in this study offers a representative subset of the lettuce germplasm for future genomic research and conservation efforts.

Keywords: *Lactuca sativa*; germplasm; genetic diversity; population structure; agro-morphological characterization; genotyping by sequencing

check for updates

Citation: Moon, S.; Hur, O.; Kim, S.-H.; Lee, Y.; Oh, H.; Yi, J.; Ko, H.-C.; Woo, H.-J.; Ro, N.; Na, Y.-W. Genetic Diversity and Evaluation of Agro-Morphological Traits in Lettuce Core Collection. *Plants* **2024**, *13*, 3552. https://doi.org/10.3390/plants13243552

Academic Editor: Andreas W. Ebert

Received: 11 November 2024
Revised: 8 December 2024
Accepted: 18 December 2024
Published: 19 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The genus *Lactuca*, belonging to the family Asteraceae, comprises approximately 150 species that are primarily distributed across warm and temperate regions of the Northern Hemisphere, including Europe, Asia, Indonesia, and North and Central America [1]. Among these species, *Lactuca sativa* L., commonly referred to as lettuce, stands out due to its widespread cultivation and significant economic value [2]. As one of the most consumed leafy vegetables globally, lettuce reached a total production of approximately 27.1 million tonnes in 2022 [3]. Additionally, several wild *Lactuca* species, such as *L. serriola* L., *L. saligna* L., and *L. virosa* L., are of commercial interest, offering advantageous traits that are utilized as genetic resources in lettuce breeding programs [4,5]. Furthermore, *L. indica* L. has been traditionally used in folk medicine due to its medicinal properties [6,7].

Lettuce is classified into six horticultural types based on morphological characteristics: leaf, crisphead, romaine, butterhead, stem, and oilseed [8]. These morphological traits, especially those related to leaves and heads, are crucial for classification and directly impact the quality and yield of lettuce crops [9]. Environmental factors, including temperature, light quality, and nutrition, along with inherent genetic differences, significantly influence these morphological characteristics [8,10–14]. Consequently, understanding the genetic

diversity within lettuce is essential for elucidating the interactions between genotypes and phenotypes, which is critical for improving various characteristics of the crop.

To assess genetic variation and diversity within *Lactuca* species, various molecular markers such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), and Simple Sequence Repeats (SSRs) have been employed [15–19]. However, these markers often provide limited genomic coverage, which can be insufficient for advanced breeding programs due to their limited number and uneven distribution across the genome [20]. Addressing these limitations, single-nucleotide polymorphisms (SNPs) have emerged as powerful tools in genomic analysis, offering extensive coverage and ubiquitous distribution across the genome [21,22]. The advent of high-throughput sequencing technologies has further facilitated the rapid and cost-effective discovery of SNPs [23]. Among these technologies, the genotyping-by-sequencing (GBS) approach has been widely applied in leafy crops, including spinach, chicory, cabbage, and kale [24–28]. This approach enables efficient genotyping of large populations, thereby advancing marker-assisted breeding and genomic studies in plant breeding programs [29,30].

In managing large germplasm collections, the presence of redundancy often poses challenges in terms of efficient management and evaluation [31]. The development of core collections has been proposed as a solution, offering a representative subset that captures the full range of genetic diversity within a germplasm collection [32,33]. These core collections facilitate in-depth genetic and phenotypic analyses while simplifying germplasm management and enhancing the efficiency of breeding programs [34].

Studies on developing core collections have been conducted in various plant species, particularly in economically important crops. For horticultural crops, core collections have been established in species such as spinach, amaranth, cabbage, and cauliflower [15,35–37]. However, research on the development of core collections for lettuce remains limited. One study aimed to enhance the understanding of genetic diversity assessment and core collection development by comparing various sampling methods using *L. sativa* accessions [38].

In this study, we investigated the genetic diversity and population structure within a 300-core collection of the lettuce germplasm. Additionally, we assessed the phenotypic diversity of this core collection by evaluating 17 qualitative and quantitative agromorphological traits. Through this comprehensive analysis, we aim to enhance the understanding of genetic and phenotypic variability within the lettuce germplasm, thereby supporting future breeding efforts and improving resource management.

2. Materials and Methods

2.1. Plant Materials and DNA Extraction

A collection of the 2001 *Lactuca* germplasm, including 2 accessions of *L. altaica* Fisch. & C.A.Mey., 3 of *L. indica* L., and 1 accession each of *L. livida* Boiss. & Reut. and *L. saligna* L., along with 1982 accessions of *L. sativa* L. and 12 of *L. serriola* L., was utilized to develop the core collection. Detailed information on sample ID, species, origin, and accession type is provided in Table S1. The lettuce was grown at the research farm of the National Agrobiodiversity Center (NAC), Rural Development Administration (RDA), Jeonju (35°49′18″ N, 127°08′56″ E), Republic of Korea, during 2022–2024, following the RDA-recommended lettuce cultivation method. Total genomic DNA was extracted from lyophilized seedling samples using a Genomic DNA Prep Kit (Inclone Biotech, Yongin, Korea) according to the manufacturer's protocol.

2.2. GBS Library Preparation and Sequencing

The GBS libraries were constructed using the ApeKI restriction enzyme, following a previously established protocol with minor modifications [39]. The Illumina DNA Prep kit (Illumina Inc., San Diego, CA, USA) was utilized for library preparation according to the manufacturer's instructions. The quality and fragment size distribution of the libraries were assessed using a Bioanalyzer equipped with a High Sensitivity DNA Chip (Agilent

Technologies, Santa Clara, CA, USA). The GBS libraries were sequenced on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA).

2.3. SNP Discovery, Filtering, and Genotype Imputation

After sequencing, raw reads were demultiplexed based on barcode sequences using Stacks v2.60 [40] with the "process_radtags" function. The reads then underwent quality control with Trimmomatic v0.39 [41] to remove adapter sequences and low-quality reads, ensuring the retention of high-quality data for subsequent analysis. The cleaned reads were aligned to the reference genome of *Lactuca sativa* [42] using the Burrows-Wheeler Aligner (BWA) v0.7.17 [43]. Picard v2.26.4 (https://broadinstitute.github.io/picard/ accessed on 11 May 2022) was employed to add read group information, which is essential for proper duplicate marking and further processing.

To enhance the accuracy of SNP identification, the Genome Analysis Toolkit (GATK) v4.2.0.0 [44] was used to perform local realignment of reads. A summary of the sequencing results is presented in Table S2. Following realignment, GATK HaplotypeCaller was utilized for variant calling, generating an initial set of SNPs and indels. The resulting variants were further filtered using VCFtools to obtain a high-confidence set of SNPs [45]. The filtering criteria included retaining only biallelic SNPs, requiring a minimum minor allele frequency (MAF) of 0.05, ensuring a mean read depth (meanDP) of at least 5, and excluding variants with more than 5% missing genotype data. The SNPs that were not assembled onto the chromosome were manually filtered out. The number of filtered SNPs and heterozygous genotype ratio for each chromosome are presented in Table S3.

To further enhance the dataset's completeness and utility for downstream analyses, genotype imputation was performed using the BEAGLE algorithm [46], with 1000 burn-in and 1000 main iterations.

2.4. Morphological Characterization

For the assessment of agro-morphological traits, ten plants per genotype were grown. Three representative plants, chosen based on their phenotypic characteristics, were evaluated for each genotype. Seventeen agro-morphological traits were recorded during field inspections, including plant growth type, leaf shape, leaf color, the density of leaf margin incisions, leaf blistering, head formation, head shape in the longitudinal section, head size, head density, leaf length (cm), leaf width (cm), leaf thickness (mm), plant weight (g), head height (cm), head diameter (cm), days to 50% bolting, and days to 50% flowering (Table S4). With the exception of days to 50% bolting and flowering, all traits were investigated at harvest stage. All plants were evaluated at the same time to ensure consistency. Qualitative traits related to the head, such as head shape in the longitudinal section, head size, and head density, were measured only in cases where the head formation trait exhibited semi-heading or heading types. To ensure accuracy and consistency in the data, accessions that exhibited poor growth were excluded from quantitative trait measurements. For each trait, the values were reported as the mean of triplicate determinations.

2.5. Development of Lettuce Core Collection

To develop the core collection of 2001 *Lactuca* germplasm, we analyzed 7136 high-quality SNPs obtained through GBS and stringent filtering, along with 9 qualitative morphological traits. We focused on these nine traits because quantitative traits are often heavily influenced by environmental factors, which can obscure the genetic relationships among accessions. Initially, 400 germplasm accessions were selected based on the genetic distance matrix using the R package Core Hunter [47]. From these 400 lettuce germplasm accessions, a representative subset of 300 accessions was selected as the core collection by considering qualitative traits and applying the advanced maximization (M) strategy, which was implemented through a modified heuristic algorithm in PowerCore software [48]. To compare the genetic diversity between entire and core collections based on qualitative agro-morphological traits, the Shannon–Weaver diversity index (I'), Nei's genetic diversity

index (*H'*), and genetic evenness (*J*) were calculated, following Hennink and Zeven [49] and Pielou [50].

2.6. Analysis of Population Structure

Bayesian-based clustering was performed using STRUCTURE v2.3.4 [51] by testing ten independent runs with K ranging from 1 to 10. Each run included a burn-in period of 10,000 iterations followed by 10,000 Markov Chain Monte Carlo (MCMC) iterations, assuming the admixture model. The optimal number of subpopulations and subtle clustering patterns were identified using StructureSelector [52] via the ΔK method [53]. A membership coefficient (q) greater than 0.6 was used to assign samples to specific clusters, while samples with membership coefficients < 0.6 were classified as 'genetically admixed'. The admixture proportions of each lettuce accession, estimated by STRUCTURE, were visualized using STRUCTURE PLOT v2.0 [54].

The genetic structure was further analyzed by the Discriminant Analysis of Principal Components (DAPC) using the R package adegenet [55]. The "find.clusters" function was used to detect the number of genetic clusters in the population. The best number of subpopulations in the DAPC was indicated by an elbow in the curve of the Bayesian information criterion (BIC). A cross-validation function "xvalDapc" was used to confirm the correct number of PCs to be retained and then the analysis was rerun with this number.

Genetic distances between pairs of accessions were calculated, and an unrooted unweighted pair group method with arithmetic means (UPGMA) was constructed using TASSEL5 [56]. The resulting unrooted UPGMA tree was visualized using Interactive Tree Of Life (iTOL) v6 [57].

2.7. Analysis of Molecular Variance (AMOVA) and Genetic Diversity Indices

The number of subpopulations determined based on the DAPC analysis was used for calculating AMOVA, the PhiPT index, and Nei's genetic distance using GenAlEx v6.51b2 [58]. To further verify the genetic differentiation, the number of different (*Na*), and effective alleles (*Ne*), the expected heterozygosity (*He*), the fixation index (*F*), Shannon's information index (*I*), and pairwise PhiPT values were calculated for each population. To determine whether variance component partitioning was significant, we estimated the probability values by using 999 permutations.

2.8. Statistical Analysis

Data summarization and descriptive statistics on agro-morphological data were performed using the R v4.4.1 [59]. Significant differences between mean values were computed using the Student–Newman–Keuls multiple comparison test ($\alpha = 0.05$).

3. Results

3.1. Distribution of SNPs in Lactuca sativa Genome

Approximately 6.9 billion reads were generated from the 2001 *Lactuca* germplasm accessions using the Illumina NovaSeq 6000 platform, with an average mapping depth of 7.52× per accession. A summary of these sequencing results is provided in Table S2. For the identification of polymorphic loci, only SNPs with a minor allele frequency (MAF) of ≥0.05 and a mean depth of coverage (meanDP) of ≥5 were considered. A total of 7136 high-quality SNPs were physically mapped across nine chromosomes (Table S3). Chromosome 4 harbored the highest number of SNPs (1076 SNPs), whereas chromosome 6 contained the fewest (451 SNPs). These high-quality SNPs were distributed across the *L. sativa* genome, with variable densities observed along different chromosomal regions, as illustrated in Figure 1.

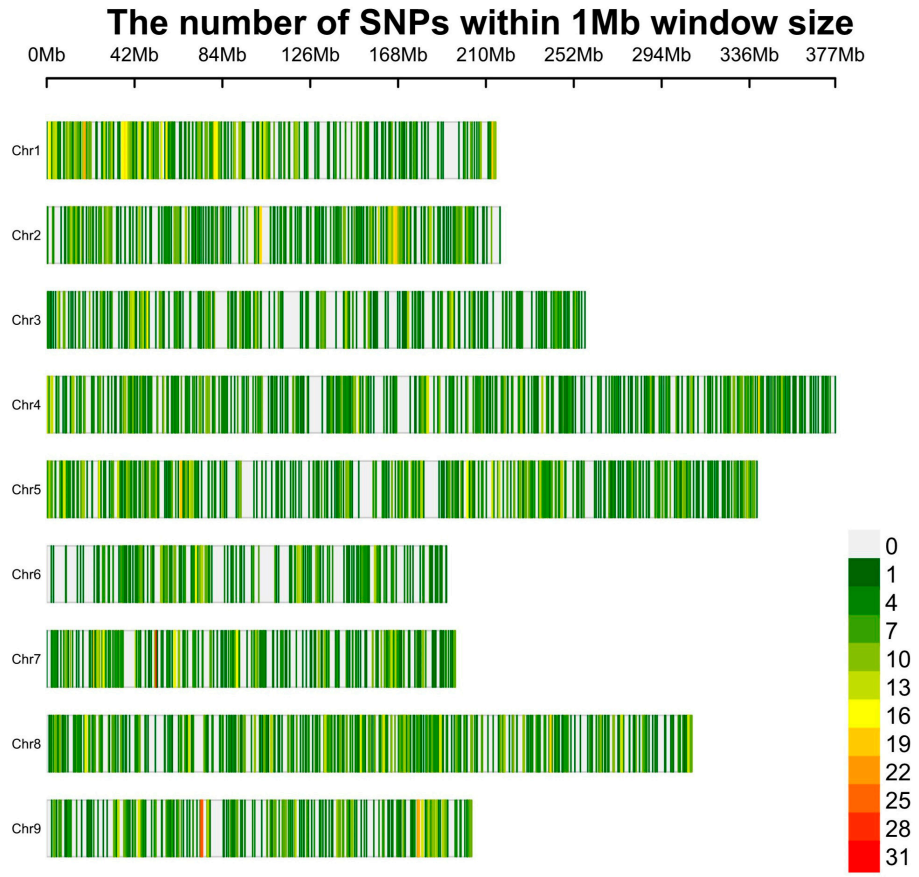


Figure 1. The distribution of 7136 high-quality SNPs across the 9 *Lactuca sativa* chromosomes. The horizontal axis represents the chromosome length in Mb. Different colors correspond to SNP density.

3.2. Development of Core Collection

Genetic distances between 2001 *Lactuca* accessions were calculated using Core Hunter based on the 7136 SNPs, leading to the initial selection of 400 representative accessions. Among these 400 accessions, accessions that were admixed or exhibited poor growth were excluded. Subsequently, qualitative phenotypic data were incorporated to refine this selection using the M strategy, resulting in a final core collection of 300 accessions. PCA of these 300 accessions in the core collection showed nearly identical patterns to those of the 2001 accessions in the entire collection (Figure 2A). The resulting core collection was then used for further analyses.

Phenotypic characterization of 17 agro-morphological traits was performed for the lettuce core collection. The frequency and relative percentage of qualitative agro-morphological traits are summarized in Table 1. The majority of the accessions were categorized as leaf type, with lesser representation from other types such as romaine, butterhead, crisphead, and stem. The predominant leaf shapes within the germplasm were elliptic and broad elliptic. Regarding leaf color, green and red were the two most common colors observed. Most accessions displayed medium blistering on leaves, followed by weak and strong blistering. In terms of head formation, accessions were classified as non-heading, heading, or semi-heading types. Traits such as head shape in the longitudinal section, head size, and head density were evaluated for the semi-heading and heading types. Circular head shapes were the most common, followed by elliptic and broad elliptic. Medium-sized heads were most prevalent, and head density ranged from absent and loose to medium, dense, and very dense categories. Additionally, the genetic diversity and evenness between the entire and core collections based on qualitative agro-morphological traits were compared (Table S5). The mean values for the Shannon–Weaver diversity index (*I′*) and Nei's genetic diversity index (*H′*) in the core collection (0.94 and 0.70, respectively) were similar to those of the entire collection

(1.09 and 0.78, respectively). Moreover, the mean genetic evenness (J) values for the core and entire collections were also comparable (0.61 and 0.69, respectively). These results demonstrate that the developed core collection adequately represents the morphological characteristics of the entire collection.

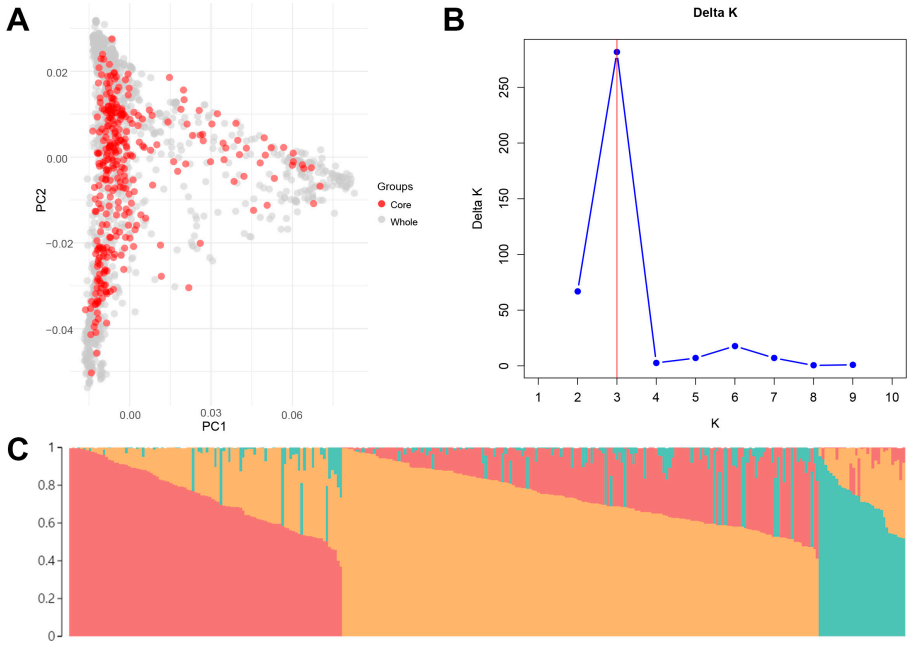


Figure 2. The development and population structure analysis of the lettuce core collection. (**A**) Principal component analysis (PCA) of the core collection. The red dots represent accessions in the core collection, while the gray dots represent accessions not included in the core collection. (**B**) Delta K values for different assumed population numbers (K) in the STRUCTURE analysis. (**C**) The inferred population structure for K = 3. Individual membership coefficients were sorted within each cluster.

For quantitative agro-morphological data, the minimum, maximum, average, standard deviation, and coefficient of variation (CV) values are presented in Table 2. The average values for leaf length, leaf width, and leaf thickness were 25.32 cm, 18.65 cm, and 0.21 mm, respectively. Plant weight showed the widest variation, ranging from 118.00 g to 1940.33 g. Regarding the height and diameter of the head, the average values were 28.92 cm and 18.87 cm, respectively. The average number of days to bolting and flowering were 79.20 and 102.58 days, respectively. Five of the eight traits had relatively high CV values above 20%, with the highest CV observed for plant weight (38.36%) and the lowest for days to flowering (8.75%).

Table 1. Summary of qualitative agro-morphological traits of lettuce core collection.

Traits	Frequency	Percentage	Traits	Frequency	Percentage
Plant growth type			Leaf blistering		
Leaf	186	62.00	Absent	7	2.33
Romaine	27	9.00	Weak	90	30.00
Butterhead	47	15.67	Medium	152	50.67
Crisphead	34	11.33	Strong	50	16.67
Stem	6	2.00	Very strong	1	0.33
Leaf shape			Head formation		
Triangular	8	2.67	Non-heading	190	67.38
Narrow elliptic	10	3.33	Semi-heading	6	2.13
Elliptic	106	35.33	Heading	86	30.50
Broad elliptic	89	29.67	Head shape in		
Circular	39	13.00	longitudinal section		

Table 1. *Cont.*

Traits	Frequency	Percentage	Traits	Frequency	Percentage
Obovate	29	9.67	Elliptic	22	25.00
Oblate	7	2.33	Broad elliptic	20	22.73
Broad oblate	5	1.67	Circular	40	45.45
Broad obtrullate	7	2.33	Oblate	5	5.68
			Other	1	1.14
Leaf color			Head size		
Yellow	3	1.00	Very small	9	10.11
Green	196	65.33	Small	31	34.83
Greyish green	4	1.33	Medium	34	38.20
Cyan	13	4.33	Large	15	16.85
Red	84	28.00	Head density		
Density of			Absent	18	20.22
leaf margin incisions			Loose	12	13.48
Absent or sparse	219	73.00	Medium	17	19.10
Medium	66	22.00	Dense	16	17.98
Dense	15	5.00	Very dense	26	29.21

Table 2. Summary of quantitative agro-morphological traits of lettuce core collection.

Traits	No. of Germplasm	Min	Max	Average	SD	CV (%)
Leaf length (cm)	262	10.17	46.00	25.32	6.46	25.51
Leaf width (cm)	262	9.00	29.67	18.65	3.40	18.23
Leaf thickness (mm)	262	0.10	0.90	0.21	0.08	38.10
Plant weight (g)	262	118.00	1940.33	518.60	198.96	38.36
Head height (cm)	261	10.67	60.00	28.92	7.65	26.45
Head diameter (cm)	262	10.00	70.33	18.87	4.78	25.33
Days to bolting	262	62.00	103.00	79.20	8.11	10.24
Days to flowering	262	82.00	133.00	102.58	8.98	8.75

CV, coefficient of variation.

3.3. Population Structure and Genetic Relationships

To investigate the population structure of the lettuce core collection, the Bayesian-based clustering analysis was conducted. The ΔK analysis showed that the highest ΔK value was K = 3, suggesting the presence of three main populations in the lettuce core collection (Figure 2B,C). Population 1, 2, and 3 consisted of 72, 130, and 24 accessions, respectively. Seventy-four accessions with a membership coefficient below 0.6 were identified as a genetically admixed population. Furthermore, population structure analysis confirmed that the three clusters observed in the core collection are also present in the whole collection (Figure S1).

The lowest BIC value was found at K = 4, but elbowed at K = 3, in the DAPC (Figure 3A). The cluster analysis classified the accessions into three clusters following the elbow method. The result of cross-validation for the remaining PCs was optimized to the 20 first components, where it had highest mean successful assignment and the lowest root mean squared error (RMSE) (Figure 3B). For the subsequent DAPC, twenty PCs and two discriminant functions were retained. The resulting scatterplot of the lettuce core collection is shown in Figure 3C. The grouping assignment for individual accessions by the DAPC is listed in Table S4. Cluster 1, 2, and 3 consisted of 91, 170, and 39 accessions, respectively. Among the three clusters, cluster 2 had the highest number of admixed individuals (39 accessions). The unweighted pair group method with arithmetic means (UPGMA) analysis of the lettuce core collection is presented in Figure 4. The resulting phylogenetic tree reveals significant genetic diversity both within and between the clusters. For further genetic analysis, the clusters of the DAPC were used because it separated the lettuce core collections from the populations of the Bayesian-based clustering in more detail.

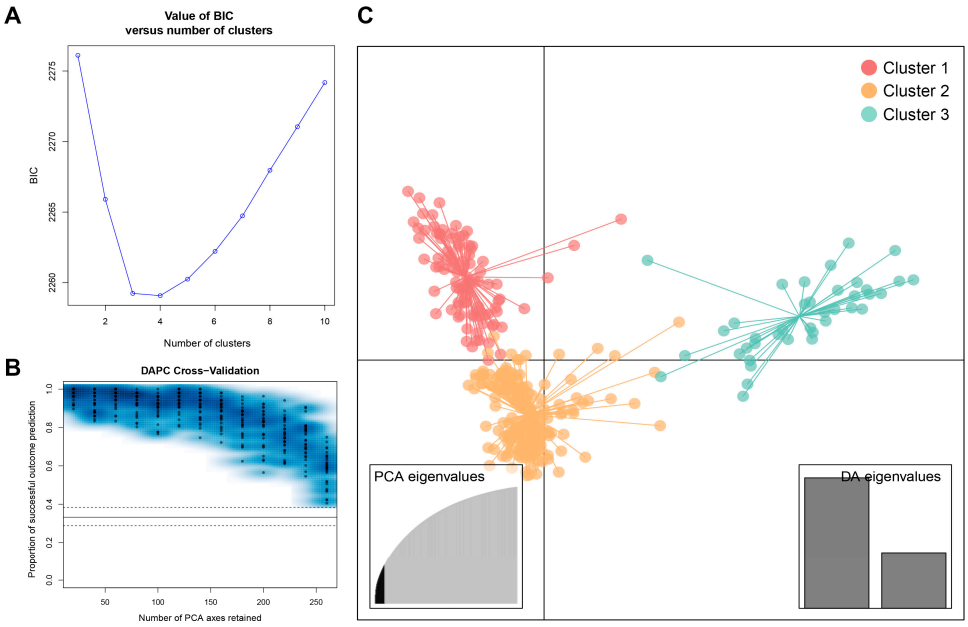


Figure 3. Discriminant analysis of principal components (DAPC) results of lettuce core collection. (**A**) Inference of optimal number of clusters for lettuce core collection based on BIC values. (**B**) DAPC cross-validation determining optimal number of principal components (PCs) retained for analysis of three predefined groups. (**C**) Scatterplot of individuals projected onto first twenty principal components (PCs) from DAPC. The colors represent the three clusters identified: red (Cluster 1), orange (Cluster 2), and teal (Cluster 3). Eigenvalues from analysis are displayed in inset.

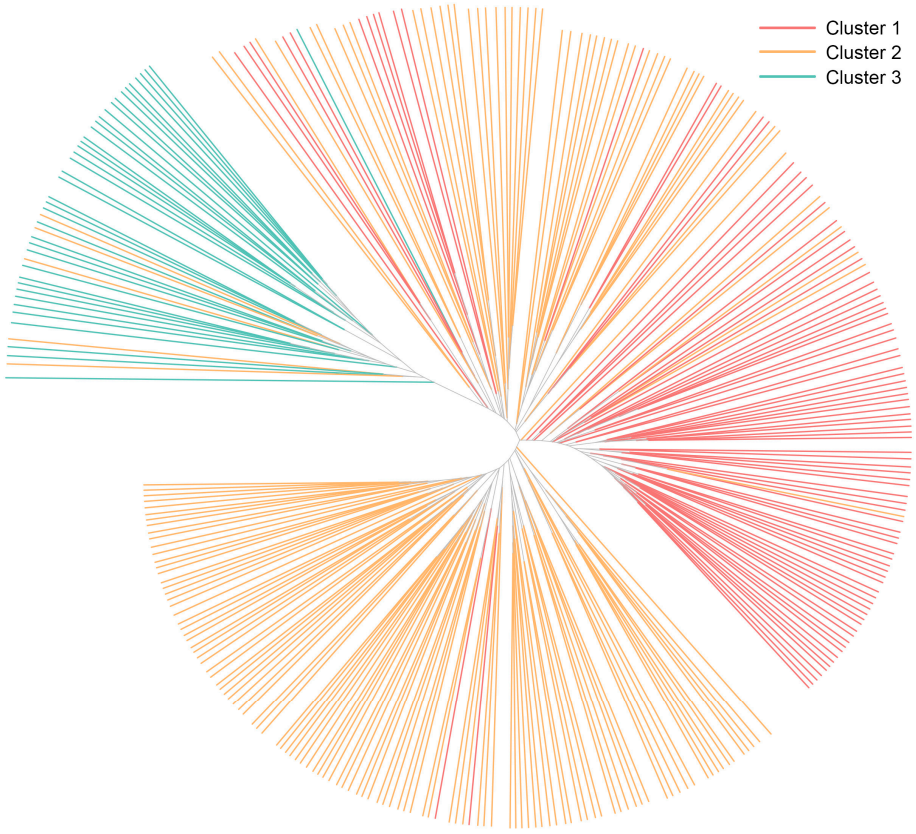


Figure 4. The unrooted UPGMA phylogenetic tree based on the genetic distance matrix representing the grouping of the lettuce core collection.

3.4. Genetic Differentiation and Genetic Diversity Across Populations

The genetic variability within and among populations was assessed by using analysis of molecular variance (AMOVA). The AMOVA results indicated that 14% of the total genetic variation was partitioned among populations, while a substantial 86% of the variation was found within populations (Table 3). The genetic differentiation among populations was found to be moderate (with PhiPT = 0.14), with significance at $p < 0.001$ and a gene flow value (Nm) of 1.56. This result suggests that genetic differentiation is more pronounced within subpopulations than among them. The pairwise PhiPT genetic distance values ranged from 0.04 between Cluster 1 and Cluster 2 to 0.12 between Cluster 1 and Cluster 3, indicating varying degrees of genetic divergence between the clusters (Table 4). Notably, the PhiPT values were statistically significant ($p < 0.001$) for all pairwise comparisons between subpopulations, underscoring the genetic distinctiveness of each cluster.

Table 3. Analysis of molecular variance (AMOVA), genetic differentiation (PhiPT), and gene flow.

Source	df	SS	MS	Est. Var.	%	PhiPT	Nm
Among Populations	2	104,454.03	52,227.02	569.25	14%	0.14 **	1.56
Within Populations	297	1,056,730.75	3558.02	3558.02	86%		
Total	299	1,161,184.78		4127.27	100%		

** *p*-value < 0.001. df, degree of freedom; SS, sum of squares; MS, mean of squares; Est. Var., estimated variance; %, percentage of genetic variation.

Table 4. Pairwise PhiPT and genetic distances between populations.

Subpopulation	Cluster 1	Cluster 2	Cluster 3
Cluster 1	-	0.04	0.12
Cluster 2	0.09	-	0.09
Cluster 3	0.25	0.18	-

PhiPT values (above diagonal) and Nei's genetic distance (blow diagonal) between.

The grand mean values for the number of different alleles (*Na*) and the number of effective alleles (*Ne*) across the three subpopulations were 1.96 and 1.45, respectively (Table 5). The mean values for the entire population in terms of the expected heterozygosity (*He*), the unbiased diversity index (*uHe*), and Shannon's information index (*I*) were 0.28, 0.28, and 0.43, respectively. Among the three clusters, Cluster 2 exhibited the highest genetic diversity, as indicated by its slightly higher values for the *He* (0.29), *uHe* (0.29), and *I* (0.44), compared to Clusters 1 and 3. The fixation index (*F*), which measures the degree of inbreeding within each population, ranged from 0.77 to 0.79, indicating a consistently high level of genetic structuring across the clusters.

Table 5. Genetic diversity parameters of lettuce core collections.

Population	*N*	*Na*	*Ne*	*He*	*uHe*	*I*	*F*
Cluster 1	91	1.98	1.43	0.26	0.27	0.41	0.77
Cluster 2	170	1.99	1.47	0.29	0.29	0.44	0.79
Cluster 3	39	1.90	1.47	0.28	0.28	0.42	0.79
Mean		1.96	1.45	0.28	0.28	0.43	0.78

N, sample size; *Na*, mean number of different alleles; *Ne*, mean number of effective alleles; *He*, expected heterozygosity; *uHe*, unbiased diversity index; *I*, Shannon's information index; *F*, fixation index.

3.5. Phenotypic Variation Analysis Across Population

The phenotypic variation among the three clusters identified in the lettuce core collection was assessed using various agro-morphological traits, as shown in Figure 5 and Table S6. Significant differences were observed across the clusters for several traits, indicating distinct phenotypic profiles. Cluster 3 exhibited the longest average leaf length (29.80 cm), which was

significantly greater than that of Clusters 1 (21.92 cm) and 2 (26.54 cm) ($p < 0.001$). However, leaf width and thickness did not show significant differences across the clusters. Cluster 2 had a higher average plant weight compared to the other two clusters, with a value of 571.96 g. Significant variation was also observed in head height, with Cluster 3 having the tallest heads on average (34.26 cm), while head diameter did not differ significantly across the clusters. Similarly, the average number of days to bolting and flowering were similar across all clusters.

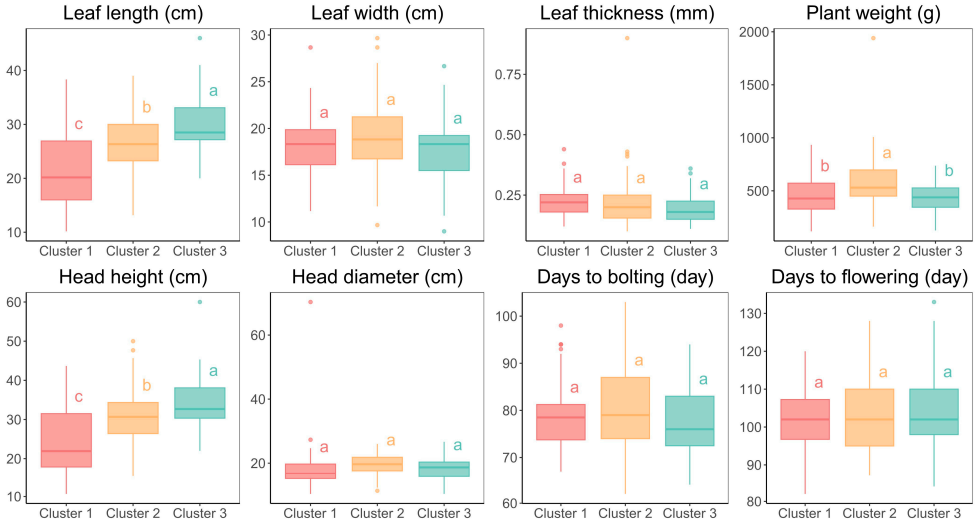


Figure 5. Boxplots showing the variation in quantitative agro-morphological traits across clusters. Different letters indicate significant differences between groups ($p < 0.05$).

4. Discussion

Studying genetic diversity and population genetics is essential for the conservation and sustainable use of crop germplasms [60,61]. In this study, a core collection of the *Lactuca* germplasm was developed by integrating high-density genotypic data with agro-morphological data. This approach enhances research efficiency and supports breeding programs, ensuring that the genetic diversity required for future agricultural innovation is preserved and utilized effectively [62,63]. Our findings provide valuable insights for future breeding efforts, laying the foundation for genome-wide association studies (GWASs) and marker-assisted selection in lettuce.

GBS analysis identified 7136 high-quality SNPs distributed across nine chromosomes, offering a comprehensive overview of genetic variation. These SNPs were crucial in revealing the genetic relationships among *Lactuca* accessions. The variability in SNP counts across studies reflects methodological differences, such as filtering stringency and data handling. Missing data are a common challenge in GBS-based research, often resulting in a significant reduction in the number of SNPs [64,65]. For instance, Park et al. [66] and Lee et al. [67] reported vastly different SNP numbers due to varied filtering criteria, particularly in the handling of missing genotype data. By applying stringent criteria, we ensured data reliability while minimizing the potential loss of data.

The first core collection in plants was established in chickpea based on phenotypic data and geographical distribution [68]. The evolution of core collections from simple phenotypic assessments to advanced genetic integrations exemplifies progress in germplasm management [69–74]. Given that using only genotypic or phenotypic information for establishing core collections may not efficiently capture the full genetic diversity of a species [75], we constructed a lettuce core collection through the integration of SNP data and agro-morphological data. Additionally, our lettuce core collection was refined by excluding admixed and poorly performing accessions, because admixed accessions may introduce genetic redundancy, while those with poor growth performance can distort the evaluation metrics of the core collection,

potentially compromising breeding outcomes [34,76]. This refinement not only preserves genetic variation but also enhances its utility for breeding [77,78].

Although all wild *Lactuca* species accessions were excluded from the final core collection due to genetic redundancy and poor growth performance, additional accessions from other *Lactuca* species were initially included to capture a broader spectrum of genetic diversity. Based on their cross-compatibility and the fertility of F1 hybrids, *Lactuca* species can be categorized into the lettuce gene pool [79]. The primary gene pool, which is fully interfertile, includes *L. sativa*, *L. altaica*, and *L. serriola*, while the secondary gene pool consists solely of *L. saligna* [5]. Historically, *L. altaica* has been considered conspecific with *L. serriola* [17,80]. However, recent findings on morphological traits, phytochemical profiles, and plastid phylogeny suggest distinct differences, warranting further phylogenetic analyses to clarify its taxonomic status [81,82]. Similarly, *L. livida* is regarded as conspecific with *L. virosa* due to the continuity of their morphometric and biological traits [83]. *L. virosa*, in turn, is classified as a member of the tertiary gene pool. For *L. indica*, it has been reported as relatively distantly related to *L. sativa*, despite some shared genetic characteristics [15,81]. However, the limited sample sizes of wild species compared to *L. sativa* may introduce bias in these conclusions. To achieve a comprehensive understanding of genetic diversity and evolutionary relationships, broader sampling of wild *Lactuca* species is necessary.

Population structure analysis, employing Bayesian clustering and discriminant analysis of principal components (DAPC), identified three distinct genetic clusters (Figures 2 and 3). Similar results have been reported with target region amplification polymorphism markers that were used to analyze fifty-four cultivars [84]. These clusters exhibited varying levels of genetic diversity, with Cluster 2 displaying the highest genetic heterogeneity and Cluster 3 emerging as the most genetically distinct. Although the UPGMA analysis also assigned the lettuce core collection to three major clusters, discrepancies were observed when compared with Bayesian clustering and DAPC regarding the assignment of accessions to specific groups (Figure 4). The distance-based nature of the UPGMA method may not effectively address compositional heterogeneity and rate variation, potentially leading to systematic clustering errors. In contrast, model-based approaches such as Bayesian clustering and DAPC compensate for these complexities, ensuring more reliable results [85,86]. The presence of admixed accessions, especially among cultivated varieties, suggests extensive hybridization and gene flow, likely driven by breeding practices and natural cross-pollination events within *L. sativa* and occasionally through hybridization with wild relatives such as *L. serriola*, *L. saligna*, and *L. virosa* [87,88].

The AMOVA revealed that the majority of genetic diversity was found within populations rather than between them (Table 3). This pattern of higher within-population variance compared to among-population variation has also been observed in several other crops, such as rutabaga, rice, and wheat [89–91]. Wright [92] noted that gene flow (Nm) values greater than 1 generally indicate sufficient migration to prevent significant genetic divergence between populations. Thus, the high gene flow (Nm = 1.56) and relatively low genetic differentiation (PhiPT = 0.14) observed in this study likely explain the distribution of genetic diversity within and among populations. Cluster-specific genetic distances further emphasize the influence of breeding history and local adaptation, with Cluster 3's distinctiveness likely stemming from its high proportion of landraces, which have adapted to specific environments.

Significant phenotypic variations in traits such as leaf length, plant weight, and head height were observed, highlighting the potential for targeted breeding. In contrast, traits like leaf width, head diameter, and the average number of days to bolting and flowering exhibited limited variation between clusters, suggesting that these traits may be less influenced by the genetic diversity observed (Figure 5, Table S6). The observed phenotypic diversity provides a foundation for future studies, including QTL mapping and GWASs, to identify loci linked to agronomically important traits.

Previous research has already demonstrated the value of these approaches in uncovering genetic determinants of traits, including post-harvest shelf life, disease resistance,

and bolting time [67,93–98]. Incorporating such methods would provide a more refined understanding of the genetic mechanisms underlying important traits, potentially opening new avenues for targeted breeding strategies.

5. Conclusions

This study demonstrates that integrating genotypic data with phenotypic data can be effectively used to develop a core collection of the *Lactuca* germplasm and to identify significant genetic diversity and population structure within it. The identification of three genetically distinct clusters, along with significant phenotypic variations, particularly in leaf length, plant weight, and head height, provides valuable insights for lettuce breeding programs. The moderate genetic differentiation among clusters and high within-population diversity suggest strong gene flow between clusters, further supporting the utility of this core collection for future genome-wide association studies and marker-assisted selection. These findings enhance our understanding of lettuce genetic diversity and will contribute to improved cultivar development and the conservation of genetic resources.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/plants13243552/s1: Figure S1. Population structure of whole 2001 Lactuca accessions. (A) Delta K values for different assumed population numbers (K) in the STRUCTURE analysis. (B) Inferred population structure for K = 3. Individual membership coefficients were sorted within each cluster. Table S1: List of 2001 *Lactuca* germplasm in National Agrobiodiversity Center (NAC). Table S2: Genotyping-by-sequencing statistics summary of 2001 *Lactuca* germplasm. Table S3: Summary statistics of the identified SNPs for each chromosome. Table S4: Phenotype data collected from the 300 core collections. Table S5. Comparison of genetic diversity for qualitative agro-morphological traits between entire and core collection of *Lactuca* germplasm. Table S6: Average cluster values of agricultural traits of lettuce core collection.

Author Contributions: Conceptualization, N.R. and Y.-W.N.; methodology, S.M., O.H. and S.-H.K.; formal analysis, S.M. and Y.L.; investigation, S.M., O.H., H.O., J.Y., H.-C.K., H.-J.W., N.R. and Y.-W.N.; data curation, S.M., O.H., N.R. and Y.-W.N.; writing—original draft preparation, S.M., O.H., N.R. and Y.-W.N.; writing—review and editing, S.M., O.H., N.R. and Y.-W.N.; supervision, N.R. and Y.-W.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Establishment of Utilization System and Advancement in Quality of Resources in Seed-Cluster Central Genebank (Project No. PJ016045) of the National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea.

Data Availability Statement: The SNP data presented in this study are deposited in the National Agricultural Biotechnology Information Center repository under accession number NV-0906.

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

References

1. Bar, F.M.A.; Fatah, N.H.A.; Amen, Y.; Halim, A.F.; Saad, H.-E.A. Genus *Lactuca* (Asteraceae): A comprehensive review. *Rec. Nat. Prod.* **2023**, *17*, 201–231. [CrossRef]
2. Nicolle, C.; Cardinault, N.; Gueux, E.; Jaffrelo, L.; Rock, E.; Mazur, A.; Amouroux, P.; Rémésy, C. Health effect of vegetable-based diet: Lettuce consumption improves cholesterol metabolism and antioxidant status in the rat. *Clin. Nutr.* **2004**, *23*, 605–614. [CrossRef] [PubMed]
3. FAOSTAT. Data/QCL. Available online: https://www.fao.org/faostat/en/ (accessed on 12 July 2024).
4. Beharav, A.; Lewinsohn, D.; Lebeda, A.; Nevo, E. New Wild *Lactuca* Genetic Resources with Resistance Against Bremia lactucae. *Genet. Resour. Crop Evol.* **2006**, *53*, 467–474. [CrossRef]
5. Lebeda, A.; Křístková, E.; Kitner, M.; Mieslerová, B.; Jemelková, M.; Pink, D.A.C. Wild *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. *Eur. J. Plant Pathol.* **2014**, *138*, 597–640. [CrossRef]
6. Ha, J.; Lee, T.; Kim, M.Y.; Oliya, B.K.; Gwag, J.-G.; Lee, Y.-H.; Lee, S.-h. Comprehensive transcriptome analysis of *Lactuca indica*, a traditional medicinal wild plant. *Mol. Breed.* **2017**, *37*, 112. [CrossRef]
7. Oliya, B.K.; Kim, M.Y.; Lee, S.-H. In vitro propagation, lactucin quantification, and antibacterial activity of Indian lettuce (*Lactuca indica* L.). *In Vitro Cell. Dev. Biol.-Plant* **2022**, *58*, 361–371. [CrossRef]
8. Křístková, E.; Doležalová, I.; Lebeda, A.; Vinter, V.; Novotná, A. Description of morphological characters of lettuce (*Lactuca sativa* L.) genetic resources. *Hortic. Sci.* **2008**, *35*, 113–129. [CrossRef]

9. Morărița, S.; Teodorescu, G.; David, I. Research on the Quality of Lettuce Production Grown in Protected Areas. *Ann. "Valahia" Univ. Târgoviște. Agric.* **2023**, *15*, 42–50. [CrossRef]
10. Dufault, R.J.; Ward, B.; Hassell, R.L. Dynamic relationships between field temperatures and romaine lettuce yield and head quality. *Sci. Hortic.* **2009**, *120*, 452–459. [CrossRef]
11. Šuštar-Vozlič, J.; Ugrinović, K.; Maras, M.; Křístková, E.; Lebeda, A.; Meglič, V. Morphological and genetic diversity of Slovene lettuce landrace 'Ljubljanska ledenka' (*Lactuca sativa* L.). *Genet. Resour. Crop Evol.* **2021**, *68*, 185–203. [CrossRef]
12. Lee, J.-S.; Chandra, D.; Son, J. Growth, Physicochemical, Nutritional, and Postharvest Qualities of Leaf Lettuce (*Lactuca sativa* L.) as Affected by Cultivar and Amount of Applied Nutrient Solution. *Horticulturae* **2022**, *8*, 436. [CrossRef]
13. Lin, K.-H.; Huang, M.-Y.; Huang, W.-D.; Hsu, M.-H.; Yang, Z.-W.; Yang, C.-M. The effects of red, blue, and white light-emitting diodes on the growth, development, and edible quality of hydroponically grown lettuce (*Lactuca sativa* L. var. capitata). *Sci. Hortic.* **2013**, *150*, 86–91. [CrossRef]
14. Choi, K.Y.; Paek, K.Y.; Lee, Y.B. Effect of Air Temperature on Tipburn Incidence of Butterhead and Leaf Lettuce in a Plant Factory. In *Transplant Production in the 21st Century, Proceedings of the International Symposium on Transplant Production in Closed System for Solving the Global Issues on Environmental Conservation, Food, Resources and Energy, Chiba, Japan, 28 February–2 March 2000*; Kubota, C., Chun, C., Eds.; Springer: Dordrecht, The Netherlands, 2000; pp. 166–171.
15. Oliya, B.K.; Kim, M.Y.; Lee, S.-H. Development of genic-SSR markers and genetic diversity of Indian lettuce (*Lactuca indica* L.) in South Korea. *Genes Genom.* **2018**, *40*, 615–623. [CrossRef] [PubMed]
16. Riar, D.S.; Rustgi, S.; Burke, I.C.; Gill, K.S.; Yenish, J.P. EST-SSR Development from 5 *Lactuca* Species and Their Use in Studying Genetic Diversity Among *L. serriola* Biotypes. *J. Hered.* **2011**, *102*, 17–28. [CrossRef]
17. Koopman, W.J.M.; Zevenbergen, M.J.; Van den Berg, R.G. Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *Am. J. Bot.* **2001**, *88*, 1881–1887. [CrossRef]
18. Tardin, F.D.; Amaral Júnior, A.T.d.; Pereira, M.G.; Vidigal, M.C.G.; Daher, R.F.; Scapim, C.A. Genetic diversity and determination of the optimum number of RAPD markers in lettuce (*Lactuca sativa* L.). *Acta Sci.-Agron.* **2003**, *25*, 1–5.
19. Kesseli, R.; Ochoa, O.; Michelmore, R. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). *Genome* **1991**, *34*, 430–436. [CrossRef]
20. Jiang, G.-L. Molecular Marker-Assisted Breeding: A Plant Breeder's Review. In *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*; Al-Khayri, J.M., Jain, S.M., Johnson, D.V., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 431–472.
21. Mammadov, J.; Aggarwal, R.; Buyyarapu, R.; Kumpatla, S. SNP Markers and Their Impact on Plant Breeding. *Int. J. Plant Genom.* **2012**, *2012*, 728398. [CrossRef]
22. Rafalski, J.A. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci.* **2002**, *162*, 329–333. [CrossRef]
23. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [CrossRef]
24. Stansell, Z.; Hyma, K.; Fresnedo-Ramírez, J.; Sun, Q.; Mitchell, S.; Björkman, T.; Hua, J. Genotyping-by-sequencing of *Brassica oleracea* vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. *Hortic. Res.* **2018**, *5*, 38. [CrossRef] [PubMed]
25. Li, X.; Kong, C.; Yu, H.; Liu, X.; Fang, Z.; Liu, Y.; Yang, L.; Zhuang, M.; Wang, Y.; Lv, H.; et al. Identification of a major QTL for seed number per silique in cabbage (*Brassica oleracea* L. var. capitata) using genotyping by sequencing. *Euphytica* **2019**, *215*, 133. [CrossRef]
26. Cai, X.; Xu, C.; Wang, X.; Wang, S.; Zhang, Z.; Fei, Z.; Wang, Q. Construction of genetic linkage map using genotyping-by-sequencing and identification of QTLs associated with leaf color in spinach. *Euphytica* **2018**, *214*, 229. [CrossRef]
27. Palumbo, F.; Qi, P.; Pinto, V.B.; Devos, K.M.; Barcaccia, G. Construction of the First SNP-Based Linkage Map Using Genotyping-by-Sequencing and Mapping of the Male-Sterility Gene in Leaf Chicory. *Front. Plant Sci.* **2019**, *10*, 276. [CrossRef]
28. Ma, J.; Shi, A.; Mou, B.; Evans, M.; Clark, J.R.; Motes, D.; Correll, J.C.; Xiong, H.; Qin, J.; Chitwood, J.; et al. Association mapping of leaf traits in spinach (*Spinacia oleracea* L.). *Plant Breed.* **2016**, *135*, 399–404. [CrossRef]
29. Rayaprolu, L.; Deshpande, S.P.; Gupta, R. Genotyping-by-Sequencing (GBS) Method for Accelerating Marker-Assisted Selection (MAS) Program. In *Genomics of Cereal Crops*; Springer: New York, NY, USA, 2022; pp. 245–257.
30. Pootakham, W. Genotyping by Sequencing (GBS) for Genome-Wide SNP Identification in Plants. In *Plant Genotyping: Methods and Protocols*; Shavrukov, Y., Ed.; Springer: New York, NY, USA, 2023; pp. 1–8.
31. Van Hintum, T.J.; Brown, A.H.D.; Spillane, C.; Hodgkin, T. Core collections of plant genetic resources. In *IPGRI Technical Bulletin*; International Plant Genetic Resources Institute: Rome, Italy, 2000; Volume 3, p. 48.
32. Frankel, O.H. Plant genetic resources today: A critical appraisal. In *Crop Genetic Resources: Conservation and Evaluation*; Allen & Unwin: London, UK, 1984; pp. 241–256.
33. Brown, A.H.D. Core collections: A practical approach to genetic resources management. *Genome* **1989**, *31*, 818–824. [CrossRef]
34. Gu, R.; Fan, S.; Wei, S.; Li, J.; Zheng, S.; Liu, G. Developments on Core Collections of Plant Genetic Resources: Do We Know Enough? *Forests* **2023**, *14*, 926. [CrossRef]
35. Wang, M.-m.; Sun, D.-l.; Chen, R.; Yang, Y.-x.; Zhang, G.; Lü, M.-j.; Wang, Q.; Xie, T.-y.; Niu, G.-b.; Shan, X.-z. Construction and evaluation of cauliflower core collection. *Acta Hortic. Sin.* **2023**, *50*, 421–431.

36. Hoshikawa, K.; Lin, Y.-P.; Schafleitner, R.; Shirasawa, K.; Isobe, S.; Nguyen, D.C.; Ohsawa, R.; Yoshioka, Y. Genetic diversity analysis and core collection construction for *Amaranthus tricolor* germplasm based on genome-wide single-nucleotide polymorphisms. *Sci. Hortic.* **2023**, *307*, 111428. [CrossRef]
37. Li, L.; He, W.; Ma, L.; Liu, P.; Xu, H.; Xu, J.; Zheng, X. Construction Chinese cabbage (*Brassica rapa* L.) core collection and its EST-SSR fingerprint database by EST-SSR molecular markers. *Genom. Appl. Biol.* **2009**, *28*, 76–88.
38. Jansen, J.; van Hintum, T. Genetic distance sampling: A novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theor. Appl. Genet.* **2007**, *114*, 421–428. [CrossRef] [PubMed]
39. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef] [PubMed]
40. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **2013**, *22*, 3124–3140. [CrossRef] [PubMed]
41. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
42. Reyes-Chin-Wo, S.; Wang, Z.; Yang, X.; Kozik, A.; Arikit, S.; Song, C.; Xia, L.; Froenicke, L.; Lavelle, D.O.; Truco, M.-J.; et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* **2017**, *8*, 14953. [CrossRef]
43. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
44. De Summa, S.; Malerba, G.; Pinto, R.; Mori, A.; Mijatovic, V.; Tommasi, S. GATK hard filtering: Tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. *BMC Bioinform.* **2017**, *18*, 119. [CrossRef]
45. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
46. Browning, B.L.; Zhou, Y.; Browning, S.R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [CrossRef]
47. De Beukelaer, H.; Davenport, G.F.; Fack, V. Core Hunter 3: Flexible core subset selection. *BMC Bioinform.* **2018**, *19*, 203. [CrossRef]
48. Kim, K.-W.; Chung, H.-K.; Cho, G.-T.; Ma, K.-H.; Chandrabalan, D.; Gwag, J.-G.; Kim, T.-S.; Cho, E.-G.; Park, Y.-J. PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* **2007**, *23*, 2155–2162. [CrossRef] [PubMed]
49. Hennink, S.; Zeven, A.C. The interpretation of Nei and Shannon-Weaver within population variation indices. *Euphytica* **1990**, *51*, 235–240. [CrossRef]
50. Pielou, E.C. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **1966**, *13*, 131–144. [CrossRef]
51. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [CrossRef] [PubMed]
52. Li, Y.-L.; Liu, J.-X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* **2018**, *18*, 176–177. [CrossRef]
53. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
54. Ramasamy, R.K.; Ramasamy, S.; Bindroo, B.B.; Naik, V.G. STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus* **2014**, *3*, 431. [CrossRef]
55. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef]
56. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]
57. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **2024**, *52*, W78–W82. [CrossRef]
58. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef] [PubMed]
59. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024.
60. Liu, Y.; Fu, B.; Zhang, Q.; Cai, J.; Guo, W.; Zhai, W.; Wu, J. Genetic diversity and population structure of wheat landraces in Southern Winter Wheat Region of China. *BMC Genom.* **2024**, *25*, 664. [CrossRef] [PubMed]
61. Wu, X.; Michael, V.N.; López-Hernández, F.; Cortés, A.J.; Morris, J.B.; Wang, M.; Tallury, S.; Miller Ii, M.C.; Blair, M.W. Genetic Diversity and Genome-Wide Association in Cowpeas (*Vigna unguiculata* L. Walp). *Agronomy* **2024**, *14*, 961. [CrossRef]
62. Cuevas, H.E.; Prom, L.K. Association analysis of grain mould resistance in a core collection of NPGS Ethiopian sorghum germplasm. In *Plant Genetic Resources: Characterization and Utilization*; Cambridge University Press: Cambridge, UK, 2024; pp. 1–10. [CrossRef]
63. Bernád, V.; Al-Tamimi, N.; Langan, P.; Gillespie, G.; Dempsey, T.; Henchy, J.; Harty, M.; Ramsay, L.; Houston, K.; Macaulay, M.; et al. Unlocking the genetic diversity and population structure of the newly introduced two-row spring European HerItage Barley collecTion (ExHIBiT). *Front. Plant Sci.* **2024**, *15*, 1268847. [CrossRef]
64. Wu, Y.; San Vicente, F.; Huang, K.; Dhliwayo, T.; Costich, D.E.; Semagn, K.; Sudha, N.; Olsen, M.; Prasanna, B.M.; Zhang, X.; et al. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor. Appl. Genet.* **2016**, *129*, 753–765. [CrossRef]

65. Elbasyoni, I.S.; Lorenz, A.J.; Guttieri, M.; Frels, K.; Baenziger, P.S.; Poland, J.; Akhunov, E. A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci.* **2018**, *270*, 123–130. [CrossRef]
66. Park, S.; Kumar, P.; Shi, A.; Mou, B. Population genetics and genome-wide association studies provide insights into the influence of selective breeding on genetic variation in lettuce. *Plant Genome* **2021**, *14*, e20086. [CrossRef]
67. Lee, O.N.; Fukushima, K.; Park, H.Y.; Kawabata, S. QTL Analysis of Stem Elongation and Flowering Time in Lettuce Using Genotyping-by-Sequencing. *Genes* **2021**, *12*, 947. [CrossRef]
68. Upadhyaya, H.D.; Bramel, P.J.; Singh, S. Development of a Chickpea Core Subset Using Geographic Distribution and Quantitative Traits. *Crop Sci.* **2001**, *41*, 206–210. [CrossRef]
69. Sa, K.J.; Kim, D.M.; Oh, J.S.; Park, H.; Hyun, D.Y.; Lee, S.; Rhee, J.H.; Lee, J.K. Construction of a core collection of native *Perilla* germplasm collected from South Korea based on SSR markers and morphological characteristics. *Sci. Rep.* **2021**, *11*, 23891. [CrossRef]
70. Santos, C.C.d.; Andrade, L.R.B.d.; Carmo, C.D.d.; Oliveira, E.J.d. Development of cassava core collections based on morphological and agronomic traits and SNPS markers. *Front. Plant Sci.* **2023**, *14*, 1250205. [CrossRef] [PubMed]
71. Bhattacharjee, R.; Khairwal, I.S.; Bramel, P.J.; Reddy, K.N. Establishment of a pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection based on geographical distribution and quantitative traits. *Euphytica* **2007**, *155*, 35–45. [CrossRef]
72. Reddy, L.J.; Upadhyaya, H.D.; Gowda, C.L.L.; Singh, S. Development of Core Collection in Pigeonpea [*Cajanus cajan* (L.) Millspaugh] using Geographic and Qualitative Morphological Descriptors. *Genet. Resour. Crop Evol.* **2005**, *52*, 1049–1056. [CrossRef]
73. Li, N.; Yang, Y.; Xu, F.; Chen, X.; Wei, R.; Li, Z.; Pan, W.; Zhang, W. Genetic Diversity and Population Structure Analysis of *Castanopsis hystrix* and Construction of a Core Collection Using Phenotypic Traits and Molecular Markers. *Genes* **2022**, *13*, 2383. [CrossRef]
74. Upadhyaya, H.D.; Ortiz, R.; Bramel, P.J.; Singh, S. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genet. Resour. Crop Evol.* **2003**, *50*, 139–148. [CrossRef]
75. Mahmoodi, R.; Dadpour, M.R.; Hassani, D.; Zeinalabedini, M.; Vendramin, E.; Leslie, C.A. Composite core set construction and diversity analysis of Iranian walnut germplasm using molecular markers and phenotypic traits. *PLoS ONE* **2021**, *16*, e0248623. [CrossRef]
76. Manrique-Carpintero, N.C.; Berdugo-Cely, J.A.; Cerón-Souza, I.; Lasso-Paredes, Z.; Reyes-Herrera, P.H.; Yockteng, R. Defining a diverse core collection of the Colombian Central Collection of potatoes: A tool to advance research and breeding. *Front. Plant Sci.* **2023**, *14*, 1046400. [CrossRef]
77. Egan, L.M.; Conaty, W.C.; Stiller, W.N. Core Collections: Is There Any Value for Cotton Breeding? *Front. Plant Sci.* **2022**, *13*, 895155. [CrossRef]
78. Ghamkhar, K.; Brown, A.H. The Development of Core Collections and Their Role in Association Mapping. In *Plant Genetic Resources for the 21st Century*; Ghamkhar, K., Williams, W.M., Brown, A.H.D., Eds.; Apple Academic Press: New York, NY, USA, 2023; pp. 215–227.
79. McGuire, P.E.; Ryder, E.J.; Michelmore, R.W.; Clark, R.L.; Antle, R.; Emery, G.; Hannan, R.M.; Kesseli, R.V.; Kurtz, E.A.; Ochoa, O.; et al. *Genetic Resources of Lettuce and Lactuca Species in California: An Assessment of the USDA and UC Collections and Recommendations for Long-Term Security*; University of California: Davis, CA, USA, 1993; Report No. 12.
80. Koopman, W.J.M.; Guetta, E.; van de Wiel, C.C.M.; Vosman, B.; van den Berg, R.G. Phylogenetic relationships among Lactuca (Asteraceae) species and related genera based on ITS-1 DNA sequences. *Am. J. Bot.* **1998**, *85*, 1517–1530. [CrossRef]
81. Wei, T.; van Treuren, R.; Liu, X.; Zhang, Z.; Chen, J.; Liu, Y.; Dong, S.; Sun, P.; Yang, T.; Lan, T.; et al. Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* **2021**, *53*, 752–760. [CrossRef]
82. Beharav, A.; Malarz, J.; Michalska, K.; Ben-David, R.; Stojakowska, A. Variation of sesquiterpene lactone contents in *Lactuca altaica* natural populations from Armenia. *Biochem. Syst. Ecol.* **2020**, *90*, 104030. [CrossRef]
83. Palomino, J.; Mejías, J.A. Morphometrics as a robust tool for disambiguation in plant taxonomy: The case of *Lactuca livida*, a commonly accepted but never delimited taxon. *Plant Biosyst.-Int. J. Deal. All Asp. Plant Biol.* **2023**, *157*, 47–60. [CrossRef]
84. Simko, I.; Hu, J. Population Structure in Cultivated Lettuce and Its Impact on Association Mapping. *J. Am. Soc. Hortic. Sci. J. Amer. Soc. Hort. Sci.* **2008**, *133*, 61–68. [CrossRef]
85. Yoshida, R.; Paul, L.; Nesbitt, P. Stochastic Safety Radius on UPGMA. *Algorithms* **2022**, *15*, 483. [CrossRef]
86. Bleidorn, C. Sources of Error and Incongruence in Phylogenomic Analyses. In *Phylogenomics: An Introduction*; Bleidorn, C., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 173–193.
87. Lebeda, A.; Dolezalová, I.; Feráková, V.; Astley, D. Geographical distribution of wild *Lactuca* species (Asteraceae, Lactuceae). *Bot. Rev.* **2004**, *70*, 328–356. [CrossRef]
88. Whitaker, T.W.; Provvidenti, R. Collecting lettuce germplasm in Turkey and Greece. *HortScience* **1983**, *18*, 285–288. [CrossRef]
89. Atsbeha, G.; Tesfaye, K.; Mekonnen, T.; Haileselassie, T.; Kebede, M. Genetic diversity and population structure analysis of bread wheat (*Triticum aestivum* L.) germplasms as revealed by inter simple sequence repeat (ISSR) markers. *Genet. Resour. Crop Evol.* **2024**, *71*, 2721–2735. [CrossRef]
90. Yu, Z.; Fredua-Agyeman, R.; Hwang, S.-F.; Strelkov, S.E. Molecular genetic diversity and population structure analyses of rutabaga accessions from Nordic countries as revealed by single nucleotide polymorphism markers. *BMC Genom.* **2021**, *22*, 442. [CrossRef]

91. Naaz, S.; Pandey, V.; Yadav, H.K. Evaluation of genetic diversity in rice (*Oryza sativa* L. ssp. Indica) accessions using SSR marker. *Vegetos* **2022**, *35*, 961–968. [CrossRef]
92. Wright, S. *Evolution and the Genetics of Populations: The Theory of Gene Frequencies*; Univ. Chicago Press: Chicago, IL, USA, 1969; Volume 2.
93. Kwon, S.; Simko, I.; Hellier, B.; Mou, B.; Hu, J. Genome-wide association of 10 horticultural traits with expressed sequence tag-derived SNP markers in a collection of lettuce lines. *Crop J.* **2013**, *1*, 25–33. [CrossRef]
94. Sthapit Kandel, J.; Peng, H.; Hayes, R.J.; Mou, B.; Simko, I. Genome-wide association mapping reveals loci for shelf life and developmental rate of lettuce. *Theor. Appl. Genet.* **2020**, *133*, 1947–1966. [CrossRef] [PubMed]
95. Simko, I.; Sthapit Kandel, J.; Peng, H.; Zhao, R.; Subbarao, K.V. Genetic determinants of lettuce resistance to drop caused by Sclerotinia minor identified through genome-wide association mapping frequently co-locate with loci regulating anthocyanin content. *Theor. Appl. Genet.* **2023**, *136*, 180. [CrossRef] [PubMed]
96. Zhang, Z.; van Treuren, R.; Yang, T.; Hu, Y.; Zhou, W.; Liu, H.; Wei, T. A comprehensive lettuce variation map reveals the impact of structural variations in agronomic traits. *BMC Genom.* **2023**, *24*, 659. [CrossRef] [PubMed]
97. Simko, I.; Peng, H.; Sthapit Kandel, J.; Zhao, R. Genome-wide association mapping reveals genomic regions frequently associated with lettuce field resistance to downy mildew. *Theor. Appl. Genet.* **2022**, *135*, 2009–2024. [CrossRef]
98. Rosental, L.; Still, D.W.; You, Y.; Hayes, R.J.; Simko, I. Mapping and identification of genetic loci affecting earliness of bolting and flowering in lettuce. *Theor. Appl. Genet.* **2021**, *134*, 3319–3337. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.